Genomic analysis of the marine yeast *Rhodotorula sphaerocarpa* ETNP2018 reveals adaptation to the open ocean

Dylan M. Lane[1], David L. Valentine[2,3] and Xuefeng Peng[1,2]*

Abstract

Background Despite a rising interest in the diversity and ecology of fungi in marine environments, there are few published genomes of fungi isolated from the ocean. The basidiomycetous yeast (unicellular fungus) genus *Rhodotorula* are prevalent and abundant in the open ocean, and they have been isolated from a wide range of other environments. Many of these environments are nutrient poor, such as the Antarctica and the Atacama deserts, raising the question as to how *Rhodotorula* yeasts may have adapted their metabolic strategies to optimize survival under low nutrient conditions. In order to understand their adaptive strategies in the ocean, the genome of *R. sphaerocarpa* ETNP2018 was compared to that of fourteen representative *Rhodotorula* yeasts, isolated from a variety of environments.

Results *Rhodotorula sphaerocarpa* ETNP2018, a strain isolated from the oligotrophic part of the eastern tropical North Pacific (ETNP) oxygen minimum zone (OMZ), hosts the smallest of the fifteen genomes and yet the number of protein-coding genes it possesses is on par with the other strains. Its genome exhibits a distinct reduction in genes dedicated to Major Facilitator Superfamily transporters as well as biosynthetic enzymes. However, its core metabolic pathways are fully conserved. Our research indicates that the selective pressures of the ETNP OMZ favor a streamlined genome with reduced overall biosynthetic potential balanced by a stable set of core metabolisms and an expansion of mechanisms for nutrient acquisition.

Conclusions In summary, this study offers insights into the adaptation of fungi to the oligotrophic ocean and provides valuable information for understanding the ecological roles of fungi in the ocean.

Keywords Marine fungi, *Rhodotorula*, Genome streamlining, Yeast, Comparative genomics, Oxygen minimum zone

Background

First described over a century ago, marine fungi have received increasing recognition for their roles in ocean biogeochemical cycles and microbial food webs [1–4]. While environmental DNA sequencing surveys have uncovered many previously unrecognized lineages of marine fungi [5, 6], understanding of the functional roles of marine fungi has been partially limited by the small number of genomic analyses. Efforts such as the 1000 Fungal Genomes Project have primarily focused on fungi isolated from terrestrial environments [7]. This bias in research efforts partially stems from the view that the

*Correspondence:
Xuefeng Peng
xpeng@seoe.sc.edu
[1] School of Earth, Ocean, and Environment, University of South Carolina, Columbia, SC, USA
[2] Marine Science Institute, University of California, Santa Barbara, CA, USA
[3] Department of Earth Science, University of California, Santa Barbara, CA, USA

Lane *et al. BMC Genomics*   (2023) 24:695

Page 2 of 14

relative low substrate availability in seawater, especially when compared to soil and other types of terrestrial systems, limits the diversity and abundance of marine fungi.

Nutrient scarcity and the intensity of competition in oligotrophic environments is proposed to impact microbial genome size and composition. This theory, known as the streamlining theory, states that microbes recovered from environments of low nutrient availability will tend to have reduced genomes with a low base-pair count, a low intergenic-to-coding DNA ratio, few paralogs or pseudogenes, and a highly conserved set of core metabolisms [8]. Selection in oligotrophic environments is thought to favor efficiency of transport and metabolism as well as a low cost of transcription and replication [8]. To minimize the cost of protein synthesis and increase efficiency, streamlined marine microorganisms are expected to primarily reduce genes for transcriptional regulations [9]. It is supposed that carbon, nitrogen, iron, and sulfur limitations all lead to a reduction in nitrogen-rich protein expression and transition to less costly amino acid synthesis, including a preferential reduction in N-containing side chains and ribosomal proteins [10, 11]. While the streamlining theory has been supported by numerous prokaryotic examples, it is unclear whether marine fungi have made similar adaptations as their prokaryotic counterparts.

We analyzed the genome of the marine yeast *Rhodotorula sphaerocarpa* ETNP2018 isolated from the water column of the eastern tropical North Pacific (ETNP) ocean. A recent metabarcoding survey showed the genus *Rhodotorula* is one of the most prevalent and abundant fungal taxa in the ETNP oxygen minimum zone (OMZ). The *Rhodotorula* strain we analyzed was isolated from the same research expedition as the metabarcoding survey [12]. *Rhodotorula* are basidiomycetous yeasts that are sometimes referred to as "red yeasts" due to their production of β-carotene, responsible for their red, pink, or orange pigmentation [13]. Carotenogenic yeasts are utilized by the biotechnological industry to produce β-carotene as well as other isoprenoid products (γ-carotene, torulene, and torularhodin) [14]. The genus is also considered oleaginous, capable of synthesizing lipids up to one third of their dry weight. They are known to inhabit a wide array of environments, both eutrophic and oligotrophic, including human and plant hosts, benthic marine sediments, Antarctic permafrost, eutrophic freshwater lakes, ocean and river waters, soil, wood pulp, and the International Space Station [15–20]. The species *R. sphaerocarpa* has been isolated from the marine environments such as the Marguerite Bay in the Antarctic Ocean and the coastal waters of Thailand [21, 22]. In this study, we used comparative genomics of the open ocean isolate *R. sphaerocarpa* ETNP2018 and fourteen other *Rhodotorula* strains from diverse environments to test the genome streamlining theory for marine fungi and further searched for evidence of adaptation of marine fungi to the oligotrophic ocean.

Methods

Sampling and isolate information
The *R. sphaerocarpa* ETNP2018 strain was isolated from seawater in the eastern tropical North Pacific (ETNP) oxygen minimum zone (OMZ) [12]. Seawater from 50 m depth at Station 1 (10°N 113°W) was collected onboard R/V Sally Ride using 30-L Niskin bottles and inoculated into aerobic artificial seawater [23] medium supplemented with cellobiose (2 g L^{-1}) and antibiotics (200 mg L^{-1} penicillin and streptomycin). The initial batch of enrichment cultures that showed growth were inoculated on agar plates. Axenic cultures were obtained after five passages from the first batch of agar plates. Routine maintenance of the culture was performed in 10-ml of medium at 20 °C without shaking. The strain is accessible in the Agriculture Research Service Culture Collection (Northern Regional Research Laboratory, NRRL 64474).

Genome sequencing and assembly
Genomic DNA from 11-day cultures of *R. sphaerocarpa* ETNP2018 (OD = 0.5) was extracted using QIAGEN DNeasy Plant Mini Kit (Cat. No. 69104) following the manufacturer's instructions. The DNA quality was checked by a NanoDrop 2000c spectrophotometer and by a Agilent TapeStation 4150 system. The DNA quantity was measured using Qubit dsDNA BR Assay Kit (Invitrogen). The DNA sample was stored at -20 °C. Sequencing libraries were constructed using the TruSeq DNA PCR Free kit (Illumina # 20015962). The draft genome of *R. sphaerocarpa* ETNP2018 was sequenced at the University of California Davis Genome Center on an Illumina HiSeq 4000 platform with 150 bp paired-end sequencing. Raw reads were filtered with the BBDuk tool in the BBMap software package (v38.73) [24] using the settings "ktrim = r ordered minlen = 51 minlenfraction = 0.33 mink = 11 tbo tpe rcomp = f k = 23 ftm = 5". Adapters were trimmed from the BBDuk-filtered reads using the tool Trimmomatic v0.39 [25] with the settings "ILLUMINACLIP:$adapters:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:100". De novo assembly of reads that passed both quality filtering and adapter trimming was performed for each individual sample using SPAdes v3.15.2 [26] with kmer lengths 21, 33, 55, 77, 99, and 127.

Lane *et al. BMC Genomics*   (2023) 24:695

Page 3 of 14

Phylogenomic reconstruction of Rhodotorula genomes

For phylogenomic analysis, all *Rhodotorula* genome assemblies from the National Center for Biotechnology Information's (NCBI) Assembly Database were retrieved. The genome of *Rhodotorula sp.* FNED7-22 was excluded due to its exceptionally low quality: it contains 2,892 scaffolds with a scaffold N50 value of 5.9 kb. Five metagenome-assembled *Rhodotorula* genomes were excluded also due to similar quality issues (large number of scaffolds and small scaffold N50 values). This leaves 167 *Rhodotorula* genomes for phylogenomic analysis, as well as two genomes (*Microbotrium intermedium* GCA 900096595.1 and *Leucosporidium creatinivorum* GCA 002105055.1) selected as part of the outgroup. Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.4.4 was used to retrieve single-copy orthologues from all 169 genomes [27, 28]. A total of 983 single-copy orthologues were present in all *Rhodotorula* genomes. They were aligned individually using MUSCLE v5.1 [29] and trimmed with trimAl v1.2 with the parameters "-gt 0.6 -w 3 -st 0.001" [30]. The dataset was concatenated with the perl script "catfasta2phyml.pl" which generated an accompanying partition file (https://github.com/nylander/). IQ-TREE v1.6.12 was used to find the best model for each partition and infer a maximum-likelihood tree with the parameters "-m MFP + MERGE -rcluster 10 -bb 1000 -alrt 1000" [31, 32]. All partitions shared the same set of branch lengths but are allowed to have its own evolution rate [33]. The tree was visualized using Interactive Tree Of Life v6 [34].

Structural and functional annotations

For genome annotations and metabolic reconstruction, 14 *Rhodotorula* genomes were selected so as to capture a broad range of native habitats, as well as different species (Table S1) [16, 35–44]. BBMap was used to determine each genome's GC content, scaffold and contiguous sequence (contig) counts, scaffold and contig N50 values, and total length [24]. Direct statistical comparison between overall averages and individual species was performed using directional one-sample t-tests at 95% confidence. Statistical analysis of differences between source categories was performed using two sample t-tests assuming unequal variance at 95% confidence.

Both structural and functional annotations were facilitated by Funannotate v1.8.14, a program for prediction, annotation, and comparison of eukaryotic genomes [45]. Each genome assembly was first soft-masked using tantan [46] and contigs < 1500 bp in length were excluded from downstream annotations. Structural annotations from BRAKER2 were generated in EPmode with OrthoDB v10 and all existing protein sequences from *Rhodotorula* genomes retrieved from the Joint Genome Institute Mycocosm as protein evidence, using ProHint as the alignment tool [7, 28, 47–49]. Because BRAKER2 utilizes GeneMark-EP +, the structural annotations from BRAKER2 are fed into the "Funannotate predict" pipeline which also uses AUGUSTUS v3.3.3 (with at least 500 training models), GlimmerHMM, and SNAP [50–52]. The SWISS-PROT protein knowledgebase, as well as existing protein sequences from *Rhodotorula* genomes retrieved from the Joint Genome Institute Mycocosm, was supplied as protein evidence to the "Funannotate predict" pipeline [7, 53]. EVidenceModeler was used to generate a weighted consensus of structural annotations from all four tools [54]. The program tRNAscan-SE was used to detect tRNA-coding genes [55].

Predicted protein sequences from structural annotations were annotated by BlastKOALA against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [56, 57], by InterProScan v5.51–85.0 [58], and by the eggnog-mapper v2.19 against the eggnog 5.0 database [59, 60]. Carbohydrate-active enzymes were identified using the dbCAN2 meta server against the Carbohydrate Enzyme (CAZy) database [61, 62]. The dbCAN2 pipeline uses three tools (DIAMOND, HMMER, and Hotpep) and we only kept annotations supported by at least two tools to be conservative and confident. To determine potential function of all well-annotated CAZymes, the CAZy family number of each reported CAZyme was compared against its entry on the online CAZypedia index [63]. Biosynthetic gene clusters (BGCs) were identified by the Antibiotics and Secondary Metabolite Analysis Shell (AntiSMASH) v6.0 using the fungal version (fungiSMASH) [64]. BGC sequences identified by AntiSMASH were submitted to the NCBI Conserved Domain Database for refined annotation of BGC genes with the location of conserved domain footprints and functional sites inferred from these footprints [65, 66].

Metabolic reconstruction

Functional annotations were manually inspected to determine key metabolic pathways conserved in the *R. sphaerocarpa* ETNP2018 genome. Metabolic pathways were reconstructed based on the presence of key enzymes using the KEGG mapper, the *Saccharomyces* Genome Database (SGD), and published literature on budding yeasts [67–69]. Cofactors were determined using the SGD, and localization was determined using the annotations given by BlastKOALA or from the literature. This information was then used to create a custom graphic depicting the intercellular transport and localized carbohydrate metabolic pathways contained within the genome of *R. sphaerocarpa* ETNP2018.

Results and discussion

Genomic characteristics

The *R. sphaerocarpa* ETNP2018 genome was assembled into 115 scaffolds, with a total size of 17.7 Mbp, and an N50 value of 377,844 (Table 1). With 6,451 gene models, the gene density is 364 genes/Mbp, on par with previously published genomes of marine fungi [70]. BUSCO estimated that the genome is 97.3% complete. Of the 1764 BUSCOs from the OrthoDB v10 database for Basidiomycota, 1713 are present as single copies, 3 as duplicated copies, and 13 as fragmented in the *R. sphaerocarpa* ETNP2018 genome. There are 5,324 eukaryotic cluster of orthologs (KOGs), 7,872 protein family (Pfam) domains, and 137 tRNAs in the assembly. A total of 3,210 (49.8%) genes were annotated by KEGG orthology, as well as 172 CAZymes. The average genome size of all 15 *Rhodotorula* strains is 20.3 ± 1.6 Mbp and the average number of proteins per genome is $3,275 \pm 154$ (Table 1). The 15 *Rhodotorula* genomes included an average of $5,610 \pm 360$ KOG assignments, $8,240 \pm 418$ Pfam domains, and 194 ± 19 CAZymes. Marine yeasts isolated from seawater including *R. sphaerocarpa* ETNP2018, *R. sphaerocarpa* GDMCC 60679, *R. diobovata* 08–225, and *R. mucilaginosa* CYJ03 have an average genome size of 19.0 ± 1.6 Mbp with $3,203 \pm 99$ annotated proteins (Table 1).

A one sample t-test found the difference in genomic size between *R. sphaerocarpa* ETNP2018 and the average of all strains to be significant ($p = 1.059 \times 10^{-5}$). *R. sphaerocarpa* ETNP2018's genome is the smallest of the marine strains; however, a one sample t-test found this difference to be insignificant ($p = 0.099$). The difference was significant when compared with the mean genome size of freshwater, endophytic, and terrestrial strains ($p = 0.04$, 0.039, and 0.0008, respectively). In contrast, the number of genes with functional annotations was not different between *R. sphaerocarpa* ETNP2018 and the average of all strains, as well as marine, freshwater, endophytic, and terrestrial source category averages ($p = 0.06$, 0.44, 0.12, 0.57, and 0.24, respectively).

The genome size of *R. sphaerocarpa* ETNP2018 was 10.8% smaller than the average genome size of five representative terrestrial *Rhodotorula* strains (Table 2). This reduction of 10.8% was the largest amongst the marine strains of *Rhodotorula*, prompting the hypothesis that the environmental pressures imparted at the OMZ favor fungal strains with smaller genomes in comparison to other marine environments. This hypothesis can be tested in future studies once additional fungal genomes from the open ocean become available. A metabarcoding survey showed that *Rhodotorula* was one of the most abundant and prevalent genera of the fungal communities in the ETNP [12], indicating that *Rhodotorula* are active in the oligotrophic ocean despite their reduced genome size.

The degree of genome reduction we observed in marine *Rhodotorula* strains was similar to that in the marine bacterium *Pelagibacter ubique* HTCC1062, a well-documented example of a streamlined microbe (Table 2) [71]. However, the genome reduction by the heterotrophic yeast and bacteria was not as large as the reduction (38%) found in the marine chemoautotrophic archaea *Nitrosopumilus maritimus* SCM1 when it was compared to five terrestrial strains of ammonia-oxidizing archaea (Table 2; Table S2).

The number of KOGs related to translation and ribosomal biogenesis, the transport and metabolism of amino acids, carbohydrates, lipids, secondary metabolites, and coenzymes were lower in the genome of *R. sphaerocarpa* ETNP2018 than the other 14 *Rhodotorula* strains (Tables S3 and S4) [12]. Nevertheless, *R. sphaerocarpa* ETNP2018 contained the core set of protein-coding genes despite nutrient scarcity and its small genome (Figure S1), which is consistent with streamlining in response to nutrient deprivation [8]. Previously conducted studies on streamlined microbes have found that the average genome size was the smallest for microorganisms isolated from oligotrophic seawater and the largest for those isolated from soil [8]. Microorganisms isolated from freshwater exhibit a broad spectrum of genome sizes [8]. We find *Rhodotorula* genomes from both ends of the size spectrum consistent with this theory. Among the 15 *Rhodotorula* strains, *R. glutinis* ZHK and *R. kratochvilovae* YM25235 were the largest in genome size, and they were isolated from eutrophic environments (the Pearl River and Chenghai Lake, respectively) [15, 18]. The genomes of two soil strains, *R. frigidialcoholis* JG-1b and *R. sp.* CCFEE5036, were smaller than the average genome size of all 15 strains. However, both of these strains were isolated from permafrost and hyper-arid soil in Antarctica's McMurdo Dry Valley [37, 39]. Their reduced genome size could be related to the extreme conditions of their environment with low nutrient availability.

Phylogeny

Phylogenomic analysis revealed multiple monophyletic clades at the level of species, including *R. sphaerocarpa*, *R. paludigena*, *R. kratochvilovae*, and *R. toruloides* (Fig. 1 and Figure S2). *R. sphaerocarpa* ETNP2018 was closely related to the *R. sphaerocarpa* strain isolated from mariculture seawater in Maoming, Guangdong, China. Together, these two *R. sphaerocarpa* genomes form a clade distinct from all other *Rhodotorula* species, demonstrating the evolutionary pressure that likely contributed to the speciation of *R. sphaerocarpa*, which has been primarily isolated from seawater [22, 72–74]. Among the *Rhodotorula*, some strains of the same species have been isolated from drastically different environments (Fig. 1).

Lane *et al. BMC Genomics* (2023) 24:695

Page 5 of 14

Table 1 General genomic characteristics of *R. sphaerocarpa* ETNP2018 and fourteen representative *Rhodotorula* strains. "I.S.S." represents International Space Station

Source	Species	Strain	Total Length (Mbp)	GC (%)	Scaffolds	Scaffold N50 (bp)	Contigs	Contig N50 (bp)	Annotated Proteins	Gene Models	tRNA
Seawater	*R. sphaerocarpa*	ETNP2018	17.7	63.4	115	377,844	120	356,447	3210	6451	137
Seawater	*R. sphaerocarpa*	GDMCC 60679	18.0	63.0	32	1,074,774	32	1,074,774	3213	6310	316
Seawater	*R. diobovata*	08–225	21.1	67.0	361	118,648	678	82,556	3315	7741	183
Seawater	*R. mucilaginosa*	CYJ03	19.1	60.5	88	420,192	88	420,192	3073	6620	137
Marine Sediment	*R. paludigena*	P4R5	21.0	64.3	277	180,700	290	171,007	3406	7430	127
Freshwater	*R. glutinis*	ZHK	22.3	67.8	30	1,466,672	49	963,562	3376	7569	154
Freshwater	*R. kratochvilovae*	YM25235	23.7	67.3	46	1,067,950	46	1,067,950	3581	8224	294
Acid Mine Drainage	*R. taiwanensis*	MD1149	19.6	61.7	181	388,693	227	345,821	3202	6961	113
Endophytic	*R. kratochvilovae*	Y14	22.0	67.5	46	1,029,848	46	1,029,848	3465	7958	194
Endophytic	*R. graminis*	WP1	21.0	67.8	26	1,420,730	322	167,431	3213	6875	152
Terrestrial (I.S.S.)	*R. mucilaginosa*	F6_4S_B_2B	20.2	60.6	199	432,962	222	353,031	3261	7040	124
Terrestrial	*R. toruloides*	NBRC 0880	20.7	61.8	30	1,390,799	30	1,390,799	3429	8284	128
Terrestrial	*R. mucilaginosa*	B1	20.0	60.6	225	256,462	225	256,462	3256	7052	113
Terrestrial	*R. frigidialcoholis*	JG-1b	19.4	67.0	156	301,937	171	280,417	3036	6396	115
Terrestrial	*Rhodotorula sp.*	CCFEE5036	19.1	60.6	155	337,802	201	256,314	3096	6624	120
Average of all Strains			20.3	64	131	684,401	183	547,774	3275	7169	160
Stdev. of all Strains			1.6	3	103	493,511	168	425,961	154	662	64
Average of seawater strains			19	63.5	149	497,865	230	483,492	3,203	6781	193
Stdev. Of seawater strains			1.5	2.7	145	407,049	301	420,520	99	653	85

Lane *et al. BMC Genomics*   (2023) 24:695

Page 6 of 14

Table 2 The percent reduction in the genomic size of *Rhodotorula* strains isolated from seawater, compared with the average of the representative terrestrial *Rhodotorula*. *Pelagibacter ubique* HTCC1062 and *Nitrosopumilis matrimis* SCM1 were chosen as well-known examples of streamlined marine microbes for comparison. *P. ubique* HTCC1062 was compared to five terrestrial strains of the alphaproteobacterial genus *Rickettsia* and *N. matrimis* SCM1 to five representative ammonia-oxidizing archaea, all retrieved from the NCBI assembly database (Table S2)

Species	Strain	Genome Size (bp)	Terrestrial Average (bp)	% Reduction
R. sphaerocarpa	ETNP2018	17,716,787	19,867,559	10.8
R. sphaerocarpa	GDMCC 60679	18,031,004	19,867,559	9.2
R. diobovata	08-225	21,142,622	19,867,559	no reduction
R. mucilaginosa	CYJ03	19,073,214	19,867,559	4.0
Pelagibacter ubique	HTCC1062	1,308,759	1,413,279	7.4
Nitrosopumilus maritimus	SCM1	1,645,259	2,670,916	38.4

This could be attributed to the ability of *Rhodotorula* yeasts to adapt to a diverse range of environmental conditions [14, 39]. *R. mucilaginosa*, for example, has been isolated from soil, both animal and plant microbiomes, industrial mineral deposits, the International Space Station, and the marine water column [19, 43]. Our phylogenomic reconstruction shows that the recently described *R. frigidialcoholis* JG-1b [20] is closely related to *R. mucilaginosa* (Fig. 1). The sister clade of the main *R. mucilaginosa* clade includes *R. frigidialcoholis* JG-1b, two genomes of undescribed species, and four *R. mucilaginosa* genomes, so it is possible that these four *R. mucilaginosa* genomes obtained from the international space station may have been misclassified and are actually *R. frigidialcoholis*.
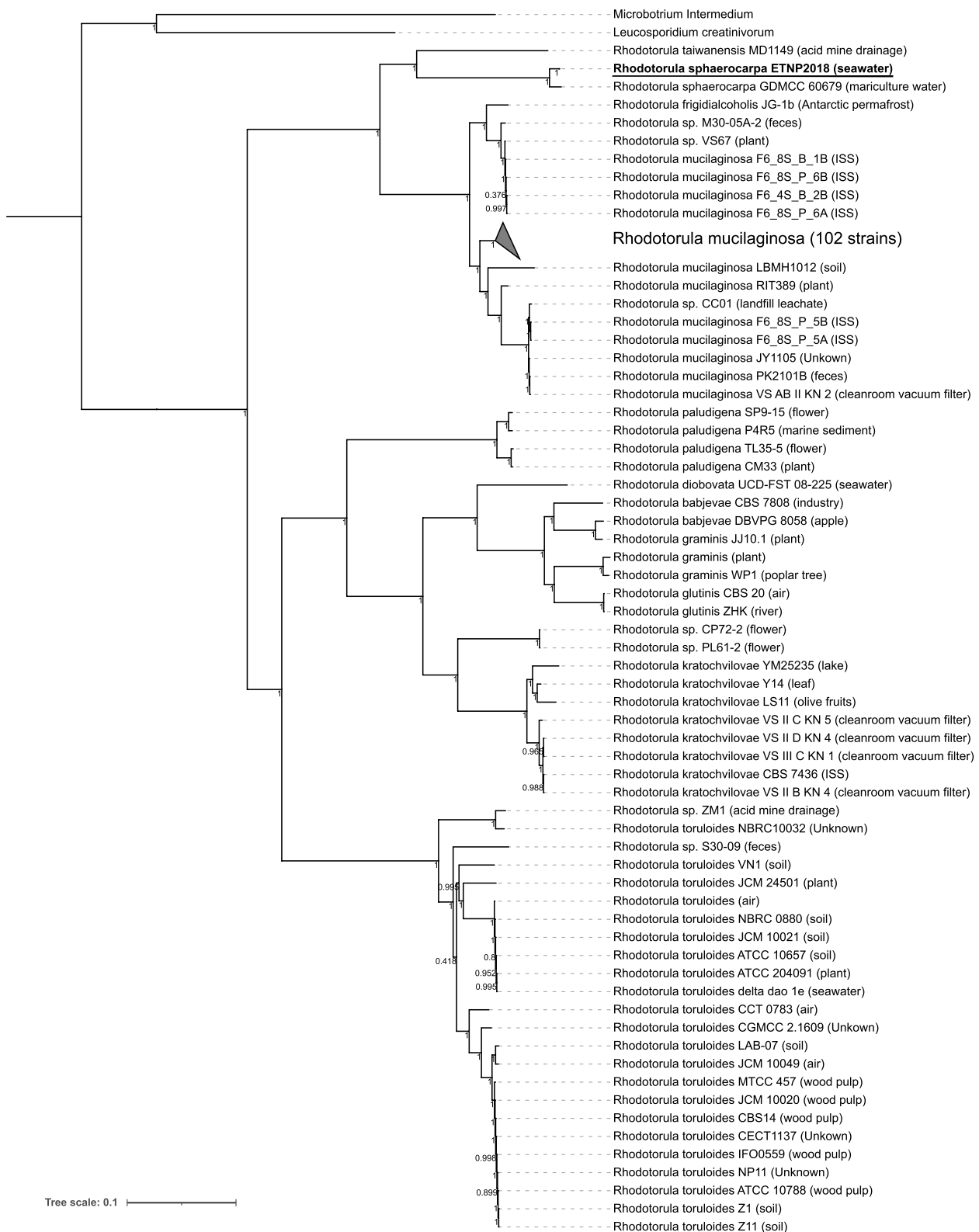
CAZymes

CAZymes identified in the *R. sphaerocarpa* ETNP2018 genome include 56 glycoside hydrolases (GH), 84 glycosyltransferases (GT), 17 related to auxiliary activities (AA), 3 carbohydrate binding modules (CBM), 8 carbohydrate esterases (CE), and 4 polysaccharide lyases (PL) (Table S5). Chitinase (GH18), xyloglucanase (GH16), β-hexosaminidase (GH20), both α- and β-glucosidase (GH3/31), invertase (GH32), α,α-trehalase (GH37), α-mannosidase (GH38/47), and cellulase (GH5) glycoside hydrolase CAZymes are all conserved across the *Rhodotorula* genus (Fig. 2).

Therefore, *Rhodotorula* yeasts have the potential to digest chitin, xylan, hexoses, trehalose, some mannose, and cellulose [63]. Trehalase and glycogen debranching CAZymes (GH13) are also present in all 15 representative members of the *Rhodotorula* genus, suggesting the ability to use both storage polysaccharides to maintain energy production in response to potential carbon deprivation (Fig. 2). The most prevalent and abundant glycosyl hydrolase families include GH5, GH16, and GH18 (Figure S3).
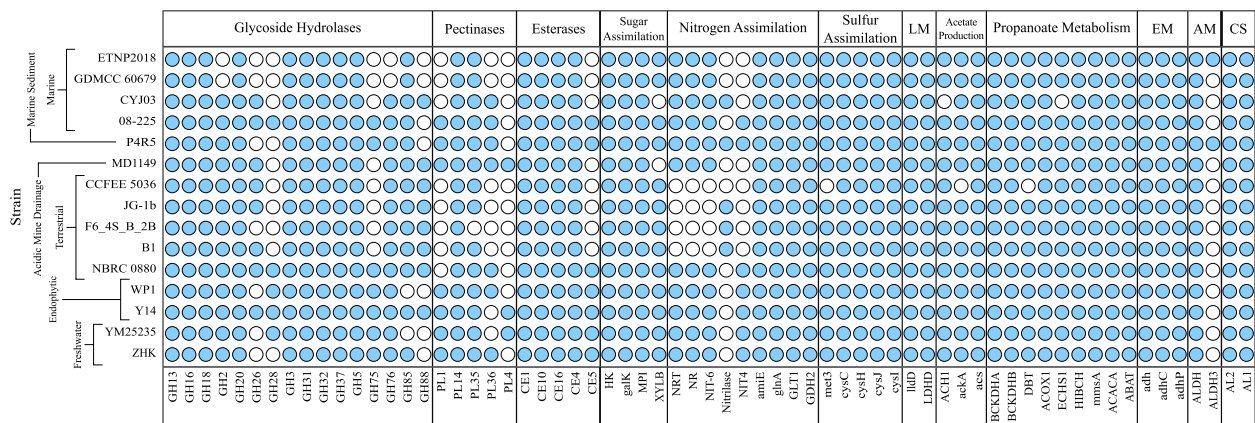
Chitin is the most abundant biopolymer found in the marine environments and thus an important source of carbon and nitrogen for marine microbes [75]. Chitin is produced throughout the water column by fungi, protists, and crustaceans, yet is utilized so rapidly that it is present only in trace concentrations in marine sediments. The importance of chitin as a source of nutrient for marine fungi is relatively understudied however, as most marine chitin degradation is attributed to bacteria [75]. Chitinase is conserved throughout the *Rhodotorula* genus. GH18 was the most prevalent in genomes of freshwater strains (7 ± 1.4) and the least in genomes of marine strains (4 ± 1.4). Chitin degradation via chitinase results in a wide array of oligomers including diacetylchitobiose. The endo-β-N-acetylglucosaminidase (GH85, EC: 3.2.1.96), which degrades diacetylchitobiose into monomeric residues of β-1,4-N-acetyl-D-glucosamine (GlcNAc) and subsequently soluble sugars and dissolved

(See figure on next page.)

Fig. 1 A phylogenomic tree constructed using single-copy orthologues shared amongst 168 *Rhodotorula* strains (as well as two outgroup species, *Microbotrium intermedium* GCA 900096595.1 and *Leucosporidium creatinivorum* GCA 002105055.1). 983 single-copy orthologs were retrieved for comparison using BUSCO v5.4.4. The isolation source is denoted in the parentheses after the strain names ("ISS" = International Space Station). Bootstrap values (*n* = 1000) are displayed at each node. The collapsed clade includes 102 strains of *Rhodotorula mucilaginosa* primarily from the International Space Station. The two side lengths of the triangle representing the collapsed clade are proportional to the distances to the node's closest and farthest child leaves. *R. sphaerocarpa* ETNP2018 is highlighted by bold font and an underline. The same tree that displays the 102 collapsed *Rhodotorula mucilaginosa* strains is included as Figure S2

Lane *et al. BMC Genomics*   (2023) 24:695

Page 7 of 14

Microbotrium Intermedium

Leucosporidium creatinivorum

Rhodotorula taiwanensis MD1149 (acid mine drainage)

Rhodotorula sphaerocarpa ETNP2018 (seawater)

Rhodotorula sphaerocarpa GDMCC 60679 (mariculture water)

Rhodotorula frigidialcoholis JG-1b (Antarctic permafrost)

Rhodotorula sp. M30-05A-2 (feces)

Rhodotorula sp. VS67 (plant)

Rhodotorula mucilaginosa F6_8S_B_1B (ISS)

Rhodotorula mucilaginosa F6_8S_P_6B (ISS)

Rhodotorula mucilaginosa F6_4S_B_2B (ISS)

Rhodotorula mucilaginosa F6_8S_P_6A (ISS)

Rhodotorula mucilaginosa (102 strains)

Rhodotorula mucilaginosa LBMH1012 (soil)

Rhodotorula mucilaginosa RIT389 (plant)

Rhodotorula sp. CC01 (landfill leachate)

Rhodotorula mucilaginosa F6_8S_P_5B (ISS)

Rhodotorula mucilaginosa F6_8S_P_5A (ISS)

Rhodotorula mucilaginosa JY1105 (Unkown)

Rhodotorula mucilaginosa PK2101B (feces)

Rhodotorula mucilaginosa VS AB II KN 2 (cleanroom vacuum filter)

Rhodotorula paludigena SP9-15 (flower)

Rhodotorula paludigena P4R5 (marine sediment)

Rhodotorula paludigena TL35-5 (flower)

Rhodotorula paludigena CM33 (plant)

Rhodotorula diobovata UCD-FST 08-225 (seawater)

Rhodotorula babjevae CBS 7808 (industry)

Rhodotorula babjevae DBVPG 8058 (apple)

Rhodotorula graminis JJ10.1 (plant)

Rhodotorula graminis (plant)

Rhodotorula graminis WP1 (poplar tree)

Rhodotorula glutinis CBS 20 (air)

Rhodotorula glutinis ZHK (river)

Rhodotorula sp. CP72-2 (flower)

Rhodotorula sp. PL61-2 (flower)

Rhodotorula kratochvilovae YM25235 (lake)

Rhodotorula kratochvilovae Y14 (leaf)

Rhodotorula kratochvilovae LS11 (olive fruits)

Rhodotorula kratochvilovae VS II C KN 5 (cleanroom vacuum filter)

Rhodotorula kratochvilovae VS II D KN 4 (cleanroom vacuum filter)

Rhodotorula kratochvilovae VS III C KN 1 (cleanroom vacuum filter)

Rhodotorula kratochvilovae CBS 7436 (ISS)

Rhodotorula kratochvilovae VS II B KN 4 (cleanroom vacuum filter)

Rhodotorula sp. ZM1 (acid mine drainage)

Rhodotorula toruloides NBRC10032 (Unknown)

Rhodotorula sp. S30-09 (feces)

Rhodotorula toruloides VN1 (soil)

Rhodotorula toruloides JCM 24501 (plant)

Rhodotorula toruloides (air)

Rhodotorula toruloides NBRC 0880 (soil)

Rhodotorula toruloides JCM 10021 (soil)

Rhodotorula toruloides ATCC 10657 (soil)

Rhodotorula toruloides ATCC 204091 (plant)

Rhodotorula toruloides delta dao 1e (seawater)

Rhodotorula toruloides CCT 0783 (air)

Rhodotorula toruloides CGMCC 2.1609 (Unkown)

Rhodotorula toruloides LAB-07 (soil)

Rhodotorula toruloides JCM 10049 (air)

Rhodotorula toruloides MTCC 457 (wood pulp)

Rhodotorula toruloides JCM 10020 (wood pulp)

Rhodotorula toruloides CBS14 (wood pulp)

Rhodotorula toruloides CECT1137 (Unkown)

Rhodotorula toruloides IFO0559 (wood pulp)

Rhodotorula toruloides NP11 (Unkown)

Rhodotorula toruloides ATCC 10788 (wood pulp)

Rhodotorula toruloides Z1 (soil)

Rhodotorula toruloides Z11 (soil)

Tree scale: 0.1

Fig. 1 (See legend on previous page.)

Lane *et al. BMC Genomics*   (2023) 24:695

Page 8 of 14

Fig. 2 Heatmap of major genes involved in carbohydrate, nitrogen, and sulfur metabolisms in the *Rhodotorula* genome. Filled circle (light blue) indicate the presence of a gene whereas empty circles indicate its absence. Abbreviations for the headers are as follows: LM, Lactate Metabolism; EM, Ethanol Metabolism; AM, Aldehyde Metabolism; CS, Carotenoid Synthesis. The described functions of glycoside hydrolases (GH), pectinases (PL) and esterases (CE) are in Table S6. Abbreviations for the gene names are as follows: HK, hexokinase; galK, galactokinase; MPI, mannose-6-phosphate isomerase; XYLB, xylulokinase; NRT, nitrate/nitrite transporter; NR, nitrate reductase (NAD(P)H); NIT-6, nitrite reductase (NAD(P)H); NIT4, nitrate assimilation; amiE, aliphatic amidase; glnA, glutamine synthetase; GLT1, glutamate synthase (NADH); GDH2, glutamate dehydrogenase; met3, sulfate adenylyltransferase; cysC, adenylylsulfate kinase; cysH, phosphoadenosine phosphosulfate reductase; cysJ, sulfite reductase (NADPH) flavoprotein alpha-component; cysI, sulfite reductase (NADPH) hemoprotein beta-component; lldD, L-lactate dehydrogenase; LDHD, D-lactate dehydrogenase; ACH1, acetyl-CoA hydrolase; ackA, acetate kinase; acs, acetyl-CoA synthetase; BCKDHA, 2-oxoisovalerate dehydrogenase alpha subunit; BCKDHB, 2-oxoisovalerate dehydrogenase beta subunit; DBT, dihydrolipoyl transacylase; ACOX1 acyl-CoA oxidase; ECHS1, enoyl-CoA hydratase; HIBCH, 3-hydroxyisobutyryl-CoA hydrolase; mmsA, acetyl-CoA carboxylase; ACACA, malonate-semialdehyde dehydrogenase; ABAT, 4-aminobutyrate aminotransferase; adh, alcohol dehydrogenase (NADP +); adhC, alcohol dehydrogenase; adhP, alcohol dehydrogenase, propanol-preferring; ALDH, aldehyde dehydrogenase (NAD +); ALDH3, aldehyde dehydrogenase (NAD(P) +); AL2, 15-cis-phytoene synthase / lycopene beta-cyclase; AL1, phytoene desaturase

organic nitrogen, was present in all *Rhodotorula* genomes except for *R. graminis* WP1 and *R. kratochvilovae* YM25235 (Fig. 2) [75, 76].

GH2, a β-mannosidase, and GH76, an α-1,6-mannanase, are absent exclusively in the two genomes of *R. sphaerocarpa* strains isolated from seawater (Fig. 2). Polygalacturonase (GH28), β-glucoronyl hydrolase (GH88), chitonsanase (GH75), and β-mannanase (GH26) CAZymes are absent in *R. sphaerocarpa* ETNP2018 as well as several other strains (Fig. 2). Mannanase and mannosidase allow yeast to ferment mannose to ethanol [77]; reduction in mannose-hydrolyzing CAZymes in the *R. sphaerocarpa* ETNP2018 genome suggests that it is not a commonly utilized substrate for the strain. Mannans are typically found in plant vacuoles and the endosperm of seeds, as well as the cell walls of certain yeasts [77]. These sources place open ocean yeast such as *R. sphaerocarpa* ETNP2018 far from a consistent supply of mannans, suggesting that its reduction in mannanase and mannosidase CAZymes is a response to the low encounter frequency for the substrate [77, 78].

A recent study demonstrated a positive correlation between a fungal strain's repertoire of CAZymes and its saprophytic tendencies [79]. This suggests that strains which encode few CAZymes, such as those isolated from the water column, have less saprophytic tendencies and encounter fewer carbohydrates than those with higher counts, such as freshwater or endophytic strains. Fungi are the dominant detritovores in eutrophic freshwater ecosystems, and endophytic fungi are reported to opportunistically utilize saprophytic feeding mechanisms after the death of their host plant [80, 81]. The availability of organic matter in these environments makes the synthesis of many different CAZymes more energetically favorable in comparison with the oligotrophic open ocean. Given the low availability of organic matter in the open ocean water column, *R. sphaerocarpa* ETNP2018 likely streamlined its genome to reduce unnecessary and biosynthetically expensive CAZymes.

Central carbon metabolisms

The Embden-Meyerhof-Parnas pathway, Tricarboxylic acid (TCA) cycle, glyoxylate cycle, and pentose phosphate pathway were present in their entirety in the *R. sphaerocarpa* ETNP2018 genome (Figure S1). Potential substrates for these pathways include glucose, the cell's preferred substrate, as well as acetate, ethanol, D-lactate, L-glutamine, and oxaloacetate (Fig. 3). The glyoxylate cycle, a secondary shunt the TCA cycle localized in the peroxisome, utilizes isocitrate lyase (EC: 4.1.3.1) to catalyze the conversion of isocitrate to glyoxylate as well as succinate to malate without requiring the energy

Lane *et al. BMC Genomics*   (2023) 24:695

Page 9 of 14
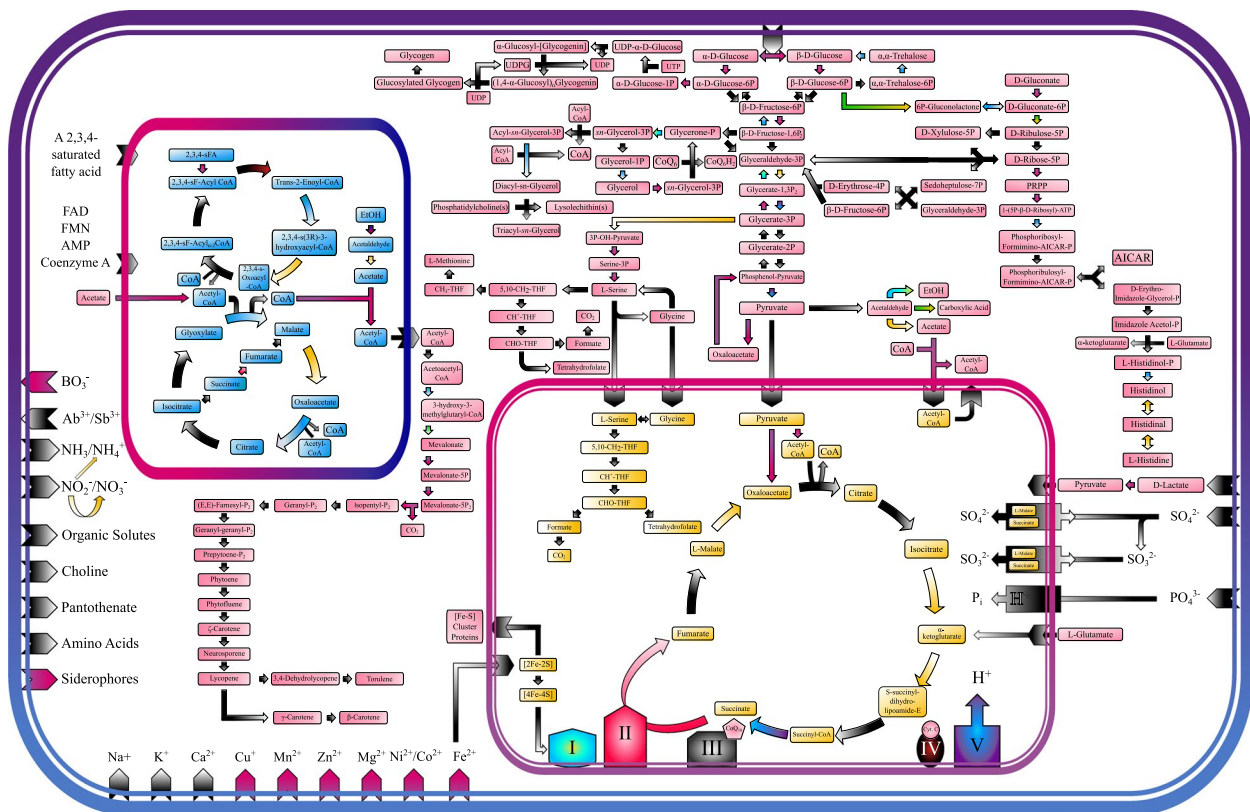
Fig. 3 Reconstructed metabolic pathways for *Rhodotorula sphaerocarpa* ETNP2018. The outer bi-layered membrane represents plasma membrane surrounding the cytosol, the inner bi-layered membrane the mitochondrion, and the inner single-layered membrane the peroxisome; shapes embedded into membranes represent transport proteins. Arrow and transport protein color gradients represent requirement for a specific cofactor in reaction or transport, with the second color of the gradient indicating preferred direction: Black/White, no cofactor required; Light Blue/White, H_2O; Purple/Pink, ATP; Purple/Blue, ADP; Yellow/White, NAD^+; Teal/Yellow, NADH; Green/Yellow, $NADP^+$; Green/White, NADPH; Red/Pink, FAD^{2+}; Black/Red, O_2; Purple/Red, H_2O_2. Abbreviations are as follows: EtOH, Ethanol; CoA, Coenzyme A; THF, tetrahydrofolate; P_i, Inorganic Phosphate; AICAR, 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranoside; PRPP, phosphoribosyl diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; UDP, uridine diphosphate; UTP, uridine triphosphate; UDPG, UDP-α -D-Glucose; sFA, Saturated Fatty Acid; FAD, Flavin Adenine Dinucleotide; FMN, Flavin Mononucleotide; CoQ_6, Ubiquinone-6; CoQ_6H_2, Ubiquinol-6; CoQ_{10}, Ubiquinone-10; Cyt. C, Cytochrome C

intensive decarboxylation steps required to form S-succi-nyl-dihydrolipoamide-E from isocitrate during the TCA cycle (Fig. 2) [82]. Glyoxylate cycle genes in yeast have been shown to upregulate in macrophage-engulfed *Candida* yeasts, concurrent with a downregulation of transcriptional machinery and glycolytic enzymes, allowing the cell to acquire carbon through alternative sources to glucose and conserve energy [82]. This suggests preferential use of the glyoxylate cycle as a response to glucose deprivation which is typical in the oligotrophic ocean.

In case physical processes transport *R. sphaerocarpa* ETNP2018 to the anoxic part of the water column, its genome shows the potential to ferment pyruvate via the enzyme pyruvate decarboxylase (EC: 4.1.1.1), creating acetaldehyde that can be further converted to acetate, ethanol, or carboxylic acids (Fig. 2). Acetate is synthesized from acetaldehyde via aldehyde dehydrogenase

(EC: 1.2.3.1) to replenish acetyl-CoA using its acetyl group (Fig. 2). Ethanol is then synthesized via alcohol dehydrogenase (EC: 1.1.1.2) alongside the interconversion of NADH to NAD^+ as a mechanism of replenishing the intracellular reducing agent (Figs. 2 and 3). It can then be excreted passively or converted to acetate in the peroxisome (Fig. 3). The small number of Pfam domains related to short chain dehydrogenase enzymes (PF00106), which are responsible for fermentative reactions on aldehydes and alcohols (Table 3), suggests the niche of *R. sphaerocarpa* ETNP2018 is not the anoxic portion of the water column. Reduced fermentative machinery may rather serve as an adaptation to their oligotrophic yet oxygenated environment, where biosynthetic resources are at a premium, and anaerobic metabolisms act as a backup for unfavorable changes in conditions.

Lane *et al. BMC Genomics*   (2023) 24:695

Page 10 of 14

Table 3 Pfam domains with major depletions in the genome of *R. sphaerocarpa* ETNP2018 as compared to the average of all *Rhodotorula* strains studied

Domain ID	Description	*R. sphaerocarpa* ETNP2018	Average	Standard Deviation
PF07690	Major Facilitator Superfamily	101	129.7	19.0
PF00172	Fungal Zn(2)-Cys(6) binuclear cluster domain	49	55.7	11.6
PF00106	Short Chain Dehydrogenase	37	45.3	6.1
PF13561	Enoyl-(Acyl carrier protein) reductase	35	44.4	6.0
PF08659	KR Domain	30	36.9	5.4
PF04082	Fungal specific transcription factor domain	21	29.7	4.0
PF12937	F-box-like	20	34.6	13.1
PF00646	F-box domain	14	22.2	4.8
PF01753	MYND finger	7	24.7	15.1
PF03171	2OG-Fe(II) oxygenase superfamily	5	9.6	3.4
PF02668	Taurine catabolism dioxygenase TauD, TfdA family	5	9.9	3.6
PF01179	Copper amine oxidase, enzyme domain	0	2.1	1.0
PF02727	Copper amine oxidase, N2 domain	0	1.1	0.7
PF03452	ANP1 (alpha-1,2-mannosyltransferase)	0	1.7	0.7
PF02194	PXA domain	0	1.7	0.7
PF07683	Cobalamin synthesis protein cobW C-terminal domain	0	1.7	0.6

Transporters

Compared to other *Rhodotorula* strains, the genomes of both *R. sphaerocarpa* ETNP2018 and GDMCC 60679 are particularly low in the number of the Pfam domain encoding the major facilitator superfamily (MFS) of transporters (PF07690) (Table 3), which play a significant role in the cross-membrane transport of organic solutes. This suggests *R. sphaerocarpa* strains isolated from seawater have streamlined their genomes given the low substrate availability in the ocean. MFS transporters also symport H^+ with siderophores, organometallic molecules formed by prokaryotes to sequester ferric iron [83, 84]. Although yeasts are considered to utilize siderophore assimilation as an opportunistic iron uptake mechanism [84], the low number of MSF transporters in the genomes of *R. sphaerocarpa* strains suggests a low competitiveness in siderophore acquisition, a tradeoff resulting from genome streamlining.

Nevertheless, the number of other metal transporters annotated by Pfam was not lower in the genomes of *R. sphaerocarpa* ETNP2018 and GDMCC 60679 in comparison to the other strains. High affinity iron permease (PF03239) was conserved in all 15 *Rhodotorula* genomes, suggesting ferrous iron intake provides them with much of the iron required for protein synthesis. One copy of PF10566, which contains natural resistance associated macrophage protein (Nramp) transporters, was conserved across all 15 *Rhodotorula* strains, aside from *R. toruloides* NBRC 0880, which contains two, and *R. kratochvilovae* YM25235, which contains zero. Nramp transporters, belonging to the Smf family of genes, are responsible for the cross-membrane transport of a variety of transition metals [85, 86]. Smf proteins demonstrate the highest affinity for Cu^{2+} and Mn^{2+} and are thought to be responsible for the high-affinity Mn^{2+} uptake system, but also show function in transporting ferrous iron, copper, nickel, cadmium, cobalt, zinc, and manganese [83, 85].

Nitrogen assimilation

Ten of the 15 representative *Rhodotorula* strains encode the genes for nitrate assimilation pathways (Fig. 2), through which nitrate and nitrite are transported into the cell by the nitrate/nitrite transporter narK and reduced to ammonium via the enzymes nitrate reductase (EC: 1.7.1.1) and nitrite reductase (EC: 1.7.1.4) (Fig. 3). Six of these ten genomes were isolated from aquatic sources: two from freshwater, *R. glutinis* ZHK and *R. kratochvilovae* YM25235, and four from the marine environment, *R. sphaerocarpa* ETNP2018, *R. paludigena* P4R5, *R. sphaerocarpa* GDMCC 60679, and *R. diobovata* 08–225 (Fig. 2). The only aquatic yeast lacking this genetic potential is *R. mucilaginosa* CYJ03, isolated from the Yellow Sea in Yunnan, China (Fig. 2) [43].

Nitrate assimilation, in particular the reduction of nitrate in the cytosol, is an energetically expensive process [87]. Strains lacking the genetic potential to assimilate nitrate were largely isolated from environments where competition for resources is less intense and alternative sources of nitrogen (e.g. ammonium, urea)

Lane *et al. BMC Genomics*   (2023) 24:695

Page 11 of 14

are likely readily available. *R. mucilaginosa* CYJ03 was isolated from the northern Yellow Sea, which has been eutrophic for decades [88]. It can therefore be inferred that *R. mucilaginosa* CYJ03 encounters comparatively high ammonium concentrations in the water column and to synthesize nitrate reductase would constitute a waste of biosynthetic resources.

Yeasts in the genus *Rhodotorula* have previously displayed the ability to grow on acetonitrile as a sole nitrogen source [89]. Nitrile hydratase (NHase) proteins, together with amidases (EC: 3.5.1.4), mediate a two-step metabolism of nitrile compounds such as acetonitrile to amides and acids; a second nitrile-hydrolyzing enzyme found in yeast, nitrilase (EC: 3.5.5.1), can perform the same reaction in one step. One of either nitrilase or cyanoalanine nitrilase (EC: 3.5.5.4) was found in all *Rhodotorula* genomes except for the two *R. sphaerocarpa* strains, *Rhodotorula sp.* CCFEE5036, and *R. taiwanensis* MD1149 (Fig. 2). However, none of the representative strains contained genes for NHase synthesis. Acetonitriles are predominately released via terrestrial biomass burning and constitute only a trace gas in the global atmosphere, placing open ocean *R. sphaerocarpa* strains far from stable sources of incorporable nitriles [90]. The lack of NHase, *CobW,* and nitrilase genes exclusively in both *R. sphaerocarpa* ETNP2018 and *R. sphaerocarpa* GDMCC 60679 suggests that as the lineage was diverging, *R. sphaerocarpa* strains did not retain *CobW* or nitrilase genes potentially due to a lack of available nitrile compounds in the ocean.

Secondary metabolisms

All *Rhodotorula* genomes contained between four and six BGCs, primarily from the categories of non-ribosomal polyketide synthase (NRPS) and Terpene synthesis. The genome of *R. sphaerocarpa* ETNP2018 included NRPS-like clusters 1.1 and 6.1 as well as terpene clusters 8.1 and 9.1. The core biosynthetic gene of terpene cluster 9.1 functions in the formation of an isoprenoid biosynthetic complex. A BLASTp search identified both lycopene β-cyclase (EC: 5.5.1.19) and phytoene synthase (EC: 2.5.1.32) domains in the complex. Both lycopene and phytoene result from the digestion of cytosolic acetyl-CoA during the mevalonate (MVA) pathway, which converts acetyl-CoA into isopentyl diphosphate (Fig. 3) [91]. Phytoene is converted to lycopene via the enzymatic action of phytoene desaturase (EC: 1.3.99.30), where it can be further metabolized to form the carotenoid β-carotene via lycopene β-cyclase [91]. Yeast carotenoids are responsible for protection from over-exposure to ultraviolet light, in addition to proposed antimicrobial activity

[92]. *Rhodotorula* have been shown to increase carotenogenesis as light intensity increases, indicating the molecules have a photoprotective role in the cell [92, 93]. *R. sphaerocarpa* ETNP2018 encodes five laccases (AA1), the most of all 15 *Rhodotorula* analyzed (Table S6). Fungal laccases can be ligninolytic enzymes, and are also known to function in plant pathogenicity, detoxification, and pigment modification [94]. Laccases degrade β-carotene and other carotenoids, so they may play a role in the breakdown of intracellular carotenoid pigments. Light is attenuated more rapidly in eutrophic lakes than in pelagic seawater due to the presence of particulate and dissolved organic matter, so lake yeasts would have less exposure to potentially harmful UV light and thus a reduced requirement for pigmented molecules [95].

Conclusions

Our analysis suggests that the marine yeast *R. sphaerocarpa* ETNP2018 has adapted to conditions in the oligotrophic marine environment through genomic and subsequent biosynthetic streamlining. Its genome is smaller than the typical *Rhodotorula* strain, allowing it to conserve limited nutrients during replication without a massive reduction in potential proteins. Reduction was expected for transcription-related genes but was primarily found in biosynthetic genes. In the genome of *R. sphaerocarpa* ETNP2018, the number of KOGs related to carbohydrate, lipid, and secondary metabolisms was lower than average, and depleted Pfam domains were mostly related to transport and non-essential biosynthetic pathways. The number of CAZymes was lower in the genome of *R. sphaerocarpa* ETNP2018 than the average of *Rhodotorula* genomes, and the reduction in CAZymes was primarily found among CAZymes involved in the metabolism of non-marine biopolymers. The conservation of core carbohydrate metabolisms in the genome of *R. sphaerocarpa* ETNP2018, including carotenoid production, suggests it has maintained an independent lifestyle despite streamlining pressures.

Supplementary Information

Additional file 1: Table S1. The NCBI accession number, origin, and isolation source/host of *Rhodotorula* genomes chosen for comparison. **Table S2.** Bacterial and archaeal genomes selected for comparison to *Pelagibacter ubique* HTCC1062 and *Nitrosopumilus maritimus* SCM1. **Table S3.** Eukaryotic clusters of orthologous groups (KOGs) assigned to each functional category for fifteen *Rhodotorula* species. **Table S4.** Statistical analysis of the depletion of specific KOG categories in *Rhodotorula sphaerocarpa* ETNP2018, performed via one sample T-test. All tests were performed using a 95% confidence interval and the average of all fifteen representative *Rhodotorula* strains. **Table S5.** CAZyme families present in 15 representative Rhodotorula strains determined using dbCAN2. The carbohydrate binding module domains (CBM) column only includes genes

---

Lane *et al. BMC Genomics*   (2023) 24:695

Page 12 of 14

containing both glycoside hydrolase (GH) and carbohydrate binding module domains (CBM). **Table S6.** All CAZymes (including their described functions) present in 15 representative Rhodotorula strains. **Figure S1.** A heatmap showing highly conserved carbohydrate metabolism pathways in the *Rhodotorula* genus. **Figure S2.** A phylogenomic tree constructed using single-copy orthologues shared amongst 168 representative *Rhodotorula* strains (as well as two outgroup species, *Microbotrium intermedium* GCA 900096595.1 and *Leucosporidium creatinivorum* GCA 002105055.1). **Figure S3.** Heatmap showing the number of major CAZymes present in the *Rhodotorula* genus.

Availability of data and materials
The genomic data for *R. sphaerocarpa* ETNP2018 is available at the NCBI with the BioSample number SAMN15201391.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References
1. Amend A, Burgaud G, Cunliffe M, Edgcomb VP, Ettinger CL, Gutiérrez MH, et al. Fungi in the marine environment: open questions and unsolved problems. mBio. 2019;10:e01189-18.
2. Grossart H-P, Rojas-Jimenez K. Aquatic fungi: targeting the forgotten in microbial ecology. Curr Opin Microbiol. 2016;31:140–5.
3. Setchell WA. The marine flora of the pacific coast. Nature and science on the pacific coast: a guide-book for scientific travelers in the west. 1915. p. 177.
4. Hassett BT, Borrego EJ, Vonnahme TR, Rämä T, Kolomiets MV, Gradinger R. Arctic marine fungi: biomass, functional genes, and putative ecological roles. ISME J. 2019;13:1484–96.
5. Richards TA, Jones MDM, Leonard G, Bass D. Marine fungi: their ecology and molecular diversity. Ann Rev Mar Sci. 2012;4:495–522.
6. Hassett BT, Vonnahme TR, Peng X, Jones EBG, Heuzé C. Global diversity and geography of planktonic marine fungi. Bot Mar. 2020;63:121–39.
7. Grigoriev IV, Nikitin R, Haridas S, Kuo A, Ohm R, Otillar R, et al. MycoCosm portal: gearing up for 1000 fungal genomes. Nucleic Acids Res. 2014;42:D699-704.
8. Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. ISME J. 2014;8:1553–65.
9. Cortez D, Neira G, González C, Vergara E, Holmes DS. A large-scale genome-based survey of acidophilic bacteria suggests that genome streamlining is an adaption for life at low pH. Front Microbiol. 2022;13:803.
10. Gilbert JD, Fagan WF. Contrasting mechanisms of proteomic nitrogen thrift in Prochlorococcus. Mol Ecol. 2011;20:92–104.
11. Grzymski JJ, Dussaq AM. The significance of nitrogen cost minimization in proteomes of marine microorganisms. ISME J. 2012;6:71–80.
12. Peng X, Valentine DL. Diversity and N2O production potential of fungi in an oceanic oxygen minimum zone. J Fungi. 2021;7:218.
13. Kot AM, Błażejak S, Kieliszek M, Gientka I, Bryś J, Reczek L, et al. Effect of exogenous stress factors on the biosynthesis of carotenoids and lipids by Rhodotorula yeast strains in media containing agro-industrial waste. World J Microbiol Biotechnol. 2019;35:157.
14. Mannazzu I, Landolfo S, da Silva TL, Buzzini P. Red yeasts and carotenoid production: outlining a future for non-conventional yeasts of biotechnological interest. World J Microbiol Biotechnol. 2015;31:1665–73.
15. Huang XP, Huang LM, Yue WZ. The characteristics of nutrients and eutrophication in the Pearl River estuary, South China. Mar Pollut Bull. 2003;47:30–6.
16. Firrincieli A, Otillar R, Salamov A, Schmutz J, Khan Z, Redman R, et al. Genome sequence of the plant growth promoting endophytic yeast Rhodotorula graminis WP1. Front Microbiol. 2015;6:978.
17. Ayaz ÇM, Gülmez D, Akdağlı S, Uzun Ö. A Rare yeast: cases of rhodotorula mucilaginosa infection followed up in a tertiary University Hospital. Mikrobiyol Bul. 2021;55:91–8.
18. Hou P, Chang F, Duan L, Zhang Y, Zhang H. Seasonal variation and spatial heterogeneity of water quality parameters in lake Chenghai in Southwestern China. Water. 2022;14:1640.
19. Simpson AC, Urbaniak C, Bateh JR, Singh NK, Wood JM, Debieu M, et al. Draft genome sequences of fungi isolated from the international space station during the microbial tracking-2 experiment. Microbiol Resour Announc. 2021;10:e00751-e821.
20. Touchette D, Altshuler I, Gostinčar C, Zalar P, Raymond-Bouchard I, Zajc J, et al. Novel Antarctic yeast adapts to cold by switching energy metabolism and increasing small RNA synthesis. ISME J. 2021;16:221–32.
21. Hoondee P, Wattanagonniyom T, Weeraphan T, Tanasupawat S, Savarajara A. Occurrence of oleaginous yeast from mangrove forest in Thailand. World J Microbiol Biotechnol. 2019;35:108.
22. Newell SY, Fell JW. The perfect form of a marine-occurring yeast of the genus rhodotorula. Mycologia. 1970;62:272–81.
23. Kester DR, Duedall IW, Connors DN, Pytkowicz RM. Preparation of artificial seawater1. Limnol Oceanogr. 1967;12:176–9.
24. Bushnell B. BBMap short-read aligner, and other bioinformatics tools. 2015.
25. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20.
26. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
27. Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, et al. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018;35:543–8.
28. Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, et al. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 2019;47:D807–11.
29. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:113.
30. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25:1972–3.
31. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9.

Lane *et al. BMC Genomics*   (2023) 24:695

Page 13 of 14

32. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.

33. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. Syst Biol. 2016;65:997–1008.

34. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.

35. Wang M, Mao W, Wang X, Li F, Wang J, Chi Z, et al. Efficient simultaneous production of extracellular polyol esters of fatty acids and intracellular lipids from inulin by a deep-sea yeast Rhodotorula paludigena P4R5. Microb Cell Fact. 2019;18:149.

36. Newell SY, Hunter IL. Rhodosporidium diobovatum sp. n., the perfect form of an asporogenous yeast (Rhodotorula sp.). J Bacteriol. 1970;104:503–8.

37. Coleine C, Masonjones S, Onofri S, Selbmann L, Stajich JE. Draft genome sequence of the yeast Rhodotorula sp. strain CCFEE 5036, isolated from McMurdo Dry Valleys, Antarctica. Microbiol Resour Announc. 2020;9:e00020–20.

38. Coradetti ST, Pinel D, Geiselman GM, Ito M, Mondo SJ, Reilly MC, et al. Functional genomics of lipid metabolism in the oleaginous yeast Rhodosporidium toruloides. eLife. 2018;7:e32110.

39. Goordial J, Raymond-Bouchard I, Riley R, Ronholm J, Shapiro N, Woyke T, et al. Improved high-quality draft genome sequence of the Eurypsychrophile Rhodotorula sp. JG1b, isolated from permafrost in the hyperarid upper-elevation McMurdo Dry Valleys, Antarctica. Genome Announc. 2016;4:10.1128/genomea.00069-16.

40. Cui J, He S, Ji X, Lin L, Wei Y, Zhang Q. Identification and characterization of a novel bifunctional Δ12/Δ15-fatty acid desaturase gene from Rhodosporidium kratochvilovae. Biotechnol Lett. 2016;38:1155–64.

41. Li C-J, Zhao D, Cheng P, Zheng L, Yu G-H. Genomics and lipidomics analysis of the biotechnologically important oleaginous red yeast Rhodotorula glutinis ZHK provides new insights into its lipid and carotenoid metabolism. BMC Genomics. 2020;21:834.

42. Simpson AC, Urbaniak C, Bateh JR, Singh NK, Wood JM, Debieu M, et al. Draft genome sequences of fungi isolated from the International Space Station during the microbial tracking-2 experiment. Microbiol Resour Announc. 2021;10:10.1128/mra.00751-21.

43. Tang W, Wang Y, Cai Y, Liu S, Zhang J, He Z. Genome sequence of a marine carotenoid producing yeast Rhodotorula mucilaginosa CYJ03. J Ocean Univ China. 2020;19:466–72.

44. Tkavc R, Matrosova VY, Grichenko OE, Gostinčar C, Volpe RP, Klimenkova P, et al. Prospects for fungal bioremediation of acidic radioactive waste sites: characterization and genome sequence of Rhodotorula taiwanensis MD1149. Front Microbiol. 2018;8.

45. Palmer JM, Stajich J. Funannotate v1.8.1: eukaryotic genome annotation. 2020.

46. Frith MC. A new repeat-masking method enables specific detection of homologous sequences. Nucleic Acids Res. 2011;39:e23–e23.

47. Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genom Bioinform. 2021;3:lqaa108.

48. Brůna T, Lomsadze A, Borodovsky M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR Genom Bioinform. 2020;2:lqaa026.

49. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.

50. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20:2878–9.

51. Wu TD, Reeder J, Lawrence M, Becker G, Brauer MJ. GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. In: Mathé E, Davis S, editors. Statistical genomics: methods and protocols. New York: Springer; 2016. p. 283–334.

52. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24:637–44.

53. Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 2003;31:365–70.

54. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 2008;9:R7.

55. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.

56. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.

57. Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol. 2016;428:726–31.

58. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.

59. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38:5825–9.

60. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019;47:D309-14.

61. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, et al. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2018;46:W95–101.

62. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 2014;42:D490–5.

63. The CAZypedia Consortium. Ten years of CAZypedia: a living encyclopedia of carbohydrate-active enzymes. Glycobiology. 2018;28:3–8.

64. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. Nucleic Acids Res. 2021;49:W29-35.

65. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C, et al. CDD: a conserved domain database for the functional annotation of proteins. Nucleic Acids Res. 2011;39 suppl_1:D225-9.

66. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, et al. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res. 2020;48:D265–8.

67. Kastanos EK, Woldman YY, Appling DR. Role of mitochondrial and cytoplasmic serine hydroxymethyltransferase Isozymes in de novo purine synthesis in saccharomyces cerevisiae. Biochemistry. 1997;36:14956–64.

68. Cherry JM, Hong EL, Amundsen C, Balakrishnan R, Binkley G, Chan ET, et al. Saccharomyces genome database: the genomics resource of budding yeast. Nucleic Acids Res. 2012;40 Database issue:D700-5.

69. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci. 2020;29:28–35.

70. Hagestad OC, Hou L, Andersen JH, Hansen EH, Altermark B, Li C, et al. Genomic characterization of three marine fungi, including Emericellopsis atlantica sp. nov. with signatures of a generalist lifestyle and marine biomass degradation. IMA Fungus. 2021;12:21.

71. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. Science. 2005;309:1242–5.

72. Breyer E, Böhm M, Reitbauer M, Amano C, Heitger M, Baltar F. Autofluorescence is a common trait in different oceanic fungi. J Fungi. 2021;7:709.

73. Yun L, Wang W, Li Y, Xie M, Chen T, Hu C, et al. Potential application values of a marine red yeast, Rhodosporidiums sphaerocarpum YLY01, in aquaculture and tail water treatment assessed by the removal of ammonia nitrogen, the inhibition to Vibrio spp., and nutrient composition. PLoS One. 2021;16:e0246841.

74. Libkind D, Buzzini P, Turchetti B, Rosa CA. Yeasts in continental and seawater. In: Buzzini P, Lachance M-A, Yurkov A, editors. Yeasts in natural ecosystems: diversity. Cham: Springer International Publishing; 2017. p. 1–61.

75. Souza CP, Almeida BC, Colwell RR, Rivera ING. The importance of chitin in the marine environment. Mar Biotechnol. 2011;13:823–30.

76. Gooday GW. Physiology of microbial degradation of chitin and chitosan. In: Ratledge C, editor. Biochemistry of microbial degradation. Dordrecht: Springer Netherlands; 1994. p. 279–312.

Lane *et al. BMC Genomics*   (2023) 24:695

Page 14 of 14

77. Ishii J, Okazaki F, Djohan AC, Hara KY, Asai-Nakashima N, Teramura H, et al. From mannan to bioethanol: cell surface co-display of β-mannanase and β-mannosidase on yeast Saccharomyces cerevisiae. Biotechnol Biofuels. 2016;9:188.

78. Yu R, Campbell K, Pereira R, Björkeroth J, Qi Q, Vorontsov E, et al. Nitrogen limitation reveals large reserves in metabolic and translational capacities of yeast. Nat Commun. 2020;11:1881.

79. Battaglia E, Benoit I, van den Brink J, Wiebenga A, Coutinho PM, Henrissat B, et al. Carbohydrate-active enzymes from the zygomycete fungus Rhizopus oryzae: a highly specialized approach to carbohydrate degradation depicted at genome level. BMC Genomics. 2011;12:38.

80. Wurzbacher C, Rösel S, Rychła A, Grossart H-P. Importance of saprotrophic freshwater fungi for pollen degradation. PLoS One. 2014;9:e94643.

81. Zhou J, Li X, Huang P-W, Dai C-C. Endophytism or saprophytism: decoding the lifestyle transition of the generalist fungus phomopsis liquidambari. Microbiol Res. 2018;206:99–112.

82. Chew SY, Chee WJY, Than LTL. The glyoxylate cycle and alternative carbon metabolism as metabolic adaptation strategies of Candida glabrata: perspectives from Candida albicans and Saccharomyces cerevisiae. J Biomed Sci. 2019;26:52.

83. Eide DJ. The molecular biology of metal ion transport in saccharomyces cerevisiae. Annu Rev Nutr. 1998;18:441–69.

84. Behnsen J, Raffatellu M. Siderophores: more than stealing iron. mBio. 2016;7:e01906-16.

85. Cohen A, Nelson H, Nelson N. The family of SMF metal ion transporters in yeast cells*. J Biol Chem. 2000;275:33388–94.

86. Liu XF, Culotta VC. Mutational analysis of Saccharomyces cerevisiae Smf1p, a member of the Nramp family of metal transporters. J Mol Biol. 1999;289:885–91.

87. Siverio JM. Assimilation of nitrate by yeasts. FEMS Microbiol Rev. 2002;26:277–84.

88. Zheng L, Zhai W. Excess nitrogen in the Bohai and Yellow seas, China: distribution, trends, and source apportionment. Sci Total Environ. 2021;794:148702.

89. Rezende RP, Dias JCT, Rosa CA, Carazza F, Linardi VR. Utilization of nitriles by yeasts isolated from a Brazilian gold mine. J Gen Appl Microbiol. 1999;45:185–92.

90. E S, Holzinger R, Kleiss B, Donoso L, Crutzen P. New insights in the global cycle of acetonitrile: release from the ocean and dry deposition in the tropical savanna of Venezuela. Atmosher Chem Phys. 2004;4.

91. Jing Y, Guo F, Zhang S, Dong W, Zhou J, Xin F, et al. Recent advances on biological synthesis of lycopene by using industrial yeast. Ind Eng Chem Res. 2021;60:3485–94.

92. Vargas-Sinisterra AF, Ramírez-Castrillón M. Yeast carotenoids: production and activity as antimicrobial biomolecule. Arch Microbiol. 2021;203:873–88.

93. Kong W, Yang S, Agboyibor C, Chen D, Zhang A, Niu S. Light irradiation can regulate the growth characteristics and metabolites compositions of Rhodotorula mucilaginosa. J Food Sci Technol. 2019;56:5509–17.

94. Kalyani D, Tiwari MK, Li J, Kim SC, Kalia VC, Kang YC, et al. A Highly efficient recombinant laccase from the yeast yarrowia lipolytica and its application in the hydrolysis of biomass. PLoS One. 2015;10:e0120156.

95. Pérez GL, Queimaliños CP, Modenutti BE. Light climate and plankton in the deep chlorophyll maxima in North Patagonian Andean lakes. J Plankton Res. 2002;24:591–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.