Revision Manuscript BIORXIV-2024-629819

Use of Genome Scale Metabolic Reconstructions of Avian Pathogenic *Escherichia coli* (APEC) phylogroups for the identification of lineage-specific metabolic pathways

Huijun Long [1], Jai W. Mehat [2], HuiHai Wu [2,3], Arnoud H. M. van Vliet [1,*], Roberto M. La Ragione [1,2*]

1. Department of Comparative Biomedical Sciences, School of Veterinary Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK
2. Discipline of Microbes, Infection and Immunity, School of Biosciences, Faculty of Health and Medical Sciences, University of Surrey, Guildford, UK
3. Bioinfopipe Ltd, London, UK

* Corresponding authors. Email: a.vanvliet@surrey.ac.uk, r.laragione@surrey.ac.uk

ORCID: Jai W. Mehat: 0000-0002-9563-6348, Arnoud H. M. van Vliet: 0000-0003-0203-1305, Roberto M. La Ragione: 0000-0001-5861-613X.

Running title: Metabolic models of Avian Pathogenic *Escherichia coli*

Keywords: Avian pathogenic *Escherichia coli* (APEC), *in silico* genome scale metabolic model, phylogroup, pathotype, 3-hydroxyphenylacetate, 3-HPAA

Data statement. All supporting data, code and protocols have been provided within the

article or through supplementary data files and/or via Figshare. Five supplementary figures and nine supplementary tables are available with the online version of this article.

Data summary. All genome sequencing data (sequencing reads and genome assemblies) are available at the Sequence Read Archive (SRA) and Genomes repositories at NCBI, in Bioproject number PRJNA523197. The information on the individual APEC isolates has been included in Supplementary Material 2 (Table S1), whereas data on gene essentiality, Biolog measurements and model growth simulations have been included in Supplementary Material 2 (Tables S2-S4), respectively. The reactions included in the APEC GEM and phylogroup-specific sub-models, the media pfile versions for submodels, carbon, sulphate, phosphate and the *lysA* and *potE* mutants have been made available via Figshare (https://doi.org/10.6084/m9.figshare.28067993), with the content of these files included as Supplementary Material 3-5 (Tables S5-S7), respectively.

Funding information. This work received no specific grant from any funding agency.

Conflict of interest. The authors declare that there are no conflicts of interest.

Author contributions. R.M.L.R and A.H.M.v.V. designed the study. H.L., J.W.M. and H.W. acquired data. H.L. and H.W. created the metabolic models, H.L., J.W.M, H.W. and A.H.M.v.V. carried out analyses and interpreted data. H.L. drafted the manuscript with input from J.W.M, H.W., A.H.M.v.V. and R.M.L.R. All authors critically reviewed and approved the final version of the manuscript.

Ethical statement. Not applicable.

51

52 **Abbreviations.** SRA, Sequence Read Archive; APEC, Avian Pathogenic *Escherichia coli*;
53 ExPEC, Extra-Intestinal Pathogenic *E. coli*; GEM, Genome scale metabolic reconstruction;
54 AMR, antimicrobial resistance; AMP, ampicillin; Cm, chloramphenicol; CAT,
55 chloramphenicol resistance cassette; PCR, Polymerase Chain Reaction; RT-qPCR, reverse
56 transcription quantitative-PCR; 3-HPAA, 3-hydroxyphenylacetate; SBC, stoichiometrically
57 balanced cycles; BCR, biomass consumption reactions; FBA, flux balance analysis.

58

59

ABSTRACT

Avian Pathogenic *Escherichia coli* (APEC) are a genetically diverse pathotype primarily associated with extra-intestinal infections in birds. APEC lineages are predicted to have unique metabolic capabilities contributing to virulence and survival in the host environment. Here we present a genome-scale metabolic model for the APEC pathotype based on 114 APEC genome sequences, and lineage-specific models for the phylogroups B2, C and G based on a representative isolate for each phylogroup. A total of 1,848 metabolic reactions were predicted in the 114 APEC isolates before gap filling and manual correction. Of these, 89% represented core reactions, whilst the 11% accessory reactions were mostly associated with carbon and nitrogen metabolism. Predictions of auxotrophy were confirmed by inactivation of the conditionally essential *lysA* and the non-essential *potE* genes. The APEC metabolic model outperformed the *E. coli* K-12 *i*JO1366 model in the Biolog Phenotypic Array platform. Sub-models specific for phylogroups B2, C and G predicted differences in the metabolism of 3-hydroxyphenylacetate (3-HPAA), a phenolic acid derived from the flavonoid quercetin, which is commonly added to poultry feed. Two 3-HPAA associated reactions/genes distinguished APEC phylogroup C from APEC phylogroups B2 and G, and 3-HPAA supported the growth of APEC phylogroup C in minimal media, but not phylogroups B2 and G. In conclusion, we have constructed genome-scale metabolic models for the three major APEC phylogroups B2, C and G, and have identified a metabolic pathway distinguishing phylogroup C APEC. This demonstrates the importance of lineage- and pathotype-specific metabolic models when investigating genetically diverse microbial pathogens.

83 **IMPACT STATEMENT**

84 Avian Pathogenic *Escherichia coli* (APEC) are the cause of colibacillosis in poultry,
85 which results in a significant economic burden to the poultry industry, and strongly affects
86 the health and welfare of flocks. APEC isolates show a high level of genetic diversity, which
87 complicates diagnostics, epidemiology and the design of prevention and treatment strategies.
88 In this study we have used genome sequences derived from 114 APEC isolates to investigate
89 their metabolic capabilities, and define the metabolic diversity of APEC within a generalised
90 APEC metabolic model, and lineage-specific metabolic models. These models have been
91 interrogated to find unique pathways that can be targeted for the development of anti-APEC
92 treatments, and one such metabolic pathway was identified as a proof of principle. This
93 approach shows great promise for the design of future strategies to prevent and deal with
94 APEC infections, and can be adapted to other genetically diverse microbial pathogens.
95

96 **INTRODUCTION**

97 Avian pathogenic *Escherichia coli* (APEC) belong to the extra-intestinal pathogenic
98 (ExPEC) group of *E. coli* [1, 2]. APEC cause extra-intestinal infections in birds, commonly
99 termed colibacillosis, a collective name describing a range of systemic and localised
100 syndromes such as pericarditis, perihepatitis, peritonitis, and airsacculitis. Colibacillosis
101 causes significant economic losses and is a major welfare concern in the poultry industry [3].
102 The APEC pathotype is defined by the diseases caused in birds, with *E. coli* isolated from
103 extra-intestinal locations in birds with colibacillosis-associated diseases being candidate
104 APEC, pending further characterisation. APEC are found in all major *E. coli* lineages and
105 phylogroups and encompass a range of serogroups and MLST sequence types [4-7], and it
106 this genetic variation and imprecise definition which complicates the development and
107 efficacy of control strategies, such as antimicrobial administration and vaccine development
108 [6]. Therefore, an improved understanding of the biology and variation of this pathogen will
109 be beneficial to the development of novel and effective control strategies.

110 It has been proposed that different pathogenic *E. coli* lineages, in addition to commensal
111 lineages, harbour variable metabolic profiles [8-10]. However, there remains a paucity of
112 knowledge regarding the unique metabolic traits of APEC isolates, as most studies have
113 focused on identification of markers allowing distinction of APEC from avian faecal *E. coli*
114 [7]. Genome scale metabolic reconstructions (GEMs) are a powerful tool that have been
115 applied in a number of fields [11]. Essentially, GEMs use genome sequences and database
116 comparisons to predict metabolic functions and responses in a range of environmental
117 conditions. To date, most *E. coli* GEMs have been constructed from a single commensal
118 isolate, such as K-12 MG1655 [12-15], probiotic isolate Nissle 1917 [16] and *E. coli* W
119 (ATCC 9637) [17]. However, as increasing numbers of *E. coli* genome sequences are now
120 available, it has become apparent that GEMs of single commensal isolates only partially

121 represent the metabolic capability of this species and its different pathotypes [9]. Similarly, it
122 was reported that GEMs based on multiple isolates were able to differentiate the metabolic
123 profile of pathogenic from commensal *E. coli* isolates [9]. Hence it is clear that single
124 laboratory strain-based models are not adequate to represent the metabolic capability and
125 variation within *E. coli* pathotypes.

126 In this study, we have constructed a comprehensive metabolic genome scale model based
127 on a panel of 114 APEC isolates. This general APEC model was then used as a template to
128 develop specific models that recapitulate the metabolism of the major phylogroups of the
129 APEC pathotype. Models were queried for auxotrophy and gene essentiality, growth
130 predictions in diverse media and used to identify a metabolic pathway distinguishing
131 phylogroup C APEC from phylogroup B2 and G APEC, demonstrating the potential for such
132 GEMs to improve our understanding of APEC biology and identification of potential targets
133 for future prevention and intervention strategies.
134
135

136 **METHODS**

137

138 **Bacterial strains and growth conditions**

139 114 APEC isolates were obtained from the SAP culture collection at the School of
140 Veterinary Medicine, University of Surrey and used for APEC model construction
141 (Supplementary Material 2, Table S1). These APEC isolates were obtained from commercial
142 laying hens or broilers with either colibacillosis, perihepatitis, yolk sac infection, pericarditis
143 or peritonitis in the UK and Germany. Isolates were collected at *post-mortem* examination per
144 the protocols used at the diagnostic services involved, and were submitted to the School of
145 Veterinary Medicine at the University of Surrey between 2014 and 2017 for further
146 diagnostic or research activity. The clinical information available is provided in
147 Supplementary Material 2, Table S1. All isolates were stored in BHI broth supplemented
148 with 50% glycerol at -80 °C until required for analysis. Prior to analysis, the APEC isolates
149 were streaked onto MacConkey agar 3 plates, Nutrient agar or Luria-Bertani (LB) and
150 incubated aerobically at 37°C for 16-24 hours, unless otherwise stated. M9 minimal medium
151 (1× M9 salt, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 0.4% glucose if required) was used for
152 specific experiments as indicated below. Chloramphenicol and ampicillin were used at final
153 concentrations of 100 µg/ml and 30 µg/ml, respectively. All growth media and antibiotic
154 supplements were purchased from Oxoid (ThermoFisher Scientific) unless otherwise stated.

155

156 **Biolog Phenotypic Microarray and M9 minimal medium test supplemented with 3-**
157 **hydroxyphenylacetate (3-HPAA)**

158 Three APEC isolates were selected from the representative phylogroups B2, C and G
159 (serogroups O2, O78, and O24, respectively) and used for Biolog analysis (N=3). Additional
160 isolates from these phylogroups were selected for M9 minimal medium cultivation analysis

161 (N=13). The isolates used for these analyses are listed in Supplementary Material 2, Table
162 S1. Three biological and three technical replicates were performed for each isolate. The
163 APEC isolates were pre-cultured on R2A agar plates (Biolog) for Biolog assays as per the
164 manufacturer's instructions [18], and subsequently tested using 190 carbon sources (PM1 and
165 PM2 MicroPlate™), as well as 59 Phosphorus and 35 Sulphur Sources (PM4A
166 MicroPlate™). The cell suspensions and plate inoculation were undertaken according to the
167 instructions from the manufacturer. The Biolog PM1, PM2 and PM4A microplates were
168 incubated and analysed in Biolog's OmniLog® instrument at 37 °C for 24h, aerobically.
169 Readings were taken every 15 min. M9 minimal media was used to test the metabolic ability
170 of isolates in a defined synthetic medium. Three concentrations (1 mM, 0.75 mM, and 0.5
171 mM) of 3-hydroxyphenylacetate (3-HPAA) were tested in this study. L-lysine or putrescine
172 (1 mM) was also tested in this study for the cultivation analysis of gene knockout mutants.
173 The polystyrene 96 well plates (ThermoFisher Scientific) were cultured and read in Spark®
174 microplate reader (TECAN) at 37 °C for 24h, aerobically. Readings were taken every 15 min.
175

176 **DNA extraction and Next generation sequencing (NGS)**

177 The high molecular weight genomic DNA of APEC isolates was extracted and purified
178 using a DNA Wizard® genomic purification kit (Promega), according to the instructions
179 from the manufacturer. The extracted DNA samples were sent to the Animal and Plant Health
180 Agency (Weybridge, UK) for sequencing on the Illumina MiSeq platform, providing 150 nt
181 paired end reads. The obtained raw DNA sequences were *de novo* assembled in contigs using
182 Shovill version 1.1.0 (https://github.com/tseemann/shovill) using version 3.14 of the Spades
183 assembler [19] with minimum coverage of 2.00× and minimum length of 200 nt. Quality
184 assessment of genome assemblies were taken by QUAST version 5.0 [20], and required the
185 following minimum metrics to be satisfied: genome size between 4.5-6.0 Mbp, number of

186 contigs < 500 and N_{50} values of > 50 kb, respectively (Supplementary Material 2, Table S1).
187 Prokka version 1.14 [21] was used for initial annotation of genome sequences, which
188 produced relevant output files compatible for downstream analyses. The phylogenetic
189 grouping was analysed on the *in-silico* based ClermonTyping program [22]. *In silico*
190 serotyping was performed using Abricate version 1.0.1
191 (https://github.com/tseemann/abricate) with the EcOH [23] database. MLST sequence typing
192 was performed using the mlst v2.23 program (https://github.com/tseemann/mlst) with the
193 Warwick scheme [24]. A phylogenetic tree based on whole genome single-nucleotide
194 polymorphisms (SNPs) was generated using ParSNP version 1.2 [25], resulting in a
195 phylogenetic tree highlighting genomic similarity and sharing of core genome of all isolates.
196 All 114 APEC genomes were screened for 11 proposed virulence markers [26] using
197 Abricate. All relevant information for the 114 genomes (assembly metrics, presence/absence
198 of virulence genes, O:H serotype information and MLST sequence types) are included in
199 Supplementary Material 2, Table S1.
200
201 **Construction of APEC metabolic models**
202 The FASTA files of 114 APEC samples generated from Prokka were submitted to the
203 RAST server (https://rast.nmpdr.org/) for genome annotation [27]. The RAST annotation was
204 automatically submitted to ModelSEED (https://modelseed.org/genomes/) for further model
205 draft construction [28]. Model gap-filling is an essential step to investigate the missing
206 reactions, which apply a mixed-integer linear program to define the minimal number of
207 reactions required for the customer-defined collection of reactions. Several algorithms have
208 been reported and applied in model gap-filling [29, 30]. COBRA Toolbox v3.0 [31] was used
209 in this study and the database was based on ModelSeed. Further manual curation was
210 performed when model was unable to be corrected by gap-filling steps, for instance:

211 unnecessary reactions must be on for model running, and to avoid stoichiometrically
212 balanced cycles which are a subset of contiguous reactions that forms a loop or internal
213 network reactions, that carry fluxes without any exchange reactions involved. The model was
214 adjusted based on the literature (basic nutrient requirement for cell growth and common
215 pathways) and reference model *i*JO1366 [14]. The correction procedure included
216 adding/deletion of reactions, adjusting of reversibility/directionality of reactions, and change
217 of biomass consumption reactions. Metabolic models have been included as Supplementary
218 Material 3 (Table S5), Supplementary Material 4 (Table S6) and Supplementary Material 5
219 (Table S7).
220
221 **GEM analysis**
222 The growth prediction analysis was performed on the SurreyFBA 2.0 beta with JyMet2
223 [32]. The model was imported as a SurreyFBA formatted file, meanwhile a specific format
224 (pfile) was also introduced, in which the media setting can be adjusted via changing the
225 upper/lower bound of each exchange reaction. The computational prediction of growth rate
226 was based on typical flux units (millimole per gram dry weight per hour, mmol gDW^{-1} h^{-1}).
227 The default nutrient sources were glucose (carbon, C), ammonia (NH_3), phosphate (PO_4^{2-})
228 and sulphate (SO_4^{2-}). When examining the usage of a particular nutrient source, the exchange
229 reaction of its corresponding default source was adjusted to zero (reaction shut down). In
230 addition, growth simulation studies were performed using the *E. coli* K-12 model *i*JO1366
231 [14] for comparative analysis. Due to the differences in default nutrient requirements in
232 modelling, two growth simulations (*i*JO1366a and *i*JO1366b) were performed within two
233 different input nutrient settings. The input nutrients of *i*JO1366a were adjusted as in the
234 APEC model used here, where the exchange reactions of selenate (SeO_4^{2-}), selenite (SeO_3^{2-})
235 and tungstate (WO_4^{2-}) were disabled as they are absent from the APEC model; whilst these

236 three exchange reactions were enabled in *iJO1366*b to perform the optimal conditions for
237 *i*JO1366. In both *i*JO1366a and *i*JO1366b, the exchange reactions of molybdate (MoO_4^{2-}) and
238 nickel (Ni^{2+}), which are also absent from the APEC model, were kept enabled, as they are
239 essential for *i*JO1366. The phenotypic and computational measurements were compared
240 using a Pearson correlation coefficient test to measure the strength and direction of these two
241 variables.
242
243 **Inactivation of genes in *E. coli***
244 Genes were inactivated using the Red Lambda Recombination assay [33]. Two gene
245 targets (*lysA* and *potE*) were selected based on gene essentially predicted by model. The
246 target genes were disrupted by deletion of approximately 450~550 bp sequence in the middle
247 part of the gene sequence, and replaced with a chloramphenicol resistance cassette (CAT) as
248 selective marker. These cloning fragments were constructed using gBlocks gene fragments in
249 tubes service (gBlocks™, Integrated DNA Technologies). APEC isolates were selected for
250 gene inactivation experiments from phylogroup B2, G and C, respectively, with 90%
251 coverage and identity for the *lysA* and *potE* genes, and no resistance to ampicillin and
252 chloramphenicol in Mueller Hinton broth, as these were used as selective markers for the
253 inactivation procedure and associated plasmids. Electrocompetent cells were generated by
254 three times washing with ice-cold 10% glycerol and electroporated with the assistant plasmid
255 pSIJ8 [33, 34], containing an ampicillin resistance gene. The transformed cells with pSIJ8
256 were selected on LB agar supplemented with ampicillin, followed by PCR using primers
257 described in Supplementary Material 2, Table S8. The expression of the pSIJ8-encoded λ-
258 Red recombination system was induced with 0.5M arabinose. The cloning fragment was then
259 introduced through electroporation, with cells recovered at 30°C (as plasmid pSIJ8 is
260 temperature sensitive) on a shaking platform for 1 -2h. Mutants were selected on

261 chloramphenicol-containing LB plates and verified by PCR. We also designed a reverse
262 transcription quantitative-PCR (RT-qPCR) to confirm the absence of *lysA* and *potE* mRNA
263 transcripts in mutant candidates. The primers and probe of the real time quantitative PCR
264 were designed to amplify the sequences deleted from the wildtype isolate during the
265 construction of the cloning fragment (Supplementary Material 2, Table S8).
266

267 **RESULTS**

268

269 **Genomic characterisation of APEC isolates for metabolic model generation**

270 A total of 114 *E. coli* isolates were included in this study, isolated from diseased birds
271 from the UK and Germany. The pathologies observed are provided in Supplementary
272 Material 1, Table S1, and ranged from colibacillosis, perihepatitis, peritonitis, pericarditis to
273 yolk sac infection. To test whether these 114 APEC isolates were representative of the APEC
274 pathotype, their genome sequences were determined using Illumina sequencing and
275 subsequent genome assembly. The total size of the genome assemblies was between 4.72-
276 5.56 Mbp (average 5.11 Mbp), with the number of contigs ranging from 104-327 (average
277 197), and the N_{50} ranging from 63-342 kbp (average 159 kbp). Subsequently their
278 phylogroups, dominant serogroups and sequence types were determined using the genome
279 sequences (Figure 1). The phylogenetic grouping of isolates based on core genome SNPs
280 matched the distinctions in phylogroups and sequence types. *In silico* phylogroup typing
281 showed that the majority of APEC isolates belonged to B2 (27.2%), followed by C (18.4%),
282 G (16.7%), A (15.8%) and B1 (13.2%). The phylogroups D, E and F were underrepresented
283 in this collection. The most prevalent serogroups were O78 (17%) and O2 (13%) and the
284 most common MLST sequence types (STs) were ST-117 (15.8%), ST-23 (12.3%) and ST-95
285 (10.5%). Screening of the genome sequences for the presence of 11 virulence markers
286 showed that 83/114 (73%) of the included APEC isolates had 5 or more of these virulence
287 markers. All these characteristics broadly matched earlier studies [4-7, 35], and hence we
288 consider this collection of isolates broadly representative for the APEC pathotype.

289

290 **Generation of the metabolic profile of the APEC pathotype**

291 The metabolic profile of the APEC isolates was predicted using ModelSEED. A total of
292 1,848 different metabolic reactions were identified and annotated in the 114 APEC isolates
293 included, and were defined as pan reactions. A total of 1,639 reactions (88.7%) were shared
294 among all isolates, defined as core reactions, whereas 209 reactions (11.3%) were defined as
295 accessory reactions. The reactions were categorised according to their predicted metabolic
296 functions using ModelSEED, as shown in Figure 2. The most highly conserved metabolic
297 reactions were responsible for lipid metabolism (96.4%), nucleotide metabolism (95.1%),
298 coenzyme metabolism (94.2%) and energy production (93.9%). Reactions for alternative
299 carbohydrate utilisation and amino acid metabolism made up 42.9% (792/1,848) of pan
300 reactions and 52.6% (110/209) of accessory reactions. The metabolic reactions were
301 subsequently clustered with the phylogroups identified (Figure 3, Supplementary Material 2,
302 Table S9). In the accessory reactions, there was clustering of metabolic profiles of APEC
303 isolates from the same phylogroups, especially in metabolism of alternative carbohydrates
304 and amino acids including their transport/exchange capabilities. This suggests that certain
305 lineages are likely adapted to niches or environments, where they may harbour an advantage
306 in the utilisation of available nutrient. Notably, phylogroup B2 APEC isolates displayed
307 distinct metabolic capabilities compared to the other phylogroups. Similar observations were
308 also reported previously [36, 37]. Metabolic capabilities absent in phylogroup B2 were
309 mainly associated with amino acids and carbohydrates, and related transport/exchange of
310 metabolites. In contrast, metabolic reactions involved in membrane biogenesis, such as N-
311 acetylneuraminic acid (Neu5Ac) metabolism, were predominantly clustered in phylogroup
312 B2 and G. These reactions have been described to contribute to the pathogenicity of Gram-
313 negative bacteria by protecting against the host immune system [38].
314
315 **Characterisation of APEC GEM and phylogroup-specific sub-models**

316 The 1,848 metabolic reactions identified in the APEC isolates were used to generate a
317 draft genome-scale metabolic model using ModelSEED. Sixteen reactions were excluded as
318 dead-end reactions. Eventually, the pan-genome metabolic model constructed in this study,
319 referred as APEC GEM, was composed of 2,923 reactions (inclusive of gap-filling reactions
320 and a total biomass reaction), 2,527 metabolites, and 2,242 genes (Figure 4). Notably, there
321 were many reactions added during gap-filling algorithms and manual curation, such as
322 spontaneous reaction for extracellular metabolites exchange, isomerization, putative
323 reactions, which significantly change the distribution of functional categories compared to
324 Figure 2. Based on KEGG annotations and the *E. coli* K-12 model [14], a large number of
325 exchange reactions of extracellular metabolites were accumulated in the transport/exchange
326 functional category (Figure 4). This contrasted with the central metabolism reactions, such as
327 those involved in energy production and nucleotide metabolism, which are generally
328 conserved across *E. coli* lineages.

329 The APEC GEM was built as a pan-genome metabolic model, to represent the full
330 metabolic capabilities in the complete APEC collection used in this study. Considering the
331 distinct metabolic variances between different APEC lineages (Figure 3, Supplementary
332 Material 2, Table S9), phylogroup B2, C and G-specific models were generated from the
333 APEC GEM to represent the metabolic capability and deficiency of each phylogroup. A brief
334 summary of the reactions associated with different metabolic categories removed for sub-
335 model construction for each phylogroup is shown in Supplementary Material 1, Figure S1. In
336 total, 133 reactions were deleted during the construction of sub-models. Notably, the same set
337 of metabolic pathways were removed from more than one phylogroup. Most of deleted
338 reactions were associated with carbohydrate metabolism, accounting for 42%, followed by
339 amino acid metabolism (14.3%) and transport/exchange reactions (9%). A relatively high
340 ratio of carbohydrate (13/19) and amino acid (45/57) metabolism reactions and

341 transport/exchange reactions (9/12) were removed in phylogroup B2 in this study, consistent
342 with prior observations. Besides, 39/57 of carbohydrate metabolism reactions were removed
343 for phylogroup G.
344
345 **Gene essentiality analysis of APEC GEM**
346 Gene essentiality analysis is often adopted to evaluate the validity of GEM by comparing
347 with experimentally obtained essential gene datasets [9, 16]. The APEC GEM was used to
348 predict growth on glucose and glycerol as carbon sources. For growth on glucose or glycerol,
349 the lower boundaries of the glucose and glycerol exchange reactions were set to -1,000,
350 respectively. All 2,242 genes in APEC GEM were inactivated *in silico* individually as time
351 and growth were simulated by flux balance analysis (FBA) (Supplementary Material 2, Table
352 S2). When inactivation of a gene resulted in a growth rate above zero, it was considered as a
353 non-essential gene. The gene essentiality predictions on glucose and glycerol in minimal
354 media showed a consistent outcome using the APEC model in this study (Supplementary
355 Material 2, Table S2). In general, *in silico* inactivation of 208 out of 2,242 genes (9.3%) led
356 to an absence of flux through the network and as such would lead to growth inhibition, and
357 thus were determined to be essential genes in the APEC model. The distribution of these
358 essential genes is shown in Supplementary Material 1, Figure S2. Of these essential genes, 57
359 (27%) were associated with more than one reaction in this model. 13% were responsible for
360 amino acid metabolism and 3% for carbohydrate metabolism. To validate these predictions,
361 we decided against the use of the Keio collection grown on glucose/glycerol, as there are
362 significant differences between *E. coli* K-12 and APEC. Hence two single gene knockout
363 mutants were constructed in APEC for this purpose, in the essential *lysA* gene and non-
364 essential *potE* gene. Both genes were present in 100% of the 114 APEC isolates analysed in
365 this study.

366 The *lysA* gene is predicted to encode a diaminopimelate decarboxylase that is responsible
367 for the biosynthesis of lysine. In the model, the *lysA* gene (model ID: peg18487) was
368 determined to be an essential gene and encoded for chemical equation (model ID:
369 R_rxn00313) as below ([c] represent for the metabolites in cytoplasm):
370 *Rxn00313: (1) H^+ [c]+(1) meso-2,6-Diaminopimelate [c] => (1) CO_2 [c]+(1) L-Lysine [c]*
371 The *potE* gene encodes a membrane protein mediating both the uptake (via proton
372 symport mechanism) and excretion of putrescine (as a putrescine and ornithine antiporter).
373 There is redundancy in putrescine transport with the PotABCD, PotFGHI, PlaP/YeeF, PuuP
374 and SapBCDF systems [39-42]. In the model, the target gene *potE* (model ID: peg6625) was
375 determined as non-essential and responsible for two putrescine transport reactions (model ID:
376 R_rxn05867 and R_rxn10182). For the modelling purpose, the chemical equations of these
377 two reactions are represented as below ([e] represents for the extracellular metabolites; [c]
378 represent for the metabolites in cytoplasm):
379 *Rxn05687: (1) H^+[e] + (1) Putrescine[e] => (1) H^+[c] + (1) Putrescine[c]*
380 *Rxn10182: (1) Ornithine[c] + (1) Putrescine[e] <=> (1) Ornithine[e] + (1) Putrescine[c]*
381 The *lysA* mutants generated in isolates SAP0010 and SAP0631 and *potE* mutant
382 generated in isolates SAP0631 were analysed using glucoside as default carbon source. Other
383 default nutrient sources were adjusted by general standard principle described in next section.
384 The *in-silico* growth prediction of knockouts grown in with and without supplementation of
385 L-lysine or putrescine were compared to the phenotypic screening using M9 minimal medium
386 as illustrated in Figure 5. The Δ*lysA* mutants (Δ*lysA* 10-9 and Δ*lysA* 631-11) were observed to
387 be auxotrophic for L-lysine (Figure 5), which indicates the fact that the *lysA* gene encodes the
388 enzyme in the final and only lysine biosynthesis pathway. This was also revealed in APEC
389 GEM knockout analysis that no biomass was produced when the rxc00313 was switched off.
390 On the contrary, the supplementation of putrescine in M9 medium did not lead any

391 significant change in growth between the Δ*potE* 631-12 mutant and its parental strain (Figure
392 5). The APEC GEMs correctly predicted the growth status of these knockout mutants, as well
393 as their wildtype, in minimal media with and without supplementation.
394

395 **Comparison of the APEC GEM with the *E. coli* K-12 *i*JO1366 GEM**

396 To validate *in silico* model prediction with phenotypic observations, the growth
397 (respiration) phenotype of three representative APEC isolates, chosen from phylogroups B2,
398 C and G, were tested using the Biolog Phenotypic Microarray analysis (PM1, PM2A: carbon
399 source plates, and PM4A: sulphate and phosphate plate). The respiration measurements are
400 available in Supplementary Material 2, Table S3**.** In total, 125 nutrient substrates were
401 analysed in the APEC model, and 119 substrates were analysed in *i*jO1366 (Supplementary
402 Material 2, Table S4). The Pearson correlation coefficient of all three comparison
403 experiments showed mostly positive correlations; the *i*JO1366a model showed a weak
404 positive correlation (r^2 values between 0.07 and 0.17, Supplementary Material 1, Figure S3),
405 while the *i*JO1366b and APEC models showed a moderate positive correlations (r^2 values
406 between 0.50 and 0.59, Figure 6). Under the optimised media settings, the APEC model
407 showed an improved performance with the representative phylogroups C (r^2=0.59, $p < 0.005$)
408 and G (r^2=0.59, $p < 0.005$), while it showed decreased performance with phylogroup B2
409 (r^2=0.50, $p < 0.005$), compared to model *i*JO1366 (Figure 6).
410

411 **Comparison of the APEC GEM with phylogroup-specific sub-models**

412 We also explored the performance of phylogroup-specific APEC model based sub
413 models. The default media setting of phylogroup based sub-models were modified from the
414 APEC GEM, with the import of Fe(II)-Nicotianamine (Fe(II)-NA or Fe^{2+}-NA) added as it
415 was required for model performance of sub-models for phylogroups C and G. In total, *in*

416 *silico* flux balance analysis (FBA) of 142 nutrient substrates was performed with phylogroup-
417 specific sub-models and then compared to growth simulation of APEC GEM according to
418 Biolog data (Supplementary Material 2, Table S4). In Figure 7, the relationship between the
419 computational prediction from the sub-models and their Biolog measurements showed
420 moderate positive correlation (r^2 values of 0.51, 0.64 and 0.3 for phylogroups C, G and B2,
421 respectively). Compared to the computational prediction by the pan model APEC GEM, the
422 sub-models of phylogroup C performed less well under similar media input, while the
423 phylogroup G and B2 sub-model showed improved performance compared to APEC model.
424

425 **Identification of 3-hydroxyphenylacetate as a phylogroup C APEC-specific nutrient**

426 During querying of the phylogroup-specific APEC GEMs, one of the reactions predicted
427 to be specific for phylogroup C APEC isolates was associated with the metabolism of 3- and
428 4-hydroxyphenylacetic acid:

429 *Rxn02363:* $NADH+O_2+H^+$*+3-hydroxyphenylacetate =>* H_2O+NAD^+*+homoprotocatechuate*

430 This reaction is catalysed by 4-hydroxyphenylacetate 3-monooxygenase oxygenase,
431 encoded by gene *hpaB* (4-hydroxyphenylacetate 3-monooxygenase, EC 1.14.13.3). The *hpaB*
432 gene was present in all phylogroup C (n=21/21) APEC, while absent in phylogroup B2
433 (n=0/31) and G (n=0/19) APEC. The *hpaB* gene was detected in phylogroups A and B1 as
434 well, but these constitute minor phylogroups for APEC. The metabolic variation on 3-HPAA
435 metabolism was measured using M9 minimal medium. As is shown in Supplementary
436 Material 1, Figure S4, five phylogroup C isolates were all able to grow with the three
437 different concentrations of 3-HPAA tested, whereas six phylogroup B2 and two phylogroup
438 G isolates were unable to grow with 3-HPAA as the single carbon source (Supplementary
439 Material 1, Figure S5), but all grew in LB medium.

440 The metabolism of 3-HPAA in *E. coli* has been described previously [43, 44]. However,
441 the intracellular/extracellular reaction of 3-HPAA in *E. coli* has not yet been documented in
442 the ModelSEED database. Therefore, we manually added the exchange reaction of 3-HPAA
443 into the sub-models, in order to perform the growth simulation with/without supplementary
444 of 3-HPAA. The additional reactions are presented as below (cpd03320 is the compound ID
445 of 3-HPAA in APEC GEM and relative sub-models, *e* represent for extracellular; *b* represents
446 for periplasm; *c* represent for cytoplasm):
447 *R_e2c_cpd03320: M_cpd03320_e = M_cpd03320_c*
448 *R_EX_cpd03320: M_cpd03320_e = M_cpd03320_b*
449 The lower bound (lb) of the 3-HPAA exchange reaction was adjusted to 0, -10, -100 and
450 -1000 in the APEC GEM sub-model B2, C and G, to simulate APEC growth in response to
451 four different concentrations of 3-HPAA. The growth prediction of different sub-groups and
452 M9 experimental observation are summarised in Figure 8. In sub-models B2 and G, the
453 relatively minimal biomass (~15) was produced with the lower bound of 3-HPAA exchange
454 reaction adjusted to 0, -10, -100, and -1000 (increasing uptake), which was assumed as no
455 growth for the modelling purpose. This agreed with the M9 minimal media experimental
456 data, where isolates from phylogroup B2 and G were unable to use 3-HPAA as the sole
457 carbon source. However, growth was predicted in sub-model C, matching the setting of
458 increased uptake of 3-HPAA (Figure 8).
459

460 **DISCUSSION**

461 In this study, we have investigated the metabolic capabilities of a representative set of
462 114 avian pathogenic *E. coli* isolates representing different phylogroups. A genome scale
463 metabolic model was constructed based on the genome sequences of these APEC isolates and
464 used to 1) perform growth simulation experiments which were validated using phenotypic
465 assays 2) perform gene essentiality analysis 3) investigate functional difference between
466 phylogroups and develop phylogroup based sub-models.

467 The comparative genomic analysis performed in this study confirmed that the APEC
468 pathotype is genetically variable, with phylogroups B2, C and G being the dominant lineages
469 [4, 5, 45]. The resulting genome scale metabolic profile of the 114 APEC isolates (Figure 3,
470 Supplementary Material 2, Table S9) showed strong conservation in central metabolism
471 involved in energy production, with the ratio of core metabolism versus accessory
472 metabolism higher than in previous studies, where multiple *E. coli* pathotypes and
473 commensal isolates were involved [9, 46]. The major metabolic variances between isolates
474 are associated with metabolism of carbohydrates and amino acids, as well as
475 exchange/transport reactions. This was confirmed using Biolog metabolic analysis performed
476 in this study. As shown in Figure 3 and Supplementary Material 2, Table S9, the phylogroup
477 B2 isolates showed distinct metabolic features (mostly missing metabolic reactions)
478 compared to others, which was previously also observed [36, 47, 48]. This is consistent with
479 the view that loss and acquisition of genes during evolution differentiates isolates into
480 different lineage with unique metabolic profiles and adaptation to specific niches.

481 When comparing the predictions of the APEC model with the Biolog phenotypic
482 measurements, the association between APEC GEM prediction and phenotypic observations
483 was moderately positive (r^2 values between 0.5 and 0.6). The vast numbers of reactions and
484 genes encompassed within metabolic models potentially leads to a high risk of gaps or errors

485 within the metabolic network. Furthermore, the reconstruction of GEMs is severely limited
486 by the paucity of verified metabolic pathways that are specific to APEC and current
487 knowledge in our database, including errors introduced by automatic annotation of genome
488 sequences in the respective databases. From the perspective of APEC diversity, the pan-
489 genome scale metabolic model is too extensive to represent the differences and specifics of
490 certain APEC lineages. Notably, in this study, we used the Biolog platform as the indicator of
491 biomass formation of APEC isolates in response to different nutrient sources, while this assay
492 is based on the measurement of microbial respiration and reflected by dye changes. These
493 two different measurements may lead to false-positive or -negative results during the
494 coefficient relationship analysis. This was addressed by changing the function from biomass
495 production to indicator dye reduction used in the Biolog plates, which leads to a significantly
496 decrease in false negative results [8]. Alternatively, a comprehensive screening of isolates
497 using different nutrient sources by M9 minimal medium may provide more direct observation
498 for model validation. Typically, the phenotype of the Keio collection was used in model
499 validation of *i*JO1366 [14]. However, this was not appropriate for our study due to the
500 significant differences between the laboratory strain K-12 and the APEC pathotype.
501 Moreover, the APEC pathotype has a diverse metabolic profile and construction of a single
502 gene knockout library representing the APEC pathotype is not possible, which made it
503 difficult to apply the APEC metabolic model for the pathotype. However, the APEC
504 metabolic model provides a comprehensive database of metabolic capabilities for a diverse
505 panel of APEC, and as such can be used for initial predictions, and can be used as the basis
506 for the construction of lineage-based strain sub-models, as performed here for the dominant
507 phylogroups B2, C and G.

508 Although no significant improvement in the association strength between computational
509 prediction of sub-models and the phenotypic measurement of the corresponding APEC

510 isolates were observed, the APEC model showed its potential and plasticity as a pan model
511 used for the construction of phylogroup B2, C and G sub-models. The phenotypic difference
512 in catabolism of 3-HPAA in phylogroup B2, C and G were successfully predicted by those
513 phylogroup based sub-models as a proof of principle. Interestingly, 3- hydroxyphenylacetate
514 and 4-hydroxyphenylacetate are the main phenolic acids derived from flavonoid quercetin
515 (which is commonly added to poultry feed) in the colorectum [49, 50]. Therefore, these
516 compounds are likely to accumulate in the avian intestine and its utilisation could be a
517 hypothetical nutrient advantage in the intestinal microbial community. The phylogroup C
518 sub-models not only predicted the growth of phylogroup C APEC isolates using 3-
519 hydroxyphenylacetate as sole carbon source, but also reflected the biomass increases
520 alongside the enhancing lower bound setting of 3- hydroxyphenylacetate (Figure 8,
521 Supplementary Material 1, Figure S4).

522 The models generated in this study can be further exploited for the identification of
523 additional pathways that are unique to APEC lineages. Although APEC are extraintestinal
524 pathogens, potential APEC isolates are present in the intestinal *E. coli* population [7], and
525 may form a reservoir for future infections within flocks. These models may inform the
526 targeted suppression of such APEC populations thereby benefitting avian health and welfare
527 in the long term. The approach taken may also be applicable for other genetically diverse
528 intestinal or extraintestinal pathogens dispersed over multiple lineages. Furthermore, in the
529 future, the rapid progress in software development (e.g. Bactabolize [51]), availability of
530 APEC genome sequences and associated repositories may expedite the generation of genome
531 scale metabolic models to revisit or improve existing metabolic models for APEC and other
532 *E. coli* pathotypes.

533

534 **ACKNOWLEDGMENTS**

The authors would like to acknowledge the diagnostic laboratories for providing the isolates used in the study.

538 **REFERENCES**

539

540 1. **Denamur E, Clermont O, Bonacorsi S, Gordon D**. The population genetics of
541 pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2021;19(1):37-54. doi:
542 10.1038/s41579-020-0416-x
543 2. **Kohler CD, Dobrindt U**. What defines extraintestinal pathogenic *Escherichia coli*?
544 *Int J Med Microbiol* 2011;301(8):642-647. doi: 10.1016/j.ijmm.2011.09.006
545 3. **La Ragione RM, Woodward MJ**. Virulence factors of *Escherichia coli* serotypes
546 associated with avian colisepticaemia. *Res Vet Sci* 2002;73(1):27-35. doi:
547 10.1016/s0034-5288(02)00075-9
548 4. **Cordoni G, Woodward MJ, Wu H, Alanazi M, Wallis T et al.** Comparative
549 genomics of European avian pathogenic *E. coli* (APEC). *BMC Genomics*
550 2016;17(1):960. doi: 10.1186/s12864-016-3289-7
551 5. **Mageiros L, Meric G, Bayliss SC, Pensar J, Pascoe B et al.** Genome evolution and
552 the emergence of pathogenicity in avian *Escherichia coli*. *Nat Commun*
553 2021;12(1):765. doi: 10.1038/s41467-021-20988-w
554 6. **Mehat JW, van Vliet AHM, La Ragione RM**. The avian pathogenic *Escherichia*
555 *coli* (APEC) pathotype is comprised of multiple distinct, independent genotypes.
556 *Avian Pathol* 2021;50(5):402-416. doi: 10.1080/03079457.2021.1915960
557 7. **Logue CM, Wannemuehler Y, Nicholson BA, Doetkott C, Barbieri NL et al.**
558 Comparative analysis of phylogenetic assignment of human and avian ExPEC and
559 fecal commensal *Escherichia coli* using the (previous and revised) Clermont
560 phylogenetic typing methods and its impact on avian pathogenic *Escherichia coli*
561 (APEC) classification. *Front Microbiol* 2017;8:283. doi: 10.3389/fmicb.2017.00283
562 8. **Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT**. The evolution of
563 metabolic networks of *E. coli*. *BMC Syst Biol* 2011;5:182. doi: 10.1186/1752-0509-5-
564 182
565 9. **Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N et al.** Genome-
566 scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-
567 specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A*
568 2013;110(50):20338-20343. doi: 10.1073/pnas.1307797110

569 10. **Nhu NTK, Rahman MA, Goh KGK, Kim SJ, Phan MD et al.** A convergent
570 evolutionary pathway attenuating cellulose production drives enhanced virulence of
571 some bacteria. *Nat Commun* 2024;15(1):1441. doi: 10.1038/s41467-024-45176-4
572 11. **Gu C, Kim GB, Kim WJ, Kim HU, Lee SY**. Current status and applications of
573 genome-scale metabolic models. *Genome Biol* 2019;20(1):121. doi: 10.1186/s13059-
574 019-1730-3
575 12. **Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR et al.** A genome-
576 scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for
577 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3:121. doi:
578 10.1038/msb4100155
579 13. **Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A et al.** iML1515, a knowledgebase
580 that computes *Escherichia coli* traits. *Nat Biotechnol* 2017;35(10):904-908. doi:
581 10.1038/nbt.3956
582 14. **Orth JD, Conrad TM, Na J, Lerman JA, Nam H et al.** A comprehensive genome-
583 scale reconstruction of *Escherichia coli* metabolism--2011. *Mol Syst Biol* 2011;7:535.
584 doi: 10.1038/msb.2011.65
585 15. **Reid CJ, Cummins ML, Borjesson S, Brouwer MSM, Hasman H et al.** A role for
586 ColV plasmids in the evolution of pathogenic *Escherichia coli* ST58. *Nat Commun*
587 2022;13(1):683. doi: 10.1038/s41467-022-28342-4
588 16. **Kim D, Kim Y, Yoon SH**. Development of a genome-scale metabolic model and
589 phenome analysis of the probiotic *Escherichia coli* strain Nissle 1917. *Int J Mol Sci*
590 2021;22(4):2122. doi: 10.3390/ijms22042122
591 17. **Archer CT, Kim JF, Jeong H, Park JH, Vickers CE et al.** The genome sequence of
592 *E. coli* W (ATCC 9637): comparative genome analysis and an improved genome-
593 scale reconstruction of *E. coli*. *BMC Genomics* 2011;12:9. doi: 10.1186/1471-2164-
594 12-9
595 18. **Bochner BR, Gadzinski P, Panomitros E**. Phenotype microarrays for high-
596 throughput phenotypic testing and assay of gene function. *Genome Res*
597 2001;11(7):1246-1255. doi: 10.1101/gr.186501
598 19. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al.** SPAdes: a new
599 genome assembly algorithm and its applications to single-cell sequencing. *J Comput*
600 *Biol* 2012;19(5):455-477. doi: 10.1089/cmb.2012.0021

601 20. **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. QUAST: quality assessment tool for
602 genome assemblies. *Bioinformatics* 2013;29(8):1072-1075. doi:
603 10.1093/bioinformatics/btt086
604 21. **Seemann T**. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*
605 2014;30(14):2068-2069. doi: 10.1093/bioinformatics/btu153
606 22. **Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O**.
607 ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus
608 strain phylotyping. *Microb Genom* 2018;4(7):e000192. doi: 10.1099/mgen.0.000192
609 23. **Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M et al.** In silico
610 serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive
611 diversity of O:H serotype combinations within and between pathogenic lineages.
612 *Microb Genom* 2016;2(7):e000064. doi: 10.1099/mgen.0.000064
613 24. **Wirth T, Falush D, Lan R, Colles F, Mensa P et al.** Sex and virulence in
614 *Escherichia coli*: an evolutionary perspective. *Mol Microbiol* 2006;60(5):1136-1151.
615 doi: 10.1111/j.1365-2958.2006.05172.x
616 25. **Treangen TJ, Ondov BD, Koren S, Phillippy AM**. The Harvest suite for rapid core-
617 genome alignment and visualization of thousands of intraspecific microbial genomes.
618 *Genome Biol* 2014;15(11):524. doi: 10.1186/s13059-014-0524-x
619 26. **Subedi M, Luitel H, Devkota B, Bhattarai RK, Phuyal S et al.** Antibiotic
620 resistance pattern and virulence genes content in avian pathogenic *Escherichia coli*
621 (APEC) from broiler chickens in Chitwan, Nepal. *BMC Vet Res* 2018;14(1):113. doi:
622 10.1186/s12917-018-1442-z
623 27. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ et al.** The SEED and the
624 Rapid Annotation of microbial genomes using Subsystems Technology (RAST).
625 *Nucleic Acids Res* 2014;42(Database issue):D206-214. doi: 10.1093/nar/gkt1226
626 28. **Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA et al.** Automated genome
627 annotation and metabolic model reconstruction in the SEED and Model SEED.
628 *Methods Mol Biol* 2013;985:17-45. doi: 10.1007/978-1-62703-299-5_2
629 29. **Kumar VS, Maranas CD**. GrowMatch: an automated method for reconciling in
630 silico/in vivo growth predictions. *PLoS Comput Biol* 2009;5(3):e1000308. doi:
631 10.1371/journal.pcbi.1000308

632 30. **Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK et al.** Systems approach to
633 refining genome annotation. *Proc Natl Acad Sci U S A* 2006;103(46):17480-17484.
634 doi: 10.1073/pnas.0603364103
635 31. **Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A et al.** Creation and
636 analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0.
637 *Nat Protoc* 2019;14(3):639-702. doi: 10.1038/s41596-018-0098-2
638 32. **Gevorgyan A, Bushell ME, Avignone-Rossa C, Kierzek AM**. SurreyFBA: a
639 command line tool and graphics user interface for constraint-based modeling of
640 genome-scale metabolic reaction networks. *Bioinformatics* 2011;27(3):433-434. doi:
641 10.1093/bioinformatics/btq679
642 33. **Jensen SI, Lennen RM, Herrgard MJ, Nielsen AT**. Seven gene deletions in seven
643 days: Fast generation of *Escherichia coli* strains tolerant to acetate and osmotic stress.
644 *Sci Rep* 2015;5:17874. doi: 10.1038/srep17874
645 34. **Jensen SI, Nielsen AT**. Multiplex genome editing in *Escherichia coli*. *Methods Mol*
646 *Biol* 2018;1671:119-129. doi: 10.1007/978-1-4939-7295-1_8
647 35. **Cummins ML, Reid CJ, Roy Chowdhury P, Bushell RN, Esbert N et al.** Whole
648 genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry
649 the class 1 integrase gene. *Microb Genom* 2019;5(2):e000250. doi:
650 10.1099/mgen.0.000250
651 36. **Fang X, Monk JM, Mih N, Du B, Sastry AV et al.** *Escherichia coli* B2 strains
652 prevalent in inflammatory bowel disease patients have distinct metabolic capabilities
653 that enable colonization of intestinal mucosa. *BMC Syst Biol* 2018;12(1):66. doi:
654 10.1186/s12918-018-0587-5
655 37. **Gonzalez-Alba JM, Baquero F, Canton R, Galan JC**. Stratified reconstruction of
656 ancestral *Escherichia coli* diversification. *BMC Genomics* 2019;20(1):936. doi:
657 10.1186/s12864-019-6346-1
658 38. **Lin BX, Qiao Y, Shi B, Tao Y**. Polysialic acid biosynthesis and production in
659 *Escherichia coli*: current state and perspectives. *Appl Microbiol Biotechnol*
660 2016;100(1):1-8. doi: 10.1007/s00253-015-7019-x
661 39. **Guerra PR, Herrero-Fresno A, Ladero V, Redruello B, Dos Santos TP et al.**
662 Putrescine biosynthesis and export genes are essential for normal growth of avian
663 pathogenic *Escherichia coli*. *BMC Microbiol* 2018;18(1):226. doi: 10.1186/s12866-
664 018-1355-9

665 40. **Guerra PR, Herrero-Fresno A, Pors SE, Ahmed S, Wang D et al.** The membrane
666 transporter PotE is required for virulence in avian pathogenic *Escherichia coli*
667 (APEC). *Vet Microbiol* 2018;216:38-44. doi: 10.1016/j.vetmic.2018.01.011
668 41. **Kurihara S, Suzuki H, Oshida M, Benno Y**. A novel putrescine importer required
669 for type 1 pili-driven surface motility induced by extracellular putrescine in
670 *Escherichia coli* K-12. *J Biol Chem* 2011;286(12):10185-10192. doi:
671 10.1074/jbc.M110.176032
672 42. **Thongbhubate K, Irie K, Sakai Y, Itoh A, Suzuki H**. Improvement of putrescine
673 production through the arginine decarboxylase pathway in *Escherichia coli* K-12.
674 *AMB Express* 2021;11(1):168. doi: 10.1186/s13568-021-01330-5
675 43. **Cooper RA, Skinner MA**. Catabolism of 3- and 4-hydroxyphenylacetate by the 3,4-
676 dihydroxyphenylacetate pathway in *Escherichia coli*. *J Bacteriol* 1980;143(1):302-
677 306. doi: 10.1128/jb.143.1.302-306.1980
678 44. **Prieto MA, Garcia JL**. Identification of the 4-hydroxyphenylacetate transport gene
679 of *Escherichia coli* W: construction of a highly sensitive cellular biosensor. *FEBS Lett*
680 1997;414(2):293-297. doi: 10.1016/s0014-5793(97)01012-0
681 45. **Chen X, Liu W, Li H, Yan S, Jiang F et al.** Whole genome sequencing analysis of
682 avian pathogenic *Escherichia coli* from China. *Vet Microbiol* 2021;259:109158. doi:
683 10.1016/j.vetmic.2021.109158
684 46. **Vieira G, Sabarly V, Bourguignon PY, Durot M, Le Fevre F et al.** Core and
685 panmetabolism in *Escherichia coli*. *J Bacteriol* 2011;193(6):1461-1472. doi:
686 10.1128/JB.01192-10
687 47. **van der Hooft JJJ, Goldstone RJ, Harris S, Burgess KEV, Smith DGE**.
688 Substantial extracellular metabolic differences found between phylogenetically
689 closely related probiotic and pathogenic strains of *Escherichia coli*. *Front Microbiol*
690 2019;10:252. doi: 10.3389/fmicb.2019.00252
691 48. **Clermont O, Dixit OVA, Vangchhia B, Condamine B, Dion S et al.**
692 Characterization and rapid identification of phylogroup G in *Escherichia coli*, a
693 lineage with high virulence and antibiotic resistance potential. *Environ Microbiol*
694 2019;21(8):3107-3117. doi: 10.1111/1462-2920.14713
695 49. **Bhutto ZA, He F, Zloh M, Yang J, Huang J et al.** Use of quercetin in animal feed:
696 effects on the P-gp expression and pharmacokinetics of orally administrated
697 enrofloxacin in chicken. *Sci Rep* 2018;8(1):4400. doi: 10.1038/s41598-018-22354-1

698 50. **Hong YJ, Mitchell AE**. Identification of glutathione-related quercetin metabolites in
699 humans. *Chem Res Toxicol* 2006;19(11):1525-1532. doi: 10.1021/tx0601758
700 51. **Vezina B, Watts SC, Hawkey J, Cooper HB, Judd LM et al.** Bactabolize is a tool
701 for high-throughput generation of bacterial strain-specific metabolic models. *Elife*
702 2023;12. doi: 10.7554/eLife.87406

703

704

705 **LEGENDS TO FIGURES**

706

707 **Figure 1.** APEC population structure for the isolates selected for GEM construction in this
708 study assigned with phylogroup (n=114). Colour code of phylogroups: A = Yellow, B1 =
709 Green, B2 = Red, C = Orange, D = Purple, E = Blue, F = Light blue, G = Black. The
710 distribution of dominant serogroups is shown in the outer ring. The colour code of serogroup:
711 O78 = Black, O2 = Blue, O18ac = Grey, O1 = Green. The distribution of dominant Sequence
712 types (STs) is shown in the middle ring. The colour code for sequence types: ST-117 =
713 Purple, ST-23 = Red, ST-95 = Black, ST-93 = Blue, ST-101 = Grey, ST-140 = Yellow. The
714 three representative APEC isolates from phylogroup B2, C and G used for Biolog phenotype
715 assays are highlighted with a red triangle symbol.

716

717 **Figure 2.** The distribution of annotated metabolic reactions screened by ModelSEED in 114
718 APEC isolates. This figure illustrates the core (Grey) and pan (Blue) metabolic content of the
719 114 APEC isolates used in this study. The core and pan content were defined by the extent of
720 content shared among all isolates. The percentage shown represents the core reactions of the
721 total.

722

723 **Figure 3.** Distribution of accessory metabolic reactions in APEC (n=114) among different
724 phylogenetic groups, clustered using a tree based on single nucleotide polymorphisms.
725 Colour code of phylogroups (from top to bottom): Yellow = A; Green = B1; Orange = C;
726 Dark blue = E; Purple = D; Light blue = F; Red = B2; Pink = G. Presence of metabolic
727 reactions is represented using a filled square, absence of the reaction by an open square. Only
728 the 209 accessory reactions of the total of 1,848 reactions (pan) are shown here; a full
729 overview of reactions and distribution in the 114 APEC is provided in Supplementary

730 Material 2, Table S9. The functional categories of accessory reactions is also shown here.
731 Colour code of functional categories of reactions (from left to right): Black = Amino acid
732 metabolism; Blue = Carbohydrate metabolism; Red = Cell wall/ membrane/ envelope
733 metabolism; Orange = Cofactor and prosthetic group metabolism; Dark blue = Energy
734 production and conversion; Light blue = Inorganic ion transport and metabolism; Beige =
735 Lipid metabolism; Green = Nucleotide metabolism; Gray = Other; Purple =
736 Transport/exchange reaction; Dark pink = unassigned (* Unassigned reactions were not
737 involved in reconstruction of APEC GEM or phylogroup-specific Sub-GEMs, as determined
738 as dead-end reactions by gap-filling).
739
740 **Figure 4.** The distribution of the final defined reactions (Grey) in the APEC model after gap-
741 filling and manual curation compared to the annotated metabolic reactions (Blue) screened by
742 ModelSEED in 114 APEC isolates. The percentage shown represents the fraction of reactions
743 retained in the model.
744
745 **Figure 5.** The computational growth prediction (Left) and phenotypic observation (Right) of
746 wildtype and single gene knockout isolates under constrained nutrient input. Figure 5A shows
747 the effect of inactivation of *lysA* with/without medium supplementation with lysine; an
748 asterisk indicates lack of growth demonstrating the essential role of the *lysA* gene. Figure 5B
749 shows the effect of inactivation of *potE* with/without medium supplementation with
750 putrescine, which had no effect as the gene is not essential. The computational prediction is
751 generated by APEC GEM and data is measured by flux unit (mmol gDW^{-1} h^{-1}); The
752 phenotypic observation is measured using M9 minimal medium and the optical density is
753 recorded after 24 hour incubation (37 °C) at a wavelength of 600 nm in 96-well plate. The
754 error bar shows standard deviation (SD) of the mean of 3 biological replicates. Asterisks

755 indicate an absence of growth.

756

757 **Figure 6.** Scatter plots comparing computational predictions from the *E. coli* K-12 *i*JO1366
758 model and the APEC model for three APEC isolates representing phylogroups B2, C and G,
759 with Biolog phenotypic observations with single nutrient source utilisation. The APEC
760 isolates used are from phylogroup C (SAP0009), G (SAP0545) and B2 (SAP0474). The
761 scatter plots show the model prediction (X-axis) using *i*JO1366b (Left) and APEC model
762 (Right) and phenotypic observation using Biolog measurement (Y-axis). The correlation
763 value shown above graphs is the Pearson correlation coefficient r^2 value.

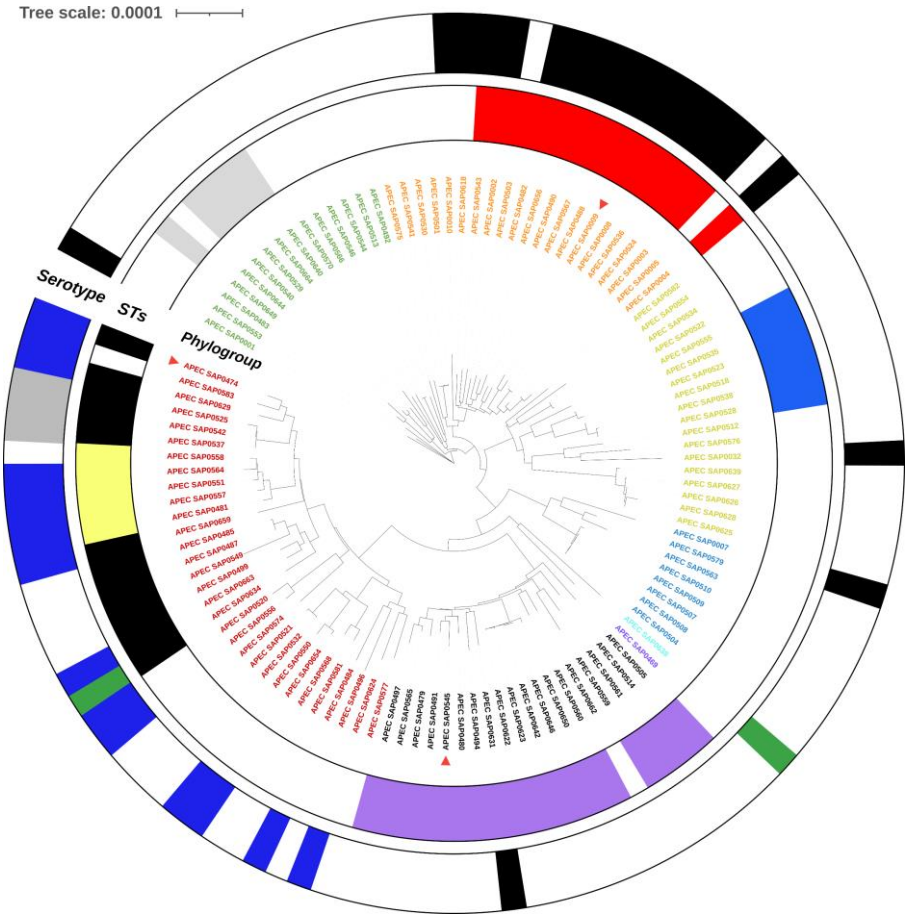
764

765 **Figure 7.** Scatter plots comparing computational predictions from the APEC phylogroup-
766 specific sub-models for phylogroups B2, C and G with the general APEC model for three
767 APEC isolates representing phylogroups B2, C and G, with Biolog phenotypic observations
768 with single nutrient source utilisation. The APEC isolates used are from phylogroup C
769 (SAP0009), G (SAP0545) and B2 (SAP0474). The scatter plots show the model prediction
770 (X-axis) using *i*JO1366b (Left) and APEC model (Right) and phenotypic observation using
771 Biolog measurement (Y-axis). The correlation value shown above graphs is the Pearson
772 correlation coefficient r^2 value.

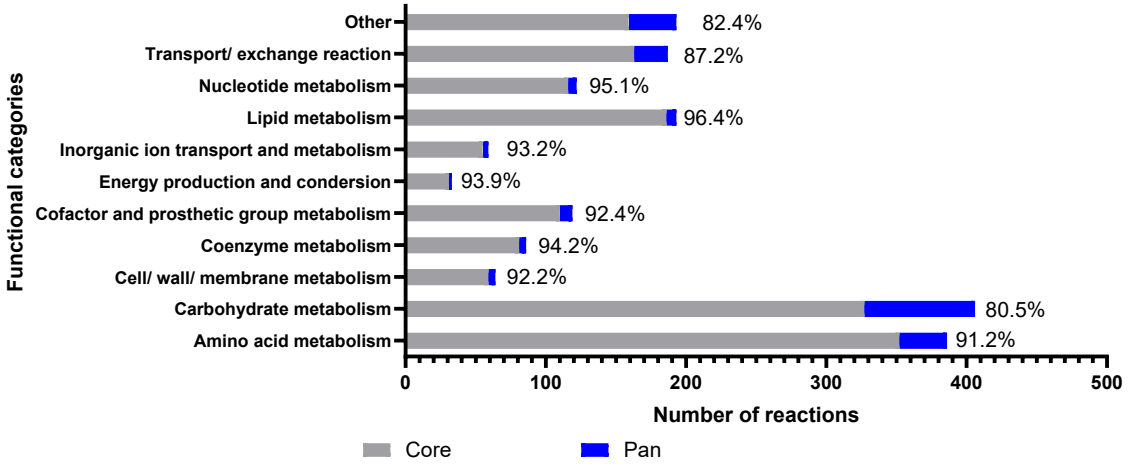
773

774 **Figure 8.** The comparison between *in silico* prediction and M9 minimal medium
775 measurements of different phylogroup APEC isolates on utilisation of different concentration
776 of 3-hydroxyphenoylacetic acid (3-HPAA). The left figures show the *in silico* prediction
777 using phylogroup-based sub-model (C, G and B2) under the constrain of 3-HPAA exchange
778 reaction lower bound adjusted to 0, -10 (low), -100 (medium) and -1000 (high), respectively
779 (-1000 refers to unlimited uptake of inputs). The right figures show the phenotypic

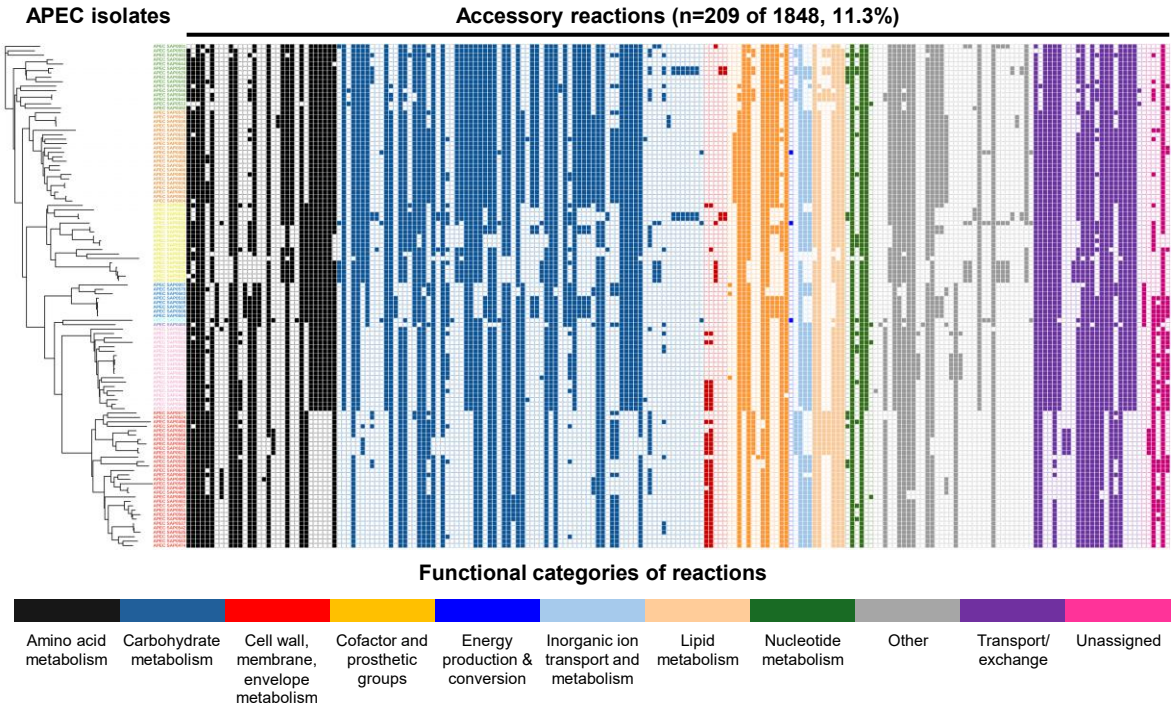
780 measurements in M9 minimal media with supplement of 0 mM (zero), 0.5 mM (low), 0.75
781 mM (medium), and 1 mM (high) 3-hydroxyphenoylacetic acid as sole carbon source. The
782 measurements indicate the mean of three biological endpoint (t=18h) readings. More detailed
783 growth curves are provided in Supplementary Material 1, Figures S4 and S5. Asterisks
784 indicate an absence of growth.
785



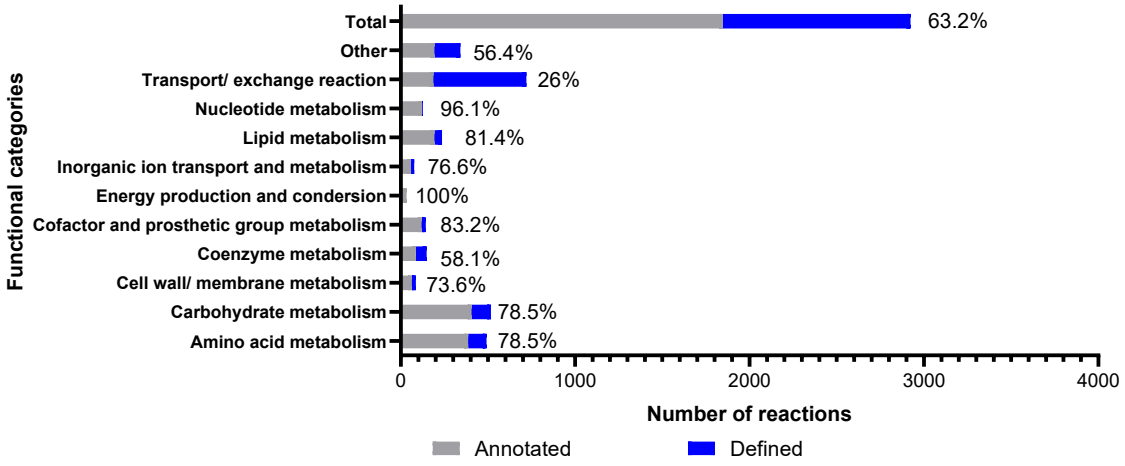
Huijun Long et al, Figure 1



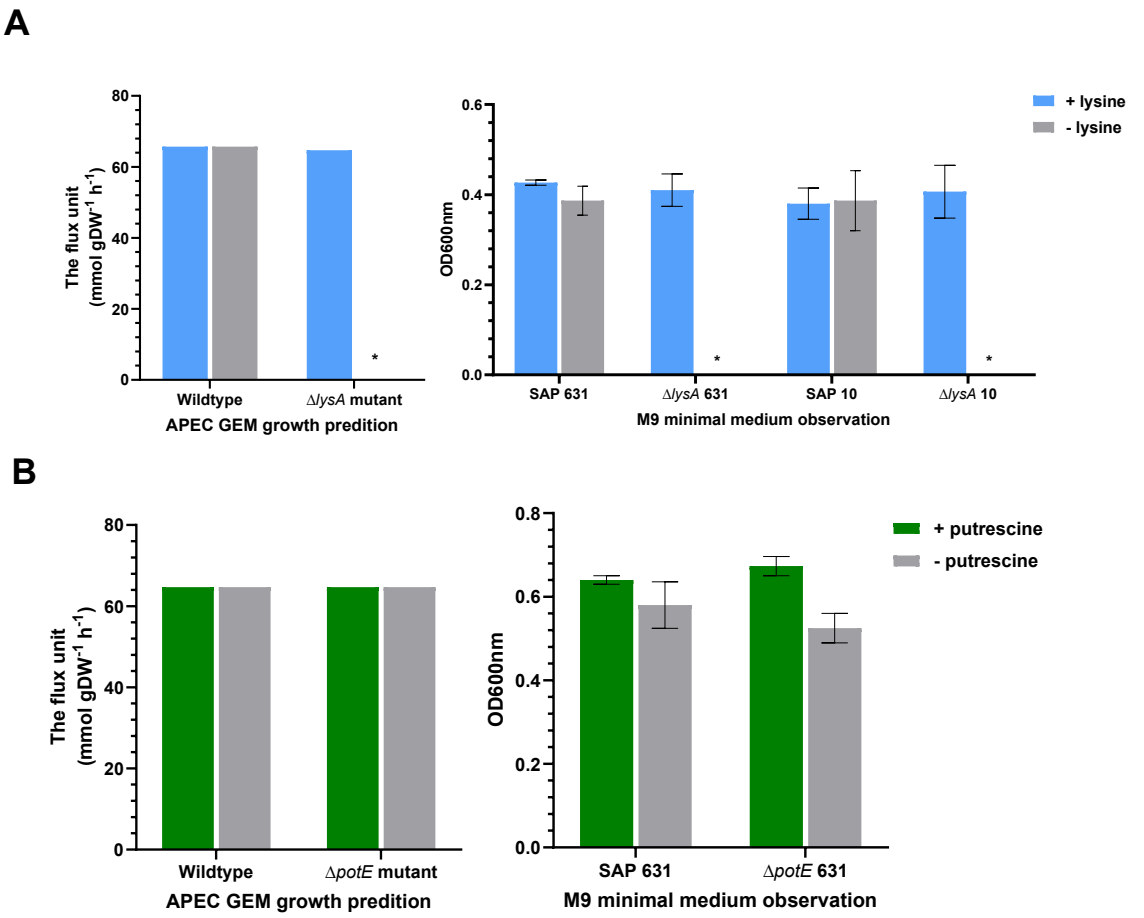
Huijun Long et al, Figure 2



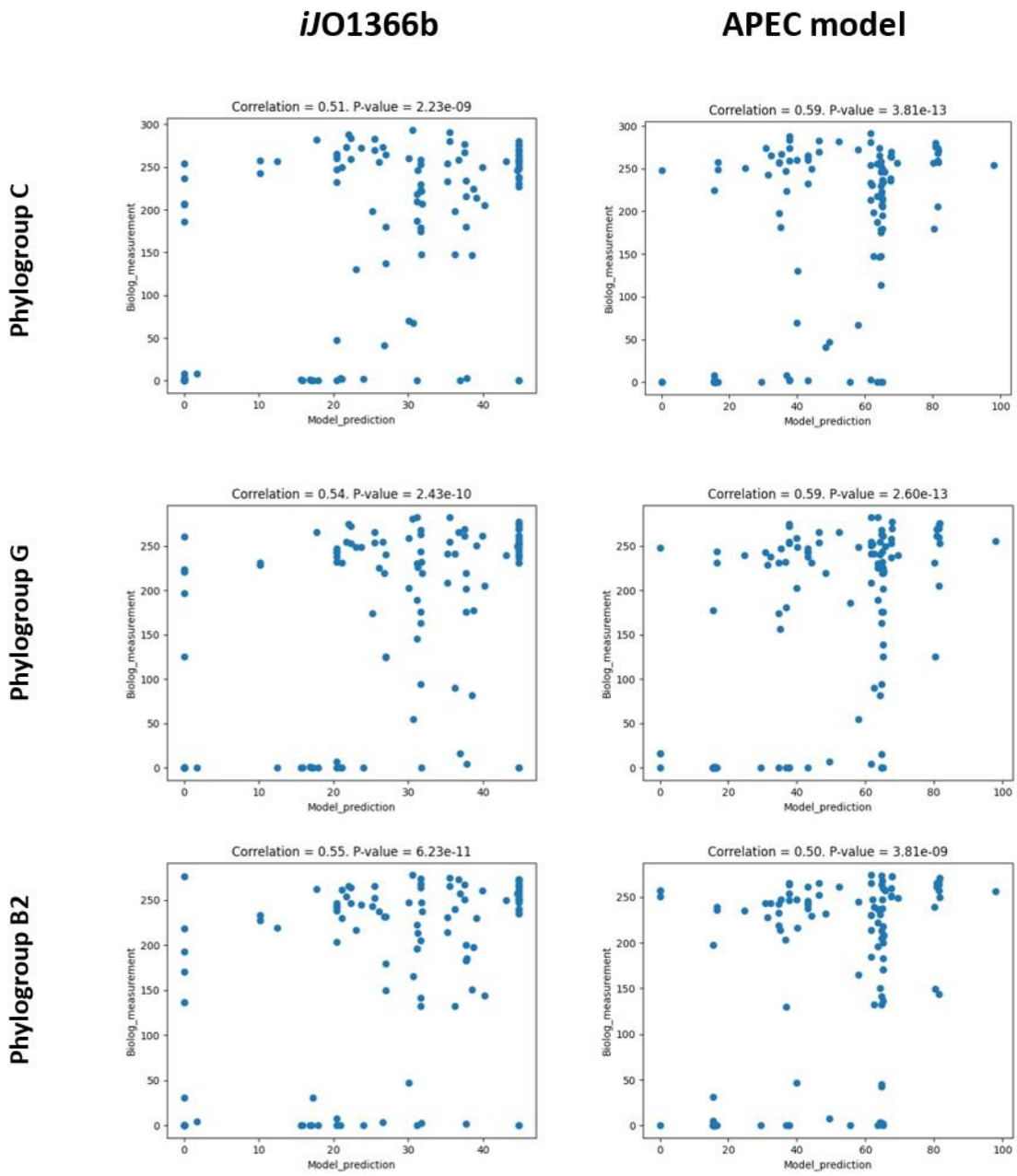
Huijun Long et al, Figure 3



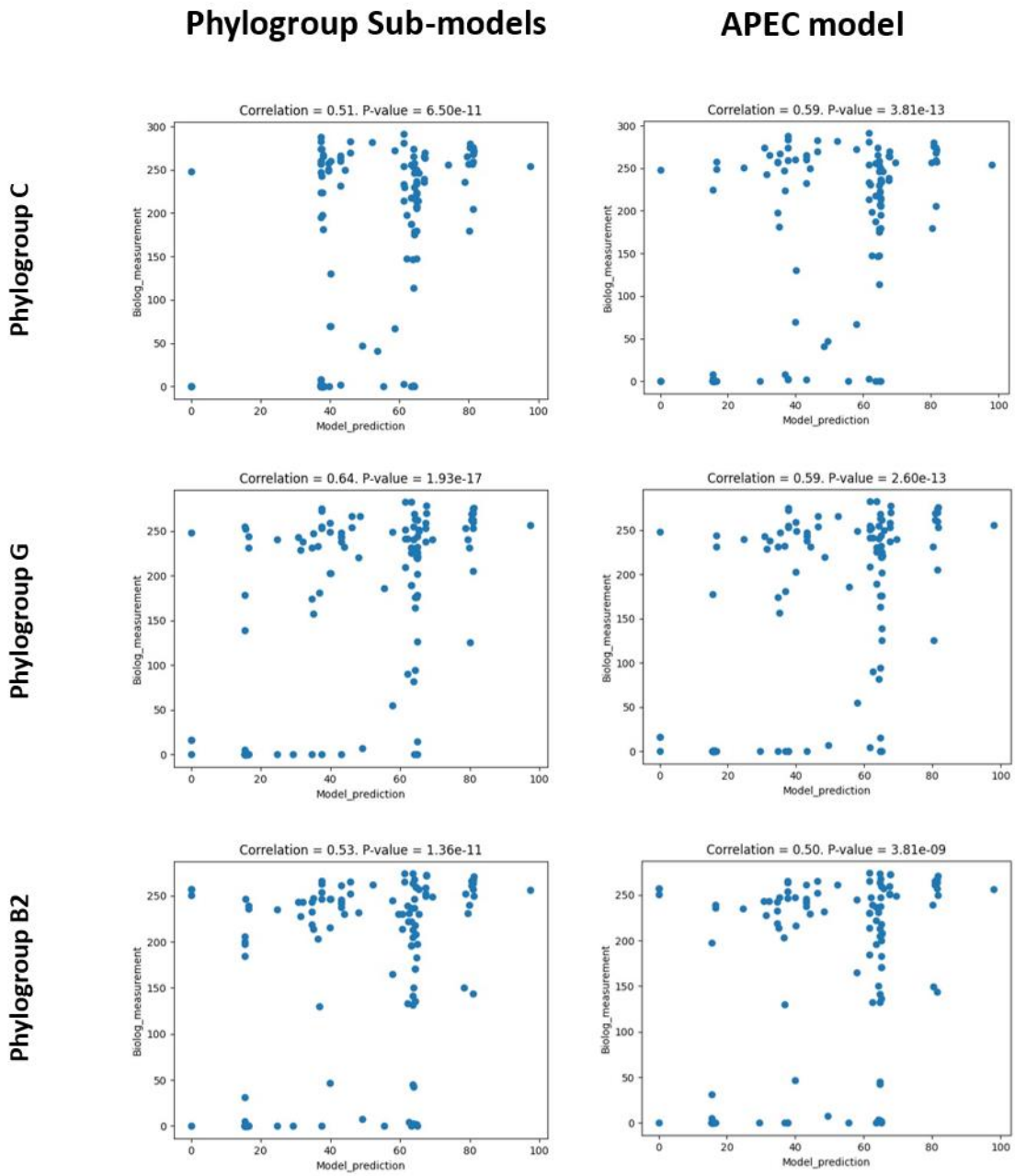
Huijun Long et al, Figure 4



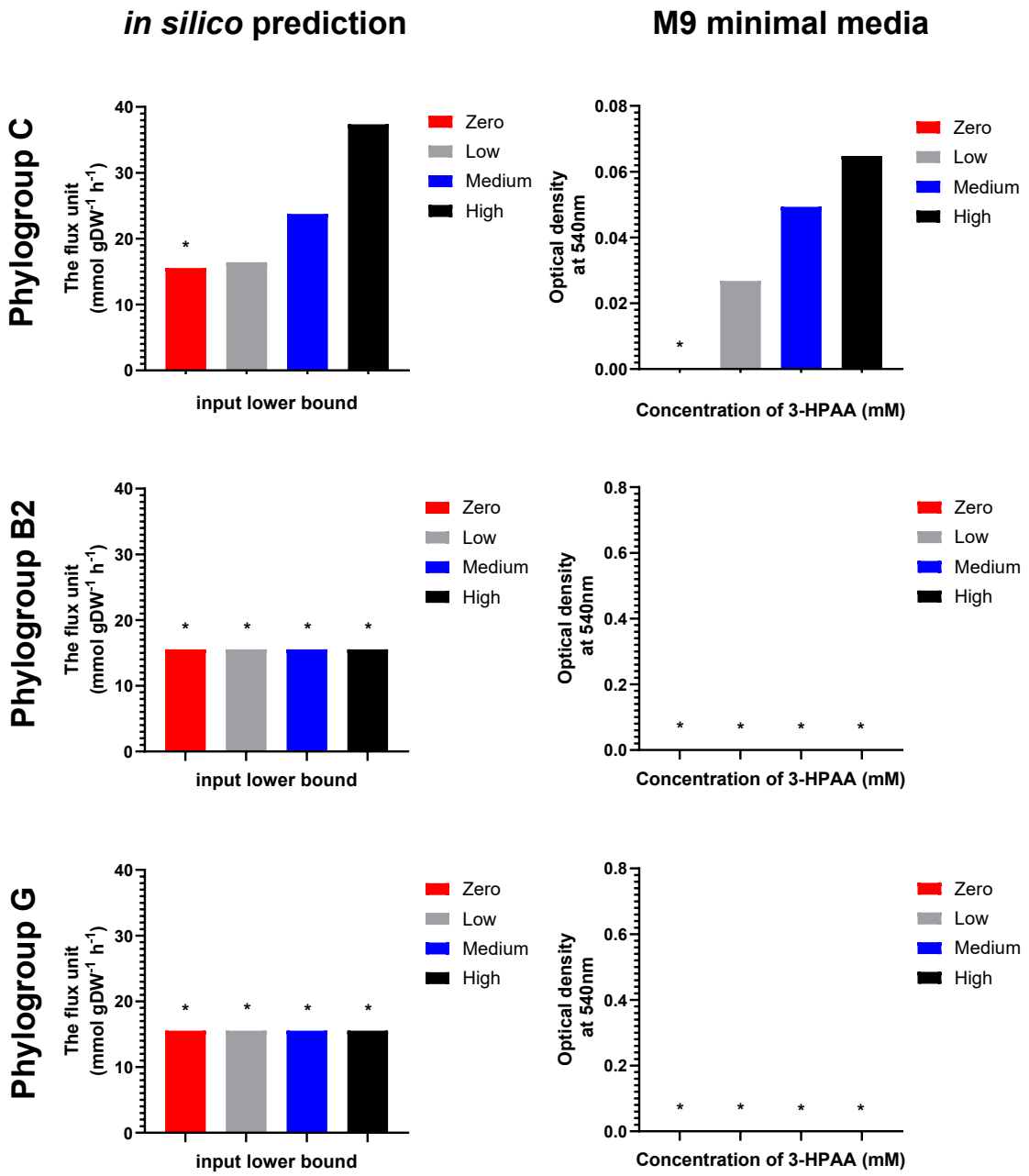
Huijun Long et al, Figure 5



Huijun Long et al, Figure 6



Huijun Long et al, Figure 7



Huijun Long et al, Figure 8