Flux modelling analysis reveals the metabolic impact of cryptic plasmids and environmental conditions in probiotic *Escherichia coli* Nissle 1917

Paola Corbín-Agustí[1*], Alba Arévalo-Lalanne[1], Patricia Álvarez[1], Daniel Ramón[2], Juli Peretó[1,3], Marta Tortajada[4*]

[1]Institute for Integrative Systems Biology, I[2]SysBio (CSIC-Universitat de València), Paterna, Spain.

[2]Department of Animal Production and Health, Faculty of Veterinary Sciences, Universidad Cardenal Herrera, Centro de Estudios Universitarios, UCH-CEU, Alfara del Patriarca, Spain.

[3]Department of Biochemistry and Molecular Biology, Universitat de València, Burjassot, Spain.

[4]Department of Biomedical Sciences, Faculty of Health Sciences, Universidad Cardenal Herrera, Centro de Estudios Universitarios, UCH-CEU, Alfara del Patriarca, Spain

*Corresponding authors: Marta Tortajada marta.tortajadaserra@uchceu.es, Paola Corbín Agustí paola.corbin@uv.es

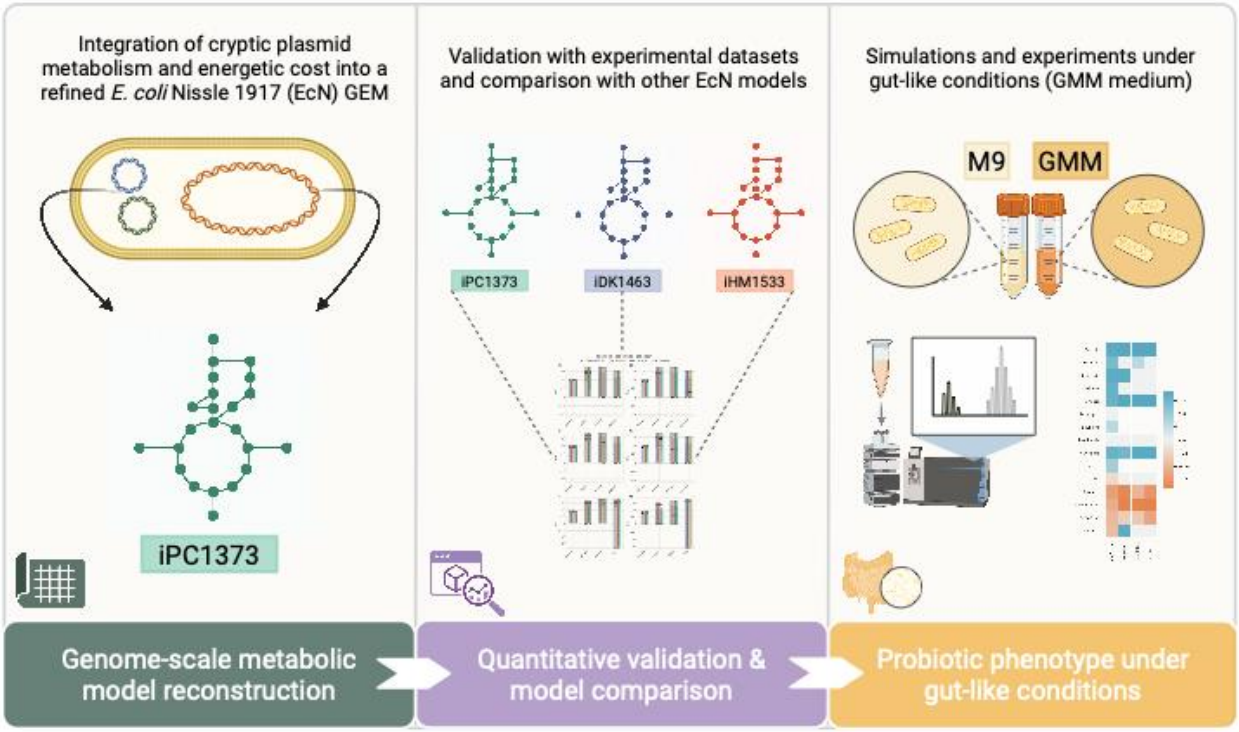
Abstract

Escherichia coli Nissle 1917 (EcN) is a well-characterized Gram-negative probiotic distinguished by its unique, strain-specific physiology. Genome-scale metabolic models (GEMs) are powerful tools for elucidating metabolic traits and predicting genotype–phenotype relationships. Although several EcN GEMs have been published, none have explicitly represented its probiotic physiology. Here, we present a manually curated GEM of EcN that, for the first time, incorporates the energetic costs associated with its cryptic plasmids. Inclusion of a plasmid-specific module improved biomass yield predictions and overall model accuracy, providing a more physiologically realistic representation of EcN metabolism. Using COBRA methodologies and possibilistic metabolic flux analysis, this model and previous EcN reconstructions were systematically compared to evaluate the trade-off between model complexity and predictive performance. The analysis revealed that increased structural detail does not necessarily enhance quantitative accuracy and that predictive reliability depends on both computational methodology and model context. Metabolomic profiling under gut-like anaerobic conditions further showed that EcN exhibits a distinctive metabolic phenotype, characterized by elevated amino acid consumption and enhanced short-chain fatty acid production. These findings highlight the unique probiotic physiology of EcN and demonstrate the utility of metabolic modeling for reproducing and exploring such traits. Overall, this study provides a quantitatively reliable and physiologically relevant framework for modeling *E. coli* Nissle 1917 and related commensal bacteria, supporting advances in probiotic engineering, synthetic biology, and bioprocess design.

Keywords: *Escherichia coli* Nissle 1917, genome-scale metabolic modeling, possibilistic metabolic flux analysis, probiotics, plasmid burden

Graphical Abstract Summary

This study presents a manually curated genome-scale model of *Escherichia coli* Nissle 1917 that accounts for the metabolic cost of its cryptic plasmids. Through systematic comparison with previous reconstructions and validation against fluxomics datasets, the models improved accuracy in predicting growth and fluxes. Simulations and experiments under gut-like conditions provide new insights into EcN's unique probiotic traits.



1. Introduction

Escherichia coli Nissle 1917 (EcN) is a well-characterized probiotic strain originally isolated by Alfred Nissle in 1917 from the faeces of a First World War soldier who appeared resistant to dysentery (Beimfohr, 2016). Since its discovery, EcN has been extensively used to treat gastrointestinal disorders such as infectious diarrhea, inflammatory bowel diseases (IBD), ulcerative colitis or Crohn's disease (Petersen *et al.*, 2011; Olier *et al.*, 2012; Maltby *et al.*, 2013). Remarkably, EcN remains the only Gram-negative strain in routine probiotic use, despite belonging to a species typically associated with pathogenic lineages.

EcN exhibits several unique, strain-specific features that underpin its probiotic efficacy. These include (i) high fitness in the intestinal environment and stable colonization through strong adhesion to the epithelium, biofilm formation, and antimicrobial resistance; (ii) competitive exclusion of enteric pathogens via production of antimicrobials such as microcins, interference with virulence mechanisms, and sequestration of key nutrients such as iron; and (iii) reinforcement of host mucosal immunity by activating TLR and NF-κB signaling, enhancing epithelial barrier function, and stimulating antimicrobial peptide production (Deriu *et al.*, 2013; Schumann *et al.*, 2014; Hare *et al.*, 2022). EcN also lacks typical virulence factors such as α-hemolysin and fimbriae and displays a semi-rough lipopolysaccharide, which may contribute to immune evasion, attenuated inflammatory response, and improved mucosal adhesion (Deriu *et al.*, 2013; Schumann *et al.*, 2014; Hare *et al.*, 2022).

Genome-scale metabolic models (GEMs) represent the metabolic capabilities of an organism through the stoichiometric formulation of its biochemical reactions. Over the past two decades, GEMs have become indispensable powerful and robust computational tools in systems biology for elucidating microbial metabolism, predicting genotype–phenotype relationships, functional characteristics, and optimizing industrial bioprocesses (Monk *et al.*, 2013, 2016, 2017; Pearcy *et al.*, 2021; Bernstein *et al.*, 2023). Due to the large number of reactions and metabolic

uncertainties, GEMs are inherently underdetermined and require constraint-based approaches—collectively known as COBRA (Constraint-Based Reconstruction and Analysis)—to predict metabolic flux distributions. Among the most widely applied methods are metabolic flux analysis (MFA), which integrates experimental flux measurements to estimate flux rates, and flux balance analysis (FBA), which uses stoichiometric constraints and optimization principles to infer metabolic states under an assumed cellular objective (Orth *et al.,* 2010; Antoniewicz, 2021). Complementary to these approaches, possibilistic MFA provides a framework for handling uncertainty by combining constraint-based models with measurement error distributions represented as possibility functions. This method enables robust flux estimation even when experimental data are incomplete, using linear programming to identify feasible flux ranges (Llaneras *et al.*, 2009; Llaneras, 2011).

Recent GEM-based studies have explored the metabolic capabilities of EcN. These models have been used to benchmark EcN against laboratory or pathogenic *E. coli* strains, identify distinctive metabolic features and explore its potential as a chassis for biosynthetic applications (Kim *et al.*, 2021; Moutinho *et al.*, 2022; van't Hof *et al.*, 2022).

EcN harbors two cryptic plasmids, pMUT1 (3.2 kb) and pMUT2 (5.5 kb), both possessing ColE1-type replication origins and maintained at high copy numbers (more than 100 copies per cell) (Zainuddin *et al.*, 2019). These plasmids are stably inherited and commonly used for strain typing (Blum-Oehler *et al.*, 2003). Although historically considered phenotypically neutral, recent transcriptomic evidence indicates that deletion of pMUT1 and pMUT2 substantially alters the expression of genes involved in cell aggregation and amino acid metabolism, suggesting possible contributions to EcN's probiotic function (Lin *et al.*, 2024). Additionally, their stable maintenance and replication machinery have been exploited as antibiotic-free expression systems for targeted intestinal delivery of heterologous proteins (Zhou *et al.*, 2024).

However, plasmid carriage often imposes a significant metabolic burden on host cells due to the energetic costs of replication and gene expression. This phenomenon is well documented in *E. coli*, particularly in recombinant protein production systems (Rozkov *et al.*, 2004; Zeng and Yang, 2019). The additional biosynthetic demand typically leads to reduced growth rate and yield compared with plasmid-free strains and is frequently accompanied by metabolic overflow, such as acetate accumulation. Even strains harboring empty, high-copy-number ColE1-derived plasmids exhibit globally altered physiological states and increased non-growth-associated maintenance energy (Ow *et al.*, 2009; Oftadeh and Hatzimanikatis, 2024). Several modeling studies have successfully incorporated plasmid-related metabolic costs into constraint-based metabolic models, including GEMs, improving their ability to reproduce experimental data (Özkan *et al.*, 2005; Ow *et al.*, 2009; Zeng and Yang, 2019; Oftadeh and Hatzimanikatis, 2024). Nevertheless, the explicit consideration of plasmid burden in probiotic EcN models remains limited.

This study presents a manually curated genome-scale metabolic model (GEM) of EcN that explicitly accounts for the energetic costs of its cryptic plasmids. Using COBRA methods and possibilistic MFA, this model and previously published EcN GEMs were systematically compared and evaluated, including prediction of biomass formation and by-product external rates, as well as their consistency with experimental metabolomics and ^{13}C-fluxomics datasets. This comparative framework enabled the assessment of how different models predict metabolic behavior under specific experimental and methodological contexts. Importantly, the experimental metabolomics data used were obtained under conditions that mimic those of the probiotic's native environment, an approach that, to our knowledge, has not been previously applied to this strain. The results provide a mechanistic framework for better understanding the metabolic determinants of EcN's probiotic properties, and a more accurate characterization of its probiotic capacity and model predictions within that physiological context.

123

124 ## 2. Materials and Methods

125 **Bacterial strains and culture conditions**

126 Two strains of *Escherichia coli* were used in this work: *E. coli* Nissle 1917 (DSM 6601) and *E. coli*
127 K12-HB101. EcN strain was isolated from Mutaflor commercial product (Ardeypharm) and grown
128 in Lysogeny Broth (LB) at 37ºC unless stated otherwise. EcN strain was routinely identified by
129 PCR amplification in cryptic plasmid pMUT2, as described in (Blum-Oehler *et al.*, 2003). *E. coli*
130 K12-HB101 strain was purchased from Promega (Madison, WI, USA). Growth media used
131 throughout experiments include M9 minimal medium, with glucose (Monk *et al*., 2016a) and
132 supplemented with casamino acids to fulfill the auxotrophic requirements of K12-HB101 strain,
133 and rich-gut microbiota medium (GMM), prepared according to Goodman *et al.* (2011) and
134 Średnicka *et al.* (2023) without agar. Composition details are provided in the Supplementary
135 Methods.

136

137 **Draft model construction and refinement**

138 The genome sequence of EcN strain was retrieved from NCBI (RefSeq accession number:
139 NZ_CP007799.1; annotated Feb 23, 2017) and used to compare with other reference *E. coli*
140 strains. For that, the protein sequences of K-12 MG1655 (NC_000913) and CFT073
141 (NC_004431.1) were used to identify the ortholog genes between EcN and these strains by a
142 reciprocal blast best hit method, with an e-value > $1x10^6$ and a percentage of homology greater
143 than 85%. In parallel, the curated models of E. coli K-12 MG1655 (iAF1260, iJO1366 and
144 iML1515) and E. coli CFT073 (ic_1306) were downloaded from the BIGG database (King *et al.*,
145 2016) in JSON format. A first EcN draft model was reconstructed incorporating all shared
146 reactions and metabolites from iAF1260 and ic_1306 metabolic models, dependent on EcN
147 homologous genes. Subsequently, the non-homologous genes of *E. coli* EcN were reviewed and

148 the associated reactions, retrieved from BiGG when possible, were added to the draft model.
149 MetaCyc (Caspi *et al.*, 2020), KEGG (Kanehisa *et al.*, 2021) and BRENDA (Schomburg *et al.*,
150 2013) databases were also consulted for the manual model curation. The biomass reaction used
151 in this EcN model corresponds to that of the iJO1366 model. Consistency analysis of the
152 topological network was conducted to identify gap metabolites and blocked reactions (Ponce-de-
153 Leon *et al.*, 2015). Finally, gaps, reaction directionality and redundancies were manually curated.
154 Reactions, metabolites, genes, and subsystems were annotated using the KEGG, BioCyc,
155 MetaNetX, SEED, Rhea, ChEBI, and NCBI databases. SBOannotator (Leonidou *et al.*, 2023) was
156 used to assign SBO terms. The quality of the model was evaluated using the MEMOTE tool
157 (v.0.13.0) (Lieven *et al.*, 2020). For the manipulation, study and development of the model Python
158 3 and the COBRApy library were used (Ebrahim *et al.*, 2013). The final GEM is provided in json,
159 mat and sbml format (Supplemental Material S1).
160
161 **Metabolic models comparison**
162 The EcN model reconstructed here was compared to the remaining available models (iDK1463
163 (Kim *et al.*, 2021) and iHM1533 (van't Hof *et al.*, 2022) in terms of structure, Memote scores, and
164 predictive analysis.
165
166 **Plasmid integration in the EcN model**
167 To incorporate the biosynthesis of the two cryptic plasmids, present in EcN into the model, two
168 reactions were introduced: one representing the consumption of deoxyribonucleotides (dNTPs)
169 and the production of the plasmids (Eq. 1), and a demand reaction (Eq. 2) to enable simulations.
170 The stoichiometric coefficients of dNTPs and pyrophosphate were calculated using the BOFdat
171 algorithm (Lachance *et al.*, 2019), based on each plasmid's nucleotide composition and its
172 estimated copy numbers (Zainuddin *et al.*, 2019). The contribution of the plasmids to the cellular

173 dry weight was computed as the DNA fraction, considering the product of plasmids size and copy
174 number relative to the chromosome genome size (Eq. 3).

175 0.0133 datp_c + 0.0125 dctp_c + 0.011 dgtp_c + 0.0137 dttp_c --> plasmid_c + 0.0504 ppi_c (Eq. 1)

176 plasmid_c --> (Eq. 2)

177 ((plasmid bp × copy number) / (chromosome bp + plasmid bp × copy number)) × 0.031 (Eq. 3)

178 To simulate plasmid replication and its associated metabolic burden, we imposed a fixed lower
179 bound on the plasmid, while keeping biomass maximization as the objective function. To select
180 this fixed value, the strategy was evaluated across a range of NGAM lower bound values,
181 considering which would best reproduce the experimental conditions (see belowe *In silico growth*
182 *conditions* and *Experimental datasets*) and to assess sensitivity to NGAM. Further details and
183 simulations are provided in the Supplementary Methods.

184

185 **ATP expenditure parameters**

186 Growth associated ATP maintenance (GAM) and non-growth associated ATP maintenance
187 (NGAM) are critical parameters for GEMs to accurately represent physiological responses. GAM
188 and NGAM values for wild-type and plasmid-bearing *E. coli* were retrieved from literature. A
189 consensus value of 59.8 $mmol \cdot gDW^{-1} \cdot h^{-1}$ for GAM was used (Kim *et al.*, 2021). The occurrence
190 of plasmids has been referred to increase NGAM energy due to the metabolic burden associated
191 with their maintenance, replication, and expression (Rozkov *et al.*, 2004; Ow *et al.*, 2009; Zeng
192 and Yang, 2019). The median of previously reported NGAM values was used to set the lower
193 bound of the ATPM reaction in aerobic (25 $mmol \cdot gDW^{-1} \cdot h^{-1}$) and anaerobic conditions (3.5
194 $mmol \cdot gDW^{-1} \cdot h^{-1}$). Further details can be found in Supplementary Methods.

195

196 **Experimental datasets employed for model validation**

197 Experimental data were compiled from literature on the growth of EcN on various carbon sources
198 (Revelles *et al.*, 2013; Millard *et al.*, 2021) and used for model validation. Each dataset contains

199 experimental measurements of several external fluxes, including specific growth rate, carbon
200 source uptake rate, and secretion rates of byproducts. These experimental measurements are
201 provided in Supplementary Methods.
202
203 ***In silico* constraints and growth conditions**
204 For reversible reactions, upper and lower bounds of 1000 and -1000 mmol·gDW^{-1}·h^{-1}, were
205 respectively applied. For irreversible reactions, the lower bound –or in specific cases, the upper
206 bound– was fixed at 0 mmol·gDW^{-1}·h^{-1}. Maximization of growth rate was set as the objective
207 function. Anaerobic conditions were imposed by constraining the lower bound of the O_2 exchange
208 reaction to 0 mmol·gDW^{-1}·h^{-1}; whereas aerobic conditions were simulated by setting this bound
209 to -20 mmol·gDW^{-1}·h^{-1} (Nogales, 2017; Blázquez *et al.*, 2023).
210 To prevent unrealistic behaviors under some conditions, certain reactions were deactivated and
211 constrained to 0 mmol·gDW^{-1}·h^{-1} as stated in reference *E. coli* models. Under all conditions, the
212 following reactions were blocked: CAT, DHPTDNR, DHPTDNRN, SPODM, SPODMpp,
213 SUCASPtpp, SUCFUMtpp, SUCMALtpp, and SUCTARTtpp, as well as PTRCORNt7pp,
214 GALCTNLt2pp, GALCTLO, EX_chitob_e, URCN, CHITPH and CHITOBpts. The following
215 reactions were kept active only in anaerobiosis: DTARTD, LCARS, SUCTARTtpp, TARTD,
216 TARTRt7pp, ARGAGMt7pp, FHL and CELBpts (Monk *et al.*, 2016). Although FHL is active
217 anaerobically, it was also blocked here to allow quantification of formate production that would
218 otherwise be fully consumed by the return reaction.
219 NGAM and GAM energy requirements were constrained according to the values specified in the
220 previous *ATP expenditure parameters* section. For the simulation of each growth media, the lower
221 bounds of exchange reactions were set to zero except for the metabolites present in the defined
222 medium (M9 minimal media or gut microbiota media, GMM). Detailed specific exchange
223 restrictions of each media are provided in Supplementary Methods. M9 minimal media was
224 employed when simulating the experimental datasets, subject to additional condition-specific

225 adjustments. A detailed summary of the constrained reactions and fluxes applied to in *silico*
226 simulations of these datasets is provided in Supplementary Methods.
227
228 **Flux Balance and Flux Variability Analyses**
229 Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA) were performed using the
230 COBRApy software (Ebrahim *et al.*, 2013) All optimization problems were solved with CPLEX and
231 the GNU Linear Programming Kit (GLPK). FBA was used to compute the maximum value of the
232 objective function and the corresponding flux distribution (Thiele and Palsson, 2010; Orth *et al.*,
233 2011). Maximization of the growth rate was set as the default cellular objective function. For
234 modelling and simulating growth specific conditions, bounds of some reactions were modified as
235 previously specified in *in silico constraints and growth conditions* section. Subsequently, FVA was
236 carried out to determine the minimum and maximum feasible fluxes for each reaction on the
237 constrained metabolic model while maintaining at least 90% of the optimal FBA objective value.
238
239 **Possibilistic Metabolic Flux and Consistency Analysis**
240 Possibilistic Metabolic Flux Analysis (MFA) was applied as previously described in (Tortajada *et*
241 *al.*, 2010; Llaneras, 2011; Morales *et al.*, 2014), using the PFAToolbox implemented in MATLAB
242 R2020b. Briefly, this methodology accounts for data uncertainty, allowing experimental
243 measurements to deviate from actual flux values, and can be represented as: wm = vm + em,
244 where em is a vector containing the measurement errors. Each candidate solution fulfilling the
245 model and experimental constraints is denoted as δ. A possibility value [0, 1] is assigned to each
246 solution based on a linear cost index $\mathbf{J}(\delta)=\alpha\cdot\varepsilon 1+\beta\cdot\mu 1$, where the possibility of each solution will
247 be $\pi(\delta)=\exp(-\mathbf{J}(\delta)\delta\subset\Delta$. Thus, a situation where vm = w is considered fully possible, and the
248 greater the deviation between vm and wm, the lower the associated possibility.
249 In order to assess consistency, the agreement between experimental datasets and the metabolic
250 network was evaluated as described in (Tortajada *et al.*, 2010; Llaneras, 2011). Briefly, this

251 method assigns a possibility value π^{mp} = [0, 1] to consistency between the most possible flux
252 distribution with the model constraints and the experimental data under the model constraints. A
253 value of 1 indicates full consistency, whereas lower values indicate discrepancies between
254 measurements and model. For each dataset, measurement uncertainty was defined so that fluxes
255 were considered fully possible ($\pi = 1$) within 5% of their reported standard deviations, while
256 deviations exceeding the larger of either 20% of the experimental mean value or two standard
257 deviations from experimental standard deviations corresponded to a possibility of $\pi \leq 0.1$.
258

259 **Flux-associated error estimation**

260 To quantify the agreement between experimental measurements and model-predicted fluxes, the
261 Relative Squared Error (RSE) was computed. For each condition and model, individual flux RSE
262 values were calculated and then summarized to obtain the median error estimate. These values
263 were used to compare the predictive performance of the different EcN GEMs across experimental
264 datasets.
265

266 **Metabolomic assays and data analysis**

267 Starting from a 10 mL overnight culture in LB medium for each strain, the cells were washed twice
268 with saline solution and inoculated into the corresponding medium at OD 0.05. Cultures were
269 incubated overnight at 37°C under anaerobic conditions (BD GasPak™ system). After incubation,
270 the cultures were collected and centrifuged for 15 minutes at 4500*g*. The supernatants were
271 filtered through a 0.22 µm membrane filter and stored at -80°C until HPLC analysis. For each
272 growth medium (GMM and M9), three replicates of both strains were performed, across two
273 independent experimental setups. Blanks were conducted as medium samples, non-inoculated,
274 independently.
275 The determinations of the metabolites listed in Supplementary Methods were carried out by high-
276 performance liquid chromatography in an Alliance 2695 HPLC equipment coupled to a refractive

277 index detector. The column used was an Aminex HPX-87H from Bio-Rad (Madrid, Spain, 300 mm
278 × 7.8 mm) at a temperature of 60°C. Isocratic elution was performed with 5 mM H_2SO_4 at a flow
279 rate of 0.6 mL min^{-1}. The identifications were performed by comparison with the retention time of
280 the standards and calibration curves were used for the quantifications.
281 For data analysis and comparison, blanks measurements were extracted for each metabolite
282 value, depending on medium. Metabolite concentrations (in mg/L) were normalized to cell growth
283 in terms of OD. This processed data was used in the subsequent analysis. For visualization of
284 each metabolite profiling, ggplot2 (v3.4.4; Wickham, 2009) was employed. Statistical analysis was
285 performed in python (v3.11.5) using a bootstrap resampling with 10000 iterations and a
286 significance level of 0.05 to evaluate differences between control and condition samples. A
287 threshold of 10^{-3} mg/L was applied to filter out low-abundance metabolites. For each metabolite
288 and condition, 95% bootstrap-based confidence intervals were computed to assess the
289 significance of differences in metabolite concentrations relative to control values.
290
291 **Code availability**
292 The code and data for reproducing this analysis are available in the Github repository:
293 https://github.com/paocorbin/EcN-GEM
294

295 # 3. Results

296 ## 3.1 Reconstruction of the EcN and structural comparison with other EcN
297 GEMs

298 ### 3.1.1 Previously published GEMs of EcN

299 Several researchers have developed genome-scale metabolic models (GEMs) specific to EcN.
300 The initial step in this research involved comparing the available reconstructions (Table 1) to

evaluate how their structural features influence predictive accuracy and performance in practical experimental scenarios.

The first GEM of EcN (iDK1463; Kim *et al.*, 2021) was derived from earlier reconstructions of *E. coli* K-12 MG1655 and two uropathogenic strains, CFT073 and ABU83972. This model was qualitatively validated using Biolog phenotypic assays, covering a broad range of carbon and nitrogen sources, and other substrates under aerobic conditions. The analysis identified gene clusters unique to EcN relative to the K-12 strain and used the metabolic model to simulate anaerobic growth in intestinal environments with various carbon sources. These simulations revealed distinct metabolic behaviors between EcN and K-12 strains.

Subsequently, Moutinho and coworkers (2022) reconstructed a smaller, more streamlined network to demonstrate their proposed reconstruction workflow, called Constraint-based Analysis Yielding reaction Usage across metabolic Networks (CANYUN). This model was validated against experimental growth data under both aerobic and anaerobic conditions.

Finally, van't Hof and coworkers (2022) developed an additional model (iHM1533), built on a high-quality genomic sequence and related to 55 other *E. coli* strains with available GEMs. This model was validated through phenome analysis and ^{13}C-fluxomics data, which provided insights into internal fluxes under aerobic conditions (Revelles *et al.*, 2013). They also refined the pathways involved in secondary metabolism to create a framework for using EcN as a synthetic biology chassis, particularly for enhancing secondary metabolite production, such as antimicrobial compounds. For this, a mutant EcN strain curated from the cryptic plasmids was used for further validation during Biolog phenotypic assays.

An analysis of the main features of these EcN models shows an average of approximately 3,064 reactions (excluding the Canyun model) and around 1,471 unique metabolites. The Canyun model differs from the others because its primary goal was to simulate qualitative growth phenotypes and validate the reconstruction workflow. As a result, it omits over 1,200 reactions that are not essential for growth. Overall, the published EcN GEMs vary in complexity and the

327 level of manual curation, which depends on the reconstruction strategy—whether based on
328 extensive reference GEMs or utilizing new reconstruction methods—and the degree of refinement
329 performed. The models listed in Table 1 are ranked from least to most complex based on these
330 criteria. Nevertheless, all reconstructions have been validated using similar experimental
331 datasets.
332

333 **3.1.2 Development of the EcN model presented in this study**

334 In this study, we present an updated GEM of EcN, designated iPC1373 (Supplementary Material
335 S1; Alvarez *et al.*, 2019). The model was developed to accurately represent the core physiology
336 and metabolic capabilities of this probiotic strain. Reconstruction was based on the annotated
337 genome of EcN, available GEMs of closely related *E. coli* strains, and manual curation.
338 Initially, a draft reconstruction was generated by incorporating all shared reactions and
339 metabolites from the *E. coli* K-12 MG1655 and the uropathogenic *E. coli* CFT073 models, based
340 on the presence of homologous genes. Manual curation was subsequently performed to refine
341 the model, addressing non-homologous genes, filling metabolic gaps, verifying reaction
342 directionality, and removing redundant reactions based on literature evidence. The final step
343 involved the annotation of reactions, metabolites, and genes using multiple databases, including
344 KEGG, BioCyc, MetaNetX, SEED, Rhea, ChEBI, and NCBI. Systems Biology Ontology (SBO)
345 terms were assigned using SBOannotator (Leonidou *et al.*, 2023).
346 Unlike previously published models, which do not include the EcN cryptic plasmids pMUT1 and
347 pMUT2, iPC1373 explicitly incorporates these genetic elements. The concept of plasmid-
348 associated metabolic burden is well established, and several studies have addressed its
349 representation in metabolic models of various scales, including GEMs (Ow *et al.*, 2009; Zeng and
350 Yang, 2019; Oftadeh and Hatzimanikatis, 2024). Notably, simulations of wild-type EcN growth on
351 glucose using previously published models overestimated the experimental biomass growth rate
352 by approximately 20% (van't Hof *et al.*, 2022), suggesting an unaccounted burden or metabolic

353 cost. To capture this effect, iPC1373 includes the maintenance and expression of both cryptic
354 plasmids, implemented as two additional reactions (Eq. 1 and Eq. 2) representing their associated
355 biosynthetic and energetic demands. Plasmid maintenance costs were parameterized using
356 literature-derived estimates for plasmid-bearing *E. coli* (Weber *et al.*, 2002; Orth *et al.*, 2011; Zeng
357 and Yang, 2019). Accordingly, the NGAM parameter was adjusted by constraining the lower
358 bound of the ATPM reaction to the median of reported values: 25 $mmol \cdot gDW^{-1} \cdot h^{-1}$ under aerobic
359 conditions and 3.5 $mmol \cdot gDW^{-1} \cdot h^{-1}$ under anaerobic conditions. For additional validation, the
360 effect of ATP parameters on reaction fluxes, particularly biomass estimation, was evaluated (see
361 Supplementary Methods). A good correspondence was found between the values selected from
362 literature, and the resulting predictions using biomass as the objective function.
363 The finalized model comprises 2,716 reactions, 1,943 metabolites (corresponding to 1,217 unique
364 metabolites), and 1,373 genes. Model quality and consistency evaluated using the Memote
365 testing suite (Lieven *et al*., 2020), yielded a total score of 91% and a consistency score of 98%
366 (Supplementary Material 2). Annotation completeness scores were 78% for metabolites, 79% for
367 reactions, and 60% for genes; the latter reflecting incomplete database coverage. The coverage
368 of SBO terms reached 93%, indicating a high level of semantic and structural annotation.
369

370 **3.1.3 Structural analysis and comparison of available EcN metabolic models**

371 Once the iPC1373 model had been established, we evaluated its structural correspondence with
372 previously published reconstructions. The EcN CANYUN model was excluded from the
373 comparison and subsequent analyses owing to its deliberately reduced size. In terms of reactions
374 (Fig. 1A, Supplementary Material 3), the models shared 2,654 reactions, 83% of the total. Notably,
375 the published models iDK1463 and iHM1533 contain substantially more reactions; over half of
376 their surplus (154 of 284; 58%) comprises extracellular exchange and transport reactions,
377 followed by 39 reactions involved in alternative carbon metabolism (Fig. S1). The iPC1373 model
378 exhibits a smaller set of 17 unique reactions, comprising: i) two biomass formulations—wt and

379 core, corresponding respectively to the full, experimentally measured macromolecular
380 composition of the wild-type strain and a simplified precursor-based biomass reaction—, ii) the
381 two reactions related to plasmid representation, iii) seven transport reactions, and iv) six reactions
382 representing the O-antigen biosynthesis module (Wzx/Wzy/WaaL pathway), responsible for
383 semirough LPS production in EcN (Grozdanov *et al.*, 2002). In iDK1463, apart from the biomass
384 reaction, the other 19 unique reactions were removed during the curation of iPC1373 as
385 duplicates or redundancies—changes consistent with the updated *E. coli* K-12 model iML1515
386 (e.g., ALAt2pp, ALAt2pp_copy1, ALAt2pp_copy2). Finally, iHM1533 includes 188 unique
387 reactions (Fig. S2), primarily in secondary-metabolite biosynthesis (102), transport and exchange
388 (43), and aromatic compounds breakdown (15).
389 For metabolites (Fig. 1B, Supplementary Material 4), the main differences occur in iHM1533 and
390 reflect the subsystem expansions aimed at optimizing specialized metabolism (Fig. S3),
391 consistent with the reaction-level disparities among the models. In terms of MEMOTE (Fig. S4),
392 the three EcN models show high consistency (97–98%). Annotation coverage is strong for
393 metabolites and reactions (78–84%), with gene annotation being the main weakness (60–68%).
394 iPC1373 has the highest SBO term coverage (93%). Overall scores are high and comparable
395 (89–91%), with iPC1373 slightly leading by two points, indicating all three reconstructions are
396 robust with minor annotation differences.
397

398 ## 3.2 Model validation with experimental flux datasets

399 The next step was to validate the phenotypic consistency of the EcN metabolic models against
400 independent experimental data, beyond their mathematical accuracy as calculated with
401 MEMOTE. Equations reflecting plasmid replication were incorporated into all EcN models
402 available for validation, comparison and predictive analyses. Unless otherwise specified, models
403 including these plasmid-associated reactions were employed.

404 A quantitative approach was adopted to rigorously evaluate model performance of the different
405 EcN metabolic models, examining whether increased model complexity improves quantitative
406 predictive accuracy (O'Brien *et al.*, 2015; Blázquez *et al.*, 2023). Since completeness and
407 mathematical consistency do not necessarily ensure accurate physiological predictions, this
408 validation aimed to test the models' ability to reproduce experimental biomass formation,
409 substrate uptake, and by-product secretion rates—key parameters for bioprocess control and
410 understanding EcN physiology. For that, several datasets from the literature were compiled (see
411 above *Experimental datasets* in Methods).
412

413 **3.2.1 Experimental growth scenarios and model consistency analysis**

414 First, the experimental datasets were used to evaluate consistency—that is, the compatibility
415 between the experimental data and the model—by applying possibilistic MFA and computing the
416 possibility (π) of the most possible flux state. As shown in Table 2, the consistency analysis
417 yielded possibility values of 1 for the Revelles and coworkers (2013) dataset, indicating full
418 agreement between the data and the models. For both fucose and rhamnose conditions,
419 consistency was marginally higher in the iHM1533 model, increasing with model complexity.
420 Conversely, randomly generated scenarios yielded possibilities lower than 0.1 on over 98% of
421 occasions.
422

423 **3.2.2. Prediction of biomass growth rate**

424 Then, we evaluated whether the incorporation of the plasmid module and the adjustment of
425 energy maintenance rates improved the prediction of biomass growth rates across the
426 experimental datasets. Introducing the plasmid module partially corrected the overestimation of
427 *in vivo* growth observed in all simulated carbon sources (Table 3). Model predictions including the
428 plasmid representation showed better agreement with experimental data, clearly decreasing the

429 relative squared error (RSE) across all scenarios and models, resulting in an average reduction
430 from 16% to 4% (Table S1). There were no wide differences in performance between the different
431 models, with the iDK1463-based reconstruction achieving the best biomass prediction, with a
432 prediction error of 3.6%, compared to 4.0% and 4.6% for iHM1533 and iPC1373, respectively.
433

434 **3.2.3 Prediction of remaining external rates**

435 Once plasmid incorporation was shown to improve biomass growth prediction, external production
436 rates were estimated using the plasmid-bearing models by constraining only the carbon source
437 uptake rate to the experimental values of each dataset. This approach reflects typical fermentation
438 conditions, where substrate uptake is measured and controlled, while other exchange fluxes are
439 rarely monitored. Predictions for all datasets were obtained using FBA, FVA, and possibilistic
440 MFA (Figure 2; Supplementary Material S5).
441 Overall, a good directional agreement was observed between experimental and *in silico* FBA-
442 predicted values (Table S2), though quantitative deviations were larger than those found for
443 biomass estimation, with an overall average RSE of 32% (Table S3). Among the models, iPC1373
444 provided in this case the most accurate predictions, achieving an RSE of 27%, compared to 35%
445 for both iDK1463 and iHM1533.
446 Additionally, the three models produced comparable external flux predictions across all
447 methodologies, with only minor variations. Noticeable discrepancies appeared for 1,2-propanediol
448 production (Fig. 2E, Fig. 2F), pyruvate production under gluconate conditions (Fig. 2C, Fig. 2D),
449 and glycogen formation (Fig. 2B). Range-based approaches (possibilistic MFA and FVA)
450 generally encompass both the FBA predictions and the experimental values within their bounds.
451 In some cases—such as 1,2-propanediol production—these range-based methods captured the
452 experimental results more accurately.
453

454 **3.2.4 Prediction of internal flux reactions rates**

455 Finally, since the complete flux distribution can be estimated, we compared the model predictions
456 with the experimental ^{13}C fluxomics data reported for the EcN wild-type strain by Revelles and
457 coworkers (2013), using glucose (Fig. 3A) and gluconate (Fig. 3B) as sole carbon sources. Both
458 FBA and possibilistic MFA were applied to analyze central metabolic pathways and reactions (Fig.
459 3, Supplementary Material S6), allowing for a comparison of the methodologies. Overall, model
460 predictions qualitatively followed the experimental intracellular flux trends, except for the TCA
461 cycle, which showed underestimated fluxes under both carbon sources–except for the iHM1533
462 model in glucose using possibilistic MFA–and reaction 7 of glycolysis (the conversion of
463 phosphoenolpyruvate to pyruvate) under glucose. Quantitative discrepancies depended on both
464 the model and the methodology, although FBA predictions were generally accurate across
465 reaction groups. However, notable deviations were observed in reactions 4 (glucose) and 14
466 (gluconate), where the iPC1373 and iHM1533 models demonstrated greater consistency with the
467 experimental findings. Overall, attending to median RSE, all models performed similarly across
468 the different scenarios with relative errors ranging 29-33% (Table S4).

469 When possibilistic MFA was applied to estimate the most probable flux value for each reaction,
470 larger differences emerged depending on the underlying model. In this case, the model iPC1373
471 achieved the highest overall accuracy, better capturing the fluxes through the TCA cycle reactions
472 and exhibiting the lowest median RSE across the 23-24 reactions analyzed, 32% (Table S4).

473

474 **3.3 EcN exhibits a distinct metabolomic profile in GMM medium, both**
475 **experimentally and *in silico***

476 Once the models were validated and compared against experimental flux datasets, we examined
477 whether they could reproduce EcN's metabolic behavior under different environmental conditions.
478 To this end, EcN and the reference *E. coli* K-12 strain were cultured anaerobically in minimal M9

479 media and Gut Microbiota Medium (GMM), which promotes the growth of beneficial gut bacteria
480 and enhance short-chain fatty acids (SCFAs) production (Park *et al*., 2024). This setup aimed to
481 determine if GMM elicits distinct metabolic behavior in the probiotic strain, as it mimics the nutrient
482 composition of the gut environment. Although previous GEM studies (Kim *et al.,* 2021) suggested
483 that EcN behaves differently under anaerobic environments, this physiologically relevant aspect
484 has not yet been experimentally confirmed best of our knowledge.
485 To explore global metabolic differences, representative metabolites–including organic acids,
486 SCFAs, and amino acids–were quantified. Samples were grouped by strain and medium,
487 revealing largely similar profiles overall, except for EcN grown in GMM, which clearly separated
488 from the rest (Fig. 4A). This separation was mainly driven by PC1, which explained 82% of the
489 total variance, followed by PC2 (10.5%). The metabolites contributing most to this separation
490 were lactate, acetate, formate, and succinate.
491 When analyzing specific patterns in the different extracellular compounds, some metabolites
492 showed no detectable production or consumption or did so only in insignificant changes in several
493 compounds, including propionate, isobutyrate, and isovalerate (Supplementary Material S7). In
494 contrast, clear differences were observed in the production of organic acids—particularly
495 SCFAs—and in amino acid consumption when EcN was grown in GMM medium (Fig. 4B-I).
496 Indeed, an overall increase in fermentation-related metabolites and SCFAs was found in EcN
497 cultures under GMM conditions (Fig. 4B-F). EcN produced acetic, butyric, formic, lactic, and
498 succinic acids at net concentrations of 1.19, 0.16, 0.49, 1.01, and 0.49 g/L, respectively
499 (Supplementary Material S7). In addition, small amounts of valeric acid (≈20 mg/L) were detected
500 exclusively in EcN cultures (Fig. 4F). Compared to the K-12 strain, EcN showed markedly higher
501 production of acetic, butyric, lactic, and succinic acids (0.08, 0.08, 0.33, and 0.35 g/L in K-12,
502 respectively). Moreover, EcN demonstrated distinct amino acid utilization patterns within GMM
503 (Fig. 4G-I), exhibiting elevated uptake of arginine and tryptophan compared to the K-12 strain or
504 EcN in M9 medium. Interestingly, ornithine consumption showed opposite behavior for the strains.

505 These observed variations seem independent of growth, as EcN achieved a lower optical density
506 under these conditions (Supplementary Material S7).
507 Finally, the models were used to simulate EcN metabolism under M9 and GMM media (Fig. 5,
508 Supplementary Material S8). As previously evaluated, all three models successfully captured the
509 overflow metabolism characteristic of EcN growth in M9 medium with glucose (Fig. 5A), including
510 the production of acetate, ethanol and formate, although they failed to reproduce lactate
511 formation. Under simulated GMM conditions (Fig. 5B), all three models accurately reproduced the
512 distinctive overproduction of succinate and acetate observed experimentally. Model iPC1373 also
513 captured the increased butyrate production, while iHM1533 predicted higher lactate formation. In
514 contrast, ethanol production was not predicted by any model —which aligns with its low
515 experimental levels when compared to M9 medium (Supplementary Material S7). Enhanced
516 consumption of key amino acids, including glutamate, arginine, and tryptophan, was also correctly
517 represented. However, none of the models reproduced ornithine uptake, with iHM1533 instead
518 predicting its secretion. Some discrepancies, such as the absence of isovaleric, valeric, and
519 GABA predictions were probably due to missing reactions or metabolites in the reconstructions.
520

521 ## 4. Discussion

522 The main objective of this study was to develop a GEM of EcN capable of accurately reproducing
523 its physiological behavior, with particular focus on its characteristic probiotic traits. Although EcN
524 is a well-characterized and commercially available probiotic strain, systems biology studies have
525 paradoxically emphasized its potential as a biosynthetic chassis rather than its probiotic
526 physiology. Owing to its strong adherence, resistance, and persistence in the gastrointestinal
527 tract, EcN represents an attractive platform for targeted drug delivery and live biotherapeutic
528 applications (Kan *et al.,* 2021; Kim *et al.,* 2021). Previous research has typically relied on specific
529 experimental constraints, media compositions, or genetically modified variants —such as the

530 plasmid-cured strain reported by van't Hof *et al.* (2022)—to characterize EcN's performance as a
531 recombinant host.
532 In constructing our model, we explicitly incorporated the presence and energetic cost of EcN's
533 cryptic plasmids. The metabolic burden associated with plasmid maintenance is well established
534 in *E. coli* (Ow *et al*., 2009; Oftadeh and Hatzimanikatis, 2024), and previous EcN GEMs have
535 tended to overestimate biomass growth rates (van't Hof *et al*., 2022). Van't Hof and colleagues
536 attributed these discrepancies to incomplete biomass formulations, underestimated maintenance
537 energy requirements, or missing by-product pathways. By introducing a plasmid-specific module
538 representing the additional energetic demands of plasmid maintenance, our model achieved more
539 realistic biomass yield estimates and accurately reproduced acetate overflow metabolism. This
540 adjustment reduced the overestimation of *in silico* growth rates to within 5% of experimental
541 measurements. Beyond improving model fidelity, this refinement is also relevant to synthetic
542 biology applications, where EcN's cryptic plasmids are often retained and exploited as stable
543 endogenous expression vectors. The systematic inclusion of endogenous plasmids could also
544 have broader implications for GEM development in other microorganisms. Many industrially and
545 clinically important bacteria —whether probiotic, pathogenic, or commensal— naturally harbor
546 plasmids (Fogarty *et al*., 2024; Garcillán-Barcia *et al.,* 2025), yet these elements are rarely
547 represented in their respective GEMs. Incorporating plasmid-associated metabolic costs and
548 interactions thus represents a promising strategy to enhance both model realism and predictive
549 power.
550 To further assess model performance, we compared several EcN GEMs after integrating the
551 plasmid module, evaluating the trade-offs between model complexity and accuracy in reproducing
552 experimental physiological behavior. Interestingly, models with differing levels of curation and
553 structural detail yielded distinct quantitative outcomes when validated against experimental flux
554 data. Notably, greater model complexity did not necessarily improve predictive accuracy —
555 particularly for external fluxes such as acetate or ethanol secretion— and similar inconsistencies

556 were observed for intracellular flux predictions. Some recurring discrepancies, including those
557 related to 1,2-propanediol, pyruvate, or glycogen production, likely reflect inherent limitations of
558 FBA, which assumes growth optimization and neglects non-essential or suboptimal fluxes. These
559 by-products are often regulated through mechanisms such as CsrA-dependent glycogen
560 accumulation, underscoring the need to integrate regulatory constraints or multi-objective
561 formulations to more faithfully reproduce EcN's observed metabolism.
562 Our analyses further demonstrated that intracellular flux prediction accuracy does not scale
563 linearly with model complexity. Predictive performance depended on both the computational
564 method and the reaction–model pairing, indicating that specific models may better capture
565 particular pathway dynamics. Under certain environmental simulations, some metabolites—such
566 as valerate, isovalerate, and GABA—were not predicted by all models, likely due to missing
567 exchange reactions or incomplete pathway representations. Other deviations, such as the
568 absence of lactate production in IDK1463 and iPC1373 or misrepresented ornithine flux in
569 iHM1533, may result from alternative flux distributions that divert carbon flow from these
570 pathways.
571 While qualitative predictions aid in interpreting general physiological trends, quantitative precision
572 is crucial for industrial applications. In probiotic manufacturing, accurate control of fermentation
573 parameters is essential to maintain cell viability and product stability. GEMs are valuable tools for
574 optimizing such processes, especially when it is difficult the direct measurement of key
575 parameters such as growth, substrate uptake, or by-product secretion. Reliable model predictions
576 can thus improve fermentation efficiency and product consistency. Accordingly, the present study
577 establishes a framework for developing and validating strain-specific GEMs with quantitative data
578 and industrial relevance.
579 From a methodological perspective, this work also illustrates how distinct metabolic flux analysis
580 approaches provide complementary insights into cellular metabolism. Possibilistic MFA proved
581 particularly valuable for incorporating measurement uncertainty and generating robust flux

582 estimates in underdetermined systems, where conventional MFA or FBA may produce less stable
583 solutions. Comparative analyses of FBA, FVA, and possibilistic MFA revealed that each method
584 captures unique aspects of metabolic behavior and uncertainty. Furthermore, as shown in this
585 study, different metabolic reconstructions—regardless of their origin or level of curation—can be
586 effectively prioritized depending on the specific biological or experimental context.
587 Finally, this study enhances our understanding of EcN's unique probiotic features. Metabolomic
588 analyses confirmed that EcN displays a distinctive metabolic profile under anaerobic conditions
589 in GMM, supporting previous *in silico* observations that it expresses unique metabolic traits under
590 gut-like environments (Kim *et al*., 2021; van't Hof *et al*., 2022). This phenotype is characterized
591 by selective consumption and secretion of key metabolites that distinguish EcN from other *E. coli*
592 strains. Compared to *E. coli* K-12, EcN consumes several amino acids at higher rates, potentially
593 conferring a competitive advantage in the intestinal niche by depleting local nutrient pools and
594 limiting resources available to competing microorganisms. Simultaneously, EcN produces
595 elevated levels of organic acids—such as lactic and acetic acid—that inhibit pathogens, as well
596 as SCFAs like butyrate and valerate, which promote the growth of beneficial anaerobes. Together,
597 these findings reinforce EcN's physiological distinctiveness and provide ecological insight into its
598 probiotic mechanisms of action. They also confirm that EcN's phenotype is strongly context-
599 dependent, particularly on growth medium composition and oxygen availability. Thus,
600 physiologically relevant conditions are essential when validating and applying GEMs to
601 commensal, probiotic, or pathogenic microorganisms.
602 Future work could extend this framework to simulate microbial interactions within the gut
603 ecosystem. Integrating EcN into community-scale metabolic models —similar to prior approaches
604 for *Bacillus* and *Lactobacillus* species (Devika *et al*., 2021)— would allow exploration of its
605 interactions with both commensal and pathogenic bacteria, offering deeper insights into its
606 ecological role and therapeutic potential.
607

5. Conclusion

In summary, this study presents a refined genome-scale metabolic model of EcN that explicitly accounts for the energetic cost of its cryptic plasmids and achieves improved quantitative agreement with experimental data. Comparative analyses with previous models revealed that increased structural complexity does not necessarily enhance predictive accuracy, underscoring the importance of contextual and condition-specific validation. Metabolomic analyses further demonstrated that EcN exhibits a distinctive metabolic phenotype under gut-like conditions, characterized by enhanced amino acid consumption and short-chain fatty acid production, reinforcing its probiotic identity. Overall, this work provides a robust framework for the quantitative and physiologically relevant modeling of EcN and related commensal bacteria, with implications for probiotic research, industrial fermentation, and synthetic biology applications.

Author Contributions

Paola Corbín Agustí: Conceptualization, Methodology, Formal analysis, Visualization, Investigation, Software, Data curation, Writing– Original draft, Reviewing and Editing. **Alba Arévalo-Lalanne:** Methodology, Investigation, Writing– Reviewing and editing. **Patricia Álvarez**: Methodology, Formal analysis, Investigation, Software, Data curation, Writing– Reviewing and editing. **Daniel Ramón**: Conceptualization, Formal analysis, Writing– Reviewing and editing. **Juli Peretó**: Conceptualization, Formal analysis, Funding acquisition, Writing– Original draft, Reviewing and Editing. **Marta Tortajada:** Conceptualization, Formal analysis, Supervision, Writing– Original draft, Reviewing and Editing.

Funding

P. Corbín-Agustí and A. Arévalo-Lalanne were recipients of predoctoral fellowships from the Generalitat Valenciana under the call “Subvenciones para la contratación de personal

investigador de carácter predoctoral (ACIF)", with reference ACIF/2021/110 and "Subvenciones para la formación de doctores y doctoras en empresas valencianas (FDGENT)", with reference FDEGENT/2020/006, respectively.

Conflicts of Interest

D. Ramón and M. Tortajada are former employees of Biopolis–Archer Daniels Midland and contributed to this study in their personal capacity.

Data Availability Statement

The code and data for reproducing this analysis are available in the Github repository: https://github.com/paocorbin/EcN-GEM

6. References

Álvarez, P., Corbín, P., Ruiz, M., Peretó, J., and Tortajada, M. (2019) Reconstruction and analysis of genome-scale metabolic models of industrial *Escherichia coli* strains. In: *Proceedings 8th International Conference on Foundations of Systems Biology in Engineering*. Valencia: FOSBE, pp. 275–276.

Antoniewicz, M.R. (2021) A guide to metabolic flux analysis in metabolic engineering: Methods, tools and applications. *Metabolic Engineering* 63: 2–12.

Beimfohr, C. (2016) A review of research conducted with probiotic *E. coli* marketed as Symbioflor. International *Journal of Bacteriology* 2016: 1–10.

Bernstein, D.B., Akkas, B., Price, M.N., and Arkin, A.P. (2023) Evaluating *E. coli* genome-scale metabolic model accuracy with high-throughput mutant fitness data. *Molecular Systems* Biology 19: e11566.

Blázquez, B., San León, D., Rojas, A., Tortajada, M., and Nogales, J. (2023) New insights on metabolic features of *Bacillus subtilis* based on multistrain genome-scale metabolic modeling. International. *Journal of Molecular Sciences* 24: 7091.

Blum-Oehler, G., Oswald, S., Eiteljörge, K., Sonnenborn, U., Schulze, J., Kruis, W., and Hacker, J. (2003) Development of strain-specific PCR reactions for the detection of the probiotic *Escherichia coli* strain Nissle 1917 in fecal samples. *Research in Microbiology* 154: 59–66.

Caspi, R., Billington, R., Keseler, I.M., Kothari, A., Krummenacker, M., Midford, P.E., et al. (2020) The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Research* 48: D445–D453.

Deriu, E., Liu, J.Z., Pezeshki, M., Edwards, R.A., Ochoa, R.J., Contreras, H., et al. (2013) Probiotic bacteria reduce *Salmonella typhimurium* intestinal colonization by competing for iron. *Cell Host & Microbe* 14: 26–37.

Devika, N.T., Jangam, A.K., Katneni, V.K., Patil, P.K., Nathamuni, S., and Shekhar, M.S. (2021) In silico prediction of novel probiotic species limiting pathogenic *Vibrio* growth using constraint-based genome scale metabolic modeling. *Front Cell Infect Microbiol* 11:.

Ebrahim, A., Lerman, J.A., Palsson, B.O., and Hyduke, D.R. (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol* 7: 74.

Fogarty, E. C., Schechter, M. S., Lolans, K., Sheahan, M. L., Veseli, I., Moore, R. M., Kiefl, E., Moody, T., Rice, P. A., Yu, M. K., Mimee, M., Chang, E. B., Ruscheweyh, H.-J., Sunagawa, S., Mclellan, S. L., Willis, A. D., Comstock, L. E., and Eren, A. M. (2024). A cryptic plasmid is among the most numerous genetic elements in the human gut. *Cell* 187: 1206-1222.e16.

Garcillán-Barcia, M. P., de la Cruz, F., and Rocha, E. P. C. (2025). The extended mobility of plasmids. *Nucleic Acids Research* 53: gkaf652.

Goodman, A.L., Kallstrom, G., Faith, J.J., Reyes, A., Moore, A., Dantas, G., and Gordon, J.I. (2011) Extensive personal human gut microbiota culture collections characterized and manipulated in gnotobiotic mice. *Proc Natl Acad Sci USA* 108: 6252–6257.

Grozdanov, L., Zähringer, U., Blum-Oehler, G., Brade, L., Henne, A., Knirel, Y.A., et al. (2002) A single nucleotide exchange in the *wzy* gene Is responsible for the semirough O6 lipopolysaccharide phenotype and serum sensitivity of *Escherichia coli* strain Nissle 1917. *J Bacteriol* 184: 5912–5925.

Güttsches, A.-K., Löseke, S., Zähringer, U., Sonnenborn, U., Enders, C., Gatermann, S., and Bufe, A. (2012). Anti-inflammatory modulation of immune response by probiotic *Escherichia coli* Nissle 1917 in human blood mononuclear cells. *Innate Immunity* 18: 204-216.

Hare, P.J., Englander, H.E., and Mok, W.W.K. (2022) Probiotic *Escherichia coli* Nissle 1917

inhibits bacterial persisters that survive fluoroquinolone treatment. *Journal of Applied Microbiology* 132: 4020–4032.

Kan, A., Gelfat, I., Emani, S., Praveschotinunt, P., and Joshi, N.S. (2021) Plasmid vectors for in vivo selection-free use with the probiotic *E. coli* Nissle 1917. *ACS Synthetic Biology* 10: 94–106.

Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* 49: D545–D551.

Kim, D., Kim, Y., and Yoon, S. H. (2021). Development of a genome-scale metabolic model and phenome analysis of the probiotic *Escherichia coli* strain Nissle 1917. *International Journal of Molecular Sciences* 22: 2122.

King, Z.A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J.A., et al. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Research* 44: D515–D522.

Leonidou, N., Fritze, E., Renz, A., and Dräger, A. (2023) SBOannotator: a Python tool for the automated assignment of systems biology ontology terms. *Bioinformatics* 39: btad437.

Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., et al. (2020) MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol* 38: 272–276.

Lin, Q., Jiang, Z., Zhong, B., Chen, J., Lv, Z., and Nie, Z. (2024) Unveiling the impact of cryptic plasmids curing on *Escherichia coli* Nissle 1917: massive increase in Ag43c expression. *AMB Expr* 14: 48.

Llaneras, F. (2011) *Interval and possibilistic methods for constraint-based metabolic models*. PhD Dissertation, Universitat Politècnica de València.

Llaneras, F., Sala, A., and Picó, J. (2009) A possibilistic framework for constraint-based metabolic flux analysis. *BMC Syst Biol* 3: 79.

Maltby, R., Leatham-Jensen, M.P., Gibson, T., Cohen, P.S., and Conway, T. (2013) Nutritional basis for colonization resistance by human commensal *Escherichia coli* strains HS and Nissle 1917 against *E. coli* O157:H7 in the mouse intestine. *PLoS ONE* 8: e53957.

Millard, P., Pérochon, J., and Létisse, F. (2021) Functional analysis of deoxyhexose sugar utilization in *Escherichia coli* reveals fermentative metabolism under aerobic conditions. *Applied and Environmental Microbiology* 87: e00719-21.

Monk, J.M., Charusanti, P., Aziz, R.K., Lerman, J.A., Premyodhin, N., Orth, J.D., et al. (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci USA* 110: 20338–20343.

Monk, J.M., Koza, A., Campodonico, M.A., Machado, D., Seoane, J.M., Palsson, B.O., et al. (2016) Multi-omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. *Cell Systems* 3: 238–251.

Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., et al. (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol* 35: 904–908.

Morales, Y., Tortajada, M., Picó, J., Vehí, J., and Llaneras, F. (2014) Validation of an FBA model for *Pichia pastoris* in chemostat cultures. *BMC Syst Biol* 8: 142.

Moutinho, T.J., Neubert, B.C., Jenior, M.L., and Papin, J.A. (2022) Quantifying cumulative phenotypic and genomic evidence for procedural generation of metabolic network reconstructions. *PLoS Comput Biol* 18: e1009341.

Nogales, J. (2017) A practical protocol for genome-scale metabolic reconstructions. In *Hydrocarbon and Lipid Microbiology Protocols: Genetic, Genomic and System Analyses of Pure Cultures*. McGenity, T.J., Timmis, K.N., and Nogales, B. (eds). Berlin, Heidelberg: Springer, pp. 197–221.

O'Brien, E.J., Monk, J.M., and Palsson, B.O. (2015) Using Genome-scale models to predict biological capabilities. *Cell* 161: 971–987.

Oftadeh, O. and Hatzimanikatis, V. (2024) Genome-scale models of metabolism and expression predict the metabolic burden of recombinant protein expression. *Metab Eng* 84: 109–116.

Olier, M., Marcq, I., Salvador-Cartier, C., Secher, T., Dobrindt, U., Boury, M., et al. (2012) Genotoxicity of *Escherichia coli* Nissle 1917 strain cannot be dissociated from its probiotic activity. *Gut Microbes* 3: 501–509.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., and Palsson, B.Ø. (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Molecular Systems Biology* 7: 535.

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010) What is flux balance analysis? *Nat Biotechnol* 28: 245–248.

Ow, D.S.-W., Lee, D.-Y., Yap, M.G.-S., and Oh, S.K.-W. (2009) Identification of cellular objective for elucidating the physiological state of plasmid-bearing *Escherichia coli* using genome-scale in silico analysis. *Biotechnology Progress* 25: 61–67.

Özkan, P., Sariyar, B., Ütkür, F.Ö., Akman, U., and Hortaçsu, A. (2005) Metabolic flux analysis of recombinant protein overproduction in *Escherichia coli*. *Biochemical Engineering* Journal 22: 167–195.

Park, H., Yeo, S., Ryu, C. B., and Huh, C. S. (2024). A streamlined culturomics case study for the human gut microbiota research. *Scientific Reports* 14: 20361.

Pearcy, N., Hu, Y., Baker, M., Maciel-Guerra, A., Xue, N., Wang, W., et al. (2021) Genome-scale metabolic models and machine learning reveal genetic determinants of antibiotic resistance in *Escherichia coli* and unravel the underlying metabolic adaptation mechanisms. *mSystems* 6: e00913-20.

Petersen, A.M., Schjørring, S., Gerstrøm, S.C., and Krogfelt, K.A. (2011) Treatment of inflammatory bowel disease associated *E. coli* with ciprofloxacin and *E. coli* Nissle in the streptomycin-treated mouse intestine. *PLoS One* 6: e22823.

Ponce-de-Leon, M., Calle-Espinosa, J., Peretó, J., and Montero, F. (2015) Consistency analysis of genome-scale models of bacterial metabolism: A metamodel approach. *PLoS One* 10: e0143626.

Revelles, O., Millard, P., Nougayrède, J.-P., Dobrindt, U., Oswald, E., Létisse, F., and Portais, J.-C. (2013) The carbon storage regulator (Csr) system exerts a nutrient-specific control over central metabolism *in Escherichia coli* strain Nissle 1917. *PLoS One* 8: e66386.

Rozkov, A., Avignone-Rossa, C. a., Ertl, P. f., Jones, P., O'Kennedy, R. d., Smith, J. j., et al. (2004) Characterization of the metabolic burden on *Escherichia coli* DH1 cells imposed by the presence of a plasmid containing a gene therapy sequence. *Biotechnology and Bioengineering* 88: 909–915.

Schomburg, I., Chang, A., Placzek, S., Söhngen, C., Rother, M., Lang, M., et al. (2013) BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res* 41: D764-772.

Schumann, S., Alpert, C., Engst, W., Klopfleisch, R., Loh, G., Bleich, A., and Blaut, M. (2014) Mild gut inflammation modulates the proteome of intestinal *Escherichia coli*. *Environmental Microbiology* 16: 2966–2979.

Średnicka, P., Roszko, M.Ł., Popowski, D., Kowalczyk, M., Wójcicki, M., Emanowicz, P., et al. (2023) Effect of in vitro cultivation on human gut microbiota composition using 16S rDNA amplicon sequencing and metabolomics approach. *Sci Rep* 13: 3026.

Thiele, I. and Palsson, B.Ø. (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5: 93–121.

Tortajada, M., Llaneras, F., and Picó, J. (2010) Validation of a constraint-based model of *Pichia pastoris* metabolism under data scarcity. *BMC Syst Biol* 4: 115.

van't Hof, M., Mohite, O. S., Monk, J. M., Weber, T., Palsson, B. O., and Sommer, M. O. A. (2022). High-quality genome-scale metabolic network reconstruction of probiotic bacterium *Escherichia coli* Nissle 1917. *BMC Bioinformatics* 23: 566.

796 von Kamp, A. and Klamt, S. (2023) Balancing biomass reaction stoichiometry and measured
797 fluxes in flux balance analysis. *Bioinformatics* 39: btad600.
798 Weber, J., Hoffmann, F., and Rinas, U. (2002) Metabolic adaptation of *Escherichia coli* during
799 temperature-induced recombinant protein production: 2. Redirection of metabolic fluxes.
800 *Biotechnol Bioeng* 80: 320–330.
801 Wickham, H. (2009) *ggplot2: Elegant graphics for data analysis*, New York, NY: Springer New
802 York.
803 Zainuddin, H.S., Bai, Y., and Mansell, T.J. (2019) CRISPR-based curing and analysis of metabolic
804 burden of cryptic plasmids in *Escherichia coli* Nissle 1917. *Engineering in Life Sciences*
805 19: 478–485.
806 Zeng, H. and Yang, A. (2019) Quantification of proteomic and metabolic burdens predicts growth
807 retardation and overflow metabolism in recombinant *Escherichia coli*. *Biotechnology and*
808 *Bioengineering* 116: 1484–1495.
809 Zhou, S., Zhao, L., Zuo, W., Zheng, Y., Zhang, P., Sun, Y., Wang, Y., Du, G., and Kang, Z. (2024).
810 Minimizing endogenous cryptic plasmids to construct antibiotic-free expression systems
811 for *Escherichia coli* Nissle 1917. *Synthetic and Systems Biotechnology* 9: 165-175.

812 **Figure Legends**

813

814 **Figure 1. Comparison of reactions (A) and metabolites (B) among available *Escherichia***
815 ***coli* Nissle 1917 GEMs.**

816

817 **Figure 2. External fluxes predicted by the models iDK1463 (blue), iHM1533 (orange), and**
818 **iPC1373 (green).** Simulations were performed under aerobic conditions by constraining substrate
819 uptake rates to the corresponding experimental values. Panels A and B show predictions for
820 glucose in the wild type and the ΔcsrA51 mutant strains, respectively; panels C and D show
821 predictions for gluconate in the wild type and the ΔcsrA51 mutant; panel E and F corresponds to
822 fucose and rhamnose as sole carbon sources, respectively. Gray internal bars indicate FVA
823 ranges; colored bars (in blue, green and orange) represent the range of possibilistic MFA
824 solutions; circles denote FBA solution; and crosses mark experimental flux measurements. Fluxes
825 are displayed using a symlog scale with a linear threshold of 0.01 to highlight differences at low
826 magnitudes. Original flux values are provided in Supplementary Material S5.

827

828 **Figure 3. Predicted flux distributions in central metabolism for the models iDK1463 (blue),**
829 **iHM1533 (orange), and iPC1373 (green) under wildtype conditions using glucose (A) and**
830 **gluconate (B).** Simulations were performed by constraining the measured exchange fluxes to
831 their experimental values (Revelles et al., 2013). Circles denote FBA solution; triangles represent
832 possibilistic MFA at $\pi = 1$; and black crosses indicate experimental ^{13}C flux measurements. Fluxes
833 are shown on a linear scale and normalized to the carbon source uptake rate (set to 100). Flux
834 values can be found in Supplementary Material S6. Central metabolism reactions are grouped
835 as: glycolysis/gluconeogenesis (Glycolysis), tricarboxylic acid (TCA) cycle, anaplerotic reactions,
836 Entner-Doudoroff (ED) pathway, pentose phosphate pathway (PPP), and acetate/pyruvate
837 production. Reaction indices (1–24) correspond to the entries in Table S5, which list the

838 metabolite conversions (Revelles *et al.,* 2013) and their corresponding BiGG database reaction
839 IDs.
840
841 **Figure 4. Metabolite profiling of *E. coli* Nissle (EcN) and K-12 in M9 medium with glucose**
842 **and Gut Microbiota Medium (GMM).** A) Principal Component Analysis of HPLC quantified
843 metabolites for *E. coli* K-12 (circle) and EcN (diamond) cultured in GMM (yellow) or M9 (gray)
844 medium. B-I) Boxplots showing metabolite levels (in mg/L·OD600) grouped by strain and medium.
845 Each group includes three biological replicates, displayed as individual points (circles) within the
846 boxplots. Succinic, lactic, acetic, butyric and valeric acids correspond to SCFAs and fermentation-
847 related metabolites; while Arg (arginine), Orn (ornithine), Trp (tryptophan) are amino acids.
848 Positive values correspond to production of metabolites and negative, to consumption.
849
850 **Figure 5. Simulation results of EcN models under M9 medium with glucose and gut**
851 **microbiota medium (GMM).** The heatmap shows the FBA predicted fluxes for models iHM1533,
852 iDK1463 and iPC1373 under the *in silico* conditions of M9 medium with glucose (A) and GMM
853 medium (B), along with the corresponding experimental results for the EcN strain. Data are
854 displayed using a symlog scale to facilitate comparison between experimental measurements and
855 model predictions, including only those metabolites for which both experimental and simulated
856 values are available. The scale indicates consumed metabolites in orange, produced metabolites
857 in blue, and those with a value of zero –neither consumed nor produced– in gray. Experimental
858 values are provided in Supplementary Material S7; original flux values, in Supplementary Material
859 S8.
860

861 **Table legends**

862

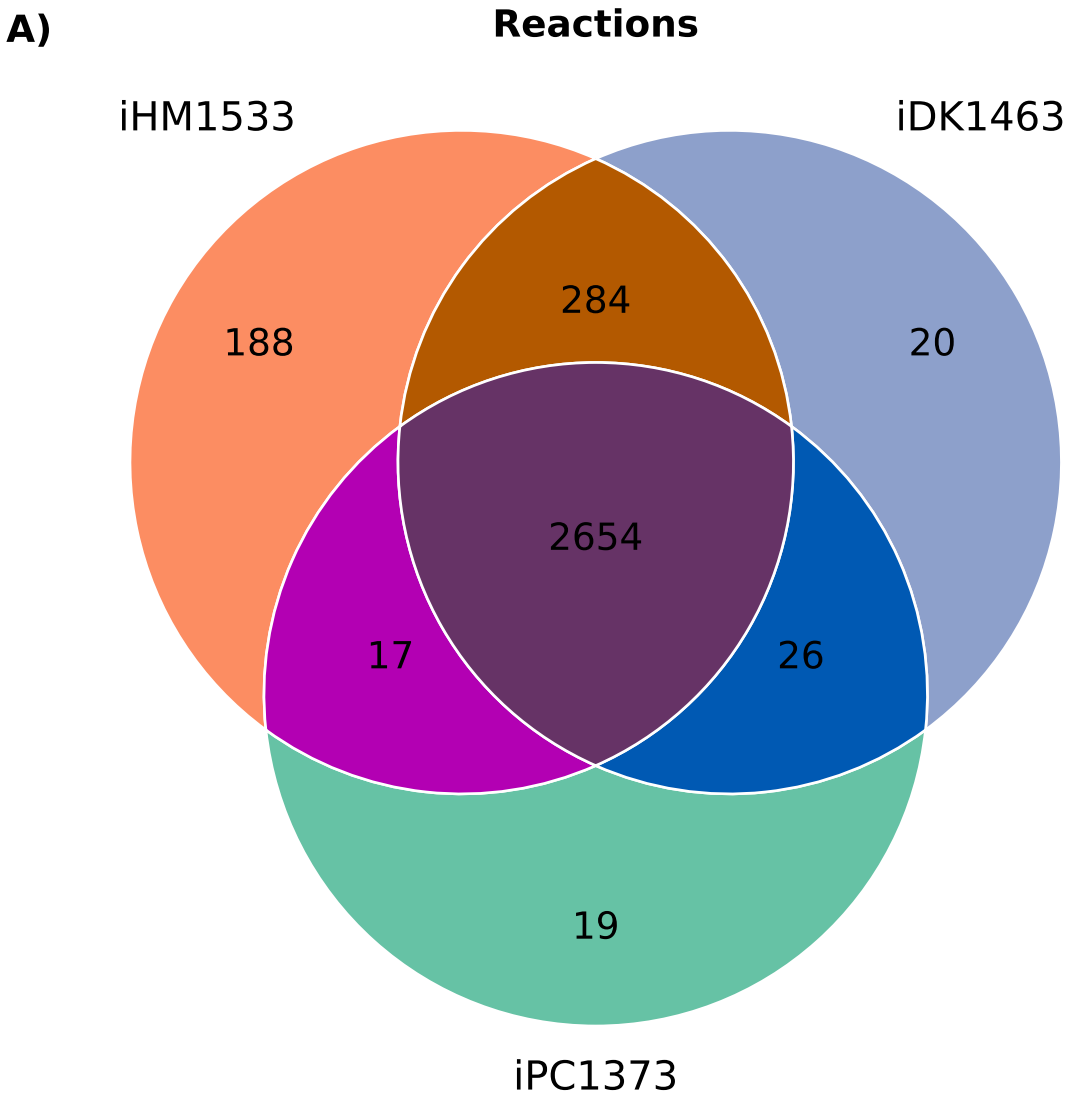
863 **Table 1. Main features of published genome-scale metabolic models (GEMs) of *Escherichia***
864 ***coli* Nissle 1917.** For each model, the table reports the total number of reactions, metabolites
865 and genes, along with a brief description of its focus and the type of experimental validation used.

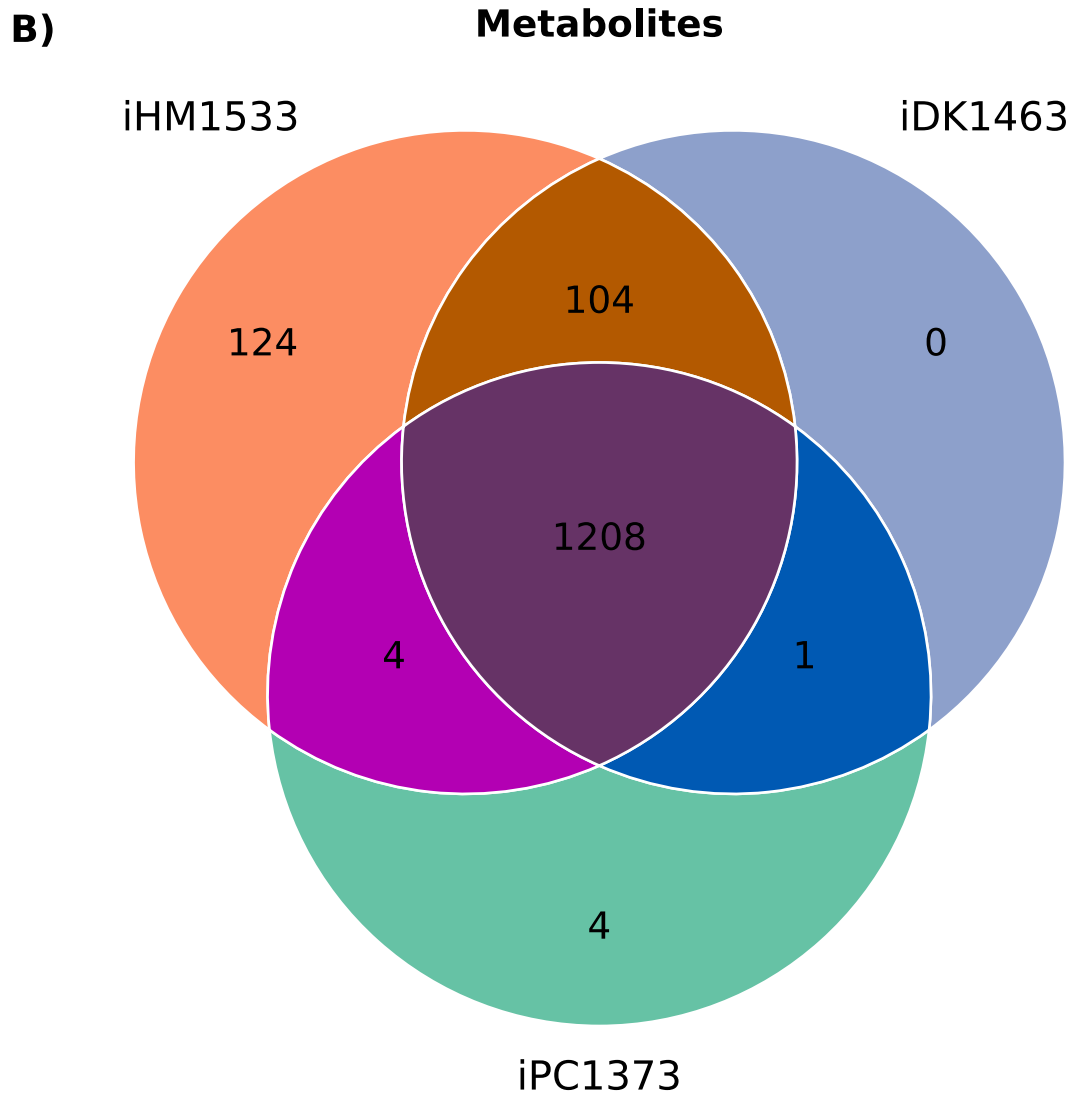
866

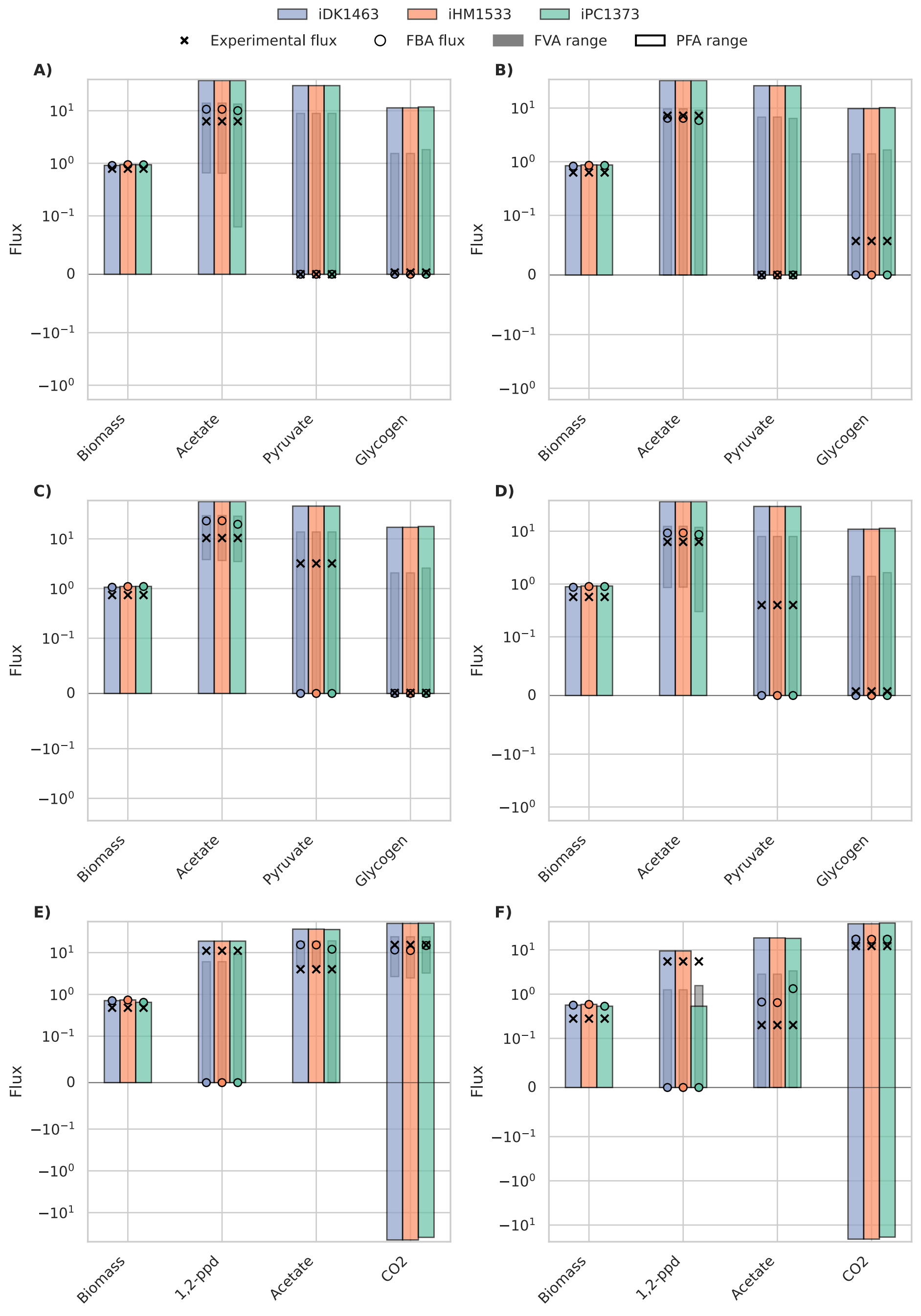
867 **Table 2. Experimental data for EcN cultures and model consistency.** Consistency values (π)
868 were calculated using the possibilistic MFA method for each dataset across the evaluated models:
869 iPC1373, iDK1463, and iHM1533. The data shown are the specific growth rate (h^{-1}), substrate
870 uptake (q_S), acetate formation (q_{ace}), pyruvate formation (q_{pyr}), CO_2 formation (q_{CO2}), glycogen
871 formation (q_{glyc}) and 1,2-propanediol formation (q_{12ppd}) obtained from datasets *(Revelles *et al*.,
872 2013) and **(Millard *et al*., 2021) under aerobic conditions.

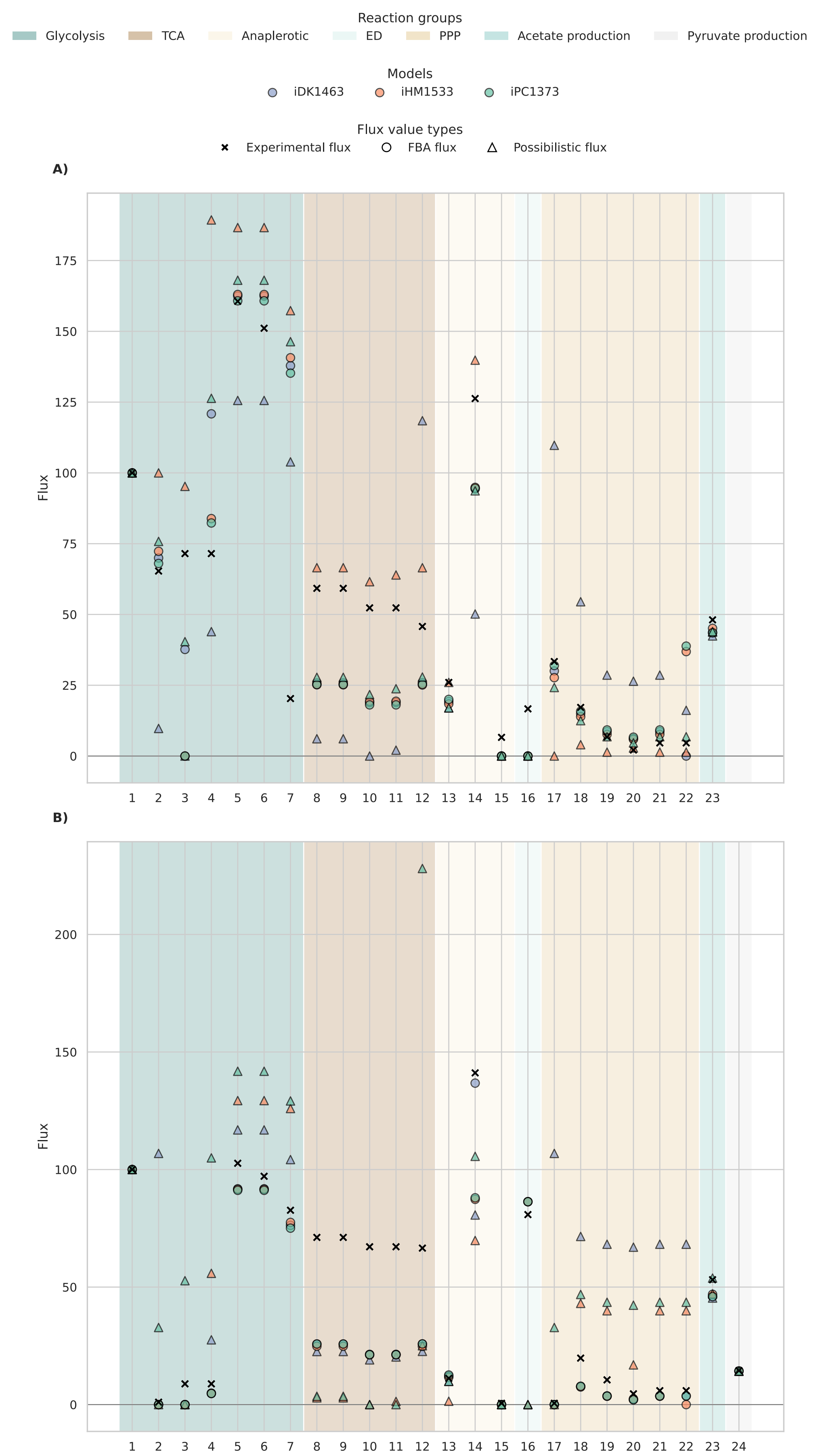
873

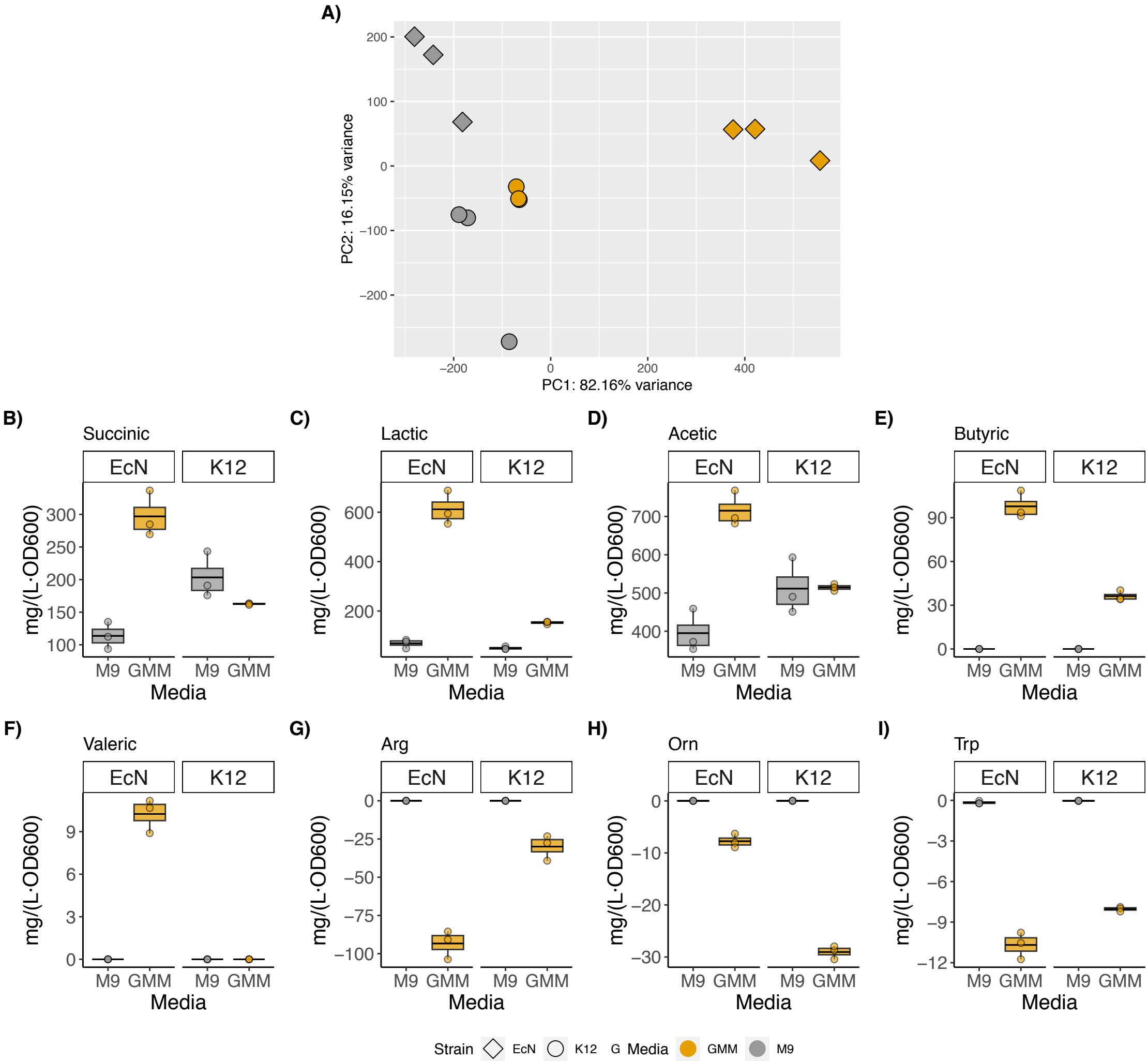
874 **Table 3. Comparison between experimental and model-predicted biomass growth rates.**
875 For each experimental dataset, all measured external fluxes—except biomass—were constrained
876 to their experimental values within the standard deviation. The table shows the FBA-optimized
877 growth rate predicted for each model, evaluated both with and without plasmid consideration.


A)
Reactions
iHM1533
iDK1463
188
284
20
2654
17
26
19
iPC1373


B)
Metabolites
iHM1533
iDK1463
124
104
0
1208
4
1
4
iPC1373


iDK1463
iHM1533
iPC1373
Experimental flux
FBA flux
FVA range
PFA range
A)
B)
C)
D)
E)
F)
Flux
10^1
10^0
10^{-1}
0
-10^{-1}
-10^0
-10^1
Biomass
Acetate
Pyruvate
Glycogen
1,2-ppd
CO2


Reaction groups
Glycolysis
TCA
Anaplerotic
ED
PPP
Acetate production
Pyruvate production
Models
iDK1463
iHM1533
iPC1373
Flux value types
Experimental flux
FBA flux
Possibilistic flux
A)
Flux
175
150
125
100
75
50
25
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
B)
Flux
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24


A)
PC2: 16.15% variance
PC1: 82.16% variance
B)
Succinic
C)
Lactic
D)
Acetic
E)
Butyric
F)
Valeric
G)
Arg
H)
Orn
I)
Trp
EcN
K12
mg/(L·OD600)
M9
GMM
Media
Strain
EcN
K12
Media
GMM
M9

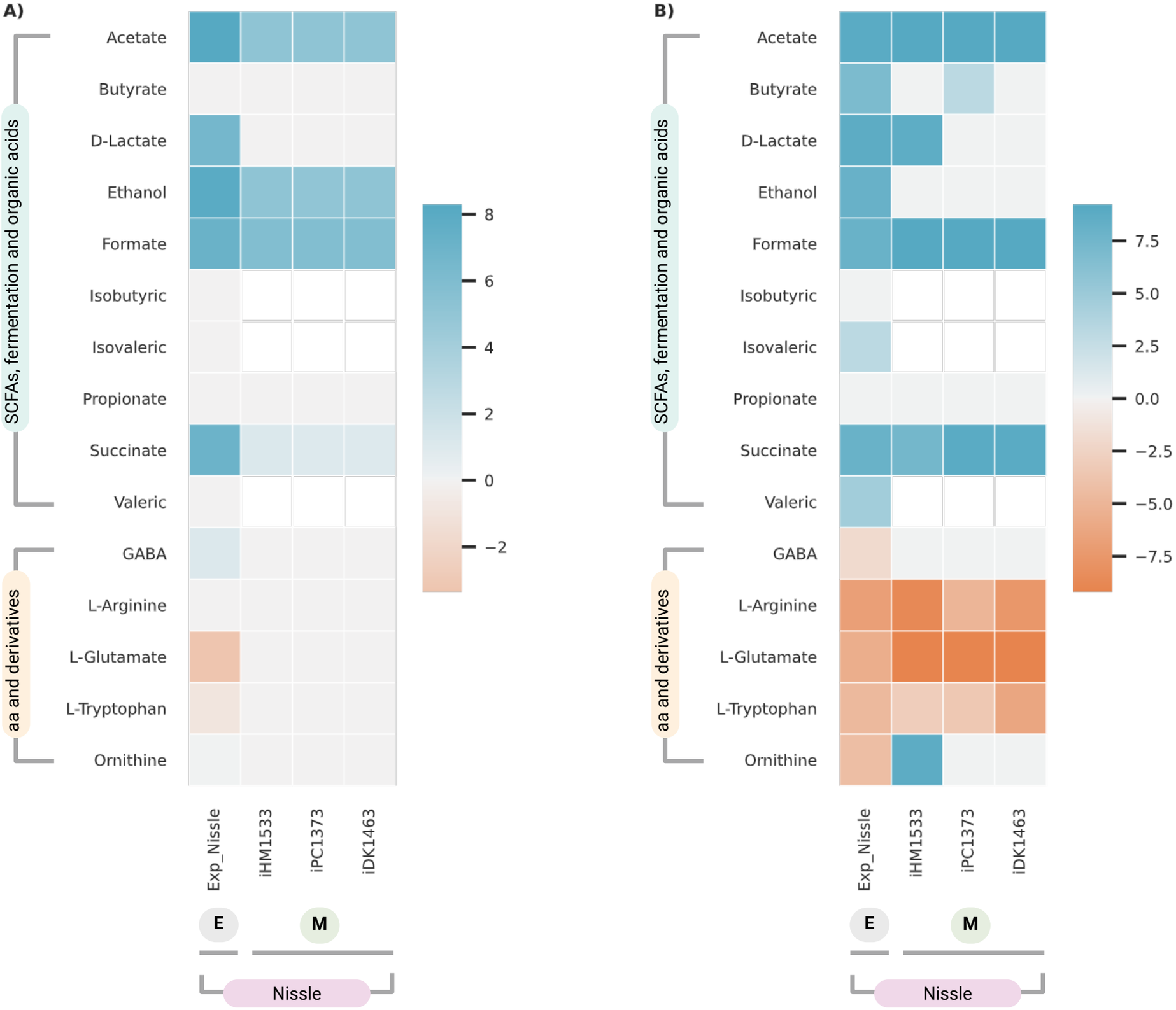

A)
B)
SCFAs, fermentation and organic acids
aa and derivatives
Acetate
Butyrate
D-Lactate
Ethanol
Formate
Isobutyric
Isovaleric
Propionate
Succinate
Valeric
GABA
L-Arginine
L-Glutamate
L-Tryptophan
Ornithine
Exp_Nissle
iHM1533
iPC1373
iDK1463
E
M
Nissle
8
6
4
2
0
−2
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5

Table 1.

Model ID	**CANYUN model**	**iDK1463**	**iHM1533**
Reference	Moutinho *et al*., 2022	Kim *et al*., 2021	van't Hoff *et al*., 2022
Reactions	737	2984	3143
Metabolites*	2607 (1661)	2112 (1313)	2261 (1440)
Genes	277	1464	1533
Experimental validation	Phenome analysis in aerobic and anaerobic conditions with Biolog assays for carbon sources.	Phenome analysis in aerobic conditions with Biolog assays for carbon, nitrogen, phosphorus, sulfur, nutrient supplements and inhibitory substrates.	Phenome analysis in aerobic conditions with Biolog assays for carbon, nitrogen, phosphorus and sulfur sources. Previously published ^{13}C fluxomics data.
Focus	CANYUN method validation and example of utility.	Comparative metabolic traits with other *E. coli* strains relevant to commensalism and gut colonization.	Optimization of secondary metabolites production.

*Numbers in parentheses indicate non-redundant metabolites, counted only once regardless of their compartmental location.

Table 2.

Condition	μ	q_s	q_{ace}	q_{CO2}	q_{pyr}	q_{glyc}	q_{12ppd}	Consistency		
	(h^{-1})	mmol·gDW^{-1}·h^{-1}						π iPC1373	π iDK1463	π iHM1533
Glucose wt*	0.79 ±0.02	12.50 ±-1.00	6.30 ±0.10	—	n.d	0.003 ±0.010	—	1	1	1
Gluconate wt*	0.74 ±0.02	20.00 ±1.70	10.40 ±0.10	—	3.20 ±0.10	0.001 ±0.010	—	1	1	1
Glucose ΔcsrA51*	0.63 ±0.20	10.80 ±1.00	7.20 ±0.10	—	n.d.	0.057 ±0.014	—	1	1	1
Gluconate ΔcsrA51*	0.57 ±0.02	12.60 ±1.70	6.30 ±0.10	—	0.40 ±0.10	0.007 ±0.010	—	1	1	1
Fucose**	0.48 ±0.01	12.80 ±0.30	4.00 ±0.20	15.30± 2.40	—	—	11.10± 0.20	0.32	0.35	0.45
Rhamnose**	0.28 ±0.01	6.90 ±0.20	0.20 ±0.10	12.30± 0.70	—	—	5.50 ±0.10	0.41	0.45	0.60
Random glucose	0-10	0-10	0-10	—	0-10	0-10	—	< 0.1: 98%	< 0.1: 99%	< 0.1: 99%

Random datasets correspond to 100 automatically generated flux datasets used for consistency testing.
n.d., not detected; —, not measured.

Table 3.

Models excluding plasmid consideration				
	Exp.	**iPC1373**	**iDK1463**	**iHM1533**
Condition	**Growth rate (h^{-1})**			
Glucose wt	0.790	1.098	1.059	1.097
Gluconate wt	0.740	1.103	1.064	1.105
Glucose ΔcsrA51	0.630	0.893	0.881	0.914
Gluconate ΔcsrA51	0.570	0.958	0.925	0.961
Fucose	0.480	0.493	0.493	0.513
Rhamnose	0.280	0.334	0.335	0.348
Models including plasmid consideration				
	Exp.	**iPC1373**	**iDK1463**	**iHM1533**
Condition	**Growth rate (h^{-1})**			
Glucose wt	0.790	0.903	0.870	0.902
Gluconate wt	0.740	0.914	0.882	0.916
Glucose ΔcsrA51	0.630	0.816	0.805	0.836
Gluconate ΔcsrA51	0.570	0.771	0.743	0.771
Fucose	0.480	0.388	0.390	0.406
Rhamnose	0.280	0.236	0.238	0.248