Whole genome sequencing of *Rhodotorula mucilaginosa* isolated from the chewing stick (*Distemonanthus benthamianus*): insights into *Rhodotorula* phylogeny, mitogenome dynamics and carotenoid biosynthesis

Han Ming Gan [Corresp., 1, 2, 3], Bolaji N Thomas [4], Nicole T Cavanaugh [5], Grace H Morales [5], Ashley N Mayers [4], Michael A Savka [5], Andre O Hudson [Corresp. 5]

1 Centre for Integrative Ecology-School of Life and Environmental Sciences, Deakin University, Victoria, Australia

2 Genomics Facility, Monash University, Selangor, Malaysia

3 School of Science, Monash University, Selangor, Malaysia

4 College of Health Science and Technology, Rochester Institute of Technology, Rochester, New York, United States

5 Thomas H. Gosnell School of School of Life Sciences, Rochester Institute of Technology, Rochester, New York, USA

Corresponding Authors: Han Ming Gan, Andre O Hudson
Email address: han.gan@deakin.edu.au, aohsbi@rit.edu

In industry, the yeast *Rhodotorula mucilaginosa* is commonly used for the production of carotenoids. The production of carotenoids is important because they are used as natural colorants in food and some carotenoids are precursors of retinol (vitamin A). However, the identification and molecular characterization of the carotenoid pathway/s in species belonging to the genus *Rhodotorula* is scarce due to the lack of genomic information thus potentially impeding effective metabolic engineering of these yeast strains for improved carotenoid production. In this study, we report the isolation, identification, characterization and the whole nuclear genome and mitogenome sequence of the endophyte *R. mucilaginosa* RIT389 isolated from *Distemonanthus benthamianus,* a plant known for its anti-fungal and antibacterial properties and commonly used as chewing sticks. The assembled genome of *R. mucilaginosa* RIT389 is 19 Mbp in length with an estimated genomic heterozygosity of 9.29%. Whole genome phylogeny supports the species designation of strain RIT389 within the genus in addition to supporting the monophyly of the currently sequenced *Rhodotorula* species. Further, we report for the first time, the recovery of the complete mitochondrial genome of *R. mucilaginosa* using the genome skimming approach. The assembled mitogenome is at least 7,000 bases larger than that of *Rhodotorula taiwanensis* which is largely attributed to the presence of large intronic regions containing open reading frames coding for homing endonuclease from the LAGLIDADG and GIY-YIG families. Furthermore, genomic regions containing the key genes for carotenoid production were identified in *R. mucilaginosa* RIT389, revealing differences in gene synteny that may play a role in the regulation of the biotechnologically important

carotenoid synthesis pathways in yeasts.

1   Running Head: *Rhodotorula mucilaginosa* RIT389 isolated from *Distemonanthus benthamianus*

3   Whole genome sequencing of *Rhodotorula mucilaginosa* isolated from the chewing stick
4   (*Distemonanthus benthamianus*): insights into *Rhodotorula* phylogeny, mitogenome dynamics
5   and carotenoid biosynthesis.

7   Han Ming Gan[1,2,3]*, Bolaji N. Thomas[4], Nicole T. Cavanaugh[5], Grace H. Morales[5], Ashley N.
8   Mayers[4], Michael A. Savka[5] and André O. Hudson[5]*

10   [1] Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University,
11   Geelong, Victoria, Australia
12   [2]Genomics Facility, Tropical and Medicine Biology Platform, Monash University Malaysia,
13   Bandar Sunway, Selangor, Malaysia
14   [3] School of Science, Monash University Malaysia, Bandar Sunway, Selangor, Malaysia
15   [4] The College of Health Science and Technology, Rochester Institute of Technology, Rochester,
16   NY, USA.
17   [5] Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY,
18   USA
19
20
21   *** Corresponding authors:**
22
23   Han Ming Gan, Ph.D.
24   Centre for Integrative Ecology,
25   School of Life and Environmental Sciences,
26   Deakin University
27   75 Pigdons Road, Waurn Ponds,
28   3216 VIC Australia
29   E-mail: han.gan@deakin.edu.au
30
31   André O. Hudson, Ph.D.
32   Thomas H. Gosnell School of Life Sciences
33   Rochester Institute of Technology
34   85 Lomb Memorial Dr.
35   Rochester, NY 14623 U.S.A.
36   Telephone: 585-475-4259
37   FAX: 585-475-5766
38   E-mail: aohsbi@rit.edu

39
40
41
42
43
44   **Abstract**

45   In industry, the yeast *Rhodotorula mucilaginosa* is commonly used for the production of

46   carotenoids. The production of carotenoids is important because they are used as natural

47   colorants in food and some carotenoids are precursors of retinol (vitamin A). However, the

48   identification and molecular characterization of the carotenoid pathway/s in species belonging to

49   the genus *Rhodotorula* is scarce due to the lack of genomic information thus potentially

50   impeding effective metabolic engineering of these yeast strains for improved carotenoid

51   production. In this study, we report the isolation, identification, characterization and the whole

52   nuclear genome and mitogenome sequence of the endophyte *R. mucilaginosa* RIT389 isolated

53   from *Distemonanthus benthamianus,* a plant known for its anti-fungal and antibacterial

54   properties and commonly used as chewing sticks. The assembled genome of *R. mucilaginosa*

55   RIT389 is 19 Mbp in length with an estimated genomic heterozygosity of 9.29%. Whole genome

56   phylogeny supports the species designation of strain RIT389 within the genus in addition to

57   supporting the monophyly of the currently sequenced *Rhodotorula* species. Further, we report for

58   the first time, the recovery of the complete mitochondrial genome of *R. mucilaginosa* using the

59   genome skimming approach. The assembled mitogenome is at least 7,000 bases larger than that

60   of *Rhodotorula taiwanensis* which is largely attributed to the presence of large intronic regions

61   containing open reading frames coding for homing endonuclease from the LAGLIDADG and

62   GIY-YIG families. Furthermore, genomic regions containing the key genes for carotenoid

63   production were identified in *R. mucilaginosa* RIT389, revealing differences in gene synteny that

64   may play a role in the regulation of the biotechnologically important carotenoid synthesis

65   pathways in yeasts.

66   **Introduction**

67   *Rhodotorula mucilaginosa* is a common saprophytic fungus that is a part of the

68   Basidiomycota phylum. The organism is typically found in soils, lakes, ocean water, milk and

69   fruit juice (Wirth & Goldani 2012). Of the numerous species in the genus *Rhodotorula,* only

70   *Rhodotorula mucilaginosa*, *Rhodotorula glutinis*, and *Rhodotorula minuta* have been known to

71   be pathogenic to humans (Wirth & Goldani 2012; Zaas et al. 2003). Despite being categorized as

72   an opportunistic and emerging pathogen, *R. mucilaginosa* from natural environments appear to

73   possess interesting biological traits ranging from indole acetic acid production (plant growth-

74   promoting), bacterial quorum sensing signal degradation (quorum quenching) to carotenoid

75   production (Ghani et al. 2014; Ignatova et al. 2015; Libkind et al. 2004). Despite its genomic

76   potential, resources for *R. mucilaginosa* are surprisingly scarce in public database. To date, the

77   only genomic resource publicly available for this species is from *R. mucilaginosa* strain C2.5t1

78   that was isolated from the seeds of the cacao plant in Cameroon (Deligios et al. 2015). Beyond

79   the NCBI database, another genome of *R. mulaginosa* (strain ATCC58901) can be found in the

80   JGI portal (https://genome.jgi.doe.gov/Rhomuc1/Rhomuc1.home.html) but a user account is

81   required to access the genome.

82   Carotenoid production in fungi has been suggested as a natural mechanism to protect

83   against photo-oxidative damage in light-intensive environments, given the known antioxidant

84   property of these lipid-soluble pigments as attributed to their chemical structure (Avalos &

85   Carmen Limon 2015; Cerdá-Olmedo 1989; Echavarri-Erasun & Johnson 2002). The biosynthetic

86   pathway of beta-carotene from phytoene has been elucidated in fungal species based on cDNA

87   cloning and enzymatic characterization and was shown to require two major proteins namely, a

88   dehydrogenase and a bifunctional enzyme, encoding both cyclase and phytoene synthase

89   activities (Sanz et al. 2011; Verdoes et al. 2003). Leveraging on the ease of mutant screening

90   based on visual inspection, the carotenoid pathway in the genus *Rhodotorula* has been

91   conveniently selected for the development of genetic manipulation tool in *Rhodotorula* (Abbott

92   et al. 2013; Koh et al. 2014; Sun et al. 2017) despite the biotechnological significance of this

93   pathway in *Rhodotorula* (Cutzu et al. 2013; Davoli et al. 2004; Libkind et al. 2004; Marova et al.

94   2012; Taccari et al. 2012). The heterologous expression of a 3-hydroxy-3-methylglutaryl

95   coenzyme A reductase from *Saccharomyces cerevisiae* substantially increased carotenoid

96   production in *R. mucilaginosa* strain KC8 (Wang et al. 2017), indicating the potential of

97   metabolic engineering as alternative and/or complementary approach to growth condition

98   optimization (Cutzu et al. 2013; Davoli et al. 2004; Marova et al. 2012) for improving carotenoid

99   production in *Rhodotorula* species.

100   The plant *Distemonanthus benthamianus* is a semi-deciduous perennial tree commonly

101   found in second-growth forests in Nigeria, Cameroon and Ghana (Adeniyi et al., 2011). *D.*

102   *benthamianus* is of interest given that the plant is used as chewing sticks for dental and oral

103   hygiene by members of Yoruba community in Nigeria. A relatively recent study showed that

104   extracts from the bark of the stems exhibit bactericidal activity against *Staphylococcus aureus*

105   and *Streptococcus mutans*, two bacteria that are often associated with skin and dental infections,

106   respectively (Adeniyi & Odumosu 2012).

107   In this study, an initial screen for endophytic bacteria that are resistant to the extracts of

108   *D. benthamianus* led to the isolation of a pink-pigmented strain subsequently identified as a

109   fungal strain belonging to the species *Rhodotorula mucilaginosa*. Given the intriguing property

110   of this fungal species and its lack of genomic resources, we sequenced its whole genome on the

111   Illumina platform and performed comparative genomic analysis to gain insight into the

112   carotenoid biosynthetic pathway of this species and more generally the genus *Rhodotorula*.

113 Notably, we also recovered the complete mitochondrial genome of *R. mucilaginosa,* the first for

114 its species and the second for its genus, using genome skimming approach.

115

116 **Materials and Methods**

117 **Strain Isolation**

118 Two grams of internal tissue obtained from surfaced sterilized stem of *Distemonanthus*

119 *benthamianus* plant was used to inoculate 25mL of half strength tryptic soy broth (TSB) medium

120 and grown overnight at 30°C. Microorganisms were isolated by plating 100 µL of 10 fold serial

121 dilutions from 10^{-5}-10^{-10} of the overnight culture on half strength tryptic soy agar.

122 **Scanning Electron Microscopy (SEM)**

123 To fix the organism, 100 µL of cells from 10^{-7} dilution from an overnight grown culture

124 was suspended in 3% glutaraldehyde in 0.1M phosphate buffer pH 7.2 for 30 minutes.

125 Following fixing, the cells were washed 3 times and pelleted in sterile water followed by a

126 secondary fixation in 2% osmium tetroxide (in H_20) for 30 minutes. The cells were washed 3

127 more times in sterile water followed by dehydration of the cells in 25%, 50%, 75%, 95% and

128 100% ethanol for 5 minutes in each ethanol concentration. The cells were filtered through a 0.22

129 micron polyethersulfone membrane and incubated at room temperature for 1 hour followed by

130 SEM stub mounting and sputter-coating using   10 nm gold/palladium.

131 **Whole genome sequencing**

132 Total DNA was extracted from a 3-day-old half strength tryptic soy agar culture of *R.*

133 *mucilaginosa* RIT389 using the MolBio DNA extraction kit according to the manufacturer's

134 instructions. The gDNA was sheared to 500 bp fragment using the Covaris ultrasonicator and

135 subsequently prepared for whole genome sequencing using NEBNext Ultra™ DNA Library Prep

136 kit for Illumina (New England BioLabs, Ipswich, MA). The generated library was subsequently

137   quantified using Qubit and sequenced on the MiSeq (Illumina, San Diego, CA) located at the

138   Monash University Malaysia Genomics Facility using the run configuration of 2×250bp.

139   **Genome assembly and annotation**

140   Genome size, heterozygosity rate and repeat content were initially estimated using

141   GenomeScope (Vurture et al. 2017). Based on the observed high genome heterozygosity of strain

142   RIT389, dipSPAdes version 3.10.1 was used to assemble the whole genome with the additional

143   option of "-expect-rearrangements" activated (Bankevich et al. 2012). Genome completeness

144   was calculated using BUSCO3 based on the Basidiomycota odb9 ortholog dataset (Simao et al.

145   2015). Then, gene prediction was performed using GeneMark-ES fungal version (Borodovsky &

146   Lomsadze 2011) with the enhanced intron submodel that can better accommodate sequences

147   with and without branch point sites in the fungal genomes.

148   Complete mitogenome was recovered by randomly sub-sampling 1/10 of the pair-end

149   reads and assembling them using SPAdes version 3.10.1 (Bankevich et al. 2012). The contig

150   corresponding to the whole mitogenome was re-circularized manually, as previously described

151   (Gan et al. 2014) and annotated automatically using MFannot

152   (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl). Additional genes coding

153   for homing endonucleases commonly found in fungal mitogenomes were identified based on the

154   presence of protein domains corresponding to the GIY-YIG catalytic domain (PF01541.23) and

155   LAGLIDADG endonuclease (PF00961.18, PF03161.12 and PF14528.5) using hmmsearch3 with

156   an E-value cutoff of 1e-5 (Eddy 2011).

157   **Phylogenomics and Comparative Genomics**

158   Pair-wise average nucleotide identity (ANIm) was calculated using JSpecies (Richter et

159   al. 2016) and subsequently visualized with the library package pheatmap in Rstudio. Single-copy

160   genes present in all selected fungal genomes were identified using BUSCO3 (Simao et al. 2015).

161   The protein sequences for each ortholog were aligned and trimmed using Muscle and trimAl (-

162   automated1), respectively (Capella-Gutierrez et al. 2009; Edgar 2004). The final trimmed

163   alignments were concatenated and used to construct a maximum likelihood tree using

164   FastTreeMP (Price et al. 2010). The reconstructed tree was visualized and annotated using

165   TreeGraph2 (Stöver & Müller 2010).

166   Identification of proteins involved in the carotenoid biosynthesis pathway was done by

167   scanning the whole predicted proteome for protein domain hits (NC cutoff for TIGRfam and 1e-

168   5 cutoff for Pfam) to lycopene cyclase (TIGR03462, CrtY), phytoene desaturase/dehydrogenase

169   (TIGR02734, CrtI), squalene/phytoene synthase (PF0494, CrtB) and isopentenyl-diphosphate

170   delta-isomerase (TIGR02150). Visualization and comparison of gene neighborhoods were

171   performed using EasyFig with the default BlastN setting (Sullivan et al. 2011). Proteins coded in

172   each genomic sub-region were functionally annotated using Interproscan5 (Jones et al. 2014).

173

174   **Results and Discussion**

175   We noticed an organism that was pink/red in color from the screen on tryptic soy agar

176   (Figure 1A). Based on our previous studies of isolating endophytic organisms, we initially

177   thought the organism belonged to genus *Serratia* or a related genus based on the color of the

178   colonies. However, based on SEM analysis, it was initially determined that the organism was

179   eukaryotic and not a bacterium based on the size and the morphology depicting cell division

180   (Figure 1B). The identification of the organism was subsequently confirmed using whole

181   genome nucleotide sequencing.

182   GenomeScope estimated a genome size of 18.6 mega base pairs (Mbp) with an estimated

183 heterozygosity of 9.29% for strain RIT389 (Figure 2). The predicted genome size is fairly close

184 to the *de novo* assembled genome length of 19.6 Mbp contained in 250 contigs. The assembled

185 genome has a GC content of 60.28% with an estimated completeness of 89.70%. *De novo*

186 assembly using sub-sampled reads enabled the recovery of the complete mitogenome of strain

187 RIT389 which is the first mitogenome reported for this genus. Approximately 4.55% of the total

188 pair-end reads mapped to the complete mitogenome with an estimated coverage of 400× (Table

189 1). The complete mitogenome length is 47,023 bp with a GC content of 40.43% which is

190 substantially lower than that of the nuclear genome.

191

192 **Genomic and genetic approaches support the species identification of strain RIT389 as**

193 ***Rhodotorula mucilaginosa***

194   The ITS region of strain RIT389 exhibits a 100% identity with the sequences of various

195 *Rhodotorula mucilaginosa* strains including the type strain *R. mucilaginosa* ATCC 201848

196 (Supplemental Table 1). At the whole genome level, it exhibits the highest average nucleotide

197 identity of 94.10% to strain C2.5t1, the only other genome-sequenced strain of this species at the

198 time of this study (Deligios et al. 2015) (Figure 3). Similar to strain RIT389, strain C2.5t1 is also

199 a plant-associated and was isolated from a cacao seeds (*Theobroma cacao* L) in Cameroon and

200 shown to produce high carotenoid levels when grown in medium supplemented with glycerol

201 (Cutzu et al. 2013). Although the plant growth-promoting activity of both strains RIT389 and

202 C2.5t1 has not been studied, a third strain of *R. mucilaginosa*, YR07 isolated from legume plant

203 rhizophere, has been shown to synthesize up to 45.3 µg of indole acetic acid (IAA) per mL of

204 culture medium (Ignatova et al. 2015). In addition, strain YR07 also exhibits antifungal activity

205   as evidenced by the formation of large inhibition zone against *Fusarium graminearum*, a

206   phytopathogenic fungi.

207

208   **Whole genome phylogeny supports the monophyly of the genus *Rhodotorula***

209   A total of 789 single-copy genes universally present in all 19 fungal strains (Figure 4)

210   were used to generate a concatenated amino acid alignment consisting of 539,792 sites (234,641

211   informative sites). By rooting *Microbotyrum violaceum* and *Microbotyrum saponariae* belonging

212   to a different order i.e. Microbotryales, as the outgroup, members of the genus *Rhodotorula*

213   formed a monophyletic group cluster with maximal SH-like support with *R. graminis* WP1 being

214   basal to the rest of the *Rhodotorula* strains (Figure 4). The lack of strong SH-like support at the

215   shallow relationship especially for members of the species *R. toluroides* is most likely due to the

216   lack of genomic differences which is expected given that some of the strain names are the

217   alternative strain name of the identical type strain. For example, the type strain designations

218   ATCC10788, IFO0559 and JCM10020 for *R. toluroides* were all derived from the original strain

219   CBS 14. Interestingly, phylogenomic analysis indicates that the currently sequenced strains of

220   *Rhodotorula toluroides* consist of two major clades with a Jspecies-calculated intraclade and

221   interclade average pair-wise ANI difference of 0.4% and 13%, respectively (Figure 3).

222   *Rhodotorula* sp. JG1b together with *R. mucilaginosa* strains RIT389 and C2.5t1 formed a

223   monophyletic group that is sister taxa to the major *R. toluroides* group. The close affinity of

224   *Rhodotorula* sp. JG-1b to *R. mucilaginosa* is interesting as it is an eurypsychrophilic yeast

225   isolated from ~150,000-year-old ice-cemented permafrost soils (Goordial et al. 2016a). Given

226   the close genomic affinity of *R*. sp. JG-1b to the currently sequenced *R. mucilaginosa* strains and

227   their diverse isolation source, comparative genomics of these strains may assist in the future

228   identification of novel cold adaptive traits at the molecular level in the genus *Rhodotorula*

229   (Goordial et al. 2016b).

230

231   **Homing endonuclease-mediated mitogenome expansion in *Rhodotorula mucilaginosa***

232   **RIT389**

233   Given the high abundance of mitochondrial organelle in an actively dividing cell, the

234   depth of mitochondrial-derived sequencing reads will be substantially higher than that of the

235   nuclear genome. Thus, by performing a shallow sequencing (genome skimming) on the organism

236   of interest, it is possible to obtain sufficient read representation of the mitochondrial genome

237   which subsequently allows complete assembly. Since the first reported success of genome

238   skimming approach in the construction of the bighorn sheep mitochondrial genome (Miller et al.

239   2012), similar success in recovering mitogenomes across different organisms has been reported

240   (Froufe et al. 2016; Gan et al. 2016; Krzeminska et al. 2016; Pavlova et al. 2017). In addition,

241   plastomes and other high copy number genes have also been routinely recovered and assembled

242   with genome skimming (Bakker 2017; Gan et al. 2014; Grandjean et al. 2017; Richter et al.

243   2015; Straub et al. 2012).

244   Contrary to genome skimming, a high coverage whole genome sequencing will generate

245   an extremely high read representation of the mitogenome that can negatively affect mitogenome

246   assembly due to the accumulation of sequencing errors, leading to the generation of fragmented

247   mitogenome (Mirebrahim et al. 2015). In this work, we show that a simple subsampling

248   approach i.e. "*in-silico* genome skimming" followed by *de novo* assembly substantially improves

249   mitogenome assembly and leads to its recovery as a circularized contig in *R. mucilaginosa*.

250   The reconstructed mitogenome of *R. mucilaginosa* RIT389 is 93% similar to that of *R.*

251   *taiwanensis* (Accession Number: HF558455), the only other publicly available complete

252   *Rhodotorula* mitogenome assembled from a low output paired-end run (470 megabases output)

253   of Roche 454 Genome Sequencer (Zhao et al. 2013) despite the availability of various

254   *Rhodotorula* whole genome sequences in the public database.

255   Future study focusing on the reconstruction of complete mitogenome using *in-silico* genome

256   skimming approach from fungal whole genome sequencing data that are publicly available in the

257   the NCBI sequence read archive (SRA) will be instructive.

258   Despite exhibiting a similar mitochondrial gene arrangement and a relatively high

259   nucleotide sequence similarity to *R. taiwanensis* RS1, the assembled complete mitogenome of

260   strain RIT389 is at least 7,000 bp larger than that of *R. taiwanensis*. Gene neighborhood analysis

261   indicates that a majority of the length difference was largely due to the presence of intronic

262   regions containing homing endonuclease genes (Figure 5) which is consistent with other studies

263   reporting the prevalence of fungal mitogenome size polymorphism among species from the same

264   genus due to intron acquisition (Joardar et al. 2012; Kanzi et al. 2016). Homing endonucleases

265   recognize and cleave target sites ranging from 14 to 40 base pairs which match the intron

266   insertion site in donor DNA (Belfort 2005). The homing endonucleases identified in both

267   *Rhodotorula* mitogenomes belong to the LADLIDADG and GIY-YIG families. Both

268   LADLIDADG and GIY-YIG endonucleases were named according to the signature motifs

269   presence in their protein sequence. For example, the GIY-YIG endonucleases are characterized

270   by the presence of a structural domain with two short motifs "GIY" and "YIG" in the N-terminal

271   (Dunin-Horkawicz et al. 2006).

272   In strain RIT389, approximately 3 kilobases of the large mitochondrial ribosomal RNA

273   gene consist of intronic regions coding for LAGLIDADG-type endonuclease and in contrast,

274   these regions are completely absent in the mitogenome of *R. taiwanensis* RS1. A GIY-YIG

275   endonuclease ORF could also be identified within the 1.5 kilobases intronic region of RIT389

276   *nad5* gene which is absent in that of *R. taiwanensis* RS1. Presence of intronic region in the

277   mitochondrial *nad5* has been previously reported in *Basidiomycota* and Ascomycota species

278   such as *Trametes cingulata, Moniliophthora perniciosa, Ustilago maydis* and *Rhynchosporium*

279   *commune* (Abbott et al. 2013; Formighieri et al. 2008; Haridas & Gantt 2010). However, in the

280   reported fungal mitogenomes, the intronic ORF(s) in *nad5* encodes for LAGLIDADG-type

281   endonuclease instead of GIY-YIG endonuclease. The first piece of evidence for a mobile intronic

282   GIY-YIG endonuclease ORF in fungi was demonstrated by the efficient transfer of the GIY-YIG

283   ORF from the second intron of mitochondrial cytochrome b gene in *Podospora curvicolla* to a

284   GIY-YIG-less allele (Saguez et al. 2000).

285

286   **Identification of a genomic region associated with carotenoid biosynthesis**

287   Essential genes required for the biosynthesis of carotenoid could be identified in strain

288   RIT389 which is consistent with its red coloration (Figure 1A, Supplemental Table 2), a visual

289   evidence for carotenoid production. Such notable phenotype associated with carotenoid

290   production presents a huge advantage for molecular cloning and characterization of this pathway.

291   As expected, genes involved in the carotenoid synthesis pathway were frequently cloned and

292   characterized from a wide variety of bacteria, archaea, fungi and plants (Li et al. 2011; Li et al.

293   1996; Misawa et al. 1995; Nupur et al. 2016; Reddy et al. 2017; Van Dien et al. 2003; Yang et al.

294   2015).

295   The genes coding for phytoene synthase (*crtB*), lycopene cyclase (*crtY*), and phytoene

296   desaturase (*crtI*) are located in relatively close proximity with one another while the gene coding

297   for the enzyme geranyl pyrophosphate synthase which is crucial for the production of an early

298   precursor for carotenoid is located on separate contig (Supplemental Table 2). As observed in

299   several fungal species, the *crtB* and *crtY* genes are fused and thus code for a bifunctional protein

300   containing both lycopene cyclase and phytoene synthase activities (Arrach et al. 2001; Sanz et al.

301   2011). Within the genus *Rhodotorula*, the gene coding for carotenoid oxygenase (*crtX*)

302   responsible for the cleavage of carotenoid to retinal (Vitamin A) and *crtBY* are located in close

303   proximity and are convergently transcribed except in the species *R. mucilaginosa*, whereby *crtX*

304   and *crtBY* are divergently transcribed and are separated by a large gene coding for OPT family

305   small oligopeptide transporter (Figure 6). In *Fusarium fujikuroi*, mutation in the *crtX* gene led to

306   the overproduction of carotenoid (Prado-Cabrero et al. 2007), suggesting that carotenoid

307   oxygenase is involved in the regulation of the carotenoid synthesis through a negative feedback

308   mechanism. The notable difference in gene arrangement and transcription orientation involving

309   *crtX* can therefore affect the regulation of carotenoid synthesis and accumulation in *R.*

310   *mucilaginosa* (Noble & Andrianopoulos 2013). It is also worth noting that the *crtI* gene in strain

311   RIT389 was predicted as two separate genes which is unexpected given that the gene region

312   exhibits high nucleotide homology and coverage to its respective orthologs in strains WP1 and

313   Ct2.5. Whole transcriptome analysis of strain RIT389 will be necessary to validate the predicted

314   spliced *crtI* gene in the future.

315

316   **Conclusion**

317   We demonstrate the feasibility of reconstructing the whole genome and complete mitogenome of

318   *Rhodotorula mucilaginosa* using only Illumina short reads. The whole genome of *R.*

319   *mucilaginosa* is the second to be reported to date for its species. Despite the availability of

320   various whole genome sequences of *Rhodotorula* in public databases, the complete and

321   annotated mitogenome of *Rhodotorula mucilaginosa* strain RIT389 is the first to be successfully

322   reconstructed via *in-silico* genome skimming and annotated for its species. We also highlight the

323   considerable dissimilarity in the syntheny of carotenoid synthesis gene cluster among

324   *Rhodotorula* strains with potential implications in the regulation of carotenoid production.

325

326   **Data Deposition**

327   This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank

328   under the accession NIUW00000000. The version described in this paper

329   is version NIUW01000000. Bioproject, Biosample and SRA accession numbers are shown in

330   Table 1.

331

332   **Acknowledgments**

333   The authors thank Dr. Richard Hailstone from the Chester F. Carlson Center for Imaging Science

334   (RIT) for assistance with SEM analysis.

335

336   **Figure Legends**

337   **Figure 1**. (A) Color/morphology of *Rhodotorula mucilaginosa* RIT389 grown on half-strength

338   tryptic soy agar (B) Scanning electron microscopy of *Rhodotorula mucilaginosa* RIT389 at

339   52.8K magnification.

340

341   **Figure 2**. GenomeScope estimation of genome size, repeat content and heterozygosity (Kmer

342   length = 21, Read length = 251 bp and Max kmer coverage = 1,000).

343

344   **Figure 3**. Pairwise average nucleotide identity calculation of *Rhodotorula* genomes. Genomes

345   with the superscript "T" are type strains.

346

347   **Figure 4**. Maximum likelihood tree of a concatenated amino acid alignment consisting of

348   537,792 sites that represent 798 universally present single-copy genes from 19 fungal strains.

349   Labels on branches indicate shimodaira-hasegawa (SH)-like local branch support values. The

350   scale bar indicates the average number of amino acid substitutions per site.

351

352   **Figure 5**. Complete mitochondrial genome of *R. mucilaginosa* RIT389 compared against that of

353   *R. taiwanensis* RS1. Orange frames indicate coding sequences commonly found in a typical

354   mitochondrial genome. Red and Blue arrows indicate transfer and ribosomal RNAs, respectively.

355   Arrow direction represents transcriptional orientation. Dotted lines indicate intronic regions.

356

357   **Figure 6**. Comparison of genomic sub-region containing the gene cluster associated with

358   carotenoid biosynthetic pathway. Orange frames within the teal arrows indicate the coding

359   sequences in the exonic regions of the corresponding genes.

360

361

362

363

364   **References**

365   Abbott EP, Ianiri G, Castoria R, and Idnurm A. 2013. Overcoming recalcitrant
366   transformation and gene manipulation in *Pucciniomycotina* yeasts. *Appl Microbiol*
367   *Biotechnol* 97. 10.1007/s00253-012-4561-7
368   Adeniyi CBA, and Odumosu BT. 2012. Antibacterial and antifungal properties of
369   *Distemonanthus benthamianus* (Fabaceae) crude extract. *The global journal of*
370   *pharmaceutical research* 1:567-574.
371   Arrach N, Fernández-Martín R, Cerdá-Olmedo E, and Avalos J. 2001. A single gene for
372   lycopene cyclase, phytoene synthase, and regulation of carotene biosynthesis in
373   Phycomyces. *Proceedings of the National Academy of Sciences* 98:1687-1692.
374   10.1073/pnas.98.4.1687
375   Avalos J, and Carmen Limon M. 2015. Biological roles of fungal carotenoids. *Curr Genet*
376   61:309-324.
377   Bakker FT. 2017. Herbarium genomics: skimming and plastomics from archival specimens.
378   *Webbia* 72:35-45. 10.1080/00837792.2017.1313383
379   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko
380   SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA,
381   and Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its
382   applications to single-cell sequencing. *J Comput Biol* 19:455-477.
383   Borodovsky M, and Lomsadze A. 2011. Eukaryotic gene prediction using GeneMark.hmm-E
384   and GeneMark-ES. *Curr Protoc Bioinformatics* 4:1-10.
385   Capella-Gutierrez S, Silla-Martinez JM, and Gabaldon T. 2009. trimAl: a tool for automated
386   alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972-
387   1973.
388   Cerdá-Olmedo E. 1989. Production of Carotenoids with Fungi. In: Vandamme EJ, ed.
389   *Biotechnology of Vitamins, Pigments and Growth Factors*. Dordrecht: Springer
390   Netherlands, 27-42.
391   Cutzu R, Coi A, Rosso F, Bardi L, Ciani M, Budroni M, Zara G, Zara S, and Mannazzu I. 2013.
392   From crude glycerol to carotenoids by using a *Rhodotorula glutinis* mutant. *World J*
393   *Microbiol Biotechnol* 29:1009-1017.
394   Davoli P, Mierau V, and Weber RWS. 2004. Carotenoids and Fatty Acids in Red Yeasts
395   *Sporobolomyces roseus* and *Rhodotorula glutinis*. *Applied Biochemistry and*
396   *Microbiology* 40:392-397. 10.1023/B:ABIM.0000033917.57177.f2
397   Deligios M, Fraumene C, Abbondio M, Mannazzu I, Tanca A, Addis MF, and Uzzau S. 2015.
398   Draft Genome Sequence of *Rhodotorula mucilaginosa*, an Emergent Opportunistic
399   Pathogen. *Genome Announcements* 3:e00201-00215. 10.1128/genomeA.00201-15
400   Dunin-Horkawicz S, Feder M, and Bujnicki JM. 2006. Phylogenomic analysis of the GIY-YIG
401   nuclease superfamily. *BMC Genomics* 7:98.
402   Echavarri-Erasun C, and Johnson EA. 2002. Fungal carotenoids. *Applied Mycology and*
403   *Biotechnology* 2:45-85. http://dx.doi.org/10.1016/S1874-5334(02)80006-5
404   Eddy SR. 2011. Accelerated Profile HMM Searches. *PLoS Comput Biol* 7:20.
405   Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and
406   space complexity. *BMC Bioinformatics* 5:113.
407   Formighieri EF, Tiburcio RA, Armas ED, Medrano FJ, Shimo H, Carels N, Goes-Neto A,
408   Cotomacci C, Carazzolle MF, Sardinha-Pinto N, Thomazella DP, Rincones J,

409   Digiampietri L, Carraro DM, Azeredo-Espin AM, Reis SF, Deckmann AC, Gramacho K,
410   Goncalves MS, Moura Neto JP, Barbosa LV, Meinhardt LW, Cascardo JC, and Pereira
411   GA. 2008. The mitochondrial genome of the phytopathogenic basidiomycete
412   *Moniliophthora perniciosa* is 109 kb in size and contains a stable integrated plasmid.
413   *Mycol Res* 112:1136-1152.
414 Froufe E, Gan HM, Lee YP, Carneiro J, Varandas S, Teixeira A, Zieritz A, Sousa R, and Lopes-
415   Lima M. 2016. The male and female complete mitochondrial genome sequences of
416   the Endangered freshwater mussel *Potomida littoralis* (Cuvier, 1798) (Bivalvia:
417   Unionidae). *Mitochondrial DNA Part A* 27:3571-3572.
418   10.3109/19401736.2015.1074223
419 Gan HM, Schultz MB, and Austin CM. 2014. Integrated shotgun sequencing and
420   bioinformatics pipeline allows ultra-fast mitogenome recovery and confirms
421   substantial gene rearrangements in Australian freshwater crayfishes. *BMC Evol Biol*
422   14:1471-2148.
423 Gan HM, Tan MH, Eprilurahman R, and Austin CM. 2016. The complete mitogenome of
424   *Cherax monticola* (Crustacea: Decapoda: Parastacidae), a large highland crayfish
425   from New Guinea. *Mitochondrial DNA Part A* 27:337-338.
426   10.3109/19401736.2014.892105
427 Ghani NA, Sulaiman J, Ismail Z, Chan X-Y, Yin W-F, and Chan K-G. 2014. *Rhodotorula*
428   *mucilaginosa*, a Quorum Quenching Yeast Exhibiting Lactonase Activity Isolated
429   from a Tropical Shoreline. *Sensors (Basel, Switzerland)* 14:6463-6473.
430   10.3390/s140406463
431 Goordial J, Davila A, Lacelle D, Pollard W, Marinova MM, Greer CW, DiRuggiero J, McKay CP,
432   and Whyte LG. 2016a. Nearing the cold-arid limits of microbial life in permafrost of
433   an upper dry valley, Antarctica. *The ISME Journal* 10:1613-1624.
434   10.1038/ismej.2015.239
435 Goordial J, Raymond-Bouchard I, Riley R, Ronholm J, Shapiro N, Woyke T, LaButti KM, Tice
436   H, Amirebrahimi M, Grigoriev IV, Greer C, Bakermans C, and Whyte L. 2016b.
437   Improved High-Quality Draft Genome Sequence of the Eurypsychrophile
438   *Rhodotorula* sp. JG1b, Isolated from Permafrost in the Hyperarid Upper-Elevation
439   McMurdo Dry Valleys, Antarctica. *Genome Announcements* 4.
440   10.1128/genomeA.00069-16
441 Grandjean F, Tan MH, Gan HM, Lee YP, Kawai T, Distefano RJ, Blaha M, Roles AJ, and Austin
442   CM. 2017. Rapid recovery of nuclear and mitochondrial genes by genome skimming
443   from Northern Hemisphere freshwater crayfish. *Zoologica Scripta*.
444 Haridas S, and Gantt JS. 2010. The mitochondrial genome of the wood-degrading
445   basidiomycete *Trametes cingulata*. *FEMS Microbiology Letters* 308:29-34.
446   10.1111/j.1574-6968.2010.01979.x
447 Ignatova LV, Brazhnikova YV, Berzhanova RZ, and Mukasheva TD. 2015. Plant growth-
448   promoting and antifungal activity of yeasts from dark chestnut soil. *Microbiological*
449   *Research* 175:78-83. http://dx.doi.org/10.1016/j.micres.2015.03.008
450 Joardar V, Abrams NF, Hostetler J, Paukstelis PJ, Pakala S, Pakala SB, Zafar N, Abolude OO,
451   Payne G, Andrianopoulos A, Denning DW, and Nierman WC. 2012. Sequencing of
452   mitochondrial genomes of nine *Aspergillus* and *Penicillium* species identifies mobile
453   introns and accessory genes as main sources of genome size variability. *BMC*
454   *Genomics* 13:698. 10.1186/1471-2164-13-698

455   Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
456   Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez
457   R, and Hunter S. 2014. InterProScan 5: genome-scale protein function classification.
458   *Bioinformatics* 30:1236-1240. 10.1093/bioinformatics/btu031
459   Kanzi AM, Wingfield BD, Steenkamp ET, Naidoo S, and van der Merwe NA. 2016. Intron
460   Derived Size Polymorphism in the Mitochondrial Genomes of Closely Related
461   Chrysoporthe Species. *PLOS ONE* 11:e0156104. 10.1371/journal.pone.0156104
462   Koh CMJ, Liu Y, Moehninsi, Du M, and Ji L. 2014. Molecular characterization of KU70 and
463   KU80 homologues and exploitation of a KU70-deficient mutant for improving gene
464   deletion frequency in *Rhodosporidium toruloides. BMC Microbiology* 14:50.
465   10.1186/1471-2180-14-50
466   Krzeminska U, Wilson R, Rahman S, Song BK, Gan HM, Tan MH, and Austin CM. 2016. The
467   complete mitochondrial genome of the invasive house crow *Corvus splendens*
468   (Passeriformes: Corvidae). *Mitochondrial DNA Part A* 27:974-975.
469   10.3109/19401736.2014.926512
470   Li M, Gan Z, Cui Y, Shi C, and Shi X. 2011. Cloning and Characterization of the -Carotene
471   Desaturase Gene from *Chlorella protothecoides* CS-41. *Journal of Biomedicine and
472   Biotechnology* 2011:7. 10.1155/2011/731542
473   Li Z-H, Matthews PD, Burr B, and Wurtzel ET. 1996. Cloning and characterization of a maize
474   cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthetic
475   pathway. *Plant Molecular Biology* 30:269-279. 10.1007/bf00020113
476   Libkind D, Brizzio S, and van Broock M. 2004. *Rhodotorula mucilaginosa*, a carotenoid
477   producing yeast strain from a Patagonian high-altitude lake. *Folia Microbiol* 49:19-
478   25.
479   Marova I, Carnecka M, Halienova A, Certik M, Dvorakova T, and Haronikova A. 2012. Use of
480   several waste substrates for carotenoid-rich yeast biomass production. *J Environ
481   Manage* 95:8.
482   Miller JM, Malenfant RM, Moore SS, and Coltman DW. 2012. Short Reads, Circular Genome:
483   Skimming SOLiD Sequence to Construct the Bighorn Sheep Mitochondrial Genome.
484   *Journal of Heredity* 103:140-146. 10.1093/jhered/esr104
485   Mirebrahim H, Close TJ, and Lonardi S. 2015. De novo meta-assembly of ultra-deep
486   sequencing data. *Bioinformatics* 31.
487   Misawa N, Satomi Y, Kondo K, Yokoyama A, Kajiwara S, Saito T, Ohtani T, and Miki W. 1995.
488   Structure and functional analysis of a marine bacterial carotenoid biosynthesis gene
489   cluster and astaxanthin biosynthetic pathway proposed at the gene level. *Journal of
490   Bacteriology* 177:6575-6584.
491   Noble LM, and Andrianopoulos A. 2013. Fungal Genes in Context: Genome Architecture
492   Reflects Regulatory Complexity and Function. *Genome Biology and Evolution* 5:1336-
493   1352. 10.1093/gbe/evt077
494   Nupur LNU, Vats A, Dhanda SK, Raghava GPS, Pinnaka AK, and Kumar A. 2016. ProCarDB: a
495   database of bacterial carotenoids. *BMC Microbiology* 16:96. 10.1186/s12866-016-
496   0715-6
497   Pavlova A, Gan HM, Lee YP, Austin CM, Gilligan DM, Lintermans M, and Sunnucks P. 2017.
498   Purifying selection and genetic drift shaped Pleistocene evolution of the
499   mitochondrial genome in an endangered Australian freshwater fish. *Heredity*
500   118:466-476. 10.1038/hdy.2016.120

501 Prado-Cabrero A, Scherzinger D, Avalos J, and Al-Babili S. 2007. Retinal biosynthesis in
502   fungi: characterization of the carotenoid oxygenase CarX from *Fusarium fujikuroi*.
503   *Eukaryot Cell* 6:650-657.
504 Price MN, Dehal PS, and Arkin AP. 2010. FastTree 2--approximately maximum-likelihood
505   trees for large alignments. *PLoS One* 5:0009490.
506 Reddy CS, Lee SH, Yoon JS, Kim JK, Lee SW, Hur M, Koo SC, Meilan J, Lee WM, Jang JK, Hur Y,
507   Park SU, and Kim AYB. 2017. Molecular Cloning and Characterization of Carotenoid
508   Pathway Genes and Carotenoid Content in *Ixeris dentata* var. albiflora. *Molecules* 22.
509 Richter M, Rosello-Mora R, Oliver Glockner F, and Peplies J. 2016. JSpeciesWS: a web
510   server for prokaryotic species circumscription based on pairwise genome
511   comparison. *Bioinformatics* 32:929-931.
512 Richter S, Schwarz F, Hering L, Böggemann M, and Bleidorn C. 2015. The Utility of Genome
513   Skimming for Phylogenomic Analyses as Demonstrated for Glycerid Relationships
514   (Annelida, Glyceridae). *Genome Biology and Evolution* 7:3443-3462.
515   10.1093/gbe/evv224
516 Saguez C, Lecellier G, and Koll F. 2000. Intronic GIY-YIG endonuclease gene in the
517   mitochondrial genome of *Podospora curvicolla*: evidence for mobility. *Nucleic Acids*
518   *Research* 28:1299-1306.
519 Sanz C, Velayos A, Álvarez MI, Benito EP, and Eslava AP. 2011. Functional Analysis of the
520   Phycomyces *carRA* Gene Encoding the Enzymes Phytoene Synthase and Lycopene
521   Cyclase. *PLOS ONE* 6:e23102. 10.1371/journal.pone.0023102
522 Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, and Zdobnov EM. 2015. BUSCO:
523   assessing genome assembly and annotation completeness with single-copy
524   orthologs. *Bioinformatics* 31:3210-3212.
525 Stöver BC, and Müller KF. 2010. TreeGraph 2: Combining and visualizing evidence from
526   different phylogenetic analyses. *BMC Bioinformatics* 11:7. 10.1186/1471-2105-11-7
527 Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, and Liston A. 2012. Navigating
528   the tip of the genomic iceberg: Next-generation sequencing for plant systematics.
529   *American Journal of Botany*. 10.3732/ajb.1100335
530 Sullivan MJ, Petty NK, and Beatson SA. 2011. Easyfig: a genome comparison visualizer.
531   *Bioinformatics* 27:1009-1010.
532 Sun W, Yang X, Wang X, Lin X, Wang Y, Zhang S, Luan Y, and Zhao ZK. 2017. Homologous
533   gene targeting of a carotenoids biosynthetic gene in *Rhodosporidium toruloides* by
534   *Agrobacterium*-mediated transformation. *Biotechnol Lett* 39:1001-1007.
535 Taccari M, Canonico L, Comitini F, Mannazzu I, and Ciani M. 2012. Screening of yeasts for
536   growth on crude glycerol and optimization of biomass production. *Bioresour*
537   *Technol* 110:488-495.
538 Van Dien SJ, Marx CJ, O'Brien BN, and Lidstrom ME. 2003. Genetic Characterization of the
539   Carotenoid Biosynthetic Pathway in *Methylobacterium extorquens* AM1 and Isolation
540   of a Colorless Mutant. *Applied and Environmental Microbiology* 69:7563-7566.
541   10.1128/aem.69.12.7563-7566.2003
542 Verdoes JC, Sandmann G, Visser H, Diaz M, van Mossel M, and van Ooyen AJJ. 2003.
543   Metabolic Engineering of the Carotenoid Biosynthetic Pathway in the Yeast
544   *Xanthophyllomyces dendrorhous* (*Phaffia rhodozyma*). *Applied and Environmental*
545   *Microbiology* 69:3728-3738. 10.1128/aem.69.7.3728-3738.2003

546 Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, and Schatz MC.
547   2017. GenomeScope: Fast reference-free genome profiling from short reads.
548   *Bioinformatics* 24.
549 Wang Q, Liu D, Yang Q, and Wang P. 2017. Enhancing carotenoid production in *Rhodotorula*
550   *mucilaginosa* KC8 by combining mutation and metabolic engineering. *Annals of*
551   *Microbiology* 67:425-431. 10.1007/s13213-017-1274-2) contains supplementary
552   material, which is available to authorized users.

553 10.1007/s13213-017-1274-2
554 Wirth F, and Goldani LZ. 2012. Epidemiology of *Rhodotorula*: An Emerging Pathogen.
555   *Interdisciplinary Perspectives on Infectious Diseases* 2012:7. 10.1155/2012/465717
556 Yang Y, Yatsunami R, Ando A, Miyoko N, Fukui T, Takaichi S, and Nakamura S. 2015.
557   Complete Biosynthetic Pathway of the C50 Carotenoid Bacterioruberin from
558   Lycopene in the Extremely Halophilic Archaeon *Haloarcula japonica*. *Journal of*
559   *Bacteriology* 197:1614-1623. 10.1128/jb.02523-14
560 Zaas AK, Boyce M, Schell W, Lodge BA, Miller JL, and Perfect JR. 2003. Risk of fungemia due
561   to Rhodotorula and antifungal susceptibility testing of *Rhodotorula* isolates. *J Clin*
562   *Microbiol* 41:5233-5235.
563 Zhao XQ, Aizawa T, Schneider J, Wang C, Shen RF, and Sunairi M. 2013. Complete
564   mitochondrial genome of the aluminum-tolerant fungus *Rhodotorula taiwanensis*
565   RS1 and comparative analysis of Basidiomycota mitochondrial genomes.
566   *MicrobiologyOpen* 2:308-317. 10.1002/mbo3.74

567

568

Strain RIT389 genome statistic and strain information

1 **Table 1**. Strain RIT389 genome statistic and strain information

Organism	*Rhodotorula mucilaginosa*
Strain Name	RIT389
SRA	SRR5860569
Bioproject	PRJNA390458
Biosample	SAMN07235707
Whole genome:	
Accession Number	NIUW01000000
Assembled genome length	19,664,434 bp
N_{50} length	194,287 bp
Number of contigs	250
GC %	60.28%
Predicted protein-coding gene	7,065
BUSCO Completeness (Basidiomycota odb9)	
Complete BUSCOs	89.70%
Complete and single-copy BUSCOs	86.70%
Complete and duplicated BUSCO	3.00%
Fragmented BUSCO	1.60%
Missing BUSCO	8.70%
Total BUSCO groups searched	1,335
Mitochondrial Genome	
Accession Number	MF694646
Genome Size	47,023 bp
GC %	40.43%
Coverage	400×
Alignment Rate	4.55%

Figure 1

(A) Color/morphology of *Rhodotorula mucilaginosa* RIT389 grown on half-strength tryptic soy agar (B) Scanning electron microscopy of *Rhodotorula mucilaginosa* RIT389 at 52.8K magnification.
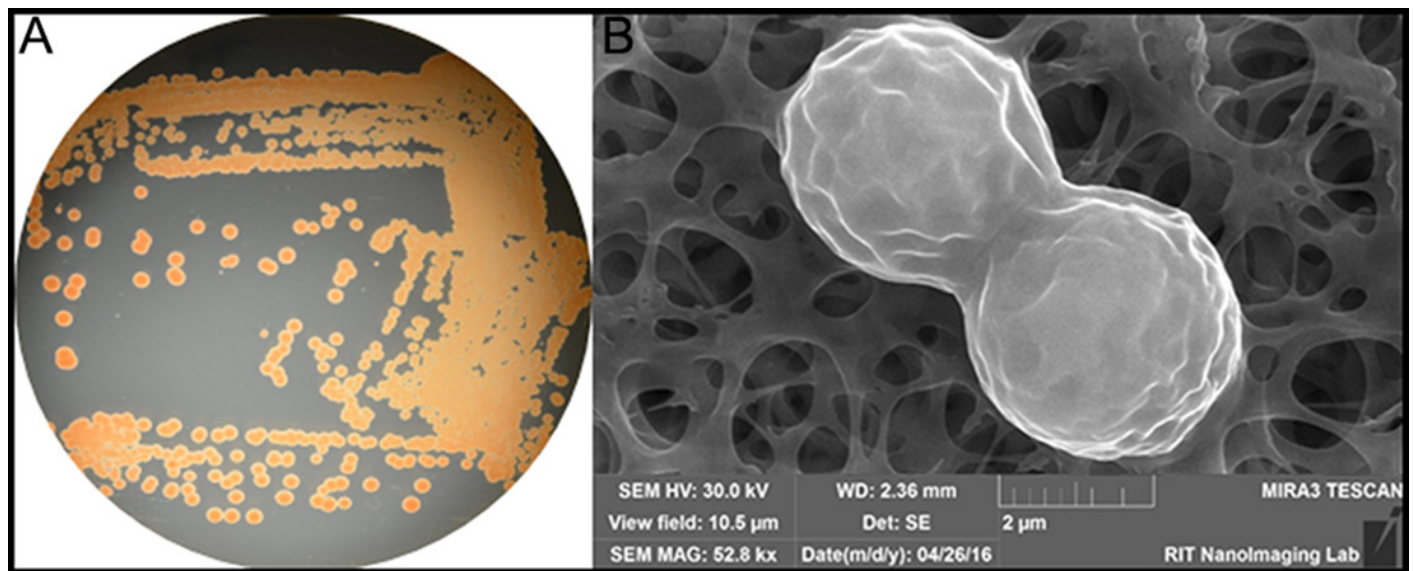
Figure 2

GenomeScope estimation of genome size, repeat content and heterozygosity (Kmer length = 21, Read length = 251 bp and Max kmer coverage = 1,000).
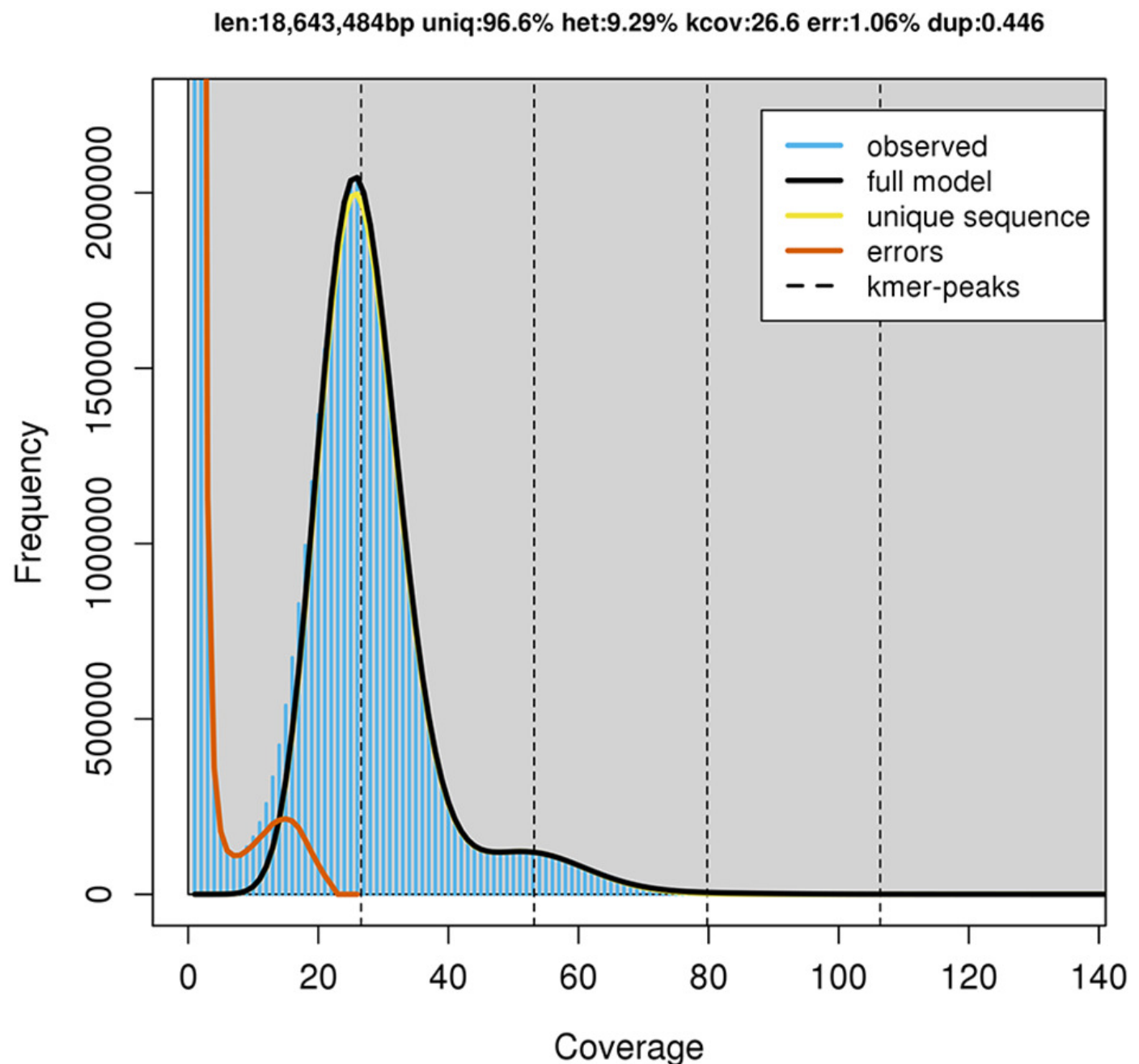
Figure 3

Pairwise average nucleotide identity calculation of Rhodotorula genomes.

Genomes with the superscript "T" are type strains.
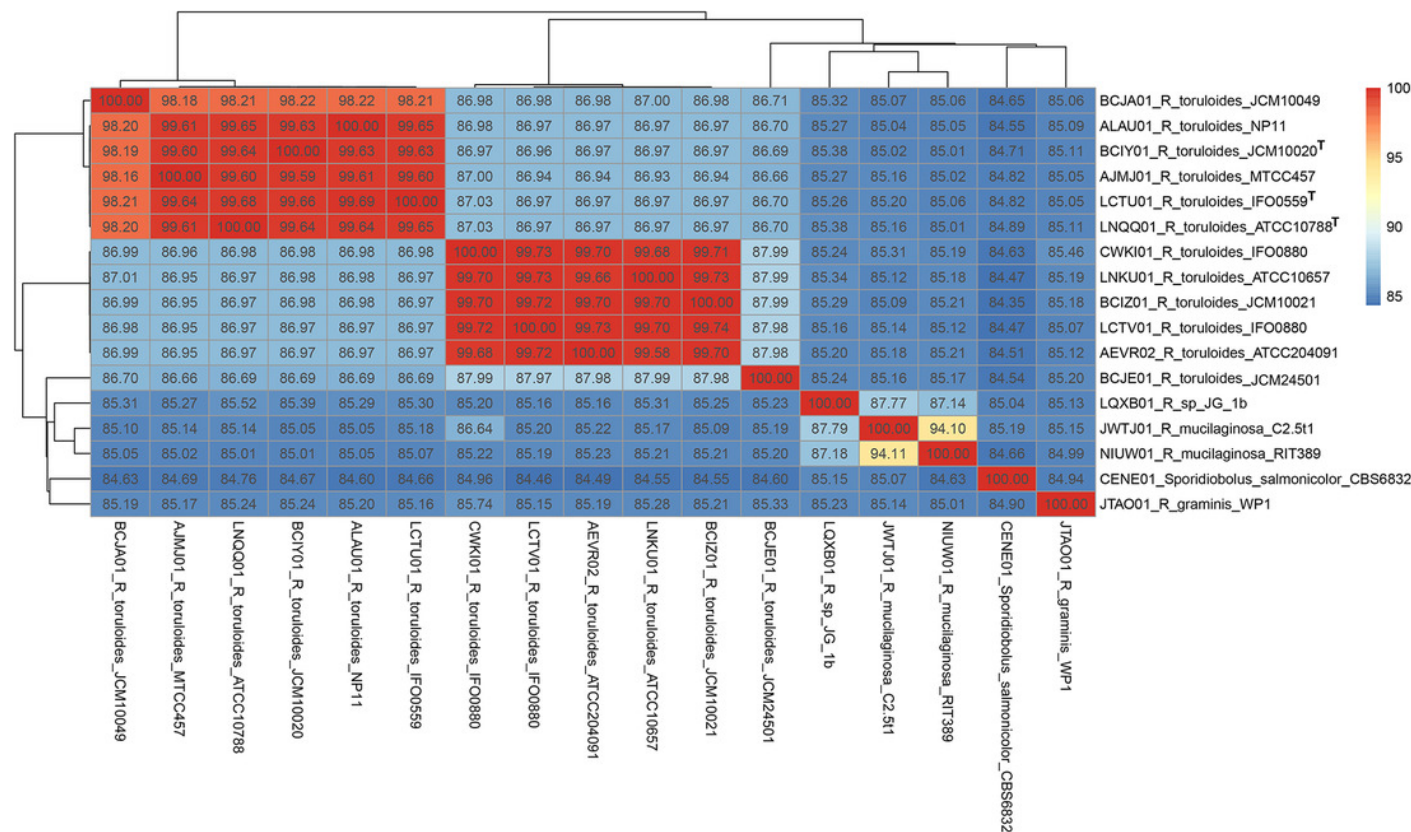
Figure 4

Maximum likelihood tree of a concatenated amino acid alignment consisting of 537,792 sites that represent 798 universally present single-copy genes from 19 fungal strains.

Labels on branches indicate shimodaira-hasegawa (SH)-like local branch support values. The scale bar indicates the average number of amino acid substitutions per site.
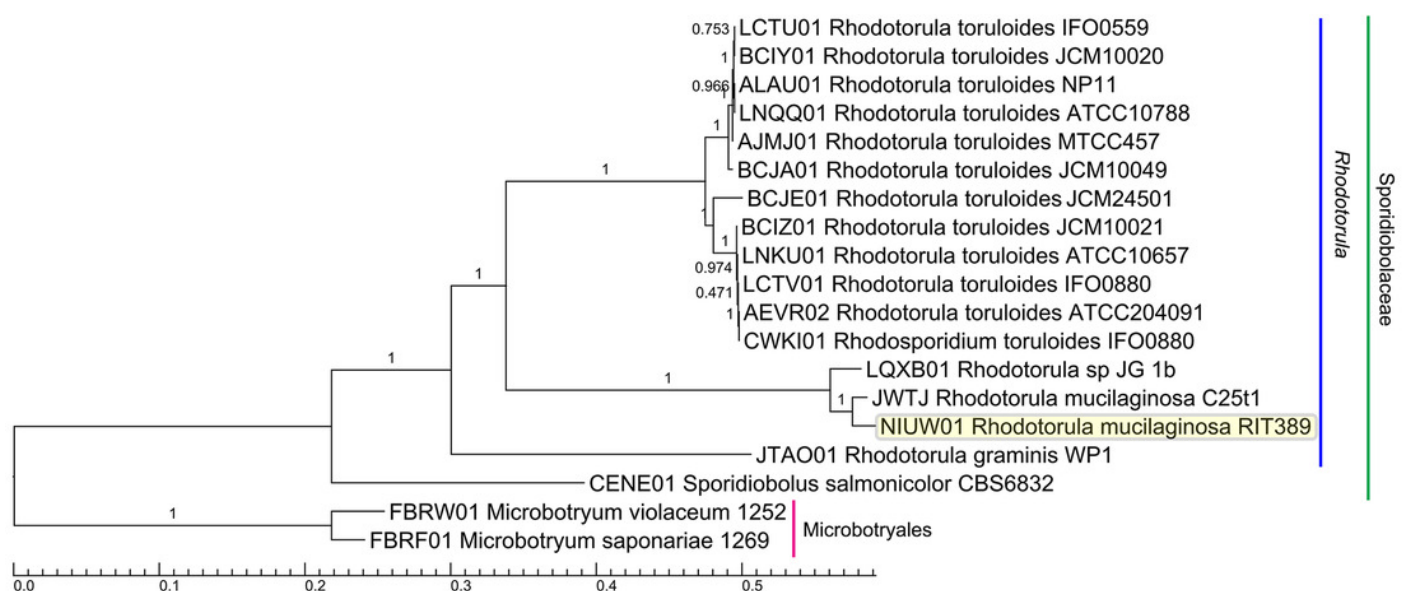
Figure 5

Complete mitochondrial genome of *R. mucilaginosa* RIT389 compared against that of *R. taiwanensis* RS1.

Orange frames indicate coding sequences commonly found in a typical mitochondrial genome. Red and blue arrows indicate transfer and ribosomal RNAs, respectively. Arrow direction represents transcriptional orientation. Dotted lines indicate intronic regions.
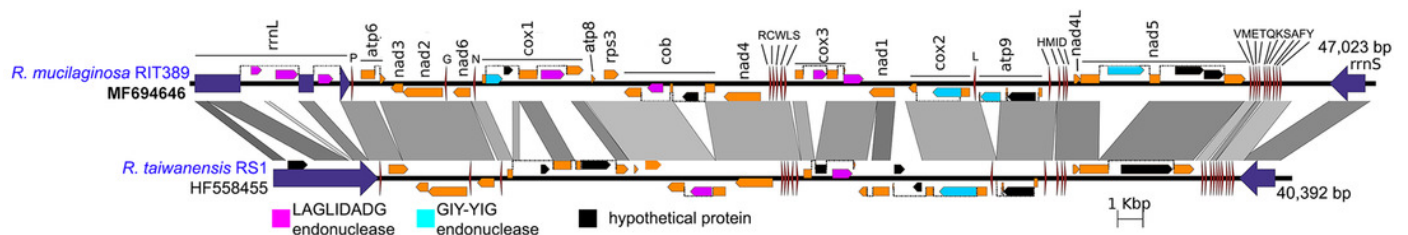
Figure 6

Comparison of genomic sub-region containing the gene cluster associated with carotenoid biosynthetic pathway.

Orange frames within the teal arrows indicate the coding sequences in the exonic regions of the corresponding genes.