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 eight 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 Table 1, whereas data on gene essentiality, Biolog measurements and model growth simulations have been included in Supplementary Tables 2-4, 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 Tables 5-7, respectively.

Funding information. None reported.

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.

Abbreviations. SRA, Sequence Read Archive; APEC, Avian Pathogenic *Escherichia coli*; ExPEC, Extra-Intestinal Pathogenic *E. coli*; GEM, Genome scale metabolic reconstruction; AMR, antimicrobial resistance; AMP, ampicillin; Cm, chloramphenicol; CAT,

chloramphenicol resistance cassette; PCR, Polymerase Chain Reaction; RT-qPCR, reverse transcription quantitative-PCR; 3-HPAA, 3-hydroxyphenylacetate; SBC, stoichiometrically balanced cycles; BCR, biomass consumption reactions; FBA, flux balance analysis.

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

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.

IMPACT STATEMENT

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

INTRODUCTION

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

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

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

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

METHODS

Bacterial strains and growth conditions

114 APEC isolates were obtained from the SAP culture collection at the School of Veterinary Medicine, University of Surrey and used for APEC model construction (Supplementary Table 1). These APEC isolates were collected from commercial laying hens or broilers in the UK and Germany at *post-mortem* examination and were submitted for diagnostic or research activity undertaken in the School of Veterinary Medicine at the University of Surrey between 2014 and 2017. All isolates were stored in BHI broth supplemented with 50% glycerol at -80 °C until required for analysis. Prior to analysis, the APEC isolates were streaked onto MacConkey agar 3 plates, Nutrient agar or Luria-Bertani (LB) and incubated aerobically at 37°C for 16-24 hours, unless otherwise stated.

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

APEC isolates were selected from the representative phylogroups B2, C and G and used for Biolog analysis (N=3), M9 minimal medium cultivation analysis (N=13) and gene inactivation analysis. The isolates used for the analysis are listed in Supplementary Table 1. Three biological and three technical replicates were performed for each isolate. The APEC isolates were pre-cultured on R2A agar plates for Biolog assays on 190 carbon sources (PM1 and PM2 MicroPlate^{TM}), as well as 59 Phosphorus and 35 Sulphur Sources (PM4A MicroPlate^{TM}). The cell suspensions and plate inoculation were undertaken according to the instructions from the manufacturer. The Biolog plates were incubated and analysed in Biolog's OmniLog® instrument at 37 °C for 24h, aerobically. Readings were taken every 15 min. M9 minimal media (1× M9 salt, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, and 0.4% glucose if

required) was used to test the metabolic ability of isolates in a defined synthetic medium. Three concentrations (1 mM, 0.75 mM, and 0.5 mM) of 3-HPAA were tested in this study. L-lysine or putrescine (1 mM) was also tested in this study for the cultivation analysis of gene knockout mutants. The 96 well plates were cultured and read in Spark® microplate reader (TECAN) at 37 °C for 24h, aerobically. Readings were taken every 15 min.

DNA extraction and Next generation sequencing (NGS)

The high molecular weight genomic DNA of APEC isolates was extracted and purified using a DNA Wizard® genomic purification kit (Promega), according to the instructions from the manufacturer. The extracted DNA samples were sent to the Animal and Plant Health Agency (Weybridge, UK) for sequencing on the Illumina MiSeq platform, providing 150 nt paired end reads. The obtained raw DNA sequences were *de novo* assembled in contigs using Shovill version 1.1.0 (https://github.com/tseemann/shovill) using version 3.14 of the Spades assembler [18] with minimum coverage of 2.00× and minimum length of 200 nt. Quality assessment of genome assemblies were taken by QUAST version 5.0 [19], with required N_{50} and L_{50} values of > 50 kb and < 20, respectively. Prokka version 1.14 [20] was used for initial annotation of genome sequences, which produced relevant output files compatible for downstream analyses. The phylogenetic grouping was analysed on the *in-silico* based ClermonTyping program [21]. *In silico* serotyping was performed using Abricate version 1.0.1 (https://github.com/tseemann/abricate) with the SerotypeFinder [22] and EcOH [23] databases. A phylogenetic tree based on whole genome single-nucleotide polymorphisms (SNPs) was generated using ParSNP version 1.2 [24], resulting in a phylogenetic tree highlighting genomic similarity and sharing of core genome of all isolates.

Construction of APEC metabolic models

The FASTA files of 114 APEC samples generated from Prokka were submitted to the RAST server (https://rast.nmpdr.org/) for genome annotation [25]. The RAST annotation was automatically submitted to ModelSEED (https://modelseed.org/genomes/) for further model draft construction [26]. Model gap-filling is an essential step to investigate the missing reactions, which apply a mixed-integer linear program to define the minimal number of reactions required for the customer-defined collection of reactions. Several algorithms have been reported and applied in model gap-filling [27, 28]. COBRA Toolbox v3.0 [29] was used in this study and the database was based on ModelSeed. Further manual curation was performed when model was unable to be corrected by gap-filling steps, for instance: unnecessary reactions must be on for model running, and to avoid stoichiometrically balanced cycles which are a subset of contiguous reactions that forms a loop or internal network reactions, that carry fluxes without any exchange reactions involved. The model was adjusted based on the literature (basic nutrient requirement for cell growth and common pathways) and reference model *i*JO1366 [14]. The correction procedure included adding/deletion of reactions, adjusting of reversibility/directionality of reactions, and change of biomass consumption reactions.

GEM analysis

The growth prediction analysis was performed on the SurreyFBA 2.0 beta with JyMet2 [30]. The model was imported as a SurreyFBA formatted file, meanwhile a specific format (pfile) was also introduced, in which the media setting can be adjusted via changing the upper/lower bound of each exchange reaction. The computational prediction of growth rate was based on typical flux units (millimole per gram dry weight per hour, mmol gDW^{-1} h^{-1}). The default nutrient sources were glucose (carbon, C), ammonia (NH_3), phosphate (PO_4^{2-}) and sulphate (SO_4^{2-}). When examining the usage of a particular nutrient source, the exchange

reaction of its corresponding default source was adjusted to zero (reaction shut down). In addition, growth simulation studies were performed using the *E. coli* K-12 model *i*JO1366 [14] for comparative analysis. Due to the differences in default nutrient requirements in modelling, two growth simulations (*i*JO1366a and *i*JO1366b) were performed within two different input nutrient settings. The input nutrients of *i*JO1366a were adjusted as in the APEC model used here, where the exchange reactions of selenate (SeO_4^{2-}), selenite (SeO_3^{2-}) and tungstate (WO_4^{2-}) were disabled as they are absent from the APEC model; whilst these three exchange reactions were enabled in *iJO1366*b to perform the optimal conditions for *i*JO1366. In both *i*JO1366a and *i*JO1366b, the exchange reactions of molybdate (MoO_4^{2-}) and nickel (Ni^{2+}), which are also absent from the APEC model, were kept enabled, as they are essential for *i*JO1366. The phenotypic and computational measurements were compared using a Pearson correlation coefficient test to measure the strength and direction of these two variables.

Inactivation of genes in *E. coli*

Genes were inactivated using the Red Lambda Recombination assay [31]. Two gene targets (*lysA* and *potE*) were selected based on gene essentially predicted by model. The target genes were disrupted by deletion of approximately 450~550 bp sequence in the middle part of the gene sequence, and replaced with a chloramphenicol resistance cassette (CAT) as selective marker. These cloning fragments were constructed using gBlocks gene fragments in tubes service (gBlocks™, Integrated DNA Technologies). APEC isolates were selected for gene inactivation based on their phylogroup and susceptibility to ampicillin (AMP) and chloramphenicol (Cm) in MH broth. Competent cells were electroporated with the assistant plasmid pSIJ8, containing the AMP resistance gene. The competent cells with pSIJ8 were selected by ampicillin plate and PCR. The expression of the pSIJ8-encoded λ-Red

recombination system was induced with 0.5M arabinose. The cloning fragment was then introduced through electroporation, with cells recovered at 30°C (as plasmid pSIJ8 is temperature sensitive) on a shaking platform for 1~ 2h. Mutants were selected on chloramphenicol-containing LB plates and verified by PCR. We also designed a reverse transcription quantitative-PCR (RT-qPCR) to confirm the absence of *lysA* and *potE* mRNA transcripts in mutant candidates. The primers and probe of the real time quantitative PCR were designed to amplify the sequences deleted from the wildtype isolate during the construction of cloning fragment (Supplementary Table 8).

RESULTS

Genomic characterisation of APEC isolates for metabolic model generation

To test whether the selected 114 APEC isolates were representative of the APEC pathotype, their phylogroups, dominant serotypes and sequence types were determined (Figure 1). The phylogenetic grouping of isolates based on core genome SNPs matched the distinctions in phylogroups and sequence types. *In silico* phylogroup typing showed that the majority of APEC isolates belonged to B2 (27.2%), followed by C (18.4%), G (16.7%), A (15.8%) and B1 (13.2%). The phylogroups D, E and F were underrepresented in this collection. The most prevalent serotypes were O78 (17%) and O2 (13%) and the most common sequence types (STs) were ST-117 (15.8%), ST-23 (12.3%) and ST-95 (10.5%). All these characteristics broadly matched earlier studies [4-7, 32], and hence we consider this collection of isolates broadly representative for the APEC pathotype.

Generation of the metabolic profile of the APEC pathotype

The metabolic profile of the APEC isolates was predicted using ModelSEED. A total of 1,848 different metabolic reactions were identified and annotated in the 114 APEC isolates included, and were defined as pan reactions. A total of 1,639 reactions (88.7%) were shared among all isolates, defined as core reactions, while 209 reactions (11.3%) were defined as accessory reactions. The reactions were categorised according to their predicted metabolic functions using ModelSEED, as shown in Figure 2. The most highly conserved metabolic reactions were responsible for lipid metabolism (96.4%), nucleotide metabolism (95.1%), coenzyme metabolism (94.2%) and energy production (93.9%). Reactions for alternative carbohydrate utilisation and amino acid metabolism made up 42.9% (792/1,848) of pan reactions and 52.6% (110/209) of accessory reactions. The metabolic reactions were

subsequently clustered with the phylogroups identified (Figure 3). In the accessory reactions, there was clustering of metabolic profiles of APEC isolates from the same phylogroups, especially in metabolism of alternative carbohydrates and amino acids including their transport/exchange capabilities. This suggests that certain lineages are likely adapted to niches or environments, where they may harbour an advantage in the utilisation of available nutrient. Notably, phylogroup B2 APEC isolates displayed distinct metabolic capabilities compared to the other phylogroups. Similar observations were also reported previously [33, 34]. Metabolic capabilities absent in phylogroup B2 were mainly associated with amino acids and carbohydrates, and related transport/exchange of metabolites. In contrast, metabolic reactions involved in membrane biogenesis, such as N-acetylneuraminic acid (Neu5Ac) metabolism, were predominantly clustered in phylogroup B2 and G. These reactions have been described to contribute to the pathogenicity of Gram-negative bacteria by protecting against the host immune system [35].

Characterisation of APEC GEM and phylogroup-specific sub-models

The 1,848 metabolic reactions identified in the APEC isolates were used to generate a draft genome-scale metabolic model using ModelSEED. Sixteen reactions were excluded as dead-end reactions. Eventually, the pan-genome metabolic model constructed in this study, referred as APEC GEM, was composed of 2,923 reactions (inclusive of gap-filling reactions and a total biomass reaction), 2,527 metabolites, and 2,242 genes (Figure 4). Notably, there were many reactions added during gap-filling algorithms and manual curation, such as spontaneous reaction for extracellular metabolites exchange, isomerization, putative reactions, which significantly change the distribution of functional categories compared to Figure 2. Based on KEGG annotations and the *E. coli* K-12 model [14], a large number of exchange reactions of extracellular metabolites were accumulated in the transport/exchange

functional category (Figure 4). This contrasted with the central metabolism reactions, such as those involved in energy production and nucleotide metabolism, which are generally conserved across *E. coli* lineages.

The APEC GEM was built as a pan-genome metabolic model, to represent the full metabolic capabilities in the complete APEC collection used in this study. Considering the distinct metabolic variances between different APEC lineages (Figure 3), phylogroup B2, C and G-specific models were generated from the APEC GEM to represent the metabolic capability and deficiency of each phylogroup. A brief summary of the reactions associated with different metabolic categories removed for sub-model construction for each phylogroup is shown in Supplementary Figure 1. In total, 133 reactions were deleted during the construction of sub-models. Notably, the same set of metabolic pathways were removed from more than one phylogroup. Most of deleted reactions were associated with carbohydrate metabolism, accounting for 42%, followed by amino acid metabolism (14.3%) and transport/exchange reactions (9%). A relatively high ratio of carbohydrate (13/19) and amino acid (45/57) metabolism reactions and transport/exchange reactions (9/12) were removed in phylogroup B2 in this study, consistent with prior observations. Besides, 39/57 of carbohydrate metabolism reactions were removed for phylogroup G.

Gene essentiality analysis of APEC GEM

Gene essentiality analysis is often adopted to evaluate the validity of GEM by comparing with experimentally obtained essential gene datasets [9, 16]. The APEC GEM was used to predict growth on glucose and glycerol as carbon sources. For growth on glucose or glycerol, the lower boundaries of the glucose and glycerol exchange reactions were set to -1,000, respectively. All 2,242 genes in APEC GEM were inactivated *in silico* individually as time and growth were simulated by flux balance analysis (FBA) (Supplementary Table 2). When

inactivation of a gene resulted in a growth rate above zero, it was considered as a non-essential gene. The gene essentiality predictions on glucose and glycerol in minimal media showed a consistent outcome using the APEC model in this study (Supplementary Table 2). In general, *in silico* inactivation of 208 out of 2,242 genes (9.3%) led to an absence of flux through the network and as such would lead to growth inhibition, and thus were determined to be essential genes in the APEC model. The distribution of these essential genes is shown in Supplementary Figure 2. Of these essential genes, 57 (27%) were associated with more than one reaction in this model. 13% were responsible for amino acid metabolism and 3% for carbohydrate metabolism. To validate these predictions, we decided against the use of the Keio collection grown on glucose/glycerol, as there are significant differences between *E. coli* K-12 and APEC. Hence two single gene knockout mutants were constructed in APEC for this purpose, in the essential *lysA* gene and non-essential *potE* gene. Both genes were present in 100% of the 114 APEC isolates analysed in this study.

The *lysA* gene is predicted to encode a diaminopimelate decarboxylase that is responsible for the biosynthesis of lysine. In the model, the *lysA* gene (model ID: peg18487) was determined to be an essential gene and encoded for chemical equation (model ID: R_rxn00313) as below ([c] represent for the metabolites in cytoplasm):

Rxn00313: (1) H^+*[c]+(1) meso-2,6-Diaminopimelate[c] => (1)* CO_2*[c]+(1) L-Lysine[c]*

The *potE* gene encodes a membrane protein mediating both the uptake (via proton symport mechanism) and excretion of putrescine (as a putrescine and ornithine antiporter). There is redundancy in putrescine transport with the PotABCD, PotFGHI, PlaP/YeeF, PuuP and SapBCDF systems [36-39]. In the model, the target gene *potE* (model ID: peg6625) was determined as non-essential and responsible for two putrescine transport reactions (model ID: R_rxn05867 and R_rxn10182). For the modelling purpose, the chemical equations of these

two reactions are represented as below ([e] represents for the extracellular metabolites; [c] represent for the metabolites in cytoplasm):

Rxn05687: (1) H^+[e] + (1) Putrescine[e] => (1) H^+[c] + (1) Putrescine[c]

Rxn10182: (1) Ornithine[c] + (1) Putrescine[e] <=> (1) Ornithine[e] + (1) Putrescine[c]

The *lysA* mutants generated in isolates SAP0010 and SAP0631 and *potE* mutant generated in isolates SAP0631 were analysed using glucoside as default carbon source. Other default nutrient sources were adjusted by general standard principle described in next section. The *in-silico* growth prediction of knockouts grown in with and without supplementation of L-lysine or putrescine were compared to the phenotypic screening using M9 minimal medium as illustrated in Figure 5. The Δ*lysA* mutants (Δ*lysA* 10-9 and Δ*lysA* 631-11) were observed to be auxotrophic for L-lysine (Figure 5), which indicates the fact that the *lysA* gene encodes the enzyme in the final and only lysine biosynthesis pathway. This was also revealed in APEC GEM knockout analysis that no biomass was produced when the rxc00313 was switched off. On the contrary, the supplementation of putrescine in M9 medium did not lead any significant change in growth between the Δ*potE* 631-12 mutant and its parental strain (Figure 5). The APEC GEMs correctly predicted the growth status of these knockout mutants, as well as their wildtype, in minimal media with and without supplementation.

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

To validate *in silico* model prediction with phenotypic observations, the growth (respiration) phenotype of three representative APEC isolates, chosen from phylogroups B2, C and G, were tested using the Biolog Phenotypic Microarray analysis (PM1, PM2A: carbon source plates, and PM4A: sulphate and phosphate plate). The respiration measurements are available in Supplementary Table 3**.** In total, 125 nutrient substrates were analysed in the APEC model, and 119 substrates were analysed in *i*jO1366 (Supplementary Table 4). The

correlation coefficient of all three comparison experiments showed mostly positive correlations; the *i*JO1366a model showed a weak positive correlation (r^2 values between 0.07 and 0.17, Supplementary Figure 3), while the *i*JO1366b and APEC models showed a moderate positive correlations (r^2 values between 0.50 and 0.59, Figure 6). Under the optimised media settings, the APEC model showed an improved performance with the representative phylogroups C (r^2=0.59, $p < 0.005$) and G (r^2=0.59, $p < 0.005$), while it showed decreased performance with phylogroup B2 (r^2=0.50, $p < 0.005$), compared to model *i*JO1366 (Figure 6).

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

We also explored the performance of phylogroup-specific APEC model based sub models. The default media setting of phylogroup based sub-models were modified from the APEC GEM, with the import of Fe(II)-Nicotianamine (Fe(II)-NA or Fe^{2+}-NA) added as it was required for model performance of sub-models for phylogroups C and G. In total, *in silico* flux balance analysis (FBA) of 142 nutrient substrates was performed with phylogroup-specific sub-models and then compared to growth simulation of APEC GEM according to Biolog data (Supplementary Table 4). In Figure 7, the relationship between the computational prediction from the sub-models and their Biolog measurements showed moderate positive correlation (r^2 values of 0.51, 0.64 and 0.3 for phylogroups C, G and B2, respectively). Compared to the computational prediction by the pan model APEC GEM, the sub-models of phylogroup C performed less well under similar media input, while the phylogroup G and B2 sub-model showed improved performance compared to APEC model.

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

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

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

This reaction is catalysed by 4-hydroxyphenylacetate 3-monooxygenase oxygenase, encoded by gene *hpaB* (4-hydroxyphenylacetate 3-monooxygenase, EC 1.14.13.3). The *hpaB* gene was present in all phylogroup C (n=21/21) APEC, while absent in phylogroup B2 (n=0/31) and G (n=0/19) APEC. The *hpaB* gene was detected in phylogroups A and B1 as well, but these constitute minor phylogroups for APEC. The metabolic variation on 3-HPAA metabolism was measured using M9 minimal medium. As is shown in Supplementary Figure 4, the five phylogroup C isolate were all able to grow with the three different concentrations of 3-HPAA tested, whereas the six phylogroup B2 and two phylogroup G isolates were unable to grow with 3-HPAA as the single carbon source (Supplementary Figure 5), but grew in LB.

The metabolism of 3-HPAA in *E. coli* has been described previously [40, 41]. However, the intracellular/extracellular reaction of 3-HPAA in *E. coli* has not yet been documented in the ModelSEED database. Therefore, we manually added the exchange reaction of 3-HPAA into the sub-models, in order to perform the growth simulation with/without supplementary of 3-HPAA. The additional reactions are presented as below (cpd03320 is the compound ID of 3-HPAA in APEC GEM and relative sub-models, *e* represent for extracellular; *b* represents for periplasm; *c* represent for cytoplasm):

R_e2c_cpd03320: M_cpd03320_e = M_cpd03320_c

R_EX_cpd03320: M_cpd03320_e = M_cpd03320_b

The lower bound (lb) of the 3-HPAA exchange reaction was adjusted to 0, -10, -100 and -1000 in the APEC GEM sub-model B2, C and G, to simulate APEC growth in response to

four different concentrations of 3-HPAA. The growth prediction of different sub-groups and M9 experimental observation are summarised in Figure 8. In sub-models B2 and G, the relatively minimal biomass (~15) was produced with the lower bound of 3-HPAA exchange reaction adjusted to 0, -10, -100, and -1000 (increasing uptake), which was assumed as no growth for the modelling purpose. This agreed with the M9 minimal media experimental data, where isolates from phylogroup B2 and G were unable to use 3-HPAA as the sole carbon source. However, growth was predicted in sub-model C, matching the setting of increased uptake of 3-HPAA (Figure 8).

DISCUSSION

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

The comparative genomic analysis performed in this study confirmed that the APEC pathotype is genetically variable, with phylogroups B2, C and G being the dominant lineages [4, 5, 42]. The resulting genome scale metabolic profile of the 114 APEC isolates (**Figure 3**) showed strong conservation in central metabolism involved in energy production, with the ratio of core metabolism versus accessory metabolism higher than in previous studies, where multiple *E. coli* pathotypes and commensal isolates were involved [9, 43]. The major metabolic variances between isolates are associated with metabolism of carbohydrates and amino acids, as well as exchange/transport reactions. This was confirmed using Biolog metabolic analysis performed in this study. As shown in Figure 3, the phylogroup B2 isolates showed distinct metabolic features (mostly missing metabolic reactions) compared to others, which was previously also observed [33, 44]. This is consistent with the view that loss and acquisition of genes during evolution differentiates isolates into different lineage with unique metabolic profiles and adaptation to specific niches.

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

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

Although no significant improvement in the association strength between computational prediction of sub-models and the phenotypic measurement of the corresponding APEC isolates were observed, the APEC model showed its potential and plasticity as a pan model

used for the construction of phylogroup B2, C and G sub-models. The phenotypic difference in catabolism of 3-HPAA in phylogroup B2, C and G were successfully predicted by those phylogroup based sub-models as a proof of principle. Interestingly, 3- hydroxyphenylacetate and 4-hydroxyphenylacetate are the main phenolic acids derived from flavonoid quercetin (which is commonly added to poultry feed) in the colorectum [45, 46]. Therefore, these compounds are likely to accumulate in the avian intestine and its utilisation could be a hypothetical nutrient advantage in the intestinal microbial community. The phylogroup C sub-models not only predicted the growth of phylogroup C APEC isolates using 3-hydroxyphenylacetate as sole carbon source, but also reflected the biomass increases alongside the enhancing lower bound setting of 3- hydroxyphenylacetate (Figure 8, Supplementary Figure 4).

The models generated in this study can be further exploited for the identification of additional pathways that are unique to APEC lineages. Although APEC are extraintestinal pathogens, potential APEC isolates are present in the intestinal *E. coli* population [7], and may form a reservoir for future infections within flocks. These models may inform the targeted suppression of such APEC populations thereby benefitting avian health and welfare in the long term. The approach taken may also be applicable for other genetically diverse intestinal or extraintestinal pathogens dispersed over multiple lineages. Furthermore, in the future, the rapid progress in software development (e.g. Bactabolize [47]), availability of APEC genome sequences and associated repositories may expedite the generation of genome scale metabolic models to revisit or improve existing metabolic models for APEC and other *E. coli* pathotypes.

REFERENCES

1. **Denamur E, Clermont O, Bonacorsi S, Gordon D**. The population genetics of pathogenic *Escherichia coli*. *Nat Rev Microbiol* 2021;19(1):37-54. doi: 10.1038/s41579-020-0416-x
2. **Kohler CD, Dobrindt U**. What defines extraintestinal pathogenic *Escherichia coli*? *Int J Med Microbiol* 2011;301(8):642-647. doi: 10.1016/j.ijmm.2011.09.006
3. **La Ragione RM, Woodward MJ**. Virulence factors of *Escherichia coli* serotypes associated with avian colisepticaemia. *Res Vet Sci* 2002;73(1):27-35. doi: 10.1016/s0034-5288(02)00075-9
4. **Cordoni G, Woodward MJ, Wu H, Alanazi M, Wallis T et al.** Comparative genomics of European avian pathogenic *E. coli* (APEC). *BMC Genomics* 2016;17(1):960. doi: 10.1186/s12864-016-3289-7
5. **Mageiros L, Meric G, Bayliss SC, Pensar J, Pascoe B et al.** Genome evolution and the emergence of pathogenicity in avian Escherichia coli. *Nat Commun* 2021;12(1):765. doi: 10.1038/s41467-021-20988-w
6. **Mehat JW, van Vliet AHM, La Ragione RM**. The Avian Pathogenic Escherichia coli (APEC) pathotype is comprised of multiple distinct, independent genotypes. *Avian Pathol* 2021;50(5):402-416. doi: 10.1080/03079457.2021.1915960
7. **Logue CM, Wannemuehler Y, Nicholson BA, Doetkott C, Barbieri NL et al.** Comparative Analysis of Phylogenetic Assignment of Human and Avian ExPEC and Fecal Commensal *Escherichia coli* Using the (Previous and Revised) Clermont Phylogenetic Typing Methods and its Impact on Avian Pathogenic *Escherichia coli* (APEC) Classification. *Front Microbiol* 2017;8:283. doi: 10.3389/fmicb.2017.00283
8. **Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT**. The evolution of metabolic networks of E. coli. *BMC Syst Biol* 2011;5:182. doi: 10.1186/1752-0509-5-182
9. **Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N et al.** Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A* 2013;110(50):20338-20343. doi: 10.1073/pnas.1307797110

10. **Nhu NTK, Rahman MA, Goh KGK, Kim SJ, Phan MD et al.** A convergent evolutionary pathway attenuating cellulose production drives enhanced virulence of some bacteria. *Nat Commun* 2024;15(1):1441. doi: 10.1038/s41467-024-45176-4
11. **Gu C, Kim GB, Kim WJ, Kim HU, Lee SY**. Current status and applications of genome-scale metabolic models. *Genome Biol* 2019;20(1):121. doi: 10.1186/s13059-019-1730-3
12. **Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR et al.** A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 2007;3:121. doi: 10.1038/msb4100155
13. **Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A et al.** iML1515, a knowledgebase that computes Escherichia coli traits. *Nat Biotechnol* 2017;35(10):904-908. doi: 10.1038/nbt.3956
14. **Orth JD, Conrad TM, Na J, Lerman JA, Nam H et al.** A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Mol Syst Biol* 2011;7:535. doi: 10.1038/msb.2011.65
15. **Reid CJ, Cummins ML, Borjesson S, Brouwer MSM, Hasman H et al.** A role for ColV plasmids in the evolution of pathogenic Escherichia coli ST58. *Nat Commun* 2022;13(1):683. doi: 10.1038/s41467-022-28342-4
16. **Kim D, Kim Y, Yoon SH**. Development of a Genome-Scale Metabolic Model and Phenome Analysis of the Probiotic Escherichia coli Strain Nissle 1917. *Int J Mol Sci* 2021;22(4):2122. doi: 10.3390/ijms22042122
17. **Archer CT, Kim JF, Jeong H, Park JH, Vickers CE et al.** The genome sequence of E. coli W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of E. coli. *BMC Genomics* 2011;12:9. doi: 10.1186/1471-2164-12-9
18. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M et al.** SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19(5):455-477. doi: 10.1089/cmb.2012.0021
19. **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29(8):1072-1075. doi: 10.1093/bioinformatics/btt086

20. **Seemann T**. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30(14):2068-2069. doi: 10.1093/bioinformatics/btu153

21. **Beghain J, Bridier-Nahmias A, Le Nagard H, Denamur E, Clermont O**. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microb Genom* 2018;4(7):e000192. doi: 10.1099/mgen.0.000192

22. **Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F**. Rapid and Easy In Silico Serotyping of *Escherichia coli* Isolates by Use of Whole-Genome Sequencing Data. *J Clin Microbiol* 2015;53(8):2410-2426. doi: 10.1128/JCM.00008-15

23. **Ingle DJ, Valcanis M, Kuzevski A, Tauschek M, Inouye M et al.** In silico serotyping of *E. coli* from short read data identifies limited novel O-loci but extensive diversity of O:H serotype combinations within and between pathogenic lineages. *Microb Genom* 2016;2(7):e000064. doi: 10.1099/mgen.0.000064

24. **Treangen TJ, Ondov BD, Koren S, Phillippy AM**. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;15(11):524. doi: 10.1186/s13059-014-0524-x

25. **Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ et al.** The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res* 2014;42(Database issue):D206-214. doi: 10.1093/nar/gkt1226

26. **Devoid S, Overbeek R, DeJongh M, Vonstein V, Best AA et al.** Automated genome annotation and metabolic model reconstruction in the SEED and Model SEED. *Methods Mol Biol* 2013;985:17-45. doi: 10.1007/978-1-62703-299-5_2

27. **Kumar VS, Maranas CD**. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS Comput Biol* 2009;5(3):e1000308. doi: 10.1371/journal.pcbi.1000308

28. **Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK et al.** Systems approach to refining genome annotation. *Proc Natl Acad Sci U S A* 2006;103(46):17480-17484. doi: 10.1073/pnas.0603364103

29. **Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A et al.** Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* 2019;14(3):639-702. doi: 10.1038/s41596-018-0098-2

30. **Gevorgyan A, Bushell ME, Avignone-Rossa C, Kierzek AM**. SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of

genome-scale metabolic reaction networks. *Bioinformatics* 2011;27(3):433-434. doi: 10.1093/bioinformatics/btq679

31. **Jensen SI, Lennen RM, Herrgard MJ, Nielsen AT**. Seven gene deletions in seven days: Fast generation of Escherichia coli strains tolerant to acetate and osmotic stress. *Sci Rep* 2015;5:17874. doi: 10.1038/srep17874

32. **Cummins ML, Reid CJ, Roy Chowdhury P, Bushell RN, Esbert N et al.** Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene. *Microb Genom* 2019;5(2):e000250. doi: 10.1099/mgen.0.000250

33. **Fang X, Monk JM, Mih N, Du B, Sastry AV et al.** Escherichia coli B2 strains prevalent in inflammatory bowel disease patients have distinct metabolic capabilities that enable colonization of intestinal mucosa. *BMC Syst Biol* 2018;12(1):66. doi: 10.1186/s12918-018-0587-5

34. **Gonzalez-Alba JM, Baquero F, Canton R, Galan JC**. Stratified reconstruction of ancestral Escherichia coli diversification. *BMC Genomics* 2019;20(1):936. doi: 10.1186/s12864-019-6346-1

35. **Lin BX, Qiao Y, Shi B, Tao Y**. Polysialic acid biosynthesis and production in Escherichia coli: current state and perspectives. *Appl Microbiol Biotechnol* 2016;100(1):1-8. doi: 10.1007/s00253-015-7019-x

36. **Guerra PR, Herrero-Fresno A, Ladero V, Redruello B, Dos Santos TP et al.** Putrescine biosynthesis and export genes are essential for normal growth of avian pathogenic Escherichia coli. *BMC Microbiol* 2018;18(1):226. doi: 10.1186/s12866-018-1355-9

37. **Guerra PR, Herrero-Fresno A, Pors SE, Ahmed S, Wang D et al.** The membrane transporter PotE is required for virulence in avian pathogenic Escherichia coli (APEC). *Vet Microbiol* 2018;216:38-44. doi: 10.1016/j.vetmic.2018.01.011

38. **Kurihara S, Suzuki H, Oshida M, Benno Y**. A novel putrescine importer required for type 1 pili-driven surface motility induced by extracellular putrescine in Escherichia coli K-12. *J Biol Chem* 2011;286(12):10185-10192. doi: 10.1074/jbc.M110.176032

39. **Thongbhubate K, Irie K, Sakai Y, Itoh A, Suzuki H**. Improvement of putrescine production through the arginine decarboxylase pathway in Escherichia coli K-12. *AMB Express* 2021;11(1):168. doi: 10.1186/s13568-021-01330-5

40. **Cooper RA, Skinner MA**. Catabolism of 3- and 4-hydroxyphenylacetate by the 3,4-dihydroxyphenylacetate pathway in Escherichia coli. *J Bacteriol* 1980;143(1):302-306. doi: 10.1128/jb.143.1.302-306.1980

41. **Prieto MA, Garcia JL**. Identification of the 4-hydroxyphenylacetate transport gene of Escherichia coli W: construction of a highly sensitive cellular biosensor. *FEBS Lett* 1997;414(2):293-297. doi: 10.1016/s0014-5793(97)01012-0

42. **Chen X, Liu W, Li H, Yan S, Jiang F et al.** Whole genome sequencing analysis of avian pathogenic Escherichia coli from China. *Vet Microbiol* 2021;259:109158. doi: 10.1016/j.vetmic.2021.109158

43. **Vieira G, Sabarly V, Bourguignon PY, Durot M, Le Fevre F et al.** Core and panmetabolism in Escherichia coli. *J Bacteriol* 2011;193(6):1461-1472. doi: 10.1128/JB.01192-10

44. **van der Hooft JJJ, Goldstone RJ, Harris S, Burgess KEV, Smith DGE**. Substantial Extracellular Metabolic Differences Found Between Phylogenetically Closely Related Probiotic and Pathogenic Strains of Escherichia coli. *Front Microbiol* 2019;10:252. doi: 10.3389/fmicb.2019.00252

45. **Bhutto ZA, He F, Zloh M, Yang J, Huang J et al.** Use of quercetin in animal feed: effects on the P-gp expression and pharmacokinetics of orally administrated enrofloxacin in chicken. *Sci Rep* 2018;8(1):4400. doi: 10.1038/s41598-018-22354-1

46. **Hong YJ, Mitchell AE**. Identification of glutathione-related quercetin metabolites in humans. *Chem Res Toxicol* 2006;19(11):1525-1532. doi: 10.1021/tx0601758

47. **Vezina B, Watts SC, Hawkey J, Cooper HB, Judd LM et al.** Bactabolize is a tool for high-throughput generation of bacterial strain-specific metabolic models. *Elife* 2023;12. doi: 10.7554/eLife.87406

LEGENDS TO FIGURES

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

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

Figure 3. Distribution of accessory metabolic reactions in APEC (n=114) among different phylogenetic groups, clustered using a tree based on single nucleotide polymorphisms. Colour code of phylogroups (from top to bottom): Yellow = A; Green = B1; Orange = C; Dark blue = E; Purple = D; Light blue = F; Red = B2; Pink = G. Presence of metabolic reactions is represented using a black mark, where the phylogroup colour is visible, this means absence of the reaction. In total, 1,848 reactions (pan) were annotated in the full panel of APEC collections, 1639 (88.7%) of which were shared among all isolates (100%) defined

as core, otherwise defined as accessory (11.3%). The functional distribution of accessory reactions was also show here. Colour code of functional categories of reactions (from left to right): Purple = Amino acid metabolism; Pink = Carbohydrate metabolism; Green = Cell wall/ membrane/ envelope metabolism; Orange = Cofactor and prosthetic group metabolism; Yellow = Energy production and conversion; Blue = Inorganic ion transport and metabolism; Beige = Lipid metabolism; Brown = Nucleotide metabolism; Gray = Other; Light blue = Transport/exchange reaction; Red = unassigned (* Unassigned reactions were not involved in reconstruction of APEC GEM or phylogroup-specific Sub-GEMs, as determined as dead-end reactions by gap-filling).

Figure 4. The distribution of the final defined reactions (Grey) in the APEC model after gap-filling and manual curation compared to the annotated metabolic reactions (Blue) screened by ModelSEED in 114 APEC isolates. The percentage shown represents the fraction of reactions retained in the model.

Figure 5. The computational growth prediction (Left) and phenotypic observation (Right) of wildtype and single gene knockout isolates under constrained nutrient input. Figure 5A shows the effect of inactivation of *lysA* with/without medium supplementation with lysine; an asterisk indicates lack of growth demonstrating the essential role of the *lysA* gene. Figure 5B shows the effect of inactivation of *potE* with/without medium supplementation with putrescine, which had no effect as the gene is not essential. The computational prediction is generated by APEC GEM and data is measured by flux unit (mmol gDW^{-1} h^{-1}); The phenotypic observation is measured using M9 minimal medium and the optical density is recorded after 24 hour incubation (37 °C) at a wavelength of 600 nm in 96-well plate. The error bar shows standard deviation (SD) of the mean of 3 biological replicates. Asterisks

indicate an absence of growth.

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

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

Figure 8. The comparison between *in silico* prediction and M9 minimal medium measurements of different phylogroup APEC isolates on utilisation of different concentration of 3-hydroxyphenoylacetic acid (3-HPAA). The left figures show the *in silico* prediction using phylogroup-based sub-model (C, G and B2) under the constrain of 3-HPAA exchange reaction lower bound adjusted to 0, -10 (low), -100 (medium) and -1000 (high), respectively (-1000 refers to unlimited uptake of inputs). The right figures show the phenotypic measurements in M9 minimal media with supplement of 0 mM (zero), 0.5 mM (low), 0.75

mM (medium), and 1 mM (high) 3-hydroxyphenoylacetic acid as sole carbon source. The measurements indicate the mean of three biological endpoint (t=18h) readings. More detailed growth curves are provided in Supplementary Figures 4 and 5. Asterisks indicate an absence of growth.

LEGENDS TO SUPPLEMENTARY FIGURES

Supplementary Figure 1. Summary of the deleted reactions for construction of phylogenetic group specific sub-models based on APEC model. A total of 133 reactions were removed in construction of phylogroups-specific sub-models for 114 APEC isolates in this study. Colour code of functional categories of reactions: Amino acid metabolism = Black; Carbohydrate metabolism = Blue; Transport/exchange reactions = Grey; Others* = Green. *Other reactions including Cell wall/ membrane/ envelope metabolism, Cofactor and prosthetic group metabolism, Energy production and conversion, Inorganic ion transport and metabolism, Lipid metabolism, Nucleotide metabolism and other described in Figure 2.

Supplementary Figure 2. The distribution of essential genes predicted by *in silico* analysis against glucose and glycerol as sole carbon sources, using the APEC model in different functional categories.

Supplementary Figure 3. Scatter plots comparing computational predictions from the *E. coli* K-12 *i*JO1366 model and the APEC model for three APEC isolates representing phylogroups B2, C and G, with Biolog phenotypic observations with single nutrient source utilisation. The APEC isolates used are from phylogroup C (SAP0009), G (SAP0545) and B2 (SAP0474). The scatter plots show the model prediction (X-axis) using *i*JO1366a (Left) and APEC model (Right) and phenotypic observation using Biolog measurement (Y-axis). The correlation value shown above graphs is the r^2 value.

Supplementary Figure 4. Growth of five representative Phylogroup C APEC isolates in M9 minimal media supplemented with 3-hydroxyphenylacetic acid (panels A-E) and in LB broth

(panel F), respectively. The growth curves show the mean of 3 biological replicates, using optical density measurements at 600 nm wavelength for every 30 min in 18h, with error bars indicating the standard error of the mean (SEM). Panels A-E illustrates the growth curves of phylogroup C APEC isolates in M9 minimal media supplemented with three concentrations of 3-hydroxyphenylacetic acid (1 mM, 0.75 mM and 0.5 mM). Panel F shows growth curves of all group C isolates tested in this study in LB broth.

Supplementary Figure 5. The growth curve of phylogroup B2 and G APEC isolates in M9 minimal media supplemented with 3-hydroxyphenylacetic acid (1 mM, 0.75 mM and 0.5 mM) and in LB broth, respectively. The growth curves show the mean of 3 biological replicates, using optical density measurements at 600 nm wavelength for every 30 min in 18h, with error bars indicating the standard error of the mean (SEM). Panels A-F shows the phylogroup B2 APEC isolates used in this study whilst panels G-H shows the phylogroup G APEC isolates used in this study. There was no growth observed with any of the phylogroup B2 and G isolates with any of the concentrations of 3-HPAA added.

SUPPLEMENTARY TABLES

Supplementary Table 1. List of Avian Pathogenic *E. coli* (APEC) isolates used in this study, including information on phylogroup, country of isolation, type of poultry, disease caused, in which analyses they were used in this study, the genome assembly accession numbers, and the SRA accession numbers of the sequencing reads (Excel spreadsheet).

Supplementary Table 2. Gene essentiality analysis data using the APEC GEM model (Excel spreadsheet).

Supplementary Table 3. Respiration measurements for APEC isolates SAP0474 (phylogroup B2), SAP0545 (phylogroup G) and SAP0009 (phylogroup C), obtained using the Biolog system, using the PM1 (carbon), PM2 (carbon) and PM4A (phosphate and sulphate) plates (Excel spreadsheet).

Supplementary Table 4. Growth predictions for the APEC GEM and submodel GEMs for phylogroups B2, C and G, expressed as a typical flux units (mmol gDW^{-1} h^{-1}) in response to alternative single nutrient source. The default sources of four main elements were glucose (C), ammonia (NH_3), phosphate (P_i) and sulphate (SO_4^{2-}) respectively (Excel spreadsheet).

Supplementary Table 5. The reactions included in the APEC GEM and the submodels for phylogroups A, B1, B2, C and G (Excel spreadsheet, individual tabs). The source files are available via Figshare (https://doi.org/10.6084/m9.figshare.28067993).

Supplementary Table 6. Contents of growth media pfiles for the submodels, carbon sources,

sulphate sources and phosphate sources, as well as the files for the *lysA* and *potE* mutants (Excel spreadsheet, individual tabs). The source files are available via Figshare (https://doi.org/10.6084/m9.figshare.28067993).

Supplementary Table 7. The reactions of the APEC submodels for phylogroups B2, C and G, used for the growth experiments with 3-HPAA supplementation (Excel spreadsheet, individual tabs). The source files are available via Figshare (https://doi.org/10.6084/m9.figshare.28067993).

Supplementary Table 8. Oligonucleotide primers used in this study.

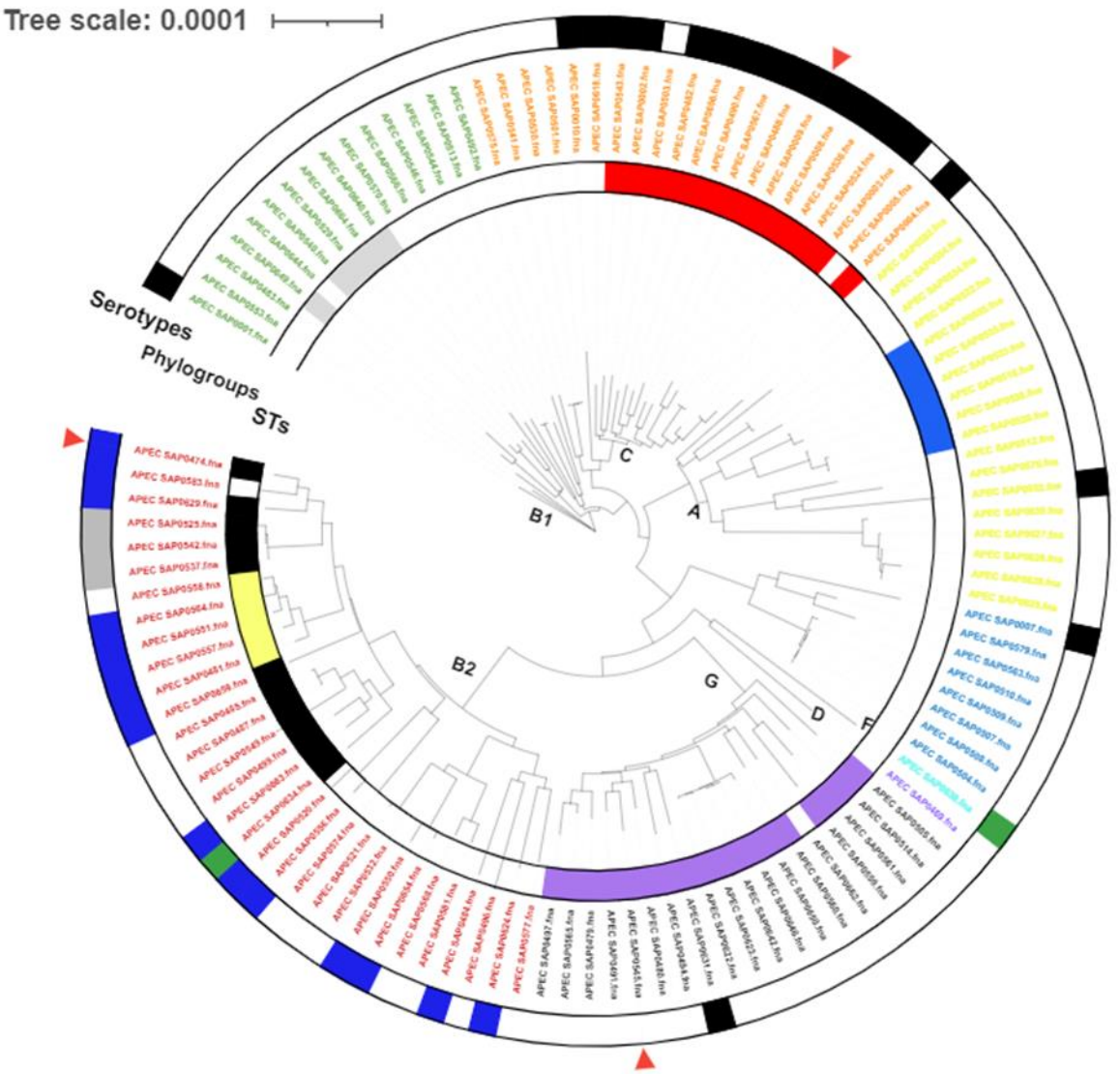

Figure 1. APEC population structure for the isolates selected for GEM construction in this study assigned with phylogroup (n=114). Colour code of phylogroups: A = Yellow, B1 = Green, B2 = Red, C = Orange, D = Purple, E = Blue, F = Light blue, G = Black. The distribution of dominant serotypes shown in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

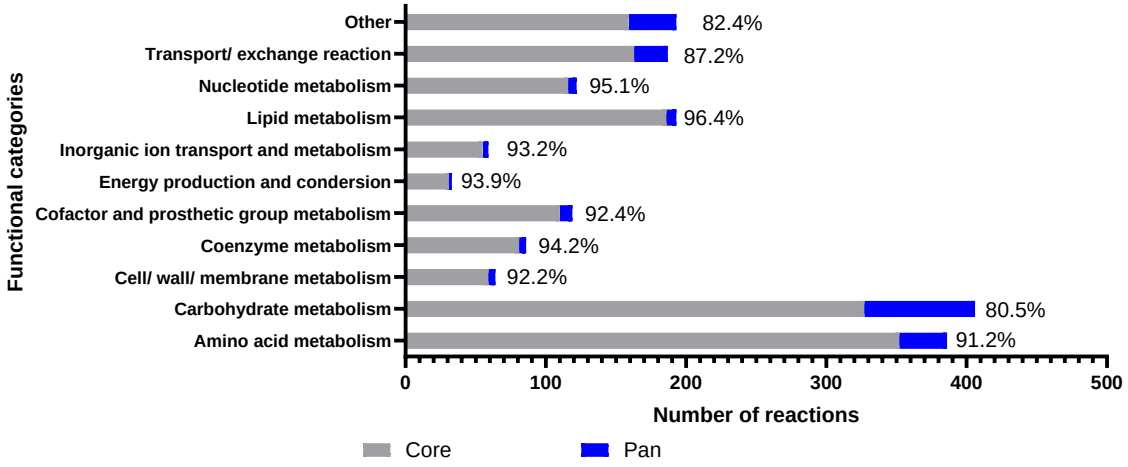

Figure 2. The distribution of annotated metabolic reactions screened by ModelSEED in 114 APEC isolates. This figure illustrates the core (Grey) and pan (Blue) metabolic content of the 114 APEC isolates used in this study. The core and pan content were defined by the extent of content shared among all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

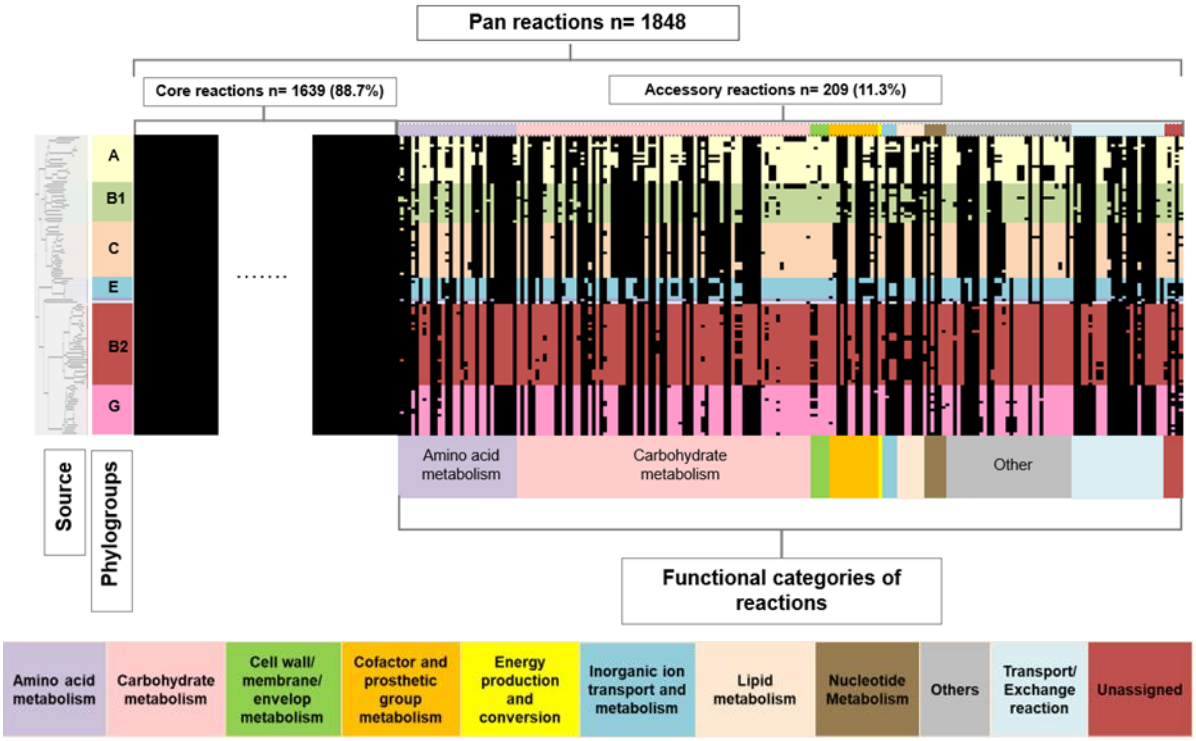

Figure 3. Distribution of accessory metabolic reactions in APEC (n=114) among different phylogenetic groups, clustered using a tree based on single nucleotide polymorphisms. Colour code of phylogroups (from top to bottom): Yellow = A; Green = B1; Orange = C; Dark blue = E; Purple = D; Light blue = F; Red = B2; Pink = G. Presence of metabolic reactions is represented using a black mark, where the colour is visible, this means absence of the reaction. In total, 1,848 reactions (pan) were annotated in the full panel of APEC collections, 1639 (88.7%) of which were shared among all isolates (100%) defined as core, otherwise defined as accessory (11.3%). The functional distribution of accessory reactions was also show here. Colour code of functional categories of reactions (from left to right): Purple = Amino acid metabolism; Pink = Carbohydrate metabolism; Green = Cell wall/ membrane/ envelope metabolism; Orange = Cofactor and prosthetic group metabolism; Yellow = Energy production and conversion; Blue = Inorganic ion transport and metabolism; Beige = Lipid metabolism; Brown = Nucleotide metabolism; Gray = Other; Light blue = Transport/exchange reaction; Red = unassigned (* Unassigned reactions were not involved in reconstruction of APEC GEM or phylogroup-specific Sub-GEMs, as determined as dead-end reactions by gap-filling).

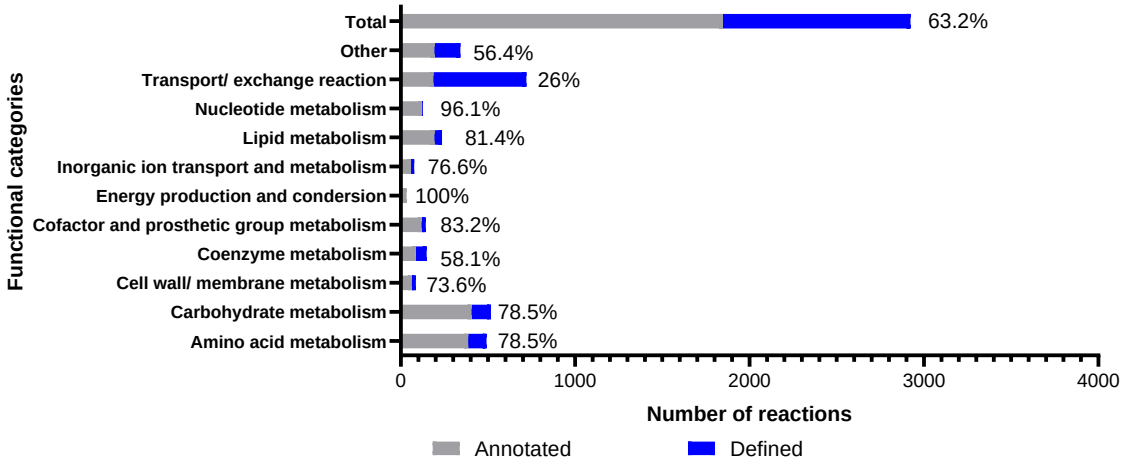

Figure 4. The distribution of the final defined reactions (Grey) in the APEC model after gap-filling and manual curation compared to the annotated metabolic reactions (Blue) screened by ModelSEED in 114 APEC isolates. The percentage shown represents the fraction of reactions retained in the model.

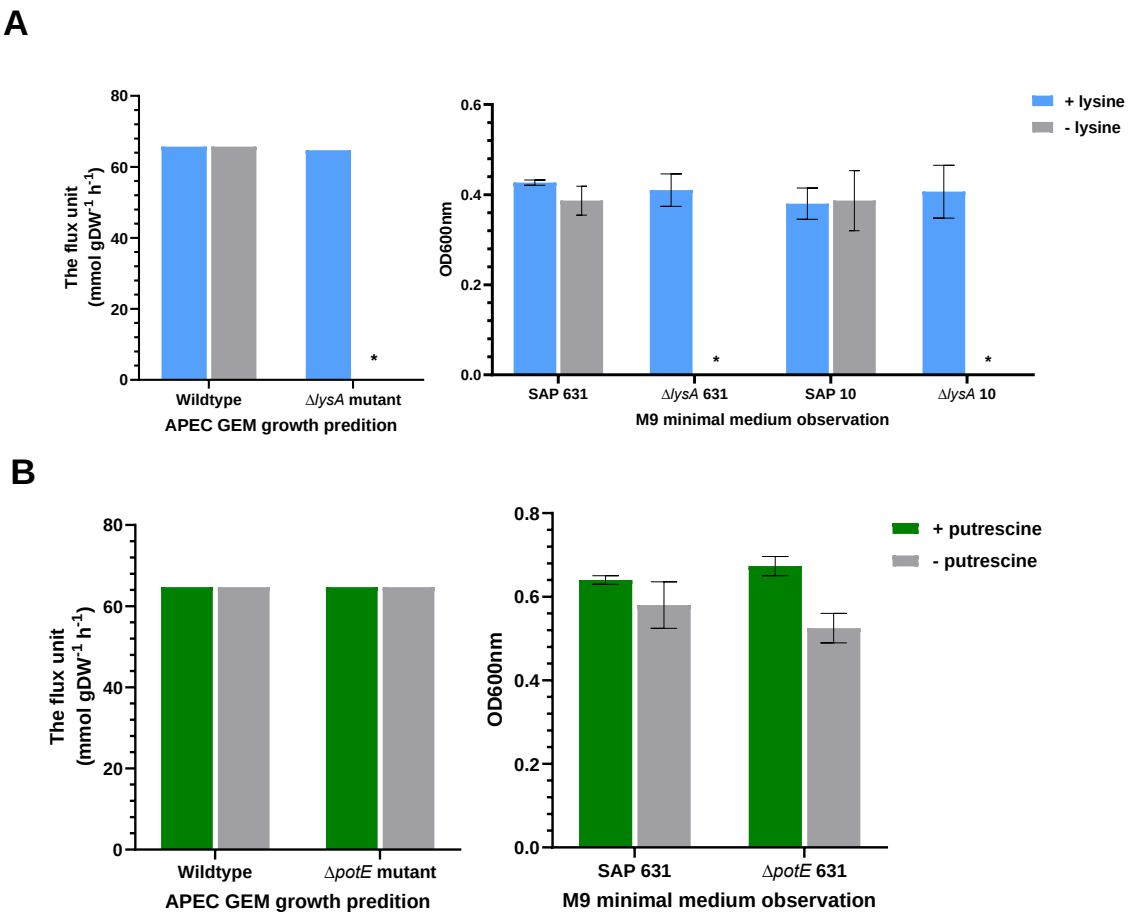

Figure 5. The computational growth prediction (Left) and phenotypic observation (Right) of wildtype and single gene knockout isolates under constrained nutrient input. Figure 5A shows the effect of inactivation of *lysA* with/without medium supplementation with lysine; an asterisk indicates lack of growth demonstrating the essential role of the *lysA* gene. Figure 5B shows the effect of inactivation of *potE* with/without medium supplementation with putrescine, which had no effect as the gene is not essential. The computational prediction is generated by APEC GEM and data is measured by flux unit (mmol gDW^{-1} h^{-1}); The phenotypic observation is measured using M9 minimal medium and the optical density is recorded after 24 hour incubation (37 °C) at a wavelength of 600 nm in 96-well plate. The error bar shows standard deviation (SD) of the mean of 3 biological replicates. Asterisks indicate an absence of growth.

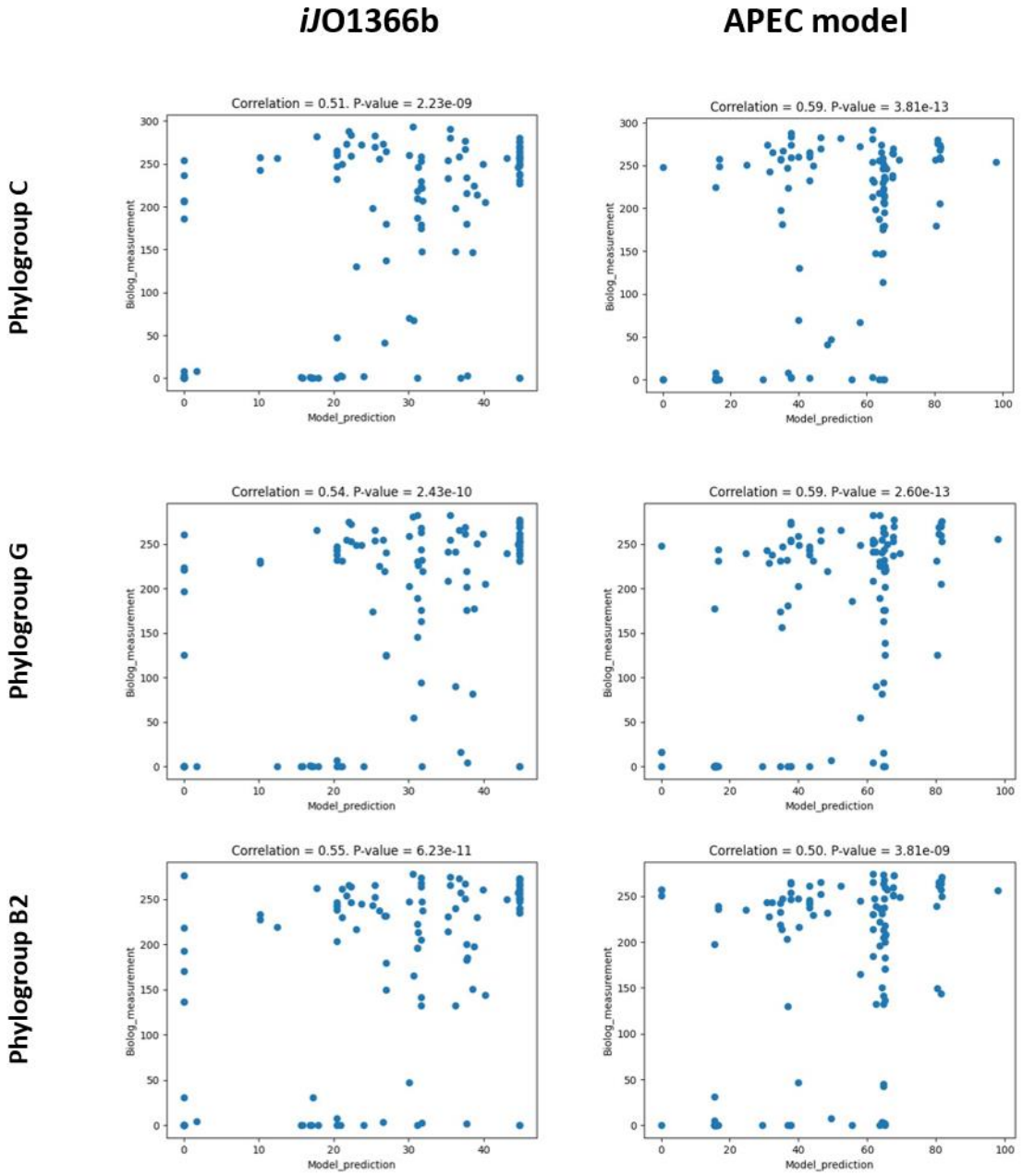

Figure 6. Scatter plots comparing computational predictions from the *E. coli* K-12 *i*JO1366 model and the APEC model for three APEC isolates representing phylogroups B2, C and G, with Biolog phenotypic observations with single nutrient source utilisation. The APEC isolates used are from phylogroup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

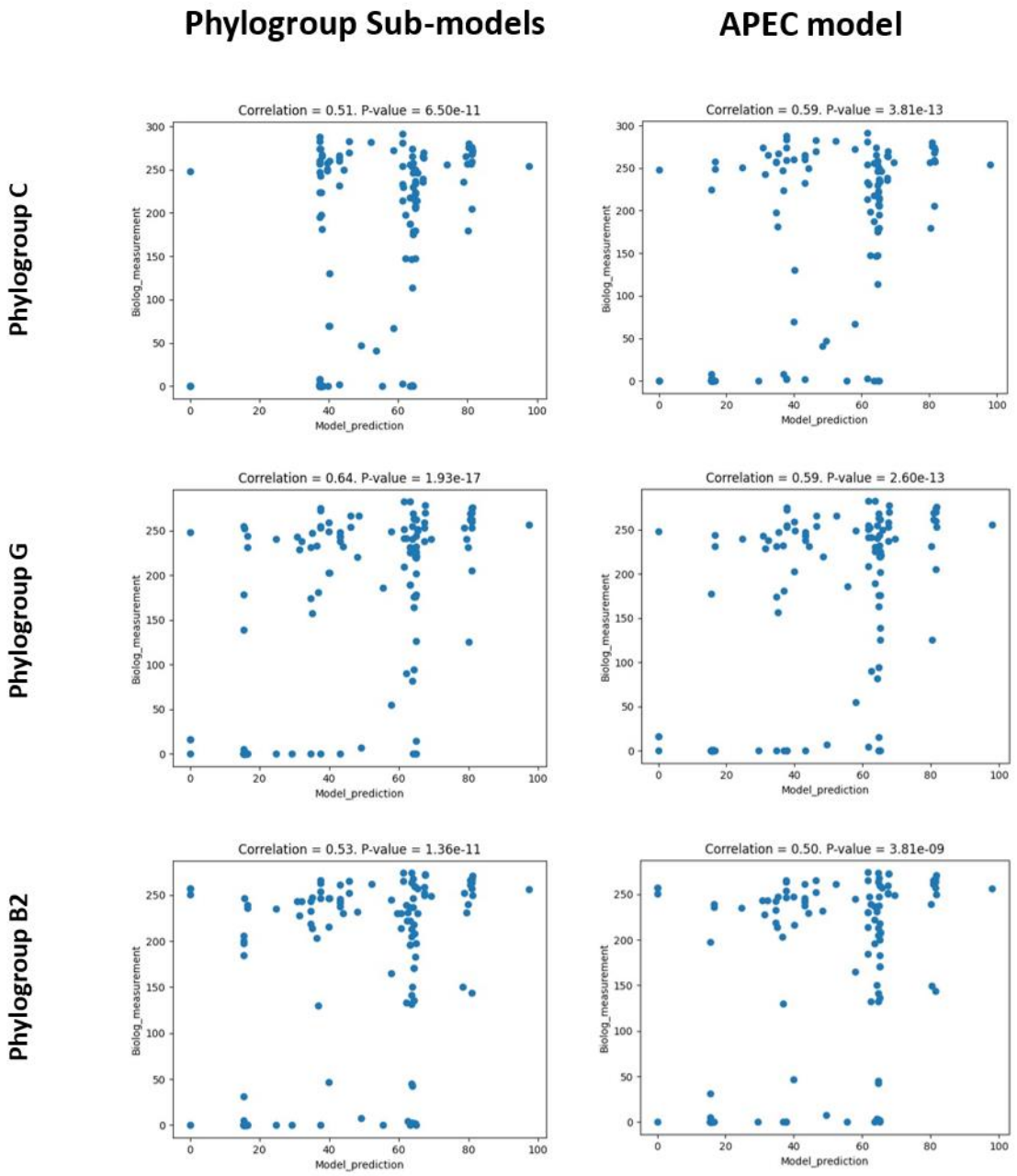

Figure 7. Scatter plots comparing computational predictions from the APEC phylogroup-specific sub-models for phylogroups B2, C and G with the general APEC model for three APEC isolates representing phylogroups B2, C and G, with Biolog phenotypic observations with single nutrient source utilisation. The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

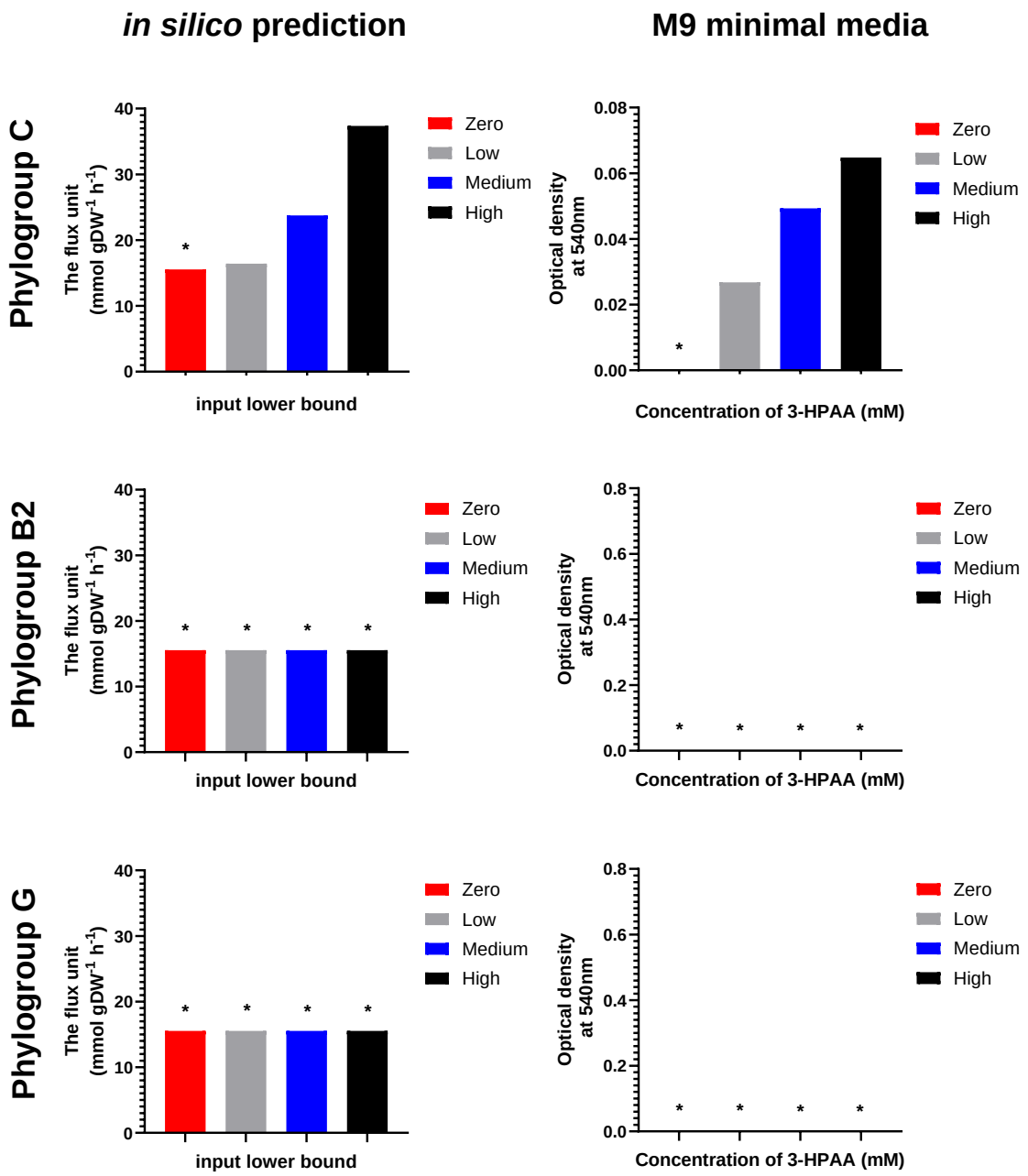

Figure 8. The comparison between *in silico* prediction and M9 minimal medium measurements of different phylogroup APEC isolates on utilisation of different concentration of 3-hydroxyphenoylacetic acid (3-HPAA). The left figures show the *in silico* prediction using phylogroup-based sub-model (C, G and B2) under the constrain of 3-HPAA exchange reaction lower bound adjusted to 0, -10 (low), -100 (medium) and -1000 (high), respectively (-1000 refers to unlimited uptake of inputs). The right figures show the phenotypic measurements in M9 minimal media with supplement of 0 mM (zero), 0.5 mM (low), 0.75 mM (medium), and 1 mM (high) 3-hydroxyphenoylacetic acid as sole carbon source. The measurements indicate the mean of three biological endpoint (t=18h) readings. More detailed growth curves are provided in Supplementary Figures 4 and 5. Asterisks indicate an absence of growth.