Genome reconstruction of a novel carbohydrate digesting bacterium from the chicken caecal microflora

Ankit T. Hinsu*[a], Ramesh J. Pandit*[b], Shriram H. Patel[a], Androniki Psifidi[c,d], Fiona M. Tomley[e], Subrata K. Das[f], Damer P. Blake[e], Chaitanya G. Joshi[b#]

[a]Department of Animal Genetics & Breeding, Anand Agricultural University, Anand, India.

[b]Department of Animal Biotechnology, Anand Agricultural University, Anand, India.

[c]The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, UK.

[d]Department of Clinical Science and Services, Royal Veterinary College, North Mymms, Hertfordshire, UK.

[e]Department of Pathobiology and Population Sciences, Royal Veterinary College, North Mymms, Hertfordshire, UK.

[f]Functional Genomics of Extremophiles, Institute of Life Sciences, Nalco Square, Bhubaneswar, India.

* Equal contribution

Address for correspondence:

Chaitanya G. Joshi,

Professor,

Department of Animal Biotechnology

College of Veterinary Science & Animal Husbandry

Anand Agricultural University,

Anand-388 001, Gujarat, India

Email- cgjoshi@rediffmail.com

Tel: +91 2692 261038

Fax: +91 2692 261486

Abstract

The advent of metagenomics using Next-Generation Sequencing (NGS) has led to the acquisition of unprecedented quantities of bacterial sequence data. One use for these data is the recovery of complete genomes of uncultivable, and thus uncharacterised, organisms. The separation of genome from metagenomics data remains challenging, however genomes of several novel lineages and organisms have been reconstructed successfully. Here, we report the use of NGS to reconstruct the draft genome of a previously undescribed bacterium from the caecal microbiome of domestic chickens. This bacterium has a genome size of 2.38Mb, a 52.13% GC content and 16S rRNA gene phylogenetic analysis indicates that it belongs to a recently recognised family of anaerobes, the *Muribaculaceae* belonging to Bacteroidetes phylum. Bacteroidetes is one of the major phyla in chicken caecum amounting to 12%-61%. The caecum is a major site of intestinal fermentation that harbours a diverse collection of anaerobic bacteria. From direct genome annotation, this newly identified bacterium is predicted to utilise starch as its primary carbon source, as evidenced by the large number of amylases encoded within its genome: we thus propose the name Candidatus *Amulumruptor caecigallinarius*. The importance of caecal carbohydrate metabolism and fermentation to the overall health, welfare and productivity of the chicken predicts an important role for this novel Candidatus bacterium in food security.

Keywords: Microbiome; caecum; chicken; genome reconstruction; Candidatus Amulumruptor caecigallinarius

Acknowledgment

The authors would like to acknowledge support from the Department of Biotechnology (DBT, India) under grant BT/IN/Indo-UK/FADH/49/CGJ/2013, and the Biotechnology and Biological Sciences Research Council (BBSRC, UK) under grant BB/L00478X/1.

Introduction

Poultry farming is one of the important livestock industry with particular relevance in South and East Asia, including India where it growing very rapidly(Chengat Prakashbabu et al. 2017). Although very few different types of chicken are raised by the majority of commercial breeders, countries such as India harbour a rich diversity of indigenous chicken breeds. In addition to a source of income for small-scale backyard farmers, these chickens are perceived to present aesthetic and medicinal values. One such example is the Kadaknath chicken, a breed characterised by fibromelanosis with hyperpigmentation throughout its body that contributes distinct characteristics to its meat and a slate-like color to all its tissues including skin, muscle, bone-marrow, tendon, nerves and blood (Arora et al. 2011). In contrast, Global Commercial Broiler (GCB) lines of hybrid birds, such as the Cobb400, are breeds specifically for meat production with the purpose of high weight gain in reduced periods of time. It has been suggested that compared to GCB birds, Kadaknath chickens are more resistant to extreme environmental conditions such as summer heat and winter cold, can thrive under poor housing and poor feeding conditions, and are more resistant to diseases encountered in the natural environment (Thakur et al. 2006). Even though the Kadaknath grows more slowly than its hybrid commercial counterparts, its meat is in high demand and has been used for medicinal purposes by tribal people (Thakur et al. 2006). As a consequence, Kadaknath chickens fetch higher prices than GCB.

The chicken caecum is a site of fermentation and is densely populated with microbes(Sergeant et al. 2014). In common with microbial populations of ruminant animals, bacteria within the chicken caecal microbiome actively ferment ingested feed and generate additional nutrients for the host (Pan and Yu 2014). Ingested feed particles remain in the caeca for 12-20 hours, allowing time for the breakdown and adsorption of nutrients (Sergeant et al. 2014). Several studies have defined the microbial population composition residing in the caeca of the chickens, and it's potential role in digestion have been explored (Ahir et al. 2010; Lee et al. 2017; Pedroso et al. 2013). In our previous study (Pandit et al. 2018) we have intensively studied the structure of microbial population in caecum from different chicken breeds/lines using Amplicon sequencing including indigenous Kadaknath and globally available GCB. However, with the advances in the informatics has opened a new way to analyze the metagenomics data. Nowadays researchers have shown interest to recover the whole genome of yet uncultured bacteria using the shotgun metagenome sequencing data. In line with this, several novel bacterial genomes have been reconstructed from the metagenome shotgun data from chicken ceacum (Sergeant et al. 2014). Therefore, in this study, we have looked in to direction and recovered a draft bacterial genome belonging to a novel genus. The binning-based reconstruction method require use of differentially abundant samples and hence, we have selected two chicken lines both reared together.

Results

Sequencing and binning

Metagenomic sequencing of caecal contents from five individual Cobb400 and five Kadaknath chickens yielded total data of 4.1Gb comprising 7.507 million paired-end reads (Table S1). De novo assembly using the MIRA assembler v4.0 yielded a total of 60,629 contigs with length greater than 500 bp, comprising 109.5 Mb data, which was used for further analysis. Data from Cobb400 and Kadaknath chickens were pooled into separate type-specific datasets and mapped back to these contigs in CLC Genomics Workbench v7.0 and the average coverage of each contig was calculated.

The methodology adopted by Albertsen et al. (2013) has been called differential-coverage based binning. As the name implies, it requires two metagenomes, both having bacterial populations with varying abundances. Binning is carried out based on differences in coverage and other information such as GC content, taxonomy of essential genes and nucleotide frequency. Initial statistics suggested 3685 total essential genes and 109 unique essential genes (Figure 1A). A cluster assigned to the Bacteroidetes phylum was selected comprising of 119 total essential genes and 101 unique essential genes with a size of 3.46 Mb (Figure 1B). This cluster was subjected to redundancy analysis using trinucleotide, tetranucleotide and pentanucleotide frequencies; and a plot of PC2vs PC3 was used to remove further contaminants (Figure 1C). Statistics after separation from the PCA plot showed the presence of 103 total essential genes and 101 unique essential genes with a size of 2.49 Mb (Figure 1D). It was considered to be the final bin. Further, paired-end tracking and manual curation was done to refine the bin as recommended by the authors (Albertsen et al. 2013).

Reads constituting binned contigs were extracted and reassembled yielding 157 contigs spanning 2.377Mb with an N50 of 41,191, a longest contig of 107,566bp and 52.13% GC content. This assembly was considered as the final assembly of the draft genome. The completeness of the reconstructed genome was verified using the CheckM tool wherein a lineage-specific workflow compared it to the order *Bacteroidales* with 90.44% completeness, 2.45% contamination and 22.22 strain heterogeneity. Performing a taxonomic-specific workflow for the family *Porphyromonadaceae* (16S rRNA gene-based taxonomy assignment) indicated 86.24% completeness, 2.10% contamination and 9.30 strain heterogeneity.

Taxonomic identification

The identity of bacteria was determined by BLAST against the NCBI 16S rRNA database which gave a positive hit in contig127. It was matched with the 16SrRNA gene of *Muribaculum intestinale* strain YL27 (Lagkouvardos et al. 2016) (recently discovered within the mouse intestine) with 99% query coverage and 89% identity, followed by *Barnesiella viscericola* (coverage:99%, identity:87%) and *Barnesiella intestinihominis* (coverage:99%, identity:87%) (Figure 2 & Figure S1A). To check for other related organisms from uncultured/environmental samples BLAST was carried out against the same database including environmental sequences. It matched most closely with the 16SrRNA sequence of an Uncultured bacterium clone DTB_G40, recovered previously from the caeca of a turkey(Scupham et al. 2008) with 100% query coverage and 99% identity, followed by other bacteria clones with identity 96% (feces of a mouse), 95% (gut of a mouse) and 92% (caeca and colon of mouse)(Stecher et al. 2007) (Figure S1B). RDP Naive Bayesian rRNA Classifier v2.11 identified the bacteria upto family level as *Porphyromonadaceae* with 99% confidence. Additionally, Pintail based checking revealed absence of chimera in 16s sequence (Figure S2).

Further, to resolve the taxonomy, Average Nucleotide Identity (ANI) and *in silico* DNA-DNA hybridisation (DDH) was performed. In contrast to the 16S rRNA gene sequence based assignment, ANI results showed 76.11% and 79.98% similarity between the genomes of *M. intestinale* and *B. viscericola*, respectively, with the genome described here (Figure 3). DDH also revealed higher genome-to-genome similarity with *M. intestinale* than *B. viscericola* (Figure S3). Additionally, CV-Tree based prediction revealed organisation of *M. intestinale* and reconstructed genome in the same clade while *Barnesiella* genomes were present as a sister clade (Figure S4). All these findings point towards M. intestinale as the closest neighbour and reconstructed genome belonging to the same family.

Genome properties

The RAST server predicted the presence of 1,989 coding sequences (CDS) which were classified into 52 RNAs and 231 subsystems. Out of 1,989 CDS, 732 (37%) were represented in subsystems. The greatest number of features were annotated within Protein metabolism (161), followed by Amino acid and derivatives (139) (Figure 4 & Figure S5).

Mapping reads from the Cobb400 and Kadaknath chicken groups covered ~99.2% and 98.7% of the reconstructed genome with total sizes of 2.359Mb and 2.346Mb, respectively. Comparing 16S rRNA gene sequences from the Cobb400 and Kadaknath consensuses with the reconstructed genome showed 96.4% and 93.1% identity, respectively.ANI between the reconstructed genome and the Cobb400 consensus genome was 99.9%, while the figure for the Kadaknath consensus genome was 99.8% (Figure S6). To check for variations in the 16S rRNA gene sequences, we repeated the mapping with different parameters (Table S2), revealing that the stringency of mapping parameters affected the 16S rRNA gene sequence more compared to other portions of genome, as evident from the ANI scores. Annotation using the subsystems database revealed very little variation between the Cobb400 and Kadaknath groups in terms of number of subsystems and proportion of annotated categories (Figure S7).

Comparing the reconstructed and consensus genomes based on their sequence revealed that from 1,990 predicted proteins in the reconstructed genome, the Cobb400 consensus retained 1,934 with >90% identity and 29were absent. The Kadaknath consensus genome included 1,911 with >90% identity and 58 were absent.

All of the protein encoding genes (PEGs) predicted by RAST were subjected to CAZy prediction using dbCAN. Of the PEGs present in the reconstructed genome, 135 CAZy domains were predicted in 115 PEGs, consisting mainly of Glycosyl transferases (GT) (56) and Glycoside hydrolases (GH) (41). Consensus genomes from the Cobb400 and Kadaknath groups showed the presence of 131 and 133 CAZy domains in 110 and 112 PEGs, respectively (Figure S8). In all three genomes within the GH families GH13 was the most abundant, followed by GH3, GH23 and GH32 (Table S3).

GH13 primarily codes for Amylase, which converts dietary starch into maltose. The maltose formed can be converted to glucose via the action of Maltase (EC 3.2.1.20), and is subsequently converted to pyruvate via glycolysis (Figure5A). The pyruvate formed under anaerobic conditions is likely to be converted to succinate via the tricarboxylic acid (TCA) cycle as shown in Figure5B. All of the genes forming the *Batl* (Bacteroides aerotolerance) operon, part of the "Aerotolerance operon in *Bacteroides* and potentially orthologous operons in other organisms" subsystem**,** along with Superoxide dismutase were also annotated (Figure S9).

Virulence associated genes

Genes belonging to the "Virulence, disease and defence" category were investigated further to explore possible differences between the Cobb400 and Kadaknath consensus genomes. All 26 genes annotated under this category were compared between the three genome assemblies. One PEG coding for *Quinolinate synthetase* (EC 2.5.1.72) was abbreviated in the Kadaknath-derived consensus genome, missing the first 99 bases.

All other genes from "Virulence, disease and defence" category were checked for variants using VDAP-GUI, using the PEG sequences from the reconstructed genome as reference. Reads from each of the consensus genomes were mapped onto the reference, and putative polymorphisms detected in more than 60% of reads were

considered to be variants. Three PEGs were found to contain polymorphisms; specifically the RNA Polymerase beta (β) and beta prime (β') subunits and the Translation elongation factor Tu. Two variants induced non-synonymous substitutions, one in each of the β and β' of RNA Polymerase subunit genes. The mutation in the β subunit gene led to a E1169D change and the substitution of a negatively charged amino acid for another negatively charged amino acid. The mutation in the β' subunit gene led to a R455K change and the substitution of a positively charged amino acid for another positively charged amino acid. The Duet server was used to predict any destabilizing effect of the non-synonymous substitutions in the protein structures predicted by I-TASSER. On superimposing structures using PyMol, the RMSD value was 0 for both proteins. However, several changes were apparent in the structures of the mutant and native forms, most notably shortening in the β-sheets in the β subunit structure and shortening of the β-sheets and A-helix in the β' subunit (Figure S10).

Discussion

At the time of writing 101,859 bacterial genomes were available through GenBank, of which 2,519 have been designated as Candidatus organisms, highlighting the potential of culture-independent recovery of complete microbial genomes. Researchers have assembled complete or draft bacterial genomes from various environments including acid mine drainage (Kadnikov et al. 2016), bioreactors (Albertsen et al. 2013), hydrothermal plumes (Anantharaman et al. 2016; Li et al. 2016), estuary sediments (Baker et al. 2015), oceans (Mehrshad et al. 2016), biogas plants (Stolze et al. 2016), birds (Sergeant et al. 2014) and the human gut (Brown et al. 2013; Gupta et al. 2016; Jeraldo et al. 2016). Genome reconstruction from metagenomic data is an emerging technique and several approaches have been described (Albertsen et al. 2013; Alneberg et al. 2014; Broeksema et al. 2017; Eren et al. 2015; Kang et al. 2015; Lin and Liao 2016; Skennerton et al. 2015; Wu et al. 2016). Most rely on conservation of features such as GC% and nucleotide frequency between contigs from a specific genome, permitting them to be binned together. Other components include variation in tetranucleotide frequency, providing a balance between speed of computation and sufficient representation of information. In this study, we have used trinucleotide, tetranucleotide and pentanucleotide frequency, supplementing the approach described in the original study (Albertsen et al. 2013). Increasing from the frequency of one to three such markers is expected to have increased specificity.

Microbes in the caeca are primarily involved in degradation and fermentation of feed. As a result, they produce a multitude of glycoside hydrolase enzymes, illustrated by the presence of 41 different GH proteins in the reconstructed genome described here. The higher abundance of GH13 family enzymes may be correlated with the feed received by chickens in the area. The feed provided to chickens in India commonly contains a high proportion of maize grains which are rich in starch, requiring amylase for effective digestion. The reconstructed genome has been found to encode multiple alpha-amylase enzymes (GH13 family), indicating a role in degradation of starch from maize. Further, maltose is the major product released after starch degradation and the reconstructed genome contained genes that are involved in converting maltose into glucose. Based on these observations, it can be predicted that the bacterium feeds on starch and uses maltose as one of its carbon sources. Based on the genomic analysis, it can be predicted that the bacterium is anaerobic in nature, supported by the presence of genes whose products are capable of converting glucose to succinate, a major intermediate in anaerobic organisms formed through the reductive branch of TCA cycle (Cheng et al. 2013). Other notable features include the presence of genes representing the Batl (Bacteroides aerotolerance) operon (Tang et al.

1999). All five core genes from the operon (*BatlA*, *BatlB*, *BatlC*, *BatlD* and *BatlE*) were present and appear likely to be functional in the reconstructed genome. The presence of the Batl operon along with Superoxide dismutase suggests that this bacterium is an obligatory anaerobic in nature.

Current conventions of taxonomic classification have recently started to shift from phenotypic to molecular approaches. Presently, widely accepted norms state that an bacteria with 16S rRNA gene identity <97% can be considered as a novel species, while 94.5% and 86.5% are respective cut-offs for novel genus and family classification (Lagkouvardos et al. 2016). Our reconstructed genome showed 89% identity with *Muribaculum intestinale* and hence, can be considered as a novel genus within the same family (i.e. *Muribaculaceae*). Additionally, classification based on all proteins of genome with other bacterial proteins also classified reconstructed genome clustered with Muribaculum and nearest to the cluster of Barnesiella members (Figure S4 and Figure 2). Based on the amylase-degrading features of this bacterium, we propose the novel genus for this bacterium to be Candidatus *Amulumruptor* and a novel species under it to be named Candidatus *Amulumruptor caecigallinarius*. Also, we suggest addition of the uncultured bacterium clone DTB_G40 (Scupham et al. 2008) and other related sequences (Stecher et al. 2007) under the same phylogeny due to their close similarity with our bacterium.

As for other animals, the chicken gastrointestinal tract is widely considered to be sterile at the time of birth (hatch)(Funkhouser and Bordenstein 2013). However, microbes commence colonisation immediately after hatching and many shows commensal activity. Here, chickens of two different genetic backgrounds were reared together, suggesting a shared environment and the likelihood of comparable enteric microbial populations. However, we found non-synonymous variations in a small number of genes from the "Virulence, disease and defence" category between the consensus reconstructed genomes recovered from Cobb400 and Kadaknath chickens. Such variation may be coincidental, reflecting the timing of bacterial exposure by different strains, or it might indicate a host genetic component in colonisation by distinct strains of the novel organism. Host genotype has previously been shown to influence the shape of enteric microbial communities not only in humans (Davenport 2016) but also in chickens (Psifidi et al. 2018; Zhao et al. 2013). Non-synonymous mutations among virulence genes were observed in the RNA polymerase β and β' subunit coding sequences. It has been reported that mutations in RNA polymerase confer resistant to some antibiotics (Kristich and Little 2012; Lee et al. 2013; Perez-Varela et al. 2017; Yuzenkova et al. 2002).

Successful genome reconstruction for a novel organism from a small number of complex metagenomic datasets indicates a relatively high level of occurrence within the caecal microbiome. The identification of an organism which is predicted to play an important role in caecal fermentation is directly relevant to feed utilisation, productivity and health, influencing economic performance, chicken welfare and contributing to food security. Reconstruction of genomes from metagenome faces several challenges; absence of culturable representative being one of those. Because of such limitations, the information about this organism is based on prediction. However, this could be overcome by directing the further research towards isolation of such candidate organisms. While the occurrence and relevance of *A. caecigallinarius* is yet to be defined in chickens raised in different regions, under different production systems, it is expected to be just one of many as yet unknown but influential bacteria.

Description of "Candidatus Amulumruptor caecigallinarius."

Based on the similarity described, it can be inferred that bacterium belongs to a novel genus within the *Muribaculaceae* family. We propose the name Candidatus *Amulumruptor caecigallinarius* (L. amulum: starch; ruptor: breaker; amulumruptor: starch-breaker; caeca: caecum; gallinarius: of a poultry; caecigallinarius: from the caecum of poultry) for this genome. The name is assigned based on the predicted ability of the organism to digest starch. It is represented by a composite genome (GenBank acc. no. PXWE01000000) obtained from the metagenome of caecum of chicken.

Material and methods

Sample and sequencing

Cobb400 (a hybrid commercial broiler line) and indigenous Kadaknath chickens were reared together at the Central Poultry Research Station (CPRS), AAU, Anand. All birds were fed a local standard maize and soybean based diet. At the age of 35 days, 5 birds of each chicken type were euthanized humanly via cervical dislocation and the caecal pouches were removed and opened using sterile scissors. Caecal contents were recovered into 2 mL cryo-vials containing 1 mL RNAprotect Bacterial Reagent (Qiagen, Germany), transported to the lab at -20^0C and thereafter stored at -80^{o}C. DNA was extracted using a Qiagen DNA Stool Mini Kit (Qiagen, Germany) following the manufacturer's protocol and quantified using a Qubit fluorometer V3.0. The DNA was then subjected to library preparation using a Nextera XT library preparation kit (Illumina Inc., CA, USA) and sequenced on an Illumina MiSeq desktop system (Illumina Inc., CA, USA) using 2x300bp chemistry.

Assembly

Sequencing data from all 10 birds were pooled and assembled using MIRA assembler v4.0 (Chevreux et al. 1999) with default parameters. Contigs greater than 500 bp in length were taken for downstream analysis. Further, pooled data from each of the Cobb400 and Kadaknath types were mapped back to the contigs using CLC Genomics Workbench v7.0 (Qiagen, Germany).

Binning

The differential-coverage based binning approach described by Albertsen et al. (Albertsen et al. 2013) was followed. In brief, information such as GC content and nucleotide frequency (trimer, tetramer and pentamer) for each contig was calculated using a perl script provided by the authors and found in the package multi-metagenome (http://github.com/MadsAlbertsen/multi-metagenome). Further, from each contig, ORFs were predicted using MetaProdigal v2.6.3 (Hyatt et al. 2012) and used to predict 110 single copy essential genes using HMMer v3.1b2 (http://hmmer.org/). Predicted genes were matched against the Refseq-protein database from NCBI using local BLAST+ v2.2.3.1 (Camacho et al. 2009) and output was used to assign taxonomy to contigs using MEGAN v5.11.3 (Huson et al. 2011). All of the information generated was subsequently transferred to R v3.3.1 and binned first based on coverage, followed by Principal Component Analysis (PCA) plot based on nucleotide frequencies. Finally, the bin was refined by paired-end tracking using Cytoscape software (Shannon et al. 2003) as described in the manual. Contigs were manually checked by matching against the NCBI nt database using BLASTn and contigs with score>200, alignment length>100 and belonging to other

phyla were removed from the bin. Reads making binned contigs were then extracted and reassembled using the GS De Novo Assembler v2.6 (Roche Diagnostics, IN, USA) (length overlap: 50% and similarity overlap: 90%).

Annotation

The reconstructed genome was matched against the NCBI 16S rRNA database using BLASTn to assign taxonomy, in addition to classification using the RDP Naive Bayesian rRNA Classifier v2.11. Further, 16S sequences from type organisms from the assigned order were obtained from RDP. All these sequences including the 16S rRNA gene sequence from the reconstructed genome and closely related genomes (as retrieved from RDP) were aligned using MAFFT v6.864 (Katoh et al. 2002), available online at GenomeNet, Kyoto University Bioinformatics Center, with the FFT-NS-i strategy. A maximum-likelihood tree was inferred using the GTR+R4 model recommended using ModelFinder (Kalyaanamoorthy et al. 2017) in IQ-Tree (Nguyen et al. 2015). Branch supports were obtained with 1000 ultrafast bootstraps (Minh et al. 2013) implemented in IQ-Tree software (Nguyen et al. 2015). Presence of chimera in 16s rRNA gene was checked using Pintail (Ashelford et al. 2005) against three nearest neighbour cultured neighbour. Genome completeness was evaluated using the CheckM tool v1.0.7 (Parks et al. 2015) and annotated on the RAST server (Aziz et al. 2008). Complete proteome was used to infer phylogeny using CV-Tree v3.0 (Qi et al. 2004; Zuo and Hao 2015). Carbohydrate Active Enzymes (CAZy) domains were predicted from the protein encoding genes (PEGs) predicted by the RAST server using an Hidden Markov Model (HMM) based approach through dbCAN (web server and DataBase for automated Carbohydrate-active enzyme Annotation) (Yin et al. 2012) using dbCAN HMMs 5.0. ANI was calculated using server at Kostas lab (http://enve-omics.ce.gatech.edu/ani/index) while *in silico* DNA-DNA hybridization was performed using tool hosted by King Abdullah University of Science and Technology.

Breed wise variability

Reads derived from both types of chicken were individually mapped onto the reconstructed genome using GS Reference Mapper v2.6 (Roche Diagnostics, IN, USA) (length overlap 60% and similarity of 90%) to obtain types specific consensus genomes, which were also annotated using the RAST server. Genes involved in the "Virulence, Disease and Defence" category according to annotation with the subsystems database were studied further to look for variants between host types. The VDAP-GUI (Menon et al. 2016) tool was used for variant calling in all genes using the multicom approach within the tool. Variants present in more than 60 % of reads were considered as positive hits. Mutants were further studied using the I-TASSER (Roy et al. 2010) server for structure prediction, Duet server (Pires et al. 2014) for inducing mutations, and PyMol (DeLano 2002) for visualisation and comparison of structures.

Data Submission

Sequences representing the reconstructed genome is submitted at NCBI (Accession:**PXWE01000000**) under Bioproject ID **PRJNA435487**.

Compliance with Ethical Standards

Funding: This study was funded by the Department of Biotechnology (DBT, India) under grant BT/IN/Indo-UK/FADH/49/CGJ/2013, and the Biotechnology and Biological Sciences Research Council (BBSRC, UK) under grant BB/L00478X/1.

Conflict of Interest: The authors declare that they have no conflict of interest.

Ethical Approval: This study was carried out using welfare standards consistent with those established under the Animals (Scientific Procedures) Act 1986, an Act of Parliament of the United Kingdom. All protocols were approved by the Ethical Review Panel of Anand Agricultural University (AAU) and the Clinical Research Ethical Review Board (CRERB) of the Royal Veterinary College.

References

Ahir, V., Koringa, P., Bhatt, V., Ramani, U., Tripathi, A., Singh, K., Dhagat, U.M., Patel, J.S., Patel, M., Katudia, K., Sajnani, M., Jakhesara, S., Joshi, C., 2010. Metagenomic analysis of poultry gut microbes. Indian J Poult Sci. 45, 111-114.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., Nielsen, P.H., 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. Nat Biotechnol. 31(6), 533-8. https://doi.org/10.1038/nbt.2579.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Lahti, L., Loman, N.J., Andersson, A.F., Quince, C., 2014. Binning metagenomic contigs by coverage and composition. Nat Methods. 11(11), 1144-6. https://doi.org/10.1038/nmeth.3103.

Anantharaman, K., Breier, J.A., Dick, G.J., 2016. Metagenomic resolution of microbial functions in deep-sea hydrothermal plumes across the Eastern Lau Spreading Center. ISME J. 10(1), 225-39. https://doi.org/10.1038/ismej.2015.81.

Arora, G., Mishra, S.K., Nautiyal, B., Pratap, S.O., Gupta, A., Beura, C.K., Singh, D.P., 2011. Genetics of hyperpigmentation associated with the Fibromelanosis gene (Fm) and analysis of growth and meat quality traits in crosses of native Indian Kadaknath chickens and non-indigenous breeds. Br Poult Sci. 52(6), 675-85. https://doi.org/10.1080/00071668.2011.635637.

Ashelford, K.E., Chuzhanova, N.A., Fry, J.C., Jones, A.J., Weightman, A.J., 2005. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol. 71(12), 7724-36. https://doi.org/10.1128/AEM.71.12.7724-7736.2005.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 9, 75. https://doi.org/10.1186/1471-2164-9-75.

Baker, B.J., Lazar, C.S., Teske, A.P., Dick, G.J., 2015. Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria. Microbiome. 3, 14. https://doi.org/10.1186/s40168-015-0077-6.

Broeksema, B., Calusinska, M., McGee, F., Winter, K., Bongiovanni, F., Goux, X., Wilmes, P., Delfosse, P., Ghoniem, M., 2017. ICoVeR - an interactive visualization tool for verification and refinement of metagenomic bins. BMC Bioinformatics. 18(1), 233. https://doi.org/10.1186/s12859-017-1653-5.

Brown, C.T., Sharon, I., Thomas, B.C., Castelle, C.J., Morowitz, M.J., Banfield, J.F., 2013. Genome resolved analysis of a premature infant gut microbial community reveals a Varibaculum cambriense genome and a shift towards fermentation-based metabolism during the third week of life. Microbiome. 1(1), 30. https://doi.org/10.1186/2049-2618-1-30.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10, 421. https://doi.org/10.1186/1471-2105-10-421.

Cheng, K.K., Wang, G.Y., Zeng, J., Zhang, J.A., 2013. Improved succinate production by metabolic engineering. Biomed Res Int. 2013, 538790. https://doi.org/10.1155/2013/538790.

Chengat Prakashbabu, B., Thenmozhi, V., Limon, G., Kundu, K., Kumar, S., Garg, R., Clark, E.L., Srinivasa Rao, A.S., Raj, D.G., Raman, M., Banerjee, P.S., Tomley, F.M., Guitian, J., Blake, D.P., 2017. Eimeria species occurrence varies between geographic regions and poultry production systems and may influence parasite genetic diversity. Vet Parasitol. 233, 62-72. https://doi.org/10.1016/j.vetpar.2016.12.003.

Chevreux, B., Wetter, T., Suhai, S. Genome sequence assembly using trace signals and additional sequence information. In: German conference on bioinformatics, 1999. vol 99. p 45-56

Davenport, E.R., 2016. Elucidating the role of the host genome in shaping microbiome composition. Gut Microbes. 7(2), 178-84. https://doi.org/10.1080/19490976.2016.1155022.

DeLano, W.L., 2002. Pymol: An open-source molecular graphics tool. CCP4 Newsletter On Protein Crystallography. 40, 82-92.
Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., Delmont, T.O., 2015. Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ. 3, e1319. https://doi.org/10.7717/peerj.1319.
Funkhouser, L.J., Bordenstein, S.R., 2013. Mom knows best: the universality of maternal microbial transmission. PLoS Biol. 11(8), e1001631. https://doi.org/10.1371/journal.pbio.1001631.
Gupta, A., Kumar, S., Prasoodanan, V.P., Harish, K., Sharma, A.K., Sharma, V.K., 2016. Reconstruction of Bacterial and Viral Genomes from Multiple Metagenomes. Front Microbiol. 7, 469. https://doi.org/10.3389/fmicb.2016.00469.
Huson, D.H., Mitra, S., Ruscheweyh, H.J., Weber, N., Schuster, S.C., 2011. Integrative analysis of environmental sequences using MEGAN4. Genome Res. 21(9), 1552-60. https://doi.org/10.1101/gr.120618.111.
Hyatt, D., LoCascio, P.F., Hauser, L.J., Uberbacher, E.C., 2012. Gene and translation initiation site prediction in metagenomic sequences. Bioinformatics. 28(17), 2223-30. https://doi.org/10.1093/bioinformatics/bts429.
Jeraldo, P., Hernandez, A., Nielsen, H.B., Chen, X., White, B.A., Goldenfeld, N., Nelson, H., Alhquist, D., Boardman, L., Chia, N., 2016. Capturing One of the Human Gut Microbiome's Most Wanted: Reconstructing the Genome of a Novel Butyrate-Producing, Clostridial Scavenger from Metagenomic Sequence Data. Front Microbiol. 7, 783. https://doi.org/10.3389/fmicb.2016.00783.
Kadnikov, V.V., Ivasenko, D.A., Beletskii, A.V., Mardanov, A.V., Danilova, E.V., Pimenov, N.V., Karnachuk, O.V., Ravin, N.V., 2016. A novel uncultured bacterium of the family Gallionellaceae: Description and genome reconstruction based on metagenomic analysis of microbial community in acid mine drainage. Microbiology. 85(4), 449-461. https://doi.org/10.1134/s002626171604010x.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14(6), 587-589. https://doi.org/10.1038/nmeth.4285.
Kang, D.D., Froula, J., Egan, R., Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 3, e1165. https://doi.org/10.7717/peerj.1165.
Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30(14), 3059-66.
Kristich, C.J., Little, J.L., 2012. Mutations in the beta subunit of RNA polymerase alter intrinsic cephalosporin resistance in Enterococci. Antimicrob Agents Chemother. 56(4), 2022-7. https://doi.org/10.1128/AAC.06077-11.
Lagkouvardos, I., Pukall, R., Abt, B., Foesel, B.U., Meier-Kolthoff, J.P., Kumar, N., Bresciani, A., Martinez, I., Just, S., Ziegler, C., Brugiroux, S., Garzetti, D., Wenning, M., Bui, T.P., Wang, J., Hugenholtz, F., Plugge, C.M., Peterson, D.A., Hornef, M.W., Baines, J.F., Smidt, H., Walter, J., Kristiansen, K., Nielsen, H.B., Haller, D., Overmann, J., Stecher, B., Clavel, T., 2016. The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota. Nat Microbiol. 1(10), 16131. https://doi.org/10.1038/nmicrobiol.2016.131.
Lee, K.C., Kil, D.Y., Sul, W.J., 2017. Cecal microbiome divergence of broiler chickens by sex and body weight. J Microbiol. 55(12), 939-945. https://doi.org/10.1007/s12275-017-7202-0.
Lee, Y.H., Nam, K.H., Helmann, J.D., 2013. A mutation of the RNA polymerase beta' subunit (rpoC) confers cephalosporin resistance in Bacillus subtilis. Antimicrob Agents Chemother. 57(1), 56-65. https://doi.org/10.1128/AAC.01449-12.

Li, M., Jain, S., Dick, G.J., 2016. Genomic and Transcriptomic Resolution of Organic Matter Utilization Among Deep-Sea Bacteria in Guaymas Basin Hydrothermal Plumes. Front Microbiol. 7, 1125. https://doi.org/10.3389/fmicb.2016.01125.

Lin, H.H., Liao, Y.C., 2016. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. Sci Rep. 6, 24175. https://doi.org/10.1038/srep24175.

Mehrshad, M., Amoozegar, M.A., Ghai, R., Shahzadeh Fazeli, S.A., Rodriguez-Valera, F., 2016. Genome Reconstruction from Metagenomic Data Sets Reveals Novel Microbes in the Brackish Waters of the Caspian Sea. Appl Environ Microbiol. 82(5), 1599-612. https://doi.org/10.1128/AEM.03381-15.

Menon, R., Patel, N.V., Mohapatra, A., Joshi, C.G., 2016. VDAP-GUI: a user-friendly pipeline for variant discovery and annotation of raw next-generation sequencing data. 3 Biotech. 6(1), 68. https://doi.org/10.1007/s13205-016-0382-1.

Minh, B.Q., Nguyen, M.A., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol. 30(5), 1188-95. https://doi.org/10.1093/molbev/mst024.

Nguyen, L.T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1), 268-74. https://doi.org/10.1093/molbev/msu300.

Pan, D., Yu, Z., 2014. Intestinal microbiome of poultry and its interaction with host and diet. Gut Microbes. 5(1), 108-19. https://doi.org/10.4161/gmic.26945.

Pandit, R.J., Hinsu, A.T., Patel, N.V., Koringa, P.G., Jakhesara, S.J., Thakkar, J.R., Shah, T.M., Limon, G., Psifidi, A., Guitian, J., Hume, D.A., Tomley, F.M., Rank, D.N., Raman, M., Tirumurugaan, K.G., Blake, D.P., Joshi, C.G., 2018. Microbial diversity and community composition of caecal microbiota in commercial and indigenous Indian chickens determined using 16s rDNA amplicon sequencing. Microbiome. 6(1), 115. https://doi.org/10.1186/s40168-018-0501-9.

Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25(7), 1043-55. https://doi.org/10.1101/gr.186072.114.

Pedroso, A.A., Hurley-Bacon, A.L., Zedek, A.S., Kwan, T.W., Jordan, A.P., Avellaneda, G., Hofacre, C.L., Oakley, B.B., Collett, S.R., Maurer, J.J., Lee, M.D., 2013. Can probiotics improve the environmental microbiome and resistome of commercial poultry production? Int J Environ Res Public Health. 10(10), 4534-59. https://doi.org/10.3390/ijerph10104534.

Perez-Varela, M., Corral, J., Vallejo, J.A., Rumbo-Feal, S., Bou, G., Aranda, J., Barbe, J., 2017. Mutations in the beta-Subunit of the RNA Polymerase Impair the Surface-Associated Motility and Virulence of Acinetobacter baumannii. Infect Immun. 85(8), https://doi.org/10.1128/IAI.00327-17.

Pires, D.E., Ascher, D.B., Blundell, T.L., 2014. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res. 42(Web Server issue), W314-9. https://doi.org/10.1093/nar/gku411.

Psifidi, A., Crotta, M., Pandit, R.J., Fosso, B., Koringa, P.G., Limon, G., Boulton, K., Banos, G., Guitian, J., Tomley, F.M., Rank, D.N., Joshi, C.G., Hume, D., Blake, D.P. Identification of SNP markers affecting gut microbiome composition in chicken. In: Proceedings of the 11th World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand, 2018.

Qi, J., Wang, B., Hao, B.I., 2004. Whole proteome prokaryote phylogeny without sequence alignment: a K-string composition approach. J Mol Evol. 58(1), 1-11. https://doi.org/10.1007/s00239-003-2493-7.

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 5(4), 725-38. https://doi.org/10.1038/nprot.2010.5.

Scupham, A.J., Patton, T.G., Bent, E., Bayles, D.O., 2008. Comparison of the cecal microbiota of domestic and wild turkeys. Microb Ecol. 56(2), 322-31. https://doi.org/10.1007/s00248-007-9349-4.

Sergeant, M.J., Constantinidou, C., Cogan, T.A., Bedford, M.R., Penn, C.W., Pallen, M.J., 2014. Extensive microbial and functional diversity within the chicken cecal microbiome. PLoS One. 9(3), e91941. https://doi.org/10.1371/journal.pone.0091941.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13(11), 2498-504. https://doi.org/10.1101/gr.1239303.

Skennerton, C.T., Ward, L.M., Michel, A., Metcalfe, K., Valiente, C., Mullin, S., Chan, K.Y., Gradinaru, V., Orphan, V.J., 2015. Genomic Reconstruction of an Uncultured Hydrothermal Vent Gammaproteobacterial Methanotroph (Family Methylothermaceae) Indicates Multiple Adaptations to Oxygen Limitation. Front Microbiol. 6, 1425. https://doi.org/10.3389/fmicb.2015.01425.

Stecher, B., Robbiani, R., Walker, A.W., Westendorf, A.M., Barthel, M., Kremer, M., Chaffron, S., Macpherson, A.J., Buer, J., Parkhill, J., Dougan, G., von Mering, C., Hardt, W.D., 2007. Salmonella enterica serovar typhimurium exploits inflammation to compete with the intestinal microbiota. PLoS Biol. 5(10), 2177-89. https://doi.org/10.1371/journal.pbio.0050244.

Stolze, Y., Bremges, A., Rumming, M., Henke, C., Maus, I., Puhler, A., Sczyrba, A., Schluter, A., 2016. Identification and genome reconstruction of abundant distinct taxa in microbiomes from one thermophilic and three mesophilic production-scale biogas plants. Biotechnol Biofuels. 9, 156. https://doi.org/10.1186/s13068-016-0565-3.

Tang, Y.P., Dallas, M.M., Malamy, M.H., 1999. Characterization of the Batl (Bacteroides aerotolerance) operon in Bacteroides fragilis: isolation of a B. fragilis mutant with reduced aerotolerance and impaired growth in in vivo model systems. Mol Microbiol. 32(1), 139-49.

Thakur, M., Parmar, S., Pillai, P., 2006. Studies on growth performance in Kadaknath breed of poultry. Livestock Research for Rural Development. 18, 116.

Wu, Y.W., Simmons, B.A., Singer, S.W., 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 32(4), 605-7. https://doi.org/10.1093/bioinformatics/btv638.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y., 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40(Web Server issue), W445-51. https://doi.org/10.1093/nar/gks479.

Yuzenkova, J., Delgado, M., Nechaev, S., Savalia, D., Epshtein, V., Artsimovitch, I., Mooney, R.A., Landick, R., Farias, R.N., Salomon, R., Severinov, K., 2002. Mutations of bacterial RNA polymerase leading to resistance to microcin j25. J Biol Chem. 277(52), 50867-75. https://doi.org/10.1074/jbc.M209425200.

Zhao, L., Wang, G., Siegel, P., He, C., Wang, H., Zhao, W., Zhai, Z., Tian, F., Zhao, J., Zhang, H., 2013. Quantitative genetic background of the host influences gut microbiomes in chickens. Scientific reports. 3,

Zuo, G., Hao, B., 2015. CVTree3 Web Server for Whole-genome-based and Alignment-free Prokaryotic Phylogeny and Taxonomy. Genomics Proteomics Bioinformatics. 13(5), 321-31. https://doi.org/10.1016/j.gpb.2015.08.004.

Figure legends

Figure 1: Screenshots of binning process. A. Initial plot between coverage of both chicken lines. Each point represents a separate contig. Size of point is relative to contig size. Contigs are coloured by their phylum level assignment. Highlighted red circle represents area enlarged in panel B. B. Enlarged area from plot having cluster of contigs. Contigs from enclosed area was taken for further process. C. PCA plot based on trinucleotide, tetranucleotide and pentanucleotide frequencies along with coverage and GC content. Highlighted PC2 vs PC3 graph was further analysed. D. Enlarged PC2 vs PC3 plot. Selected portion represents contigs taken for further analysis. APK=Kadaknath breed/line; APC=GCB breed/line.

Figure 2: Maximum-likelihood tree as inferred by IQ-Tree. Bootstrap support values are shown at branch nodes in percentages. Parentheses represent the family within the order *Bacteroidales*. *Lactobacillus helveticus*¸ an abundant organism in the chicken caecum microbiome was taken as an outgroup. Species from same genus with more than 90% support values were compressed.

Figure 3: Average-nucleotide identity comparison between the reconstructed genome and A. *Muribaculum intestinale* and B. *Barnesiella viscericola*.

Figure 4: Circular tree view of the reconstructed genome. From the outside to the inside: contigs arranged in descending order and coloured alternatively; PEGs located on forward orientation; PEGs located on reverse orientation; rRNA genes; tRNA genes; GC-ratio; GC-skew. In the GC-ratio and GC-skew rings the colour purple indicates below average, while the colour green indicates above average.

Figure 5: KEGG pathways highlighting enzymes present (green coloured) in A. Glycolysis and B. the Citrate cycle (TCA cycle).