Acta Scientific MICROBIOLOGY (ISSN: 2581-3226)

Volume 5 Issue 2 February 2022

Research Article

Significant Differences in Media Components and Predicted Growth Rates of 58 *Escherichia coli* Genome-scale Models

Felicia Leyi Tan, Zhi Jue Kuan, Nabil Amir-Hamzah, Xander Kng, Yik Yew Wee, Si Xian Sor and Maurice HT Ling*

School of Applied Sciences, Temasek Polytechnic, Singapore

***Corresponding Author:** Maurice HT Ling, School of Applied Sciences, Temasek Polytechnic, Singapore.

Received: January 03, 2022

Published: January 27, 2022

Abstract

Escherichia coli is a common host for metabolite production and genome-scale metabolic models (GSMs) is an important computational tool to aid in such experimental design. As of September 30, 2021; 58 GSMs have been registered with BiGG database. However, these GSMs had been built for different applications and no large-scale comparative study had been performed to-date. In this study, we examine the media components and predicted growth rates of these 58 GSMs using flux balance analysis across various glucose uptake rates. Only 5 out of 29 uptake rates (as proxy for media components) are common in all 58 GSMs; namely, proton, water, ammonium, oxygen, and phosphate. 74.25% (2370 of the 3192) pairwise comparisons of predicted growth rates show significant differences (p-value < 0.05) and 34 of 42 pairwise comparisons of predicted growth rates within the same strain are significantly different. Hence, our results demonstrated substantial differences in media components and significant differences in predicted growth rates between the GSMs and even within GSMs constructed for the same strain.

Keywords: *Escherichia coli;* Genome-scale Metabolic Models (GSMs); Ammonium

Introduction

Escherichia coli is a fast-growing bacterium in chemically defined media and with extensive molecular tools available [1], it has been the linchpin in discovering many important findings in molecular biology and cell physiology [2]. The emergence of *E. coli* as a notable host for natural product biosynthesis [2] also led to successful engineering of *E. coli* for metabolite production [3]. However, a challenge faced by metabolic engineering is the complexity of pathway optimization [4,5]. This is due to the myriad of regulatory systems that control natural metabolic pathways. The precise control required over the expression of several natural and heterologous genes to avoid limiting the desired product yield [6] is an arduous process when done manually. Furthermore, each optimization is dependent on multiple factors such as the pathway and compound, making this a multivariate problem that poses as a difficulty to many researchers [7,8].

A solution to this challenge involves the usage of a Genome-Scale Metabolic Model (GSM), which is a fundamental framework built upon extensive collection and curation of biological data of gene annotation, gene functions, metabolites, metabolic reactions, enzymes, and their interactions inside a targeted organism [9-12]. These GSMs can be analyzed using computation algorithms such as constraint-based flux balance analysis (FBA) to comprehend the functions and objectives of the metabolic network [13]. Therefore, GSMs can be an *in silico* platform for lessening the difficulty and burden on the researchers [7], and had been used in many studies [12]. For example, iBsu1144 (a *Bacillus subtilis* GSM) was used to identify the effects of oxygen transfer rates on the production of serine alkaline protease and recombinant proteins [14], and iEK1101 (a *Mycobacterium tuberculosis* GSM) was used to study the metabolic status of *M. tuberculosis* under hypoxic conditions [15].

Since the 1st *E. coli* GSM, iJE660 [16], many *E. coli* GSMs have been developed with increasing availability of experimental data. The most recent *E. coli* GSM is the iML1515 model, which has been used to maximize lysine production [17]. Another commonly used model is iAF160 which had been used in various studies [18-20] including to optimize the yield of violacein [21]. As of September 30, 2021; BiGG database [22] lists 58 *E. coli* GSMs[1]. Hence, it is difficult to know which model to use, and what the similarities and differences are in terms of media components and predicted growth rate. Here, we review, and differentiate these 58 *E. coli* GSMs. Our results demonstrated substantial differences in media components and significant differences in predicted growth rates between the GSMs and even within GSMs constructed for the same strain.

[1]http://bigg.ucsd.edu/search?query=Escherichia+coli

Materials and Methods

Models

58 GSMs from BiGG database [22]; namely, (A) iJR904, (B) iAF1260, (C) e_coli_core, (D) iAF1260b, (E) iJO1366, (F) iEC042_1314, (G) iECP_1309, (H) iEC55989_1330, (I) iECABU_c1320, (J) iAPECO1_1312, (K) iEcolC_1368, (L) iECB_1328, (M) iB21_1397, (N) iECD_1391, (O) iECBD_1354, (P) iBWG_1329, (Q) ic_1306, (R) iECDH1ME8569_1439, (S) iEcDH1_1363, (T) iECED1_1282, (U) iETEC_1333, (V) iEcHS_1320, (W) iECIAI1_1343, (X) iECIAI39_1322, (Y) iECOK1_1307, (Z) iEKO11_1354, (AA) iLF82_1304, (AB) iECNA114_1301, (AC) iECO103_1326, (AD) iECO111_1330, (AE) iE2348C_1286, (AF) iEcE24377_1341, (AG) iECH74115_1262, (AH) iZ_1308, (AI) iECs_1301, (AJ) iECSP_1301, (AK) iECO26_1355, (AL) iG2583_1286, (AM) iNRG857_1313, (AN) iECS88_1305, (AO) iECSE_1348, (AP) iECSF_1327, (AQ) iEcSMS35_1347, (AR) iECDH10B_1368, (AS) iY75_1357, (AT) iEC1372_W3110, (AU) iUMN146_1321, (AV) iECUMN_1333, (AW) iUMNK88_1353, (AX) iUTI89_1310, (AY) iWFL_1372, (AZ) iECW_1372, (BA) iEC1349_Crooks, (BB) iEC1356_Bl21DE3, (BC) iEC1344_C, (BD) iEC1368_DH5a, (BE) iEC1364_W, and (BF) iML1515; were used.

Comparison of media components

Media components for each GSM could be extracted using Cameo [7], which was available via cameo-medium-cpds command from AdvanceSyn Toolkit [8], and compared across GSMs for commonality.

Comparison of growth rates

Predicted growth rates at various glucose intakes (10, 9, 8, 7, 6, 5, 4, 3, 2, and 1 millimole per gram dry weight per hour of glucose intake) were given as proxy as output from the objective function [23] after flux balance analysis [24] using Cameo [7], which was available via cameo-medium-fba command from AdvanceSyn Toolkit [8], by changing the glucose uptake rates (EX_glc__D_e) while maintaining the rest of the media components unchanged. Predicted growth rates from each GSM were normalized into normalized predicted growth rates by calculating each predicted growth rate as a ratio of predicted growth rate at 10 millimole per gram dry weight per hour of glucose intake. Normalized predicted growth rates between any two GSMs were compared using paired t-test to identify differences in predicted growth rates across different glucose uptake rates under the null hypothesis of no average differences in predicted growth rates.

Results and Discussion

A cursory examination suggests that these 58 GSMs were built for specific purposes (Table 1), such as metabolite production [25] or examining gene essentiality [26], or for specific strains, such as *E. coli* APEC O1 [27] or *E. coli* W3110 [28]. However, a review of the applications of each GSM is likely to be extensive not the intent of this study as we are interested only in similarities and differences of each GSM in terms of media components, predicted growth rate and fluxes.

GSM Model IDs	Purposes and Applications
(A) iJR904	Integrate and analyze the diverse datasets, such as 'omics' data and to provide a more chemically accurate description of *E. coli* metabolism over GSM iJE660a [29]. Investigating the predictive power of the constraint-based modelling approach of flux distribution in comparison to the kinetic modelling approach [30]. Test DEF, an automated gap filling approach based on the endosymbiosis theory to fill gaps and gain more insights for the genomic annotation and modelling [31].
(B) iAF1260	Based off iJR904, iAF1260 serves as a BIGG database as it contains the current knowledge of *E. coli* metabolism and this GSM is also a framework for mathematical analysis and computational predictions. iAF1260 showed an overall increase of 4 and 16% over iJR904 predictions provide a broader perspective of cellular metabolism for *E. coli* [32]. Investigate the efficiency of GridProd [33], a method of calculating parsimonious metabolic networks and the production of metabolites while involving several reactions that were included in the iAF1260 model. Repurposed into an extended version of the model, iAF1260vio, to include violacein production [21].
(C) e_coli_core	Subset of the iAF1260 containing the central metabolism of *E. coli* [34]. Repurposed as a stoichiometric model by adding eight different pathways to produce butanol [35]. Used in comparison to the iAF1260 model to test a new pessimistic optimization framework for the identification of the optimal knock out strategies for maximum targeted bio-production under model uncertainty and have derived the robustness and stability of the metabolic strain perturbation through the modelling [36].

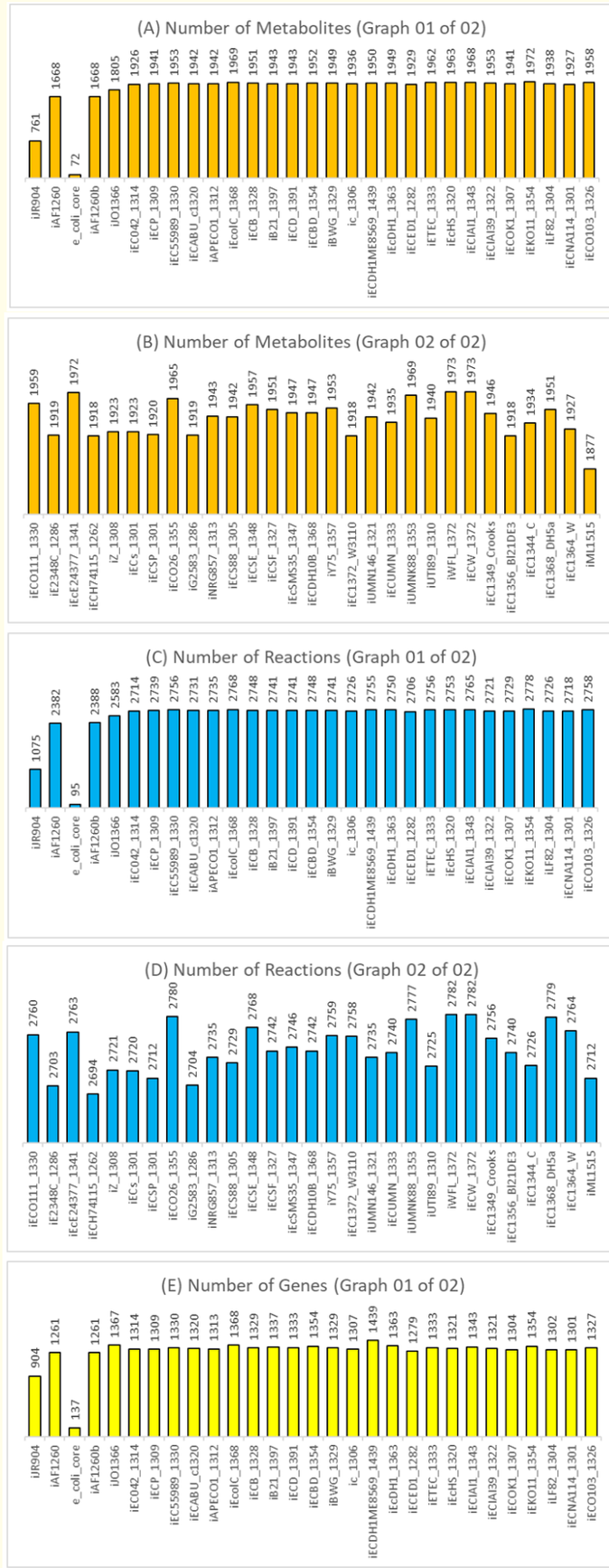
(D) iAF1260b	A slightly improved and updated version of the iAF1260 to address incorrect model predictions and most genes in the iAF1260b model were experimentally determined for conditional essentiality [26,37].
(E) iJO1366	Updated version of iAF1260 to include newly characterized genes and reactions where gaps in iAF1260 network were identified [26]. Used to evaluate RIPTiDe (Reaction Inclusion by Parsimony and Transcript Distribution), a method that utilizes the transcriptomic abundances and parsimony of overall flux for the identification of the most cost-effective usage of metabolism reflecting the cell's investment into transcription with further addition of identifying the activity of context-specific metabolic pathways without knowledge of the media conditions [38]. Comparison with its repurposed versions created from different methods and tool developments to identify key properties and differences; thus, emphasizing the importance of considering enzyme constraints in enhancements [39].
(F) iEC042_1314	GSM for *E. coli* 042 [27].
(G) iECP_1309	GSM for *E. coli* 536 [27].
(H) iEC55989_1330	GSM for *E. coli* 55989 [27].
(I) iECABU_c1320	GSM for *E. coli* ABU 83972 [27] and used to analyze probiotic *E. coli* Nissle 1917 [40].
(J) iAPECO1_1312	GSM for *E. coli* APEC O1 [27].
(K) iEcolC_1368	Predicting growth and acetate production rate of *E. coli* strains and adjusted to minimize errors in acetate production and growth rates [25].
(L) iECB_1328	GSM for *E. coli* B str. REL606 [27] and used to examine adaptive evolution of carbon source utilization [41].
(M) iB21_1397	Model growth rates of the *E. coli* strain BL21(DE3) [42].
(N) iECD_1391	GSM for *E. coli* BL21(DE3) [43].
(O) iECBD_1354	GSM for *E. coli* 'BL21-Gold(DE3)pLysS AG' strain [27].
(P) iBWG_1329	GSM for *E. coli* BW2952 [44].
(Q) ic_1306	GSM for uropathogenic *E. coli* CFT073 [27].
(R) iECDH1ME8569_1439	GSM for *E. coli* DH1 [27].
(S) iEcDH1_1363	GSM for *E. coli* DH1 [27] for fatty acids production [45].
(T) iECED1_1282	GSM for *E. coli* ED1a [27].
(U) iETEC_1333	GSM for *E. coli* ETEC H10407 [27].
(V) iEcHS_1320	GSM for *E. coli* HS [27].
(W) iECIAI1_1343	GSM for *E. coli* IAI1 [27].
(X) iECIAI39_1322	GSM for *E. coli* IAI39 [27].
(Y) iECOK1_1307	GSM for *E. coli* IHE3034 [27].
(Z) iEKO11_1354	GSM for *E. coli* KO11FL [27].
(AA) iLF82_1304	GSM for *E. coli* LF82 [27].
(AB) iECNA114_1301	GSM for *E. coli* NA114 [27].
(AC) iECO103_1326	GSM for *E. coli* O26:H11 str. 11368 [27].
(AD) iECO111_1330	GSM for *E. coli* O111:H- str. 11128 [27].
(AE) iE2348C_1286	GSM for *E. coli* O127:H6 str. E2348/69 [27].
(AF) iEcE24377_1341	GSM for *E. coli* O139:H28 str. E24377A [27].
(AG) iECH74115_1262	GSM for *E. coli* O157:H7 str. EC4115 [27].
(AH) iZ_1308	GSM for *E. coli* O157:H7 str. EDL933 [27] and has been utilised in gene knock out optimization and is comparable quantitatively with the model iJO1366 [46].
(AI) iECs_1301	GSM for enterohemorrhagic *E. coli* O157:H7 str. Sakai and has been used to test the robustness of the GSMs [47].
(AJ) iECSP_1301	GSM for enterohemorrhagic *E. coli* O157:H7 str. TW14359 [27].

(AK) iECO26_1355	GSM for *E. coli* O26:H11 str. 11368 [27].
(AL) iG2583_1286	GSM for *E. coli* O55:H7 str. CB9615 [27].
(AM) iNRG857_1313	GSM for *E. coli* O83:H1 str. NRG 857C [27].
(AN) iECS88_1305	GSM for *E. coli* S88 [27].
(AO) iECSE_1348	GSM for *E. coli* SE11 [27].
(AP) iECSF_1327	GSM for *E. coli* SE15 [27].
(AQ) iEcSMS35_1347	GSM for *E. coli* S-M-S-3-5 [27].
(AR) iECDH10B_1368	GSM for *E. coli* K-12 str. DH10B [27].
(AS) iY75_1357	GSM for *E. coli* K-12 str. W3110 for production of 2-keto-4-hydroxybutyrate and 1,3-propanediol [28].
(AT) iEC1372_W3110	GSM for *E. coli* K-12 str. W3110 [27].
(AU) iUMN146_1321	GSM for *E. coli* UM146 [27].
(AV) iECUMN_1333	GSM for *E. coli* UMN026 [27].
(AW) iUMNK88_1353	GSM for *E. coli* UMNK88 [27].
(AX) iUTI89_1310	GSM for *E. coli* UTI89 [27].
(AY) iWFL_1372	GSM for *E. coli* W [27].
(AZ) iECW_1372	GSM for *E. coli* W [27].
(BA) iEC1349_Crooks	GSM for *E. coli* ATCC 8739 [48].
(BB) iEC1356_Bl21DE3	GSM for *E. coli* BL21 (DE3) [48].
(BC) iEC1344_C	GSM for *E. coli* C [48].
(BD) iEC1368_DH5a	GSM for *E. coli* DH5a [48].
(BE) iEC1364_W	GSM for *E. coli* W [48].
(BF) iML1515	Most complete genome-scale reconstruction of the metabolic network in *E. coli* K-12 MG1655 to carry out a comparative structural proteome analysis of 1122 *E. coli* strains and identify multi-strain sequence variations, as well as providing a knowledge base for integrating modelling framework bridging systems and structural biology linked to 1515 protein structures. Investigate the creation of extended metabolic models through augmenting the iML1515 model with reactions from promiscuous enzyme activity using PROXIMAL and EMMA as a prediction tool and a way to compare the generated putative derivatives with a set of metabolites documented in ECMDB [50]. Repurposed as an enzyme-constrained model, ec_iML1515, using GECKO method to assist in the clarification of intracellular mechanisms and improve the production titre of lysine while accurately predicting and designing cellular phenotypes [17].

Table 1: Purposes and Applications of Each GSM.

Number of metabolites and reactions

The average number of metabolites (Figure 1A, and 1B; 5% trimmed mean) is 1898. Models iWFL_1372 [27] and iECW_1372 [27] have the highest number of metabolites, 1973 metabolites; while e_coli_core has the lowest number of metabolites, 72 metabolites. The average number of reactions (Figure 1C, and 1D; 5% trimmed mean) is 2697, with the highest (n = 2782) in iWFL_1372 [27] and iECW_1372 [27], and lowest (n = 95) in e_coli_core [34]. Model e_coli_core is the subset of iAF1260 [32], containing only the central metabolism of *E. coli* [34], which is often used for educational purposes [51] and tool testing [52-54]. Another notable left skewed number would be 761 metabolites found in iJR904 [29], which was used to integrate and analyse the diverse datasets and to provide a more chemically accurate description of *E. coli* metabolism over GSM iJE660a. The average number of genes (Figure 1E, and 1F; 5% trimmed mean) is 1322, with iML1515 [49] being the highest (n = 1516) and e_coli_core [34] being the lowest (n = 137). Model iML1515 [49] is the most recent *E. coli* GSM model catalogued in BiGG [22], and there is a trend of increasing number of genes over the years – from 904 genes in iJR904 [29] to 1515 genes in iML1515 [49]. Hence, it is likely that this trend will continue as iML1515 [49] contains only 34.4% of the 4401 genes in *E. coli* K-12 genome [55].


(A) Number of Metabolites (Graph 01 of 02)
761 1668 72 1668 1805 1926 1941 1953 1942 1942 1969 1951 1943 1943 1952 1949 1936 1950 1949 1929 1962 1963 1968 1953 1941 1972 1938 1927 1958
iJR904 iAF1260 e_coli_core iAF1260b iJO1366 iEC042_1314 iECP_1309 iEC55989_1330 iECABU_c1320 iAPECO1_1312 iEcolC_1368 iECB_1328 iB21_1397 iECD_1391 iECBD_1354 iBWG_1329 ic_1306 iECDH1ME8569_1439 iEcDH1_1363 iECED1_1282 iETEC_1333 iEcHS_1320 iECIAI1_1343 iECIAI39_1322 iECOK1_1307 iEKO11_1354 iLF82_1304 iECNA114_1301 iECO103_1326
(B) Number of Metabolites (Graph 02 of 02)
1959 1919 1972 1918 1923 1920 1965 1919 1943 1942 1957 1951 1947 1947 1953 1918 1942 1935 1969 1940 1973 1973 1946 1918 1934 1951 1927 1877
iECO111_1330 iE2348C_1286 iECE24377_1341 iECH74115_1262 iZ_1308 iECs_1301 iECSP_1301 iECO26_1355 iG2583_1286 iNRG857_1313 iECS88_1305 iECSE_1348 iECSF_1327 iEcSMS35_1347 iECDH10B_1368 iY75_1357 iEC1372_W3110 iUMN146_1321 iECUMN_1333 iUMNK88_1353 iUTI89_1310 iWFL_1372 iECW_1372 iEC1349_Crook5 iEC1356_Bl21DE3 iEC1344_C iEC1368_DH5a iEC1364_W iML1515
(C) Number of Reactions (Graph 01 of 02)
1075 2382 95 2388 2583 2714 2739 2756 2731 2735 2768 2748 2741 2741 2748 2741 2726 2755 2750 2706 2756 2753 2765 2721 2729 2778 2726 2718 2758
iJR904 iAF1260 e_coli_core iAF1260b iJO1366 iEC042_1314 iECP_1309 iEC55989_1330 iECABU_c1320 iAPECO1_1312 iEcolC_1368 iECB_1328 iB21_1397 iECD_1391 iECBD_1354 iBWG_1329 ic_1306 iECDH1ME8569_1439 iEcDH1_1363 iECED1_1282 iETEC_1333 iEcHS_1320 iECIAI1_1343 iECIAI39_1322 iECOK1_1307 iEKO11_1354 iLF82_1304 iECNA114_1301 iECO103_1326
(D) Number of Reactions (Graph 02 of 02)
2760 2703 2763 2694 2721 2720 2712 2780 2704 2735 2729 2768 2742 2746 2742 2759 2758 2735 2740 2777 2725 2782 2782 2756 2740 2726 2779 2764 2712
iECO111_1330 iE2348C_1286 iECE24377_1341 iECH74115_1262 iZ_1308 iECs_1301 iECSP_1301 iECO26_1355 iG2583_1286 iNRG857_1313 iECS88_1305 iECSE_1348 iECSF_1327 iEcSMS35_1347 iECDH10B_1368 iY75_1357 iEC1372_W3110 iUMN146_1321 iECUMN_1333 iUMNK88_1353 iUTI89_1310 iWFL_1372 iECW_1372 iEC1349_Crook5 iEC1356_Bl21DE3 iEC1344_C iEC1368_DH5a iEC1364_W iML1515
(E) Number of Genes (Graph 01 of 02)
904 1261 137 1261 1367 1314 1309 1330 1320 1313 1368 1329 1337 1333 1354 1329 1307 1439 1363 1279 1333 1321 1343 1321 1304 1354 1302 1301 1327
iJR904 iAF1260 e_coli_core iAF1260b iJO1366 iEC042_1314 iECP_1309 iEC55989_1330 iECABU_c1320 iAPECO1_1312 iEcolC_1368 iECB_1328 iB21_1397 iECD_1391 iECBD_1354 iBWG_1329 ic_1306 iECDH1ME8569_1439 iEcDH1_1363 iECED1_1282 iETEC_1333 iEcHS_1320 iECIAI1_1343 iECIAI39_1322 iECOK1_1307 iEKO11_1354 iLF82_1304 iECNA114_1301 iECO103_1326

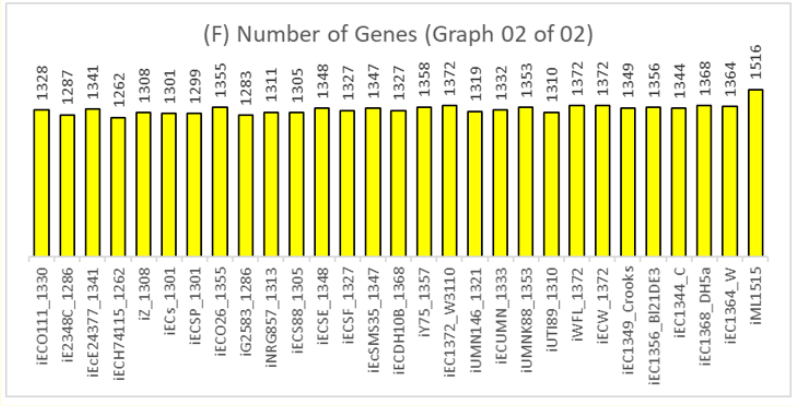


Figure 1: Statistics (Number of Metabolites, Reactions, and Genes) in Each GSM. The GSMs are arranged by year, from earliest (iJR904) to latest (iML1515).

Five media components are common in All GSMs

Five media components are common in all GSMs (Table 2) out of 29; namely, EX_h_e (proton), EX_h20_e (water), EX_nh4_e (ammonium), EX_o2_e (oxygen), and EX_pi_e (phosphate); all of which can be found in M9 minimum media [18] under oxic environments. This suggests that these 5 components are critical for aerobic functions of *E. coli*. Not taking account of core metabolism (e_coli_core) and the earliest GSM (iJR904), 14 other media components are common in the remaining 56 GSMs.

Of which, 6 media components (calcium, chloride, magnesium, potassium, sodium, sulphate] are found in M9 minimum media. The remaining 8 media components (copper, cobalt, ferrous ion, ferric ion, manganese, molybdate, tungstate, and zinc) were known to affect *E. coli* growth. Copper alone shows slight growth inhibition of *E. coli* O157:H7 but more pronounced growth inhibition in the presence of lactic acid [56]. Cobalt has been shown to induce stress to *E. coli* [57] and toxicity at high concentration [58]. An early study by Ratledger and Winder 1964 had demonstrated effects or iron (ferric and ferrous ions) and zinc on *E. coli* growth [59] and the growth supporting effects of iron had also been re-demonstrated in a more recent study [60]. The presence of manganese is suggested to have a protective effect of *E. coli* to high oxidative stress [61] by activation of manganese-superoxide dismutase [62]. Similarly, molybdate [63] and tungstate [64] has also been shown to inhibit *E. coli* growth.

Substantial differences in the media components in various models lends difficulties in conducting complete comparative study such as that of Cheong., *et al*. [18] across multiple GSM models. As a result, only the effects of glucose on growth across models is studied. However, it is surprising that glucose (EX_glc__D_e) is not found in iECIAI1_1343, which is GSM for *E. coli* IAI1 [27]. As we are interested in examining the predicted growth rate and fluxed under varying glucose condition, iECIAI1_1343 was removed from study as the model cannot respond to glucose.

On the other hand, Ex_leu_L_e (leucine), EX_thm_e (thiamine), and Ex_trp_L_e (tryptophan) is found in only one GSM each: iECDH10B_1368 [27], iECIAI39_1322 [27], and ic_1306 [27] respectively. The reason for having these specific nutrients in the media is due to auxotrophic nature of these strains from experimental or predicted findings. Leucine is found only in iECDH10B_1368, a GSM for *E. coli* K-12 substr. DH10B [27], which lacks leucine synthesis pathway; thus, requires leucine for growth on minimal medium [65]. Thiamine is found only in iECIAI39_1322, a GSM for *E. coli* IAI39, which is predicted to lack thiamine synthesis [27]. This has resulted in the routine addition of thiamine into minimum media to support growth [66]. Tryptophan is found only in ic_1306, a GSM for uropathogenic *E. coli* CFT073, which is similarly to be predicted to be auxotrophic to tryptophan [27]. This is partly supported by an earlier study suggesting that *E. coli* CFT073 catalyzes tryptophan during serine and aspartate depletion [67].

Substantial differences between predicted growth rates from varying glucose uptake rates across 58 GSMs

Of the 3192 pairwise permutation of possible non-self-comparison paired t-tests (Figure 2), 74.25% (n = 2370) are significant (p-value < 0.05); of which, 2036 paired t-tests are highly significant (p-value < 0.01). Hence, only 25.75% (n = 822) are not sig-

nificant (p-value > 0.05). This suggests substantial differences between normalized predicted growth rates across the 57 GSMs. As a whole, the average number of GSMs not significantly different from a specific GSM in terms of predicted growth rate is 14.4 (Figure 3), with iECP_1309 [27] and ic_1306 [27] having the least differences (not significantly different with 49 other GSMs) and most differences (significantly different from all 57 other GSMs) respectively. In terms of phylogram analysis based on Manhattan distance [68], it is interesting to note that iAF1260 [32] and iAF1260b [26,37] are most different to the other 56 GSMs (Figure 4). However, iAF1260b [26,37] can be seen as a slight improvement and correction of iAF1260 [32]; hence, not surprising that they are clustered together. Nevertheless, these results suggest significant differences between the 58 examined GSMs. This suggests that GSMs of the same species may vary substantially, which is supported Nouri., *et al*. [69] compared the prediction of 3 GSMs for *Zymomonas mobilis* ZM4 and reported demonstrating differences between various models of the same species.

Media Components	Name	Found or Not Found in GSMs in this Review
EX_h_e	H^+ exchange	Found in all GSMs
EX_h2o_e	H_2O exchange	Found in all GSMs
EX_nh4_e	Ammonia exchange	Found in all GSMs
EX_o2_e	O_2 exchange	Found in all GSMs
EX_pi_e	Phosphate exchange	Found in all GSMs
EX_co2_e	CO_2 exchange	Found in all GSMs except iECIAI1_1343
EX_fe2_e	Fe^{2+} exchange	Found in all GSMs except e_coli_core
EX_glc__D_e	D-Glucose exchange	Found in all GSMs except iECIAI1_1343
EX_k_e	K^+ exchange	Found in all GSMs except e_coli_core
EX_na1_e	Sodium exchange	Found in all GSMs except e_coli_core
EX_so4_e	Sulfate exchange	Found in all GSMs except e_coli_core
EX_ca2_e	Calcium exchange	Found in all GSMs except e_coli_core, and iJR904
EX_cl_e	Chloride exchange	Found in all GSMs except e_coli_core, and iJR904
EX_cobalt2_e	Co^{2+} exchange	Found in all GSMs except e_coli_core, and iJR904
EX_cu2_e	Cu^{2+} exchange	Found in all GSMs except e_coli_core, and iJR904
EX_fe3_e	Fe^{3+} exchange	Found in all GSMs except e_coli_core, and iJR904
EX_mg2_e	Mg exchange	Found in all GSMs except e_coli_core, and iJR904
EX_mn2_e	Mn^{2+} exchange	Found in all GSMs except e_coli_core, and iJR904
EX_mobd_e	Molybdate exchange	Found in all GSMs except e_coli_core, and iJR904
EX_tungs_e	Tungstate exchange	Found in all GSMs except e_coli_core, and iJR904
EX_zn2_e	Zinc exchange	Found in all GSMs except e_coli_core, and iJR904
EX_cbl1_e	Cob(I)alamin exchange	Found in all GSMs except e_coli_core, iJR904, iML1515, and iECIAI1_1343
EX_ni2_e	Ni^{2+} exchange	Found in all GSMs except e_coli_core, iJR904, iAF1260, and iAF1260b
EX_sel_e	Selenate exchange	Found in all GSMs except e_coli_core, iJR904, iAF1260, and iAF1260b
EX_slnt_e	Selenite exchange	Found in all GSMs except e_coli_core, iJR904, iAF1260, and iAF1260b
EX_nac_e	Nicotinate exchange	Not found in all GSMs except iECDH10B_1368, and iECUMN_1333
EX_leu__L_e	L-Leucine exchange	Not found in all GSMs except iECDH10B_1368
EX_thm_e	Thiamin exchange	Not found in all GSMs except iECIAI39_1322
EX_trp__L_e	L-Tryptophan exchange	Not found in all GSMs except ic_1306

Table 2: Comparison of Media Components.

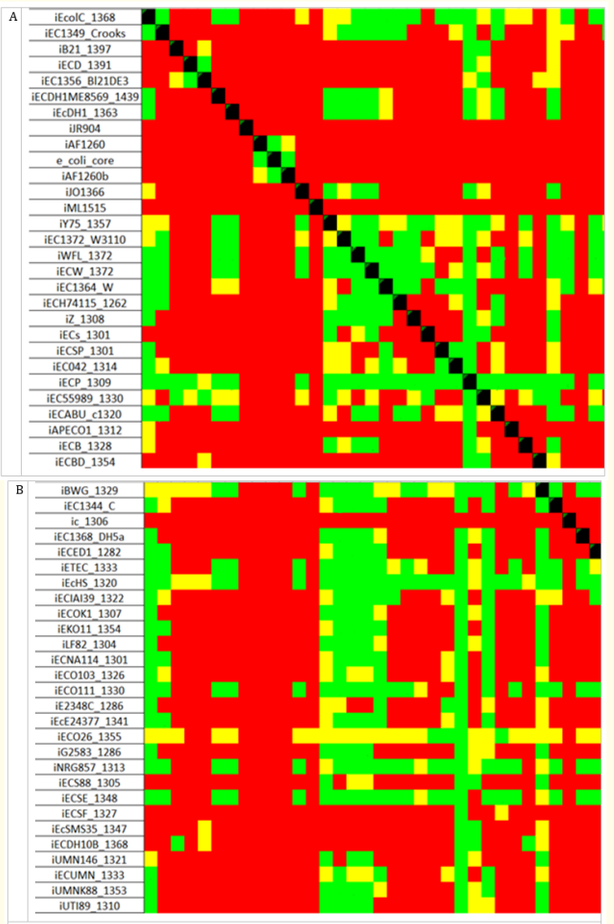


Figure 2: 3192 Paired t-test P-values of Predicted Normalized Growth Rates Across 57 GSMs. Panel A and B are consecutive images. Green represents p-value more than 0.05. Yellow represents p-value between 0.01 and 0.05. Red represents p-value less than 0.01. Black represents self-comparison.

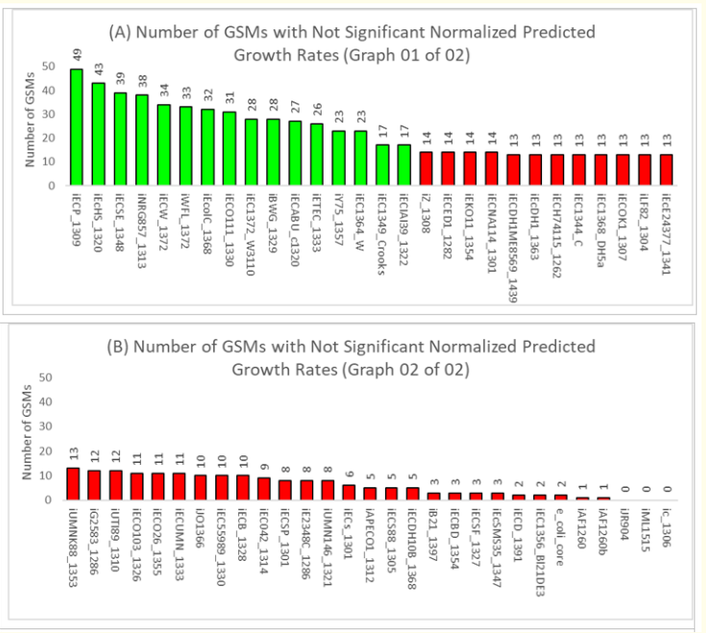


Figure 3: Number of GSMs Not Significantly Different to a Specific GSM by Normalized Predicted Growth Rates. The average number of non-significances is 14.4; hence, the green bars are GSMs with more similarity than average while the red bars are GSMs with less similarity than average in terms of normalized predicted growth rates.

Figure 4: Phylogram of GSMs Based on Normalized Growth Rates. Distance matrix was calculated using Manhattan distance [68].

34 of 42 pairwise comparisons show substantial differences between predicted growth rates from GSMs for the same strain

Paired t-test analyses between predicted growth rates within the same strain demonstrated 34 of the 42 paired t-tests to be significant (Table 3). The two GSMs (iEcolC_1368, and iEC1349_Crooks) for *E. coli* ATCC 8739 shows insignificant difference in normalized predicted growth rate (p-value = 0.350). Of the 3 GSMs (iB21_1397, iECD_1391, and iEC1356_Bl21DE3) for *E. coli* BL21(DE3), only between iECD_1391 and iEC1356_Bl21DE3 is not significant (p-value = 0.232) – the other 2 pairwise comparisons are significant (p-value ≤ 0.0240). The two GSMs (iECDH1ME8569_1439, and iEcDH1_1363) for *E. coli* ATCC DH1 shows significant difference in normalized predicted growth rate (p-value = 0.00112). Between the 8 GSMs (iJR904, iAF1260, e_coli_core, iAF1260b, iJO1366, iML1515, iY75_1357, and iEC1372_W3110) for *E. coli* K-12; only 3 comparisons, (a) iAF1260 and e_coli_core (p-value = 0.287), (b) e_coli_core and iAF1260b (p-value = 0.293), and (c) iJO1366 and iY75_1357 (p-value = 0.158); are not significant – the rest of the 25 pairwise comparisons are significant (p-value ≤ 0.0421). The normalized predicted growth rates of the 3 GSMs (iWFL_1372, iECW_1372, and iEC1364_W) for *E. coli* W are not significant (p-value ≥ 0.101). Finally, the normalized predicted growth rates of the 4 GSMs (iECH74115_1262, iZ_1308, iECs_1301, and iECSP_1301) for *E. coli* O157:H7 are significant (p-value ≤ 0.00112).

These findings are consistent with our result comparing across all 58 GSMs (Figure 2 to 4). More importantly, these findings also show that GSMs for the same bacterial strain may not yield similar results. This may be a result of differences in experimental setup to obtain the required metabolomic data for model construction despite using the same strain as Shiratsubaki., *et al*. [70] had demon-

strated that GSMs built for specific developmental stages can result in result in different prediction results. Furthermore, differences between *in vivo* experimental outcomes and *in silico* predictions from GSMs are commonplace [18,19,69,70], which had resulted in methods aiming at reconciliating such differences [71]. However, such differences in GSMs may also provide a fertile basis for examining the differences between various *E. coli* strains [70] or to construct a strain-independent *E. coli* model [69,72].

***E. coli* Strain**	**Comparison of Growth Rates Across Various Glucose Uptake Rates**		**Paired t-test p-value**
	GSM 1	**GSM 2**	
ATCC 8739	iEcolC_1368	iEC1349_Crooks	0.350
BL21 (DE3)	iB21_1397	iECD_1391	0.001
	iB21_1397	iEC1356_Bl21DE3	0.024
	iECD_1391	iEC1356_Bl21DE3	0.232
DH1	iECDH1ME8569_1439	iEcDH1_1363	0.001
K-12	iJR904	iAF1260	0.003
		e_coli_core	0.001
		iAF1260b	0.003
		iJO1366	0.001
		iML1515	0.001
		iY75_1357	0.001
		iEC1372_W3110	0.001
	iAF1260	e_coli_core	0.287
		iAF1260b	0.022
		iJO1366	1.77E-05
		iML1515	1.17E-04
		iY75_1357	1.71E-05
		iEC1372_W3110	1.78E-05
	e_coli_core	iAF1260b	0.293
		iJO1366	0.001
		iML1515	0.001
		iY75_1357	0.001
		iEC1372_W3110	0.001
	iAF1260b	iJO1366	1.72E-05
		iML1515	1.12E-04
		iY75_1357	1.65E-05
		iEC1372_W3110	1.72E-05
	iJO1366	iML1515	0.001
		iY75_1357	0.158
		iEC1372_W3110	0.036
	iML1515	iY75_1357	0.001
		iEC1372_W3110	0.001
	iY75_1357	iEC1372_W3110	0.042
W	iWFL_1372	iECW_1372	0.302
		iEC1364_W	0.286
	iECW_1372	iEC1364_W	0.102
O157:H7	iECH74115_1262	iZ_1308	0.001
		iECs_1301	0.001
		iECSP_1301	0.001
	iZ_1308	iECs_1301	0.001
		iECSP_1301	0.001
	iECs_1301	iECSP_1301	0.001

Table 3: Paired t-test Analysis of Normalized Growth Rates Within Each Strain.

Conclusion

In this study, 58 GSMs for *E. coli* were compared for their media components and predicted growth rates across various glucose uptake rates. Substantial differences in media components were found and our results suggest significant differences in predicted growth rates between the GSMs and even within GSMs constructed for the same strain.

Supplementary Materials

Data files for this study can be downloaded from https://bit.ly/CompareECO_GSM.

Acknowledgement

This work was conducted as part of the Temasek Polytechnic School of Applied Science Differential Research Project and Major Project under the Student Project fund (TP_PR1052).

Conflict of Interest

The authors declare no conflict of interest.

Bibliography

1. Idalia V-MN and Bernardo F. "Escherichia coli as a Model Organism and Its Application in Biotechnology". In: Samie A, editor. Escherichia coli - Recent Advances on Physiology, Pathogenesis and Biotechnological Applications. InTech (2017).

2. Yang D., *et al*. "Metabolic Engineering of Escherichia coli for Natural Product Biosynthesis". *Trends in Biotechnology* 38.7 (2020): 745-765.

3. García-Granados R., *et al*. "Metabolic Engineering and Synthetic Biology: Synergies, Future, and Challenges". *Frontiers in Bioengineering and Biotechnology* 7 (2019): 36.

4. Murthy MV., *et al*. "UniKin1: A Universal, Non-Species-Specific Whole Cell Kinetic Model". *Acta Scientific Microbiology* 3.10 (2020): 04-08.

5. Cho JL and Ling MH. "Adaptation of Whole Cell Kinetic Model Template, UniKin1, to Escherichia coli Whole Cell Kinetic Model, ecoJC20". *EC Microbiology* 17.2 (2021): 254-260.

6. Comba S., *et al*. "Emerging Engineering Principles for Yield Improvement in Microbial Cell Design". *Computational and Structural Biotechnology Journal* 3.4 (2012): e201210016.

7. Cardoso JGR., *et al*. "Cameo: A Python Library for Computer Aided Metabolic Engineering and Optimization of Cell Factories". *ACS Synthetic Biology* 7.4 (2018): 1163-1166.

8. Ling MH. "AdvanceSyn Toolkit: An Open Source Suite for Model Development and Analysis in Biological Engineering". *MOJ Proteomics Bioinformation* 9.4 (2020): 83-86.

9. O'Brien EJ., *et al*. "Using Genome-scale Models to Predict Biological Capabilities". *Cell* 161.5 (2015): 971-987.

10. Srinivasan S., *et al*. "Constructing Kinetic Models of Metabolism at Genome-Scales: A Review". *Biotechnology Journal* 10.9 (2015): 1345-59.

11. Simeonidis E and Price ND. "Genome-Scale Modeling for Metabolic Engineering". *Journal of Industrial Microbiology and Biotechnology* 42.3 (2015): 327-338.

12. Gu C., *et al*. "Current Status and Applications of Genome-Scale Metabolic Models". *Genome Biology* 20.1 (2019): 121.

13. Kim B., *et al*. "Applications of Genome-Scale Metabolic Network Model in Metabolic Engineering". *Journal of Industrial Microbiology and Biotechnology* 42.3 (2015): 339-348.

14. Kocabaş P., *et al*. "Analyses of Extracellular Protein Production in Bacillus subtilis - II: Responses of Reaction Network to Oxygen Transfer at Transcriptional Level". *Biochemical Engineering Journal* 127 (2017): 242-261.

15. Kavvas ES., *et al*. "Updated and Standardized Genome-Scale Reconstruction of Mycobacterium tuberculosis H37Rv, iEK1011, Simulates Flux States Indicative of Physiological Conditions". *BMC Systems Biology* 12.1 (2018): 25.

16. Edwards JS and Palsson BO. "The Escherichia coli MG1655 In Silico Metabolic Genotype: Its Definition, Characteristics, and Capabilities". *Proceedings of the National Academy of Sciences of the United States of America* 97.10 (2000): 5528-5533.

17. Ye C., *et al*. "Improving Lysine Production Through Construction of an Escherichia coli Enzyme-Constrained Model". *Biotechnology and Bioengineering* 117.11 (2020): 3533-3544.

18. Cheong KC., *et al*. "A Simulation Study on the Effects of Media Composition on the Growth Rate of Escherichia coli MG1655 Using iAF1260 Model". *Acta Scientific Microbiology* 3.8 (2020): 40-44.

19. Chang ED and Ling MH. "Explaining Monod in terms of Escherichia coli metabolism". *Acta Scientific Microbiology* 2.9 (2019): 66-71.

20. Mienda BS. "Escherichia coli Genome-Scale Metabolic Gene Knockout of Lactate Dehydrogenase (ldhA), Increases Succinate Production from Glycerol". *Journal of Biomolecular Structure and Dynamics* 36.14 (2018): 3680-3686.

21. Immanuel SRC., *et al*. "Integrated Constraints Based Analysis of An Engineered Violacein Pathway in Escherichia coli". *Biosystems* 171 (2018): 10-19.

22. King ZA., *et al*. "BiGG Models: A Platform for Integrating, Standardizing and Sharing Genome-Scale Models". *Nucleic Acids Research* 44.D1 (2016): D515-522.

23. Feist AM and Palsson BO. "The Biomass Objective Function". *Current Opinion in Microbiology* 13.3 (2010): 344-349.

24. Orth JD., *et al*. "What is Flux Balance Analysis?" *Nature Biotechnology* 28.3 (2010): 245-248.

25. Peterson JR., *et al*. "Parametric Studies of Metabolic Cooperativity in Escherichia coli Colonies: Strain and Geometric Confinement Effects". *PLoS ONE* 12.8 (2017): e0182570.

26. Orth JD., *et al*. "A Comprehensive Genome-Scale Reconstruction of Escherichia coli Metabolism — 2011". *Molecular Systems Biology* 7.1 (2011): 535.

27. Monk JM., *et al*. "Genome-Scale Metabolic Reconstructions of Multiple Escherichia coli Strains Highlight Strain-Specific Adaptations to Nutritional Environments". *Proceedings of the National Academy of Sciences of the United States of America* 110.50 (2013): 20338-20343.

28. Wang C., *et al*. "An Aldolase-Catalyzed New Metabolic Pathway for the Assimilation of Formaldehyde and Methanol To Synthesize 2-Keto-4-hydroxybutyrate and 1,3-Propanediol in Escherichia coli". *ACS Synthetic Biology* 8.11 (2019): 2483-2493.

29. Reed JL., *et al*. "An Expanded Genome-Scale Model of Escherichia coli K-12 (iJR904 GSM/GPR)". *Genome Biology* 4.9 (2003): R54.

30. Costa RS and Vinga S. "Assessing Escherichia coli Metabolism Models and Simulation Approaches in Phenotype Predictions: Validation Against Experimental Data". *Biotechnology Progress* 34.6 (2018): 1344-1354.

31. Liu L., *et al*. "DEF: An Automated Dead-End Filling Approach Based On Quasi-Endosymbiosis". *Bioinformatics* 33.3 (2017): 405-413.

32. Feist AM., *et al*. "A Genome-Scale Metabolic Reconstruction for Escherichia coli K-12 MG1655 that Accounts for 1260 ORFs and Thermodynamic Information". *Molecular Systems Biology* 3 (2007): 121.

33. Tamura T. "Grid-Based Computational Methods for the Design of Constraint-Based Parsimonious Chemical Reaction Networks to Simulate Metabolite Production: GridProd". *BMC Bioinformatics* 19.1 (2018): 325.

34. Orth J., *et al*. "Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide". *EcoSal Plus* 4.1 (2010): ecosalplus.10.2.1.

35. de Arroyo Garcia L and Jones PR. "In Silico Co-Factor Balance Estimation Using Constraint-Based Modelling Informs Metabolic Engineering in Escherichia coli". *PLOS Computational Biology* 16.8 (2020): e1008125.

36. Apaydin M., *et al*. "Robust Mutant Strain Design by Pessimistic Optimization". *BMC Genomics* 18.S6 (2017): 677.

37. Feist AM., *et al*. "Model-Driven Evaluation of the Production Potential for Growth-Coupled Products of Escherichia coli". *Metabolic Engineering* 12.3 (2010): 173-186.

38. Jenior ML., *et al*. "Transcriptome-Guided Parsimonious Flux Analysis Improves Predictions with Metabolic Networks in Complex Environments". *PLOS Computational Biology* 16.4 (2020): e1007099.

39. Bekiaris PS and Klamt S. "Automatic Construction of Metabolic Models with Enzyme Constraints". *BMC Bioinformatics* 21.1 (2020): 19.

40. Kim D., *et al*. "Development of a Genome-Scale Metabolic Model and Phenome Analysis of the Probiotic Escherichia coli Strain Nissle 1917". *International Journal of Molecular Sciences* 22.4 (2021): 2122.

41. Chu HY., *et al*. "Assessing the Benefits of Horizontal Gene Transfer by Laboratory Evolution and Genome Sequencing". *BMC Evolutionary Biology* 18.1 (2018): 54.

42. Hosseini S-R and Wagner A. "Genomic Organization Underlying Deletional Robustness in Bacterial Metabolic Systems". *Proceedings of the National Academy of Sciences of the United States of America* 115.27 (2018): 7075-7080.

43. Kim H., *et al*. "Metabolic Network Reconstruction and Phenome Analysis of the Industrial Microbe, Escherichia coli BL21 (DE3)". *PLOS ONE* 13.9 (2018): e0204375.

44. Zabeti H., *et al*. "A Duality-Based Method for Identifying Elemental Balance Violations in Metabolic Network Models". In: 18th International Workshop on Algorithms in Bioinformatics (WABI 2018). Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany (2018): 1-13. (Leibniz International Proceedings in Informatics (LIPIcs)).

45. Wu G. "Revelation of Yin-Yang Balance in Microbial Cell Factories by Data Mining, Flux Modeling, and Metabolic Engineering [Doctor of Philosophy]". [School of Engineering and Applied Science]: Washington University (2016).

46. Patané A., *et al*. "Multi-Objective Optimization of Genome-Scale Metabolic Models: The Case of Ethanol Production". *Annals of Operations Research* 276.1-2 (2019): 211-227.

47. Gerstl MP., *et al*. "Exact Quantification of Cellular Robustness in Genome-Scale Metabolic Networks". *Bioinformatics* 32.5 (2016): 730-737.

48. Monk JM., *et al*. "Multi-Omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes". *Cell System* 3.3 (2016): 238-251.e12.

49. Monk JM., *et al*. "iML1515, A Knowledgebase that Computes Escherichia coli Traits". *Nature Biotechnology* 35.10 (2017): 904-908.

50. Amin SA., *et al*. "Towards Creating An Extended Metabolic Model (EMM) for E. coli Using Enzyme Promiscuity Prediction and Metabolomics Data". *Microbe Cell Factories* 18.1 (2019): 109.

51. Hädicke O and Klamt S. "EColiCore2: A Reference Network Model of the Central Metabolism of Escherichia coli and Relationships to its Genome-Scale Parent Model". *Scientific Report* 7.1 (2017): 39647.

52. Clement TJ., *et al*. "Unlocking Elementary Conversion Modes: ecmtool Unveils All Capabilities of Metabolic Networks". *Patterns* 2.1 (2021): 100177.

53. Heinonen M., *et al*. "Bayesian Metabolic Flux Analysis Reveals Intracellular Flux Couplings". *Bioinformatics* 35.14 (2019): i548-557.

54. Venayak N., *et al*. "MoVE Identifies Metabolic Valves to Switch Between Phenotypic States". *Nature Communication* 9.1 (2018): 5332.

55. Serres MH., *et al*. "A Functional Update of the Escherichia coli K-12 Genome". *Genome Biology* 2.9 (2001): RESEARCH0035.

56. Gyawali R., *et al*. "Antimicrobial Activity of Copper Alone and in Combination with Lactic Acid against Escherichia coli O157:H7 in Laboratory Medium and on the Surface of Lettuce and Tomatoes". *Journal of Pathogens* 2011 (2011): 650968.

57. Ranquet C., *et al*. "Cobalt Stress in Escherichia coli". *Journal of Biological Chemistry* 282.42 (2007): 30442-30451.

58. Majtan T., *et al*. "Effect of Cobalt on Escherichia coli Metabolism and Metalloporphyrin Formation". *BioMetals. An International Journal on the Role of Metal Ions in Biology, Biochemistry and Medicine* 24.2 (2011): 335-347.

59. Ratledge C., *et al*. "Effect of Iron and Zinc on Growth Patterns of Escherichia coli in Iron-Deficient Medium". *Journal of Bacteriology* 87 (1964): 823-827.

60. Appenzeller BMR., *et al*. "Advantage Provided by Iron for Escherichia coli Growth and Cultivability in Drinking Water". *Applied and Environmental Microbiology* 71.9 (2005): 5621-5623.

61. McEwan AG. "New Insights into the Protective Effect of Manganese Against Oxidative Stress". *Molecular Microbiology* 72.4 (2009): 812-814.

62. Baez A and Shiloach J. "Escherichia coli Avoids High Dissolved Oxygen Stress by Activation of SoxRS and Manganese-Superoxide Dismutase". *Microbe Cell Factories* 12.1 (2013): 23.

63. Mardare CC., *et al*. "Growth Inhibition of Escherichia coli by Zinc Molybdate with Different Crystalline Structures: Growth Inhibition of Escherichia coli by Zinc Molybdate". *Physica Status Solidi* 213.6 (2016): 1471-1478.

64. Gates AJ., *et al*. "Properties of the Periplasmic Nitrate Reductases from Paracoccus pantotrophus and Escherichia coli after Growth in Tungsten-Supplemented Media". *FEMS Microbiology Letter* 220.2 (2003): 261-269.

65. Durfee T., *et al*. "The Complete Genome Sequence of Escherichia coli DH10B: Insights into the Biology of a Laboratory Workhorse". *Journal of Bacteriology* 190.7 (2008): 2597-2606.

66. Skovgaard O., *et al*. "Genome-Wide Detection of Chromosomal Rearrangements, Indels, and Mutations in Circular Chromosomes by Short Read Sequencing". *Genome Research* 21.8 (2011): 1388-1393.

67. Anfora AT., *et al*. "Uropathogenic Escherichia coli CFT073 is Adapted to Acetatogenic Growth But Does Not Require Acetate During Murine Urinary Tract Infection". *Infection and Immunity* 76.12 (2008): 5760-5767.

68. Fürnkranz J, ., *et al*. "Manhattan Distance". In: Sammut C, Webb GI, editors. Encyclopedia of Machine Learning. Boston, MA: Springer US (2011): 639-639.

69. Nouri H., *et al*. "A Reconciliation of Genome-Scale Metabolic Network Model of Zymomonas mobilis ZM4". *Scientific Report* 10.1 (2020): 7782.

70. Shiratsubaki IS., *et al*. "Genome-Scale Metabolic Models Highlight Stage-Specific Differences in Essential Metabolic Pathways in Trypanosoma cruzi". *PLOS Neglected Tropical Diseases* 14.10 (2020): e0008728.

71. van Duuren JBJH., *et al*. "Reconciling In Vivo and In Silico Key Biological Parameters of Pseudomonas putida KT2440 During Growth on Glucose Under Carbon-Limited Condition". *BMC Biotechnology* 13 (2013): 93.

72. Contador CA., *et al*. "Use of Genome-Scale Models to Get New Insights into the Marine Actinomycete Genus Salinispora". *BMC Systems Biology* 13.1 (2019): 11.