



Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism

Mol, Viviënne; Bennett, Martyn; Sánchez, Benjamín J.; Lisowska, Beata K.; Herrgård, Markus J.; Nielsen, Alex Toftgaard; Leak, David J.; Sonnenschein, Nikolaus

Published in:
Metabolic Engineering

Link to article, DOI:
[10.1016/j.ymben.2021.03.002](https://doi.org/10.1016/j.ymben.2021.03.002)

Publication date:
2021

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):

Mol, V., Bennett, M., Sánchez, B. J., Lisowska, B. K., Herrgård, M. J., Nielsen, A. T., Leak, D. J., & Sonnenschein, N. (2021). Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism. *Metabolic Engineering*, 65, 123-134.
<https://doi.org/10.1016/j.ymben.2021.03.002>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Journal Pre-proof

Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism

Viviënne Mol, Martyn Bennett, Benjamín J. Sánchez, Beata K. Lisowska, Markus J. Herrgård, Alex Toftgaard Nielsen, David J. Leak, Nikolaus Sonnenschein

PII: S1096-7176(21)00038-0

DOI: <https://doi.org/10.1016/j.ymben.2021.03.002>

Reference: YMBEN 1774

To appear in: *Metabolic Engineering*

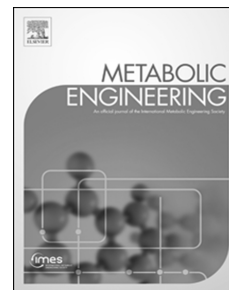
Received Date: 2 February 2021

Accepted Date: 1 March 2021

Please cite this article as: Mol, Viviënne, Bennett, M., Sánchez, Benjamín J., Lisowska, B.K., Herrgård, M.J., Nielsen, A.T., Leak, D.J., Sonnenschein, N., Genome-scale metabolic modeling of *P. thermoglucosidasius* NCIMB 11955 reveals metabolic bottlenecks in anaerobic metabolism, *Metabolic Engineering* (2021), doi: <https://doi.org/10.1016/j.ymben.2021.03.002>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier Inc. on behalf of International Metabolic Engineering Society.



1 Genome-scale metabolic modelling of *P. thermoglucosidasius*
2 NCIMB 11955 reveals metabolic bottlenecks in anaerobic
3 metabolism
4

5 Viviënne Mol^{*a}, Martyn Bennett^{*bc}, Benjamín J. Sánchez^{ad}, Beata K. Lisowska^b, Markus J.
6 Herrgård^{a,e}, Alex Toftgaard Nielsen^{a†}, David J. Leak^{#†bc}, Nikolaus Sonnenschein^{#†d}

7 * These authors contributed equally to this work.

8 #Joint senior authors

9 ^a The Novo Nordisk Foundation Center for Biosustainability, Technical University of
10 Denmark, Kongens Lyngby, Denmark

11 ^b The Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2
12 7AY, United Kingdom

13 ^c The Centre for Sustainable Chemical Technologies (CSCT), University of Bath, Claverton
14 Down, Bath BA2 7AY, United Kingdom;

15 ^d Department of Biotechnology and Biomedicine, Technical University of Denmark, Kongens
16 Lyngby, Denmark

17 ^e BioInnovation Institute, Copenhagen N, Denmark

18 † Corresponding authors: Alex Toftgaard Nielsen (atn@biosustain.dtu.dk), David J. Leak
19 (djl36@bath.ac.uk) and Nikolaus Sonnenschein (niso@dtu.dk)

20
21 **Author contributions**

22 Viviënne Mol: Methodology, validation, investigation, formal analysis, data curation, writing.
23 Martyn Bennett: Methodology, investigation, data curation, writing. Benjamín J. Sánchez:
24 Methodology, validation, data curation, editing. Beata K. Lisowska: Conceptualization,
25 methodology, investigation. Markus J. Herrgård: Conceptualization. Alex Toftgaard Nielsen:
26 Conceptualization, editing, supervision. David J. Leak: Conceptualization, editing,
27 supervision. Nikolaus Sonnenschein: Methodology, editing, supervision

28 All authors read and approved the final manuscript.

29 Declarations of interest: none

30

31 Highlights

- 32 • A validated genome-scale metabolic model for *Parageobacillus thermoglucosidasius*
- 33 NCIMB 11955 is presented.
- 34 • The model recapitulates strain engineering designs and predicts carbon utilization.
- 35 • Internal flux simulations closely match experimental values.
- 36 • The model reveals previously unknown bottlenecks in anaerobic metabolism.
- 37 • Iron(III), biotin and thiamin represent a minimal supplied nutrient set required for
- 38 anaerobic growth.

39 Abstract

40 *Parageobacillus thermoglucosidasius* represents a thermophilic, facultative anaerobic
41 bacterial chassis, with several desirable traits for metabolic engineering and industrial
42 production. To further optimize strain productivity, a systems level understanding of its
43 metabolism is needed, which can be facilitated by a genome-scale metabolic model. Here,
44 we present *p-thermo*, the most complete, curated and validated genome-scale model (to
45 date) of *Parageobacillus thermoglucosidasius* NCIMB 11955. It spans a total of 890
46 metabolites, 1175 reactions and 917 metabolic genes, forming an extensive knowledge base
47 for *P. thermoglucosidasius* NCIMB 11955 metabolism. The model accurately predicts
48 aerobic utilization of 22 carbon sources, and the predictive quality of internal fluxes was
49 validated with previously published ¹³C-fluxomics data. In an application case, *p-thermo* was
50 used to facilitate more in-depth analysis of reported metabolic engineering efforts, giving
51 additional insight into fermentative metabolism. Finally, *p-thermo* was used to resolve a
52 previously uncharacterised bottleneck in anaerobic metabolism, by identifying the minimal
53 required supplemented nutrients (thiamin, biotin and iron(III)) needed to sustain anaerobic
54 growth. This highlights the usefulness of *p-thermo* for guiding the generation of experimental
55 hypotheses and for facilitating data-driven metabolic engineering, expanding the use of *P.*
56 *thermoglucosidasius* as a high yield production platform.

57 Keywords

58 Genome-scale metabolic model, systems metabolic engineering, thermophile, *in silico* strain
59 design, anaerobic metabolism, flux balance analysis

60 1. Introduction

61 As the global transition away from petroleum-derived feedstocks continues, the need to
62 produce commodity and fine chemicals using sustainable feedstocks has accelerated the
63 interest in establishing microbial bioprocesses with lower environmental footprints¹⁻³. The
64 microbial 'chassis' organisms of these bioprocesses have been developed through modern
65 metabolic engineering strategies. Such strategies have enabled the redirection of carbon flux
66 in metabolic pathways of the corresponding microbes towards target products, in what are
67 commonly termed 'microbial cell factories'⁴.

68 Without an accurate picture of how cellular metabolism operates as a whole, metabolic
69 engineering strategies can produce flux imbalances, resulting in the accumulation of carbon
70 intermediates, metabolic bottlenecks and/or imbalances in the overall cellular redox ratio^{5,6}.
71 As a result, there can be large upfront costs in microbial strain engineering to ensure
72 economically viable biochemical product yields^{3,7}. To bolster traditional metabolic
73 engineering efforts and help elucidate genotype-phenotype relationships, systems metabolic
74 engineering aims to describe a more holistic representation of cellular metabolism through
75 the integration of stoichiometric modelling and -omics data analyses⁸.

76 In particular, the advent of cheaper DNA sequencing has given rise to genome-scale
77 metabolic models (GEMs), *in silico* reconstructions of the metabolic reaction networks of a
78 given organism, derived from its annotated genome sequence⁹. In addition to operating as a
79 knowledge base of metabolic information for a particular organism, GEMs can be used via
80 constraint-based flux balance analysis to simulate carbon flux through metabolic reaction
81 networks, enabling the rapid screening of metabolic behaviours under a range of
82 environmental variables and biological contexts¹⁰. Through comprehensive *in silico*
83 predictions of metabolic phenotypes under target conditions, GEMs can also identify
84 potential cellular redox imbalances¹¹ and metabolic bottlenecks and generate hypotheses for
85 rational, targeted genetic modifications for improved performance⁸. GEMs can even guide

86 the construction and optimization of carbon flux for either endogenous or novel heterologous
87 microbial strain pathways towards high yields of desired products^{12,13}.

88 *Parageobacillus thermoglucosidasius* NCIMB 11955 represents a Gram-positive, facultative
89 anaerobic, thermophilic bacterial chassis with several advantageous traits for industrial
90 bioprocesses when compared to many model bacterial chassis such as *Escherichia coli* and
91 *Bacillus subtilis*^{14,15}. Firstly, the thermophilicity of *Parageobacillus* spp. enables fermentations
92 between 48-70°C^{16,17} at growth rates surpassing other thermophilic organisms¹⁸ and
93 comparable to that of *E. coli*¹⁹. Compared to equivalent mesophilic fermentations, these
94 process temperatures enable a reduction in both the cooling costs of large-scale exothermic
95 fermentations, and a reduction in the risk of contamination from mesophilic microbes^{8,20}.
96 Furthermore, for industrial bioprocesses aiming for simultaneous saccharification and
97 fermentation (SSF), the thermophilicity of *P. thermoglucosidasius* is complemented by a
98 catabolic versatility. Through extracellular secretions of thermostable amylases¹⁴,
99 xylanases²¹⁻²³ and other hemicellulases²⁴⁻²⁶, *Parageobacillus* spp. are able to metabolise a
100 wide range of C₅ and C₆ sugar monomers. Notably, they are able to transport then
101 metabolize complex hemicellulosic²⁶ and cellulosic¹⁴ polysaccharides derived from
102 hydrolysates of lignocellulosic biomass, potentially reducing the reliance on externally
103 supplied hydrolases involved in lignocellulosic pre-treatment.

104 A number of synthetic biology tools applicable to *P. thermoglucosidasius* have been devised
105 including: shuttle vectors for reliable transformation^{21,27}, chromosomal integration strategies²⁸
106 promoter and RBS libraries to enable tuneable gene expression and validated reporter
107 genes²⁹⁻³². Such tools have enabled *P. thermoglucosidasius*, and genetically similar
108 (*Para*)*geobacillus* spp., to be used in the production of fuels such as bioethanol^{33,34},
109 isobutanol³⁵ and hydrogen gas^{36,37} and also in fine chemicals including 2-3 butanediol^{38,39},
110 riboflavin⁴⁰ and isoprenoids⁴¹. *Parageobacillus* spp. and *Geobacillus* spp. have also been
111 the source of thermostable variants of industrially useful proteases⁴², carboxyl esterases⁴³⁻⁴⁵,

112 lipases^{44,46} along with a thermostable DNA polymerase I from *G. stearothermophilus*
113 GIM1.543⁴⁷.

114 In spite of these advances, (with the exception of natural end-products of glycolytic
115 metabolism, such as ethanol) none of these engineered pathways have approached their
116 potential maximum yields. In general, they have relied on natural flux to their metabolic
117 precursors and its inherent control. The availability of a reliable GEM would enable a
118 systems metabolic engineering approach of *P. thermoglucosidasius* to address the
119 optimisation of flux through central metabolic pathways to balance the requirements of both
120 production and growth. At present, only one publicly available GEM of a *P.*
121 *thermoglucosidasius* exists, the related strain *P. thermoglucosidasius* C56-YS93 (denoted
122 iGT736)⁴⁸. While comprising 1159 reactions and 1163 metabolites, analysis of iGT736 using
123 the GEM assessment tool Memote developed by Lieven et al.⁴⁹ suggests that it currently
124 lacks some fundamental features, including a biomass equation, transport reactions and
125 stoichiometric balance (Supplementary File 1), preventing meaningful application for
126 quantitative analysis. Additionally, a few examples exist of smaller central carbon
127 metabolism scale models derived from experimental ¹³C isotopic tracer experiments. This
128 includes models representing *P. thermoglucosidasius* M10EXG⁷ under aerobic and
129 anaerobic growth conditions, and similar *Geobacillus* spp. *G. icigianus*³⁸ and *Geobacillus*
130 LC300⁵⁰. However, they are less useful for illustrating the scale and complexity of whole cell
131 metabolism.

132 The newly constructed genome-scale metabolic model of *P. thermoglucosidasius* NCIMB
133 11955 presented herein (named hereafter as *p-thermo*) represents 917 genes and
134 comprises of 890 metabolites and 1175 reactions across two compartments: cytosolic and
135 extracellular space (representing the medium). After iterative cycles of manual curation,
136 model refinement and analysis with Memote⁴⁹, *p-thermo* exhibits a 100% stoichiometric
137 consistency, 100% charge balance and a 99.9% mass balance. It accurately captures
138 experimentally determined utilization of 22 carbon sources using the sole input of measured

139 production and consumption rates⁵¹ and is represented in the Systems Biology Markup
140 Language (SBML)⁵² compliant format, making it compatible with commonly used constraint-
141 based modeling software such as COBRApy⁵³ and the COBRA Toolbox v3.0⁵⁴ as well as
142 more specialised software facilitating systems metabolic engineering^{55,56}. Validation of the
143 predictive quality of *p-thermo* under aerobic, oxygen limited and anaerobic conditions is
144 demonstrated through mapping the resulting *in silico* fluxes to experimentally determined
145 ¹³C-flux data obtained from ¹³C-isotopic labelling experiments of the genetically and
146 metabolically similar *P. thermoglucosidasius* M10EXG strain⁷. The predictive power of *p-*
147 *thermo* is further demonstrated through recapitulation of a metabolically engineered
148 homoethanogenic strain of *P. thermoglucosidasius*³³. Lastly, *p-thermo* was used to
149 investigate the fundamental requirements and metabolic bottlenecks of *P.*
150 *thermoglucosidasius* during anaerobic growth.

151 Currently, *p-thermo* represents the most complete, curated and experimentally validated
152 genome-scale metabolic model for a *Parageobacillus* sp, and will be a foundational platform
153 for guiding rational metabolic engineering strategies, -omic data integration, and strain
154 optimization to further the potential of *P. thermoglucosidasius* NCIMB 11955 to operate as
155 microbial chassis for sustainable bioprocesses.

156

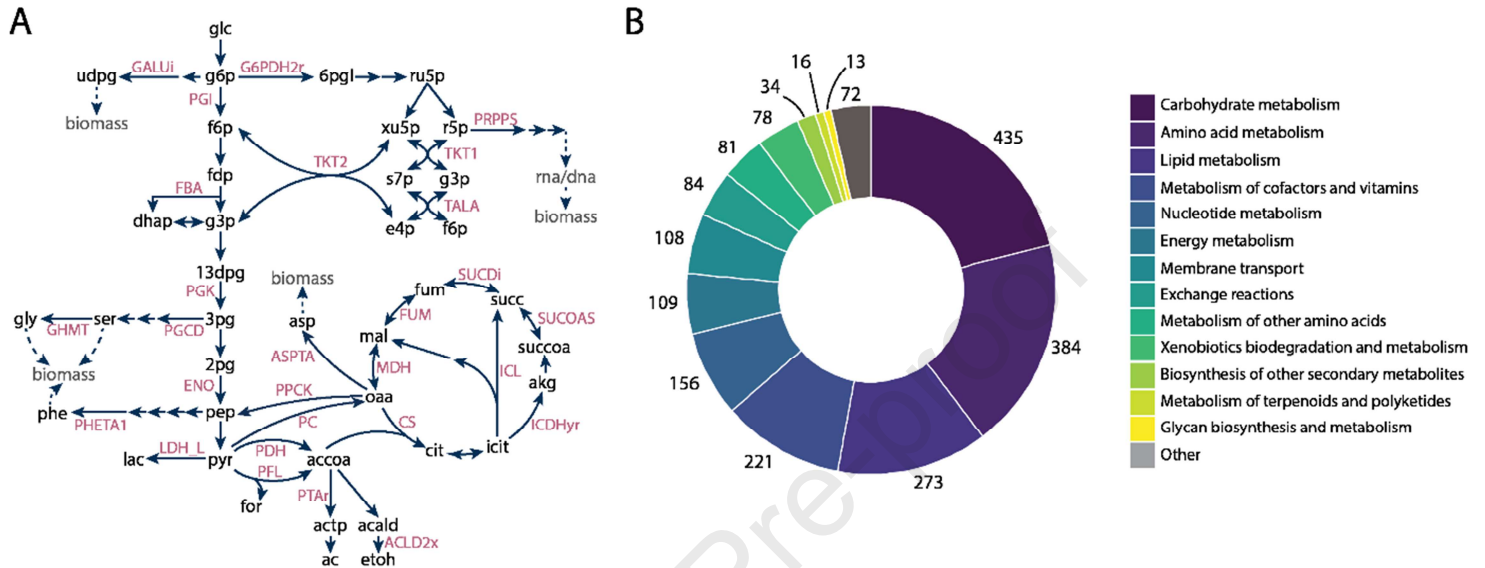
2. Results

2.1 Model reconstruction

The presented genome-scale metabolic reconstruction of *P. thermoglucosidasius* NCIMB 11955 is based on genome sequencing by ERGO™ Integrated Genomics⁵⁷ and Sheng *et al.*⁵⁸. Genome annotation was performed through the ERGO™ Integrated Genomics suite⁵⁷ and the RAST annotation server⁵⁹, followed by gap filling with Pathway Booster⁶⁰. The reconstruction was extensively manually curated using available literature and databases (KEGG, BRENDA, MetaCyc, MetaNetX and EC2PDB), according to benchmark approaches⁶¹. Detailed manual curation and refinement can be followed in Lisowska¹⁴ and in the GitHub repository. Specific attention was given to ensure nucleotide, amino acid and co-factor metabolism was accurate based on the current available knowledge. Additionally, extra care was paid to the selection and stoichiometry of the biomass precursors and their respective biosynthetic pathways, with particular focus on a more accurate representation of fatty acid and lipid biosynthesis⁶².

This metabolic model consists of 890 metabolites, involved in a total of 1175 reactions, encoded for by 917 genes, across two compartments: cytosolic and extracellular space (representing the medium). Manual curation was critical to ensure complete consistency of the model (Supplementary report 1). Central carbon metabolism of the model resembles that of previously reported *(Para)geobacillus* spp^{7,63} (Figure 1A). Of all reactions, 9.3% are involved in transport or exchange, highlighting the flexibility of the strain to grow on various carbon sources (Figure 1B). Predominantly manual gap filling, based on available literature, was used to ensure correct active metabolic pathways, accounting for 20.6% of the final *p-thermo* reactions. After gap filling, 17 dead end and 22 orphan metabolites remained, resulting in 148 universally blocked reactions (12.6% of total reactions). All dead end and orphan metabolites were manually reviewed, but a lack of conclusive evidence surrounding them with respect to *Parageobacillus* sp. prevented these gaps from being filled with sufficient accuracy and so have been deliberately unmodified until further knowledge is accrued.

185 The model as well as scripts used in the reconstruction and manual curation are made
 186 publicly available through Github, at <https://github.com/biosustain/p-thermo/releases/v1.0>.
 187 The model is stored using the community-standard SMBL format (Level 3, FBC Version 2)⁶⁴
 188 and can additionally be accessed as Supplementary File 2.



189
 190 Figure 1: A) Central carbon metabolism map, with several reaction IDs highlighted. For a more
 191 detailed overview, see Supplementary File 3. B) The number of reactions in the model for several
 192 reaction class types. See section 5 for abbreviations used.

193

194 2.2 Biomass composition and growth energetics

195 To capture biological growth in stoichiometric models, a demand reaction referred to as a
 196 biomass pseudo-reaction, was added. An overview of how the biomass pseudo-reaction was
 197 defined is explained in Materials & Methods, with the final reaction components and
 198 associated stoichiometry given in Supplementary Table 1. Energetic parameters were fitted
 199 from aerobically grown chemostat experiments. The energy required to maintain cellular
 200 homeostasis is reflected in the non-growth associated maintenance (NGAM) and was found
 201 to be $3.141 \text{ mmol}_{\text{ATP}}/\text{g}_{\text{DW}}\text{h}^{-1}$ in *p-thermo*. The growth associated maintenance, (GAM), was
 202 estimated as $152.3 \text{ mmol}_{\text{ATP}}/\text{g}_{\text{DW}}$ and reflects the energy needed for cell replication, including
 203 macromolecule synthesis. The contribution of polymerization energy, required for

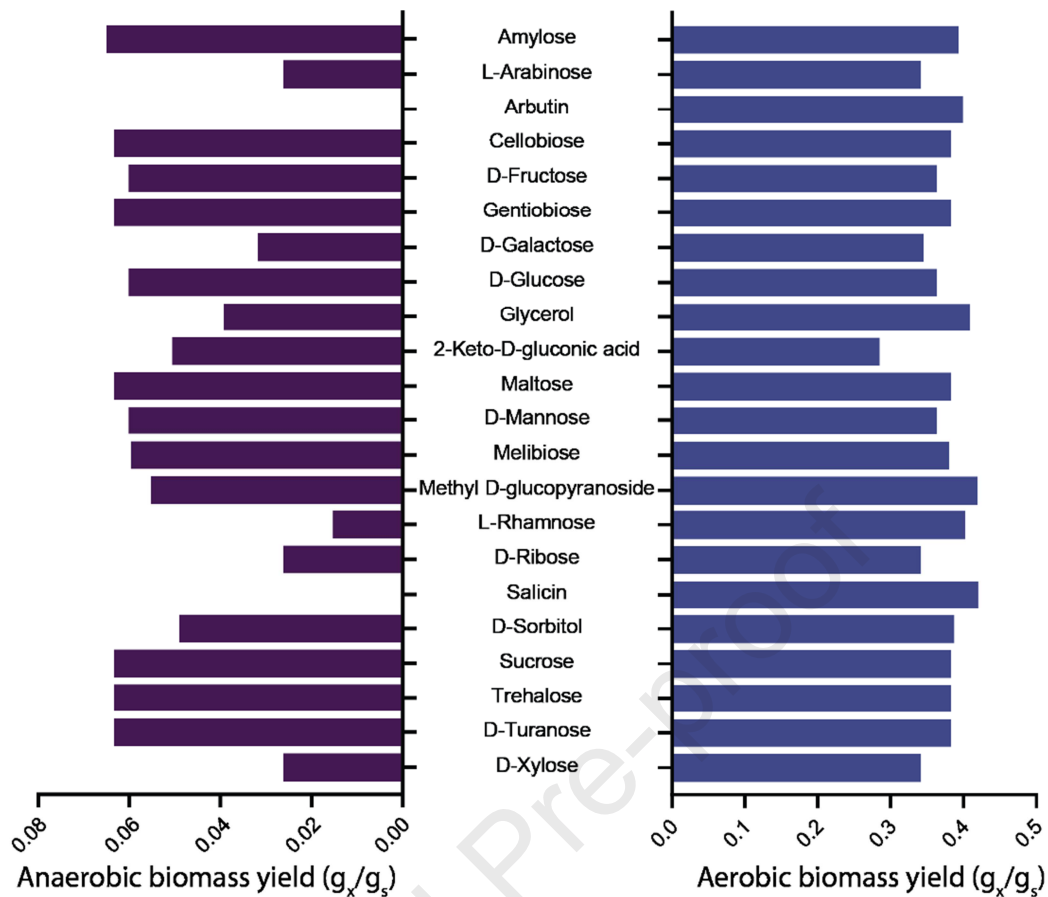
204 macromolecule synthesis, to the obtained GAM was estimated to be approximately 20%
205 (Supplementary table 2); relatively low compared to previously reported mesophiles (30-
206 40%)⁶⁵⁻⁶⁸. It was previously observed that thermophilic organisms tend to require higher
207 levels of energy for growth and homeostasis at elevated temperatures and thus have a
208 reduced growth efficiency, shown in the high maintenance estimated^{18,69}. This trait of
209 thermophiles makes them valuable hosts for bioproduction as it leads to higher production
210 rates of catabolic products compared to other organisms.

211 **2.3 Overview of metabolism**

212 To provide a comprehensive overview, two pathway maps of the model were drawn using
213 Escher⁷⁰ corresponding to central carbon and amino acid metabolism (Supplementary Files
214 3 and 4) and deposited in the GitHub repository at p-thermo/maps. Traits specific to
215 *Geobacillus* spp. and *P. thermoglucosidasius* NCIMB 11955, presented in the literature,
216 were used to validate the model's metabolism. Detailed step-by-step decisions that were
217 made, can be followed in the GitHub repository at "p-thermo/notebooks". As an example, in
218 central carbon metabolism research has shown that *Geobacillus* spp., unlike many
219 mesophilic *Bacillus* species, lack genes for a 6-phosphogluconolactonase (6PGL),
220 responsible for part of the oxidative pentose phosphate pathway (PPP)¹⁴. Instead, the
221 reaction can occur spontaneously and, at thermophilic temperatures, has been hypothesized
222 to be sufficiently rapid to maintain the requisite PPP flux⁷¹. The absence of 6PGL was
223 captured in the model, but to reflect the active PPP pathway, a pseudo-reaction was added
224 to allow the complete oxidative PPP to function.

225 (*Para*)*geobacillus* species are known to be capable of growth on a wide range of
226 carbohydrates, and have been shown to secrete various polysaccharide degrading enzymes
227 such as xylanases and other hemicellulose degrading enzymes^{14,24-26}. To assess the
228 metabolic capacity of the model, growth on various carbon sources was simulated (Figure
229 2). The choice of carbon sources was made based on previous qualitative growth
230 experiments which demonstrate a range of sole carbon sources which allow aerobic growth

231 of *P. thermoglucosidasius* NCIMB 11955⁵¹. Additionally, anaerobic growth on these
232 substrates was computationally predicted. In both cases, carbon supply was normalized to
233 30 Cmol_s/g_{DW}h, to accommodate different polymeric substrate forms being present in the
234 data set. Initially, the model showed no aerobic growth on arbutin, salicin and rhamnose, due
235 to dead-end metabolites being formed as side products in the first steps of their break down.
236 Available literature was used to fill the gaps in the catabolic pathways, which enabled
237 aerobic growth on all three carbon sources. Anaerobically, *in silico* growth on arbutin and
238 salicin was unfeasible, as current knowledge suggests that their catabolism is oxygen
239 dependent. Both arbutin and salicin are non-conventional carbon sources and are
240 glycosides, composed of either a hydroquinone or salicyl alcohol functional group attached
241 to glucose, respectively. It is known that metabolism of these glycosides occurs through
242 splitting of the glycosidic bond, with the two functional groups being catabolized individually.
243 With currently available knowledge, the further breakdown of the salicyl alcohol and
244 hydroquinone functional groups is dependent on oxygen, deeming the *in silico* prediction of
245 anaerobic growth unfeasible. As there is little knowledge about microbial catabolism of these
246 carbon substrates, this hypothesis would warrant experimental validation.



247

248

249

250

251

Figure 2: Anaerobic (left) and aerobic (right) predicted biomass yields for 22 different carbon sources, for which aerobic growth has been experimentally confirmed⁵¹. Carbon substrates were all supplied in the model at 30 Cmol/g_{DWh} to account for differences in composition between the carbon sources.

252

2.4 Assessment of predictive power through ¹³C-flux fitting

253

254

255

256

257

258

259

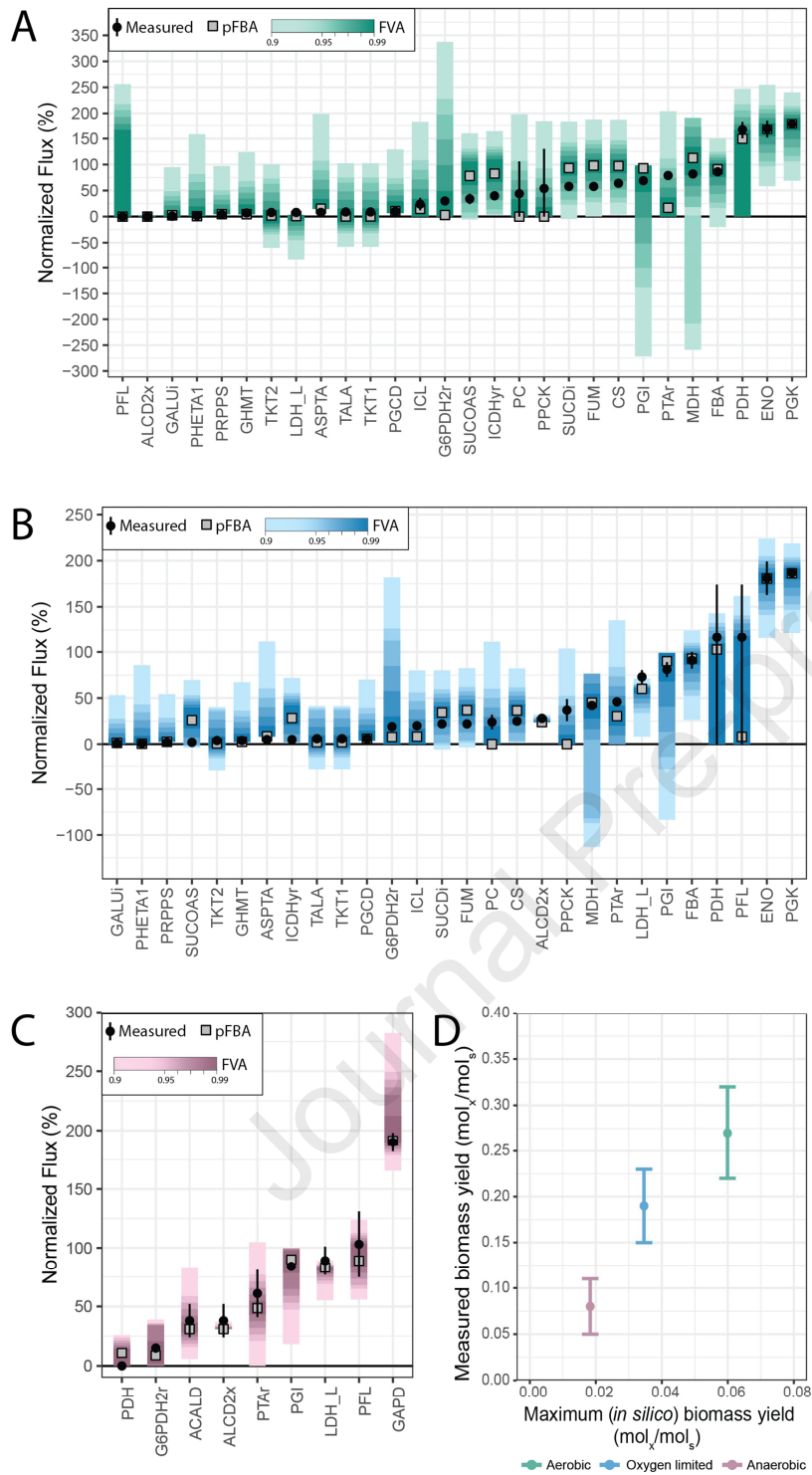
Prior to using a genome-scale model for metabolic analyses or *ab initio* predictions, it is critical to validate its predictive power based on previously attained experimental data. This was done by analysis of how well the simulated fluxes of *p-thermo* match published flux distributions of *P. thermoglucosidasius*. ¹³C-isotopic labelling is a standard tool used to elucidate intracellular fluxes in central carbon metabolism, through extensive experimental work and data analysis. Flux variability analysis (FVA) is an *in silico* approach that can allow *ab initio* analysis of metabolism without the need for laborious experimental data⁷².

260 Comparing the two data types can give insights into metabolism and allow the generation of
261 hypotheses for metabolic engineering purposes.

262 To make this comparison, ^{13}C -flux data from *P. thermoglucosidasius* M10EXG subject to
263 varying oxygen conditions was used to qualitatively assess the predictive quality of *p-*
264 *thermo*⁷. Whole proteome analysis (on a sequence basis) of the *P. thermoglucosidaius*
265 M10EXG and NCIMB 11955 strains shows that the ORFs between the two strains are highly
266 similar (Supplementary figure 1, Supplementary table 3). Importantly, considering the
267 metabolic genes that would be captured as reactions in a metabolic model, there are only 11
268 and 12 unique reactions in *P. thermoglucosidasius* NCIMB 11955 and *P.*
269 *thermoglucosidasius* M10EXG respectively (Supplementary tables 3, 4 and 5). Therefore,
270 based on the overall metabolic similarity between the two strains, we assume that the ^{13}C -
271 flux data from *P. thermoglucosidasius* M10EXG can be utilized for a qualitative assessment
272 of *p-thermo*.

273 In order to test if *p-thermo* can predict intracellular fluxes close to the ^{13}C -flux data, the
274 measured production and consumption rates were fixed in the model as exchange rates, and
275 internal fluxes were predicted in a sensitivity analysis with FVA. Parsimonious enzyme
276 usage flux balance analysis (pFBA), which has previously been shown to predict fluxes that
277 correlate with experimental measurements, was also performed⁷³. The *in silico* fluxes and
278 pFBA results were mapped to experimentally determined fluxes in aerobic, oxygen limited
279 and anaerobic conditions (Figure 3A, B and C respectively). pFBA showed good correlation
280 to the measured data for each condition (Supplementary Figure 2), and together with FVA
281 showed accurate predictions of the internal central carbon fluxes (Figure 3A, B, and C).
282 Biomass yields (Figure 3D) and oxygen consumption rates (Supplementary Figure 3) were
283 adequately predicted as well. These analyses validate the predictive quality of the created
284 model and highlight the power of using metabolic models for understanding intracellular
285 fluxes when only extracellular consumption or production rates are available.

286



287

288 Figure 3: Results of fixing experimentally measured exchange rates and predicting intracellular flux
 289 distributions⁷ in aerobic (A), oxygen limited (B) and anaerobic (C) conditions, normalized to the
 290 glucose uptake rate. Figure 1A shows the stoichiometry of all the reactions shown on the x-axis. See
 291 Section 5 for abbreviations. FVA sensitivity analysis is shown in line ranges. Predicted and measured
 292 maximum biomass yields, for a FVA threshold set at 99% of optimum biomass, are shown in (D).

293 2.5 Recapitulating & interpreting knockout physiology

294 To evaluate the utility of *p-thermo* for metabolic engineering applications, we recreated
295 previously reported homoethanogenic mutants of *P. thermoglucosidasius* NCIMB 11955 *in*
296 *silico*. Cripps *et al*⁶³ engineered lactate dehydrogenase (*ldh*) and pyruvate formate lyase (*pfl*)
297 knockouts *in vivo*, and supplemented their ethanol yields with an upregulation of pyruvate
298 dehydrogenase expression (PDH_{up}). Using *p-thermo*, the wild type (WT), Δldh and
299 $\Delta ldh\Delta pfl$ (PDH_{up}) strains were recreated, as stoichiometric modeling cannot distinguish
300 between upregulated expression levels (i.e. between $\Delta ldh\Delta pfl$ and $\Delta ldh\Delta pfl$ PDH_{up}).
301 Exchange rates of the main fermentation metabolites were predicted using *p-thermo* and
302 their accuracy evaluated based on measured data (Figure 4). In performing the analysis, two
303 distinct thresholds for Flux Variability Analysis (FVA) were selected, 95% and 99% of
304 optimum biomass production⁷², to assess the flexibility of exchange rates to the simulated
305 conditions.

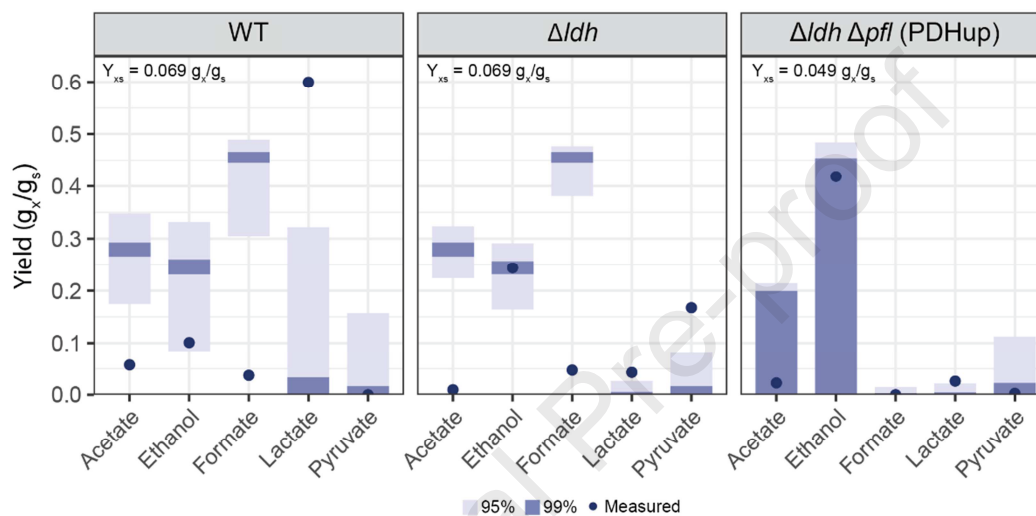
306 The performed simulations show a substantial discrepancy between predicted and
307 measured yields in the WT and Δldh strains, whereas simulations tightly match the
308 measured yields in the $\Delta ldh\Delta pfl$ PDH_{up} strain. Still, the mismatch between the experimental
309 and *in silico* data, *p-thermo* can be used to understand metabolic branch points. The main
310 discrepancy observed lies in the lactate and formate yields for the WT and Δldh strains,
311 which can be traced to the cellular decision of what to do with the synthesized pyruvate
312 (Figure 4). In this regard, there are three options: 1) conversion into lactate by lactate
313 dehydrogenase (LDH), 2) anaerobic conversion into acetyl-CoA by pyruvate formate lyase
314 (PFL) or 3) aerobic conversion into acetyl-CoA by pyruvate dehydrogenase (PDH). In both
315 the WT and Δldh strain, *p-thermo* showed flux from pyruvate to acetyl-CoA to be exclusively
316 carried through PFL, fitting with experimental expectations under anaerobic conditions due
317 to high [NADH]⁷⁴. Additionally, the conversion of pyruvate into acetate and ethanol results in
318 one additional ATP per glucose, compared to converting pyruvate into lactate⁷⁵. Therefore,
319 from a stoichiometric perspective, *p-thermo* predicts this to be the most optimal pathway for

320 growth, explaining the high concentrations of formate, ethanol and acetate predicted in the
321 simulation.

322 However, this was not observed in the experimental yields, presumably because of subtle
323 differences in dissolved oxygen availability in the experimental setup that influence multiple
324 levels of regulation *in vivo*, intrinsically not accurately captured by stoichiometric models. In
325 the experimental dataset, undefined oxygen limited conditions were used in which a gradual
326 decline in available dissolved oxygen concentration would have occurred during growth,
327 whereas simulations were performed under anaerobic conditions. Under oxygen-limited
328 conditions, PDH is expressed in the wild type *P. thermoglucosidasius*³³, where PFL is
329 typically only active under completely anaerobic conditions⁷⁶. The transition of physiological
330 states in response to decreasing oxygen availability results in excess NADH and creates a
331 redox imbalance in the cell which is alleviated through production of lactate as the
332 production of formate by PFL is restricted. This could explain the discrepancy between the
333 experimentally measured low formate and high lactate production in the WT strain and the
334 prediction by *p-thermo*. In the LDH knockout at low dissolved oxygen conditions, which
335 prevents PFL activity, PDH instead predominantly carries flux to acetyl-CoA. In this instance,
336 in order to maintain cellular redox balance, the Δldh cells increase the produced
337 ethanol/acetate ratio. This picture highlights the complexity of cellular and enzymatic
338 regulation that is poorly captured in stoichiometric models, as well as the difficulty in
339 simulating uncontrolled environments accurately.

340 However, the performed simulations can still be used to visualize and understand the burden
341 that lactate production can have on cellular growth. The inability to induce PFL at moderate
342 levels of oxygen limitation, puts a larger reliance on fermentative metabolism to lactate,
343 providing less energy. We used *p-thermo* to investigate the possible impact this has. First, all
344 measured exchange rates for the three strains were fitted to the model and used in
345 subsequent determination of predicted biomass yields. This showed that the model is
346 physiologically capable of capturing the measured data, albeit with a lower predicted

347 biomass yield than was experimentally measured, suggesting that stoichiometrically sub-
 348 optimal fermentation pathways were active *in vivo* (Supplementary Figure 4A). Finally, the
 349 effect of increasing lactate production on biomass yield was computed, showing the
 350 energetic loss that occurs from lactate production (Supplementary Figure 4B). Overall, this
 351 highlights the importance of complex regulation in dictating *in vivo* metabolism, over pure
 352 stoichiometric optima per se.



353
 354 Figure 4: Comparison of *in silico* predictions of fermentation product yields in three engineered strains
 355 with experimentally determined data from Cripps *et al.*³³, when solely the carbon uptake rate and
 356 knockouts were fixed in the model. Yield (g_x/g_s) is shown for predicted and measured exchange rates.
 357 Each panel highlights a different strain: wild type (WT), Δldh and $\Delta ldh \Delta pfl$ (PDHup). Two varying
 358 thresholds for FVA were run: 95% and 99% of the optimum biomass production. The *in silico*
 359 predicted biomass yield (Y_{xs}) for 99% of the optimum biomass production is shown for each condition.

360 2.6 Genome-scale metabolic modeling allows the elucidation of metabolic bottlenecks

361 The availability of a comprehensive GEM can also facilitate the elucidation of metabolic
 362 bottlenecks and identification and optimization of chemically defined growth media⁷⁷. Thus,
 363 *p-thermo* was used to help resolve known issues of the anaerobic metabolic physiology of *P.*
 364 *thermoglucosidasius*. Although *P. thermoglucosidasius* is clearly capable of classical mixed
 365 acid fermentation and shows elements of a regulated aerobic-anaerobic switch as revealed
 366 by transcriptomic analysis⁷⁸ (although the aerobic respiratory electron transport chain

367 remained active in an oxygen-scavenging state under fermentative conditions), it has long
368 been known that growth under anaerobic conditions on existing minimal defined growth
369 medias requires additional supplements in comparison to growth under aerobic conditions.
370 Typically, this was resolved by supplementation with a small amount of oxygen or yeast
371 extract^{14,33,51}. Therefore, here we used simulations of *p-thermo* to find a minimal set of
372 defined nutrients that can achieve anaerobic growth of *P. thermoglucosidasius*. As a first
373 observation, when fed true minimal, anaerobic medium, the model predicted no growth, in
374 accordance with experimental observations. However, fermentative energy generation was
375 observed, which highlights that oxygen requirement comes from critical secondary
376 metabolites or cofactors that cannot be synthesized anaerobically, which is corroborated by
377 previous observations¹⁴. By minimizing the oxygen uptake in the model, a critical reaction set
378 requiring oxygen was generated (Table 1). This analysis highlighted a complex combination
379 of components that cannot be synthesized anaerobically: thiamine, biotin, folate, vitamin
380 B12, spermine, spermidine and hemin. Additionally, iron(III) must be available in the medium
381 to allow porphyrin biosynthesis.

382 As expected, *in silico* supplementation of these components rescued anaerobic growth,
383 providing a combination of candidates for experimental validation. *P. thermoglucosidasius*
384 DSM 2542 was obtained from the DSM stock center⁷⁹. While DSM 2542 carries several
385 mutations (11 SNPs and 2 indels) compared to the NCIMB 11955 sequence⁵⁸, the only
386 differences found in coding sequences are annotated as hypothetical proteins. Thus, it is
387 expected that they would not impact the predictive outcome of the model between the two
388 strains. The simulated essential components were experimentally added together in trace
389 amounts to form a supplementation mix (see Materials and Methods); to assess if it would
390 allow anaerobic growth. This was compared to Wolfe's vitamin solution, a commonly used
391 mix of vitamins in base thermophilic minimal medium (TMM)^{31,80}. It should be noted that
392 Wolfe's vitamin solution contains thiamin, biotin, folate and vitamin B12, amongst other
393 nutrients, and that TMM contains trace amounts of iron(III). To uncover the minimal sets of

394 components needed to rescue anaerobic growth, eight different conditions were tested, all
 395 composed of base TMM with 10 g/l glucose: 1) no added nutrients, 2) 0.2% yeast extract, 3)
 396 biotin, 4) thiamin, 5) biotin and thiamin, 6) Wolfe's vitamins, 7) Supplementation mix and 8)
 397 Wolfe's vitamins plus the unique components of the supplementation mix (spermine,
 398 spermidine and heme) (Figure 5, Supplementary figure 5).

399 **Table 1:** Overview of critical reactions that require oxygen to allow growth in p-thermo.

Reaction ID	EC-code	Reaction	<i>In silico</i> requirement
CAT	1.11.1.6 & 1.11.1.21	Catalase-peroxidase	Glyoxylate and dicarboxylate metabolism
PMES	1.14.14.46	Pimeloyl-[acyl-carrier protein] synthase	Biotin biosynthesis
BLUB	1.13.11.79	5,6-dimethylbenzimidazole synthase	Vitamin B12 biosynthesis
GLYCTO1	1.1.3.15	(S)-2-hydroxy-acid oxidase	Folate biosynthesis
PPPGO_1	1.3.3.4	Protoporphyrinogen oxidase	Porphyrin (heme) biosynthesis
FERO	1.16.3.1	Ferroxidase	Porphyrin (Fe ³⁺) metabolism
ACDO	1.13.11.54	Acireductone dioxygenase	Aliphatic polyamine biosynthesis
GLYHOR	1.4.3.19	Glycine oxidase	Thiamine biosynthesis

400

401 Experimental observations suggested that a combination of thiamin, biotin and iron(III) were
 402 the minimal required supplementation set needed to sustain anaerobic growth, as no
 403 difference was observed when additional defined supplementation was added (Figure 5).
 404 Yeast extract also contains significant amounts of amino acids and other components and so
 405 provides an additional growth advantage, as expected. However, this highlights a
 406 discrepancy with the model predictions, as a larger minimal supplementation set was
 407 originally predicted (Table 1). Finally, as expected, base TMM can support aerobic growth, at
 408 a maximum rate of $0.267 \pm 0.021 \text{ h}^{-1}$ (Supplementary figure 5A), confirming the synthesis of
 409 the critical components in the presence of oxygen.

410 There are several reasons that can explain the differences between the *in silico* and
 411 experimentally determined minimal supplementation set. The incomplete understanding of
 412 thermophilic life introduces additional levels of complexity that are typically not captured by
 413 automatic annotation pipelines dependent on predominantly mesophilic datasets^{59,81} leading
 414 to errors in the annotation of thermophilic traits⁸². For example, genomes of thermophilic

415 organisms show a correlation with higher G/C content, less intergenic regions and a higher
416 functional stability (reflected by the lower ratio of non-synonymous to synonymous
417 substitutions over time)⁸³⁻⁸⁵. Additionally, thermostable proteins can have significantly altered
418 structure compared to their mesophilic counterparts performing the same reaction,
419 confounding homology-based annotation⁸⁶. Through the observed discrepancies, we can
420 unveil additional insights of anaerobic metabolism of *P. thermoglucosidasius*.

421 In the first place, the *in silico* dependence on vitamin B12 highlights the inaccuracy of
422 annotation pipelines. Vitamin B12 synthesis is classically divided into two routes: canonical
423 (aerobic) and non-canonical (anaerobic)⁸⁷. Although the genome annotation of *P.*
424 *thermoglucosidasius* NCIMB 11955 reveals parts of either pathway, neither is complete. It
425 has been proposed that possibly a novel, blended pathway may be present; however, this
426 may arise from incorrect annotations based on lacking knowledge of thermophilic vitamin
427 B12 biosynthesis genes^{51,88,89}. The possibility to grow without vitamin B12 supplementation
428 does highlight both an aerobic and anaerobic functional pathway in *P. thermoglucosidasius*
429 NCIMB 11955. To further understand the *de novo* biosynthesis of vitamin B12, experimental
430 validation would be required.

431 In *p-thermo*, the *in silico* oxygen requirement for spermine and spermidine biosynthesis
432 comes from the downstream recycling of a biosynthetic by-product: 5'-methylthioadenosine
433 (5-MTA). 5-MTA recycling is also important in a novel, oxygen independent MTA-isoprenoid
434 shunt, involved in the methionine salvage pathway⁹⁰. This pathway has been characterized
435 in *Rhodospirillum rubrum* and orthology analysis highlights the possible presence of parts of
436 this pathway in various facultative anaerobic *Bacillus* spp⁹¹. This presents the possibility of
437 an alternate 5-MTA recycling pathway, explaining the independence of anaerobic growth to
438 spermine or spermidine addition.

439 Similarly to spermine and spermidine, the oxygen requirement in the *in silico* folate
440 biosynthesis pathway stems from the formation of glycolaldehyde as a side product, which is
441 further oxidized to glyoxylate. The *in silico* oxidase responsible for glyoxylate formation

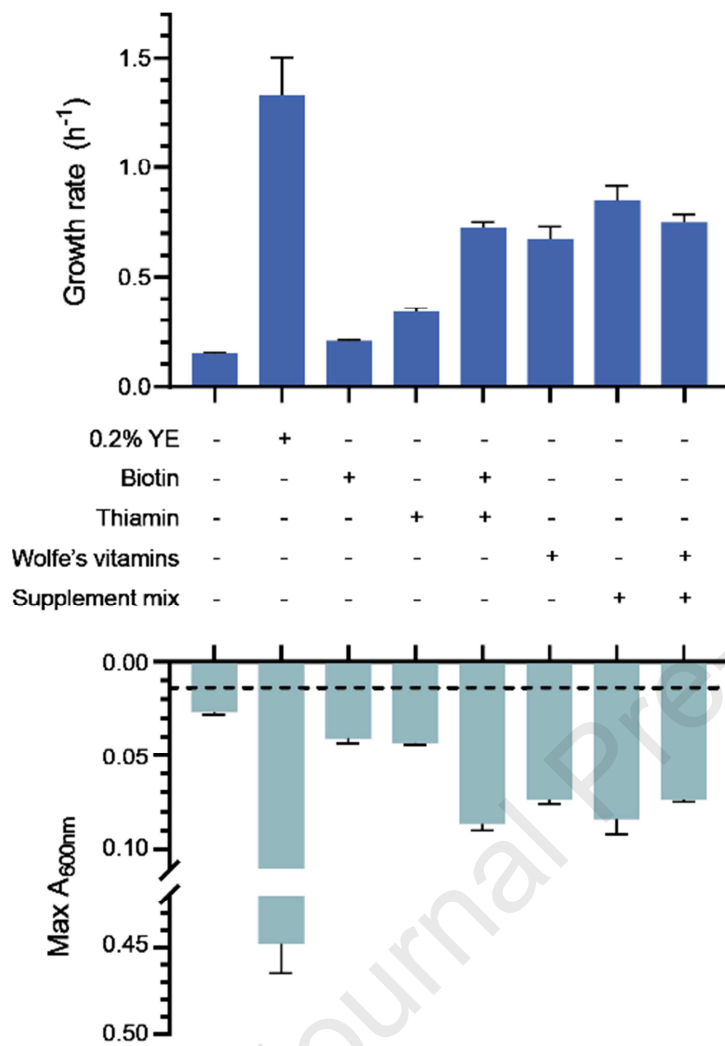
442 requires oxygen (EC 1.1.3.15). However, reports show that *Moorella thermoacetica*, a
443 thermophilic obligate anaerobe, can grow on glycolate through the formation of glyoxylate,
444 highlighting the possibility for a (to date) unknown alternate electron acceptor^{92,93}

445 Finally, heme is suggested to be synthesized in *P. thermoglucosidasius* from glycine using a
446 5-aminolevulinic acid synthase⁵¹ and notably using an oxygen-dependent
447 protoporphyrinogen oxidase. Both *E. coli* and *B. subtilis* have an oxygen independent
448 coproporphyrinogen-III oxidase (hemN), known to be responsible for anaerobic heme
449 biosynthesis, using other electron acceptors such as fumarate, or nitrate over oxygen⁹⁴⁻⁹⁷.
450 While the current genome annotation of *P. thermoglucosidasius* NCIMB 11955⁵⁸ suggests
451 that only the oxygen dependent path is present, the data presented herein suggest that
452 supplementation with hemin is not required for growth (Figure 5). One possible explanation
453 for this discrepancy can be found when performing a tBLASTn with the *B. subtilis* hemN
454 (NCBI accession CAB61616) against the *P. thermoglucosidasius* NCIMB 11955 genome.
455 This highlighted a significant hit (CP016622 region 3448674..3449762, 52% identity, E-
456 value: 10^{-115}). This suggests the possibility that some form of this oxygen independent heme
457 biosynthesis route could also be present highlighting the need to better understand the
458 metabolism of non-model organism chassis.

459 This identification of the minimal, defined anaerobic medium highlights how GEMs can be
460 used to facilitate experimental hypotheses, where previous hypotheses have failed. The
461 result, a defined minimal anaerobic medium is valuable for further investigation into
462 anaerobic metabolism through ¹³C-characterization studies, where defined media are critical.
463 Additionally, this identification of a series of components which support anaerobic growth of
464 *P. thermoglucosidasius* at a minimal medium level can further help inform the development
465 of industrial growth media for other microbial chassis used in anaerobic bioprocesses
466 improving growth and chemical product yields.

467

468



469

470 **Figure 5:** Experimental growth rates calculated and maximum observed absorbance values when *P.*
 471 *thermoglucosidasius* NCIMB 11955 was grown anaerobically in a microtiter plate reader in TMM base
 472 medium, supplemented with various nutrients, as indicated. Dashed line indicates inoculation
 473 absorbance, when an inoculation optical density of 0.05 was used.

474

475 3. Discussion

476 *Parageobacillus* spp. represent valuable microbial chassis for metabolic engineering and
477 fermentative bioproduction. Many advantages derive from their thermophilic character, with
478 additional advantages coming from species specific traits. However, to further develop
479 *Parageobacillus* spp. into fully optimized microbial cell factories, additional in-depth and
480 systems level understanding of metabolism is required, for which omic analyses and
481 genome scale metabolic models are critical. Currently, various automatic pipelines exist for
482 generating metabolic models on the sole basis of a genome sequence. Yet, for thermophilic
483 organisms, significant faults resulting from automatic annotation pipelines are evident, as
484 these are based on predominantly mesophilic datasets. Thermophilic genomes show
485 different characteristics, the effect of which on metabolism is still poorly understood, making
486 translation into a predicted function difficult. Thus, significant manual curation is needed in
487 the generation of GEMs for thermophilic organisms, which is limited by the availability of
488 knowledge on thermophilic metabolism. This is reflected by the relatively high need for
489 manual gap-filling, and a high resulting percentage of blocked reactions in *p-thermo*. To
490 increase the understanding of genotype-phenotype relationships in thermophilic hosts, the
491 availability of a GEM acts as a considerable step facilitating systems level studies. As a
492 result, with more knowledge arising, iterative rounds of model improvement are possible.

493 Therefore, in this work, we developed *p-thermo*, to date the most complete, curated and
494 validated genome-scale metabolic model for a facultative anaerobic *Parageobacillus* sp. In it,
495 genomic and biochemical knowledge were combined into a single powerful knowledge base,
496 providing a critical tool for data-driven metabolic engineering, -omic data integration, process
497 design and optimization. The model accurately captured substrate usage *in silico*, showing
498 the metabolic flexibility of the strain for production with alternative carbon sources (Lisowska,
499 2016). Furthermore, ¹³C-isotopic data verified the quality of *p-thermo* for predicting central
500 internal fluxes of the model, when solely production and consumption rates are measured, a
501 common practice when evaluating metabolic engineering designs.

502 Going beyond validation, *p-thermo* was used to provide more in-depth analysis of previously
503 reported metabolic engineering approaches³³. Initial *p-thermo* simulations did not completely
504 match experimental data, as stoichiometric models are incapable of capturing complex
505 levels of regulation that play a dominant role in *in vivo* metabolism. Nonetheless, *p-thermo*
506 was used to investigate the pyruvate branch point in central metabolism and allowed
507 additional insights into metabolic flux distributions in various genetic backgrounds. With *p-*
508 *thermo* further insights into metabolism can be gained, allowing improved targeted metabolic
509 engineering in subsequent designs.

510 Finally, we used *p-thermo* to generate hypothesis driven experiments to alleviate a
511 bottleneck in anaerobic metabolism, where previous experimental design was
512 unsuccessful¹⁴. Doing so revealed fundamental insights into the metabolism of *P.*
513 *thermoglucosidasius*, and also demonstrated that significant knowledge gaps still exist. This
514 analysis, in combination with the ¹³C-based verification, highlights an additional obstacle in
515 working with thermophilic GEMs, where annotation pipelines are less precise: information on
516 central carbon metabolism can be inferred with relative accuracy, whereas peripheral
517 metabolic pathways are not significantly understood and require further systems-level
518 investigation.

519 Overall, *p-thermo*, together with other systems level and omics based approaches, act as a
520 tool to improve our understanding of genotype-phenotype relationships. Genome-scale
521 metabolic models are in this way a critical part of an iterative cycle, and are essential to the
522 use and efficacy of thermophilic hosts for metabolic engineering and industrial bioproduction.

523

524

525 4. Materials & Methods

526 4.1 Model construction & curation

527 Genome sequencing of *Parageobacillus thermoglucosidasius* NCIMB 11955 was initially
528 performed by ERGOTM Integrated Genomics⁵⁷ (funded by TMO Renewables Ltd) and
529 subsequently updated using the published *P. thermoglucosidasius* NCIMB 11955 genome
530 sequence⁵⁸ (NCBI accession CP016622[chromosome], CP016623[pNCI001], and
531 CP016624[pNCI002]). Genome annotation was performed through the ERGOTM Integrated
532 Genomics suite⁵⁷ and the RAST server⁹⁸. Pathway Booster was used for gap filling,
533 resolving gaps through comparisons with evolutionarily-related genomes⁶⁰. Upon base
534 construction of the model, further manual curation was done following standard
535 procedures⁶¹. To do so, missing information was primarily obtained from literature and using
536 various databases: BRENDA, EC2PDB, KEGG, MetaNetX or MetaCyc⁵¹. Whenever
537 information on *P. thermoglucosidasius* was lacking, available references from other
538 (*Para*)*geobacillus* spp or *Bacillus* spp were added. All further manual curation and
539 refinement can be found in the GitHub repository. Model improvement was ensured by
540 running Memote⁴⁹ after each modification.

541 4.2 Biomass composition and growth energetics

542 To model growth, a biomass pseudo-reaction was added to the model. The reaction pools
543 metabolites needed for growth into a biomass metabolite. Base biomass composition was
544 previously determined experimentally according to reported practices^{9,51}. Lipid composition
545 was obtained from previous reports⁷, and was incorporated into the model according to a
546 restrictive approach⁶², in which a determined acyl chain length is assumed for all lipid
547 species. Further fine-tuning of biomass composition was performed based on available
548 enzymatic and metabolic requirements of the strain, with case-by-case justification given in
549 the GitHub repository. Critical metabolites known to be required for catabolic functions of
550 essential enzymes, such as heme, were added at trace stoichiometries based on knowledge
551 from related organisms and scaled to ensure that all biomass components added up to 1

552 $\text{g/g}_{\text{DW}}^{99}$. Growth energetics (ATP cost of growth-associated maintenance and ATP
553 requirement for non-growth associated maintenance) were estimated by minimizing the
554 prediction error of the specific substrate consumption rate and the specific growth rate of
555 glucose fed, aerobic chemostats^{51,61}. The P/O ratios were obtained from the given data for *B.*
556 *subtilis*¹⁰⁰. The contribution of polymerization of each metabolite type to the total growth
557 associated maintenance was estimated based on previously reported polymerization
558 energies⁶⁵.

559 **4.3 Transport reactions**

560 The model has two compartments: extracellular and intracellular. Transport reactions were
561 inferred from genome annotations and homology to known transporters. Additionally,
562 knowledge about growth on various substrates was used to validate the presence of the
563 corresponding transporters.

564 **4.4 Stoichiometric modeling & applied constraints**

565 In traditional flux balance analysis¹⁰, reaction stoichiometries are converted into a
566 stoichiometric matrix (**S**), with $m \times n$ dimensions, where m represents the various
567 metabolites and n represents the number of reactions. Coefficients in the matrix are either
568 positive or negative, reflecting production and consumption, respectively. Stoichiometric
569 modeling works under the assumption of a pseudo-steady-state, represented as:

$$570 \quad \mathbf{S} \cdot \mathbf{v} = 0$$

571 Where the vector **v** contains the fluxes of all reactions, given in units of $\text{mmol/g}_{\text{DW}}\text{h}$. As there
572 are more metabolites than reactions, to solve this underdetermined system, linear
573 programming is used by formulating an objective function (z), per default set as biomass
574 accumulation. Reversibility of reactions is set based on thermodynamic prediction and a
575 default medium is defined; in the case of *p-thermo*, minimal medium, with D-glucose as
576 default carbon source is used.

577 Quantification of metabolic fluxes was performed using flux variability analysis (FVA)⁷².
578 When running FVA, a threshold below the optimum is used to represent the metabolic
579 freedom that is given to a model. In this study, a sensitivity analysis was run with FVA
580 thresholds from 90 to 99%, to evaluate at what sensitivity level the model better matches the
581 experimental data. Additionally, parsimonious flux balance analysis (pFBA) was run, by
582 conducting a bilevel linear programming optimization that computes the optimum (growth)
583 solution of the network, whilst minimizing the sum of all fluxes⁷³. In doing so, this
584 optimization predicts the most stoichiometrically efficient pathway set, and captures the
585 maximum biomass per unit flux objective that has previously been described to be well
586 supported by proteomic and transcriptomic data^{73,101}.

587 **4.5 Genome comparison**

588 For an unbiased genome comparison, the *P. thermoglucosidasius* NCIMB 11955 genome
589 was obtained from NCBI (Accession: CP016622), and the *P. thermoglucosidasius* M10EXG
590 genome was obtained from the Integrated Microbial Genome database (ID 2501416905).
591 Genome annotation was performed using RASTk⁹⁸, after which the two proteomes were
592 compared through blast bi-directional best hits to create a homology matrix between the
593 strains, based on a published pipeline¹⁰². To filter for metabolic genes, any ORF associated
594 to a predicted EC code was considered metabolic. The exact workflow can be followed in the
595 GitHub repository.

596 **4.6 Experimental procedures**

597 The *P. thermoglucosidasius* NCIMB 11955 (DSM2542) strain was obtained from DSMZ⁷⁹.
598 The strain was grown in either 2SPY medium or base thermophile minimal medium (TMM),
599 modified from Fong *et al.*⁸⁰. 2SPY was used for a first preculture, and contains per liter, 16g
600 soy peptone, 10 g yeast extract and 5 g NaCl, adjusted to pH 6.8. Base TMM contains, per
601 liter: 930 ml Six salts solution (SSS), 40 ml 1M MOPS (pH 8.2), 10 ml 1 mM FeSO₄ in 0.4 M
602 tricine, 10 ml 0.132 M K₂HPO₄, 10 ml 0.953 M NH₄Cl, 0.5 ml 1M CaCl₂ and trace element

603 solution, adjusted to a final pH of 6.8. SSS contains, per 930 ml: 4.6 g NaCl, 1.35 g Na₂SO₄,
604 0.23 g KCl, 0.037 g KBr, 1.72 g MgCl₂·6 H₂O and 0.83g NaNO₃. The trace element solution
605 contained, per liter, 1g FeCl₃·6 H₂O, 0.18 g ZnSO₄·7 H₂O, 0.12 g CuCl₂·2 H₂O, 0.12 g
606 MnSO₄·H₂O and 0.18 g CoCl₂·6 H₂O. D-glucose to a final concentration of 10g/L was added
607 to the base TMM.

608 When indicated, the base TMM was supplemented with one of the following, to the indicated
609 final concentrations: 0.2% (w/v) yeast extract, 2 µg/L biotin, 5 µg/L thiamine-HCl, 1x Wolfe's
610 vitamins, or 1x supplementation mix. 1000x Wolfe's vitamins consist of, per liter, 10 mg
611 pyridoxine hydrochloride, 5.0 mg thiamine-HCl, 5.0 mg riboflavin, 5.0 mg nicotinic acid, 5.0
612 mg calcium D-(+)-pantothenate, 5.0 mg p-aminobenzoic acid, 5.0 mg thioctic acid, 2.0 mg
613 biotin, 2.0 mg folic acid and 0.1 mg vitamin B12. The 1000x supplementation mix contained,
614 per liter, 2.0 mg biotin, 5.0 mg thiamine-HCl, 2.0 mg folic acid, 0.1 mg vitamin B12, 127 µg/L
615 spermidine, 174 µg/L spermine tetrahydrochloride and 0.7 mg hemin.

616 The aerobic cultures were inoculated to a starting OD₆₀₀ of around 0.05, after an overnight
617 culture on the base TMM medium. Growth was monitored through OD₆₀₀ measurements,
618 during growth at 60°, 200 rpm in baffled shake flasks. Anaerobic medium was prepared
619 similarly to aerobic medium, but 1 µg/L resazurin was added to ensure complete anaerobic
620 conditions. The medium was flushed with nitrogen gas prior to use. All anaerobic work was
621 performed in an anaerobic chamber. Overnight cultures were run in anaerobic serum flasks
622 at 60°C, 200 rpm and used to inoculate a microtiter plate to a final OD of 0.05 in 200 µL
623 volume. After sealing, the OD₆₀₀ was measured every 15 minutes for 10 hours in a Biotek
624 Epoch2 microplate spectrophotometer, placed in an anaerobic chamber (run at 60°C, with
625 linear shaking).

626

627 **5. Abbreviations**628 **Metabolites**

629	13dpg	3-Phosphoglyceroyl phosphate
630	2pg	2-phosphoglycerate
631	3pg	3-phosphoglycerate
632	6pgl	6-phosphogluconolactone
633	ac	acetate
634	acald	acetaldehyde
635	accoa	acetyl-CoA
636	actp	acetyl phosphate
637	akg	α -ketoglutarate
638	asp	aspartate
639	cit	citrate
640	dhp	dihydroxyacetone phosphate
641	e4p	erythrose 4-phosphate
642	etoh	ethanol (etoh)
643	f6p	fructose-6-phosphate
644	fdp	fructose 1,6-bisphosphate
645	for	formate
646	fum	fumarate
647	g3p	glyceraldehyde-3-phosphate
648	g6p	glucose-6-phosphate
649	glc	glucose
650	gly	glycine
651	icit	iso-citrate
652	lac	L-lactate
653	mal	malate
654	oaa	oxaloacetate
655	pep	phosphoenolpyruvate
656	phe	phenylalanine
657	pyr	pyruvate
658	r5p	ribose-5-phosphate
659	ru5p	ribulose-5-phosphate
660	s7p	sedoheptulose 7-phosphate
661	ser	serine
662	succ	succinate
663	succoa	succinyl-CoA
664	udpg	uridine diphosphate glucose
665	xu5p	xylulose-5-phosphate

666

667

668

669	Reactions	
670	ACKr	Acetate kinase
671	ACONTa	Aconitase
672	ALCD2x	Alcohol dehydrogenase
673	ASPTA	Aspartate transaminase
674	CS	Citrate synthase
675	DDPA	3-deoxy-D-arabino-heptulosonate 7-phosphate synthetase
676	ENO	Enolase
677	FBA	Fructose-bisphosphate aldolase
678	FUM	fumarate hydratase
679	G6PDH2r	Glucose 6-phosphate dehydrogenase
680	GALUi	UTP-glucose-1-phosphate uridylyltransferase
681	GHMT	Glycine hydroxymethyltransferase
682	GLCtpts	Glucose phosphotransferase transporter
683	GLUSy	Glutamate synthase
684	ICL	Isocitrate lyase
685	LDH_L	L-lactate dehydrogenase
686	MDH	Malate dehydrogenase
687	PC	Pyruvate carboxylase
688	PDH	Pyruvate dehydrogenase
689	PFL	Pyruvate formate lyase
690	PGCD	Phosphoglycerate dehydrogenase
691	PGI	Glucose-6-phosphate isomerase
692	PGK	Phosphoglycerate kinase
693	PPCK	Phosphoenolpyruvate carboxykinase
694	PRPPS	Phosphoribosylpyrophosphate synthetase
695	PSCVT	3-phosphoshikimate 1-carboxyvinyltransferase
696	SUCDi	Succinate dehydrogenase
697	SUCOAS	Succinyl-CoA synthetase
698	TALA	Transaldolase
699	TKT1	Transketolase
700	TKT2	Transketolase
701		
702		

703 **Availability of data and materials**

704 The metabolic model, scripts and corresponding datasets generated during the study are all
705 freely available, under an Apache 2.0 license, at the GitHub repository:
706 <https://github.com/biosustain/p-thermo/releases/v1.0>.

707 **Funding**

708 VM was funded by the Novo Nordisk Foundation through NNF18CC0033664; MB was
709 supported by the EPSRC EP/L016354/1; BJS, MJH and NS acknowledge funding from the
710 European Union's Horizon 2020 research and innovation program under grant agreement
711 No 686070; NS furthermore acknowledges support from the Novo Nordisk Foundation
712 (NNF17SA0031362).; BKL was funded by a BBSRC-CASE studentship (Project code
713 1100372); ATN was supported through NNF16OC0021814 and NNF20CC0035580; DJL
714 also acknowledges support from BBSRC (BB/J001120/2).

715

716 **Acknowledgements**

717 We thank Dr. Shyam Maskapalli for help with biomass composition analysis.

718

719 6. References

- 720 1. Steen, E. J. *et al.* Microbial production of fatty-acid-derived fuels and chemicals from
721 plant biomass. *Nature* **463**, 559–563 (2010).
- 722 2. Lee, S. Y. & Kim, H. U. Systems strategies for developing industrial microbial strains.
723 *Nat. Biotechnol.* **33**, 1061–1072 (2015).
- 724 3. Nielsen, J. & Keasling, J. D. Engineering Cellular Metabolism. *Cell* **164**, 1185–1197
725 (2016).
- 726 4. Hollinshead, W., He, L. & Tang, Y. J. Biofuel production: an odyssey from metabolic
727 engineering to fermentation scale-up. *Front. Microbiol.* **5**, 1–8 (2014).
- 728 5. Liu, Y., Li, J., Du, G., Chen, J. & Liu, L. Metabolic engineering of *Bacillus subtilis*
729 fueled by systems biology: Recent advances and future directions. *Biotechnol. Adv.*
730 **35**, 20–30 (2017).
- 731 6. Ma, W. *et al.* Metabolic engineering of carbon over flow metabolism of *Bacillus subtilis*
732 for improved N-acetyl-glucosamine production. *Bioresour. Technol.* **250**, 642–649
733 (2018).
- 734 7. Tang, Y. J. *et al.* Analysis of metabolic pathways and fluxes in a newly discovered
735 thermophilic and ethanol-tolerant geobacillus strain. *Biotechnol. Bioeng.* **102**, 1377–
736 1386 (2009).
- 737 8. Choi, K. R. *et al.* Systems metabolic engineering strategies: integrating systems and
738 synthetic biology with metabolic engineering. *Trends Biotechnol.* **37**, 817–837 (2019).
- 739 9. Durot, M., Bourguignon, P. & Schachter, V. Genome-scale models of bacterial
740 metabolism: reconstruction and applications. *FEMS Microbiol. Rev.* **33**, 164–190
741 (2009).
- 742 10. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.*
743 **28**, 245–248 (2010).
- 744 11. Zhang, J. *et al.* Engineering an NADPH/NADP⁺ Redox Biosensor in Yeast. *ACS*
745 *Synth. Biol.* **5**, 1546–1556 (2016).
- 746 12. Pharkya, P., Burgard, A. P. & Maranas, C. D. OptStrain: A computational framework
747 for redesign of microbial process. *Genome Res.* **14**, 2367–2376 (2004).
- 748 13. Jensen, K., Broeken, V., Hansen, A. S. L., Sonnenschein, N. & Herrgård, M. J.
749 OptCouple: Joint simulation of gene knockouts, insertions and medium modifications
750 for prediction of growth-coupled strain designs. *Metab. Eng. Commun.* **8**, (2019).
- 751 14. Hussein, A. H., Lisowska, B. K. & Leak, D. J. *The genus Geobacillus and their*
752 *biotechnological potential. Advances in Applied Microbiology* vol. 92 (Elsevier, 2015).
- 753 15. Wada, K. & Suzuki, H. Biotechnological platforms of the moderate thermophiles ,
754 *Geobacillus* species: notable properties and genetic tools. in *Physiological and*
755 *Biotechnological Aspects of Extremophiles* 195–218 (INC, 2018). doi:10.1016/B978-0-
756 12-818322-9.00015-0.
- 757 16. Zeigler, D. R. The *Geobacillus* paradox: why is a thermophilic bacterial genus so
758 prevalent on a mesophilic planet? *Microbiology* 1–11 (2014)
759 doi:10.1099/mic.0.071696-0.
- 760 17. Suzuki, H. Peculiarities and biotechnological potential of environmental adaptation by
761 *Geobacillus* species. *Appl. Microbiol. Biotechnol.* 10425–10437 (2018).

- 762 18. Dahal, S., Poudel, S. & Thompson, R. A. Genome-Scale Modeling of Thermophilic
763 Microorganisms. *Adv. Biochem. Eng. Biotechnol.* (2016) doi:10.1007/10.
- 764 19. Panikov, N. S., Popova, N. A., Dorofeev, A. G., Nikolaev, Y. A. & Verkhovtseva, N. V.
765 Growth of the Thermophilic Bacterium *Geobacillus uralicus* as a Function of
766 Temperature and pH: An SCM-Based Kinetic Analysis. *Microbiology* **72**, 320–327
767 (2003).
- 768 20. Krüger, A., Schäfers, C., Schröder, C. & Antranikian, G. Towards a sustainable
769 biobased industry – Highlighting the impact of extremophiles. *N. Biotechnol.* **40**, 144–
770 153 (2018).
- 771 21. Bartosiak-Jentys, J., Hussein, A. H., Lewis, C. J. & Leak, D. J. Modular system for
772 assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*.
773 *Microbiology* **159**, 1267–1275 (2013).
- 774 22. Huang, D. *et al.* Synergistic hydrolysis of xylan using novel xylanases, β -xylosidases
775, and an α -L-arabinofuranosidase from *Geobacillus thermodenitrificans* NG80-2.
776 *Appl. Microbiol. Biotechnol.* **101**, 6023–6037 (2017).
- 777 23. Bibra, M., Kunreddy, V. R. & Sani, R. K. Thermostable Xylanase Production by
778 *Geobacillus* sp. Strain DUSELR13, and Its Application in Ethanol Production with
779 Lignocellulosic Biomass. *Microorganisms* **93**, 1–25 (2018).
- 780 24. Liu, B. *et al.* Characterization of a recombinant thermostable xylanase from hot spring
781 thermophilic *Geobacillus* sp. TC-W7. *J. Microbiol. Biotechnol.* **22**, 1388–1394 (2012).
- 782 25. Balazs, Y. S. *et al.* Identifying critical unrecognized sugar – protein interactions in
783 GH10 xylanases from *Geobacillus stearothermophilus* using STD NMR. *FEBS J.* **280**,
784 4652–4665 (2013).
- 785 26. Maayer, P. De, Brumm, P. J., Mead, D. A. & Cowan, D. A. Comparative analysis of
786 the *Geobacillus* hemicellulose utilization locus reveals a highly variable target for
787 improved hemicellulolysis. *BMC Genomics* **15**, 1–17 (2014).
- 788 27. Taylor, M. P., Esteban, C. D. & Leak, D. J. Development of a versatile shuttle vector
789 for gene expression in *Geobacillus* spp. *Plasmid* **60**, 45–52 (2008).
- 790 28. Bacon, L. F., Hamley-Bennett, C., Danson, M. J. & Leak, D. J. Development of an
791 efficient technique for gene deletion and allelic exchange in *Geobacillus* spp. *Microb.*
792 *Cell Fact.* **16**, 1–8 (2017).
- 793 29. Kananavičiūtė, R. & Čitavičius, D. Genetic engineering of *Geobacillus* spp. *J.*
794 *Microbiol. Methods* **111**, 31–39 (2015).
- 795 30. Reeve, B., Martinez-Klimova, E., Jonghe, J. de, Leak, D. J. & Ellis, T. The *Geobacillus*
796 Plasmid Set: A Modular Toolkit for Thermophile Engineering. *ACS Synth. Biol.* **5**,
797 1342–1347 (2016).
- 798 31. Pogrebnyakov, I., Jendresen, C. B. & Nielsen, A. T. Genetic toolbox for controlled
799 expression of functional proteins in *Geobacillus* spp. *PLoS One* **12**, (2017).
- 800 32. Drejer, E. B., Hakvåg, S., Irla, M. & Brautaset, T. Genetic tools and techniques for
801 recombinant expression in thermophilic bacillaceae. *Microorganisms* **6**, 1–19 (2018).
- 802 33. Cripps, R. E. *et al.* Metabolic engineering of *Geobacillus thermoglucosidasius* for high
803 yield ethanol production. *Metab. Eng.* **11**, 398–408 (2009).
- 804 34. Niu, H., Leak, D., Shah, N. & Kontoravdi, C. Metabolic characterization and modeling
805 of fermentation process of an engineered *Geobacillus thermoglucosidasius* strain for
806 bioethanol production with gas stripping. *Chem. Eng. Sci.* **122**, 138–149 (2015).

- 807 35. Lin, P. P. *et al.* Isobutanol production at elevated temperatures in thermophilic
808 *Geobacillus thermoglucosidasius*. *Metab. Eng.* **24**, 1–8 (2014).
- 809 36. Mohr, T. *et al.* CO-dependent hydrogen production by the facultative anaerobe
810 *Parageobacillus thermoglucosidasius*. *Microb. Cell Fact.* **17**, 1–12 (2018).
- 811 37. Aliyu, H., Mohr, T., Cowan, D. & Maayer, P. De. Time-course transcriptome of
812 *Parageobacillus thermoglucosidasius* DSM 6285 grown in the presence of carbon
813 monoxide and air. *Int. J. Mol. Sci.* **21**, (2020).
- 814 38. Kulyashov, M., Peltek, S. E. & Akberdin, I. R. A genome-scale metabolic model of 2,3-
815 butanediol production by thermophilic bacteria *Geobacillus icigianus*. *Microorganisms*
816 **8**, (2020).
- 817 39. Zhou, J., Lian, J. & Rao, C. V. Metabolic engineering of *Parageobacillus*
818 *thermoglucosidasius* for the efficient production of (2R,3R)-butanediol. *Appl.*
819 *Microbiol. Biotechnol.* **104**, 4303–4311 (2020).
- 820 40. Yang, Z. *et al.* Engineering thermophilic *Geobacillus thermoglucosidasius* for
821 riboflavin production. *Microb. Biotechnol.* **0**, 1–11 (2020).
- 822 41. Styles, M. Q. *et al.* The heterologous production of terpenes by the thermophile
823 *Parageobacillus thermoglucosidasius* in a consolidated bioprocess using waste bread.
824 *Metab. Eng.* (2020) doi:doi.org/10.1016/j.ymben.2020.11.005.
- 825 42. Zhu, W., Cha, D., Cheng, G., Peng, Q. & Shen, P. Purification and characterization of
826 a thermostable protease from a newly isolated *Geobacillus* sp . YMTC 1049. *Enzyme*
827 *Microb. Technol.* **40**, 1592–1597 (2007).
- 828 43. Ewis, H. E., Abdelal, A. T. & Lu, C. Molecular cloning and characterization of two
829 thermostable carboxyl esterases from *Geobacillus stearothermophilus*. *Gene* **329**,
830 187–195 (2004).
- 831 44. Zhu, Y. *et al.* Molecular cloning and characterization of a thermostable lipase from
832 deep-sea thermophile *Geobacillus* sp . EPT9. *World J. Microbiol. Biotechnol.* **31**, 295–
833 306 (2015).
- 834 45. Chen, P., Liu, C., Chen, Y., Hsu, F. & Shaw, J. Isolation, expression and
835 characterization of the thermophilic recombinant esterase from *Geobacillus*
836 *thermodenitrificans* PS01. *Appl. Biochem. Biotechnol.* **191**, 112–124 (2020).
- 837 46. Fotouh, D. M. A., Bayoumi, R. A. & Hassan, M. A. Production of thermoalkaliphilic
838 lipase from *Geobacillus thermoleovorans* DA2 and application in leather industry.
839 *Enzyme Res.* 1–9 (2016).
- 840 47. Ma, Y. *et al.* Enhancement of polymerase activity of the large fragment in DNA
841 Polymerase I from *Geobacillus stearothermophilus* by site-directed mutagenesis at
842 the active site. *Biomed Res. Int.* 1–8 (2016).
- 843 48. Ahmad, A. *et al.* A Genome Scale Model of *Geobacillus thermoglucosidasius* (C56-
844 YS93) reveals its biotechnological potential on rice straw hydrolysate. *J. Biotechnol.*
845 **251**, 30–37 (2017).
- 846 49. Lieven, C. *et al.* MEMOTE for standardized genome-scale metabolic model testing.
847 *Nat. Biotechnol.* **38**, 272–276 (2020).
- 848 50. Cordova, L. T., Long, C. P., Venkataramanan, K. P. & Antoniewicz, M. R. Complete
849 genome sequence, metabolic model construction and phenotypic characterization of
850 *Geobacillus* LC300, an extremely thermophilic, fast growing, xylose-utilizing
851 bacterium. *Metab. Eng.* **32**, 74–81 (2015).

- 852 51. Lisowska, B. K. Genomic analysis and metabolic modelling of *Geobacillus*
853 *thermoglucosidasius* NCIMB 11955. (University of Bath, 2016).
- 854 52. Hucka, M. *et al.* *The Systems Biology Markup Language (SBML): Language*
855 *Specification for Level 3 Version 2 Core Release 2. Journal of integrative*
856 *bioinformatics* vol. 16 (2019).
- 857 53. Ebrahim, A., Lerman, J. A., Palsson, B. O. & Hyduke, D. R. COBRApy: COntstraints-
858 Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* **7**, 1–6 (2013).
- 859 54. L, H. *et al.* Creation and analysis of biochemical constraint-based models: the COBRA
860 Toolbox v3.0. *Nat. Protoc.* **14**, 639–702 (2019).
- 861 55. Cardoso, J. G. R. *et al.* Cameo: A Python Library for Computer Aided Metabolic
862 Engineering and Optimization of Cell Factories. *ACS Synth. Biol.* **7**, 1163–1166
863 (2018).
- 864 56. Rocha, I. *et al.* OptFlux: An open-source software platform for in silico metabolic
865 engineering. *BMC Syst. Biol.* **4**, 1–12 (2010).
- 866 57. Overbeek, R. *et al.* The ErgoTM genome analysis and discovery system. *Nucleic*
867 *Acids Res.* **31**, 164–171 (2003).
- 868 58. Sheng, L., Zhang, Y. & Minton, N. P. Complete genome sequence of *Geobacillus*
869 *thermoglucosidasius* NCIMB 11955, the progenitor of a bioethanol production strain.
870 *Genome Announc.* **4**, 4–5 (2016).
- 871 59. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology.
872 *BMC Genomics* **9**, 1–15 (2008).
- 873 60. Liberal, R., Lisowska, B. K., Leak, D. J. & Pinney, J. W. PathwayBooster: A tool to
874 support the curation of metabolic pathways. *BMC Bioinformatics* **16**, 4–9 (2015).
- 875 61. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale
876 metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
- 877 62. Sánchez, B. J., Li, F., Kerkhoven, E. J. & Nielsen, J. SLIMER: Probing flexibility of lipid
878 metabolism in yeast with an improved constraint-based modeling framework. *BMC*
879 *Syst. Biol.* **13**, 1–9 (2019).
- 880 63. Cordova, L. T. & Antoniewicz, M. R. ¹³C metabolic flux analysis of the extremely
881 thermophilic, fast growing, xylose-utilizing *Geobacillus* strain LC300. *Metab. Eng.* **33**,
882 148–157 (2015).
- 883 64. Olivier, B. G. & Bergmann, F. T. The Systems Biology Markup Language (SBML)
884 Level 3 Package: Flux Balance Constraints. *J. Integr. Bioinform.* **12**, (2015).
- 885 65. Verduyn, C., Stouthamer, A. H., Scheffers, W. A. & van Dijken, J. P. A theoretical
886 evaluation of growth yields of yeasts. *Antonie Van Leeuwenhoek* **59**, 49–63 (1991).
- 887 66. Förster, J., Famili, I., Fu, P., Palsson, B. Ø. & Nielsen, J. Genome-Scale
888 Reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. *Genome Res.*
889 **13**, 244–253 (2003).
- 890 67. Monk, J. M. *et al.* iML1515, a knowledgebase that computes *Escherichia coli* traits.
891 *Nat. Biotechnol.* **35**, 8–12 (2017).
- 892 68. Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. Chapter 3: *Escherichia coli* and
893 *Salmonella*: Cellular and Molecular Biology. in *Physiology of the Bacterial Cell: a*
894 *Molecular Approach* (American Society of Microbiology (ASM) Press, 1996).

- 895 69. Robb, F., Antranikian, G., Grogan, D. & Driessen, A. *Thermophiles: Biology and*
896 *Technology at High Temperatures*. (CRC Press, 2007).
- 897 70. King, Z. A. *et al.* Escher: a web application for building, sharing, and embedding data-
898 rich visualizations of biological pathways. *PLoS Comput. Biol.* 1–13 (2015)
899 doi:10.1371/journal.pcbi.1004321.
- 900 71. Miclet, E., Michels, P. A. M., Opperdoes, F. R., Lallemand, J. & Duffieux, F. NMR
901 spectroscopic analysis of the first two steps of the pentose-phosphate pathway
902 elucidates the role of 6-phosphogluconolactonase. *J. Biol. Chem.* **276**, 34840–34846
903 (2001).
- 904 72. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in
905 constraint-based genome-scale metabolic models. *Metab. Eng.* **5**, 264–276 (2003).
- 906 73. Lewis, N. E. *et al.* Omic data from evolved *E. coli* are consistent with computed
907 optimal growth from genome-scale models. *Mol. Syst. Biol.* (2010)
908 doi:10.1038/msb.2010.47.
- 909 74. Kim, Y., Ingram, L. O. & Shanmugam, K. T. Dihydrolipoamide dehydrogenase
910 mutation alters the NADH sensitivity of pyruvate dehydrogenase complex of
911 *Escherichia coli* K-12. *J. Bacteriol.* **190**, 3851–3858 (2008).
- 912 75. Wang, Q. *et al.* Metabolic flux control at the pyruvate node in an anaerobic
913 *Escherichia coli* strain with an active pyruvate dehydrogenase*. *Appl. Environ.*
914 *Microbiol.* **76**, 2107–2114 (2010).
- 915 76. Sawers, G. & Bock, A. Anaerobic regulation of pyruvate formate-lyase from
916 *Escherichia coli* K-12. *J. Bacteriol.* **170**, 5330–5336 (1988).
- 917 77. Branco dos Santos, F. *et al.* Probing the Genome-Scale Metabolic Landscape of
918 *Bordetella pertussis*, the Causative Agent of Whooping Cough. *Appl. Environ.*
919 *Microbiol.* **83**, 1–19 (2017).
- 920 78. Loftie-Eaton, W. *et al.* Balancing redox cofactor generation and ATP synthesis: Key
921 microaerobic responses in thermophilic fermentations. *Biotechnol. Bioeng.* **110**,
922 1057–1065 (2013).
- 923 79. Leibniz Institute. DSMZ-German Collection of Microorganisms and Cell Cultures
924 GmbH. <https://www.dsmz.de/> (2021).
- 925 80. Fong, J. C. N. *et al.* Isolation and characterization of two novel ethanol-tolerant
926 facultative-anaerobic thermophilic bacteria strains from waste compost. *Extremophiles*
927 **10**, 363–372 (2006).
- 928 81. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
929 2069 (2014).
- 930 82. Mendoza, S. N., Olivier, B. G., Molenaar, D. & Teusink, B. A Systematic Assessment
931 Of Current Genome-Scale Metabolic Reconstruction Tools. *Genome Biol.* **20**, 1–20
932 (2019).
- 933 83. Sabath, N., Ferrada, E., Barve, A. & Wagner, A. Growth temperature and genome
934 size in bacteria are negatively correlated, suggesting genomic streamlining during
935 thermal adaptation. *Genome Biol. Evol.* **5**, 966–977 (2013).
- 936 84. Wang, Q., Cen, Z. & Zhao, J. The survival mechanisms of thermophiles at high
937 temperatures: an angle of omics. *Physiology* 97–106 (2015)
938 doi:10.1152/physiol.00066.2013.
- 939 85. Lusk, B. G. Thermophiles; or, the Modern Prometheus: the importance of extreme

- 940 microorganisms for understanding and applying extracellular electron transfer. *Front.*
941 *Microbiol.* **10**, 1–10 (2019).
- 942 86. Huang, P. *et al.* Evaluating protein engineering thermostability prediction tools using
943 an independently generated dataset. *ACS Omega* 6487–6493 (2020)
944 doi:10.1021/acsomega.9b04105.
- 945 87. Fang, H., Kang, J. & Zhang, D. Microbial production of vitamin B 12: a review and
946 future perspectives. *Microb. Cell Fact.* **16**, 1–14 (2017).
- 947 88. Raux, E., Schubert, H. L. & Warren, M. J. Biosynthesis of cobalamin (vitamin B12): a
948 bacterial conundrum. *Cell. Mol. Life Sci.* **57**, 1880–1893 (2000).
- 949 89. Gajcy, H. Biosynthesis of Vitamin B12 by mesophili and themophilic strains of *Bacillus*
950 *megaterium*. *Polish J. Microbiol.* **5**, 119–127 (1973).
- 951 90. North, J. A. *et al.* Metabolic regulation as a consequence of anaerobic 5-
952 methylthioadenosine recycling in *Rhodospirillum rubrum*. *MBio* **7**, 1–12 (2016).
- 953 91. North, J. A., Miller, A. R., Wildenthal, J. A., Young, S. J. & Tabita, F. R. Microbial
954 pathway for anaerobic 5'methylthioadenosine metabolism coupled to ethylene
955 formation. *PNAS* 10455–10464 (2017) doi:10.1073/pnas.1711625114.
- 956 92. Seifritz, C., Fröstl, J. M., Drake, H. L. & Daniel, S. L. Glycolate as a metabolic
957 substrate for the acetogen *Moorella thermoacetica*. *FEMS Microbiol. Lett.* **170**, 399–
958 405 (1999).
- 959 93. Sakai, S., Inokuma, K., Nakashimada, Y. & Nishio, N. Degradation of glyoxylate and
960 glycolate with ATP Synthesis by a thermophilic anaerobic bacterium, *Moorella* sp.
961 strain HUC22-1. *Appl. Environ. Microbiol.* **74**, 1447–1452 (2008).
- 962 94. Jacobs, N. J. & Jacobs, J. M. Nitrate, Fumarate and oxygen as electron acceptors for
963 a late step in microbial heme synthesis. *Biochim. Biophys.* **449**, 1–9 (1976).
- 964 95. Hippler, B. *et al.* Characterization of *Bacillus subtilis* hemN. *J. Bacteriol.* **179**, 7181–
965 7185 (1997).
- 966 96. Layer, G., Verfürth, K., Mahlitz, E. & Jahn, D. Oxygen-independent
967 Coproporphyrinogen-III Oxidase HemN from *Escherichia coli*. *J. Biol. Chem.* **277**,
968 34136–34142 (2002).
- 969 97. Möbius, K. *et al.* Heme biosynthesis is coupled to electron transport chains for energy
970 generation. *PNAS* **107**, 10436–10441 (2010).
- 971 98. Brettin, T. *et al.* RASTtk: A modular and extensible implementation of the RAST
972 algorithm for building custom annotation pipelines and annotating batches of
973 genomes. *Sci. Rep.* **5**, (2015).
- 974 99. Chan, S. H. J., Cai, J., Wang, L., Simons-Senftle, M. N. & Maranas, C. D.
975 Standardizing biomass reactions and ensuring complete mass balance in genome-
976 scale metabolic models. *Syst. Biol. (Stevenage)*. **33**, 3603–3609 (2017).
- 977 100. Oh, Y. K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-
978 scale reconstruction of metabolic network in *Bacillus subtilis* based on high-
979 throughput phenotyping and gene essentiality data. *J. Biol. Chem.* **282**, 28791–28799
980 (2007).
- 981 101. Schuetz, R., Kuepfer, L. & Sauer, U. Systematic evaluation of objective functions for
982 predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.* (2007)
983 doi:10.1038/msb4100162.

984 102. Norsigian, C. J., Fang, X., Seif, Y., Monk, J. M. & Palsson, B. O. A workflow for
985 generating multi-strain genome-scale metabolic models of prokaryotes. *Nat. Protoc.*
986 **15**, 1–14 (2020).

987 103. Yoshida, K.-I., Aoyama, D., Ishio, I., Shibayama, T. & Fujita, Y. Organization and
988 transcription of the myo-inositol operon, *iol*, of *Bacillus subtilis*. *J. Bacteriol.* **179**,
989 4591–4598 (1997).

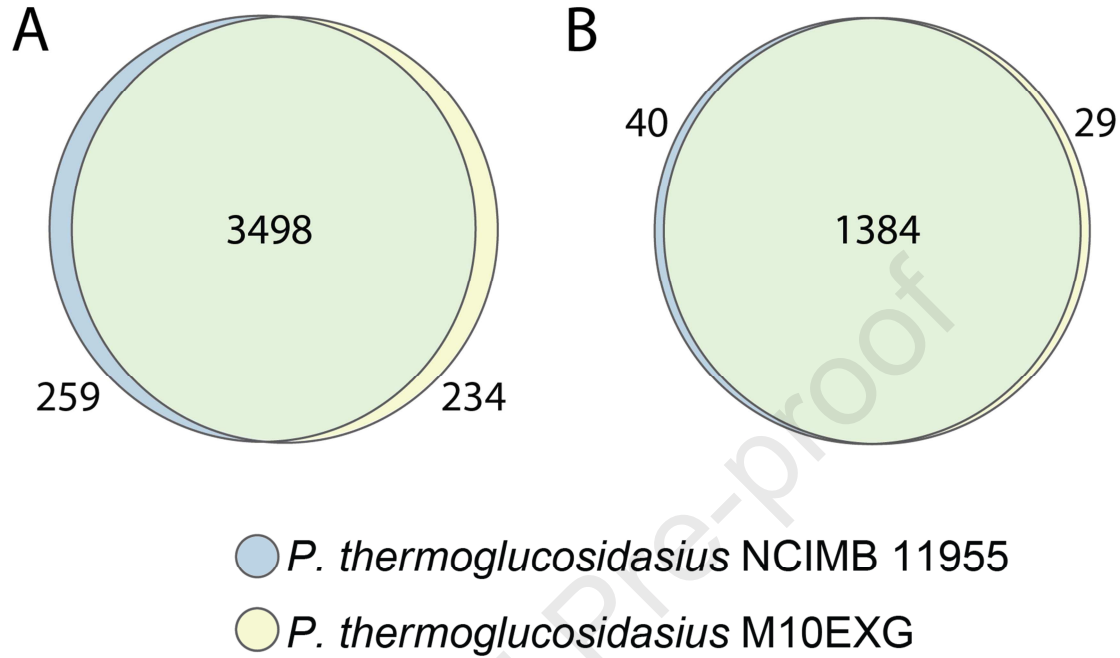
990

991

Journal Pre-proof

992 **Supplementary Figures**

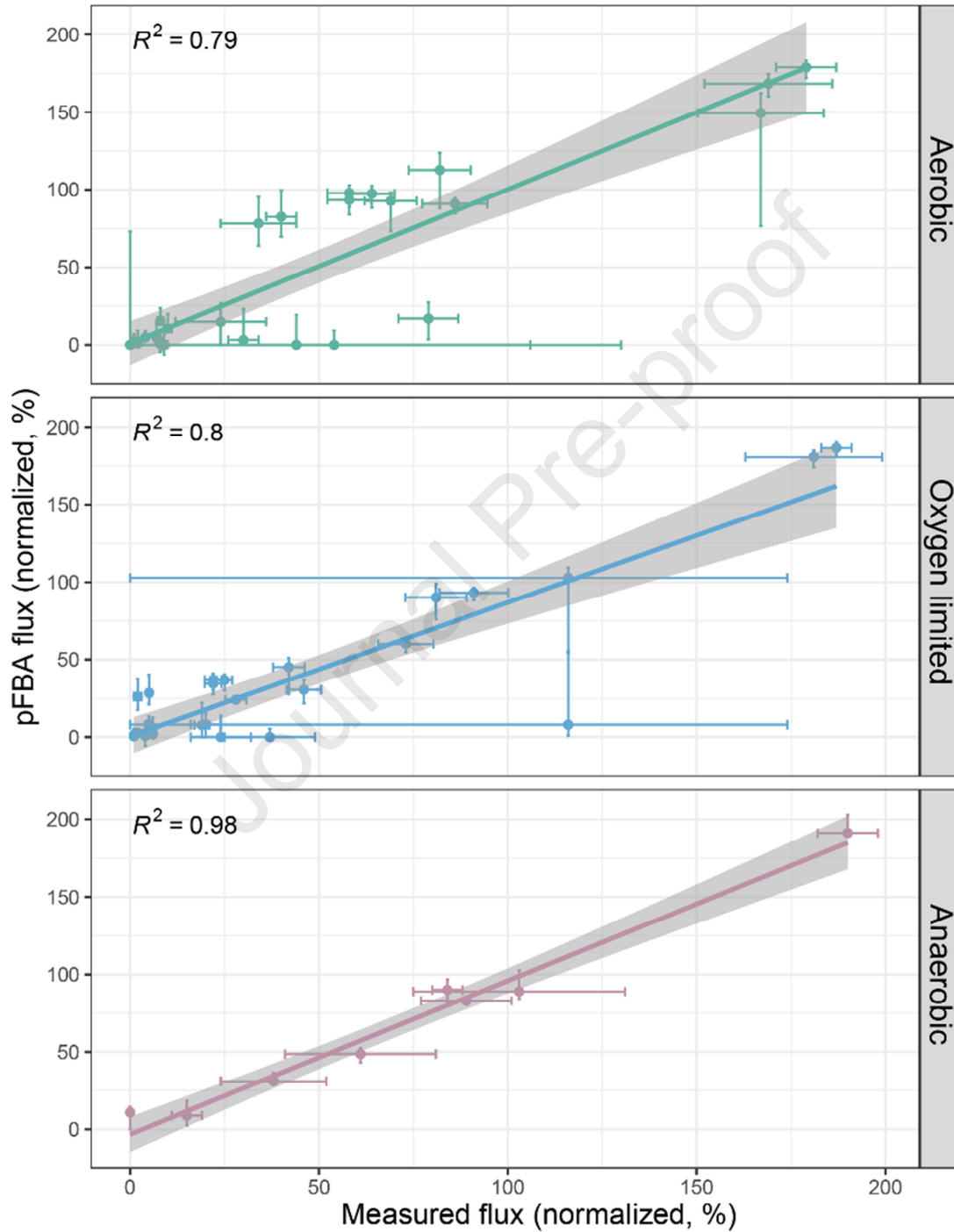
993 **Supplementary figure 1:** Whole proteome comparison between *P. thermoglucosidasius*
994 NCIMB 11955 and *P. thermoglucosidaius M10EXG*, for all ORFs (A) and when filtered for
995 metabolic genes (B, i.e. genes with an EC number associated to them). Supplementary
996 tables 4 and 5 list the unique metabolic ORFs between the two strains.



997

998

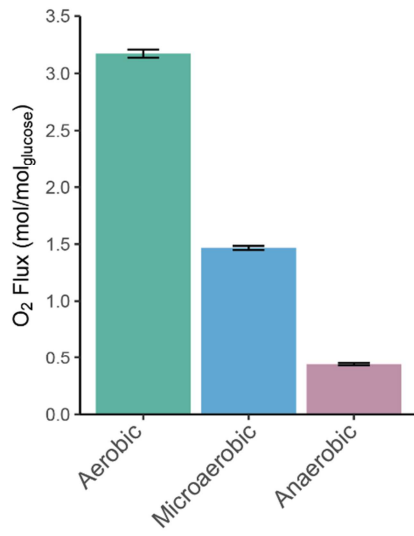
999 **Supplementary figure 2:** Correlation between pFBA analysis of the model and
1000 experimentally derived data⁷, normalized to glucose uptake rate, in aerobic, oxygen limited
1001 and anaerobic conditions. Errors for measured fluxes and variability in pFBA fluxes are
1002 shown. A linear fit has been applied to assay correlation, with the R value indicated per
1003 condition.



1004

1005

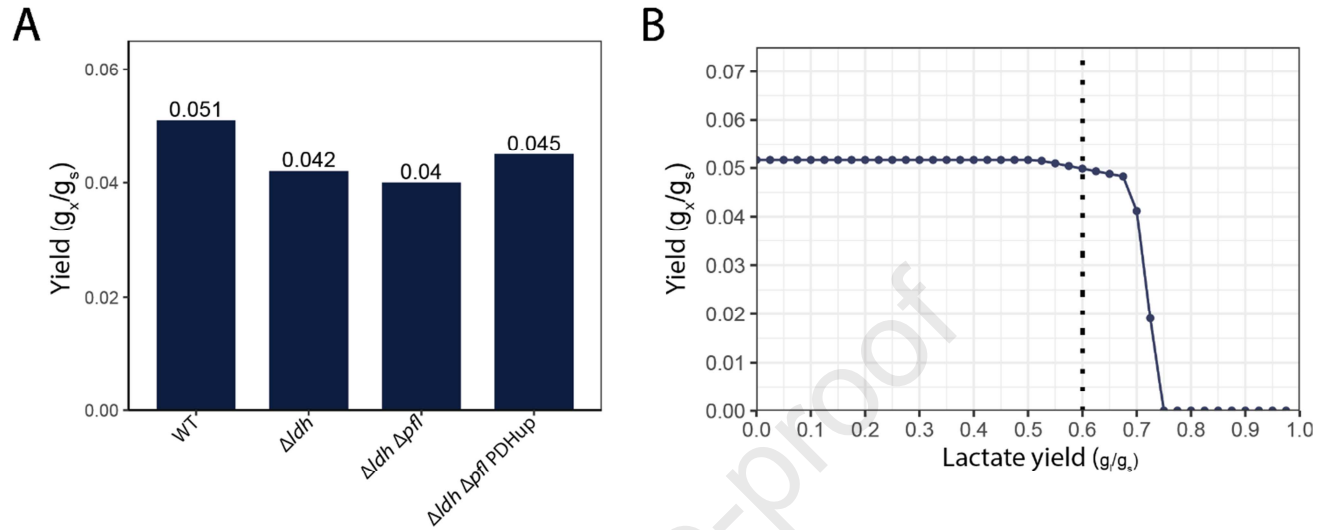
1006 **Supplementary figure 3:** Predicted oxygen consumption rates for the three conditions,
1007 when measured exchange rates of fermentation products were fit to the model.



1008

1009

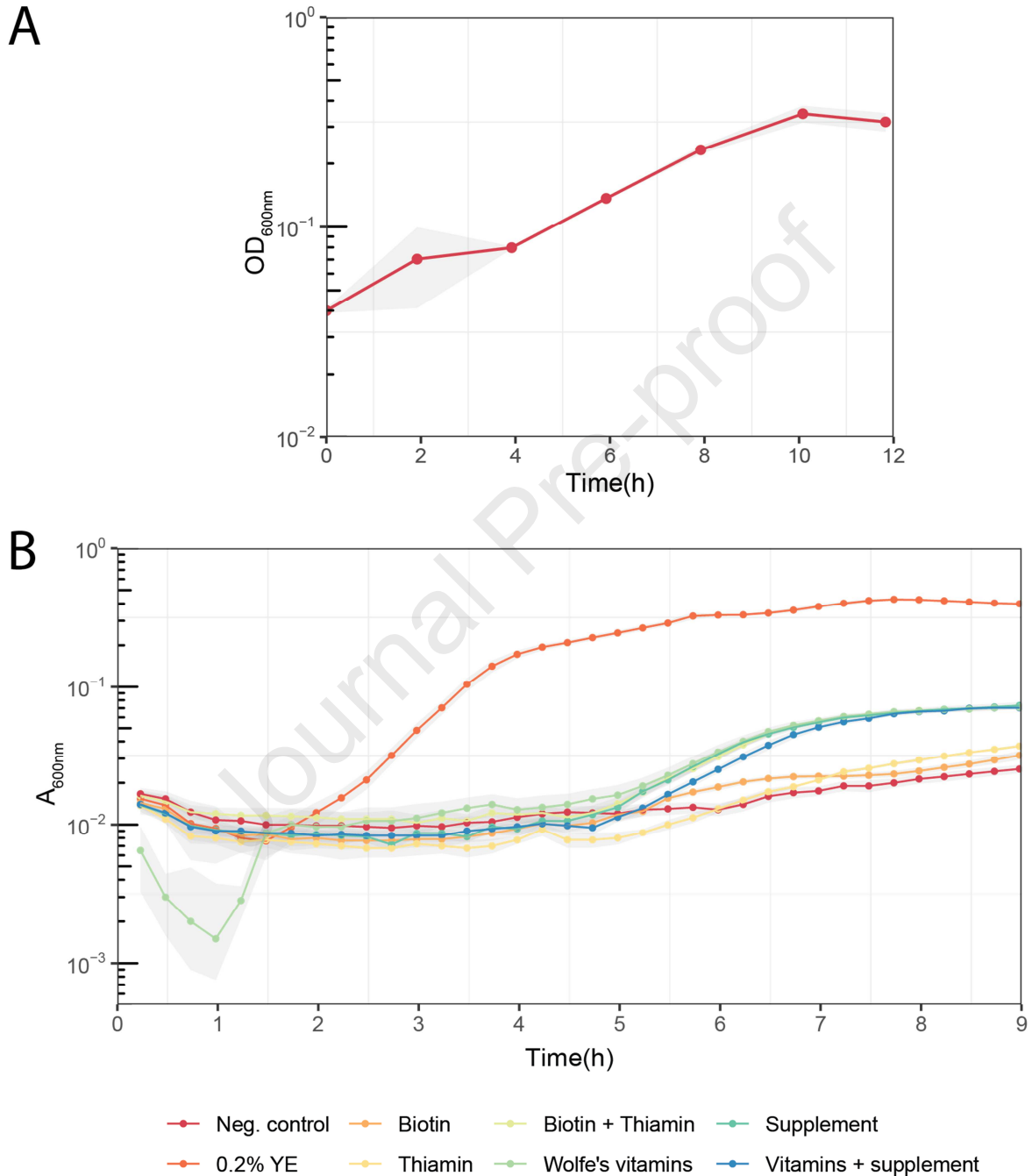
1010 **Supplementary figure 4:** A) result of computing predicted *in silico* biomass yield, when
1011 measured exchange rates, carbon uptake rates and genetic manipulations (i.e. knockouts)
1012 have been introduced. B) Effect of lactate production on biomass yield when all other
1013 measured exchange rates are fixed for the WT strain. The dotted line indicates the
1014 measured lactate production in these conditions.



1015

1016

1017 **Supplementary figure 5:** A) Aerobic shake flask experiment of *P. thermoglucosidasius*
 1018 NCIMB 11955 on TMM base medium. Shaded area shows standard deviation between three
 1019 biological replicates. B) Anaerobic growth curves of *P. thermoglucosidasius* NCIMB 11955
 1020 grown on base TMM supplemented with various nutrients. Experiment was performed in a
 1021 microtiter plate reader, and shaded area represents standard deviation of measurements
 1022 over quadruplicates. (YE = yeast extract)



1023

1024

1025

1026

1027

1028 Supplementary Tables

1029 **Supplementary Table 1:** Stoichiometry of the biomass reaction in *p-thermo*.

	Compound	mmol/gDCW
Proteins	Ala	0.5142
	Arg	0.1831
	Asp	0.2219
	Asn	0.2219
	Cys	0.1079
	Glu	0.322
	Gln	0.322
	Gly	0.4077
	His	0.0778
	Ile	0.2728
	Leu	0.3475
	Lys	0.3172
	Met	0.0832
	Phe	0.1451
	Pro	0.1665
	Ser	0.1811
	Thr	0.2688
	Trp	0.1026
	Tyr	0.1036
Val	0.3761	
Nucleic Acids	<i>Ribonucleic acid (RNA)</i>	
	AMP	0.1193
	CMP	0.0915
	GMP	0.0915
	UMP	0.1193
	<i>Deoxyribonucleic acid (DNA)</i>	
	dAMP	0.0077
	dCMP	0.0059
	dGMP	0.0059
	dTMP	0.0077
Lipids	<i>Phospholipids</i>	
	PE	0.0290
	PG	0.0277
	CLPN	0.0296
Carbohydrates	D-Fructose	0.1048
	GDP-Mannose	0.0057
	UDP-D-Galactose	0.1895
	UDP-D-Xylose	0.1825
	UDP-D-Glucose	0.0096
	UDP-L-Arabinose	0.0407

	Compound	mmol/gDCW
Salts	Phosphorous	0.0420
	Calcium	0.0028
	Potassium	0.6323
	Magnesium	0.0875
	Iron	0.0304
Vitamins & cofactors	Thiamin B1	0.0002
	Riboflavin B2	0.0002
	Vitamin B12	0.0002
	Vitamin B6	0.0002
	10-Formyltetrahydrofolate	0.0004
	Biotin	1.79E-06
	Bacillithiol	6.23E-05
	Bacillithiol disulfide	1.56E-07
	Chorismate	0.0002
	FAD	0.0002
	FMN	0.0002
	Heme	0.0002
	Proto-heme	0.0002
	Siroheme	0.0002
	Menaquinol	0.0003
	NAD ⁺	0.0145
	NADH	0.0267
	NADP	0.0042
	NADPH	0.0027
	S-Adenosyl-L-methionine	0.0002
	Spermidine	0.0011
	Spermine	0.0025
	Putrescine	0.0054
CoA	0.0002	
Intracellular Metabolites	Acetyl-CoA	0.0002
	Succinyl-CoA	8.75E-05
Energy Requirement	ADP	- 152.28
	Pi	-152.27
	Ppi	0.0011
	H	- 152.28
	H ₂ O	152.28
	ATP	152.31

1030

1031

1032 **Supplementary table 2:** Estimation of polymerization energy needed to form biomass from
 1033 the different monomer classes present in the biomass reaction. This energy fraction
 1034 constitutes part of the growth associated maintenance that is present in the biomass
 1035 reaction. Polymerization energy per molecule was obtained from literature⁶⁵.

	Cellular content (w/w)	Polymerization energy	
		Per molecule (mmATP/g polymer)	Total (mmol ATP/gcell)
Protein	0.52	37.7	19.604
Carbohydrates	0.1	12.8	1.28
RNA	0.16	26	4.16
DNA	0.1	26	2.6
Lipids	0.09	25.6	2.304
Sum			29.948

1036

1037 **Supplementary table 3:** Overview of an analyses of filtering the unique ORFs detected in
 1038 the genome analyses between two (*Para*)*geobacillus* strains, when various levels of filtering
 1039 are applied to elucidate how many reactions would be unique in models made of each strain,
 1040 and finally which would be connected to any pre-existing metabolites in the network.
 1041 Supplementary table 4 and 5 highlight the unique metabolic ORFs identified.

1042

	<i>P. thermoglucosidasius</i> NCIMB 11955	<i>P. thermoglucosiodasius</i> M10EXG
Unique ORFs	259	234
Unique metabolic ORFs ¹	40	29
Unique ORFs as reactions ²	18	13
Unique reactions ³	11	12
Connected unique reactions ⁴	8	6

1043

1044 ¹ Unique metabolic ORFs are ORFs with an E.C. code associated to them

1045 ² Subset of unique metabolic ORFs that would actually be captured as a reaction in the model

1046 ³ Subset of unique ORFs that would cause new reactions to the model

1047 ⁴ Subset of unique reactions that are connected by a main metabolite to one or more pre-existing
 1048 metabolites in the model. Note, all these reactions are still blocked (i.e. no two main metabolites found
 1049 in model).

1050

1051

1052 **Supplementary table 4:** Metabolic ORFs unique to *P. thermoglucosidasius* NCIMB 11955,
 1053 detected in the genome comparison.

1054

Gene annotation	EC	Annotated Kegg Ontology
abfA	3.2.1.55	Alpha-L-arabinofuranosidase
adk	2.7.4.3	adenylate kinase
araA	5.3.1.4	L-arabinose isomerase
cocE	3.1.1.84	Carboxylesterase
ecfA2	3.6.3.-	Hydrolase, involved in transmembrane movement
glf	5.4.99.9	UDP-galactopyranose mutase
gltX	6.1.1.17	Glutamyl-tRNA synthetase
hdl IVa	3.8.1.2	2-haloacid dehalogenase
helD_2	3.6.4.12	DNA helicase
ispD	2.7.7.60	2-C-methyl-D-erythritol 4-phosphate cytidyltransferase
ispF	4.6.1.12	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase
kpsU	2.7.7.38	3-deoxy-manno-octulosonate cytidyltransferase (CMP-KDO synthetase)
lytC_3	3.5.1.28	N-acetylmuramoyl-L-alanine amidase
mazF	3.1.27.-	Esterase (component of type II toxin-antitoxin system)
mcsB	2.7.14.1	Protein arginine kinase
mngB	3.2.1.170	Mannosylglycerate hydrolase
mrnC	3.1.26.-	Esterase (ribonuclease)
mtID	1.1.1.17	Mannitol-1-phosphate 5-dehydrogenase
ppk	2.7.4.1	Polyphosphate kinase
ppx	3.6.1.11	Exopolyphosphatase
pseG	3.6.1.57	UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase
rapA_1	3.6.4.-	RNA Polymerase associated hydrolase
rbsD	5.4.99.62	D-ribose pyranase
rhpA	3.6.4.13	ATP-dependent helicase
rpoA	2.7.7.6	DNA-directed RNA polymerase
rpoB	2.7.7.6	DNA-directed RNA polymerase
rpoC	2.7.7.6	DNA-directed RNA polymerase
srlB	2.7.1.198	Glucitol/sorbitol PTS system EIIA component
srlE	2.7.1.198	Glucitol/sorbitol PTS system EIIA component
tagD	2.7.7.39	Glycerol-3-phosphate cytidyltransferase
trmB	2.1.1.33	tRNA (guanine-N7-)-methyltransferase
truA	5.4.99.12	tRNA pseudouridine synthase
wecA	2.7.8.40	UDP-GalNAc:undecaprenyl-phosphate GalNAc-1-phosphate transferase
xylA_1	3.2.1.37	Xylan 1,4-beta-xylosidase
xynA	3.2.1.8	Endo-1,4-beta-xylanase
xynB	3.2.1.37	Xylan 1,4-beta-xylosidase
YDAf_2	2.3.1.-	Acyltransferase
yeeF	3.1.-.-	Ribonuclease
yjjG	3.1.3.5	5'-nucleotidase
yobL	3.1.-.-	Ribonuclease

1055

1056

1057 **Supplementary table 5:** Metabolic ORFs unique to *P. thermoglucosidasius* M10EXG,
 1058 detected in the genome comparison. Interestingly, it appears the M10EXG strain encodes a
 1059 complete myo-inositol utilizing operon (*iol*), where some components are lacking in NCIMB
 1060 11955¹⁰³. Data shows NCIMB 11955 indeed lacks the capability to grow on inositol⁵¹, but
 1061 data on M10EXG is lacking.

Gene annotation	EC	Annotated Kegg Ontology
cocE	3.1.1.84	Carboxylesterase
cwlK	3.4.-.-	Peptidase
cynS	4.2.1.104	Cyanate lyase
derK	2.7.1.210	D-erythrulose 4-kinase
dhaL	2.7.-.-	Phosphotransferase
fdtA	5.3.2.3	TDP-4-oxo-6-deoxy-alpha-D-glucose-3,4-oxoisomerase
hsdM_1	2.1.1.72	Site-specific DNA-methyltransferase (adenine-specific)
hsdM_2	2.1.1.72	Site-specific DNA-methyltransferase (adenine-specific)
hsdR_1	3.1.21.3	Type I restriction enzyme
hsdR_2	3.1.21.3	Type I restriction enzyme
iolB	5.3.1.30	5-deoxy-glucuronate isomerase
iolE	4.2.1.44	Inosose dehydratase
iolI	5.3.99.11	2-keto-myo-inositol isomerase
iolX	1.1.1.370	Scyllo-inositol 2-dehydrogenase (NAD+)
lhgO	1.1.3.15	Glycolate oxidase
lipA	2.8.1.8	Lipoyl synthase
lsrF	2.3.1.245	3-hydroxy-5-phosphonooxypentane-2,4-dione thiolase
mutT4	3.6.1.-	Esterase
NA	2.4.1.161	Oligosaccharide 4-alpha-D-glucosyltransferase
radD	3.6.4.12	DNA repair helicase
recD	3.1.11.5	Exodeoxyribonuclease V
rfbC	5.1.3.13	dTDP-4-dehydrorhamnose 3,5-epimerase
sqhC	4.2.1.137	Sporulenol synthase
sunS	2.4.1.-	Glycosyltransferase
tatD	3.1.21.-	Endodeoxyribonucleases
uvrD1	3.6.4.12	ATP-dependent DNA helicase
wapA_1	3.1.-.-	tRNA(Glu)-specific nuclease
wapA_2	3.1.-.-	tRNA(Glu)-specific nuclease
wapA_3	3.1.-.-	tRNA(Glu)-specific nuclease

1062