

RESEARCH ARTICLE

Open Access



Genomics and lipidomics analysis of the biotechnologically important oleaginous red yeast *Rhodotorula glutinis* ZHK provides new insights into its lipid and carotenoid metabolism

Chun-Ji Li^{1,2}, Die Zhao³, Ping Cheng^{1,2}, Li Zheng^{1,2} and Guo-Hui Yu^{1,2*} 

Abstract

Background: *Rhodotorula glutinis* is recognized as a biotechnologically important oleaginous red yeast, which synthesizes numerous meritorious compounds with wide industrial usages. One of the most notable properties of *R. glutinis* is the formation of intracellular lipid droplets full of carotenoids. However, the basic genomic features that underlie the biosynthesis of these valuable compounds in *R. glutinis* have not been fully documented. To reveal the biotechnological potential of *R. glutinis*, the genomics and lipidomics analysis was performed through the Next-Generation Sequencing and HPLC-MS-based metabolomics technologies.

Results: Here, we firstly assemble the genome of *R. glutinis* ZHK into 21.8 Mb, containing 30 scaffolds and 6774 predicted genes with a N50 length of 14, 66,672 bp and GC content of 67.8%. Genome completeness assessment (BUSCO alignment: 95.3%) indicated the genome assembly with a high-quality features. According to the functional annotation of the genome, we predicted several key genes involved in lipids and carotenoids metabolism as well as certain industrial enzymes biosynthesis. Comparative genomics results suggested that most of orthologous genes have underwent the strong purifying selection within the five *Rhodotorula* species, especially genes responsible for carotenoids biosynthesis. Furthermore, a total of 982 lipids were identified using the lipidomics approaches, mainly including triacylglycerols, diacylglyceryltrimethylhomo-ser and phosphatidylethanolamine.

Conclusion: Using whole genome shotgun sequencing, we comprehensively analyzed the genome of *R. glutinis* and predicted several key genes involved in lipids and carotenoids metabolism. By performing comparative genomic analysis, we show that most of the ortholog genes have undergone strong purifying selection within the five *Rhodotorula* species. Furthermore, we identified 982 lipid species using lipidomic approaches. These results provided valuable resources to further advance biotechnological applications of *R. glutinis*.

Keywords: *Rhodotorula glutinis*, Lipids, Carotenoids, Genome assembly, Lipidomics

* Correspondence: ygh76411@zhku.edu.cn

¹Innovative Institute for Plant Health, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, People's Republic of China

²College of Agriculture and Biology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, People's Republic of China

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Rhodotorula is a genus of oleaginous pigmented yeasts, part of the phylum Basidiomycota, class Microbotryomycetes, and order Sporidiobolales [1]. *R. glutinis* is recognized as a representative species of genus *Rhodotorula*, and it was reported firstly by F.C. Harrison in 1928 [2]. This species grows on a broad spectrum of ecological environments, ranging from air, soil, and ocean, as well as in the bodies of animals, plants, and lower organisms [3]. Most of these strains are aerobic, mesophilic, and spherical, ellipsoidal, or elongated in shape [4]. *R. glutinis* is of great industrial importance as it synthesizes numerous valuable compounds, such as, lipids (SCO, single-cell oils), carotenoids, and enzymes [5, 6].

R. glutinis strains are exceptionally robust in converting low-cost carbohydrates into lipids, such as triglycerides (TAGs) [7]. This species also produces several microbial oils, mainly including oleic, linoleic, palmitic, and stearic acid [8, 9]. These lipid metabolites account for up to 72% of its cells dry mass, and could be used as food additives, diet supplements and raw material to produce second-generation biodiesel [10]. Depending on growth conditions, its colony color presents yellow, pink, and blood red, largely due to the relative proportion of the carotenoids produced, including torulene, torularhodin, and β -carotene [11, 12]. Due to their health-promoting properties, these carotenoids are of great biotechnological potential could be used in the food, pharmaceutical, cosmetics and feed industries [13]. Previous studies concluded that both of torulene and torularhodin possess stronger anti-oxidative properties than β -carotene [14, 15]. Furthermore, the growing scientific evidences suggested that torulene and torularhodin may have potential benefits in the prevention of tumors, especially prostate and liver cancer [16–18]. Additionally, torularhodin is capable of enhancing the antimicrobial ability of TiO₂/Ti materials, and it may become a novel natural antimicrobial agent [19, 20]. Torularhodin supplementations significantly alleviates the alcoholic liver disease via decreasing the levels of ethanol-induced aspartate transaminase (ALT), aspartate transaminase (AST) and low density lipoprotein (LDL) [21]. Due to its higher biosynthetic efficiency, *R. glutinis* has been considered as one of the most important producers of torulene and torularhodin, especially the latter [22, 23]. The toxicological evaluation concluded that the pigments torulene and torularhodin extract from *R. glutinis* yeasts could be used safely as food additives, which retain anti-oxidative properties as well as colorant role [24].

Besides, biomass of *R. glutinis* strains could be used as the natural sources of lipases, α -L-arabinofuranosidase, invertase, pectinases, and tannin acyl hydrolase, particularly phenylalanine ammonia lyase (PAL) [25, 26]. Furthermore, *R. glutinis* has been regarded as a biocontrol agent for post-harvest microbial diseases of fruit [27],

possibly due to its antagonistic attachment capability to pathogenic bacteria and production of torularhodin [28, 29]. More importantly, lipids, carotenoids and industrial enzymes synthesized by *R. glutinis* strains have advantages over others, mainly due to a higher biotransformation rate, low-costs and independent of climates [30, 31]. Therefore, the usage of *R. glutinis* strains as bioreactors for the production of industrial bio-products using low-cost natural substrates have given rise to a strong interest currently [32].

Nevertheless, despite a considerable number of literatures presenting the industrial applications of *R. glutinis* strains, little is known about its basic genomic characteristics (formerly *R. glutinis* ATCC 204091 has been designated as *R. toruloides*) [33, 34]. Genetic backgrounds involved in the biosynthesis of lipids, carotenoids, enzymes and other precious metabolites in *R. glutinis* still remain poorly studied. The PacBio single-molecule long-read sequencing [35] and LC-MS/MS-based lipidomics approaches offer the advantage of increased precision of genome assembly and lipid metabolites identification [36]. Here, to better enable the continued industrial applications of this versatile single-cell yeast, we present the *de novo* genome assembly and shotgun lipidomics evaluation of the *R. glutinis* strain ZHK. We also qualitatively and quantitatively determined the carotenoid contents in the strain ZHK. Subsequently, we carried out a comparative analysis to investigate orthologous and species-specific genes between *R. glutinis* and its closely related species. Through comparative genome analysis of *R. glutinis*, we explored the evolutionary dynamics with its close related *Rhodotorula* species. Continuously, this work significantly enriched our understanding of the molecular basis underlying the industrial bio-products synthesis of this species. This should promote the development of the genetic engineering for the overproducing of bioactive natural products in *R. glutinis* strains.

Results

Genome sequencing, assembly and assessment

Here, the whole genome shotgun sequencing of oleaginous red yeast *R. glutinis* ZHK was accomplished using the PacBio Sequel system with a Single Molecule Real-Time (SMRT) sequencer and an Illumina Hiseq 2500 system. As a result, a total of 6.62 Gb polymerase reads from a 20 kb library was generated by SMRT sequencing. After removing adapters and low-quality reads, we obtained 6.57 Gb (~300 \times) subreads for whole genome assembly. Primary contigs were assembled and corrected from PacBio long reads using the program MECAT. The obtained contigs were further polished using 3.6 Gb (~170 \times) short paired-end reads (PE150, Illumina) by the program Pilon. As shown in Fig. 1, our final assembled genome of the strain *R. glutinis* ZHK consists of 30

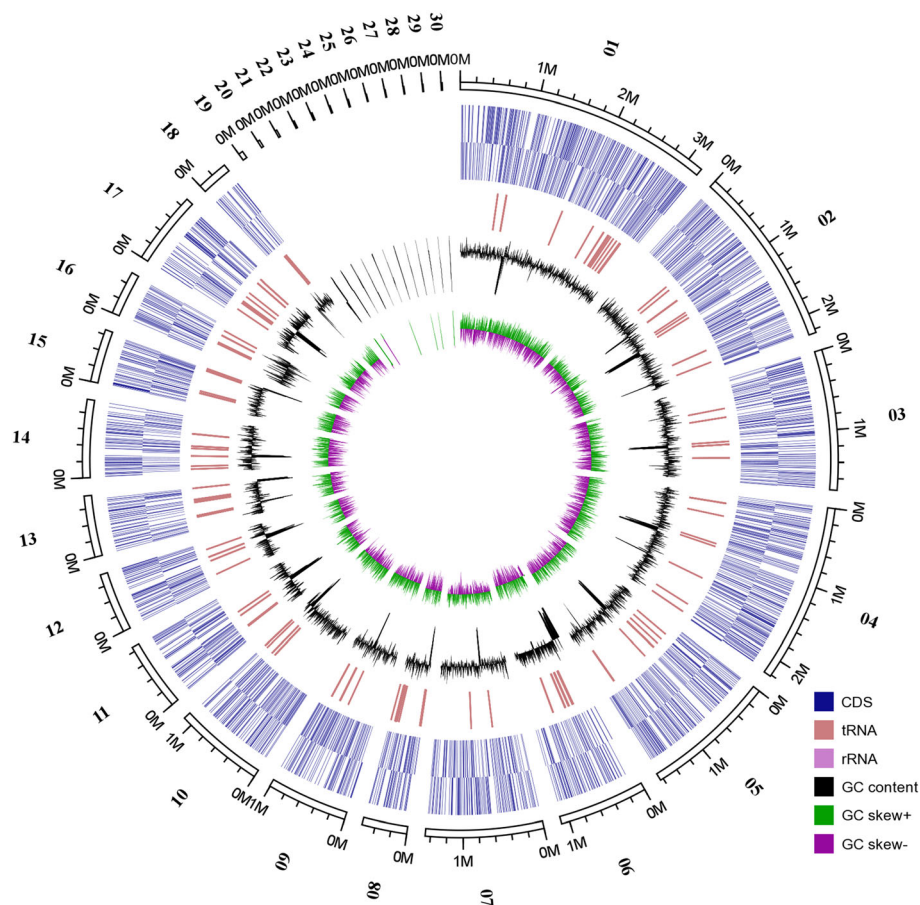


Fig. 1 Genomic features of *R. glutinis* strain ZHK. The 21.8 Mb genome of strain ZHK, containing 30 scaffolds, the longest length scaffold of 3,195,425 bp, the shortest length scaffold of 10,281 bp, a N50 length of 1,466,672 bp, a GC content of 67.8%. The ZHK genome encodes 6774 predicted proteins and 156 transfer RNAs (tRNAs), which are validated by RNA-seq. From the outer circle to the inner, it represents the length of scaffolds, coding sequence (CDS), tRNAs, GC content (black) and GC skew curve (Green: positive GC skew; Violet: negative GC skew), respectively

scaffolds, a N50 length of 1,466,672 bp, the longest length scaffold of 3,195,425 bp, the shortest length scaffold of 10,281 bp, a GC content of 67.8%, and a size of 21.8 Mb. Using the program GeneMark-ES, we predicted 6774 genes in the ZHK genome with an average length of 1813 bp and a mean GC content of 69.59%, which occupies 55.0% of the whole genome sequence. In addition, a total of 156 transfer RNAs (tRNAs) were identified in the ZHK genome using the program tRNAscan-SE, with an average length of 90 bp. The full-length tRNAs totally comprised 14053 bp, accounting for 0.06% of the whole genome sequence. The results of the BUSCO alignment showed that the assembled genome contains 1272 complete BUSCOs (94.4%), of which 1260 were single-copy, while 12 were duplicated (Additional file 1). For the transcriptome sequencing (RNA-seq) results, a total of 3.54 Gb raw reads (Q20: 95%, Q30: 90%, and GC-content: 66.32%) were generated. Regarding the results of RNA-seq, we found that 6642 (98.05%) genes predicted in the ZHK genome

regions and 143 novel genes were expressed (Additional file 2). Furthermore, the RNA-seq results showed that 93.92% reads were matched to exon regions, 2.18% to intron regions, and 3.91% to intergenic regions. These reads are aligned to the intron region, possibly due to the intron retention or alternative splicing events. In addition, a total of 1513 SNPs/InDel (Additional file 3) were identified through comparing RNA-seq data with the ZHK whole genome sequence. From the RNA-seq results, we also obtained the boundaries of 5'UTR and 3'UTR of 1814 predicted genes in the ZHK genome (Additional file 4). Consequently, both of BUSCO alignment and transcriptome mapping indicated that our current genome assembly with a high-quality of completeness. Up to now, except for our *R. glutinis* ZHK, there are eleven whole genome sequences of *Rhodotorula* species have been sequenced. In general, the genome size of these *Rhodotorula* species is ~20 Mb, which represents a smaller grade when compared to other classes in order Pucciniomycotina [37]. The phenomenon of the

decrease in genome size is considered as a shared feature in class Microbotryomycetes.

Functional annotation

As for the 6774 predicted genes, 6560 (96.84%) genes could be annotated using NCBI Nr databases based on sequence homology. Among these genes annotated to Nr database, the top three species of matched gene number are *Rhodotorula toruloides* NP11 (3141, 47.88%), *Rhodotorula graminis* WP1 (521, 7.94%), and *Rhodotorula* sp. JG-1b (479, 7.30%). Moreover, 3831 (56.55%), 3345 (50.12%), and 2880 (42.52%) genes could be annotated according to the Swissport, KOG and KEGG databases, respectively. Of which, these genes assigned to KOG categories, are involved in 31082 proteins and domains. Top five KOG categories of annotated gene number are “General function prediction only (799, 23.88%)”, “Posttranslational modification, protein turnover, chaperones (689, 20.59%)”, “Signal transduction mechanisms (640, 19.13%)”, “RNA processing and modification (406, 12.13%)”, and “Energy production and conversion (380, 11.36%)”. In order to further understand the coding genes of candidate for biotechnological potential in the ZHK genome, the 2880 genes have been successfully assigned to their orthologous in the KEGG Pathways database. Of which, there are 1609 genes have been enriched in 21 categories of KEGG_B_class, including 120 KEGG pathways. Top three KEGG pathways of annotated genes number are “Biosynthesis of antibiotics (ko01130; 200, 12.43%)” (Additional file 5), “Biosynthesis of amino acids (ko01230; 100, 6.22%)”, and “Carbon metabolism (ko01200; 85, 5.28%)”. Based on the KEGG pathways mapping, we annotated several candidate genes for biotechnological potential as following: 1) lipid metabolism, including genes encoding ACACa (acetyl-CoA carboxylase), ACOX3 (acyl-CoA oxidase), and PDAT (phospholipid:diacylglycerol acyltransferase); 2) biosynthesis of carotenoids, including *crtE* (geranylgeranyl diphosphate synthase), *crtI* (phytoene desaturase), and *crtYB* (bifunctional, phytoene synthase/lycopene beta-cyclase); 3) biosynthesis of enzymes, including genes encoding PAL (phenylalanine ammonia-lyase), TGL2 (triacylglycerol lipase), and MGLL (acylglycerol lipase).

Furthermore, there are 1069 predicted genes could be classified into three Gene Ontology (GO) categories: Cellular Component (581 genes), Biological Process (812 genes), and Molecular Function (657 genes). These 1069 genes mainly distributed across five functional entries, including “Cellular Process (558, 52.20%)”, “Metabolic Process (555, 51.92%)”, “Cell (512, 47.90%)”, “Catalytic Activity (467, 43.69%)”, and “Single-organism Process (421, 39.38%)”. In addition, we identified 43 interspersed repetitive sequences and 6348 tandem repeats, including

103 macro-satellites DNA, 2870 mini-satellite DNA and 1234 micro-satellite DNA in the ZHK genome. Moreover, a total of 1,599,119 bp full-length TEs were predicted in the ZHK genome. As shown in Fig. 2, these TEs mainly include 316 (0.73%) LTR-REs, 325 (0.93%) LINE-Res, 370 (0.64%) DNA transposons, 325 LINE-REs and 952 (5.22%) unknowns, of which 16.04% are Class LTR element, mainly assigned to Copia (112), Gypsy (160) and Pao (19). The full-length TEs totally accounts for 7.16% of the ZHK whole genome sequence.

Analysis of lipidomics and carotenoids

In order to investigate the lipid metabolites in *R. glutinis* ZHK cell, we performed the shotgun lipidomics profiling for qualitative and quantitative characterization of lipidome in mixed samples acquired from three growth periods (24 h, 48h, and 72h). Samples were processed for chromatographic separation and mass spectrometry. As shown in Table 1, a total of 982 lipid species were identified using data-dependent MS/MS scans in both positive (POS) and negative (NEG) mode, mainly including 325 Triacylglycerols (TAGs), 95 Diacylglyceryltrimethylhomo-Ser (DGTS), 64 Phosphatidylethanolamine (PE), 62 Phosphatidylcholine (PC), 64 Ceramide (Cer), 40 Hexosylceramide (HexCer), 35 Sulfoquinovosyl-diacylglycerol (SQDG), 31 Glucuronosyl-diacylglycerol (GlcADG), 27 Diacylglycerols (DAG), 26 Sulfatide (SHexCer), 26 Phosphatidylinositol (PI) and 23 Fatty acid (FA) (Comprehensive information of all identified lipid metabolites is given in Additional file 6). To gather further information of carotenoids biosynthesis in *R. glutinis* ZHK, we employed the HPLC analysis to evaluate carotenoid contents in the same mixed samples with lipidomics profiling, both qualitatively and quantitatively. It is shown that the total amount of carotenoids in *R. glutinis* ZHK reaches to 1276.02 $\mu\text{g/g}_{\text{dw}}$. The relative proportion of torularhodin, torulene and β -carotene of total carotenoids accounts for 66.64% (890.25 $\mu\text{g/g}_{\text{dw}}$), 14.07% (179.51 $\mu\text{g/g}_{\text{dw}}$), and 19.29% (206.36 $\mu\text{g/g}_{\text{dw}}$), respectively.

Analysis of phylogenetic and syntenic relationships

Taxonomically, a total of 42 yeast species have been accepted in order Sporidiobolales (*Rhodospodiobolus*: 9; *Rhodotorula*: 15; *Sporobolomyces*: 18) [38]. These strains are commonly known as the oleaginous red yeasts because of their synthesis of lipid droplets full of carotenoids. In order to investigate the phylogenetic relationships between these oleaginous red yeasts assigned to order Sporidiobolales, we constructed the phylogenetic tree with the available 26S rDNA sequences from NCBI Nucleotide database. As shown in Fig. 3, as for the genus *Rhodotorula*, the strain *R. glutinis* ZHK showed a closer evolutionary relationship with *R. babjaveae*, *R. graminis* and *R. diobovata* than the other species. Basing on the results of phylogenetic analysis, we carried out the syntenic analysis of genome between *R.*

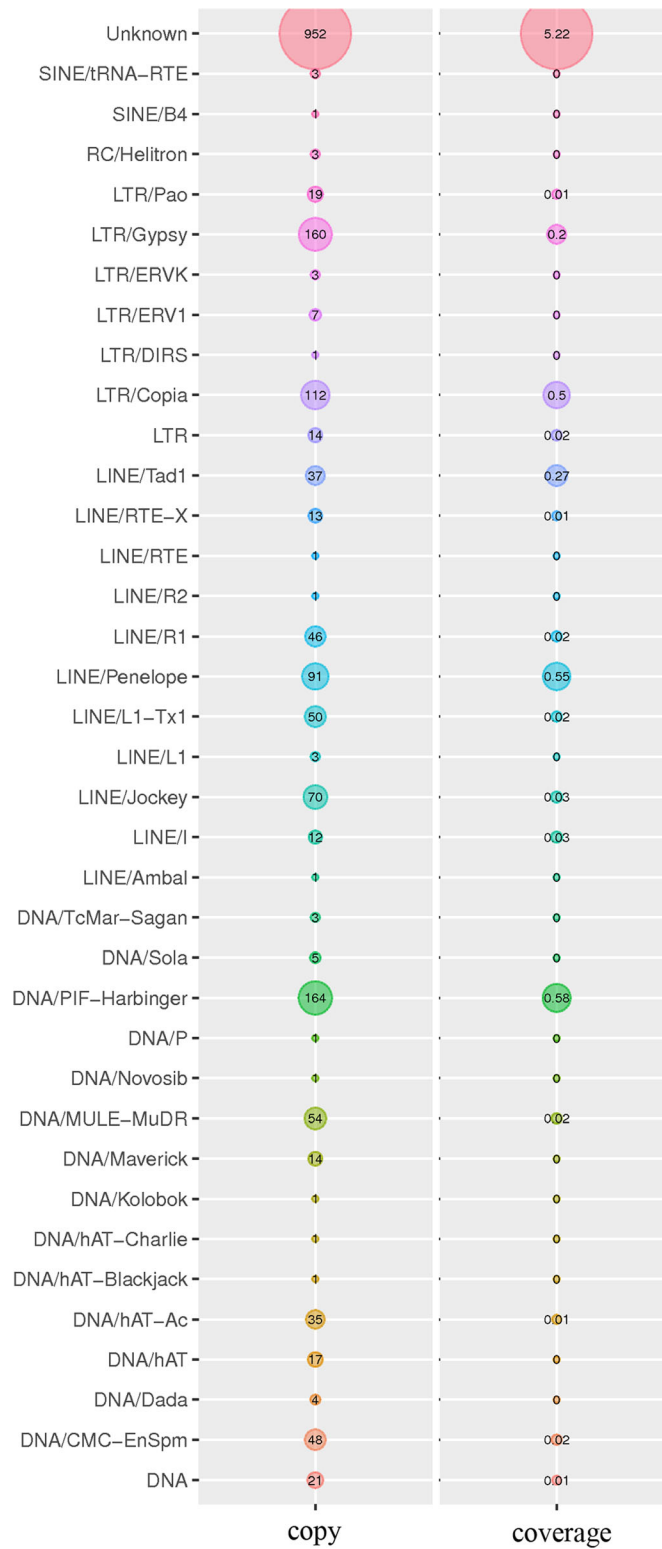


Fig. 2 Distribution of transposable element (TE) families in the assembled *R. glutinis* strain ZHK genome. Copy: total copy number per TE family; Coverage: TE coverage (%) in whole genome assemblies. The TE types and TE family names are listed at left

Table 1 Lipid metabolites were identified using shotgun lipidomics profiling in *R. glutinis* ZHK

| Lipids name | POS | NEG | Percentage (%) |
|---------------------------------------------------|-----|-----|----------------|
| Triacylglycerols (TAGs) | 325 | – | 33.09 |
| Diacylglyceryltrimethylhomo-Ser (DGTS) | 95 | – | 9.67 |
| Phosphatidylcholine (PC) | 38 | 24 | 6.31 |
| Phosphatidylethanolamine (PE) | 28 | 36 | 6.52 |
| Diacylglycerols (DAG) | 27 | – | 2.75 |
| Apocarotenoids (ACar) | 20 | – | 2.04 |
| Phosphatidylmethanol (PMeOH) | 8 | 3 | 1.12 |
| Phosphatidyl glycerol (PG) | 6 | 12 | 1.83 |
| Sulfatide (SHexCer) | 3 | 23 | 2.65 |
| Phosphatidylserine (PS) | 3 | 6 | 0.92 |
| Lysophosphatide (LPE) | 3 | 3 | 0.61 |
| Lyso-phosphatidylcholine (LPC) | 2 | 2 | 0.4 |
| Hexosylceramide (HexCer) | 2 | 40 | 4.28 |
| Ceramide (Cer) | 4 | 60 | 6.52 |
| Sulphoquinovosyldiglyceride (SQDG) | – | 35 | 3.56 |
| Glucuronosyldiacylglycerol (GlcADG) | – | 31 | 3.16 |
| Phosphatidylinositol (PI) | – | 26 | 2.65 |
| Fatty acids (FA) | – | 23 | 2.34 |
| Phosphatidic acid (PA) | – | 16 | 1.63 |
| fatty acid esters of hydroxyl fatty acids (FAHFA) | – | 14 | 1.43 |
| Oxidized phosphatidylethanolamine (OxPE) | – | 8 | 0.81 |
| Acyl-Glucuronosyldiacylglycerol (AcylGlcADG) | – | 6 | 0.61 |
| Oxidized phosphatidylinositol (OxPI) | – | 5 | 0.5 |
| Sphingomyelin (SM) | – | 3 | 0.31 |
| Ganglioside (GM3) | – | 3 | 0.31 |
| Oxidized phosphatidyl glycerol (OxPG) | – | 2 | 0.2 |
| Others | 30 | 7 | 3.78 |
| Total | 594 | 388 | 100 |

glutinis ZHK with its closely related species with available genome sequence, namely *R. graminis* and *R. diobovata*. As shown in Fig. 4, the results revealed that the ZHK genome situates a higher syntenic relationship with *R. graminis* than *R. diobovata*. There are 9737 collinear blocks between the whole genome sequence of *R. glutinis* ZHK and *R. graminis*, the proportion of the total base length in collinear blocks account for their total gene length are higher than 75%. Moreover, a total of 11413 collinear blocks was identified between *R. glutinis* ZHK and *R. diobovata*. The full length of collinear blocks accounts for 19.82% and 20.7% in the whole genome of *R. glutinis* ZHK and *R. diobovata*, respectively.

Comparative analysis of protein families and genes

Here, a total of 6774 protein-coding genes have been predicted in the ZHK genome, and the top four *Rhodotorula* species of these protein-coding genes annotated

into Nr database were *R. toruloides*, *R. graminis*, *R. diobovata* and *R. taiwanensis*. Hence, we performed a comparative genomics analysis between *R. glutinis* ZHK with the four relative closely related species. As shown in Fig. 5a, we compared the distribution of predicted genes among the five oleaginous red yeasts. To investigate the species-specific gene/protein families, the pairwise comparisons have been carried out via a series of BLASTX searches within the five oleaginous red yeasts. As shown in Fig. 5b, we identified 11915 protein families regarding the similarities between gene sequences from the five yeast species (6601 families for *R. glutinis* ZHK, 8020 families for *R. toruloides*, 7135 families for *R. graminis*, 7652 families for *R. diobovata*, and 6873 families for *R. taiwanensis*). Of which, 435 (451 genes), 1467 (1474 genes), 682 (723 genes), 1143 (1247 genes), and 886 (886 genes) protein families were species-specific in *R. glutinis* ZHK, *R. toruloides*, *R. graminis*, *R. diobovata* and *R. taiwanensis*, respectively.

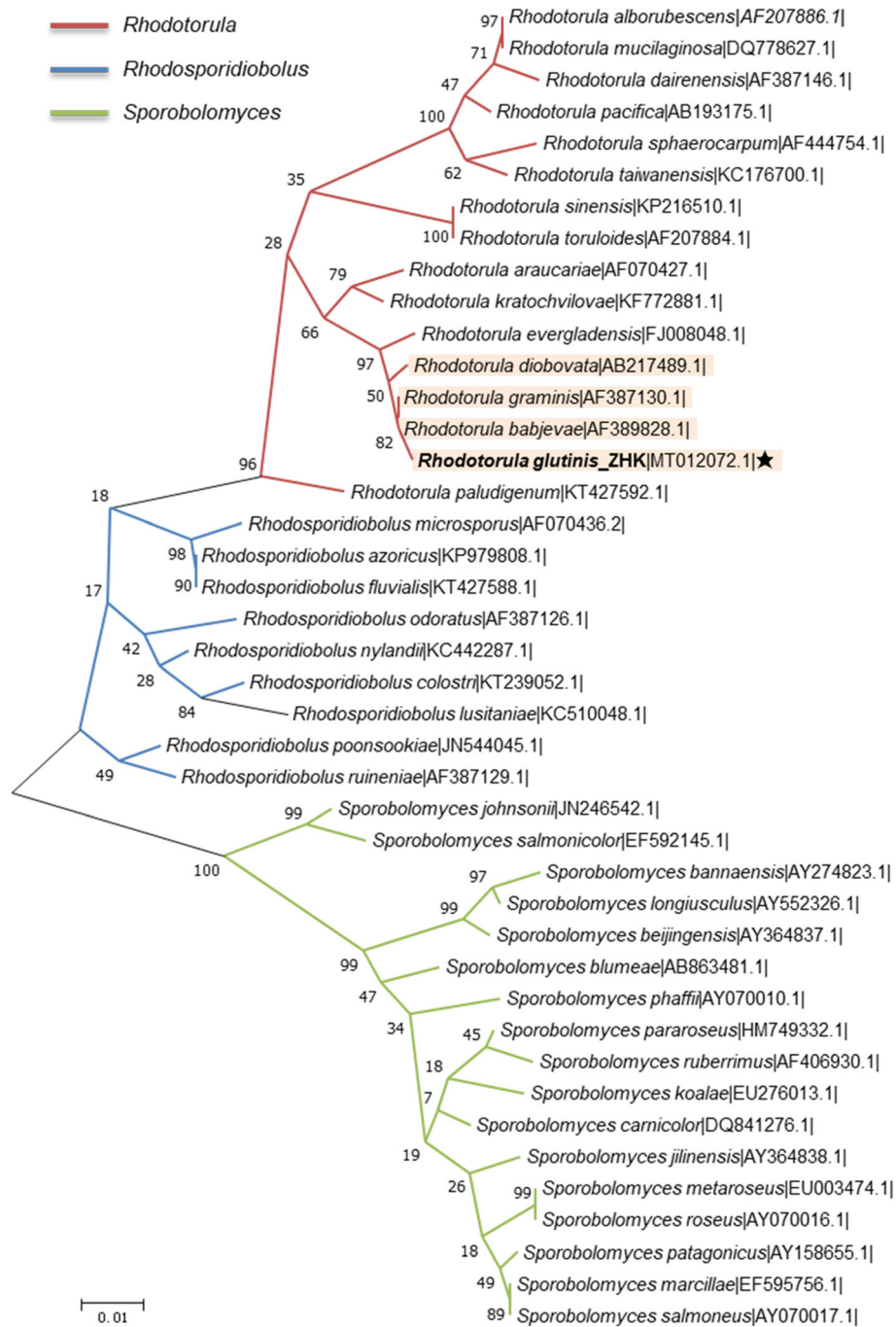
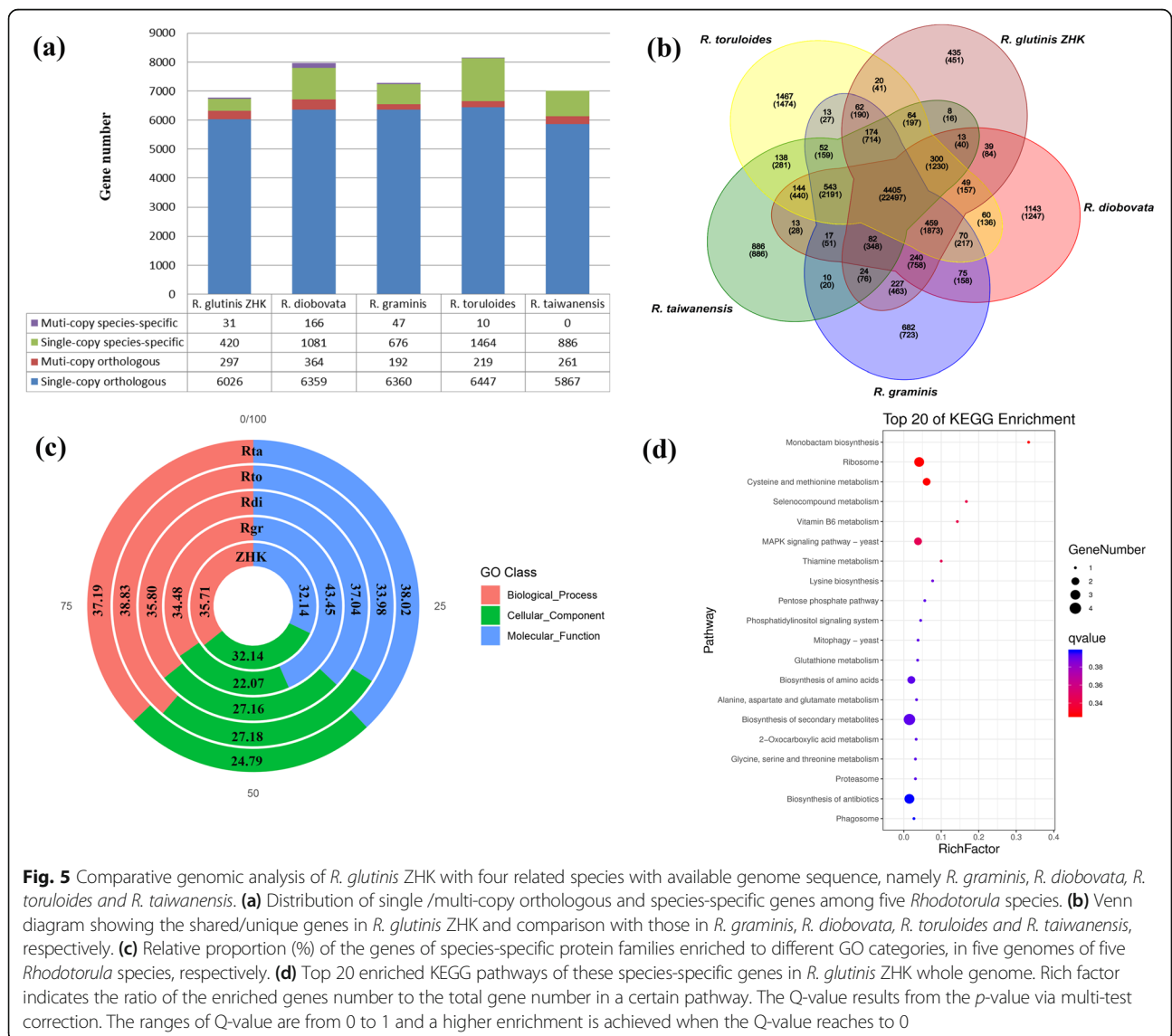
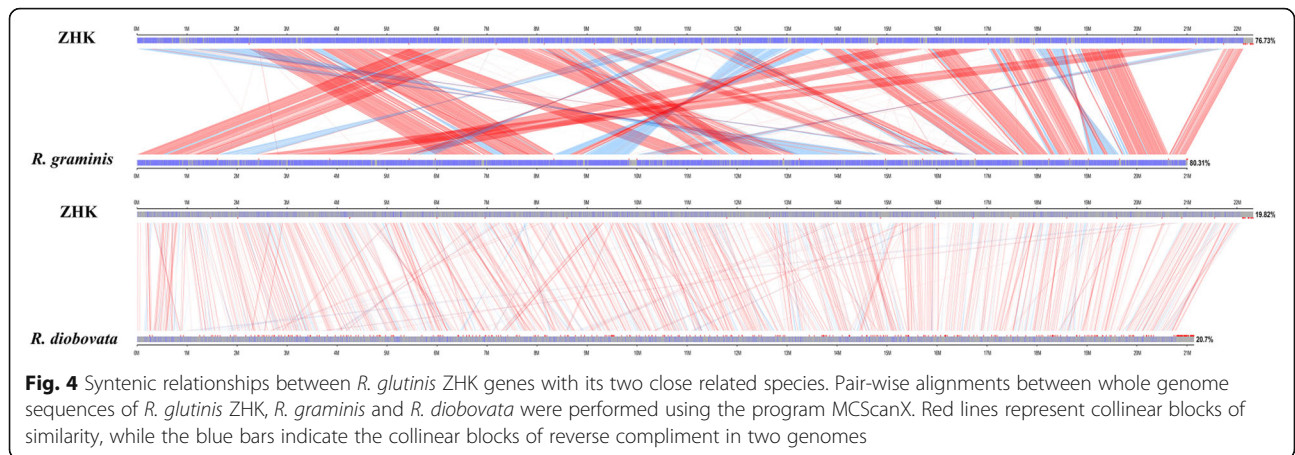


Fig. 3 The phylogeny and evolutionary analysis between the 42 red yeast species of order Sporidiobolales. The phylogenetic tree was constructed using the software MEGA 7.0 based on the alignment of the 26S rDNA sequences combined with the Neighbor-joining method and Bootstrap analysis of 1000 replicates. The strain *R. glutinis* ZHK front is marked in bold. The strain ZHK and its closely related species are highlighted with pink area. The number at each branch of phylogenetic tree indicates the bootstrap value (1000 replicates). Red, blue, and green lines indicate the different clusters of *Rhodotorula*, *Rhodosporidiobolus* and *Sporobolomyces* isolates, respectively

In order to better understand the functional classification of these species-specific genes, we conducted the GO and KEGG enrichment analysis. As shown in Fig. 5c, we performed the GO analysis using their respective species-specific genes of the five oleaginous red yeasts.

Among the species-specific genes of the *R. glutinis* ZHK (Additional file 7), 49 (37.98%), 26 (20.16%) and 54 (41.86%) GO terms were enriched in three categories Biological Process (BP), Molecular Function (MF), and Cellular Component (CC), respectively. It was found that



the significantly enriched GO terms of the species-specific genes in *R. glutinis* ZHK, including BP: metabolic process (GO:0008152), organic substance metabolic process (GO:0071704), cellular process (GO:0009987), single-organism cellular process (GO:0044763) and single-organism process (GO:0044699); MF: catalytic activity (GO:0003824), binding (GO:0005488), hydrolase activity (GO:0016787), small molecule binding (GO:0036094), and heterocyclic compound binding (GO:1901363); CC: intracellular (GO:0005622), cell (GO:0005623), organelle (GO:0043226), membrane (GO:0016020), and membrane-bounded organelle (GO:0043227). After that, we carried out the KEGG pathway mapping using the species-specific genes of the *R. glutinis* ZHK. As shown in Fig. 5d, the significantly enriched KEGG pathways of the species-specific genes in *R. glutinis* ZHK mainly contain biosynthesis of secondary metabolites (ko01110), ribosome (ko03010), biosynthesis of antibiotics (ko01130), cysteine and methionine metabolism (ko00270), MAPK signaling pathway–yeast (ko04011), and biosynthesis of amino acids (ko01230).

In order to investigate the evolutionary dynamics, we identified 3952 one-to-one orthologous genes between *R. glutinis* ZHK and its four closely related *Rhodotorula* species (Additional file 8). The K_a and K_s calculation of the single-copy orthologous genes is widely regarded as an indicator of selective pressure during biological evolution [39]. In order to evaluate the general variation in the selective restriction within the five *Rhodotorula* species at gene levels, we calculated the substitution rate (K_a/K_s) for each orthologous gene using the free ratio model. The results showed that the most of these orthologous genes exhibit a relative low substitution rate ($K_a/K_s < 0.5$), which indicates that these orthologous genes have been retained via a series of strong purifying natural selection. Of which, we found that there are 1 pairs with a K_a/K_s value > 1.0 (strong positive selection), 3 pairs with a K_a/K_s value between 0.5 and 1.0 (positive selection), 621 pairs with a K_a/K_s value between 0.1 and 0.5 (weak positive selection), and 3327 pairs with a K_a/K_s value < 0.1 (purifying selection). The categories of KEGG_B_class enriched among the positively selected genes contain “Amino acid metabolism”, “Folding, sorting and degradation”, and “Translation”. Meanwhile, the categories of KEGG_B_class enriched among the purifying selected genes mainly include “Amino acid metabolism”, “Biosynthesis of other secondary metabolites”, “Carbohydrate metabolism”, “Energy metabolism”, “Folding, sorting and degradation”, “Glycan biosynthesis and metabolism”, “Lipid metabolism”, “Metabolism of cofactors and vitamins”, “Metabolism of terpenoids and polyketides”, “Transport and catabolism”, “Translation”, “Signal transduction”, “Replication and repair”, and “Nucleotide metabolism” (Additional file 9).

Discussion

R. glutinis, one of the most representative species of genus *Rhodotorula*, was recognized as a ubiquitous yeast ranging from contrasting ecosystems such as marine, soil, lake, plant leaf and even polar ice. *R. glutinis* is versatile oleaginous red yeast capable of producing several valuable compounds including microbial lipids, carotenoids and enzymes and therefore has been regarded as a promising host for bio-refinery. Notwithstanding a considerable amount of literature has documented the industrial applications of lipids, carotenoids and enzymes production by *R. glutinis*, and the molecular basis unraveling the biosynthetic mechanism of these valuable compounds still remains largely limited. Mainly because the genomics backgrounds of this species has not yet been studied. In addition, the understanding of the lipid composition is of significant foundation to genetically enhance microbial oils. The advanced Next-Generation Sequencing and HPLC-MS-based metabolomics technology have been widely used to study genetic and metabolic systems of microorganisms without available genome information.

Here, we firstly present the results of whole-genome sequencing and shotgun lipidomics profiling of *R. glutinis* ZHK, to identify its genomics features and intracellular lipid species. Lipids, generally including phospholipids, sphingolipids, fatty acids, sterols, and triacylglycerols (TAGs), are important biomolecules for all biological vitalities [40]. Currently, the issue of excessive consumption of non-renewable fuels has brought about some concerns of energy crisis and environmental pollution worldwide. Microbial lipids are also used as substrates in the third generation biodiesel feedstock, and possess advantages over the first and second generation biofuels [41]. Many oleaginous red yeasts of order Sporidiobolales produce lipids, especially TAGs [42]. TAGs, a group of neutral lipids, are usually used as the food additives, feed supplement and feedstock for chemical syntheses [43]. Oleaginous red yeasts are classified as strains that accumulate high lipids content, and therefore, considered as the potential oil resources for renewable biodiesel feedstock [42]. Furthermore, because of the advantages over the conventional resources, the lipids biosynthesis from oleaginous red yeasts has attracted increasing attention recently. Therefore, using metabolic engineering strategies to enhance lipids production is of great significance for economic and ecological sustainable development. These strategies are usually divided into following orientations that are directly or indirectly related to the biosynthesis of fatty acid and TAG [42]: 1) overexpression of key enzyme genes; 2) transcriptional regulation of bypass pathways; 3) restriction of competing pathways. As a type species of oleaginous yeast, *R. glutinis* is a robust platform organism for these lipids production, because it's high biomass and

multiple substrate availability. We found some candidate key enzyme genes which are involved in lipid metabolism in the ZHK genome. Some of key enzymes involved in the lipid metabolism, have been documented, such as ACC (acetyl-CoA carboxylase), AOX (acyl-CoA oxidase), PDAT (phospholipid: diacylglycerol acyltransferase), and GPDH (glycerol 3-phosphate dehydrogenase). ACC catalyzes acetyl-CoA to form malonyl-CoA; AOX catalyzing the first step in the pathway of fatty acid β -oxidation; PDAT catalyzes the acyl-CoA-independent synthesis of cholesterol esters; GPDH provides the activated glycerol backbone for TAG synthesis. Furthermore, we also found some genes related to lipases biosynthesis. Because of these lipases are closely related to lipid metabolism [44, 45], and therefore, also vital for the production of engineered lipid in *R. glutinis* ZHK. These genes could be the promising targets for genetic manipulation to enhance the production of lipid metabolites.

Additionally, the total carotenoid contents of *R. glutinis* ZHK were determined qualitatively and quantitatively. Torularhodin and torulene as the dominated carotenoid contents of *R. glutinis* ZHK collectively constitute 80.71% of total carotenoids, while the β -carotene only accounts for 19.29%. Torulene and torularhodin represent two of the principal carotenoids in *R. glutinis* strains and exhibit the similar chemical structure to that of super antioxidant lycopene. The earliest literature reported that the presence of torulene and torularhodin in microbial cells could be traced back to the 1930s and 1940s, respectively. But only in the last few decades, the amount of literature picturing their properties has gradually increased. Previous studies revealed that both of torularhodin and torulene possess considerable strong properties, such as anti-oxidative, anti-cancerous, antimicrobial and food safety. In order to realize the commercial application of these two carotenoids, it is essential to obtain highly efficient yeast strains of *R. glutinis*. However, until now, there have been no *R. glutinis* strains capable of producing torularhodin in high-yields required for industrial scale use. The rapid development of the gene editing and synthetic biology approaches allows us to construct an engineering strain that overproduces torularhodin and torulene. Moreover, the general pathways for carotenoid synthesis of oleaginous yeasts have proposed previously. However, the precise nature of the coding-genes involved in the bioconversion from torulene to torularhodin in these fungi still remains unclear. It is a bottleneck that currently blocks the industrial development of microbial torulene and torularhodin.

Generally, the bifunctional lycopene cyclase/phytoene synthase is capable of catalyzing 3,4-didehydrolycopene to form torulene in *Neurospora crassa* [46]. The 16'-hydroxytorulene could be regarded as an intermediate

product of the biosynthetic pathway from torulene to torularhodin in the red yeasts *Cystofilobasidium infirmominatum* [47]. Nevertheless, the enzymes involved in the transformation from torulene to torularhodin in this species have not been elucidated. Based on previous studies, as shown in Fig. 6, we propose the bioconversion process from torulene to torularhodin is as follows: 1) the carotene hydroxylase (CrtZ) catalyzes the hydroxylation of torulene to form 16'-hydroxytorulene; 2) the carotene ketolase (CrtA) or monooxygenase (CrtO) or both of them catalyzes the carboxylation of 16'-hydroxytorulene to form torularhodin. In the ZHK genome, we found some candidate genes (Table 2) putatively encoding GGPP synthase, lycopene cyclase/phytoene synthase, phytoene desaturase, hydroxylase, monooxygenase and ketolase, which may be related to the carotenoids biosynthesis. Interestingly, the comparative genomics results showed that these genes encoding GGPP synthase, phytoene desaturase and lycopene cyclase/phytoene synthase, has underwent the strong purifying natural selection within the five *Rhodotorula* species. Subsequently, we would like to further verify the function of these genes through a group of heterologous complementary experiments as described in our previous studies [48, 49]; hope to finally fill in the gaps in the synthetic pathway of torulene and torularhodin. Besides, the *R. glutinis* strains are regarded as the resources of various kinds of industrial enzymes, especially the phenylalanine ammonia lyase (PAL) [25]. The PAL is usually considered as the key limiting-enzyme in the biosynthesis of phenylpropanoids and flavones [50, 51]. Particularly, PAL has been used as an enzyme substitution therapy in medicine to treat phenylketonuria (PKU) [52]. Our genomics results reported here should be helpful for the further understanding the biosynthesis of PAL in *R. glutinis* ZHK. In brief, the present study will lay a theoretical foundation for increasing lipids, carotenoids, enzymes, and other bio-products production in *R. glutinis* strains, which would have an immense industrial significance and application.

Conclusions

Here, the high-quality genome of *R. glutinis* ZHK was reported. It constructed a genetic basis for further research on its lipids, carotenoids, and industrial enzymes metabolism. Moreover, a total of 982 intracellular lipid species were identified in ZHK by lipidomic profiling, mainly including TAGs, DGTS and PE. In conclusion, our work provides new insights into these candidate genes and metabolites with potential biotechnological applications in *R. glutinis* ZHK. Furthermore, our results also lay the foundation for *R. glutinis* ZHK as a microbial cell factory to produce engineered compounds and facilitate the comparative genomics studies to elucidate evolutionary dynamics of the *Rhodotorula* species.

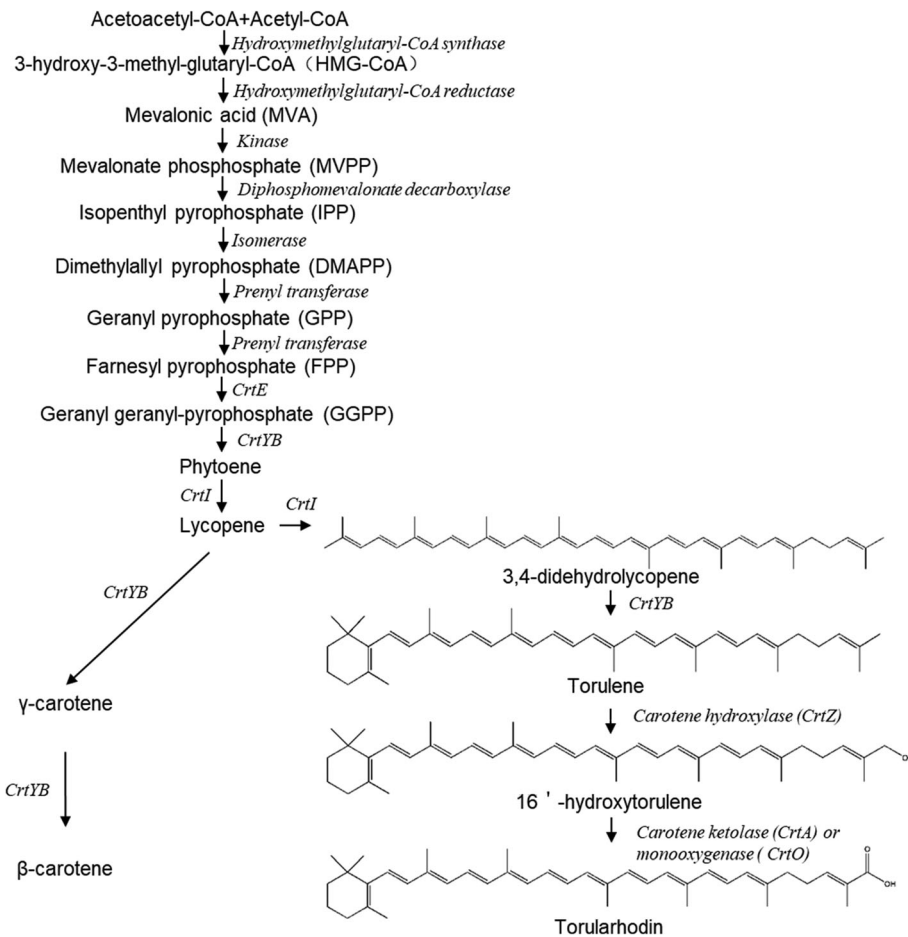


Fig. 6 Schematic of the proposed biosynthetic pathways from acetyl-CoA to torularhodin and β-carotene in *R. glutinis* ZHK. These genes-encoding proteins annotated by orthology for enzymes which can catalyze pathway steps. Carotenogenic pathway: CrtE: GGPP synthase, CrtYB: bifunctional lycopene cyclase/phytoene synthase, CrtI: phytoene desaturase. Metabolites are shown in bold, while the enzymes responsible for relative bioconversions are indicated in italics at corresponding arrows

Table 2 The candidate genes putatively encoding CrtE, CrtYB, CrtI, CrtZ, CrtO and CrtA in *R. glutinis* ZHK genome

| Gene name | Gene ID | Description | Gene number |
|------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|-------------|
| GGPP synthase | R_glutinis_G00415 | <i>crtE</i> | 1 |
| lycopene cyclase/phytoene synthase | R_glutinis_G00924 | <i>crtYB</i> | 1 |
| phytoene dehydrogenase | R_glutinis_G00921 | <i>crtI</i> | 1 |
| hydroxylase | R_glutinis_G02426 R_glutinis_G04459 R_glutinis_G05828 R_glutinis_G04456 R_glutinis_G04508 R_glutinis_G01719 | <i>crtZ</i> | 6 |
| monooxygenase | R_glutinis_G03787 R_glutinis_G05690 R_glutinis_G04646 R_glutinis_G00553 R_glutinis_G05814 R_glutinis_G03223 R_glutinis_G04524 R_glutinis_G03006 | <i>crtO</i> | 8 |
| ketolase | R_glutinis_G03471 R_glutinis_G01624 | <i>crtA</i> | 2 |

Methods

Yeast strain and growth conditions

In this study, the *Rhodotorula glutinis* strain ZHK was isolated from water of the Pearl River (23°06'N, 113°17'E), in Guangzhou City, Guangdong Province of China. Complete species identification was performed through morphology and molecular characterization, the GenBank Accession number of *R. glutinis* ZHK 26S rDNA is MT012072. The strain *R. glutinis* ZHK is cultured in 250 mL Erlenmeyer baffle flasks containing 50 mL of the YEPD medium (yeast extract: 5 g/L, peptone: 10 g/L, dextrose: 10 g/L, pH 6.5). After inoculation with the pre-cultured cell suspension of *R. glutinis* ZHK, cultures were transferred into flasks and incubated at 28°C on a rotary shaker at 180 rpm for 72 h. Fresh cells of *R. glutinis* ZHK were harvested by centrifugation (3000×g, 4°C, 10 min), immediately frozen in liquid nitrogen, and then stored at -80°C refrigerator for further extraction of genomic DNA, total RNA and metabolites.

DNA extraction and genome sequencing

Genomic DNA was extracted using Universal Genomic DNA Kit (Cowin Bio., Beijing, China) according to the manufacturer's instructions. The extracted DNA quality was detected using Qubit 2.0 Fluorometer (Life Technologies, USA) and Nanodrop (Thermo Fisher Scientific, USA) accordingly. The obtained genomic DNA with high-quality (≥ 100 ng/ μ L) was used for whole-genome sequencing through both the single molecule real time (SMRT) sequencing platform (Pacific Biosciences, USA) and the paired-end sequencing platform (Illumina, USA). Qualified genomic DNA was fragmented with G-tubes (Covaris, USA) and randomly end-repaired to prepare SMRTbell DNA template libraries (with an average fragment size of 30 kb) according to the manufacturer's specification. Library quality was detected by Qubit 2.0 Fluorometer and average fragment size was estimated on a Bioanalyzer 2100 (Agilent Technologies, USA). SMRT sequencing was performed on the PacBio Sequel sequencer according to standard protocols (MagBead Standard Seq v2 loading, 1×180 min movie). Subsequently, the qualified genomic DNA was also used to construct a paired-end library with insert sizes of 300 bp using the Paired—End Genomic DNA Sample Prep Kit (Illumina, USA). These paired-end libraries were sequenced using the HiSeq 2500 sequencer with PE150 strategies according to the manufacturer's protocols.

Genome assembly using PacBio and Illumina data

Firstly, the PacBio long-reads was corrected for random errors in the long seed reads by aligning shorter reads from the same library using the program MECAT (<https://github.com/xiaochuanle/MECAT>) (version 1.0) as described previously [53, 54]. The corrected long

reads were used for *de novo* genome assembly to generate contigs using the mecat2canu module of MECAT with an overlap-layout-consensus (OLC) strategy (parameters: min overlap length = 500; min read length = 1000) [53, 55]. After that, the Illumina paired-end short reads were used for polishing the resulting contigs through the program Pilon (version 1.22) with default parameters [56]. Finally, these polished contigs were further assembled to generate scaffolds using genomic synteny analysis (Additional file 10) with the closely related species *Rhodotorula graminis* [57]. Genome completeness of assembly was evaluated using the program BUSCO (version 3.0.1) with parameters: reference gene set of basidiomycota_odb9 [58].

Phylogenetic and syntenic analysis

Phylogenetic analysis was performed by comparing the *R. glutinis* ZHK 26S rDNA sequence to other related species of the order Sporidiobolales. All 26S rDNA sequences were obtained from the NCBI Nucleotide database (<https://www.ncbi.nlm.nih.gov/nucleotide>). MEGA 7.0 was used for processing these 26S rDNA sequences using Muscle alignment with UPGMB clustering method [59]. Phylogeny was tested using Bootstrap method with default parameters. Bootstrap values expressed as percentages of 1,000 replications were shown at the branching points. Phylogenetic tree was constructed using these aligned sequences by the Neighbor-Joining method in the software MEGA 7.0. Candidate species for syntenic analysis were selected based on the results of phylogenetic analysis. Syntenic analysis between the genome sequences of *Rhodotorula* species were performed by the program MCSanX with default parameters [60].

RNA extraction and transcriptome sequencing

The extraction of total RNA was achieved with the TRIzol Kit (Invitrogen, USA). RNA quality was evaluated using the Bioanalyzer 2100 (Agilent Technologies, USA) and agarose gel electrophoresis (RNase-free) [61]. The high-quality RNA samples were used to construct cDNA libraries for transcriptome sequencing. The workflow of cDNA library construction was performed as described previously [61, 62]. After that, the cDNA library was sequenced using HiSeq 2500 sequencer (Illumina, USA) to generate 150 bp paired-end reads. Raw reads were processed and further filtered through the FASTP program (version 0.18.0) to remove reads with following characters [63]: 1) containing adapters; 2) containing more than 10% of unknown nucleotides (N); 3) containing more than 50% of low quality (Q value ≤ 20) bases. After that, a reference genome index was built, and these paired end clean reads were mapped to the assembled *R. glutinis* ZHK genome using the program HISAT (version 2.2.4) with parameters: -rna-strandness

RF and others set as default [64]. Mapped reads were assembled by using the software StringTie (version 1.3.1) in a reference-based approach [65]. For each transcription region, we used the software StringTie (version 1.3.1) (with default parameters) to quantify expression abundance and variations using a fragment per kilobase of transcript per million mapped reads (FPKM) method [66].

Gene prediction and functional annotation

The open reading frame (ORF) was predicted using the program GeneMark-ES with default parameters (Additional file 11) [67]. Integration of mapped RNA-seq reads with genomic data to predict gene models was performed using the program GeneMark-ET with default parameters [68]. After using the program Stringtie to reconstruct these transcripts, the genes were expressed in the RNA-seq results but not included in the genome to be defined as the novel genes. Repetitive elements were identified by RepeatMasker (version v4.0.7) with default parameters [69]. Noncoding RNAs, such as rRNAs prediction was carried out using the program rRNAmmer (version 1.2) with default parameters and tRNAs were identified by the program tRNAscan-SE (version 2.0.4) with default parameters [70]. Functional annotation of the predicted genes was performed by aligning with diverse public databases [71], such as National Center for Biotechnology Information (NCBI) non-redundant protein (Nr) database, UniProt/Swiss-Prot, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Cluster of Orthologous Groups of proteins (COG), and Protein families (Pfam) (Additional file 12).

Identification of orthologous genes

Annotation information of coding sequences and proteins of *R. graminis*, *R. diobovata*, *R. toruloides*, and *R. taiwanensis* were downloaded from the NCBI Genome database (<https://www.ncbi.nlm.nih.gov/genome/>). Genome sequence alignments were performed in an all-against-all comparison using the MUMmer 3 package (<http://mummer.sourceforge.net/>, version 3.2.2) with default parameters [72]. Comparative genomics analysis of this study was performed at protein level. Software OrthoMCL (version 2.0) (<https://orthomcl.org/orthomcl/>) was used to generate core-orthologs for the *R. glutinis* ZHK, *R. graminis*, *R. diobovata*, *R. toruloides*, and *R. taiwanensis* proteomes datasets with default parameters [73]. Subsequently, all the putative proteins of the five yeast species and core-orthologs were aligned (all against all) using BLASTP (<http://www.ncbi.nlm.nih.gov/BLAST/>) and a score for each pair of proteins with significant matches was assigned with a cut-off value of 1×10^{-7} . GO and KEGG enrichment analyses of orthologous and species-specific genes were performed using DAVID

functional annotation tool (<https://david.ncifcrf.gov/tools.jsp>, version 6.8) with default parameters [74].

Substitution rate estimation

The substitution rates Ka/Ks (the number of nonsynonymous substitutions per nonsynonymous site to the number of synonymous substitutions per synonymous site) were averaged for all of the pairwise comparisons of each single-copy orthologous gene, using the free ratio Model of the software KaKs_Calculator Toolbox (version 2.0) with default parameters [75, 76]. These genes showed a higher Ka/Ks value (Ka/Ks > 1 or 0.5 and *p*-value < 0.05) were suggested to be the positively selected genes [77].

Lipidomics analysis

To perform the lipid profiling, the methyl tert-butyl ether (MTBE) method was used for lipid metabolites extraction [78]. Briefly, there are about 1×10^7 cells from three growth periods (24 h, 48h, and 72h) of each sample, which was added to 450 μ L of extraction liquid ($V_{\text{MTBE}}: V_{\text{methanol}} = 5: 1$) and further vortexed for 30 s. The mixture was then centrifuged at $3000 \times g$, 4°C, for 15 min. The supernatant (organic layer) was transferred to a clean vial and evaporated to dryness in a vacuum concentrator. Dried extract was reconstituted in 100 μ L dichloromethane/methanol liquid (1:1, v/v). 60 μ L liquid extract of each sample was collected for the following analysis. Lipid metabolites profiling was performed by an Ultra-high-performance liquid chromatography system (UHPLC 1290 series, Agilent Technologies, USA) with a C_{18} column (1.7 μ m, 100 mm \times 2.1 mm, 100 Å) (Phenomen Kinetex, USA) coupled with a quadruple time-of-flight mass spectrometer (Triple TOF 6600, AB SCIEX, USA). The acquisition software Analyst TF (version 1.7.1, AB SCIEX, USA) continuously evaluated the full scan survey MS data with default parameters. Detailed settings of the UHPLC-MS system were used as described previously [79]. Raw data files acquired from UHPLC-MS/MS were transformed into the mzXML format using the program ProteoWizard (version 3.0.4472) with default parameters [80], and further analyzed by R package XCMS (version 3.2) with default parameters [81, 82]. Peak annotation of the XCMS pre-processed data was performed using R package CAMERA with default parameters [82, 83]. Metabolites were identified through matching the acquired MS/MS spectral against the LIPID MAPS® (<https://www.lipidmaps.org/>) and an in-house standard MS/MS database (Biotree Biotech Co., Ltd., Shanghai, China) with the parameters ($|m/z \text{ errors}| < 25$ ppm, cutoff of match score=0.6, and minfrac=0.5) as described previously [79]. To ensure the accuracy of lipidomics data, the quality control (QC) samples were used for data evaluation [84]. All lipidomics analysis was performed with three independent biological replicates.

Carotenoids extraction and quantification

Carotenoids were extracted at 65°C for 30 min from freeze-dried *R. glutinis* ZHK cells using DMSO (dimethylsulfoxide)-acetone (1/3, v/v) [49]. After centrifugation of 6000×g, the pigmented supernatant (organic layer) was pipetted off and the extraction was repeated until it became entirely colorless. High performance liquid chromatography (HPLC) analysis of carotenoids was performed on an Agilent 1100 series system (Agilent Technologies, USA). A reverse-phase C₁₈ column (5 μm, 150×4.6 mm) (Thermo Fisher Scientific, USA) was used for the separation of carotenoid extracts. HPLC analysis in an isocratic elution system: acetonitrile-methanol-isopropanol (85:10:5, v/v/v) as the mobile phase at a column temperature of 32°C; an injection volume of 20 μL; a flow rate of 1.0 mL/min; the UV-visible spectra were obtained at 450 nm (for β-carotene), 484 nm (for torulene), and 507 nm (for torularhodin), respectively. Carotenoid quantification was performed with three independent biological replicates. Standard substance of β-carotene was bought from Sigma-Aldrich (St. Louis, MO, USA). Standards of torularhodin and torulene were purchased from CaroteNature GmbH (Münsingen, Switzerland).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-020-07244-z>.

Additional file 1. Genome completeness analysis through aligning the orthologs of the ZHK to a reference gene set of basidiomycota_odb9.

Additional file 2. All genes expression profile resulted from RNA-seq analysis.

Additional file 3. SNP/InDel annotations resulted from RNA-seq data.

Additional file 4. Gene structure optimization resulted from RNA-seq data.

Additional file 5. All the genes involved in the biosynthesis of antibiotics in the ZHK genome.

Additional file 6. Comprehensive information of all identified lipid metabolites.

Additional file 7. Annotations of species-specific genes of the ZHK.

Additional file 8. One-to-one orthologous genes between *R. glutinis* ZHK and its close related species *R. toruloides*, *R. graminis*, *R. diobovata* and *R. taiwanensis*.

Additional file 9. Ka/Ks results of single-copy orthologous genes.

Additional file 10. Genomic synteny analysis between the ZHK and its closely related species *Rhodotorula graminis*.

Additional file 11. CDS/cDNA sequences resulted from genome analysis.

Additional file 12. Annotations of all gene resulted from assembled ZHK genome.

Abbreviations

BP: Biological Process; CC: Cellular Component; GO: Gene Ontology; HPLC: High Performance Liquid Chromatography; KEGG: Kyoto Encyclopedia of Genes and Genomes; COG: Cluster of Orthologous Groups of proteins; MF: molecular function; MS: Mass Spectrum; Nr: NCBI's non-redundant protein database; Pfam: Protein Families; Rfam: RNA families

Acknowledgements

We thank the Guangzhou Genedenovo Biotechnology Co. Ltd. for assisting in bioinformatics analysis.

Authors' contributions

CJL conducted the experiments, analyzed the data, contributed yeast materials and analysis tools, wrote and revised the manuscript. DZ provided multiple technical supports in bioinformatics analysis. CJL conceived the study ideas and designed the experiments. PC and GHY supervised the overall written process of the paper. LZ revised the draft manuscript. All authors have approved the final manuscript.

Funding

This work was supported by the Starting Research Fund from Zhongkai University of Agriculture and Engineering (Grant No. KA190577885 & KA200540843). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All the raw sequence data are available via GenBank under the SRA accessions SRR11648405 (Illumina paired-end raw data of RNA-seq), SRR11637747 (Illumina paired-end raw data of genome sequencing) and SRR11611234 (PacBio long-read raw data of genome sequencing). The *R. glutinis* strain ZHK Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAAGPT000000000. The version described in this paper is version JAAGPT010000000.

Ethics approval and consent to participate

The strain ZHK used in this study was obtained from river water, and it is not an endangered species. The collection of the fungal material was complied with institutional and national guidelines of China.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Innovative Institute for Plant Health, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, People's Republic of China. ²College of Agriculture and Biology, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, People's Republic of China. ³National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, People's Republic of China.

Received: 27 May 2020 Accepted: 18 November 2020

Published online: 26 November 2020

References

- Gadanh M, Sampaio JP. Polyphasic taxonomy of the basidiomycetous yeast genus *Rhodotorula*: *Rh. glutinis* sensu stricto and *Rh. dairenensis* comb. nov. *FEMS Yeast Res.* 2002;2(1):47–58.
- Sampaio JP: Chapter 155 - *Rhodotorula* Harrison (1928). In Kurtzman CP, Fell JW, Boekhout T (eds): the yeasts (5th edition). In. London: Elsevier; 2011: 1873–1927.
- Tang W, Wang Y, Zhang J, Cai Y, He Z. Biosynthetic pathway of carotenoids in *Rhodotorula* and strategies for enhanced their production. *J Microbiol Biotechnol.* 2019;29(4):507.
- Lyman M, Urbin S, Stroutand C, Rubinfeld B. The oleaginous red yeast *Rhodotorula/Rhodospiridium*: a factory for industrial bioproducts: yeasts in biotechnology. In. London: IntechOpen; 2019.
- Kuan I, Kao W, Chen C, Yu C. Microbial biodiesel production by direct transesterification of *Rhodotorula glutinis* biomass. *Energies.* 2018;11(5):1036.
- Martínez JM, Schottroff F, Haas K, et al. Evaluation of pulsed electric fields technology for the improvement of subsequent carotenoid extraction from dried *Rhodotorula glutinis* yeast. *Food Chem.* 2020;323:126824.
- Gong G, Gong G, Liu L, et al. Multi-omics metabolism analysis on irradiation-induced oxidative stress to *Rhodotorula glutinis*. *Appl Microbiol Biot.* 2019;103(1):361–74.

8. Easterling ER, French WT, Hernandez R, Licha M. The effect of glycerol as a sole and secondary substrate on the growth and fatty acid composition of *Rhodotorula glutinis*. *Bioresour Technol*. 2009;100(1):356–61.
9. Zhang Z, Zhang X, Tan T. Lipid and carotenoid production by *Rhodotorula glutinis* under irradiation/high-temperature and dark/low-temperature cultivation. *Bioresour Technol*. 2014;157:149–53.
10. Mast B, Zöhrens N, Schmidl F, et al. Lipid production for microbial biodiesel by the oleaginous yeast *Rhodotorula glutinis* using hydrolysates of wheat straw and miscanthus as carbon sources. *Waste Biomass Valor*. 2014;5(6):955–62.
11. Kot AM, Błażej S, Kurcz A, et al. Effect of initial pH of medium with potato wastewater and glycerol on protein, lipid and carotenoid biosynthesis by *Rhodotorula glutinis*. *Electron J Biotechnol*. 2017;27:25–31.
12. Latha BV, Jeevaratnam K, Murali HS, Manja KS. Influence of growth factors on carotenoid pigmentation of *Rhodotorula glutinis* DFR-PDY from natural source. *Indian J Biotechnol*. 2014;4(3):353–7.
13. Li C, Zhang N, Li B, et al. Increased torulene accumulation in red yeast *Sporidiobolus pararoseus* NGR as stress response to high salt conditions. *Food Chem*. 2017;237:1041–7.
14. Sakaki H, Nochide H, Komemushi S, Miki W. Effect of active oxygen species on the productivity of torularhodin by *Rhodotorula glutinis* no. 21. *J Biosci Bioeng*. 2002;93(3):338–40.
15. Galano A, Francisco-Marquez M. Reactions of OOH radical with β -carotene, lycopene, and torulene: hydrogen atom transfer and adduct formation mechanisms. *J Phys Chem B*. 2009;113(32):11338–45.
16. Du C, Guo Y, Cheng Y, Han M, Zhang W, Qian H. Torulene and torularhodin, protects human prostate stromal cells from hydrogen peroxide-induced oxidative stress damage through the regulation of Bcl-2/Bax mediated apoptosis. *Free Radic Res*. 2017;51(2):113–23.
17. Du C, Li Y, Guo Y, Han M, Zhang W, Qian H. The suppression of torulene and torularhodin treatment on the growth of PC-3 xenograft prostate tumors. *Biochem Bioph Res Co*. 2016;469(4):1146–52.
18. Du C, Li Y, Guo Y, Han M, Zhang W, Qian H. Torularhodin, isolated from *Sporidiobolus pararoseus*, inhibits human prostate cancer LNCaP and PC-3 cell growth through Bcl-2/Bax mediated apoptosis and AR down-regulation. *RSC Adv*. 2015;5:106387–95.
19. Keceli TM, Erginkaya Z, Turkan E, Kaya U. Antioxidant and antibacterial Effects of carotenoids extracted from *Rhodotorula glutinis* strains. *Asian J Chem*. 2013;25(1):42–6.
20. Ungureanu C, Dumitriu C, Popescu S, et al. Enhancing antimicrobial activity of TiO₂/Ti by torularhodin bioinspired surface modification. *Bioelectrochemistry*. 2016;107:14–24.
21. Li J, Liu C, Guo Y, et al. Determination of the effects of torularhodin against alcoholic liver diseases by transcriptome analysis. *Free Radical Bio Med*. 2019;143:47–54.
22. Kot AM, Błażej S, Gientka I, Kieliszek M, Bryś J. Torulene and torularhodin: "new" fungal carotenoids for industry? *Microb Cell Factories*. 2018;17(1):49.
23. Zoz L, Carvalho JC, Socolo VT, Casagrande TC, Cardoso L. Torularhodin and torulene: bioproduction, properties and prospective applications in food and cosmetics - a review. *Braz Arch Biol Technol*. 2015;58(2):278–88.
24. Latha BV, Jeevaratnam K. Thirteen-week oral toxicity study of carotenoid pigment from *Rhodotorula glutinis* DFR-PDY in rats. *Indian J Exp Biol*. 2012;50(9):645–51.
25. Barron CC, Sponagle BJD, Arivalagan P, D Cunha GB. Optimization of oligomeric enzyme activity in ionic liquids using *Rhodotorula glutinis* yeast phenylalanine ammonia lyase. *Enzyme Microb Tech* 2017; 96:151–156.
26. Zhu L, Cui W, Fang Y, Liu Y, Gao X, Zhou Z. Cloning, expression and characterization of phenylalanine ammonia-lyase from *Rhodotorula glutinis*. *Biotechnol Lett*. 2013;35(5):751–6.
27. Zhang H, Wang L, Ma L, et al. Biocontrol of major postharvest pathogens on apple using *Rhodotorula glutinis* and its effects on postharvest quality parameters. *Biol Control*. 2009;48(1):79–83.
28. Li B, Peng H, Tian S. Attachment capability of antagonistic yeast *Rhodotorula glutinis* to *Botrytis cinerea* contributes to biocontrol efficacy. *Front Microbiol*. 2016;7:601.
29. Sen T, Barrow CJ, Deshmukh SK. Microbial pigments in the food industry—challenges and the way forward. *Front Nutr*. 2019;6:7.
30. Braunwald T, Schwemlein L, Graeff-Hönninger S, et al. Effect of different C/N ratios on carotenoid and lipid production by *Rhodotorula glutinis*. *Appl Microbiol Biot*. 2013;97(14):6581–8.
31. Saenge C, Cheirsilp B, Suksaroge TT, Bourtoom T. Potential use of oleaginous red yeast *Rhodotorula glutinis* for the bioconversion of crude glycerol from biodiesel plant to lipids and carotenoids. *Process Biochem*. 2011;46(1):210–8.
32. Kot AM, Błażej S, Kurcz A, Gientka I, Kieliszek M. *Rhodotorula glutinis*—potential source of lipids, carotenoids, and enzymes for use in industries. *Appl Microbiol Biot*. 2016;100(14):6103–17.
33. Paul D, Magbanua Z, Il MA et al. Genome sequence of the oleaginous yeast *Rhodotorula glutinis* ATCC 204091. *Genome Announc* 2014; 2(1):e14–e46.
34. Zhang S, Skerker JM, Rutter CD, Maurer MJ, Arkin AP, Rao CV. Engineering *Rhodospiridium toruloides* for increased lipid production. *Biotechnol Bioeng*. 2016;113(5):1056–66.
35. Sossah F, Liu Z, Yang C, et al. Genome sequencing of *Cladobotryum protrusum* provides insights into the evolution and pathogenic mechanisms of the cobweb disease pathogen on cultivated mushroom. *Genes-Basel*. 2019;10(2):124.
36. Mi S, Shang K, Li X, Zhang C, Liu J, Huang D. Characterization and discrimination of selected China's domestic pork using an LC-MS-based lipidomics approach. *Food Control*. 2019;100:305–14.
37. Sen D, Paul K, Saha C, et al. A unique life-strategy of an endophytic yeast *Rhodotorula mucilaginosa* JGTA-S1—a comparative genomics viewpoint. *DNA Res*. 2019;26(2):131–46.
38. Urbina H, Aime MC. A closer look at Sporidiobolales: ubiquitous microbial community members of plant and food biospheres. *Mycologia*. 2018;110(1):79–92.
39. Feng X, Jia Y, Zhu R, Chen K, Chen Y. Characterization and analysis of the transcriptome in *Gymnocypris selincuoensis* on the Qinghai-Tibetan plateau using single-molecule long-read sequencing and RNA-seq. *DNA Res*. 2019; 26(4):353–63.
40. Zhang X, Liu M, Zhang X, Tan T. Microbial lipid production and organic matters removal from cellulosic ethanol wastewater through coupling oleaginous yeasts and activated sludge biological method. *Bioresour Technol*. 2018;267:395–400.
41. Wang Y, Ho S, Yen H, et al. Current advances on fermentative biobutanol production using third generation feedstock. *Biotechnol Adv*. 2017;35(8):1049–59.
42. Liang M, Jiang J. Advancing oleaginous microorganisms to produce lipid via metabolic engineering technology. *Prog Lipid Res*. 2013;52(4):395–408.
43. Zhu Z, Zhang S, Liu H, et al. A multi-omic map of the lipid-producing yeast *Rhodospiridium toruloides*. *Nat Commun*. 2012;3(1):1112.
44. Pohanka M. Biosensors and bioassays based on lipases, principles and applications, a review. *Molecules*. 2019;24(3):616.
45. Maharana AK, Singh SM. A cold and organic solvent tolerant lipase produced by Antarctic strain *Rhodotorula* sp. Y-23. *J Basic Microb*. 2018; 58(4):331–42.
46. Hausmann A, Sandmann G. A single five-step desaturase is involved in the carotenoid biosynthesis pathway to β -carotene and torulene in *Neurospora crassa*. *Fungal Genet Biol*. 2000;30(2):147–53.
47. Herz S, Weber RWS, Anke H, Mucci A, Davoli P. Intermediates in the oxidative pathway from torulene to torularhodin in the red yeasts *Cystofilobasidium infirmominatum* and *C. capitatum* (Heterobasidiomycetes, fungi). *Phytochemistry*. 2007;68(20):2503–11.
48. Li C, Zhang N, Song J, et al. A single desaturase gene from red yeast *Sporidiobolus pararoseus* is responsible for both four- and five-step dehydrogenation of phytoene. *Gene*. 2016;590(1):169–76.
49. Li C, Li B, Zhang N, Wang Q, Wang W, Zou H. Comparative transcriptome analysis revealed the improved β -carotene production in *Sporidiobolus pararoseus* yellow mutant MuY9. *J Gen Appl Microbiol*. 2019;65(3):121–8.
50. Wang G, Wu L, Zhang H, et al. Regulation of the phenylpropanoid pathway: a mechanism of selenium tolerance in peanut (*Arachis hypogaea* L.) seedlings. *J Agr Food Chem*. 2016;64(18):3626–35.
51. Li J, Tian C, Xia Y, Mutanda I, Wang K, Wang Y. Production of plant-specific flavones baicalein and scutellarein in an engineered *E. coli* from available phenylalanine and tyrosine. *Metab Eng*. 2019;52:124–33.
52. Kawatra A, Dhankhar R, Mohanty A, Gulati P. Biomedical applications of microbial phenylalanine ammonia lyase: current status and future prospects. *Biochimie*. 2020;177:142–52.
53. Cai Y, Cai X, Wang Q, et al. Genome sequencing of the Australian wild diploid species *Gossypium australe* highlights disease resistance and delayed gland morphogenesis. *Plant Biotechnol J*. 2019;18(3):814–28.

54. Xiao C, Chen Y, Xie S, et al. MECAT: fast mapping, error correction, and *de novo* assembly for single-molecule sequencing reads. *Nat Methods*. 2017; 14(11):1072–4.
55. Myers EW. A whole-genome assembly of drosophila. *Science*. 2000; 287(5461):2196–204.
56. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963.
57. Firrincieli A, Otilar R, Salamov A, et al. Genome sequence of the plant growth promoting endophytic yeast *Rhodotorula graminis* WP1. *Front Microbiol*. 2015;6:978.
58. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
59. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33(7):1870–4.
60. Wang Y, Tang H, DeBarry JD, et al. MCSscanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;40(7):e49.
61. Xiong Q, Zhong L, Du J, et al. Ribosome profiling reveals the effects of nitrogen application translational regulation of yield recovery after abrupt drought-flood alternation in rice. *Plant Physiol Bioch*. 2020;155:42–58.
62. Yang B, Wang N, Wang S, et al. Network-pharmacology-based identification of caveolin-1 as a key target of *Oldenlandia diffusa* to suppress breast cancer metastasis. *Biomed Pharmacother*. 2019;112:108607.
63. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884–90.
64. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60.
65. Pertea M, Pertea GM, Antonescu CM, Chang T, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33(3):290–5.
66. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc*. 2016;11(9):1650–67.
67. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 2008;18(12):1979–90.
68. Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*. 2016;32(5):767–9.
69. Tarailo Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009;25(1):4–10.
70. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25(5): 955–64.
71. Sun Y, Luo H, Li Y, et al. Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics*. 2011;12:533.
72. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5(2):R12.
73. Li L, Stoeckert CJ, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003;13(9):2178–89.
74. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.
75. Li C, Zhao D, Li B, Zhang N, Yan J, Zou H. Whole genome sequencing and comparative genomic analysis of oleaginous red yeast *Sporobolomyces pararoseus* NGR identifies candidate genes for biotechnological potential and ballistospores-shooting. *BMC Genomics*. 2020;21(1):181.
76. Zhang Z, Li J, Zhao XQ, Wang J, Wong GK, Yu J. KaKs_Calculator: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics*. 2006;4(4):259–63.
77. Wang Y, Yang L, Zhou K, Zhang Y, Song Z, He S. Evidence for adaptation to the Tibetan plateau inferred from Tibetan loach transcriptomes. *Genome Biol Evol*. 2015;7(11):2970–82.
78. Matyash V, Liebisch G, Kurzchalia TV, Shevchenko A, Schwudke D. Lipid extraction by methyl-*tert*-butyl ether for high-throughput lipidomics. *J Lipid Res*. 2008;49(5):1137–46.
79. Chen Y, Ma Z, Shen X, et al. Serum lipidomics profiling to identify biomarkers for non-small cell lung cancer. *Biomed Res Int*. 2018;2018:1–16.
80. Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012;30(10):918–20.
81. Smith CA, Want EJ, O’Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem*. 2006;78(3):779–87.
82. Wen B, Mei Z, Zeng C, Liu S. metaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinformatics*. 2017;18(1):183.
83. Kuhl C, Tautenhahn R, Böttcher C, Larson TR, Neumann S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal Chem*. 2011; 84(1):283–9.
84. Lam SM, Tian H, Shui G. Lipidomics, en route to accurate quantitation. *BBA-Mol Cell Bio Lip*. 2017;1862(8):752–61.

Publisher’s Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

