A genome-scale metabolic reconstruction provides insight into the metabolism of the thermophilic bacterium *Rhodothermus marinus*

Thordis Kristjansdottir(1,2,*), Gudmundur O. Hreggvidsson(1,3), Sigmar Karl Stefansson(3), Elisabet Eik Gudmundsdottir(1), Snaedis H. Bjornsdottir(3), Olafur H. Fridjonsson(1), Eva Nordberg Karlsson(4), Justine Vanhalst(1), Birkir Reynisson(1), Steinn Gudmundsson(2)

(1) Matis, Vinlandsleid 12, 113. Reykjavik, Iceland

(2) University of Iceland, Department of Computer Science, School of Engineering and Natural Sciences, Dunhagi 5, 107 Reykjavik, Iceland

(3) University of Iceland, Department of Biology, School of Engineering and Natural Sciences, Sturlugata 7, 102 Reykjavík, Iceland

(4) Lund University, Department of Chemistry, Division of Biotechnology, Box 124, 221 00 Lund, Sweden

(*) Corresponding author, thordis@matis.is

Abstract

The thermophilic bacterium *Rhodothermus marinus* has mainly been studied for its thermostable enzymes. More recently, the potential of using the species as a cell factory and in biorefinery platforms has been explored, due to the elevated growth temperature, native production of compounds such as carotenoids and EPSs, the ability to grow on a wide range of carbon sources including polysaccharides, and available genetic tools. A comprehensive understanding of the metabolism of production organisms is crucial. Here, we report a genome-scale metabolic model of *R. marinus* DSM 4252^T. Moreover, the genome of the genetically amenable *R. marinus* ISCaR-493 was sequenced and the analysis of the core genome indicated that the model could be used for both strains. Bioreactor growth data was obtained, used for constraining the model and the predicted and experimental growth rates were compared. The model correctly predicted the growth rates of both strains. During the reconstruction process, different aspects of the *R. marinus* metabolism were reviewed and subsequently, both cell densities and carotenoid production were investigated for strain ISCaR-493 under different growth conditions. Additionally, the *dxs* gene, which was not found in the *R. marinus* genomes, from *Thermus thermophilus* was cloned on a shuttle vector into strain ISCaR-493 resulting in a higher yield of carotenoids.

Keywords: *Rhodothermus marinus*, Genome-scale metabolic model, Carotenoids

Importance

A biorefinery converting biomass into fuels and value-added chemicals is a sustainable alternative to fossil fuel-based chemical synthesis. *Rhodothermus marinus* is a bacterium that is potentially well suited for biorefineries. It possesses various enzymes that degrade biomass, such as macroalgae and parts of plants (e.g. starch and xylan) and grows at high temperatures (55-77°C) which is beneficial in biorefinery processes. In this study, we reviewed the metabolism of *R. marinus* and constructed a metabolic model. Such a model can be used to predict phenotypes, e.g. growth under different

environmental and genetic conditions. We focused specifically on metabolic features that are of interest in biotechnology, including carotenoid pigments which are used in many different industries. We described cultivations of *R. marinus* and the resulting carotenoid production in different growth conditions, which aids in understanding how carotenoid yields can be increased in the bacterium.

1. Introduction

Rhodothermus marinus is an aerobic bacterium that belongs to the phylum *Rhodothermaeota* [1]. It grows optimally at 65°C on various proteinaceous and carbohydrate substrates and was first isolated from a submarine hot spring in Iceland [2]. The *R. marinus* genome encompasses several gene clusters encoding pathways for utilization and cellular import of diverse carbohydrates. Several enzymes from *R. marinus* have been characterized, and many have biotechnological potential, including a number of polysaccharide degrading enzymes, such as cellulase [3], laminarinase [4] and xylanase [5], [6] (see [7] for a review). *R. marinus* grows on a wide range of sugars from second and third generation biomass and can be cultivated to relatively high yields [8]. It has anabolic pathways and precursor pools for production of various biotechnologically interesting primary and secondary compounds, such as polyamines [9], exopolysaccharides [10], carotenoids [11], compatible solutes [12] and lipids [13], [14].

Anaerobic fermentative organisms are generally preferred to produce low-value commodity chemicals, simple organic acids and alcohols that are typically catabolic waste products from incomplete oxidation of substrates. Conversely, heterotrophic aerobes such as *R. marinus*, typically oxidize their carbon substrates completely under optimized conditions, and therefore utilize organic substrates more efficiently for both energy and carbon. Consequently, aerobes can carry a greater metabolic burden and are the preferred organisms for the anabolic production of complex secondary metabolites.

Cultivation at high temperatures (60-70°C) may be beneficial in bioreactors as it reduces the cost of cooling, and higher temperatures protect the cultures from mesophilic spoilage bacteria. High temperature also increases the solubility of polysaccharides and leads to reduced viscosity of the

fermentation broth. This may alleviate scale up problems of mixing and aeration to a significant extent and enable greater substrate loadings. Elevated temperatures may also enable cost-effective recovery of volatile products by distillation or gas stripping, reducing product inhibition and prolonging the production phase of the culture [15].

R. marinus has the potential to serve as a robust production organism in the emerging biorefinery industry and as a chassis species that can be metabolically engineered for the production of novel chemical compounds of industrial interest. For this purpose, comprehensive knowledge and understanding of its metabolism are needed. A genome-scale metabolic reconstruction describes the metabolic network of a given organism based on both genomic information and available physiological data. A well-curated reconstruction contains a comprehensive overview of the metabolism of the organism in question. A reconstruction can be converted to a computational model, where phenotypic features are predicted under different conditions. The model can for instance be used to predict growth capabilities on different substrates and guide genetic engineering efforts [16]. Reconstructing the network of a poorly studied organism can result in gaps in the model. While this may skew predictions, it can nevertheless be useful for focusing future research efforts. Although *R. marinus* is not as well studied as common model organisms, it has been the subject of several studies, including the development and application of genetic tools [17], particularly for the engineering of the carotenoid pathway [18].

Carotenoids are pigments produced by many plants, fungi, algae and bacteria. Some non-photosynthetic bacteria produce carotenoids to stabilize cell membrane fluidity in response to extreme environments (high/low temperatures, pH, salinity, etc.) and to protect themselves against UV radiation and oxidative stress [19]. Carotenoids are in demand for different applications, such as the food, feed and cosmetic industries. In a previous study we engineered the carotenoid biosynthetic pathway in *R. marinus* to produce the industrially relevant carotenoid lycopene, instead of native γ -carotenoids [18]. In another, sequential batch cultivation resulted in higher carotenoid production

than shake flask cultivation [8]. Drawing upon the metabolic reconstruction of the current study, we further investigated the effects of culture conditions on carotenoid production and growth of *R. marinus*.

Here we reconstructed the genome-scale metabolic network of the *R. marinus* type strain DSM 4252^T, which has been the subject of most of the published studies so far and for which an annotated genome sequence is available [20]. However, the type strain is not amenable to genetic manipulation as it aggregates in liquid cultures and harbors a highly active DNA restriction enzyme with a 4 base recognition site [21]. Therefore, existing genetic tools were developed for another *R. marinus* strain, ISCaR-493 (DSM 16675), which was selected after screening of numerous *R. marinus* strains for a restriction-deficient phenotype [22]. Here, the genome of strain *R. marinus* ISCaR-493 was sequenced and the genomes of the two strains were compared. This analysis was used to find if any model genes from DSM 4252^T could not be found in ISCaR-493, and subsequently if the DSM 4252^T model could be extrapolated to strain ISCaR-493. Growth curves and uptake and secretion rates of the main metabolites from bioreactor cultivations were obtained for both strains and the data used to validate the model. During the reconstruction process, several interesting features related to carotenoid production were identified and investigated further. This included heterologous expression of a gene from *Thermus thermophilus* encoding the terpenoid biosynthetic enzyme 1-deoxy D-xylulose 5-phosphate (DXP) synthase, which was not identified in the genomes of *R. marinus*.

2. Methods

2.1 Strains, media and culture conditions

Three *R. marinus* strains were used in this study, DSM 4252^T, ISCaR-493 (DSM 16675) and the mutant strain TK-4 (ISCaR-493 derivative, $\Delta trpB\Delta purA::trpBdxs_{T.thermophilus}$, section 2.3). All cultivations were at 65°C and liquid cultures were set to shaking at 200 rpm. For each cultivation, *R. marinus* was first streaked on an agar plate containing rich medium, 10% medium 162 [23], with modifications (2 mM

MgSO₄ and 0.2 mM CaCl₂ in final volume) and addition of 1% NaCl, 0.03% K₂HPO₄, 0.1% yeast extract, 0.1% tryptone, 0.1% peptone, 0.05% glucose, 0.05% starch, 0.06% pyruvate and 0.018% Na₂CO₃. Utilization of different carbon sources (section 3.1) was examined on defined medium agar plates containing 10% medium 162 [23] with addition of 8 mM phosphate buffer (KH₂PO₄, Na₂HPO₄, pH=7.2), 0.1% vitamin solution [23] and 0.4% of each carbon source, except the amino acids which were 0.2%. *R. marinus* strain DSM 4252^T from a rich medium agar plate (see above) was resuspended in a drop of 0.9% NaCl solution and subsequently streaked on agar plates containing the different carbon sources. The plates were incubated at 65°C and growth was examined after 1, 3, 5 and 7 days.

The cultivations used to validate the model (section 3.3) were carried out in bioreactors (Labfors 5, Infors HT, Bottmingen, Switzerland), where pO₂ was kept at 40% with stirrer speed at 200-500 rpm and airflow as needed, pH at 7.2 with addition of 16.5% NH₄OH and the temperature at 65°C. Defined medium, which contained 10% modified medium 162 with the addition of 1% NaCl, 8 mM phosphate buffer (KH₂PO₄, Na₂HPO₄, pH=7.2), 10 mM NH₄Cl, 0.02% asparagine, 0.02% glutamine, 0.1% vitamin solution [23] and the carbon sources 1% glucose and 0.09% pyruvate, was used. The cultures were inoculated with 10% of pre-culture and were performed in duplicates. Samples were taken for OD (620 nm) and HPLC measurements every hour, until a stationary phase was reached (18 – 32h).

The cultivations for examining cell density and carotenoid production under different conditions (section 3.4) were performed in shake flasks under light and dark conditions. All cultures were exposed to day light and additional light from a halogen lamp, except when grown under dark conditions where the flasks were covered from the light. A defined medium (same as in section 3.3, except Wolf's vitamin- and trace elements solution from [24] were used), was used for these cultivations, but with different carbon sources: 1% glucose, 1% glucose and 0.09% pyruvate, 1% alginate, 0.5% glucose and 0.25% pyruvate, 0.09% pyruvate, 0.18% pyruvate, and without a carbon source for a negative control. For cultivation of strain TK-4, adenine (0.0025%) was also supplemented. Two pre-cultures were prepared for each liquid culture to be monitored. First, cells were transferred from a fresh rich medium

agar plate (see above) to defined liquid medium containing 0.5% glucose and 0.018% pyruvate and grown overnight (16h). This culture was used to inoculate (10%) a fresh defined liquid medium containing 1% glucose and 0.09% pyruvate, which was also grown overnight (16h). All the monitored cultures were inoculated (10%) with the second pre-culture. The cultures were stopped after 24h and cell density was estimated by measuring OD at 620nm in a spectrophotometer (Novaspec III⁺, Biochrom, Harvard Bioscience Inc., Holliston, Massachusetts, US).

All media components were autoclaved, except for the vitamin solution and the trace element solution, which were filter sterilized. The alginate was not sterilized due to browning and degradation when autoclaved. The probability of contamination in the alginate was low, because of the high growth temperature and because alginate is not a trivial carbon source for most bacteria. The alginate cultures were plated to verify that contamination had not occurred during growth. Individual colonies were obtained, which all had the characteristic red color, and were subsequently identified as *R. marinus* using MALDI-TOF (Microflex, Bruker, Billerica, Massachusetts, US). MALDI-TOF was used according to the manufacturer's instructions.

2.2 Analytical methods

The DNA content in *R. marinus* cells was estimated from exponential bioreactor cultures using the fluorometric Quant-iT[™] PicoGreen[™] dsDNA Assay Kit (ThermoFisher, Waltham, Massachusetts, US). A sample of freeze-dried biomass (approx. 10⁷ cells/ml) was dissolved in water and sonicated for 10 sec to lyse the cells. A standard curve was obtained from PicoGreen measurements of λ-dsDNA and used to estimate DNA concentration in *R. marinus* cells.

Glucose, lactate and acetate concentrations in the samples taken during bioreactor cultivations were measured using high-performance liquid chromatography (HPLC). The samples were filtered through 0.2 µm filters (Phenomenex) and metabolites subsequently quantified using the Dionex 2000 HPLC system (Dionex, Idstein, Germany) with a Rezex ROA-Organic Acid H + (8%, Phenomenex, Aschaffenburg, Germany) and a RI-101 detector (Shodex, München, Germany). Chromeleon evaluation

software version 6.80 (Dionex, Idstein, Germany) was used. Separation was obtained using 60°C column temperature with 0.2 mM sulfuric acid (Carl Roth, Karlsruhe, Germany) as the eluent at a flow rate of 600 µl/min for 30 min. External standards of HPLC grade (Merck, Darmstadt, Germany, Sigma-Aldrich, St. Louis, USA) were used. The pyruvate concentration was estimated using a pyruvic acid kit (Megazyme, Bray, Ireland). The two amino acids in the defined medium, asparagine and glutamine, were not measured. Their concentration in the medium was low and did therefore not influence the model predictions to a significant extent. OD (620 nm) values from bioreactor cultivations were converted to CDW (g/l) based on data obtained elsewhere [25], [26] ($CDW = 0.75 \times OD$).

For the carotenoid measurements, the cultures were diluted to OD = 1 at 620 nm. The cell pellets from 1 mL of the diluted cultures were mixed well with 1 mL of acetone and incubated in a sonication bath for 20 min. The samples were centrifuged at 16,000 g for 7 min, resulting in a colorless cell pellet and acetone supernatant containing the extracted carotenoids. The *R. marinus* carotenoids in acetone have maximum absorbance at ~480 nm (data not shown). The OD at 480 nm of the acetone extracts was measured in a 1 cm quartz cuvette in a spectrophotometer (Novaspec III⁺, Biochrom).

2.3 Genetic modification of *R. marinus*

The genomic DNA from *T. thermophilus* HB8 was isolated using the MasterPure Complete DNA purification Kit (Lucigen, Middleton, Wisconsin, US) and used as template for the amplification of the *dxs* (1-deoxy D-xylulose 5-phosphate synthase) gene (TTHA0006) (1848bp) by PCR. The amplification was performed using the Q5 High-Fidelity DNA polymerase (New England BioLabs, Ipswich, Massachusetts, US), according to the manufacturer's instructions. The primers were designed to support HiFi DNA assembly (NEBuilder HiFi DNA Assembly Master Mix, New England BioLabs): *dxs_thermus_F* (5'-AAACCATGGAGGTGCGCGATATGATCTTGGACAAGGTGAAC-3') and *dxs_thermus_R* (5'-AGCAGTTCGGTCTCGGCGGTGACATTCAGGCCCGTTCATGCACCAAG-3'), with the underlined bases matching the *dxs* gene and the rest matching a *Sall* (New England BioLabs) digested *R. marinus* shuttle vector pRM3000.0. The pRM3000.0 vector is the pRM3000 vector [17] with an ampicillin resistance

gene added, for selection in *E. coli*. The pRM3000.0-*dxs* vector was introduced into chemically competent *E. coli* (NEB 5-alpha) and plated on L-medium [27] with 100 µg/mL ampicillin. After incubation overnight at 37°C, positive clones were identified by amplifying the *dxs* gene, using the *Taq* DNA polymerase (New England BioLabs) according to the manufacturer's instructions. The primers *dxs_thermus_F* and *R* were used. The pRM3000.0-*dxs* vector was isolated from positive *E. coli* clones using the Monarch Plasmid Miniprep Kit (New England BioLabs).

The pRM3000.0-*dxs* vector was introduced into *R. marinus* SB-62 (ISCaR-493 derivative, $\Delta trpB \Delta purA$) by transformation [28], using the *trpB* gene for Trp⁺ selection. Competent *R. marinus* cells were prepared and transformed as described elsewhere [22]. The GenePulser Xcell electroporation system (Bio-Rad) was used for the transformation, with pulses delivered at 20 kV/cm. For each transformation, 40 µl of washed cells and 1 µg of DNA in ≤ 5 µL volume were used. The pRM3000.0 vector was used as a positive control and sterile MilliQ water as a negative control. Transformed cells were spread on agar plates containing medium 162 [23], with modifications (2 mM MgSO₄ and 0.2 mM CaCl₂ in final volume) and additions of 1% NaCl, 0.053% NH₄Cl, 0.2% soluble starch, 0.2% casamino acids, vitamin solution [23], 0.0025% adenine and 2.5% agar. Plates were incubated at 65°C for 5 days. Positive *R. marinus* clones were verified using the same PCR as positive *E. coli* clones.

2.4 Genome sequencing and analysis

R. marinus strain ISCaR-493 was cultured in rich liquid medium (see section 2.1). For sequencing by short-read technology, genomic DNA was extracted using the MasterPure Complete DNA purification Kit (Lucigen) and sequencing libraries made by both the Nextera XT (FC-131-1024) and Nextera Mate Pair (FC-132-1001) methods. The two resulting libraries were sequenced on the MiSeq sequencing platform using V3 2x 300bp and V2 2x 250bp chemistry, respectively. The sequence reads were quality assessed using FastQC (v0.11.7) [29] and trimmed using Trimmomatic (v0.39) [30]. A genome was assembled using SPAdes (v3.14.0) [31] with flag for isolates. Genome polishing was done by mapping all shotgun short-reads (Nextera XT) to the largest contig from the *de novo* assembly and generating a

consensus sequence using the highest quality bases, using Geneious (v9.1.4). The genome was annotated using the PGAP pipeline from NCBI [32].

2.5 Metabolic reconstruction and simulations

A protocol for reconstructing genome-scale metabolic models [33] was used to guide the reconstruction process. A draft reconstruction, based solely on the annotated genome sequence, was obtained using the Model SEED [34]. Since no well-curated model of a closely related bacterium was available at the time of the reconstruction, the obtained draft was neither comprehensive nor functional. It was therefore used as a reference model, along with the well-curated model of *E. coli* (iJO1366) [35]. The *R. marinus* model was manually reconstructed based on these two reference models and *R. marinus* experimental data, obtained here and from the literature. The BiGG [36], BRENDA [37], KEGG [38] and MetaCyc [39] databases and the BLAST tool [40] were extensively used.

Cobrapy [41] was used in all model simulations, along with the GLPK solver. The corresponding code can be found as a Jupyter notebook on Github (<https://github.com/steinng/rmarinus>). For all simulations, flux balance analysis (FBA) was used [42], [43]. Exchange reactions corresponding to metabolites taken up from the media (glucose and pyruvate) and secreted (lactate and acetate) during growth, were constrained with experimentally obtained rates (Supplementary file 1). FBA was subsequently used to optimize for growth by maximizing flux through the biomass reaction.

For accurate growth rate predictions, the biomass reaction should ideally be based on data obtained for the target organism. Here, the biomass reaction was formulated based mostly on available data on *R. marinus* (Supplementary file 2). Separate biosynthetic reactions for each group of macromolecules (protein, lipid, DNA, etc.) were formulated, describing the ratio of the building blocks (amino acids, fatty acids, nucleotides, etc.) and the energy required. Sensitivity analysis, which shows how much variation in each macromolecule affects the predicted growth rate, was performed. This analysis helps to identify which biomass components most urgently need to be accurately measured.

3. Results and discussion

3.1 Reconstruction of a genome-scale metabolic model of *R. marinus* DSM 4252^T

A genome-scale metabolic model of *R. marinus* DSM 4252^T was reconstructed, named Rmarinus_578 (<https://github.com/steinng/rmarinus>). The main features of the reconstruction are listed in Table 1. The Memote tool [44] was used to help guide the reconstruction process, by verifying stoichiometric consistency, mass and charge balance and annotation quality (Supplementary file 3). Reactions and metabolites were usually abbreviated in accordance with the BiGG database and annotations with links to external databases are included. The genome sequence for strain DSM 4252^T was obtained from GenBank (accession nr: NC_013501). The genes in the reconstruction were identified with the locus tags from the GenBank file. They were annotated with the old gene locus tag from the GenBank file, the protein ID, protein annotation and protein sequence. Experimental data on *R. marinus* obtained in this study and the available literature was used to curate reactions, genes and gene-protein-reaction (GPR) rules. Several metabolic features were reviewed during the reconstruction process. In the following, we highlight a few, which are of interest for biotechnological application of *R. marinus*.

Sugar metabolism

R. marinus produces pyruvate from glucose through the Embden-Meyerhof-Parnas (EMP) pathway. A ¹³C metabolic flux study of the central metabolism in *R. marinus* [45] showed that the EMP pathway and the TCA cycle are both highly active while metabolizing glucose. The oxidative pentose phosphate pathway and the glyoxylate shunt had very low activity and the Entner-Doudoroff (ED) pathway, malic enzyme and phosphoenolpyruvate carboxykinase were inactive.

Growth of *R. marinus* strain DSM 4252^T was tested on many different carbon sources, both *in vivo* and *in silico* (Table 2). Growth has been shown on several mono-, di- and polysaccharides, which was also observed *in silico*. However, growth on cellulose was predicted *in silico* while not observed *in vivo*. *R. marinus* does contain genes encoding a endocellulase (EC 3.2.1.4) [3], [46], which can degrade cellulose

into differently sized cellooligosaccharides and cellobiose. However, low specific activities were reported compared to benchmark enzymes. The strain does not contain genes encoding specific exocellulases (EC 3.2.1.91), which degrade glucans into β -cellobioses. However, it does contain several genes encoding glycoside hydrolases, that can degrade different oligosaccharides, including cellobiose and cellooligosaccharides, to monosaccharides [47]. Although *R. marinus* might be able to use some of the cellulose partially degraded by the cellulase and glycoside hydrolases, as the model predicts, the cellulose-specific activity of the enzymes is probably not high enough for it to grow solely on cellulose. *R. marinus* was unable to grow *in vivo* on several of the tested carbon sources, especially organic- and amino acids (Table 2). The model predictions usually showed the same results. However, when transport reactions were added to the model, growth was often observed *in silico* for carbon sources that did not result in growth *in vivo*. These carbon sources might be good targets for adaptive evolution experiments. Promiscuous enzymes might adapt to transport these compounds into the cell, where they can presumably be used for growth [48].

Lactose and galactose are examples of the di- and monosaccharides that *R. marinus* can use for growth (Table 2). The gene encoding a β -galactosidase (EC 3.2.1.23), which hydrolyzes lactose into glucose and galactose, was found in the genome. Three steps are needed to convert galactose to glucose-6-phosphate, which then enters the glycolysis EMP pathway. The genes encoding galactokinase and galactose-1-phosphate uridylyltransferase, catalyzing the first two steps (galactose $\rightarrow$ galactose-1-phosphate $\rightarrow$ glucose-1-phosphate), were found in the genome. The third step, where glucose-1-phosphate is turned into glucose-6-phosphate, is usually performed by the enzyme phosphoglucomutase (EC 5.4.2.2). The gene for this enzyme was not found in the genome. However, a homology search showed similarity between known phosphoglucomutase genes from other bacteria and genes RMAR_RS01880 (E value $2e-37$) and RMAR_RS08875 (E value $1e-25$) which are annotated as phosphomannomutase (EC 5.4.2.8) and phosphoglucosamine mutase (EC 5.4.2.10), respectively. The enzyme phosphoglucosamine mutase in *E. coli*, which usually catalyzes the interconversion of glucosamine-6-phosphate and glucosamine-1-phosphate, was also shown to be able to catalyze the

interconversion of glucose-6-phosphate and glucose-1-phosphate, at a lower rate [49]. The phosphorylation site of this enzyme in *E. coli* is Ser102 and a mutational change of Ser100 to a threonine residue increased the phosphoglucomutase activity significantly. Gene RMAR_RS08875 from *R. marinus* was investigated and the serine residue responsible for the phosphorylation was found to be residue number 103. The corresponding residue to Ser100 in the *E. coli* enzyme was found to be a threonine, which indicates that this *R. marinus* enzyme may be responsible for the interconversion of glucose-6-phosphate and glucose-1-phosphate in *R. marinus*.

R. marinus possesses several genes encoding polysaccharide degrading enzymes. As a marine bacterium, seaweed is common in its natural environment. Alginate and laminarin are major polysaccharides of brown algae, which *R. marinus* can break down and use as sole carbon sources for growth [50]. Alginate is a structural component of brown algae and can comprise up to 40% of its dry matter [51]. It is a polyuronate that consists of β -D-mannuronate (M) and α -L-guluronate (G) units forming (1 $\rightarrow$ 4) linked G-, M- and mixed blocks in the polysaccharide chain. The *R. marinus* genome has four genes encoding alginate lyases [52], [53] that, together, depolymerize alginate into the same unsaturated mono-uronate derivative of the M and G units. The *R. marinus* genome also possesses the genes encoding for the remaining enzymes of the alginate catabolic pathway enabling its utilization. The unsaturated monouronate is converted to 4-deoxy l-erythro 5-hexoseulose uronic acid (DEH) by a spontaneous reaction and further catalyzed to 2-keto 3-deoxygluconate (KDG) by an aldose reductase [54]. KDG enters the partial ED pathway in *R. marinus*, where it is catalyzed to 2-keto-3-deoxygluconate 6-phosphate (KDPG) by 2-keto 3-deoxygluconokinase (EC 2.7.1.45) and then by 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14), to pyruvate and glyceraldehyde 3-phosphate which enter central metabolism [54], [55] (Figure 1). *R. marinus* does not possess a key enzyme in the ED pathway, phosphogluconate dehydrogenase (EC 4.2.1.12), which is needed to metabolize glucose. This explains why the ED pathway is not active when *R. marinus* is grown on glucose [45], while the partial pathway is essential for utilization of alginate (Figure 1).

Terpenoid and carotenoid metabolism

The carotenoids produced by several *R. marinus* strains have been characterized [11], [56]. A carotenoid pathway was proposed in [11] using structural and bioinformatic data, and later refined [18]. This pathway is included in the reconstruction along with candidate genes. Terpenoids, which serve as precursors to carotenoids, can be produced through two different pathways, the mevalonate [57], [58] and the non-mevalonate (MEP) [59] pathways. In the latter, which is used by most bacteria including *R. marinus*, geranylgeranyl diphosphate (GGDP) is produced from pyruvate and glyceraldehyde 3-phosphate in multiple steps. *R. marinus* possesses genes coding for all the enzymes in the MEP pathway, except for 1-deoxy D-xylulose 5-phosphate (DXP) synthase (DXS, EC 2.2.1.7), which catalyzes the first step in the pathway. Studies of the MEP pathway in other bacteria have shown that DXS is not strictly necessary for the synthesis of DXP. Examples include a mutated form of pyruvate dehydrogenase that is known to rescue *E. coli* cells defective in DXS [60] and a mutated RibB protein and a YajO protein that synthesized DXP from ribulose 5-phosphate, also in *E. coli* [61]. Another possibility to bypass DXS is via the MTA-isoprenoid shunt, as has been shown in *Rhodospirillum rubrum* [62]. Here, the dead-end metabolite of polyamine biosynthesis, 5-methylthioadenosine (MTA) is metabolized by an alternative methionine salvage pathway, which produces DXP as a side-product. Phylogenetic analysis showed that the genes in this pathway are partially present in the *R. marinus* genome. At present, it is not known how DXP is produced in *R. marinus* and without further evidence of alternative pathways, the DXP synthase reaction is present in the reconstruction without any gene candidates assigned. The absence of DXS in *R. marinus* directly suggests heterologous expression of a thermostable DXS as means to increase flux through the terpenoid pathway.

Light-inducible carotenoid production has been observed in many organisms, including non-photosynthetic bacteria, and the regulatory mechanisms have been studied in some of them, including *Myxococcus xanthus* [63], *Thermus thermophilus* [64], *Streptomyces coelicolor* [65] and *Bacillus megaterium* [66]. The MerR family transcriptional regulator, LitR, acts as a repressor in the carotenoid gene cluster. Its activity is dependent on the binding with adenosyl B₁₂ and the LitR-AdoB₁₂ complex

becomes inactivated when illuminated. This causes cell cultures to become colorless under dark conditions while producing carotenoids in light. A homologue for the *T. thermophilus litR* gene was previously identified in the carotenoid gene cluster in *R. marinus* [18], located upstream of the carotenoid gene *crtB* (phytoene synthase). This suggests that light might help increase carotenoid yields in *R. marinus*.

Other pathways

The respiratory chain in *R. marinus* has been extensively studied. The first two complexes, NADH dehydrogenase and succinate dehydrogenase, have been characterized [67], [68] and found to be similar to those of other bacteria. The third complex, cytochrome dehydrogenase, is not the typical *bc₁* but an alternative complex [69]. It has an entirely different structure but carries out the same function, oxidizing reduced menaquinones-7 [70] and reducing High Potential iron-sulfur Protein (HiPIP) and cytochrome *c*. Finally, three different types of the fourth complex have been characterized in *R. marinus*, *cbb3* [71], *caa3* [72] and *ba3* [73]. These reactions and associated genes are included in the reconstruction.

Polyamines are alkaline organic compounds with at least two primary amino groups. They can be found in most forms of life and have diverse functions. Some polyamines are biotechnologically interesting because they can be used to produce plastics and are used as curing agents in polymer applications [74]. In thermophilic bacteria unusual, long and branched polyamines have been observed [75]. They are believed to have protective effects on nucleic acids and proteins under high-temperature conditions. Seven different polyamines have been characterized in *R. marinus* to date, putrescine, spermidine, cadaverine, spermine, thermopentamine, N4-aminopropylspermidine and N4-bis(aminopropyl)spermidine [9], and their biosynthesis is included in the reconstruction. *R. marinus* produces compatible solutes to protect the cell against sudden osmotic changes. They include amino acids, monosaccharides such as trehalose, small peptides and, most abundantly, mannosylglycerate [76]. Mannosylglycerate in *R. marinus* is synthesized via two pathways [77], which have been studied

in detail, along with the corresponding enzymes and genes [14], [77], [78]. These pathways are present in the metabolic reconstruction, but as they are produced in response to stress, they are not included in the biomass reaction and the pathways therefore not active during growth simulations where the biomass is maximized.

Genomic information indicates that *R. marinus* can synthesize all the amino acids needed for protein synthesis. This is supported by growth experiments which show that *R. marinus* can grow in defined medium, without any addition of amino acids [79]. Biosynthetic pathways for all the 20 amino acids are included in the reconstruction. The fatty acid and lipid composition of *R. marinus* have been characterized [13], [14]. The dominating fatty acids are iso- and anteiso-C15 and iso- and anteiso-C17, with iso-C16 and iso-C18 as minor components. The major polar lipids are phosphatidylethanolamine, diphosphatidylglycerol and one unidentified lipid and phosphatidylglycerol was identified as a minor lipid [14], [70]. The biosynthesis of the fatty acids and lipids are also included in the reconstruction.

The biomass objective function

In genome-scale metabolic models a biomass objective function (BOF) is used to simulate growth. The BOF describes the ratio between the macromolecules (protein, DNA, RNA, lipids, etc.), the composition of each macromolecule (amino acids, nucleotides, fatty acids, etc.) and the energy that the cell requires to grow and maintain itself. We collected data on several studies that describe different components of the biomass in *R. marinus*. Protein, RNA, lipid and glycogen content and amino acid ratio in the biomass were based on [45], carotenoid composition on [56], EPS on [10], lipids on [13], [14] and polyamines on [9]. Quantification of DNA in the biomass was measured in this study (Supplementary file 2). Nucleotide compositions of DNA and RNA were derived from the DSM 4252^T genome. The RNA estimate assumes equal transcription of all genes and is therefore expected to be somewhat inaccurate. Growth associated maintenance (GAM) was estimated using experimental data obtained here (supplementary file 1). The remainder of the biomass components was adopted from *E. coli* [80].

A detailed overview of the biomass, sources of information and calculations can be found in supplementary file 2.

Sensitivity analysis was performed to investigate how sensitive the predicted growth rate was to changes in biomass and energy components. The growth rate was predicted while varying each component by 50%, for multiple glucose uptake rates. The components tested were protein, DNA, RNA, lipid, exopolysaccharide, lipopolysaccharide, glycogen, peptidoglycan, GAM and non-growth associated maintenance (NGAM) (Supplementary file 2). The predicted growth rate was most sensitive to changes in the protein component, followed by the lipid component.

3.2 Genome comparison of strains DSM 4252^T and ISCaR-493

The metabolic model was reconstructed based on genomic information from *R. marinus* DSM 4252^T. This strain is, however, not amenable to genetic manipulation as it aggregates in liquid cultures and shows high DNA degrading activity due to the presence of a 4 cutter restriction enzyme, *RmaI* [21]. Therefore, genetic tools were developed for a derivative of the strain ISCaR-493 [22]. The 16S rRNA sequence similarity between the two strains is 97.92%.

Core genome

The genome of ISCaR-493 was sequenced using Illumina MiSeq and assembled using SPAdes. Subsequently, the 60 resulting contigs were annotated using the NCBI annotation software PGAP. The resulting draft genome was compared to the genome of strain DSM 4252^T using the in-house pan-genomic software Genset. The genome of ISCaR-493 had a calculated G+C% of 64.6% compared to 64.3% for strain DSM 4252^T. The Average Nucleotide Similarity (ANI value) [81] was 95.5%, indicating a relatively high sequence divergence while the genome metrics were very similar for the two strains (Supplementary file 4).

The number of total genes was 2890 and 2937 for ISCaR-493 and DSM 4252^T, respectively and protein-coding genes were about 98.3% thereof for each strain. 2609 protein-coding genes belonged to the

common core having 50% identity across at least 50% of the protein. 74% of the core protein-coding genes could be assigned to COG functional categories [82]. The remaining protein-coding core genes (26%) did not get COG IDs and had unknown (S) or poorly characterized functions (R). 230 protein-coding genes were unique to ISCaR-493 and 275 genes were unique to DSM 4252^T, a total of 505 genes that comprised the peripheric gene fraction. About 50% of the protein-coding genes in the peripheric fractions of both strains could not be assigned function compared with approximately 26% in the common core (Figure 2).

Accessory genome

Compared with the core genome the peripheric fractions of both strains were proportionally enriched in the following COG subcategories: V (Defense), M (Cell wall/membrane/envelop biogenesis), and X (Mobilome: Prophages and transposons) (Figure 2). Two restriction enzymes were shown to be only in the peripheric genome of DSM 4252^T, including the *RmaI* 4 cutter. This reflects the different restriction phenotypes observed [21] and the probably explains the facilitated uptake of foreign DNA in ISCaR-493 compared to DSM 4252^T.

A total of 578 genes were used in the metabolic network modelling of *R. marinus* 4252^T and the great majority of them belonged to the core genome. Only 7 model genes were not found in the genome of ISCaR-493 and only three of them did not have isozymes in the genome that showed high similarity to DSM 4252^T genes (Table 3).

Strain DSM 4252^T contains two genes encoding xylanases while strain ISCaR-493 contains only one homologue. Both strains grow on xylan as the sole carbon source (Table 2 for DSM 4252^T, data not shown for ISCaR-493). Strain DSM 4252^T contains four genes encoding alginate lyases and one of them is missing in ISCaR-493. The latter strain grows well in a medium with alginate as the sole carbon source [50], suggesting that the three alginate lyases are sufficient to degrade alginate for utilization.

Many enzymes take part in EPS (the envelope polysaccharides) biosynthesis and assembly, and their corresponding genes were all found in the core genome. A putative o-antigen polymerase (*wzy*) found

in the accessory genome of strain ISCaR-493 showed a low similarity to a functionally corresponding protein in DSM 4252^T (E-value 0.004). However, the encoded gene showed high similarity to genes in more distantly related bacteria annotated as o-antigen ligase and o-antigen polymerase. This enzyme activity is essential for EPS synthesis and must be present in ISCaR-493 as EPS is produced [10]. The relatively high content of genes in the accessory genomes belonging to subcategory M (Cell wall/membrane/envelop biogenesis), including glycosyl transferases and sulfotransferases, indicates that the EPS of the two strains may be different in sugar composition and sulfatation patterns and this may explain the differently aggregating phenotypes of the strains.

The above analysis suggests that the Rmarinus_578 model can be used for both strains DSM 4252^T and ISCaR-493.

3.3 Model validation for strains DSM 4252^T and ISCaR-493

Experimental growth data was obtained for strains DSM 4252^T and ISCaR-493 in bioreactors, along with measured uptake of glucose and pyruvate and secretion of acetate and lactate (Supplementary file 1). The average rates of two replicates (Supplementary file 1) was used to constrain the model to validate the accuracy of growth predictions (Figure 3a). The experimental growth data showed that the growth of *R. marinus* DSM 4252^T did not follow the typical batch growth curve of bacteria. A true exponential phase was not observed throughout the growth phase. Instead, the apparent specific growth rate decreased over time until stationary phase was reached (Figure 3b). The specific growth rate in strain ISCaR-493 was closer to being constant, with exponential growth during a longer period (Figure 3b). Data from time points 3-6 for strain DSM 4252^T and 1-5 for strain ISCaR-493 was used here. This analysis showed that the model accurately predicts growth for both strains (Figure 3a).

For both strains, but more so for DSM 4252^T, secretion rates of lactate and acetate increased during the growth phase (Supplementary file 1). A decrease in growth rate during batch cultivations has been observed in other bacteria, such as *E. coli* [83] where the main reason was oxygen limitation that could also lead to an increase in organic acid secretion. The cultivations here were carried out with high

aeration as oxygen levels were kept fixed at 40% pO₂. A plausible explanation for why the cells would experience oxygen limitation in a medium with excess oxygen levels is local limitation due to cell aggregation [84]. Aggregation of several *R. marinus* strains has been reported previously [22], especially in DSM 4252^T and *R. marinus* is also shown to produce exopolysaccharides [10], which can cause cells to aggregate [85].

When the model was optimized for growth, without oxygen limitation and free secretion of acids, it did not predict any acid production and the predicted growth rate was slightly higher than observed *in vivo*. When oxygen was limited in the model, the predicted growth rate decreased, and the model predicted lactate secretion (data not shown). Experimental data showed that lactate was first secreted, followed by acetate (Supplementary file 1). The model predicted slightly higher growth rate when lactate was the sole acid produced, opposed to when it was forced to also produce acetate.

3.4 Carotenoid production and growth of *R. marinus* ISCaR-493

To better understand the carotenoid production in *R. marinus*, a cultivation experiment comparing different conditions was performed. Besides obtaining high yields of carotenoids per cell, high cell density is important for achieving high yields of carotenoids. Therefore, both extracted carotenoids (from 1 mL of cells diluted to OD_{620 nm} = 1) and cell densities were measured from cultivations after 24 hours (Figure 4). The ISCaR-493 strain was used in this experiment, as it can be genetically modified and thus likely to be used for future cell factory designs

Glucose and pyruvate

R. marinus can grow on several monosaccharides, as predicted by the model. However, we have often observed better growth on oligo- and polysaccharides (data not shown). Growth of strain ISCaR-493 in defined medium with glucose (1%) as the sole carbon source resulted neither in high cell density nor high carotenoid production (Figure 4). *R. marinus* can utilize pyruvate as the sole carbon source (Table 2). Pyruvate is used in several pathways essential for growth and is the substrate, together with

glyceraldehyde 3-phosphate, in the first step of the MEP terpenoid pathway (Figure 1 To increase both cell density and carotenoid production, pyruvate (0.09%) was added to the glucose-based medium. This resulted in increased carotenoid production and highly increased cell density (Figure 4). Visually, these cultures exhibited much stronger red color than the glucose cultures, which can be explained by both increased carotenoids yields and higher cell densities.

Impact of light

During the model reconstruction, a homolog for a light dependent regulatory gene was found in the carotenoid gene cluster in *R. marinus*. This indicates that carotenoid production in *R. marinus* is light induced (section 3.1). This was investigated here by cultivating ISCaR-493 in glucose (1%) and pyruvate (0.09%) medium in the dark and compared to the corresponding cultures grown in the light. The former cultures were colorless and the lack of carotenoids was confirmed by measurements (Figure 4).

Alginate

R. marinus can grow on many different polysaccharides (Table 2), making it an interesting candidate for processing 2nd or 3rd generation biomass, such as seaweed. Alginate is one of the major polysaccharides of brown algae. The products from alginate degradation are pyruvate and glyceraldehyde 3-phosphate (Figure 1), which are the same metabolites as used in the first step of the MEP terpenoid pathway. This raised the question whether *R. marinus* produces more carotenoids when grown on alginate, since it produces the two metabolites needed for the biosynthesis concurrently and in equal amounts. Cultivations in defined medium with alginate (1%) as the sole carbon source showed less cell density compared to glucose and pyruvate, but highly increased carotenoid production (Figure 4).

Glucose and pyruvate in equal quantities

To further examine if the availability of glyceraldehyde 3-phosphate and pyruvate in equal amounts results in higher carotenoid production, cultivation in defined medium with glucose (0.5%) and pyruvate (0,25%) was investigated. These cultivations showed lower cell density and higher carotenoid

production compared to growth on glucose (1%) and pyruvate (0.09%) (Figure 4). The increased carotenoid production could be due to the equal availability of the two metabolites. Another possibility is that increased concentration of pyruvate alone in the medium caused higher carotenoid production.

Pyruvate

To examine if pyruvate alone affects the carotenoid production, two additional cultivations were set up, with pyruvate (0.09% and 0.18%) as the sole carbon source. The cell density in these cultures was low, only increased slightly after inoculation. This indicated that ISCaR-493 struggles to grow in liquid defined medium with pyruvate as the sole carbon source, which was surprising as growth was observed on agar medium (Table 2). The carotenoids per fixed cell density in the pyruvate cultures were much higher compared to cultures on glucose (1%) and pyruvate (0.09%). Additionally, increased pyruvate concentration resulted in increased carotenoid production (Figure 3). This suggests that the pyruvate is used for carotenoid production. Producing glyceraldehyde 3-phosphate from pyruvate costs energy (gluconeogenesis) and it cannot be determined from this data if this is the case for the observed growth. However, glycogen is an alternative source of glyceraldehyde 3-phosphate. The amount of glycogen in the biomass of *R. marinus* has been estimated as 14% [45] and is relatively high compared to other bacteria. Inclusion that could possibly contain glycogen can be discerned on electron micrographs of *R. marinus* [7]. Considering the natural habitat of *R. marinus* in coastal hot springs, it is not unreasonable to assume that it accumulates high levels of glycogen. Due to tides, the availability of nutrients in the surroundings of *R. marinus* varies widely and it is likely that the bacterium stores carbohydrates when they are in abundance in the environment. Since little or no growth was observed on pyruvate in liquid cultures it is likely that the cells experienced starvation and therefore started the breakdown of glycogen and carotenoid production. This was also seen for the negative control cultures without a carbon source (Figure 4). The cell density did not increase from inoculation, while the carotenoid production did.

Addition of the *dxs* gene from *T. thermophilus*

In an effort to increase carotenoid yields, the *dxs* gene from *T. thermophilus* was cloned on a shuttle vector into *R. marinus* strain SB-62 (ISCaR-493 derivative, $\Delta trpB\Delta purA$), resulting in the mutant strain TK-4 ($\Delta trpB\Delta purA::trpBdxs_{T.thermophilus}$) (Supplementary file 5). The *dxs* gene encodes 1-deoxy-D-xylulose-5-phosphate synthase (DXS), which catalyzes the first step in the MEP terpenoid pathway (section 3.1) and could not be identified in the genomes of *R. marinus*. Compared to ISCaR-493, cultivation of TK-4 resulted in lower cell density but highly increased carotenoid production. Presumably the added *dxs* gene resulted in a higher flux of carbons through the terpenoid and carotenoid pathways. However, it is also possible that this strain struggles to grow and responds by producing carotenoids. The dramatically lower cell density compared to ISCaR-493 can most likely be explained by the metabolic burden caused by the replication of the shuttle vector and the expression of its genes. Inserting the *dxs* gene in to the chromosome could reduce such effects.

In summary, these experiments showed that the highest cell density was obtained in glucose medium supplemented with pyruvate, while higher carotenoid production was observed during growth on alginate, with pyruvate added to a glucose-based medium and in the presence of light. It also showed that the carotenoid production per cell increased during starvation, indicating that yields can potentially be increased by either allowing the culture to reach and stay in stationary phase or transfer the cells after growth to new medium with limited or no carbon source. The motivation for the latter is that after the growth phase, the medium might not be optimal, e.g. due to accumulation of by-products that alter the pH, and the cells might stay alive and produce carotenoids longer in fresh medium. Finally, cloning the *dxs* gene from *T. thermophilus* in *R. marinus* resulted in the highest yields of carotenoids, but much lower cell density than the wild type strain ISCaR-493.

4. Conclusions

A manually curated genome-scale metabolic model of *R. marinus* DSM 4252^T was reconstructed and made publicly available (<https://github.com/steinng/rmarinus>). Experimental data from the literature and from this study was used to curate and validate the model. This includes growth data on various carbon sources, bioreactor cultivations and HPLC measurements of main metabolites, used for model validation, multiple studies on different metabolic pathways, components, genes and enzymes, and data on biomass components, which was used to formulate a species-specific biomass objective function.

The model was also evaluated for use with *R. marinus* ISCaR-493, from which the genetically modified SB-62 ($\Delta trpB \Delta purA$) was derived. The genome of strain ISCaR-493 was sequenced and the resulting draft genome was compared to that of strain DSM 4252^T. This analysis showed that only seven model genes were absent in strain ISCaR-493 and four of them were replaced by genes encoding isozymes that exhibited high similarity to the DSM 4252^T enzymes. The remaining three genes are involved in EPS formation and xylan- and alginate degradation. EPSs of both strains have been previously studied [10] and shown to be of similar structures. It was also observed that strain ISCaR-493 grows well in defined medium with xylan and alginate as the sole carbon sources. In conclusion, this analysis suggests that the model is applicable for both strains DSM 4252^T and ISCaR-493. Both strains should be considered when any future changes or additions to the model reconstruction are made. Data on growth and metabolites was used to constrain the model and compare the experimental and simulated growth rates. This revealed that the model predicts correct growth rates for both strains.

Different aspects of the metabolism of *R. marinus* were reviewed during the reconstruction process. Here, an emphasis was on those with a potential biotechnological aspect, carotenoids in particular. Cell density and carotenoid production of strain ISCaR-493 grown at different conditions were investigated. Pyruvate addition to a glucose-based medium, highly increased cell density was obtained. Carotenoid production varied considerably under different growth conditions. Higher carotenoid

yields were observed when pyruvate was present in the growth medium, alginate was used as the sole carbon source, cultivating the cells in light conditions and the cells experienced starvation. Additionally, we cloned the *dxs* gene from *T. thermophilus* on a shuttle vector into *R. marinus* and cultivation of the resulting mutant showed low cell density compared to ISCaR-493, but higher carotenoid production.

With its thermostable enzymes, wide range of potential carbon sources for growth and marketable products, *R. marinus* application potential is highly relevant in biotechnology and biorefineries. A genome-scale metabolic model helps us to understand its metabolism and should be useful in future strain designs.

596 Declaration of competing interest

597 None

598 CRediT author statement

599 **Thordis Kristjansdottir:** Conceptualization, Methodology, Investigation, Writing – Original Draft,
600 Writing – Review and Editing, Visualization. **Gudmundur O. Hreggvidsson:** Conceptualization,
601 Investigation, Supervision, Funding acquisition, Writing – Original Draft, Writing – Review and Editing.
602 **Sigmar Karl Stefansson:** Methodology, Investigation, Writing – Review and Editing, Visualization.
603 **Elisabet Eik Gudmundsdottir:** Methodology, Investigation, Writing – Review and Editing. **Snaedis H.**
604 **Bjornsdottir:** Methodology, Investigation, Writing – Review and Editing. **Olafur H. Fridjonsson:**
605 Conceptualization, Writing – Review and Editing. **Eva Nordberg Karlsson:** Conceptualization, Writing –
606 Review and Editing. **Justine Vanhalst:** Investigation, Writing – Review and Editing. **Birkir Reynisson:**
607 Investigation, Writing – Review and Editing. **Steinn Gudmundsson:** Conceptualization, Supervision,
608 Project administration, Funding acquisition, Writing – Review and Editing.

609 Funding

610 This work was supported by the Marine Biotechnology ERA-NET, *ThermoFactories*, project grant
611 number 5178–00003B; the Technology Development fund in Iceland, grant number 159004-0612; the
612 Icelandic Research fund, *ThermoExplore*, project grant number 207088-051 and the Novo Nordisk
613 Foundation (NNF18OC0034792).

614 Acknowledgements

615 The authors would like to thank Emanuel Y. C. Ron and Roya R. R. Sardari for the advice on how to
616 cultivate *R. marinus* in bioreactors and to extract carotenoids, respectively, and Louna Maignien for
617 assisting with the bioreactor cultivations.

618 References

- 619 [1] R. Munoz, R. Rosselló-Móra, and R. Amann, "Revised phylogeny of Bacteroidetes and proposal
620 of sixteen new taxa and two new combinations including Rhodothermaeota phyl. nov.," *Syst.*
621 *Appl. Microbiol.*, vol. 39, no. 5, pp. 281–296, 2016.
- 622 [2] G. A. Alfredsson, J. K. Kristjansson, S. Hjorleifsdottir, and K. O. Stetter, "Rhodothermus marinus,
623 gen.nov., sp. nov., a Thermophilic, Halophilic Bacterium from Submarine Hot Springs in
624 Iceland," *J. Gen. Microbiol.*, no. 134, pp. 299–306, 1988.
- 625 [3] S. Halldórsdóttir *et al.*, "Cloning, sequencing and overexpression of a Rhodothermus marinus
626 gene encoding a thermostable cellulase of glycosyl hydrolase family 12," *Appl. Microbiol.*
627 *Biotechnol.*, vol. 49, no. 3, pp. 277–284, 1998.
- 628 [4] M. Krah, R. Misselwitz, O. Politz, K. K. Thomsen, H. Welfle, and R. Borriss, "The laminarinase
629 from thermophilic eubacterium Rhodothermus marinus," *Eur. J. Biochem.*, vol. 257, no. 1, pp.
630 101–111, 1998.
- 631 [5] E. Nordberg Karlsson, E. Bartonek-Roxå, and O. Holst, "Cloning and sequence of a thermostable
632 multidomain xylanase from the bacterium Rhodothermus marinus," *Biochim. Biophys. Acta*,
633 vol. 1353, no. 2, pp. 118–124, Aug. 1997.
- 634 [6] E. Nordberg Karlsson, E. Bartonek-Roxå, and O. Holst, "Evidence for substrate binding of a
635 recombinant thermostable xylanase originating from Rhodothermus marinus," *FEMS Microbiol.*
636 *Lett.*, vol. 168, no. 1, pp. 1–7, Nov. 1998.
- 637 [7] S. H. Bjornsdottir *et al.*, "Rhodothermus marinus: physiology and molecular biology,"
638 *Extremophiles*, vol. 10, pp. 1–16, 2006.
- 639 [8] E. Y. C. Ron, R. R. R. Sardari, R. Anthony, E. W. J. van Niel, G. O. Hreggvidsson, and E. Nordberg-
640 Karlsson, "Cultivation technology development of Rhodothermus marinus DSM 16675,"
641 *Extremophiles*, vol. 23, no. 6, pp. 735–745, 2019.
- 642 [9] K. Hamana, H. Hamana, M. Niitsu, K. Samejima, and S. Matsuzaki, "Distribution of unusual long
643 and branched polyamines in thermophilic eubacteria belonging to "Rhodothermus, " Thermus
644 and Thermonema," *J. Gen. Appl. Microbiol.*, vol. 38, no. 6, pp. 575–584, 1992.
- 645 [10] R. R. R. Sardari *et al.*, "Evaluation of the production of exopolysaccharides by two strains of the
646 thermophilic bacterium Rhodothermus marinus," *Carbohydr. Polym.*, vol. 156, pp. 1–8, 2017.
- 647 [11] E. Y. C. Ron *et al.*, "Characterization of carotenoids in Rhodothermus marinus,"
648 *Microbiologyopen*, vol. 7, no. 1, 2018.
- 649 [12] O. C. Nunes, C. M. Manaia, M. S. Da Costa, and H. Santos, "Compatible Solutes in the
650 Thermophilic Bacteria Rhodothermus marinus and 'Thermus thermophilus,'" *Appl. Environ.*
651 *Microbiol.*, vol. 61, no. 6, pp. 2351–2357, Jun. 1995.
- 652 [13] L. Moreira, M. F. Nobre, I. Sa-correia, and M. S. Da Costa, "Genomic Typing and Fatty Acid
653 Composition of Rhodothermus marinus," *Syst. Appl. Microbiol.*, vol. 19, no. 1, pp. 83–90, 1996.
- 654 [14] O. C. Nunes, M. M. Donato, C. M. Manaia, and M. S. Da Costa, "The Polar Lipid and Fatty Acid
655 Composition of Rhodothermus Strains," *Syst. Appl. Microbiol.*, vol. 15, no. 1, pp. 59–62, 1992.
- 656 [15] A. López-Contreras *et al.*, "Biorefinery Approach to the Use of Macroalgae as Feedstock for
657 Biofuels," in *Algal Biofuels*, L. Pereira, Ed. CRC Press, 2017.
- 658 [16] E. J. O'Brien, J. M. Monk, and B. O. Palsson, "Using Genome-scale Models to Predict Biological

- 659 Capabilities,” *Cell*, vol. 161, no. 5, pp. 971–987, May 2015.
- 660 [17] S. H. Bjornsdottir, O. H. Fridjonsson, J. K. Kristjansson, and G. Eggertsson, “Cloning and
661 expression of heterologous genes in *Rhodothermus marinus*,” *Extremophiles*, vol. 11, no. 2, pp.
662 283–293, 2007.
- 663 [18] T. Kristjansdottir *et al.*, “Engineering the carotenoid biosynthetic pathway in *Rhodothermus*
664 *marinus* for lycopene production,” *Metab. Eng. Commun.*, p. e00140, 2020.
- 665 [19] A. Vershinin, “Biological functions of carotenoids - Diversity and evolution,” *BioFactors*, vol. 10,
666 no. 2–3, pp. 99–104, 1999.
- 667 [20] M. Nolan *et al.*, “Complete genome sequence of *Rhodothermus marinus* type strain (R-10 T),”
668 *Stand. Genomic Sci.*, vol. 1, no. 3, pp. 283–291, 2009.
- 669 [21] J. Rönkä, S. Hjörleifsdóttir, T. Tenkanen, K. Pitkänen, P. Mattila, and J. K. Kristjansson, “RmaI, a
670 type II restriction endonuclease from *Rhodothermus marinus* which recognizes 5’ CTAG 3’,”
671 *Nucleic Acids Res.*, vol. 19, no. 10, p. 2789, May 1991.
- 672 [22] S. H. Bjornsdottir, S. H. Thorbjarnardottir, and G. Eggertsson, “Establishment of a gene transfer
673 system for *Rhodothermus marinus*,” *Appl. Microbiol. Biotechnol.*, vol. 66, no. 6, pp. 675–682,
674 2005.
- 675 [23] E. Degryse, N. Glansdorff, and A. Piérard, “A comparative analysis of extreme thermophilic
676 bacteria belonging to the genus *Thermus*,” *Arch. Microbiol.*, vol. 117, no. 2, pp. 189–196, 1978.
- 677 [24] E. A. Wolin, M. J. Wolin, and R. S. Wolfe, “Formation of Methane by Bacterial Extracts,” *J. Biol.*
678 *Chem.*, vol. 238, no. 8, pp. 2882–2886, 1963.
- 679 [25] A. Blücher, E. N. Karlsson, and O. Holst, “Substrate-dependent production and some properties
680 of a thermostable, α -galactosidase from *Rhodothermus marinus*,” *Biotechnol. Lett.*, vol. 22, no.
681 8, pp. 663–669, 2000.
- 682 [26] L. Dahlberg, O. Holst, and J. K. Kristjansson, “Thermostable xylanolytic enzymes from
683 *Rhodothermus marinus* grown on xylan,” *Appl. Microbiol. Biotechnol.*, vol. 40, no. 1, pp. 63–68,
684 1993.
- 685 [27] J. H. Miller, *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, 1972.
- 686 [28] S. H. Bjornsdottir, O. H. Fridjonsson, G. O. Hreggvidsson, and G. Eggertsson, “Generation of
687 targeted deletions in the genome of *Rhodothermus marinus*,” *Appl. Environ. Microbiol.*, vol. 77,
688 no. 15, pp. 5505–5512, 2011.
- 689 [29] S. Andrews, “FastQC, Babraham Bioinformatics,” 2010. [Online]. Available:
690 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- 691 [30] A. M. Bolger, M. Lohse, and B. Usadel, “Trimmomatic: a flexible trimmer for Illumina sequence
692 data,” *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, Aug. 2014.
- 693 [31] S. Nurk *et al.*, “Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads BT -
694 Research in Computational Molecular Biology,” 2013, pp. 158–170.
- 695 [32] T. Tatusova *et al.*, “NCBI prokaryotic genome annotation pipeline,” *Nucleic Acids Res.*, vol. 44,
696 no. 14, pp. 6614–6624, Aug. 2016.
- 697 [33] I. Thiele and B. Ø. Palsson, “A protocol for generating a high-quality genome-scale metabolic
698 reconstruction,” *Nat. Protoc.*, vol. 5, no. 1, pp. 93–121, 2010.
- 699 [34] C. S. Henry, M. Dejongh, A. A. Best, P. M. Frybarger, B. Lindsay, and R. L. Stevens, “High-

- throughput generation , optimization and analysis of genome-scale metabolic models,” *Nat. Biotechnol.*, vol. 28, no. 9, pp. 969–974, 2010.
- [35] J. D. Orth *et al.*, “A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011,” *Mol. Syst. Biol.*, vol. 7, no. 535, pp. 1–9, 2011.
- [36] Z. A. King *et al.*, “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D515–D522, Oct. 2015.
- [37] L. Jeske, S. Placzek, I. Schomburg, A. Chang, and D. Schomburg, “BRENDA in 2019: a European ELIXIR core data resource,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D542–D549, Nov. 2018.
- [38] M. Kanehisa and S. Goto, “KEGG: kyoto encyclopedia of genes and genomes,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000.
- [39] R. Caspi *et al.*, “The MetaCyc database of metabolic pathways and enzymes,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D633–D639, Oct. 2017.
- [40] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.
- [41] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, “COBRApy: CONstraints-Based Reconstruction and Analysis for Python,” *BMC Syst. Biol.*, vol. 7, 2013.
- [42] D. A. Fell and J. R. Small, “Fat synthesis in adipose tissue. An examination of stoichiometric constraints,” *Biochem. J.*, vol. 238, pp. 781–786, 1986.
- [43] J. M. Savinell and B. O. Palsson, “Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism,” *J. Theor. Biol.*, vol. 154, no. 4, pp. 421–454, 1992.
- [44] C. Lieven *et al.*, “MEMOTE for standardized genome-scale metabolic model testing,” *Nat. Biotechnol.*, vol. 38, no. 3, pp. 272–276, 2020.
- [45] L. T. Cordova, R. M. Cipolla, A. Swarup, C. P. Long, and M. R. Antoniewicz, “¹³C metabolic flux analysis of three divergent extremely thermophilic bacteria: *Geobacillus* sp. LC300, *Thermus thermophilus* HB8, and *Rhodothermus marinus* DSM 4252,” *Metab. Eng.*, vol. 44, no. October, pp. 182–190, 2017.
- [46] G. O. Hreggvidsson, E. Kaiste, O. Holst, G. Eggertsson, A. Palsdottir, and J. K. Kristjansson, “An extremely thermostable cellulase from the thermophilic eubacterium *Rhodothermus marinus*,” *Appl. Environ. Microbiol.*, vol. 62, no. 8, pp. 3047–3049, 1996.
- [47] K. Z. G. Ara *et al.*, “Characterization and diversity of the complete set of GH family 3 enzymes from *Rhodothermus marinus* DSM 4253,” *Sci. Rep.*, vol. 10, no. 1, p. 1329, 2020.
- [48] G. I. Guzmán *et al.*, “Enzyme promiscuity shapes adaptation to novel growth substrates,” *Mol. Syst. Biol.*, vol. 15, no. 4, p. e8462, Apr. 2019.
- [49] L. Jolly, P. Ferrari, D. Blanot, J. Van Heijenoort, F. Fassy, and D. Mengin-Lecreux, “Reaction mechanism of phosphoglucosamine mutase from *Escherichia coli*,” *Eur. J. Biochem.*, vol. 262, no. 1, pp. 202–210, 1999.
- [50] L. Allahgholi *et al.*, “Composition analysis and minimal treatments to solubilize polysaccharides from the brown seaweed *Laminaria digitata* for microbial growth of thermophiles,” *J. Appl. Phycol.*, 2020.
- [51] Y. Qin, “Alginate fibres: an overview of the production processes and applications in wound management,” *Polym. Int.*, vol. 57, no. 2, pp. 171–180, Feb. 2008.

- 742 [52] G. O. Hreggvidsson *et al.*, “International Publication Number EP 3092247 A1.” 2016.
- 743 [53] G. O. Hreggvidsson and O. H. Fridjonsson, “International Publication Number WO 2020/044379
744 A1.” 2020.
- 745 [54] J. Preiss and G. Ashwell, “Alginic Acid Metabolism in Bacteria. The Enzymatic Reduction of 4-
746 deoxy-L-erythro-5-hexoseulose uronic acid to 2-keto-3-deoxy-D-guconic acid,” *J. Biol. Chem.*,
747 vol. 237, no. 2, pp. 317–321, 1962.
- 748 [55] R. Takase, A. Ochiai, B. Mikami, W. Hashimoto, and K. Murata, “Molecular identification of
749 unsaturated uronate reductase prerequisite for alginate metabolism in *Sphingomonas* sp. A1,”
750 *Biochim. Biophys. Acta*, vol. 1804, no. 9, pp. 1925–1936, Sep. 2010.
- 751 [56] B. F. Lutnaes, Å. Strand, S. K. Pétursdóttir, and S. Liaaen-Jensen, “Carotenoids of thermophilic
752 bacteria - *Rhodothermus marinus* from submarine Icelandic hot springs,” *Biochem. Syst. Ecol.*,
753 vol. 32, no. 5, pp. 455–468, 2004.
- 754 [57] H. Katsuki and K. Bloch, “Studies on the biosynthesis of ergosterol in yeast. Formation of
755 methylated intermediates,” *J. Biol. Chem.*, vol. 242, no. 2, pp. 222–227, Jan. 1967.
- 756 [58] F. Lynen, “Biosynthetic pathways from acetate to natural products,” *Pure Appl. Chem.*, vol. 14,
757 no. 1, pp. 137–167, 1967.
- 758 [59] M. Rohmer, M. Knani, P. Simonin, B. Sutter, and H. Sahm, “Isoprenoid biosynthesis in bacteria:
759 a novel pathway for the early steps leading to isopentenyl diphosphate,” *Biochem. J.*, vol. 295 (Pt 2,
760 no. Pt 2, pp. 517–524, Oct. 1993.
- 761 [60] S. Sauret-Güeto, E. M. Urós, E. Ibáñez, A. Boronat, and M. Rodríguez-Concepción, “A mutant
762 pyruvate dehydrogenase E1 subunit allows survival of *Escherichia coli* strains defective in 1-
763 deoxy-d-xylulose 5-phosphate synthase,” *FEBS Lett.*, vol. 580, no. 3, pp. 736–740, 2006.
- 764 [61] J. Kirby *et al.*, “Enhancing Terpene Yield from Sugars via Novel Routes to 1-Deoxy-d-Xylulose 5-
765 Phosphate,” *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 130 LP – 138, Jan. 2015.
- 766 [62] T. J. Erb *et al.*, “A RubisCO like protein links SAM metabolism with isoprenoid biosynthesis,” *Nat*
767 *Chem Biol.*, vol. 8, no. 11, pp. 926–932, 2013.
- 768 [63] M. C. Pérez-Marín, S. Padmanabhan, M. C. Polanco, F. J. Murillo, and M. Elías-Arnanz, “Vitamin
769 B12 partners the CarH repressor to downregulate a photoinducible promoter in *Myxococcus*
770 *xanthus*,” *Mol. Microbiol.*, vol. 67, no. 4, pp. 804–819, 2008.
- 771 [64] H. Takano *et al.*, “Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in
772 light-induced carotenoid production in *Thermus thermophilus*,” *J. Bacteriol.*, vol. 193, no. 10,
773 pp. 2451–2459, 2011.
- 774 [65] H. Takano, S. Obitsu, T. Beppu, and K. Ueda, “Light-induced carotenogenesis in *Streptomyces*
775 *coelicolor* A3(2): identification of an extracytoplasmic function sigma factor that directs
776 photodependent transcription of the carotenoid biosynthesis gene cluster,” *J. Bacteriol.*, vol.
777 187, no. 5, pp. 1825–1832, Mar. 2005.
- 778 [66] H. Takano *et al.*, “Role and function of LitR, an adenosyl B12-bound light-sensitive regulator of
779 *Bacillus megaterium* QM B1551, in regulation of carotenoid production,” *J. Bacteriol.*, vol. 197,
780 no. 14, pp. 2301–2315, 2015.
- 781 [67] A. S. Fernandes, M. M. Pereira, and M. Teixeira, “The succinate dehydrogenase from the
782 thermohalophilic bacterium *Rhodothermus marinus*: Redox-Bohr effect on heme bL1,” *J.*
783 *Bioenerg. Biomembr.*, vol. 33, no. 4, pp. 343–352, 2001.

- 784 [68] A. S. Fernandes, M. M. Pereira, and M. Teixeira, "Purification and characterization of the
785 complex I from the respiratory chain of *Rhodothermus marinus*," *J. Bioenerg. Biomembr.*, vol.
786 34, no. 6, pp. 413–421, 2002.
- 787 [69] M. M. Pereira, P. N. Refojo, G. O. Hreggvidsson, S. Hjorleifsdottir, and M. Teixeira, "The
788 alternative complex III from *Rhodothermus marinus* - A prototype of a new family of
789 quinol:electron acceptor oxidoreductases," *FEBS Lett.*, vol. 581, no. 25, pp. 4831–4835, 2007.
- 790 [70] B. J. Tindall, "Lipid composition of *Rhodothermus marinus*," *FEMS Microbiol. Lett.*, vol. 80, pp.
791 65–68, 1991.
- 792 [71] M. M. Pereira, J. N. Carita, R. Anglin, M. Saraste, and M. Teixeira, "Heme centers of
793 *Rhodothermus marinus* respiratory chain. Characterization of its cbb3 oxidase," *J. Bioenerg.
794 Biomembr.*, vol. 32, no. 2, pp. 143–152, Apr. 2000.
- 795 [72] M. M. Pereira *et al.*, "The *caa3* terminal oxidase of the thermohalophilic bacterium
796 *Rhodothermus marinus*: a HiPIP:oxygen oxidoreductase lacking the key glutamate of the D-
797 channel," *Biochim. Biophys. Acta - Bioenerg.*, vol. 1413, no. 1, pp. 1–13, 1999.
- 798 [73] A. F. Veríssimo, M. M. Pereira, A. M. P. Melo, G. O. Hreggvidsson, J. K. Kristjansson, and M.
799 Teixeira, "A *ba3* oxygen reductase from the thermohalophilic bacterium *Rhodothermus
800 marinus*," *FEMS Microbiol. Lett.*, vol. 269, no. 1, pp. 41–47, Apr. 2007.
- 801 [74] J. Schneider and V. F. Wendisch, "Biotechnological production of polyamines by bacteria: recent
802 achievements and future perspectives," *Appl. Microbiol. Biotechnol.*, vol. 91, no. 1, pp. 17–30,
803 Jul. 2011.
- 804 [75] W. Fukuda, R. Hidese, and S. Fujiwara, "Long-Chain and Branched Polyamines in Thermophilic
805 Microbes BT - Polyamines: A Universal Molecular Nexus for Growth, Survival, and Specialized
806 Metabolism," T. Kusano and H. Suzuki, Eds. Tokyo: Springer Japan, 2015, pp. 15–25.
- 807 [76] Z. Silva, N. Borges, L. O. Martins, R. Wait, M. S. da Costa, and H. Santos, "Combined effect of the
808 growth temperature and salinity of the medium on the accumulation of compatible solutes by
809 *Rhodothermus marinus* and *Rhodothermus obamensis*," *Extremophiles*, vol. 3, no. 2, pp. 163–
810 172, May 1999.
- 811 [77] N. Borges, J. D. Marugg, N. Empadinhas, M. S. da Costa, and H. Santos, "Specialized roles of the
812 two pathways for the synthesis of mannosylglycerate in osmoadaptation and
813 thermoadaptation of *Rhodothermus marinus*," *J. Biol. Chem.*, vol. 279, no. 11, pp. 9892–9898,
814 Mar. 2004.
- 815 [78] M. M. Nielsen *et al.*, "Substrate and metal ion promiscuity in mannosylglycerate synthase," *J.
816 Biol. Chem.*, vol. 286, no. 17, pp. 15155–15164, Apr. 2011.
- 817 [79] I. J. Mukti, R. R. R. Sardari, T. Kristjansdottir, G. O. Hreggvidsson, and E. Nordberg-Karlsson,
818 "Growth and production of carotenoids and exopolysaccharides by the extremophile
819 *Rhodothermus marinus* DSM16675 in defined media," 2021.
- 820 [80] A. M. Feist *et al.*, "A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655
821 that accounts for 1260 ORFs and thermodynamic information," *Mol. Syst. Biol.*, vol. 3, p. 121,
822 2007.
- 823 [81] M. Richter and R. Rosselló-Móra, "Shifting the genomic gold standard for the prokaryotic
824 species definition," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 19126 LP – 19131, Nov. 2009.
- 825 [82] R. L. Tatusov, M. Y. Galperin, D. A. Natale, and E. V. Koonin, "The COG database: a tool for
826 genome-scale analysis of protein functions and evolution," *Nucleic Acids Res.*, vol. 28, no. 1, pp.

827 33–36, Jan. 2000.

828 [83] M. Berney, H.-U. Weilenmann, J. Ihssen, C. Bassin, and T. Egli, “Specific growth rate determines
829 the sensitivity of *Escherichia coli* to thermal, UVA, and solar disinfection,” *Appl. Environ.*
830 *Microbiol.*, vol. 72, no. 4, pp. 2586–2593, Apr. 2006.

831 [84] A. K. Wessel *et al.*, “Oxygen limitation within a bacterial aggregate,” *MBio*, vol. 5, no. 2, pp.
832 e00992–e00992, Apr. 2014.

833 [85] U. U. Nwodo, E. Green, and A. I. Okoh, “Bacterial exopolysaccharides: functionality and
834 prospects,” *Int. J. Mol. Sci.*, vol. 13, no. 11, pp. 14002–14015, Oct. 2012.

835

836

Tables

Table 1. Main features of the genome-scale metabolic model of *R. marinus* DSM 4252^T, Rmarinus_578.

Genome features	
Genome size	3.26 Mb
Protein coding sequences	2889
Model features	
Genes	578
Metabolites (unique)	871 (784)
Reactions (with GPR)	929 (771)
Memote total score	55%

Table 2: Growth of *R. marinus* strain DSM 4252^T on different carbon sources investigated *in silico* and *in vivo*. Green indicates growth, light green indicates weak growth, ‘w’ indicates white colonies (instead of the characteristic red), orange indicates no growth *in silico* except if a transport reaction was added to the model, red indicates no growth and no color indicates no data available.

	Metabolite	In Silico	In Vivo		Metabolite	In Silico	In Vivo
Mono- and Disaccharides	Glucose ¹			Amino acids	Glutamate ¹		
	Galactose ¹				Aspartate ²		
	Sucrose ¹				Glutamine ¹		
	Lactose ¹				Asparagine ²		
	Raffinose ¹				Arginine ²		
	Maltose ¹				Leucine ²		
	Fructose ¹				Phenylalanine ²		
	Mannose ¹				Proline ²		
	Ribulose ¹				Serine ²		
	Xylose ¹				Threonine ²		
Other	Pyruvate ¹		w		Valine ²		
	Acetate ¹		w		Alanine		
	Lactate				Histidine		
	Formate				Glycine		

Polysaccharides	Starch ¹	w	Lysine	
	Xylan ¹		Tryptophan	
	Laminarin ¹		Tyrosine	
	Alginate ¹		Cysteine	
	Cellulose (CMC) ¹		Methionine	
Citric- and Urea cycles	Malate ¹		Isoleucine	
	Fumarate ²			
	Succinate ²			
	Citrate ²			
	Ornithine			
	Oxaloacetate			
	2-oxoglutarate ²			

¹Data from this study.

²Data from [2].

Table 3. The model genes that are not present in *R. marinus* strain ISCaR-493. The model gene IDs from DSM 4252^T are listed, with corresponding annotations and the gene IDs of the isozymes, when applicable.

DSM 4252 ^T gene ID	Annotation	Isozyme present
RMAR_RS02620	Dihydroneopterin triphosphate pyrophosphatase	Yes (RMAR_RS06180)
RMAR_RS14780	O-antigen polymerase (wzy)	No
RMAR_RS05305	Xylanase	No
RMAR_RS14585	Alginate lyase	No
RMAR_RS11910	Asparagine synthase (glutamine-hydrolyzing)	Yes (RMAR_RS10310)
RMAR_RS12230	Beta-ketoacyl-ACP synthase	Yes (RMAR_RS14045)
RMAR_RS12885	N-acetylglutamate synthase	Yes (RMAR_RS13420)

Figures

Figure 1. The alginate degradation pathway of *R. marinus* and its connections to the Embden-Meyerhof-Parnas (EMP) pathway, the partial Entner-Doudoroff (ED) pathway and the MEP terpenoid pathway. Names of substrates, products and energy molecules (ATP, NADH, NADPH) are shown, some abbreviated: 4-deoxy-L-erythro 5-hexoseulose uronic acid (DEH), 2-keto 3-deoxygluconate (KDG) and 2-keto 3-deoxygluconate 6-phosphate (KDPG). The molecular structure of a partial alginate molecule is shown: β-D-mannuronate (M) and α-L-guluronate (G). The missing reaction of the ED pathway in *R. marinus* is represented by a grey dotted arrow.

859

860 **Figure 2.** Functional annotations of core and accessory genes in the *R. marinus* strains ISCaR-493 and
861 DSM 4252^T. The main functional categories are four: Cellular processes and signalling (CP) red stacks;
862 Information storage and processing (IS) blue stacks; (MB) green stacks; poorly characterized (PC), grey
863 stacks: Single letters designate the COG subcategories. The core genome consists of the predicted
864 genes that are common in both strains. Each category or subcategory is graphed as a percentage of
865 the total number of genes in the core or accessory genomes.

866

867 **Figure 3.** Experimental growth data for strains DSM 4252^T and ISCaR-493 was used to validate model
868 growth predictions. The model was constrained with uptake (glucose and pyruvate) and secretion
869 (lactate and acetate) rates observed *in vivo* (Supplementary file 1) and optimized for growth. Predicted
870 growth rates were compared to growth rates observed *in vivo* (a). Experimental data from the
871 exponential growth phase of two replicates (▲ and ■), obtained early in the growth phase (time points
872 3-6 for strain DSM 4252 and 1-5 for strain ISCaR-493), was used (b).

Figure 4. Cell density (OD620nm) and carotenoid (OD480nm) production following growth of *R. marinus* strain ISCaR-493 for 24 hours on glucose; mixture of glucose and pyruvate under light and dark conditions; alginate; pyruvate and without any carbon sources. Additionally, the modified strain TK-4 ($\Delta trpB \Delta purA :: trpB dxs_{T.thermophilus}$) was grown on a mixture of glucose and pyruvate. The carotenoids were always extracted from a fixed number of cells (1 mL of OD620nm = 1). The total amount of carotenoids was calculated by multiplying the cell density by the measured carotenoids. The average cell density of each culture condition is represented by a blue bar, the average measured carotenoid value as a red bar and the average total carotenoids by an orange bar. Dots represent individual replicates.