BMC Microbiology

RESEARCH

Open Access

A rapid detection of Avian Pathogenic *Escherichia coli* (APEC) strains based on minimal number of virulence markers identified by whole genome sequencing

Joanna Kazimierczak[1*], Karolina Pospiech[1], Patrycja Sowińska[1], Anna Pękala[1], Paulina Borówka[2], Ewelina A. Wójcik[1], Błażej Marciniak[2], Marcin W. Lis[3], Dominik Strapagiel[2] and Jarosław Dastych[1]

Abstract

Background Colibacillosis is an important epidemiological and economic issue in poultry farming and breeding. A common problem with avian pathogenic *Escherichia coli* strains (APEC) that cause this disease is a lack of an uniform identification system, resulting from a variety of serotypes, phylogenetic groups, sequence types and combinations of virulence factors. There are no clearly defined features that can be associated with pathogenicity. Therefore, without precise identification of pathogenic strains and differentiation from commensal strains, there is no possibility of appropriate selection of targeted therapy.

The widespread use of whole genome sequencing (WGS) in recent years creates new possibilities in diagnostics. Therefore, the purpose of this study was to select features defining the APEC pathotype, based on next generation sequencing (NGS), and design a diagnostic test based on selected factors.

Results A PCR diagnostic test is proposed. Three predictors of virulence were chosen according to in silico analysis: two virulence genes: *iroC* and *hlyF,* as well as one molecular marker of O78 serotype (*wzx*—O-antigen flippase of the O78 serotype). A choice of markers was supported by a chicken embryo model.

Conclusions Whole genome sequencing of *E. coli* genomes allowed for the development of a rapid diagnostic method identifying pathogenic strains for poultry: APEC. The developed test can support field observations connected with the strain isolation source and clinical symptoms of the disease.

Keywords APEC, Colibacillosis, Whole genome sequencing, Virulence factors, Chicken embryo

Background

Colibacillosis, A common bacterial disease causing high mortality, is a significant problem in poultry farming, Contributing to large economic losses for farmers. The disease occurs as a result of infection with avian pathogenic *E. coli* (APEC) [1, 2]. However, it rarely occurs as an independent disease entity. It is most often preceded by: weakened immunity of birds (accompanying, among others: Gumboro disease, Marek's disease, mycoplasmosis, coccidiosis, infectious bronchitis, laryngotracheitis,

*Correspondence:
Joanna Kazimierczak
jkazimierczak@proteonpharma.com
[1] Proteon Pharmaceuticals, Lodz, Poland
[2] Centre for Digital Biology and Biomedical Science – Biobank Lodz, Faculty of Biology and Environmental Protection, University of Lodz, Lodz, Poland
[3] Department of Zoology and Animal Welfare, Univeristy of Agriculture, Cracow, Poland

Newcastle disease, infectious anemia of chickens), damage to the respiratory system, stress, poor environmental conditions (in particular: poor ventilation of rooms, presence of dust and ammonia), herd density and deficiency of certain nutrients. However, it is now believed that APEC strains are very well prepared to function as pathogens, which may indicate that these infections do not always have to be opportunistic [1, 3].

APECs are classified as extra-intestinal pathogenic *E. coli* (ExPEC), which carry important virulence factors for pathogenesis, such as: adhesins, invasins, toxins, protectins or factors responsible for iron acquisition mechanisms. APECs are extremely diverse and can be characterized by the presence of different virulence factors. This is due to the fact that in its genome *E. coli* has a large mobile gene pool acquired as a result of HGT (horizontal gene transfer) localized on plasmids, chromosomal islands and prophages. It is estimated that around 40% of protein-coding genes are encoded on mobile elements [4].

Therefore, the strains adapt to a changing environment, and show increased virulence through new combinations of virulence genes. Moreover, it is hypothesized that the presence of certain pathogenic factors varies with the age of birds and it is difficult to identify genes that are responsible for virulence, regardless of a developmental stage. Virulence factors are most often recognized on pathogenicity islands and transposons, which may be encoded on the bacterial chromosome or plasmids. Conserved Virulence Plasmid (CVP) regions are common and encode groups of genes responsible for virulence. Table S1 in Additional file 1 summarizes the most common and characteristic virulence genes of APEC strains and potential virulence factors, along with a description of their functions and division into the groups described above [1, 5–8].

Despite the increasing number of studies on molecular determinants of virulence in APEC strains, an important element of diagnosis is also serotyping. Serotype information is important because the bird's immune response is directed against the O antigens of the bacteria. In addition, the serotypes of *E. coli* strains vary greatly, depending on a pathogenicity of the strain and a geographic region, from which they originate [1, 3, 9]. So far, it has been established that among *E. coli* serotypes associated with colibacillosis in Europe, serotypes O1, O2, O5, O8, O18 and O78 are most frequently identified, which correspond to more than 50% of poultry infections. Data from eastern China point to strains with serotypes O1, O2, O18 and O78 as the cause of more than 85% of infections, and in Jordan, serotypes O78, O1 and O2 are identified for more than half of the strains causing colibacillosis. However, many strains cannot be typed by traditional serological methods, and there are also rare serotypes (e.g. O6, O24) that cannot be directly linked with the pathogenicity of the strains [1, 3, 10, 11].

Apart from serotype information, antibiotic resistance seems to be important in the context of strains identification. Acquisition and dissemination of resistance genes occur in the HGT process. Interestingly, antibiotic resistance is higher among pathogenic strains than commensal ones, therefore it is assumed that there is a correlation between the bacteria having virulence factors and the antibiotic resistance they acquire. These interactions are not fully explained, they probably depend on the phylogenetic structure of the strain, as well as the geographical region in which the strain occurs. It is possible that in the environment, under the influence of an antibiotic, plasmids that encode virulence and antibiotic resistance genes, undergo the same selection, thereby allowing the maintenance of pathogenicity factors.

The identification of strains responsible for the infection is very difficult due to the fact that APEC strains constitute a heterogeneous group with a multifactorial mechanism of pathogenesis [12]. In addition, a prevalence of commensal *E. coli* in the environment causes problems when isolating strains and assessing their pathogenic potential, as these strains can also cause infection under certain conditions. There is no uniform diagnostic scheme; various techniques are used to assess the pathogenicity of strains, including: serotyping, phylogenetic analysis, or assessment of the occurrence of virulence genes [9].

According to the hypotheses of many research groups, under appropriate conditions, APEC strains may cause infections typical of extraintestinal pathogenic *E. coli* strains for humans (e.g. UPEC) [13, 14]. Research carried out by Mitchell et al. and Hussain et al. show that poultry meat may be a source of ExPEC strains, potentially pathogenic to humans [15, 16]. The research of Jakobsen et al. showed that *E. coli* strains originating from meat were virulent in the UTI model in mice, and their presence was found in urine and bladder and kidneys [17]. Therefore, it is extremely important to develop a diagnostic tool that will quickly and unambiguously identify pathogenic strains that cause infection in the herd. Traditional methods may be supported by in silico analyses performed on genome sequences obtained as a result of next generation sequencing (NGS). With access to information about genome sequences, it is possible to perform serotyping, phylogenetic analyses, MLST, antibiotic resistance and the specific virulence associated genes (VAGs) occurrence.

The popularization of NGS in recent years creates new possibilities in diagnostics. Comparative analyses of genomes and factors responsible for virulence can be

performed on an unprecedented scale. There are already reports on the use of WGS to analyse genetic diversity among APEC strains and the selection of virulence determinants in ExPEC strains [18–21]. This gives hope for a clear selection of features defining the APEC pathotype and a design of a diagnostic test based on selected factors.

Methods

Bacterial strains

E. coli strains from Proteon Pharmaceuticals bacterial collection used in the studies came from poultry (chicken broilers, laying hens, turkeys, geese), both from animals with colibacillosis ($n=101$) and from healthy individuals ($n=32$) (original classification presented in Table S2 in Additional file 2).

DNA extraction

Genomic DNA from bacteria was isolated using a commercial kit (Wizard SV Genomic DNA Purification System; Promega) following a protocol modified by Proteon Pharmaceuticals. Briefly, bacteria plated on a solid LB medium were taken from the plate (one eye 10 µl loop) and suspended in PBS buffer until a uniform suspension was obtained. 200 µl Nuclei Lysis Solution, 50 µl 0.5 M EDTA pH 8.0, 2 µl Proteinase K (≥10 mg/ml) and 5 µl Rnase A (4 mg/ml) were added to the tube, and gently mixed by inversion. Then, a 15-min incubation at 55 °C was performed and manufacturer's instructions were followed. DNA concentrations were measured using a BioSpectrometer® (Eppendorf, Hamburg, Germany) and DNA was stored at -20 °C.

Phylogenetic analysis

Phylogenetic analysis was performed according to modified method by Clermont *et. al.*, i.e. PCR typing, which assigned the strains to groups A, B1, B2, D, C, E, F and clade I according to the scheme [22].

NGS sequencing

The genomes of the strains classified as distinct by the MP PCR method, according to Krawczyk et al. (Melting Profile PCR; data not shown) [23], were subjected to whole genome sequencing (WGS). For this purpose, genomic libraries were prepared for each strain DNA sample using the Nextera XT DNA Library Preparation Kit (Illumina, USA). The correctness of library preparation was assessed based on the results of capillary electrophoresis. In order to properly normalize libraries and ensure the correct number of reads during sequencing for each library, the molarity of the libraries was assessed using the qPCR method. Libraries were sequenced using NGS technology on the Illumina NextSeq 500 platform in paired 2×150 bp reads.

Bioinfomatics analysis

De novo assembly of *E. coli* genomes was performed using the SPAdes 3.7.1. and 3.8.0 [24] with necessary manual editing. The quality of the assembly was verified each time in the QUAST program [25] taking into account, among others genome length, number of contigs, N50. Genomes were annotated in RAST [26]. SerotypeFinder 2.0 was used for in silico serotyping of *E. coli* strains [27]. The following settings were used: *E. coli* organism, sequence identity threshold: 85%, minimum gene sequence coverage: 60%. The presence of the 194 amino acid sequences of selected virulence factors (Table S1, Additional file 1) was assessed in the annotated sequences of the *E. coli* strains. The blastp algorithm was used with the following parameters: sequence coverage: 70%, sequence identity threshold: 70%. Then, using an original script, a presence of virulence factors was visualized on a heat map. In strain clustering, the farthest point algorithm was used. Independent strain clusters similar to each other were marked with different colors.

Primers' design and PCR

Primers for the PCR-based diagnostic method were designed using the Primer-BLAST, selecting the size parameters of the amplicons and melting points to perform a multiplex reaction (Table 1) [28]. Multiplex PCR reaction was performed according to the following parameters: initial denaturation in 95°C for 5 min and then 30 cycles of denaturation in 95°C for 60 s, annealing in 57°C for 60 s, elongation in 72°C for 120 s.

In ovo tests

The experimental design during *in ovo* tests was based on the work of Polakowska et al., where chicken embryos were infected with *S. aureus* strains and their mortality was assessed over time [29].

Table 1 Primers used in the study

Primer name	Sequence 5′→3′	Tm [°C]	Expected product size (bp)
iroC-F	ACTATGTGCGCCGTGGTTAT	51.3	732
iroC-R	GTGAACGGGTGTCGATCAGT	53.8	
hlyF-F	GAGCACCTACTCCACAAGCG	55.9	458
hlyF-A-R	TCGGGCAACCAACAAAGGTA	51.8	
O78-A-F	CACAACTCTCGGCAATATATC ATCA	57.0	994
O78-A-R	TATGGGTTTGGTGGTACGTAGT	57.0	

Experiments were carried out on hatching chicken eggs, which were incubated in commercial incubators (PreformPoldrob) for 8 days preceding the experiments. Then, only eggs with properly developing embryos were selected for testing. On the 9th day of incubation, a challenge with *E. coli* strains was performed by inoculation of bacterial strains intraallantoically in a single infecting dose of 5×10^4 CFU/embryo (the first study involved optimizing the infecting dose). The hatched eggs were divided into groups of 30 eggs. Each time, several groups were infected with the tested strains and additional 2 groups were used as controls: a negative zero group (not treated with anything) and a negative control group (treated with a sterile solution of 0.85% NaCl, in which the strains for infection were suspended). Before injection, egg shells were disinfected with a 70% ethyl alcohol solution and a hole was made with a 1.2 mm needle. Suspensions were injected with a 0.5 mm needle and 1 ml insulin syringes. Thereafter, the hole in each egg was aseptically sealed with hot wax and incubation was continued in Mesalles 65 DIGIT laboratory incubators at 37.8 °C and 50% relative humidity. The mortality of embryos was assessed daily on the basis of candling with an ovoscope. Experiments were terminated before hatching began and all live embryos were euthanized. A developmental phase of the embryo at the time of death was assessed on the basis of embryopathological analysis.

Fisher's exact test

The Fisher's exact test was used to verify the null hypothesis of the same frequency of occurrence of a given gene in the studied populations (pathogenic and non-pathogenic strains). For this purpose, the RStudio (v 1.1.456, v 3.5.1), the *stats* package, version 3.6.2 was used. In addition, an analysis of the frequency of particular serotypes in *E. coli* strains was assessed. Then, in order to verify the accuracy of individual classifiers (predictions), the original algorithm was used, thanks to which the ROC (Receiver Operating Characteristic) curves were plotted. For each combination of genes, the AUC (Area Under Curve) was calculated, taking values from the range [0,1].

Results

Characteristics of *E.coli* strains

Bacterial strains from both healthy and sick birds, defined as *E. coli,* were differentiated in order to select unique strains (data not shown). The genomes of these strains were sequenced, and 109 unique strains were identified (sequences deposited in the NCBI database, BioProjects PRJNA1126135 and PRJNA319144; additional supplementary data with basic sequence metrics are provided in Table S3, Additional file 3). They were subjected to further in silico and *in ovo* analysis.

It was shown by in silico analysis that the most common among tested strains were the following serotypes: O78 ($n=21$), O50 ($n=8$) and O88 ($n=6$). For 12 genomes, information on the O serotype could not be obtained. For the 97 identified strains, the above-mentioned serotypes accounted for 36.1% of all cases, confirming the importance of their occurrence in Europe. For the remaining 63.9% of strains, the identified serotypes are not characteristic and it would be difficult to classify them as pathogenic or non-pathogenic on this basis.

In order to examine the presence of selected virulence factors (Table S1, Additional file 1), an analysis of the presence of amino acid sequences in the annotated sequences of tested *E. coli* strains was carried out (described in Methods section). Results were recorded in a binary manner, where 0 meant the absence of a given factor, and 1 the presence of a given factor in the strain. Then, the binary table for all strains was visualized as a heat map (Figure S1 in Additional file 8 — Analysis of the presence of selected virulence factors in *E. coli* strains (heat map)). Only virulence factors for which selected sequences' coverage and similarity were greater than or equal to 70% were colored on the heat map. The intensity of the blue color corresponds to the percent sequences' similarity and is according to the scale shown on the right side of the heat map.

In ovo results

Taking into account the lethality of chicken embryos depending on the strain used, it was observed that not in every case the biological data are consistent with the original classification of the pathogenicity of the strains or the results of these experiments are not reproducible. Therefore, the original classification of the strains into pathogenic and non-pathogenic was verified, based on the health status of the birds from which the strains were isolated (birds with or without infection). The results obtained in the lethality test were treated as superior while results obtained from the phylogenetic analysis and the presence of characteristic APEC serotypes were treated as supplementary information. Based on the results of the *in ovo* test, the strain collection was reclassified according to the following assumptions:

- for strains for which CV (coefficient of variation) < 30%, only mortality and average death days were taken into account [30],
- non-pathogenic strains were those with observed mortality rates below 75% on day 19th (at the end of the experiment) and reproducible results (i.e. at least two repetitions were required to classify a given strain as non-pathogenic); an additional deci-

sion criterion in cases where when the mortality rate was < 75%, 85% > the average death day was > 12th,
- strains with mortality rates exceeding 85% on day 19th (at the end of the experiment) and reproducible results, were considered pathogenic (i.e. at least two repetitions were required to classify a given strain as pathogenic); an additional decision criterion in cases where the mortality rate was within the < 75%, 85% > range was the average death day ≤ 12,
- for strains for which a consistent classification was obtained for two out of three replicates, a third experiment was not performed,
- for strains for which no consistent classification was obtained from two replicates, further experiments were performed and inconsistent replicates were discarded,
- for strains for which no consistent classifications were obtained from individual replicates or for which replicates were rejected due to CV > 30%, the strain was considered unreliably classified and removed from the pool of reclassified strains.

The final group after reclassification, for which analyzes were carried out, consisted of 100 strains (83 pathogenic "P" strains and 17 non-pathogenic "NP" strains). This collection was used to develop a diagnostic method allowing for the identification of APEC. 25 strains classified in the final collection had their original classification changed. For 9 strains it was not possible to determine the group (P/NP). Results summarizing re-classification of all strains in the study with information about their origin (bird with or without infection), phylogenetic analysis, serotypes (characteristic/non-characteristic) and *in ovo* results with final classification are presented in Table S4 (Additional file 4).

Fisher's exact test

To verify the assumption that it is possible to select a few genes that will allow to determine the pathogenicity of *E. coli* strains, the Fisher's exact test was used to verify the null hypothesis of the same frequency of a given gene in both populations. Statistically significant differences ($p < 0.05$) were observed in the frequency of 51 genes (including 27 at $p < 0.001$) in pathogenic and non-pathogenic strains. In addition, an analysis of the presence of certain serotypes was performed. Among the analyzed populations, 40 different serotypes were observed, with a clear predominance of the O78 serotype ($n = 21$); followed by O50 ($n = 8$), O88 ($n = 6$), O161, O2, and O8 ($n = 5$), and O117 ($n = 4$). For the most common serotypes, the relationship between their occurrence and the pathogenicity of the strains was tested (Fisher's exact test). This analysis showed statistically significant differences only in the serotype O78 ($p = 0.020$).

In addition, this serotype was strongly correlated with pathogenicity (21/21), and in 14%

of cases, it was independent of the presence of virulence genes. Therefore it was included in the analysis and considered as one of the determinant factors.

Results of the frequency of occurrence of the analyzed factors in all strains together with the *p*-value (*p*-value) are presented in Table S5, Additional file 5 (genes from Table S1 in Additional file 1, which were not present in any of the strains, are not included). Based on the analysis of the results, it can be concluded that the best discriminants in the predictive model of pathogenicity of *E. coli* strains are genes related to iron metabolism and genes encoding toxins. To avoid the situation where in the prediction model we are dealing with genes of one operon, the following genes were selected for the final prediction model: *iroC, hlyF, etsA, iucB, ompT, iutA, fyuA, ybtE, irp1* and the gene encoding flippase of O-antigen in serotype O78 (*wzx*).

Since the number of pathogenic strains and commensals was not comparable in the defined collection, to check whether the discrepancy in the number of tested groups was important for the results obtained in the Fisher test, an experiment with equal numbers of sets was performed, sampling 17 out of 83 pathogenic strains ten times and using all 17 non-pathogenic strains for analysis (data not shown). It was concluded that the results obtained in the first Fisher's test carried out on non-equal sets are reliable and can be used to develop predictive models of the pathogenicity of *E. coli* strains.

Then, various combinations of selected 11 virulence factors were tested. The effectiveness of predicting the pathogenicity of *E. coli* strains with selected genes was evaluated. All analyzed combinations and the described test evaluation parameters are presented in Table S6, while the ROC curves and areas under the AUC graph are presented in Figure S2 in Additional File 6.

Based on the abovementioned results, the analysis of the presence of the following genes: *iroC*, *hlyF* and *wzx*—O-antigen flippase of the O78 serotype was selected, where the presence of any of them indicates that the strain is classified as pathogenic. The diagnostic test composed of these three genes, would facilitate analysis and reduce costs. Table S7 in Additional file 7 shows the analysis of genes presence and the classification of each strain compared to the *in ovo* test. The selected model is characterized by high sensitivity (98.80%), positive and negative predictive value (93.18% and 91.67%) and accuracy (93.00%). The only test parameter with lower values is specificity (64.71%,

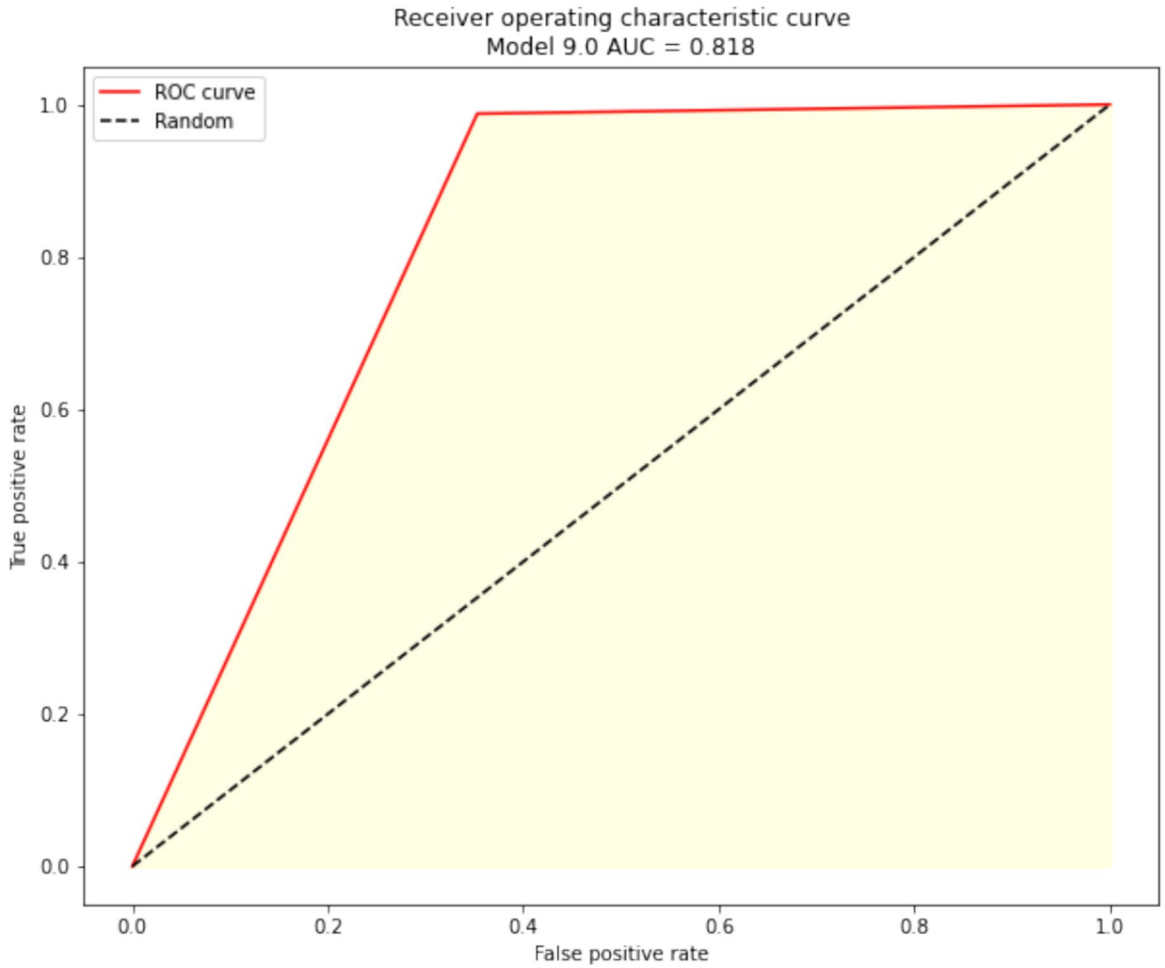

Fig. 1 ROC of a chosen model with 98.80% sensitivity and 64.71% specificity

6 false positives). AUC equals to 0.818 in this case (plot presented in Fig. 1).

Verification of PCR diagnostic method

For verification of in silico observation for chosen genes (*iroC*, *hlyF* and *wzx*), the PCR analysis was performed. To this end PCR primers were designed (Table 1) and the amplification of selected virulence genes was carried out as described in Methods Sect. 100% accordance with the in silico results was obtained, including 93 correct results and 7 incorrect results (6 false positives and 1 false negative) relating to classification based on the *in ovo* model.

An example of a PCR electropherogram for selected strains is presented in Fig. 2. According to the presented results, *E. coli* strains: 002PP2015, 009PP2015 and 126PP2016 are treated as pathogenic while the strain 082PP2016 is classified as non-pathogenic.

Discussion

In the presented work, whole genome sequencing (WGS) of *E. coli* strains was used to conduct a bioinformatics analysis to identify selected virulence factors, design a PCR-based diagnostic method and evaluate its effectiveness. Additionally, the assignment of strains to the appropriate pathogenicity groups was verified based on embryo lethal tests in the *in ovo* model. The method used in this work is an innovative approach to assess the pathogenicity of strains and select appropriate VAGs.

Despite its increasing popularity, WGS is not the first choice method for analyzing the virulence of APEC strains, mainly due to costs but also due to the complexity of the pathogenicity mechanism of APEC strains and the lack of direct in silico algorithms allowing the classification of *E. coli* as APEC. There are many diagnostic approaches in the literature based on the PCR method, which usually focus on the analysis of only a few pathogenicity genes [7, 31–34].

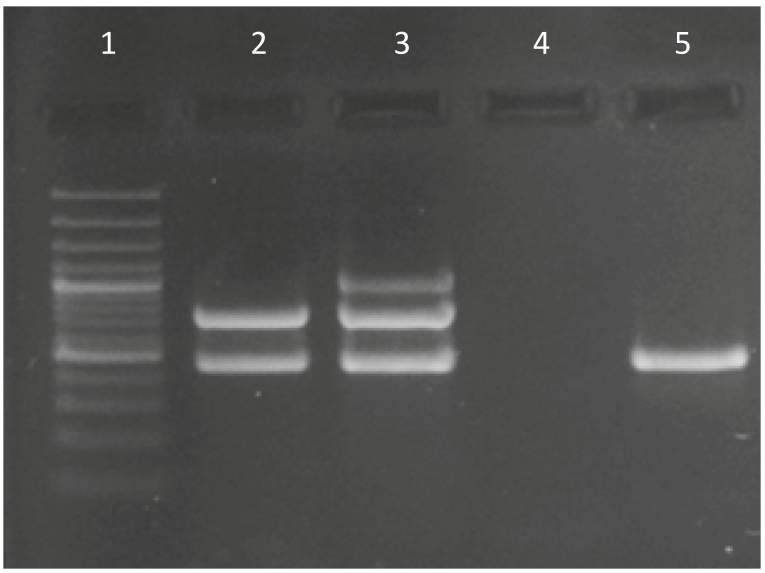

Fig. 2 An example of PCR electropherogram. Description of the lanes: 1: GeneRuler 100 bp DNA Ladder (Thermo Scientific™) (with the following distribution – 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1500, 2000, 3000 bp); 2–5: PCR products for *E.coli* strains: 002PP2015, 009PP2015, 082PP2016 and 126PP2016

A similar approach to the problem used in the presented work is described in the publication by Ronco et al. [19]. The authors of this work sequenced the genomes of 114 *E. coli* isolates, focusing mainly on the analysis of one serotype O78:H4, sequence type ST117 and several virulence genes. The group of Cordoni et al. also sequenced the genomes of 95 APEC strains originating from Europe and designed PCR tests defining these strains based on in silico analyses [18]. Azam et al. studied 75 sequences of *E.coli* strains coming from dead broilers in Pakistan [21] while Chen et al. analyzed 125 sequences of APEC isolates form birds with colibacillosis symptoms [35]. In our study, an extensive bioinformatics analysis of *E. coli* genomes was performed. It is worth noting that when assessing the presence of pathogenicity determinants in the strains, amino acid sequences were analyzed to assess significant changes that may affect protein function, rather than synonymous changes, which was an innovative approach.

According to the results, the most differentiating factors indicating pathogenicity were genes encoding mechanisms of iron acquisition: siderophores *iro, iuc, fyuA, ybt,* transporters *ets*, hemolysin *hlyF* and outer membrane protein *ompT*. Many reported PCR tests for determining the pathogenicity of APEC strains focus on the amplification of some of the aforementioned genes. Among studies, such an approach can be found, among others: in Subedi et al., who examined the presence of 11 virulence genes, including *iucD, ompT* and *hlyF* [36]. The work of De Carli et al. focused on the analysis of 10 genes, including: *iroN, fyuA, ompT* and *hlyF* [34]. However, no studies have been found in the available literature in which several hundred virulence factors were analyzed simultaneously, examining their impact on the pathogenicity of the strain. In the work presented here, an in silico analysis was performed based on NGS sequencing of the genomes of all strains.

In our study we proved that the presence of serotype O78 was the additional pathogenicity factor. It was found that, among many others, only this serotype is present in APEC strains with a statistically higher frequency and that it can itself be a determinant of pathogenicity—in 14% of cases it occurred independently of other virulence factors. This observation is consistent with literature data on the presence of serotype O78 in APEC strains. In the work of Ronco et al., similarly to the results presented in this study, the predominance of the O78 serotype compared to other serotypes defined for APEC strains was noted (presence in 43% of the isolates tested in Ronco et al. and in 20% of the isolates in this work) [19]. Chen et al. notified the predominance of O78 serogroup as well (35.46%) [35]. In the study by Huja et al., it was noted that among the *E. coli* serotypes most frequently causing colibacillosis (O1, O2 and O78), the strains defined as O78 differ from the two other groups. They do not have genes encoding the polysaccharide capsular antigen K1; their envelope is structured similarly to the somatic O antigen, which seems to be important for pathogenesis [37]. The

analysis of the genomes of *E. coli* strains of serotype O78 allowed us to conclude that they are more closely related to enterotoxigenic intestinal strains than to extraintestinal pathogens of serotype O1. A recent work of Biran et al. showed that *Escherichia coli* of O78 group are septicemic by producing O-antigen capsule (group 4 capsule) which is an essential virulence factor for the spread of bacteria in the blood [38]. However, in the works published so far, there is no information about combining the analysis of virulence genes and the occurrence of appropriate serotypes in one PCR test. Only in the research published by Dissanayake et al., PCR-based detection of the presence of 13 virulence genes and, in parallel, the O78 serotype in pathogenic strains was used to analyze the population of APEC strains from the territory of Sri Lanka [33].

The literature reports numerous attempts to develop a diagnostic test for detecting APEC strains, however, in the vast majority of cases, there is no verification of the pathogenicity of the strains analyzed in experiments. The pathogenicity of a strain is most often assessed based on its source and clinical symptoms (healthy/sick bird) [7, 32, 34, 35, 39]. Some researchers attempted to evaluate the virulence of *E. coli* strains in in vitro or in vivo models, such as Stromberg et al., who infected 5-week-old chickens with selected strains through air sacs and assessed their impact on the development of the birds' internal organs [40].

In the presented work, a preliminary classification of bacterial strains was also made based on the source of the isolates—from birds with colibacillosis or healthy ones. However, in the next an experimental assessment of the pathogenicity of all strains was carried out. It was decided to adapt the method described in the publication by Polakowska et al. [29]. On the basis of the results obtained from experiments on infection of embryos with *E. coli* strains, the final decision was made regarding the classification of strains into pathogenic and non-pathogenic, and the development of a diagnostic method began. It was noticed that some strains changed their assignment, which is consistent with publications indicating the pathogenic potential among commensal strains, as well as the possibility of isolating commensal strains from the organs of birds affected by colibacillosis [1, 40]. However, although the *in ovo* embryonic lethality test was accepted as conclusive, this model had some limitations. Chicken embryos were very sensitive to any manipulation, even in saline controls, so a relatively high cut-off threshold was used to determine that the strain was pathogenic (mortality > 85%). However, even with this value selected, it was not possible to obtain a larger number of non-pathogenic strains in the final classification. Therefore, it is possible that the developed method indicates the pathogenic potential of APEC strains, and some of the strains classified as pathogens may cause disease only under favorable conditions (e.g. in weakened individuals).

As part of the research, apart from assessing the presence of selected virulence factors in the strains, a phylogenetic analysis according to Clermont et al. was performed to verify the possibility of classifying strains into pathogenic and non-pathogenic with this method [22]. After comparing the obtained results with the *in ovo* analysis, it was found that some of the strains ultimately classified as pathogenic actually belonged to "pathogenic" phylogenetic groups, i.e. B2 and F. However, some of the samples assigned to commensal groups B1 and C, were in fact pathogens. This applies in particular to strains of the O78 serotype, which is confirmed by literature to have a higher similarity to intestinal pathogens than extraintestinal ones [37]. Additionally, research conducted by Rodriguez-Siek et al. revealed that most of the 524 APEC isolates tested were assigned to phylogenetic groups typically represented by commensal strains [31]. Moreover, it is believed that due to the occurrence of hybrid groups, up to 80–85% of *E. coli* strains may be incorrectly assigned to phylogenetic groups [8, 41]. Therefore, phylogenetic analysis should only be treated as a complement to the characterization of *E. coli*. Classifying a strain as APEC based solely on such an analysis may lead to many false-negative results.

Serotyping using the classical or in silico method is also a method supporting the diagnosis of APEC strains. It may be particularly helpful in the case of characteristic serotypes, such as O1, O2, O8, or O78. However, as research has shown, many *E. coli* strains pathogenic to poultry have uncharacteristic or unknown serotypes. Therefore, basing diagnosis only on serotyping does not provide reliable results. Additionally, understanding the antibiotic resistance profile of strains is a valuable cognitive element in the characterization of APEC, but in the era of increasing multidrug resistance of bacteria, it is not possible to conclude on the pathogenicity of a strain on this basis.

Fisher's exact test was then used to assess the selection of appropriate virulence factors for the diagnostic test. In this way, genes were selected for which statistically significant differences were found in the pathogenic and non-pathogenic groups. Among them, several sets of genes were selected that could constitute discriminants in the PCR test. Each gene set was assessed in silico by calculating the test sensitivity, specificity, accuracy, and positive and negative predictive values, as well as by analyzing the area under the plot (AUC) for the ROC curve. Ultimately, the version with the smallest number of analyzed genes was selected—including 2 virulence genes (*iroC* and *hlyF*)

and the gene encoding the O antigen flippase of serotype O78 (*wzx*). This was due to the practical advantage of this gene combination over others due to the number of genes analyzed, while maintaining acceptable method parameters. A small number of genes analyzed is beneficial from the point of view of speed of diagnosis and reducing its costs.

The developed diagnostic test has high accuracy (93.00%) and sensitivity (98.80%) parameters in parallel with lower specificity (64.71%). It was decided to leave the method in its current form and shift the cut-off point towards sensitivity, because colibacillosis is a serious disease of poultry, causing huge losses and the high sensitivity of the method increases the chance of detecting the infection very early, which is important for effective therapy. It should be added that the division of groups into pathogenic and non-pathogenic indicates the potential of the analyzed *E. coli* strain to cause disease, and not the disease itself. Moreover, the research included the already mentioned much larger number of virulent strains than commensal ones (83 to 17), and in order to make a proper comparison, their number should be similar. It is possible that the false-positive rate would be lower with more non-pathogenic strains.

In the original classification, there were more commensal strains and therefore the number of strains in the P and NP groups was more balanced. After reclassifying the strains, it was not possible to obtain a larger number of non-pathogenic representatives. The approach used in the described work is not unique. The problem of unbalanced sets can also be found in the literature data. Silveira et al., comparing APEC and AFEC strains from different regions of Brazil, had 91 pathogenic and 29 non-pathogenic strains [42]. The predominance of virulent strains over commensal ones may be explained by the fact that even isolates obtained from the intestinal flora of healthy birds may carry important virulence factors, as demonstrated in the study of Stromberg et al. [40]. In this work, an intestinal *E. coli* isolate MM218 from a healthy chicken caused bacteremia and meningitis in a mouse model, and another isolate MM29 caused sepsis in chickens in less than 20 h. Due to the difficulty of finding standard non-virulent strains, the presented work allowed for the possibility of imbalance between pathogenic and non-pathogenic sets. Additionally, it proves the importance of the research conducted in the context of food safety and its impact on human health.

Conclusions

The results obtained from the conducted research allowed for the development of a quick and inexpensive diagnostic method based on multiplex PCR, which effectively allows the detection of Avian pathogenic *Escherichia coli* (APEC). However, before introducing the test into practice, it is necessary to conduct additional validation on an independent dataset using experimental and clinical data extended to models other than the chicken embryo viability test (e.g. infecting birds into air sacs), because it has its limitations described above.

Abbreviations

APEC	Avian Pathogenic *Escherichia coli*
AUC	Area under the curve
CV	Coefficient of variation
CVP	Conserved Virulence Plasmid
ExPEC	Extraintestinal Pathogenic *Escherichia coli*
HGT	Horizontal Gene transfer
LB	Lysogeny broth
MLST	Multi-Locus Sequence Typing
MP PCR	Melting Profile PCR
NGS	Next Generation Sequencing
PCR	Polymerase Chain Reaction
ROC	Receiver Operating Characteristics
UPEC	Uropathogenic *Escherichia coli*
VAG	Virulence associated gene
WGS	Whole Genome Sequencing

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12866-025-03861-4.

Supplementary Material 1.
Supplementary Material 2.
Supplementary Material 3.
Supplementary Material 4.
Supplementary Material 5.
Supplementary Material 6.
Supplementary Material 7.
Supplementary Material 8.

Acknowledgements

Not applicable.

Authors' contributions

Conceptualization: JD and DS; methodology: JK, PS, AP, MWL; investigation: JK, KP; data analysis: JK, KP, PB, BM; writing— original draft preparation: JK; writing—review and editing: KP, EAW; project administration: EAW. All authors have read and agreed to the published version of the manuscript.

Funding

Research founded by project POIR.01.01.01–00-0149/16 "The integrated system for diagnostics and prevention of pathogenic E. coli infections in poultry flocks", co-financed by The National Centre for Research and Development from " Measure 1.1 R&D projects of enterprises Sub-measure 1.1.1 Industrial research and development work implemented by enterprises. Smart Growth Operational Program 2014–2020.

Data availability

The datasets generated and analysed during the current study are available in the NCBI repository, BioProject no PRJNA1126135 and PRJNA319144 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1126135, https://www.ncbi.nlm.nih.gov/bioproject/319144).

Declarations

Ethics approval and consent to participate

This study is complied with relevant institutional, national, and international guidelines and legislation (i.e. in accordance with the Act on the Protection

of Animals Used for Scientific or Educational Purposes in Poland which implements Directive 2010/63/EU of the European Parliament and the Council of 22 September 2010 on the protection of animals used for scientific purposes).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 2 December 2024 Accepted: 3 March 2025
Published online: 17 March 2025

References

1. Barnes HJ, Nolan LK, Vaillancourt J-P, et al. Colibacillosis. In: Saif YM, Fadly AM, Glisson JR, et al., editors. Diseases of Poultry. 12th ed. AMES, Iowa, USA: Blackwell Publishing; 2008. p. 691–737.
2. Kathayat D, Lokesh D, Ranjit S, Rajashekara G. Avian Pathogenic *Escherichia coli* (APEC): an overview of virulence and pathogenesis factors, zoonotic potential, and control strategies. Pathogens. 2021;10(4):467. https://doi.org/10.3390/pathogens10040467.
3. Guabiraba R, Schouler C. Avian colibacillosis: still many black holes. FEMS Microbiol Lett. 2015;362(15):fnv118. https://doi.org/10.1093/femsle/fnv118.
4. Davids W, Zhang Z. The impact of horizontal gene transfer in shaping operons and protein interaction networks–direct evidence of preferential attachment. BMC Evol Biol. 2008;8:23. https://doi.org/10.1186/1471-2148-8-23. Published 2008 Jan 24.
5. Dziva F, Stevens MP. Colibacillosis in poultry: unravelling the molecular basis of virulence of avian pathogenic *Escherichia coli* in their natural hosts. Avian Pathol. 2008;37(4):355–66. https://doi.org/10.1080/03079450802216652.
6. Baldy-Chudzik K, Bok E, Mazurek J. Znane i nowe warianty patogennych *Escherichia coli* jako konsekwencja plastycznego genomu. [Known and new variants of pathogenic *Escherichia coli* as a consequence of a plastic genome]. Postepy Hig Med Dosw (online). 2015;69:345–61.
7. Paixão AC, Ferreira AC, Fontes M, et al. Detection of virulence-associated genes in pathogenic and commensal avian *Escherichia coli* isolates. Poult Sci. 2016;95(7):1646–52. https://doi.org/10.3382/ps/pew087.
8. Sarowska J, Futoma-Koloch B, Jama-Kmiecik A, et al. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. Gut Pathog. 2019;11:10. https://doi.org/10.1186/s13099-019-0290-0.
9. Fratamico PM, DebRoy C, Liu Y, Needleman DS, Baranzoni GM, Feng P. Advances in molecular serotyping and subtyping of *Escherichia coli*. Front Microbiol. 2016;7:644. https://doi.org/10.3389/fmicb.2016.00644.
10. Wang S, Meng Q, Dai J, et al. Development of an allele-specific PCR assay for simultaneous sero-typing of avian pathogenic *Escherichia coli* predominant O1, O2, O18 and O78 strains. PLoS One. 2014;9(5):e96904. https://doi.org/10.1371/journal.pone.0096904.
11. Ibrahim RA, Cryer TL, Lafi SQ, Basha EA, Good L, Tarazi YH. Identification of *Escherichia coli* from broiler chickens in Jordan, their antimicrobial resistance, gene characterization and the associated risk factors. BMC Vet Res. 2019;15(1):159. https://doi.org/10.1186/s12917-019-1901-1.
12. Gambi L, Rossini R, Menandro ML, et al. Virulence factors and antimicrobial resistance profile of *Escherichia coli* isolated from laying hens in Italy. Animals (Basel). 2022;12(14):1812. https://doi.org/10.3390/ani12141812.
13. Manges AR. *Escherichia coli* and urinary tract infections: the role of poultry-meat. Clin Microbiol Infect. 2016;22(2):122–9. https://doi.org/10.1016/j.cmi.2015.11.010.
14. Jørgensen SL, Stegger M, Kudirkiene E, et al. Diversity and population overlap between avian and human *Escherichia coli* belonging to sequence type 95. mSphere. 2019;4(1):e00333-18. https://doi.org/10.1128/mSphere.00333-18.
15. Mitchell NM, Johnson JR, Johnston B, Curtiss R 3rd, Mellata M. Zoonotic potential of *Escherichia coli* isolates from retail chicken meat products and eggs. Appl Environ Microbiol. 2015;81(3):1177–87. https://doi.org/10.1128/AEM.03524-14.
16. Hussain A, Shaik S, Ranjan A, et al. Risk of transmission of antimicrobial resistant *Escherichia coli* from commercial broiler and free-range retail chicken in India. Front Microbiol. 2017;8:2120. https://doi.org/10.3389/fmicb.2017.02120.
17. Jakobsen L, Garneau P, Bruant G, et al. Is *Escherichia coli* urinary tract infection a zoonosis? Proof of direct link with production animals and meat. Eur J Clin Microbiol Infect Dis. 2012;31(6):1121–9. https://doi.org/10.1007/s10096-011-1417-5.
18. Cordoni G, Woodward MJ, Wu H, Alanazi M, Wallis T, La Ragione RM. Comparative genomics of European avian pathogenic *E. coli* (APEC). BMC Genomics. 2016;17(1):960. https://doi.org/10.1186/s12864-016-3289-7.
19. Ronco T, Stegger M, Olsen RH, et al. Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production. BMC Genomics. 2017;18(1):13. https://doi.org/10.1186/s12864-016-3415-6.
20. Lindstedt BA, Finton MD, Porcellato D, Brandal LT. High frequency of hybrid *Escherichia coli* strains with combined Intestinal Pathogenic *Escherichia coli* (IPEC) and Extraintestinal Pathogenic *Escherichia coli* (ExPEC) virulence factors isolated from human faecal samples. BMC Infect Dis. 2018;18(1):544. https://doi.org/10.1186/s12879-018-3449-2.
21. Azam M, Mohsin M, Johnson TJ, et al. Genomic landscape of multi-drug resistant avian pathogenic *Escherichia coli* recovered from broilers. Vet Microbiol. 2020;247:108766. https://doi.org/10.1016/j.vetmic.2020.108766.
22. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: improvement of specificity and detection of new phylo-groups. Environ Microbiol Rep. 2013;5(1):58–65. https://doi.org/10.1111/1758-2229.12019.
23. Krawczyk B, Stojowska K, Leibner-Ciszak J. Opracowanie zestawu diagnostycznego do genetycznego typowania szczepów bakteryjnych metodą PCR MP [Development of a diagnostic kit for genetic typing of bacterial strains using the PCR MP method]. Med Dośw Mikrobiol. 2008;60:139–54.
24. Nurk S, Bankevich A, Antipov D, et al. Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads. In: Deng M, Jiang R, Sun F, Zhang X, editors. Research in Computational Molecular Biology. RECOMB 2013. Lecture Notes in Computer Science(), vol 7821. Berlin, Heidelberg: Springer; 2013. https://doi.org/10.1007/978-3-642-37195-0_13.
25. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072–5. https://doi.org/10.1093/bioinformatics/btt086.
26. Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:75. https://doi.org/10.1186/1471-2164-9-75. Published 2008 Feb 8.
27. 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):2410–26. https://doi.org/10.1128/JCM.00008-15.
28. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13:134. https://doi.org/10.1186/1471-2105-13-134. Published 2012 Jun 18.
29. Polakowska K, Lis MW, Helbin WM, et al. The virulence of *Staphylococcus aureus* correlates with strain genotype in a chicken embryo model but not a nematode model. Microbes Infect. 2012;14(14):1352–62. https://doi.org/10.1016/j.micinf.2012.09.006.
30. Stefaniuk E, Bosacka K, Hryniewicz W. Walidacja i weryfikacja metod i testów diagnostycznych w laboratorium mikrobiologicznym [Validation and verification of diagnostic methods and tests in a microbiological laboratory]. Post Mikrobiol. 2015;54(4):415–24.
31. Rodriguez-Siek KE, Giddings CW, Doetkott C, Johnson TJ, Nolan LK. Characterizing the APEC pathotype. Vet Res. 2005;36(2):241–56. https://doi.org/10.1051/vetres:2004057.
32. Johnson TJ, Wannemuehler Y, Doetkott C, Johnson SJ, Rosenberger SC, Nolan LK. Identification of minimal predictors of avian pathogenic *Escherichia coli* virulence for use as a rapid diagnostic tool. J Clin Microbiol. 2008;46(12):3987–96. https://doi.org/10.1128/JCM.00816-08.
33. Dissanayake DR, Octavia S, Lan R. Population structure and virulence content of avian pathogenic *Escherichia coli* isolated from outbreaks in Sri Lanka. Vet Microbiol. 2014;168(2–4):403–12. https://doi.org/10.1016/j.vetmic.2013.11.028.

34. De Carli S, Ikuta N, Lehmann FK, et al. Virulence gene content in *Escherichia coli* isolates from poultry flocks with clinical signs of colibacillosis in Brazil. Poult Sci. 2015;94(11):2635–40. https://doi.org/10.3382/ps/pev256.
35. Chen X, Liu W, Li H, et al. Whole genome sequencing analysis of avian pathogenic *Escherichia coli* from China. Vet Microbiol. 2021;259:109158. https://doi.org/10.1016/j.vetmic.2021.109158.
36. Subedi M, Luitel H, Devkota B, et al. Antibiotic resistance pattern and virulence genes content in Avian Pathogenic *Escherichia coli* (APEC) from broiler chickens in Chitwan. Nepal BMC Vet Res. 2018;14(1):113. https://doi.org/10.1186/s12917-018-1442-z.
37. Huja S, Oren Y, Trost E, et al. Genomic avenue to avian colisepticemia. mBio. 2015;6(1):e01681-14. https://doi.org/10.1128/mBio.01681-14. Published 2015 Jan 13.
38. Biran D, Rosenshine I, Ron EZ. Escherichia coli O-antigen capsule (group 4) is essential for serum resistance. Res Microbiol. 2020;171(2):99–101. https://doi.org/10.1016/j.resmic.2019.12.002.
39. LeStrange K, Markland SM, Hoover DG, Sharma M, Kniel KE. An evaluation of the virulence and adherence properties of avian pathogenic *Escherichia coli*. One Health. 2017;4:22–6. https://doi.org/10.1016/j.onehlt.2017.08.001.
40. Stromberg ZR, Johnson JR, Fairbrother JM, et al. Evaluation of *Escherichia coli* isolates from healthy chickens to determine their potential risk to poultry and human health. PLoS One. 2017;12(7):e0180599. https://doi.org/10.1371/journal.pone.0180599.
41. Ewers C, Li G, Wilking H, et al. Avian pathogenic, uropathogenic, and newborn meningitis-causing *Escherichia coli*: how closely related are they? Int J Med Microbiol. 2007;297(3):163–76. https://doi.org/10.1016/j.ijmm.2007.01.003.
42. Silveira F, Maluta RP, Tiba MR, de Paiva JB, Guastalli EA, da Silveira WD. Comparison between avian pathogenic (APEC) and avian faecal (AFEC) *Escherichia coli* isolated from different regions in Brazil. Vet J. 2016;217:65–7. https://doi.org/10.1016/j.tvjl.2016.06.007.

Publisher's Note