



ΙΟΝΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ

Σχολή Θετικών Επιστημών και Τεχνολογίας  
Βιοπληροφορική και Νευροπληροφορική

Διπλωματική Εργασία

Θέμα Διπλωματικής Εργασίας: Συγκριτική γονιδιωματική και μελέτη  
συσχέτισης σε όλο το γονιδίωμα σε παθογόνα και μη παθογόνα  
στελέχη *Escherichia coli* απομονωμένων από πτηνά

Ηλίας Αποστολάκος

Επιβλέπων καθηγητής: Ηλίας Λαγκουβάρδος



Θεσσαλονίκη, Ιανουάριος 2023

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή («συγγραφέας/δημιουργός») που την εκπόνησε. Στο πλαίσιο της δημόσιας ανοικτής πρόσβασης ο συγγραφέας/δημιουργός εκχωρεί στο ΕΑΠ, μη αποκλειστική άδεια χρήσης του δικαιώματος αναπαραγωγής, προσαρμογών, δημόσιου δανεισμού, παρουσίασης στο κοινό και ψηφιακής διαχύσεως τους διεθνούς, σε ηλεκτρονική μορφή και σε κάθε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος και για όλο το χρόνο των δικαιωμάτων πνευματικής ιδιοκτησίας. Η ανοικτή πρόσβαση στο πλήρες κείμενο για μελέτη και ανάγνωση δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητική ιδιοκτησία του συγγραφέα/δημιουργού ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, μετάδοση, διανομή, έκδοση, εκτέλεση, «μεταφόρτωση» (κατέβασμα), «ανάρτηση» (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναλλαγή του συγγραφέα/δημιουργού. Ο συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.



ΕΛΛΗΝΙΚΟ  
ΑΝΟΙΚΤΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ



ΙΟΝΙΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ

Συγκριτική γονιδιωματική και μελέτη συσχέτισης σε όλο το  
γονιδίωμα σε παθογόνα και μη παθογόνα στελέχη *Escherichia coli*  
απομονωμένων από πτηνά

Ηλίας Αποστολάκος

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής

Ηλίας Λαγκουβάρδος

Συν-Επιβλέπων Καθηγητής:

Αλκίνοος Αθανασίου

Θεσσαλονίκη, Ιανουάριος 2023

## **Ευχαριστίες**

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή Ηλία Λαγκουβάρδο για την καθοδήγηση και τις συμβουλές του κατά την διάρκεια της εκπόνησης της διπλωματικής μου εργασίας. Επίσης, ευχαριστώ θερμά τον συνάδελφο Nicola Palmieri, ερευνητή στο πανεπιστήμιο της Βιέννης, για την συνδρομή και πολύτιμη βοήθειά του, ιδιαίτερα στην συλλογή των πρωτογενών δεδομένων και στην συζήτηση/επεξήγηση των αποτελεσμάτων.

## Περίληψη

Τα παθογόνα στελέχη *Escherichia coli* των πτηνών (APEC) είναι ο αιτιολογικός παράγοντας της κολιβακίλλωσης, μιας ασθένειας με σημαντικές οικονομικές απώλειες για τα οικόσιτα πουλερικά. Τα μη-παθογόνα στελέχη *E. coli* των πτηνών (AFEC) κατοικούν συμβιωτικά στο έντερο των πτηνών. Οι υποκείμενοι μηχανισμοί παθογένειας των APEC και η διάκρισή τους από τα AFEC παραμένουν σε μεγάλο βαθμό ασαφείς και κανένα σύνολο γονιδίων δεν έχει συσχετιστεί αποκλειστικά με τα APEC. Μια ολοκληρωμένη βάση δεδομένων υψηλής-ποιότητας γονιδιωμάτων στελεχών *E. coli* απομονωμένων από πτηνά, δημιουργήθηκε εξάγοντας δεδομένα από μελέτες που δημοσιεύθηκαν την περίοδο 2000-2022. Τα συλλεχθέντα γονιδιώματα αναλύθηκαν μέσω μιας μελέτης συσχέτισης σε επίπεδο μικροβιακού γονιδιώματος (mGWAS), για τον εντοπισμό γονιδίων που σχετίζονται σημαντικά με τον παθογόνο φαινότυπο των στελεχών *E. coli* από πτηνά. Στη συνέχεια, το σύνολο των σημαντικών γονιδίων χαρακτηρίστηκε σε ένα δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης με τη βάση δεδομένων STRING για τη διερεύνηση των βιολογικών διεργασιών που εμπλέκονται στην κολιβακίλλωση. Η ανάλυση mGWAS εντόπισε 21 γονίδια που σχετίζονται σημαντικά με το φαινότυπο APEC. Τα αποτελέσματα υποδεικνύουν ότι η αλληλεπίδραση μεταξύ των μεταβολικών οδών που σχετίζονται με την πρόσληψη θρεπτικών συστατικών από τα κύτταρα του ξενιστή καθώς και η ρύθμιση της άμυνας στην ανοσολογική απόκριση του ξενιστή είναι απαραίτητη για την εξέλιξη της νόσου της κολιβακίλλωσης. Επιπρόσθετα, εντοπίστηκαν σημαντικά γονίδια που πιθανώς συνεπιλέγονται με άλλους φαινοτύπους *E. coli* (π.χ. μικροβιακή αντοχή). Επιπλέον, αναπτύχθηκε ένας ταξινομητής δένδρων, ο οποίος χρησιμοποιώντας ένα σύνολο 12 γονιδίων και ένα σύνολο κανόνων απόφασης, μπορεί να διακρίνει τα γονίδια APEC από τα AFEC με ακρίβεια 79.5%. Συμπερασματικά, αυτό το έργο παρείχε πληροφορίες για τις βιολογικές διεργασίες που εμπλέκονται στην κολιβακίλλωση. Ταυτόχρονα, η ολοκληρωμένη βάση δεδομένων των γονιδιωμάτων *E. coli* από πτηνά αντιπροσωπεύει μια πολύτιμη πηγή για μελλοντικές μελέτες στο συγκεκριμένο πρόβλημα.

**Λέξεις-Κλειδιά:** μελέτη συσχέτισης σε επίπεδο μικροβιακού γονιδιώματος, Εσερίχια κόλι, πτηνά, αλληλουχία ολόκληρου του γονιδιώματος

## Abstract

Avian pathogenic *Escherichia coli* (APEC) is the causative agent of colibacillosis, a disease with significant economic losses for domestic poultry. Avian fecal *E. coli* (AFEC) are the commensal counterparts of APEC, which symbiotically inhabit the gut of birds. The underlying pathogenicity mechanisms of APEC and their distinction from AFEC remain largely elusive and no set of genes has been exclusively associated with APEC. A comprehensive database of high-quality avian *E. coli* genomes was built by data-extraction from studies published in the period 2000-2022. The collected genomes were analyzed through a microbial genome-wide association study (mGWAS) framework, to identify genes significantly associated with the pathogenic phenotype of avian *E. coli*. Then, the set of significant genes was characterized in a protein-protein interaction network with the STRING database, to explore the biological processes involved in colibacillosis. Overall, mGWAS analysis identified 21 genes significantly associated with the APEC phenotype. Results suggested that the interplay between pathways associated with the scavenging of nutrients from the host's cells as well as regulation of defense from the host's immune response are essential for the establishment of colibacillosis. Additionally, significant genes that are likely co-selected with other *E. coli* phenotypes (e.g., antimicrobial resistance) were identified. Moreover, a decision-tree classifier was developed, which by using a set of 12 genes and a set of decision rules, it can distinguish APEC genomes from their commensal counterparts with an accuracy of 79.5%. In conclusion, this project provided insights for the biological processes involved in colibacillosis. At the same time, the comprehensive database of avian *E. coli* genomes represents a valuable resource for future studies on this topic.

**Keywords:** Avian pathogenic *E. coli*, APEC, whole-genome sequencing, WGS, microbial genome-wide association study, mGWAS, poultry, avian.

## Πίνακας περιεχομένων

Ευχαριστίες .....	4
Περίληψη .....	i
Abstract .....	ii
Πίνακας περιεχομένων.....	iii
Πίνακας Εικόνων .....	v
Λίστα Πινάκων .....	vi
Κατάλογος συντομογραφιών .....	vii
1. Εισαγωγή .....	1
1.1 Παγκόσμια σημασία της παραγωγής πουλερικών ως οικονομική πηγή πρωτεΐνης .....	1
1.2 <i>Escherichia coli</i> ως συμβιωτικός μικροοργανισμός και ευκαιριακό παθογόνο .....	2
1.3 Παθογόνα στελέχη <i>E. coli</i> των πτηνών (APEC).....	3
1.4 Μελέτη συσχέτισης μικροβιακού γονιδιώματος.....	5
1.5 Σκοπός της διατριβής.....	6
2. Μέθοδοι .....	7
2.1 Δημιουργία της βάσης δεδομένων γονιδιωμάτων στελεχών <i>E. coli</i> από πτηνά.....	7
2.2 Μελέτη συσχέτισης σε επίπεδο μικροβιακού γονιδιώματος (mGWAS).....	9
2.3 Δημιουργία ενός ταξινομητή δέντρου αποφάσεων (decision tree classifier) για τη διάκριση των APEC από τα AFEC .....	10
3. Αποτελέσματα.....	12
3.1 Η βάση δεδομένων γονιδιώματος <i>E. coli</i> των πτηνών.....	12
3.2 Περιγραφική στατιστική της βάσης δεδομένων .....	15
3.3 Αποτελέσματα της ανάλυσης mGWAS και χαρακτηρισμός γονιδίων που σχετίζονται σημαντικά με τον φαινότυπο APEC .....	17
3.4 Ταξινομητής δέντρου αποφάσεων για τη διάκριση μεταξύ APEC και AFEC .....	21
4. Συζήτηση.....	23

Παραρτήματα.....	29
Παράρτημα Α: Κώδικας .....	29
Παράρτημα Β: Συμπληρωματικός Πίνακας 1.....	33
Βιβλιογραφικές αναφορές.....	34

## Πίνακας Εικόνων

Εικόνα 1. Σχηματική αναπαράσταση του πλαισίου ανάλυσης που χρησιμοποιήθηκε για τον εντοπισμό σημαντικών γονιδίων που σχετίζονται με την παθογένεια του <i>E. coli</i> των πτηνών. Λεπτομέρειες παρέχονται στο κεφάλαιο 2, Μέθοδοι. ....	13
Εικόνα 2. Φυλογενετικό δέντρο όλων των γονιδιωμάτων APEC και AFEC ως κυκλικό κλαδόγραμμα. Η επισήμανση του φαινοτύπου και της φυλογενετικής ομάδας των απομονώσεων απεικονίζεται σύμφωνα με το υπόμνημα. ....	16
Εικόνα 3. Πλαίσιο A: Κατανομή φυλογενετικών ομάδων σε παθογόνα (APEC) και μη παθογόνα (AFEC) γονιδιώματα <i>E. coli</i> πτηνών. Πλαίσιο B: Τα 10 πιο διαδεδομένα <i>E. coli</i> MLST με τουλάχιστον 10 γονιδιώματα, που βρίσκονται σε γονιδιώματα AFEC (επάνω πλαίσιο) και APEC (κάτω πλαίσιο). ....	17
Εικόνα 4. Δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης των γονιδίων που σχετίζονται σημαντικά με τον φαινότυπο APEC, που δημιουργήθηκε μέσω της διαδικτυακής πλατφόρμας STRING (Szkarczyk et al., 2021). Τα σημαντικά γονίδια σημειώνονται με μαύρα ορθογώνια. Το μεγάλο δίκτυο στο κέντρο του σχήματος υπολογίστηκε με το στέλεχος <i>E. coli</i> K12 MG1655. Τα γονίδια χρωματίζονται ανάλογα με την ομάδα (cluster) που ανήκουν (ομάδα 1 – 7). Τα τρία υπο-δίκτυα που βρέθηκαν στα πάνελ A, B και Γ υπολογίστηκαν με το στέλεχος <i>E. coli</i> CFT073. ....	19
Εικόνα 5. Γονιδιωματική θέση των τριών υποθετικών γονιδίων τρανσποζάσης (group_2364, group_180, group_6516) και μίας υποθετικής πρωτεΐνης (group_6516) στο γονιδίωμα αναφοράς <i>E. coli</i> APEC O1. ....	21
Εικόνα 6. Ταξινομητής δέντρου απόφασης για την διάκριση APEC και AFEC. Πλαίσιο A: Σε συνδυασμό με ένα σύνολο κανόνων (παρουσία/απουσία γονιδίων), χρησιμοποιούνται 12 γονίδια για την ταξινόμηση των γονιδιωμάτων ως APEC ή AFEC. Πλαίσιο B: Το True Positive Rate (TPR) απεικονίζεται σε σχέση με το False Positive Rate (FPR) για μια σειρά από χρωματικά κωδικοποιημένα κατώφλια απόφασης για τη δημιουργία της τυπικής χαρακτηριστικής καμπύλης λειτουργίας του δέκτη (καμπύλη ROC). Η βέλτιστη τιμή κατωφλιού ήταν 0.63. ....	22

## Λίστα Πινάκων

Πίνακας 1. Στοιχεία παγκόσμιας αγοράς κρέατος σύμφωνα με τον FAO. Πίνακας προσαρμοσμένος από (OECD-FAO, 2017).....	2
Πίνακας 2. Κατάλογος μελετών που περιλαμβάνεται στη βάση δεδομένων γονιδιωμάτων <i>E. coli</i> πτηνών. Αναφέρεται επίσης ο αριθμός των γονιδιωμάτων και του φαινοτύπου(ων) που αναφέρθηκαν από κάθε μελέτη.....	13
Πίνακας 3. Τα γονίδια που βρέθηκαν να σχετίζονται σημαντικά με τον φαινότυπο APEC με βάση την ανάλυση mGWAS.....	18

## Κατάλογος συντομογραφιών

AFEC: Μη παθογόνα στελέχη *E. coli* από πτηνά

APEC: Παθογόνα στελέχη *E. coli* από πτηνά

MLST: Multi-locus sequence type

ΓΛΔ: Γονίδια λοιμογόνου δράσης

## 1. Εισαγωγή

### 1.1 Παγκόσμια σημασία της παραγωγής πουλερικών ως οικονομική πηγή πρωτεΐνης

Τα πουλερικά είναι είδη πτηνών που έχουν εξημερωθεί και εκτρέφονται για να παράγουν κυρίως αυγά και κρέας αλλά και φτερά. Το κρέας και τα αυγά που προέρχονται από πουλερικά είναι τα ζωικά προϊόντα με τη μεγαλύτερη κατανάλωση παγκοσμίως και είναι ιδιαίτερα δημοφιλή ακόμη και μεταξύ διαφορετικών πολιτισμών και θρησκειών. Πολλά είδη πτηνών αναφέρονται με τον όρο «πουλερικά», αλλά τα πιο κοινά είναι τα κοτόπουλα (όρνιθες), οι γαλοπούλες (ινδόρνιθες), και οι πάπιες. Σήμερα, πάνω από 26 δισεκατομμύρια κοτόπουλα εκτρέφονται παγκοσμίως και τα είδη πουλερικών αντιπροσωπεύουν περίπου το 75% της συνολικής βιομάζας των πτηνών στη γη (Mageiros et al., 2021). Η αύξηση του πληθυσμού και η αύξηση του μέσου εισοδήματος στις αναπτυσσόμενες χώρες, πέρα από τις σημαντικές βελτιώσεις στις κτηνοτροφικές πρακτικές, έχουν οδηγήσει σε αυξημένη κατανάλωση προϊόντων πουλερικών τα τελευταία 50 χρόνια, με την παγκόσμια, κατά κεφαλήν κατανάλωση αυγών να διπλασιάζεται και αυτή του κρέατος πουλερικών να αυξάνεται πάνω από πέντε φορές (OECD-FAO, 2017). Σύμφωνα με στοιχεία του FAO (OECD-FAO, 2017), το κρέας πουλερικών αντιπροσωπεύει το ~40% της παγκόσμιας παραγωγής κρέατος (Πίνακας 1), καθιστώντας το την κυρίαρχη πηγή ζωικής πρωτεΐνης παγκοσμίως, ακολουθούμενο από το χοιρινό κρέας. Μεταξύ των ειδών πουλερικών, το κρέας κοτόπουλου είναι ιδιαίτερα κυρίαρχο καθώς είναι γενικά φθινό και έχει χαμηλή περιεκτικότητα σε λιπαρά. Αυτή η χαμηλή τιμή συνδέεται με τον σύντομο κύκλο παραγωγής κοτόπουλων κρεοπαραγωγής που σε συνδυασμό με τις κάθετα ολοκληρωμένες γεωργικές επιχειρήσεις (δηλαδή που κατέχουν τα εκκολαπτήρια, τις εκτροφές πάχυνσης καθώς και τα σφαγεία) μπορούν να οδηγήσουν σε οικονομίες κλίμακας (Hafez & Attia, 2020). Ωστόσο, αυτά τα συστήματα εντατικής γεωργικής παραγωγής δημιουργούν προκλήσεις για την υγεία και την καλή διαβίωση των ζώων λόγω της ταχείας διάδοσης ασθενειών των πουλερικών, όπως η νόσος του Newcastle, καθώς και της εμφάνισης ζωνοόσων που προκαλούν ανησυχία για τη δημόσια υγεία, όπως π.χ. *Salmonella* spp., *Campylobacter* spp., *Escherichia coli* και ο εξαιρετικά παθογόνος ιός γρίπης των πτηνών (HPAI) (EFSA/ECDC, 2019; OECD-FAO, 2017). Προκειμένου να περιοριστούν αυτές οι ασθένειες και να διαφυλαχθεί τόσο η υγεία των ζώων όσο και η δημόσια υγεία, μεγάλες ποσότητες αντιβιοτικών χορηγούνται στα συστήματα παραγωγής πουλερικών, με αποτέλεσμα αυξημένα επίπεδα μικροβιακής αντοχής (EFSA/ECDC, 2016, 2019).

Πίνακας 1. Στοιχεία παγκόσμιας αγοράς κρέατος σύμφωνα με τον FAO. Πίνακας προσαρμοσμένος από (OECD-FAO, 2017)

	2019	2020	2021
	Εκατομμύρια τόνοι (ισοδύναμο βάρους σφαγίου)		
Βόειο κρέας	72.4	71.6	72.4
<b>Κρέας πουλερικών</b>	<b>131.6</b>	<b>133.4</b>	<b>135.2</b>
Χοιρινό κρέας	110.1	109.8	114.4
Προβειο κρέας	16.2	16.2	16.5
<i>Συνολική παραγωγή</i>	<i>337.2</i>	<i>338.1</i>	<i>345.6</i>

## 1.2 *Escherichia coli* ως συμβιωτικός μικροοργανισμός και ευκαιριακό παθογόνο

Το *E. coli* είναι ένα Gram-αρνητικό, σε σχήμα ράβδου, κολοβακτηρίδιο του γένους *Escherichia*, το οποίο ως προαιρετικός αναερόβιος μικροοργανισμός κατοικεί στο κατώτερο γαστρεντερικό σωλήνα των θηλαστικών και των πτηνών λίγο μετά τη γέννησή τους. Το *E. coli* δρα κυρίως ως συμβιωτικός μικροοργανισμός, συμμετέχοντας στη βιοσύνθεση βιταμινών (Sarowska et al., 2019). Ο Γερμανός γιατρός Theodor Escherich ήταν ο πρώτος που περιέγραψε το *E. coli* ως *Bacterium coli commune*.

Ως ποικιλόμορφος μικροοργανισμός που χαρακτηρίζεται από γονιδιωματική πλαστικότητα, το *E. coli* μπορεί να εμπλουτίσει το βοηθητικό του γονιδίωμα (accessory genome) με γονίδια λοιμογόνου δράσης (ΓΛΔ) που επιτρέπουν την προσαρμογή του σε περιβάλλοντα όπου υπάρχουν δυσμενείς συνθήκες και οι πηγές θρεπτικών συστατικών είναι σπάνιες. Αυτό το χαρακτηριστικό επιτρέπει στο *E. coli* να αποικίσει διάφορους ιστούς ξενιστή, οδηγώντας περιστασιακά σε δυσμενείς επιπτώσεις στην υγεία των ξενιστών τους (Denamur et al., 2021). Τα στελέχη που προκαλούν εντερικές ή εξωεντερικές ασθένειες ονομάζονται παθογόνα *E. coli*. Τα παθογόνα *E. coli* χωρίζονται σε δύο κύριες ομάδες ανάλογα με τα όργανα που αποικίζουν και στα οποία ασκούν τις λοιμογόνες ιδιότητές τους. Τα εντερικά παθογόνα *E. coli* (Intestinal Pathogenic *E. coli*, InPEC) και τα εξω-εντερικά παθογόνα *E. coli* (Extra-Intestinal Pathogenic *E. coli*, ExPEC) (Kaper et al., 2004). Τα πιο διαδεδομένα στελέχη *E. coli* που προκαλούν εντερικές ασθένειες είναι το εντεροπαθογόνο *E. coli* (EPEC), το εντεροαιμορραγικό *E. coli* (EHEC), το εντεροτοξιγονικό *E. coli* (ETEC), το εντεροσυσσωματικό *E. coli* (EAEC), το εντεροδηθητικό *E. coli* (EIEC) και το διάχυτα προσκολλημένο *E. coli* (DAEC) (Kaper et al., 2004). Ασθένειες όπως λοιμώξεις του ουροποιητικού συστήματος, σήψη και μηνιγγίτιδα

προκαλούνται από ExPEC (Singer, 2015). Το παθογόνο *E. coli* των πτηνών (APEC), μια υποομάδα του ExPEC, μολύνει τα πτηνά και προκαλεί την κολιβακίλλωση, μια ασθένεια η οποία είναι η κύρια αιτία σοβαρών οικονομικών απωλειών για την πτηνοτροφική βιομηχανία (Poulsen et al., 2020).

Προτού οι προηγμένες μοριακές μέθοδοι και τα δεδομένα γονιδιοματικής αλληλουχίας (NGS data) διαδοθούν σαν διαγνωστικά εργαλεία, η κύρια προσέγγιση για τη διάκριση των στελεχών *E. coli* ήταν ο ορότυπος του σωματικού (O) αντιγόνου και του μαστιγιακού (H) αντιγόνου και λιγότερο συχνά ο ορότυπος του αντιγόνου της κάψας (K). Οι ορότυποι O104:H4 και O157:H7 είναι διαβόητοι για την πρόκληση περιστατικών τροφικής δηλητηρίασης, ενώ ο ορότυπος O78:K80 συνήθως σχετίζεται με το APEC (Allocati et al., 2013). Ενώ ο ορότυπος εξακολουθεί να χρησιμοποιείται σήμερα για τον εντοπισμό κύριων παθογόνων, οι μοριακές μέθοδοι επικρατούν για τον προσδιορισμό του γονότυπου (genotyping) στελεχών *E. coli*, προσφέροντας βαθύτερη φυλογενετική ταξινόμηση και χαρακτηρισμό υψηλής ανάλυσης των ΓΔ, των καθοριστικών παραγόντων μικροβιακής αντοχής και των κινητών γενετικών στοιχείων (πλασμίδια και αλληλουχίες εισαγωγής). (Abram et al., 2019). Συχνά χρησιμοποιούμενες μέθοδοι προσδιορισμού γονότυπου είναι η αναγνώριση των φυλογενετικών ομάδων (phylogroups) *E. coli*, ο τύπος αλληλουχίας πολλαπλών θέσεων (MLST) και η αλληλουχία ολόκληρου του γονιδιώματος (WGS) των στελεχών *E. coli* (Clermont et al., 2013; Wirth et al., 2006).

### 1.3 Παθογόνα στελέχη *E. coli* των πτηνών (APEC)

Τα στελέχη *E. coli* που απομονώνονται από τα όργανα των νοσούντων πτηνών ονομάζονται παθογόνα *E. coli* των πτηνών (APEC). Τα APEC είναι ο αιτιολογικός παράγοντας της κολιβακίλλωσης, μιας ασθένειας με σημαντικές οικονομικές απώλειες για την πτηνοτροφία. Το APEC μπορεί να λειτουργήσει ως πρωτογενές παθογόνο (η κύρια αιτία της νόσου) ή δευτερογενές παθογόνο, το οποίο προκαλεί ασθένεια όταν το ανοσοποιητικό σύστημα του ξενιστή τίθεται σε κίνδυνο από άλλο παράγοντα ασθένειας ή για άλλους λόγους (π.χ. στρες). Το APEC μπορεί να προκαλέσει εντοπισμένες λοιμώξεις σε συγκεκριμένα όργανα (π.χ. περικαρδίτιδα, ηπατίτιδα) ή να οδηγήσει σε συστηματική κολιβακίλλωση (λοίμωξη της κυκλοφορίας του αίματος), η οποία συνήθως οδηγεί στο θάνατο των προσβεβλημένων πτηνών (Nolan et al., 2019). Η ασθένεια επηρεάζει όλα τα στάδια παραγωγής της πτηνοτροφικής βιομηχανίας (δηλαδή, εκτροφεία αναπαραγωγής, εκκολαπτήριο, εκτροφεία πάχυνσης και

σφαγεία). Στα κοτόπουλα κρεοπαραγωγής, η ομφαλίτιδα και η σηψαιμία είναι συχνές. Η λοίμωξη της αναπνευστικής οδού (πνευμονία) και η περικαρδίτιδα είναι πολύ συχνές σε όλα τα στάδια παραγωγής (Kaper et al., 2004; Nolan et al., 2019). Στις εκτροφές αναπαραγωγής και στις αυγοπαραγωγικές εκτροφές, η κολιβακίλλωση συνήθως εκδηλώνεται ως σαλπινγίτιδα ή/και περιτονίτιδα που οδηγεί σε μείωση της παραγωγής αυγών και σε αυξημένη θνησιμότητα, το κόστος της οποίας μπορεί να φτάσει έως και 1,90 ευρώ ανά στεγασμένη όρνιθα (Landman et al., 2013; Landman & van Eck, 2015). Η απώλεια των σφαγίων πουλερικών στο σφαγείο λόγω αλλοιώσεων στο δέρμα ή συστηματικών λοιμώξεων οργάνων, έχει ως αποτέλεσμα μέχρι και 1,5% απώλειες κρέατος πουλερικών (de Oliveira et al., 2020). Αξίζει να σημειωθεί ότι εκτός από σημαντικό παθογόνο για τα πουλερικά, το APEC θεωρείται πιθανό ζωνοσογόνο που μπορεί να προκαλέσει ασθένεια στον άνθρωπο μέσω του μολυσμένου κρέατος πουλερικών (Manges, 2016). Πράγματι, τα απομονωμένα στελέχη APEC παρουσιάζουν ομοιότητα στο περιεχόμενο γονιδίων ΓΛΔ και μικροβιακής αντοχής με τα ανθρώπινα στελέχη ExPEC, και επίσης ανήκουν σε στενά συγγενείς φυλογενετικές ομάδες *E. Coli* (Manges, 2016; Mellata, 2013a).

Υπάρχουν τέσσερις κύριοι μηχανισμοί μέσω των οποίων το APEC θεωρείται ότι αναδύεται και μολύνει τα πτηνά-ξενιστές: α) οριζόντια μεταφορά γονιδίων (π.χ. ανταλλαγή πλασμιδίων) ΓΛΔ σε μη παθογόνα στελέχη *E. coli* του εντέρου των πτηνών (AFEC), καθιστώντας τα παθογόνα, β) επιλογή εξαιρετικά λοιμοδών στελεχών APEC από τον ξενιστή και γ) η εξάπλωση κλώνων APEC υψηλής λοιμογόνου δράσης που φέρουν συγκεκριμένα ΓΛΔ, σε εξωεντερικά όργανα (Mageiros et al., 2021). Το APEC που εμπλέκεται σε περιστατικά κολιβακίλλωσης συνήθως ανήκει στους ορότυπους O78, O1 και O2, στις φυλογενετικές ομάδες *E. coli* B2 και D και στους τύπους αλληλουχίας πολλαπλών τόπων (STs) ST23, ST117 και ST95 (Mellata, 2013b). Ωστόσο, τα APEC που απομονώνονται από κλινικά περιστατικά εμφανίζουν εξαιρετική γονοτυπική ποικιλομορφία μεταξύ χωρών, περιοχών καθώς και εντός των ίδιων σμηνών, γεγονός που δημιουργεί δυσκολίες για την πρόληψη, τη διάγνωση και τη θεραπεία της νόσου (Ewers et al., 2009; Nolan et al., 2019). Αν και το APEC είναι ένα από τα πιο μελετημένα παθογόνα πουλερικών, οι υποκείμενοι μηχανισμοί παθογονικότητας δεν είναι ακόμη πλήρως κατανοητοί. Έχουν περιγραφεί αρκετά ΓΛΔ που απαιτούνται για τον αποικισμό, την επιβίωση και την εξάπλωση του APEC στον ξενιστή (Dziva & Stevens, 2008; Nolan et al., 2019). Ωστόσο, η παρουσία συγκεκριμένων ΓΛΔ δεν μπορεί να διαφοροποιήσει αναμφισβήτητα τα APEC από τα μη παθογόνα (commensal) στελέχη του εντέρου (AFEC).

Για παράδειγμα, σε μια μελέτη (Kemmett et al., 2014), διαπιστώθηκε ότι ο γαστρεντερικός σωλήνας των κοτόπουλων κρεοπαραγωγής είναι εμπλουτισμένος με στελέχη AFEC που διαθέτουν πολλά από τα ΓΛΔ που συνήθως βρίσκονται στα APEC. Από την άλλη πλευρά, απομονώσεις APEC που ανακτήθηκαν από κλινικά περιστατικά κολιβακίλλωσης έχει βρεθεί ότι ανήκουν σε φυλογενετικές ομάδες *E. coli* (π.χ. A, E, B1) που συνήθως συνδέονται με μη-παθογόνα στελέχη *E. coli*. Αυτά τα ευρήματα υπογραμμίζουν το γεγονός ότι οι μηχανισμοί παθογένειας αυτού του σημαντικού παθογόνου παραμένουν ακόμη άγνωστοι (Collingwood et al., 2014). Για παράδειγμα, οι βακτηριακές κροσσικές δομές (bacterial fimbrial structures) όπως οι κροσσοί τύπου 1 (type 1 fimbriae), καθώς και οι προσκολλητίνες (adhesins), είναι απαραίτητες για την προσκόλληση και τον αποικισμό των ιστών του ξενιστή από το APEC (Antão et al., 2009). Ωστόσο, αρκετές έρευνες ανέφεραν ανοδική ή καθοδική ρύθμιση (up/down regulation) πολλαπλών γονιδίων που απαιτούνται για την προσαρμογή και το μεταβολισμό του APEC στους ιστούς ξενιστή, αλλά δεν εντοπίστηκαν υπερεκφρασμένα fimbrial γονίδια (J. Gao et al., 2018; Kathayat et al., 2021). Επιπλέον, μελέτες που διερευνούν την έκφραση γονιδίων APEC χρησιμοποιώντας διαφορετικές τεχνολογίες [π.χ. signature-tagged mutagenesis (STM), recombination based *in vivo* expression technology (RIVET)] ανέφεραν αντικρουόμενα ευρήματα (Antão et al., 2009; Dozois et al., 2003; Tuntufye et al., 2012). Ως εκ τούτου, πρόσφατες μελέτες προτείνουν ότι η παθογένεια του APEC είναι εξαιρετικά πολύπλοκη και ταυτόχρονα συντονίζεται από πλασμιδιακά και χρωμοσωμικά ΓΛΔ που βρίσκονται σε συγκεκριμένες φυλογενετικές ομάδες *E. coli*, τα οποία μπορούν να ανταλλάσσονται μεταξύ πληθυσμών *E. coli* με ανταλλαγή πλασμιδίων ή μετατόπιση στοιχείων αλληλουχίας εισαγωγής (insertion sequence (IS) elements) (Johnson et al., 2022; Mageiros et al., 2021).

#### 1.4 Μελέτη συσχέτισης μικροβιακού γονιδιώματος

Η μελέτη συσχέτισης σε επίπεδο μικροβιακού γονιδιώματος (mGWAS) είναι μια προσέγγιση για την κατανόηση του τρόπου με τον οποίο οι παραλλαγές στα μικροβιακά γονιδιώματα επηρεάζουν τους φαινοτύπους του ξενιστή ή των βακτηρίων (San et al., 2020). Οι αναλύσεις GWAS σε ανθρώπους, αναπτύχθηκαν από την υπόθεση common disease common variant (CDCV), η οποία υποδηλώνει ότι οι παραλλαγές με υψηλή συχνότητα αλλά χαμηλή επίδραση σχετίζονται με τον κίνδυνο ασθένειας (Power et al., 2017). Όπως και με το ανθρώπινο GWAS, ο σκοπός του mGWAS είναι να πραγματοποιήσει μια έρευνα για στατιστικά σημαντικές

συσχετίσεις μεταξύ του φαινοτύπου και του γονότυπου (SNPs και μοτίβα παρουσίας/απουσίας γονιδίων) του εν λόγω μικροοργανισμού. Μια σημαντική πρόκληση της προσέγγισης mGWAS είναι ότι για να εντοπιστούν πραγματικά σημαντικές συσχετίσεις, πρέπει κανείς να ελέγξει για την επιρροή συγχυτικών παραγόντων (confounding factors) λόγω της δομής του κλωνικού πληθυσμού (clonal population structure) και του ομόλογου ανασυνδυασμού (homologous recombination). Ο λόγος για αυτό είναι ότι ψευδείς συσχετίσεις μπορεί να προκύψουν από συγχυτικούς παράγοντες όπως η συσχέτιση μεταξύ της γενετικής συγγένειας και του φαινοτύπου των βακτηριακών στελεχών (δηλαδή διαφορές στις συχνότητες αλληλόμορφων μεταξύ ομάδων/υποπληθυσμών που δεν σχετίζονται με τον φαινότυπο) (Collins & Didelot, 2018; Power et al., 2017). Αυτή η διαδικασία διόρθωσης γίνεται αυτόματα στον αλγόριθμο treeWAS (βλ. κεφάλαιο 2.2, Μελέτη συσχέτισης σε επίπεδο μικροβιακού γονιδιώματος), παρέχοντας ένα ακριβές (robust) φυλογενετικό δέντρο που έχει διορθωθεί για συμβάντα ανασυνδυασμού στα γονιδιώματα (Collins & Didelot, 2018).

Η μέθοδος mGWAS είναι απαραίτητη για την ανάλυση μικροβιακών γονιδιωμάτων στο πλαίσιο της δημόσιας υγείας. Για παράδειγμα, έχει χρησιμοποιηθεί για τον εντοπισμό γονιδίων που σχετίζονται με τη μικροβιακή αντοχή, την παθογένεια του *Staphylococcus aureus* και του *Streptococcus pneumoniae* καθώς και τις αλληλεπιδράσεις ξενιστή-μικροβίου για το *Helicobacter pylori* (Power et al., 2017; San et al., 2020). Στο κεφάλαιο 4: Συζήτηση, γίνεται σύγκριση με αποτελέσματα mGWAS από άλλες μελέτες σε APEC και ExPEC.

## 1.5 Σκοπός της διατριβής

Από τα προαναφερθέντα είναι σαφές ότι οι υποκείμενοι μηχανισμοί της παθογονικότητας του APEC καθώς και η γενετική διάκριση μεταξύ AFEC και APEC παραμένει σε μεγάλο βαθμό ασαφής. Αρκετές μελέτες έχουν προσπαθήσει να ταξινομήσουν τα APEC και AFEC προσδιορίζοντας ομάδες γονιδίων που σχετίζονται σημαντικά με τους δύο φαινοτύπους [π.χ.(Ewers et al., 2005; Johnson et al., 2008, 2022)]. Μια μελέτη ορόσημο ήταν αυτή των Johnson et al. (2008), η οποία αναγνώρισε ένα σύνολο πέντε γονιδίων (*iroN*, *hlyF*, *iutA*, *iss* και *ompT*), που τοποθετούνταν σε ένα πλασμίδιο αντιγραφής ColV, ως ελάχιστους προγνωστικούς παράγοντες της παθογονικότητας APEC. Ωστόσο, μια πρόσφατη μελέτη αντέκρουσε αυτά τα αποτελέσματα βρίσκοντας αυτό το σύνολο γονιδίων σε αφθονία τόσο στα APEC όσο και στα AFEC (Mageiros et al., 2021). Επιπλέον, μια άλλη μεγάλης κλίμακας

μελέτη δεν βρήκε κάποιο συγκεκριμένο γονίδιο ή σύνολο γονιδίων που να σχετίζονται σημαντικά με τον φαινότυπο APEC (Johnson et al., 2022).

Η mGWAS είναι μια μέθοδος που μπορεί να βοηθήσει στην προσπάθεια αποσαφήνισης των γενετικών μηχανισμών που κρύβονται πίσω από την παθογένεια του APEC. Στη μελέτη των Mageiros et al. (2021), η mGWAS χρησιμοποιήθηκε για την αναγνώριση γενετικών δεικτών που διαφοροποιούν το APEC και το AFEC, χρησιμοποιώντας 550 γονιδιώματα. Οι συγγραφείς εντόπισαν γονίδια που εμπλέκονται στον μεταβολισμό, την μικροβιακή αντοχή, την απόκριση σε θερμικό σοκ και τη σύνθεση πολυσακχαριτών που σχετίζονται με το APEC. Ωστόσο, μια ανασκόπηση σχετικά με τις αναλυτικές προκλήσεις του mGWAS, τόνισε ότι η στατιστική ισχύς αυτής της μεθόδου εξαρτάται σε μεγάλο βαθμό από το μέγεθος του δείγματος, φτάνοντας σε plateau στα περίπου 3000 βακτηριακά δείγματα (γονιδιώματα).

Ως εκ τούτου, οι στόχοι αυτής της διατριβής ήταν 1) η διεξαγωγή μιας συστηματικής αναζήτησης για τον εντοπισμό και τη συλλογή όλων των διαθέσιμων γονιδιωμάτων *E. coli* των πτηνών (AFEC και APEC) προκειμένου να δημιουργηθεί μια ολοκληρωμένη βάση δεδομένων και 2) να αναλυθούν τα συλλεγμένα γονιδιώματα μέσω mGWAS σε συνδυασμό με ανάλυση δικτύου αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης, προκειμένου να εντοπιστούν οι γενετικοί μηχανισμοί που εμπλέκονται στην παθογένεια APEC. Ένας πρόσθετος παράπλευρος στόχος, ήταν μια προσπάθεια δημιουργίας ενός ταξινομητή (classifier) που θα διακρίνει με ακρίβεια τις απομονώσεις APEC και AFEC μέσω του γονιδιακού τους περιεχομένου.

## 2. Μέθοδοι

### 2.1 Δημιουργία της βάσης δεδομένων γονιδιωμάτων στελεχών *E. coli* από πτηνά

Για τον εντοπισμό μελετών που ανέφεραν αλληλουχίες ολόκληρου του γονιδιώματος APEC και/ή AFEC που απομονώθηκαν από κοινά είδη πουλερικών (κυρίως όρνια και ινδόρνια), πραγματοποιήθηκε μια συστηματική αναζήτηση βιβλιογραφίας. Οι αναζητήσεις πραγματοποιήθηκαν στα αποθετήρια PubMed και Web of Science (WoS) από τον Απρίλιο έως τον Μάιο του 2022. Χρησιμοποιήθηκε η ακόλουθη Boolean συμβολοσειρά αναζήτησης:

*(chicken\* OR poultry\* OR broiler\* OR turkey\* OR hen\* OR fowl\* OR geese OR goose OR duck\*) AND ("Escherichia coli" OR "E. coli" OR APEC OR "pathogenic E. coli" OR*

“pathogenic *Escherichia coli*” OR ExPEC OR colibacillosis OR colisepticaemia OR colisepticemia OR peritonitis OR salpingitis OR cellulitis) AND (genom\* OR wgs)

Εφαρμόστηκαν τα ακόλουθα κριτήρια αποκλεισμού: 1) μελέτες μη γραμμένες στα αγγλικά, 2) μελέτες που δημοσιεύθηκαν πριν από το έτος 2000, 3) μη επιτόπιες μελέτες (πειραματικές μελέτες) που αναφέρουν αποτελέσματα από μοντέλα μόλυνσης πτηνών (infection models). Ο επιθυμητός τύπος μελέτης ήταν μελέτες πεδίου παρατήρησης (observational studies) που ανέφεραν γονιδιώματα AFEC από τη γαστρεντερική οδό υγιών πτηνών ή/και APEC που απομονώθηκαν από περιστατικά κολιβακίλλωσης.

Οι μελέτες που πληρούσαν τα κριτήρια επιλογής λήφθηκαν από το PubMed και το WoS σε μορφή BibTeX και μεταφορτώθηκαν στο online πρόγραμμα συστηματικής ανασκόπησης Rayyan (<https://rayyan.ai/>) (Ouzzani et al., 2016). Οι διπλότυπες μελέτες αφαιρέθηκαν χρησιμοποιώντας την ενσωματωμένη λειτουργία του Rayyan. Στη συνέχεια, διαβάστηκαν οι τίτλοι και οι περιλήψεις των μη διπλότυπων μελετών και όσες δεν πληρούσαν τα κριτήρια αποκλείστηκαν. Στην επόμενη φάση αξιολογήθηκαν σε επίπεδο πλήρους κειμένου οι μελέτες που πληρούσαν τα κριτήρια και αυτές που παρέμεναν ασαφείς. Οι μελέτες που έχουν επιβεβαιωθεί ως επιλέξιμες παρουσιάζονται στον Πίνακα 2 (βλ. κεφάλαιο 3, Αποτελέσματα).

Για κάθε επιλέξιμη μελέτη, αναζητήθηκαν οι αντίστοιχοι αριθμοί προσχώρησης (accession numbers) των γονιδιωμάτων. Εάν τα accession numbers ήταν διαθέσιμα, οι συναρμολογήσεις γονιδιώματος (assemblies) λήφθηκαν από τη βάση δεδομένων NCBI (<https://www.ncbi.nlm.nih.gov/assembly>). Εάν τα accession numbers δεν ήταν διαθέσιμα, συλλέχτηκαν τα accession numbers των ακατέργαστων ακολουθιών (raw sequence reads) από τη βάση δεδομένων NCBI SRA (<https://www.ncbi.nlm.nih.gov/sra>). Στη συνέχεια, τα SRA accessions εισήχθησαν στην υπηρεσία Comprehensive Genome Analysis της βάσης δεδομένων PATRIC (Davis et al., 2019), χρησιμοποιώντας τις προεπιλεγμένες επιλογές assembly [το ελάχιστο μήκος contig ήταν 300 ζεύγη βάσεων (bp)]. Προκειμένου να αξιολογηθεί η ποιότητα όλων των assemblies, ο αλγόριθμος CheckM (Parks et al., 2015) χρησιμοποιήθηκε με τις προεπιλεγμένες παραμέτρους. Εφαρμόστηκαν κριτήρια ποιοτικού ελέγχου και γονιδιώματα *E. coli* με  $\leq 95\%$  πληρότητα γονιδιώματος και/ή με  $\geq 5\%$  μόλυνση γονιδιώματος ή/και με  $\geq 800$  contigs (υψηλά κατακερματισμένα γονιδιώματα) ή/και με στατιστική  $N50 < 20$  Kbp, αποκλείστηκαν από περαιτέρω ανάλυση. Επιπλέον, τα γονιδιώματα του *E. coli* επισημάνθηκαν (annotated) με τον αλγόριθμο PROKKA (Seemann, 2014), χρησιμοποιώντας τις προεπιλεγμένες παραμέτρους. Επίσης, ο αλγόριθμος ClermontTyping

(Beghain et al., 2018) χρησιμοποιήθηκε για να επισημανθεί κάθε γονιδίωμα με την φυλογενετική ομάδα *E. coli*. Τέλος, η βάση δεδομένων pubMLST (<https://pubmlst.org/>) χρησιμοποιήθηκε για την εκχώρηση MLST σε κάθε γονιδίωμα. Δημιουργήθηκε ένας πίνακας με όλα τα πεδία μεταδεδομένων, δηλαδή τον φαινότυπο APEC ή AFEC, τη φυλογενετική ομάδα, το MLST και την αντίστοιχη μελέτη (έτος δημοσίευσης και πρώτος συγγραφέας) (Παράρτημα Β: Συμπληρωματικός Πίνακας 1).

## 2.2 Μελέτη συσχέτισης σε επίπεδο μικροβιακού γονιδιώματος (mGWAS)

Ένα διάγραμμα ροής της ανάλυσης mGWAS παρουσιάζεται στο Σχήμα 1 (βλ. κεφάλαιο 3, Αποτελέσματα). Το πανγονιδίωμα (pangenome) των υψηλής ποιότητας γονιδιωμάτων υπολογίστηκε με τον αλγόριθμο Panaroo (Tonkin-Hill et al., 2020) (παράμετροι: `--clean-mode strict -a core --aligner mafft --core_threshold 0.95`), χρησιμοποιώντας ως είσοδο τα αρχεία `.gff` (επισημάνσεις γονιδιωμάτων) του PROKKA. Η έξοδος του Panaroo αποτελείται από ένα πίνακα παρουσίας/απουσίας γονιδίων που υπάρχουν σε όλα γονιδιώματα *E. coli*. Ένα επιπλέον αποτέλεσμα του προγράμματος είναι η ευθυγράμμιση του γονιδιώματος-πυρήνα (core-genome alignment), δηλαδή των γονιδίων που υπήρχαν σε  $\geq 95\%$  όλων των γονιδιωμάτων. Χρησιμοποιώντας το core-genome alignment, ένα φυλογενετικό δέντρο υπολογίστηκε με τον αλγόριθμο FastTree (Price et al., 2009) (παράμετροι: `-nt -nopr -cat 20 -nosupport -fastest`). Το παραγόμενο δέντρο διορθώθηκε για ομόλογο ανασυνδυασμό χρησιμοποιώντας τον αλγόριθμο ClonalFrameML (Didelot & Wilson, 2015). Πραγματοποιήθηκαν δύο ξεχωριστές αναλύσεις mGWAS για τον εντοπισμό α) γονιδίων και β) μονονουκλεοτιδικών πολυμορφισμών (SNPs) που σχετίζονται σημαντικά με τον παθογόνο φαινότυπο (APEC). Και οι δύο αναλύσεις mGWAS υλοποιήθηκαν με τον αλγόριθμο treeWAS (Collins & Didelot, 2018) στη γλώσσα προγραμματισμού R (R Core Team, 2020). Και στις δύο αναλύσεις, το φυλογενετικό δέντρο χρησιμοποιήθηκε για τη διόρθωση των συγχυτικών επιδράσεων της φυλογενετικής δομής του μικροβιακού πληθυσμού (δηλαδή, διορθώθηκαν σημαντικές συσχετίσεις που καθοδηγούνται από τη φυλογενετική σχέση του μικροβιακού πληθυσμού και όχι από γονίδια ή SNP που επηρεάζουν πραγματικά τον εν λόγω φαινότυπο) (San et al., 2020; Sul et al., 2018).

Για την mGWAS που βασίστηκε σε γονίδια, οι είσοδοι στο TreeWas (παράμετροι: `p-value = 0.05`) ήταν α) ο πίνακας παρουσίας/απουσίας γονιδίων, β) το φυλογενετικό δέντρο και γ) ο φαινότυπος κάθε απομόνωσης με τη μορφή δυαδικού διανύσματος (binary vector). [0 για μη παθογόνα (AFEC) και 1 για παθογόνα στελέχη (APEC)]. Το αποτέλεσμα του αλγορίθμου ήταν

μια λίστα με σημαντικά εμπλουτισμένα γονίδια στον φαινότυπο APEC. Αν και η βασική τιμή  $p$  ορίστηκε στο 0.05, ο αλγόριθμος προσαρμόζει αυτόματα τις τιμές  $p$  για διόρθωση πολλαπλών συγκρίσεων με βάση τη μέθοδο Bonferroni. Επιπλέον, για όλα τα σημαντικά γονίδια, υπολογίστηκε ο σχετικός επιπολασμός τους σε APEC και AFEC. Στις μεταγενέστερες αναλύσεις, χρησιμοποιήθηκαν γονίδια με επιπολασμό  $> 50\%$  σε APEC και αναλογία επιπολασμού  $APEC/AFEC > 1$  (Galardini et al., 2020).

Για το mGWAS που βασίζεται σε SNPs, τα γονιδιώματα του *E. coli* των πτηνών ευθυγραμμίστηκαν (aligned) με τον αλγόριθμο parsnp (Treangen et al., 2014) χρησιμοποιώντας τις προεπιλεγμένες παραμέτρους και το γονιδίωμα *E. coli* APEC O1 (NCBI accession: GCF\_902880315.1) ως αναφορά. Τα SNP υποβλήθηκαν σε επεξεργασία σύμφωνα με ένα προηγούμενος δημοσιευμένο πρωτόκολλο (Treangen et al., 2014) ως εξής; το αρχείο εξόδου .vcf με τα SNPs χρησιμοποιήθηκε ως είσοδος στον αλγόριθμο bcftools (παραμέτροι: view m2 -M2) (Treangen et al., 2014) για τον διαχωρισμό διαλληλικών (biallelic) και πολυαλληλικών (multiallelic) SNP. Τα biallelic SNP μετατράπηκαν σε δυαδικό διάνυσμα (1 για το αλληλόμορφο αναφοράς και 0 για το εναλλακτικό). Κάθε SNP ονομάστηκε με τη μορφή snp\_xxxx όπου xxxx είναι ένας ακέραιος αριθμός της γονιδιωματικής θέσης στην ευθυγράμμιση του γονιδιώματος (genome alignment). Τα multiallelic SNP αναπαραστάθηκαν ως δυαδικό διάνυσμα και ονομάστηκαν με τη μορφή snp\_xxxx\_y όπου xxxx είναι ένας ακέραιος αριθμός της γονιδιωματικής θέσης στο genome alignment και y είναι ένας προοδευτικός ακέραιος αριθμός που χρησιμοποιείται για τη δημιουργία ενός μοναδικού αναγνωριστικού. Και τα δύο σετ SNP συγχωνεύτηκαν σε ένα αρχείο και χρησιμοποιήθηκαν ως είσοδοι στο treeWAS (παραμέτροι: p-value = 0.05) μαζί με το δέντρο και το διάνυσμα φαινοτύπου (APEC = 1, AFEC = 0) που χρησιμοποιήθηκαν στην προηγούμενη ανάλυση, και ο αλγόριθμος απέδωσε τα στατιστικά σημαντικά SNP που σχετίζονται με τον φαινότυπο APEC.

### **2.3 Δημιουργία ενός ταξινομητή δέντρου αποφάσεων (decision tree classifier) για τη διάκριση των APEC από τα AFEC**

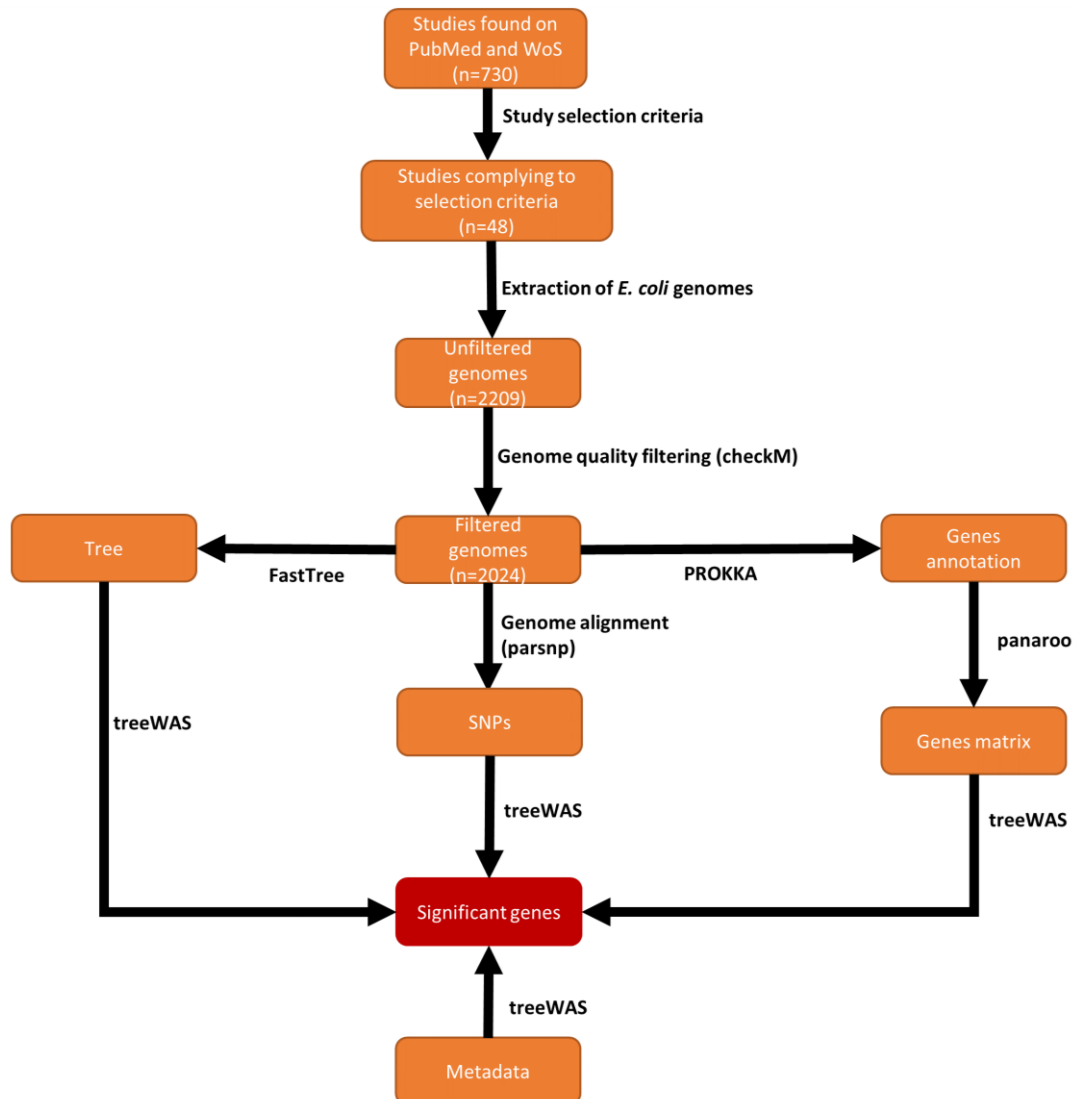
Το δυαδικό διάνυσμα φαινοτύπου των γονιδιωμάτων *E. coli* [APEC (1), AFEC (0)] και ο πίνακας παρουσίας/απουσίας γονιδίων χρησιμοποιήθηκαν ως είσοδοι για την εκπαίδευση ενός ταξινομητή δέντρου απόφασης. Τα δεδομένα ανακατεύτηκαν (shuffled) και έγινε ένας τυχαίος διαχωρισμός 80% - 20% για να δημιουργηθεί το σετ εκπαίδευσης (train) και δοκιμής (test),

αντίστοιχα. Δεδομένου ότι ο αριθμός των μεταβλητών (αριθμός γονιδίων) ήταν μεγάλος (29,262 γονίδια στο πανγονιδίωμα), ο αλγόριθμος επιλογής μεταβλητών (feature selection algorithm) Boruta (Treangen et al., 2014) στην γλώσσα R, χρησιμοποιήθηκε για τη διατήρηση μόνο των στατιστικά σημαντικών μεταβλητών (Παράρτημα Α: Κώδικας). Μετά από αυτή τη διαδικασία, παρέμειναν 729 σημαντικές μεταβλητές. Επιπλέον, γονίδια με μη τυπικά ονόματα (group\_xxx), που κωδικοποιούν υποθετικές πρωτεΐνες, αφαιρέθηκαν προκειμένου να διατηρηθούν τυπικά γονίδια που κωδικοποιούν πρωτεΐνες, τα οποία μπορούν να επισημανθούν και επομένως να χρησιμοποιηθούν με τον ταξινομητή σε ένα εξωτερικό σύνολο δεδομένων. Τελικά, χρησιμοποιήθηκαν 234 μοναδικά γονίδια για την εκπαίδευση του ταξινομητή. Το δέντρο αποφάσεων εκπαιδεύτηκε με την βιβλιοθήκη rpart της R (<https://cran.r-project.org/web/packages/rpart/>) χρησιμοποιώντας ως μεταβλητή στόχο τον φαινότυπο και τα επιλεγμένα γονίδια ως μεταβλητές πρόβλεψης. Χρησιμοποιώντας το σετ δοκιμών, υπολογίστηκε ένας πίνακας σύγχυσης (confusion matrix), συμπεριλαμβανομένων των αληθών θετικών (TP), αληθών αρνητικών (TN), ψευδώς θετικών (FP) και ψευδώς αρνητικών (FN). Το πακέτο ROCR R (<https://cran.r-project.org/web/packages/ROCR/>) χρησιμοποιήθηκε για τον υπολογισμό δεικτών απόδοσης (performance metrics):  $accuracy = (TP + TN) / (TP + TN + FP + FN)$ ,  $precision = TP / (TP + FP)$ ,  $recall = TP / (TP + FN)$ , καθώς και της χαρακτηριστικής καμπύλης λειτουργίας δέκτη [receiver operating characteristic curve (ROC) curve].

### 3. Αποτελέσματα

#### 3.1 Η βάση δεδομένων γονιδιώματος *E. coli* των πτηνών

Μια σχηματική αναπαράσταση των αναλύσεων, που απεικονίζει την επιλογή μελετών, την εξαγωγή δεδομένων, το φιλτράρισμα των γονιδιωμάτων και την ανάλυση mGWAS, απεικονίζεται στην Εικόνα 1. Η συστηματική βιβλιογραφική αναζήτηση εντόπισε 730 μελέτες που ήταν πιθανό να αναφέρουν γονιδιώματα AFEC ή/και APEC από στελέχη *E. coli* που απομονώθηκαν από είδη πουλερικών. Μετά την αφαίρεση των διπλότυπων μελετών στο Rayyan και τη επιθεώρηση του τίτλου και των περιλήψεων (φάση 1) και στη συνέχεια του πλήρους κειμένου των μελετών (φάση 2), παρέμειναν 48 μελέτες που πληρούσαν τα κριτήρια επιλογής. Αυτές οι μελέτες ανέφεραν συλλογικά 2209 WGS AFEC και APEC. Μετά τον αποκλεισμό γονιδιωμάτων χαμηλής ποιότητας (n=185), 2024 γονιδιώματα (1089 APEC και 935 AFEC) συγκεντρώθηκαν στην τελική βάση δεδομένων. Ο Πίνακας 2 συνοψίζει τις δημοσιεύσεις που περιλαμβάνονται στη βάση δεδομένων και τους αντίστοιχους αριθμούς γονιδιωμάτων από κάθε μελέτη.



Εικόνα 1. Σχηματική αναπαράσταση του πλαισίου ανάλυσης που χρησιμοποιήθηκε για τον εντοπισμό σημαντικών γονιδίων που σχετίζονται με την παθογένεια του *E. coli* των πτηνών. Λεπτομέρειες παρέχονται στο κεφάλαιο 2, Μέθοδοι.

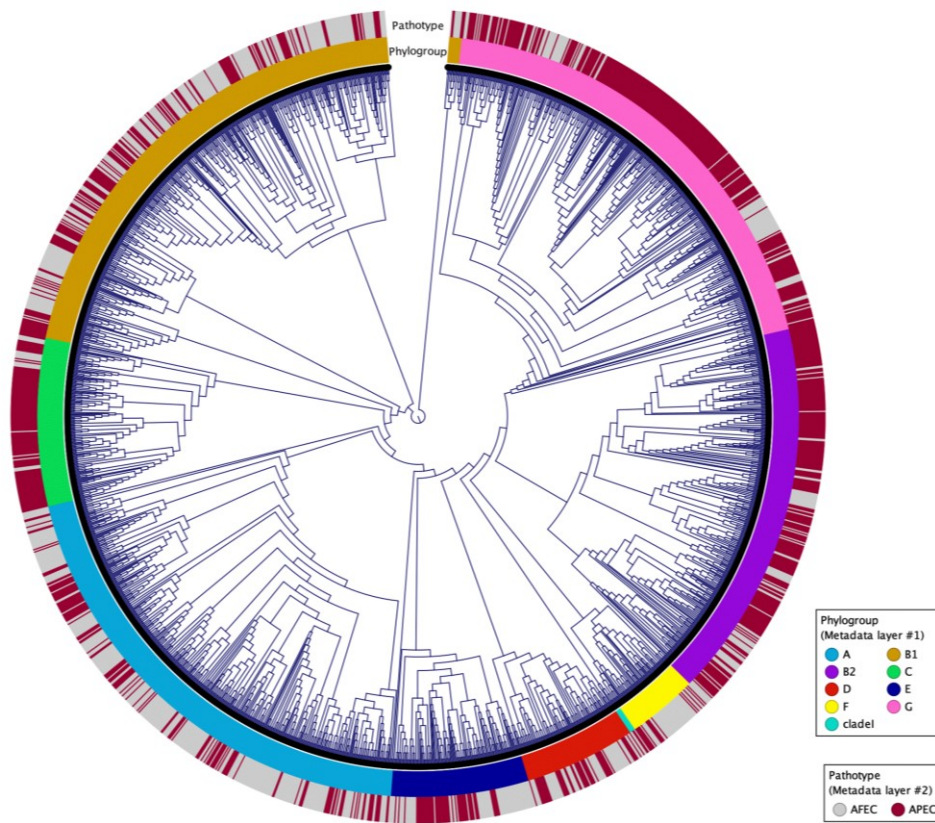
Πίνακας 2. Κατάλογος μελετών που περιλαμβάνεται στη βάση δεδομένων γονιδιωμάτων *E. coli* πτηνών. Αναφέρεται επίσης ο αριθμός των γονιδιωμάτων και του φαινοτύπου(ων) που αναφέρθηκαν από κάθε μελέτη.

Φαινότυποι <i>E. coli</i>	Αριθμός γονιδιωμάτων	Μελέτη
APEC	1	(Rojas et al., 2012)
APEC	2	(Dziva et al., 2013)
APEC	3	(Rojas et al., 2013)
AFEC	4	(de Been et al., 2014)
APEC	1	(Zhu Ge et al., 2014)
AFEC+APEC	2	(Giufre et al., 2015)
APEC	1	(Huja et al., 2015)
APEC	89	(Cordoni et al., 2016)
APEC	1	(Ewers et al., 2016)
APEC	1	(Maluta et al., 2016)

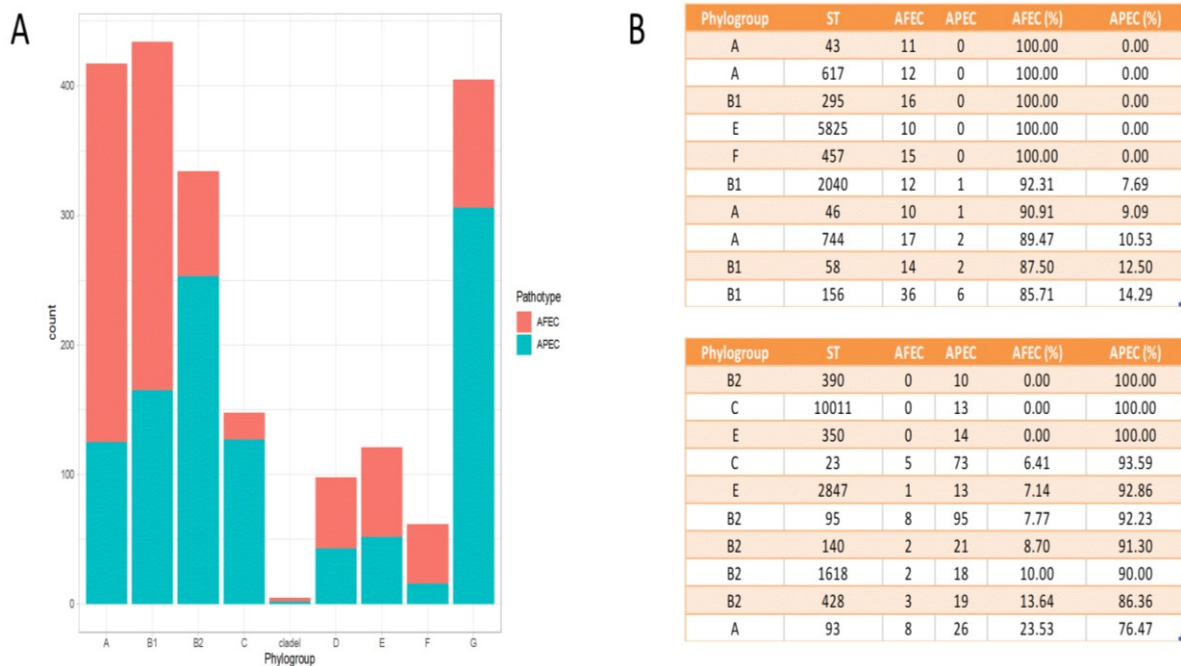
APEC	1	(Nicholson et al., 2016)
APEC	2	(Ronco et al., 2016)
APEC	1	(X. Wang et al., 2016)
APEC	1	(Jørgensen et al., 2017)
APEC	12	(Kólsut et al., 2017)
AFEC+APEC	106	(Ronco et al., 2017)
AFEC	99	(Y. Wang et al., 2017)
AFEC	28	(Alba et al., 2018)
APEC	10	(L. Chen et al., 2018)
AFEC	79	(Falgenhauer et al., 2019)
APEC	1	(Nielsen et al., 2018)
APEC	8	(Poulsen et al., 2018)
AFEC	7	(Roschanski et al., 2018)
AFEC	64	(Wu et al., 2018)
AFEC	1	(Abraham et al., 2019)
APEC	97	(Cummins et al., 2019)
AFEC	2	(J. Li et al., 2019)
AFEC	4	(Liu et al., 2019)
AFEC	94	(Maciuca et al., 2019)
AFEC	77	(Zajac et al., 2019)
APEC	10	(Zhuge, Jiang, et al., 2019)
AFEC	31	(Ahmed et al., 2020)
AFEC	60	(Apostolakos et al., 2020)
APEC	55	(Azam et al., 2020)
APEC	3	(Flament-Simon et al., 2020)
AFEC	90	(Kaspersen et al., 2020)
APEC	1	(T. Li et al., 2020)
AFEC	31	(Mo et al., 2020)
APEC	26	(Papouskova et al., 2020)
APEC	62	(Poulsen et al., 2020)
APEC	92	(Rafique et al., 2020)
APEC	23	(Apostolakos et al., 2021)
APEC	51	(X. Chen et al., 2021)
AFEC	45	(Ewers et al., 2021)
APEC	112	(Lozica et al., 2021)
AFEC+APEC	476	(Mageiros et al., 2021)
APEC	54	(Z. Wang et al., 2022)
APEC	3	(Yin et al., 2021)

### 3.2 Περιγραφική στατιστική της βάσης δεδομένων

Το φυλογενετικό δέντρο που παρουσιάζεται ως κυκλικό κλαδόγραμμα απεικονίζεται στην Εικόνα 2 και αναδεικνύει τη συσχέτιση των δύο φαινοτύπων (APEC και AFEC) με τις φυλογενετικές ομάδες *E. coli*. Η κατανομή των φυλογενετικών ομάδων σε κάθε φαινότυπο ήταν σημαντικά διαφορετική (δοκιμή Chi-Square –  $P = 4.43 \times 10^{-76}$ ) (Εικόνα 3A). Πιο αναλυτικά, οι φυλογενετικές ομάδες B2, C και G συσχετίστηκαν σημαντικά με τον φαινότυπο APEC (Chi-Square post-hoc test –  $PB_2 = 0$ ,  $PC = 0$  και  $PG = 0$ ) ενώ οι φυλογενετικές ομάδες A και F συσχετίστηκαν σημαντικά με τον φαινότυπο AFEC (Chi-Square post-hoc test –  $PA = 0$  και  $PF = 0.0001$ ). Επιπλέον, το μήκος του γονιδιώματος των γονιδιωμάτων APEC και AFEC ήταν σημαντικά διαφορετικό. Τα γονιδιώματα AFEC ήταν σημαντικά μεγαλύτερα από το APEC (διάμεσο μήκος APEC = 5,107 Mbp, διάμεσο μήκος AFEC = 5,142 Mbp). Ωστόσο, μετά τη διαίρεση των δεδομένων ανά φυλογενετική ομάδα, τα γονιδιώματα AFEC ήταν σημαντικά μεγαλύτερα μόνο για τις φυλογενετικές ομάδες B1, E και G (δοκιμή Mann-Whitney U –  $PB_1 = 1.05 \times 10^{-11}$ ,  $PE = 1.25 \times 10^{-5}$ ,  $PG = 3.79 \times 10^{-8}$ ). Αυτό το εύρημα υποδεικνύει ότι το μεγαλύτερο γονιδίωμα AFEC (ή μικρότερο APEC) δεν είναι γενικά αληθές και εξαρτάται από τη φυλογενετική ομάδα του *E. coli*. Το ίδιο μοτίβο παρατηρήθηκε για τον αριθμό των γονιδίων στα γονιδιώματα APEC και AFEC. Η κατανομή τους ήταν σημαντικά διαφορετική (δοκιμή Mann-Whitney U –  $P = 0.0108$ ), καθώς τα γονιδιώματα AFEC είχαν περισσότερα γονίδια από τα APEC (διάμεσος αριθμός γονιδίων APEC = 4888, διάμεσος αριθμός γονιδίων AFEC = 4916). Όπως και στην προηγούμενη ανάλυση, αυτό το πρότυπο επιβεβαιώθηκε μόνο για τις φυλογενετικές ομάδες B1, E, G (δοκιμή Mann-Whitney U –  $PB_1 = 9.92 \times 10^{-11}$ ,  $PE = 1.25 \times 10^{-5}$ ,  $PG = 3.79 \times 10^{-8}$ ). Η εικόνα 2B απεικονίζει την κατανομή των 10 κορυφαίων MLST και φυλογενετικών ομάδων, με τουλάχιστον 10 γονιδιώματα, ταξινομημένα κατά φθίνουσα επικράτηση σε απομονώσεις APEC και AFEC.



Εικόνα 2. Φυλογενετικό δέντρο όλων των γονιδιωμάτων APEC και AFEC ως κυκλικό κλαδόγραμμα. Η επισήμανση του φαινοτύπου και της φυλογενετικής ομάδας των απομονώσεων απεικονίζεται σύμφωνα με το υπόμνημα.



Εικόνα 3. Πλαίσιο A: Κατανομή φυλογενετικών ομάδων σε παθογόνα (APEC) και μη παθογόνα (AFEC) γονιδιώματα *E. coli* πτηνών. Πλαίσιο B: Τα 10 πιο διαδεδομένα *E. coli* MLST με τουλάχιστον 10 γονιδιώματα, που βρίσκονται σε γονιδιώματα AFEC (επάνω πλαίσιο) και APEC (κάτω πλαίσιο).

### 3.3 Αποτελέσματα της ανάλυσης mGWAS και χαρακτηρισμός γονιδίων που σχετίζονται σημαντικά με τον φαινότυπο APEC

Διεξήχθησαν δύο ξεχωριστές αναλύσεις mGWAS για τον εντοπισμό σημαντικών συσχετίσεων του φαινοτύπου APEC με α) το πανγονιδίωμα των *E. coli* από πτηνά, που αντιπροσωπεύεται από ένα πίνακα παρουσίας/απουσίας γονιδίων στα γονιδιώματα και με β) SNPs στα γονιδιώματα *E. coli*, που αντιπροσωπεύεται από ένα πίνακα παρουσίας/απουσίας SNPs και δημιουργήθηκε από την ευθυγράμμιση όλων (n=2024) των γονιδιωμάτων AFEC και APEC. Ο Πίνακας 3, συνοψίζει τις πληροφορίες για τα 17 γονίδια που βρέθηκαν να είναι σημαντικά εμπλουτισμένα στον φαινότυπο APEC (από τα 29,262 γονίδια που δοκιμάστηκαν) από την ανάλυση mGWAS που βασίζεται σε γονίδια. Επιπλέον, 5 SNPs (από τα 22,710 που δοκιμάστηκαν) που αντιστοιχούν σε 4 μοναδικά γονίδια, συσχετίστηκαν σημαντικά με τον φαινότυπο APEC (Πίνακας 3). Συγκεκριμένα ήταν τα: snp\_197289, που βρίσκεται στο γονίδιο pyrH, snp\_6681911 και snp\_668914, που βρίσκονται στο γονίδιο ybeZ. Το snp\_956770 που βρίσκεται στο γονίδιο rflB και το snp\_2448166\_1 που βρίσκεται στο γονίδιο gyrA. Συνολικά, οι δύο αναλύσεις απέδωσαν 21 γονίδια σημαντικά συνδεδεμένα με τον φαινότυπο APEC (Πίνακας 3). Λαμβάνοντας υπόψη τόσο το GWAS με βάση τα γονίδια όσο και το GWAS που βασίζεται σε SNPs, βρέθηκαν 21 γονίδια σημαντικά συνδεδεμένα με τον φαινότυπο APEC.

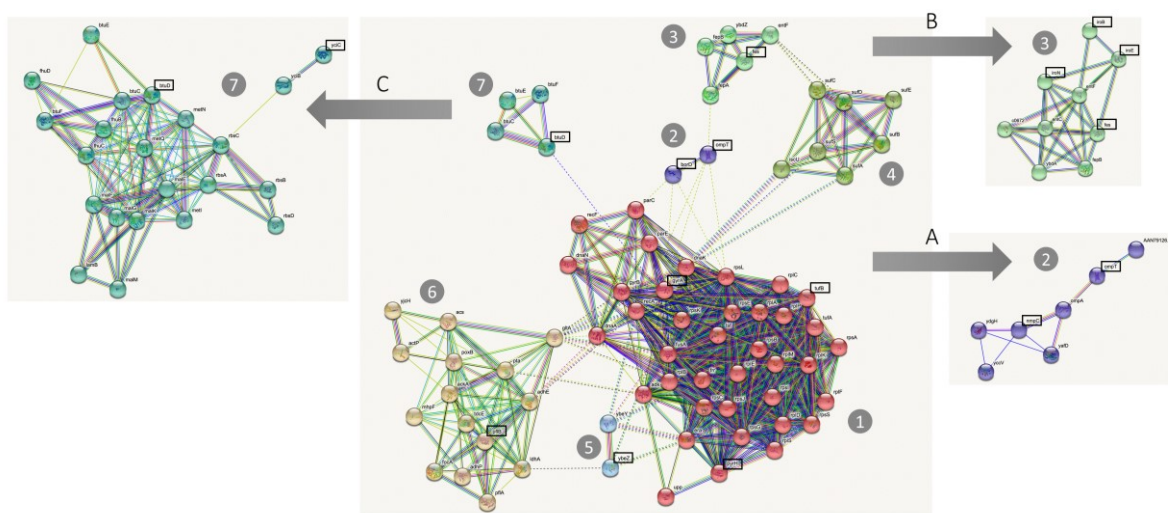
Πίνακας 3. Τα γονίδια που βρέθηκαν να σχετίζονται σημαντικά με τον φαινότυπο APEC με βάση την ανάλυση mGWAS.

Όνομα γονιδίου	Εναλλακτικό όνομα γονιδίου	Επισήμανση (annotation)	Επιπολασμός σε AFEC (%)	Επιπολασμός σε APEC (%)	Αναλο για APEC/AFEC	Βρίσκεται σε πλασμίδιο ColV
<b>iroE</b>	besA	Iron enterobactin esterase	51.12	84.30	1.649	Ναι
<b>ompT</b>		OmpT protease 7	58.07	84.30	1.452	Ναι
<b>iroN</b>	fepA	Siderophore receptor	51.02	84.02	1.647	Ναι
<b>fes</b>		Iron enterobactin esterase	51.12	84.02	1.644	Όχι
<b>btuD</b>		Vitamin B12 import ATP-binding protein	51.23	84.02	1.640	Όχι
<b>yciC</b>		Uncharacterized multi-pass membrane protein	50.05	83.84	1.675	Όχι
<b>iroB</b>	elmGT	Enterobactin C-glucosyltransferase	50.91	83.75	1.645	Ναι
<b>group_2_364</b>		IS481 family transposase ISKpn28	49.30	82.37	1.671	Ναι
<b>group_1_80</b>		IS481 family transposase ISKpn28	51.55	82.09	1.592	Ναι
<b>group_6_989</b>		Hypothetical protein	57.22	80.17	1.401	Ναι
<b>hlyF</b>	lgrD	SDR family oxidoreductase	57.01	78.88	1.384	Ναι
<b>ydhA</b>	yuaT	Putative membrane lipoprotein	55.51	77.69	1.400	Όχι
<b>ompD</b>	nmpC	Outer membrane porin protein	49.09	77.50	1.579	Όχι
<b>iss</b>	borD	Increased serum survival gene	47.49	76.68	1.615	Ναι
<b>traQ</b>		Protein involved in conjugation	39.14	55.19	1.410	Ναι
<b>tufB</b>		Elongation factor Tu 2	30.91	51.52	1.667	Όχι

Για να χαρακτηριστούν τα γονίδια που σχετίζονται με τον φαινότυπο APEC, ερευνήθηκαν στο πλαίσιο ενός δικτύου αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης, χρησιμοποιώντας τη βάση δεδομένων STRING (Treangen et al., 2014) και τα στελέχη *E. coli* K12 MG1655 και CFT073 ως γονιδιώματα αναφοράς. Τα τέσσερα γονίδια με μη τυπικά ονόματα γονιδίων (group\_xxxx), που αντιστοιχούν σε γονίδια που κωδικοποιούν υποθετικές πρωτεΐνες, εξαιρέθηκαν από αυτήν την ανάλυση, καθώς η βάση δεδομένων δέχεται μόνο τυπικά (standard) ονόματα γονιδίων. Από τα υπόλοιπα γονίδια (n=17), δίκτυα αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης αναγνωρίστηκαν για 14 γονίδια.

Είναι ενδιαφέρον ότι η ανάλυση έδειξε ότι αυτό το σύνολο γονιδίων σχηματίζει το ίδιο πρωτεϊνικό δίκτυο (Εικόνα 4 – κεντρικό πλαίσιο) μαζί με τρία υπο-δίκτυα (Εικόνα 4 – πλαίσια A, B, C), τα οποία μαζί αποτελούνται από συνολικά 108 γονίδια. Για να περιγραφεί

λεπτομερώς το μεγάλο δίκτυο (Εικόνα 3 – κεντρικό πλαίσιο), εφαρμόστηκε ομαδοποίηση (clustering) Markov Cluster Algorithm (MCL) με παράμετρο inflation = 3. Επίσης, κάθε κόμβος (node) του δικτύου επισημάνθηκε με ξεχωριστό χρώμα με βάση την ομάδα του και οι ομάδες αριθμήθηκαν από το 1 έως το 7.



Εικόνα 4. Δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης των γονιδίων που σχετίζονται σημαντικά με τον φαινότυπο APEC, που δημιουργήθηκε μέσω της διαδικτυακής πλατφόρμας STRING (Szklarczyk et al., 2021). Τα σημαντικά γονίδια σημειώνονται με μαύρα ορθογώνια. Το μεγάλο δίκτυο στο κέντρο του σχήματος υπολογίστηκε με το στέλεχος *E. coli* K12 MG1655. Τα γονίδια χρωματίζονται ανάλογα με την ομάδα (cluster) που ανήκουν (ομάδα 1 – 7). Τα τρία υπο-δίκτυα που βρέθηκαν στα πάνελ A, B και Γ υπολογίστηκαν με το στέλεχος *E. coli* CFT073.

Η ομάδα 1 περιλαμβάνει τα γονίδια elongation factor Tu 2 (*tufB*), DNA gyrase subunit A (*gyrA*) και uridylylate kinase (*pyrH*). Αυτό το σύμπλεγμα είναι εμπλουτισμένο για τις μοριακές λειτουργίες: DNA topoisomerase type II ( $P = 5.11 \times 10^{-5}$ ), DNA negative supercoiling activity ( $P = 0.0170$ ), large ribosomal subunit rRNA binding ( $P = 0.00029$ ), translation elongation factor activity ( $P = 0.00029$ ), structural constituent of ribosome ( $P = 6.85 \times 10^{-23}$ ), rRNA binding ( $P = 1.72 \times 10^{-18}$ ) και purine ribonucleotide triphosphate binding ( $P = 0.0011$ ). Το σύμπλεγμα 2 μαζί με το υπο-δίκτυο A περιλαμβάνει τα γονίδια ompT protease (*ompT*), increased serum survival gene (*borD* – πιο κοινά γνωστό ως *iss*) outer membrane porin (*nmpC*, επίσης γνωστό ως *ompD*).

Η ομάδα 3 μαζί με το υπο-δίκτυο B περιλαμβάνει τα γονίδια: iron enterobactin esterase (*fes*) enterobactin C-glucosyltransferase (*iroB*), iron enterobactin esterase (*iroE*) και siderophore receptor (*iroN*). Αυτό το σύμπλεγμα είναι εμπλουτισμένο για τις μοριακές λειτουργίες: enterobactin transmembrane transporter activity ( $P = 0.0240$ ), colicin transmembrane

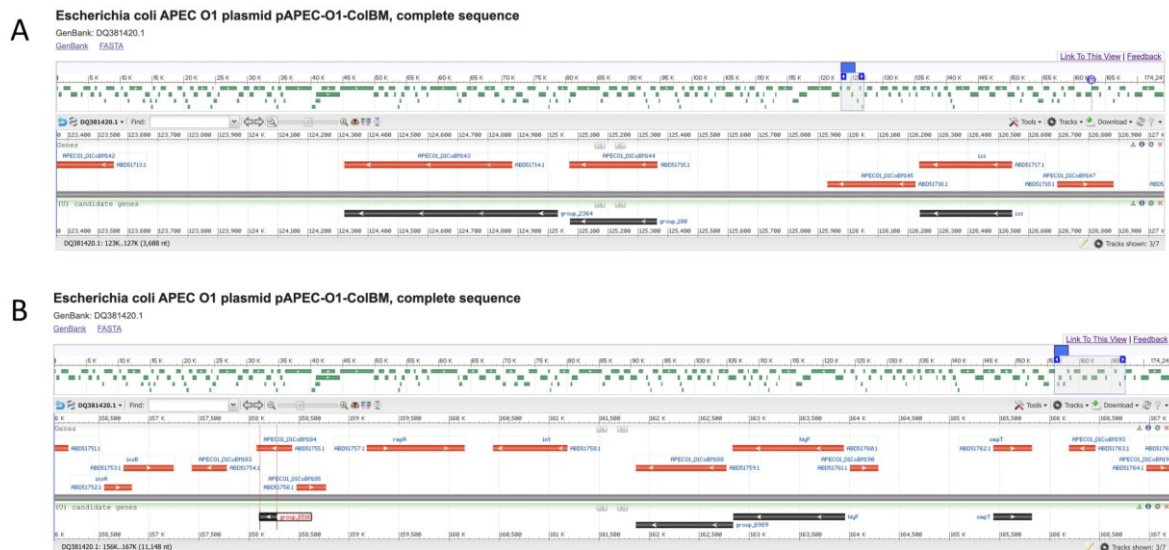
transporter activity ( $P = 0.0240$ ), ferric-enterobactin transmembrane transporter activity ( $P = 0.0240$ ), protein domain specific binding ( $P = 0.0240$ ), ligand-gated channel activity ( $P = 0.0240$ ), signalling receptor activity ( $P = 0.0272$ ) και siderophore uptake transmembrane transporter activity ( $P = 0.0279$ ). Η ομάδα 4 δεν περιλάμβανε σημαντικά γονίδια αλλά συνέδεε τις ομάδες 1 και 3. Αυτό το σύμπλεγμα είναι εμπλουτισμένο για την λειτουργία sulfur incorporation into metallo-sulfur cluster ( $P = 0.0042$ ) και iron-sulfur cluster assembly. Η ομάδα 5 περιλαμβάνει μόνο ένα από τα σημαντικά γονίδια, το phosphate starvation-inducible protein (*ybeZ*).

Η ομάδα 6 μαζί με το υπο-δίκτυο C περιλαμβάνει τα γονίδια: vitamin B12 import ATP-binding protein (*btuD*) and uncharacterized multi-pass membrane protein (*yciC*) και είναι εμπλουτισμένη για: ATPase-coupled maltose transmembrane transporter activity ( $P = 0.00033$ ), D-ribose transmembrane transporter activity ( $P = 0.0157$ ), ATPase-coupled vitamin B12 transmembrane transporter activity ( $P = 0.0242$ ) and L-methionine transmembrane transporter activity ( $P = 0.0242$ ). Η ομάδα 7 περιλαμβάνει το γονίδιο formate acetyltransferase (*pflB*) και είναι εμπλουτισμένο για: acetate metabolic process and ethanol biosynthetic process ( $P = 6.75 \times 10^{-9}$ ), pyruvate metabolism ( $P = 1.45 \times 10^{-12}$ ), glycyl-radical enzyme activating activity ( $P = 0.0283$ ), acetate transport ( $P = 0.0283$ ), citrate cycle ( $P = 6.45 \times 10^{-12}$ ) carbon metabolism και propionate metabolic process ( $1.58 \times 10^{-12}$ ).

Τα τρία γονίδια που δεν συμπεριλήφθηκαν στο δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης, περιλαμβάνουν το F pilin synthesis protein (*traQ*), putative membrane lipoprotein (*ydhA*) και το SDR family oxidoreductase (*hlyF*). Μια πρόσθετη προσπάθεια να επισημανθούν αυτά τα γονίδια στη βάση δεδομένων KEGG (Treangen et al., 2014) για συμμετοχή σε μεταβολικές οδούς *E. coli* δεν απέδωσε κανένα αποτέλεσμα.

Τέλος, τα 4 σημαντικά γονίδια που κωδικοποιούν υποθετικές πρωτεΐνες, είναι τρία υποθετικά γονίδια τρανσποζάσης (transposase) (*group\_2364*, *group\_180*, *group\_6516*) και μία υποθετική πρωτεΐνη (*group\_6516*) με άγνωστη πιθανή λειτουργία. Ερευνήθηκε η γονιδιωματική θέση αυτών των γονιδίων στο γονιδίωμα αναφοράς APEC O1 και βρέθηκε ότι όλα βρίσκονταν στο πλασμίδιο ColV, ένα κινητό γενετικό στοιχείο που είναι γνωστό ότι φέρει πολλαπλά γονίδια αντοχής στα αντιβιοτικά και ΓΛΔ στα APEC (Azam et al., 2020; Collingwood et al., 2014; Hujia et al., 2015; Schouler et al., 2012). Πιο συγκεκριμένα, οι υποθετικές τρανσποζάσες *group\_2364* και *group\_180* (που εμφανίζουν παρόμοια ταυτότητα αλληλουχίας με την οικογένεια τρανσποζασών IS418), τοποθετήθηκαν upstream του σημαντικού γονιδίου *iss*

(Εικόνα 5A). Επιπλέον, η υποθετική τρανσποζάση group\_6516 (που δείχνει παρόμοια ταυτότητα αλληλουχίας με την οικογένεια τρανσποζασών IS21) εντοπίστηκε upstream σε σχέση με τα υποψήφια γονίδια hlyF και ompT (Εικόνα 5B). Τέλος, η υποθετική πρωτεΐνης group\_6989 βρίσκεται δίπλα στο υποψήφιο γονίδιο hlyF (Εικόνα 5B).

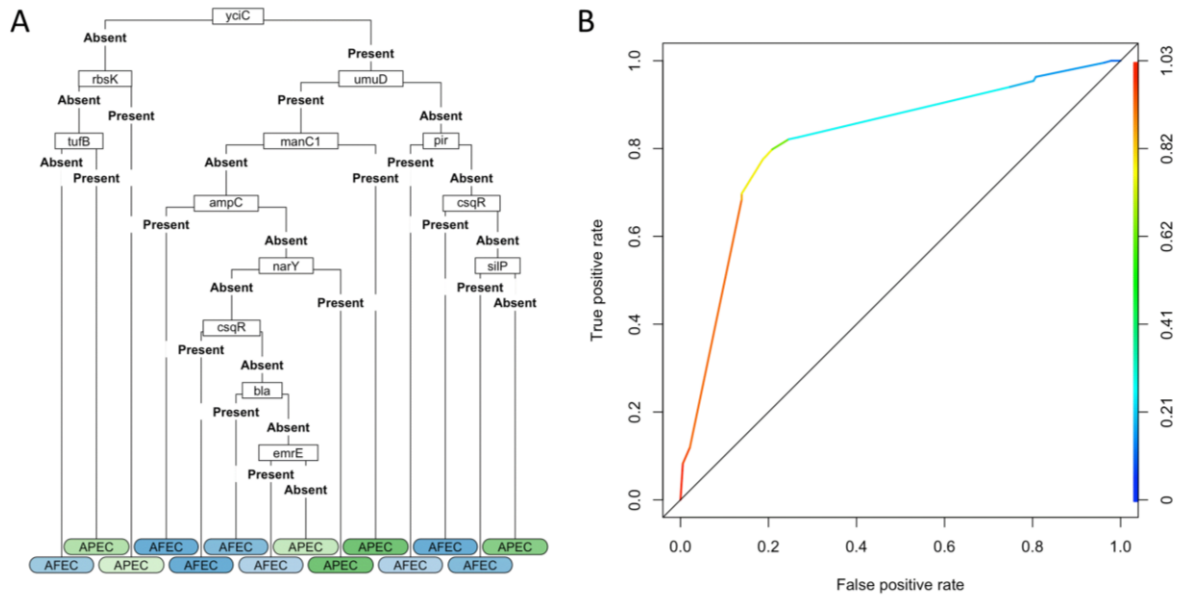


Εικόνα 5. Γονιδιοματική θέση των τριών υποθετικών γονιδίων τρανσποζάσης (group\_2364, group\_180, group\_6516) και μίας υποθετικής πρωτεΐνης (group\_6516) στο γονιδίωμα αναφοράς *E. coli* APEC O1.

### 3.4 Ταξινομητής δέντρου αποφάσεων για τη διάκριση μεταξύ APEC και AFEC

Χρησιμοποιήθηκε μια μέθοδος μηχανικής μάθησης για τη δημιουργία ενός αλγόριθμου δέντρου αποφάσεων ικανού να ταξινομήσει με ακρίβεια τα γονιδιώματα των *E. coli* ως AFEC ή APEC. Είναι σημαντικό να σημειωθεί ότι αυτή η μέθοδος είναι θεμελιωδώς διαφορετική από τις προαναφερθείσες αναλύσεις mGWAS καθώς στοχεύει στον εντοπισμό ενός ελάχιστου συνόλου γονιδίων που μαζί με ένα σύνολο κανόνων απόφασης μπορούν να διακρίνουν τις δύο μεταβλητές-στόχος (target variables), δηλαδή APEC και AFEC. Η Εικόνα 6 απεικονίζει τα 12 γονίδια που προσδιορίζονται ως σημαντικές μεταβλητές πρόβλεψης στον αλγόριθμο του ταξινομητή. Αυτά τα 12 γονίδια είναι: uncharacterized multi-pass membrane protein (yciC), ribokinase (rbsK), DNA polymerase V protein (umuD), elongation factor Tu (tufB), mannose-1-phosphate guanylyltransferase (manC1), PI protein (pir), beta-lactamase (ampC), DNA-binding transcriptional dual regulator (csqR), respiratory nitrate reductase 2 beta chain (narY), silver-translocating P-type ATPase (silP), beta-lactamase (bla) και multidrug transporter (emrE).

Οι μετρήσεις απόδοσης (performance metrics) του ταξινομητή ήταν οι εξής: accuracy = 79.5%, precision = 81.7%, recall = 79.8%. Η Εικόνα 6B απεικονίζει την τυπική καμπύλη ROC η οποία εμφανίζει τη διακύμανση του True Positive Rate (TPR) έναντι του False Positive Rate (FPR) για διαφορετικά κατώφλια (probability cut-offs). Η βέλτιστη τιμή κατωφλιού ήταν 0.63.



Εικόνα 6. Ταξινομητής δέντρου απόφασης για την διάκριση APEC και AFEC. Πλαίσιο A: Σε συνδυασμό με ένα σύνολο κανόνων (παρουσία/απουσία γονιδίων), χρησιμοποιούνται 12 γονίδια για την ταξινόμηση των γονιδιωμάτων ως APEC ή AFEC. Πλαίσιο B: Το True Positive Rate (TPR) απεικονίζεται σε σχέση με το False Positive Rate (FPR) για μια σειρά από χρωματικά κωδικοποιημένα κατώφλια απόφασης για τη δημιουργία της τυπικής χαρακτηριστικής καμπύλης λειτουργίας του δέκτη (καμπύλη ROC). Η βέλτιστη τιμή κατωφλιού ήταν 0.63.

#### 4. Συζήτηση

Σκοπός της διατριβής ήταν η διερεύνηση των υποκείμενων γενετικών μηχανισμών των στελεχών *E. coli* από πτηνά που μπορούν να προκαλέσουν κολιβακίλλωση, μια από τις σημαντικότερες ασθένειες με σημαντικές οικονομικές απώλειες για τη βιομηχανία των πουλερικών. Για το λόγο αυτό, διεξήχθη μια συστηματική αναζήτηση για τη συλλογή όλων των διαθέσιμων, υψηλής ποιότητας γονιδιωμάτων *E. coli* και τη δημιουργία μιας ολοκληρωμένης βάσης δεδομένων των APEC και AFEC. Ο στόχος ήταν να αξιοποιηθεί αυτό το μεγάλο σύνολο δεδομένων, προκειμένου να διεξαχθεί μια ισχυρή ανάλυση mGWAS (Power et al., 2017) και έτσι να προσδιοριστούν τα γενετικά στοιχεία που σχετίζονται με την παθογένεια του *E. coli*.

Από την περιγραφική στατιστική ανάλυση της βάσης δεδομένων αποκαλύφθηκαν αρκετά ενδιαφέροντα μοτίβα. Η ανάλυση για την κατανομή των φυλογενετικών ομάδων στα γονιδιώματα AFEC και APEC έδειξε ότι το APEC μπορεί να προκύψει από οποιαδήποτε φυλογενετική ομάδα *E. coli*, αν και οι φυλογενετικές ομάδες *E. coli* B2, C και G είναι πιο πιθανό να περιέχουν παθογόνα στελέχη. Αυτό το εύρημα επιβεβαιώνει τα αποτελέσματα των Johnson et al. (2022). Ένα άλλο ενδιαφέρον εύρημα ήταν ότι τα APEC έχουν σημαντικά μικρότερο γονιδίωμα καθώς και λιγότερα γονίδια από τα AFEC. Μια πρόσφατη μελέτη σχετικά με αυτό το θέμα πρότεινε ότι τα παθογόνα βακτηριακά στελέχη έχουν γενικά μικρότερο γονιδίωμα σε σύγκριση με τα μη-παθογόνα στελέχη συγγενή τους, ένα χαρακτηριστικό που ακόμα δεν έχει εξηγηθεί πλήρως (Weinert & Welch, 2017). Περαιτέρω έρευνα θα μπορούσε να αναδείξει τη σημασία αυτού του χαρακτηριστικού στα APEC.

Η ανάλυση μέσω mGWAS υπέδειξε ότι η παθογένεια στα στελέχη *E. coli* από πτηνά οφείλεται σε διακυμάνσεις τόσο στο επίπεδο SNP όσο και στο επίπεδο των γονιδίων. Συνολικά, 21 γονίδια συσχετίστηκαν σημαντικά με τον φαινότυπο APEC, 17 ταυτοποιήθηκαν από το mGWAS με βάση τα γονίδια και τέσσερα από το mGWAS που βασίστηκε στα SNPs. Στην ανάλυση δεδομένων αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης αυτών των 21 γονιδίων, βρέθηκε ένα μεγάλο πρωτεϊνικό δίκτυο που αποτελείται από 108 γονίδια, στο οποίο συγκεντρώθηκαν 14/21 σημαντικά γονίδια. Ως εκ τούτου, η παθογένεια του *E. coli* των πτηνών πιθανότατα ρυθμίζεται από την αλληλεπίδραση στενά συνδεδεμένων μεταβολικών οδών. Τα σημαντικά γονίδια των APEC που βρέθηκαν στο πρωτεϊνικό δίκτυο ομαδοποιήθηκαν σε πέντε ευρύτερες κατηγορίες.

Η πρώτη κατηγορία αποτελούνταν από γονίδια που εμπλέκονται στην πρόσληψη θρεπτικών συστατικών από τον ξενιστή. Το γεγονός ότι η πρόσληψη σιδήρου είναι απαραίτητη για την παθογένεια των APEC έχει αποδειχθεί στο παρελθόν (Q. Gao et al., 2012) και πρόσφατα αποδείχθηκε σε μια μελέτη mGWAS για στελέχη ExPEC (Galardini et al., 2020). Το operon *iroBCDEN*, που βρίσκεται συνήθως σε πλασμίδια ColV και σπανιότερα στο χρωμόσωμα των APEC, ρυθμίζει τη γλυκοζυλίωση (γονίδιο *iroB*), την επεξεργασία (γονίδιο *iroE*) και τη μεταφορά (γονίδιο *iroN*) του salmocheline siderophore, μιας πρωτεΐνης που απομονώνει (sequesters) τον σίδηρο από τα κύτταρα του ξενιστή (Caza et al., 2008). Μια λειτουργική πρωτεϊνική ανάλυση αυτών των γονιδίων στο στέλεχος αναφοράς APEC O78 αποκάλυψε τον κρίσιμο ρόλο τους στην παθογένεια των APEC (Caza et al., 2008). Το γονίδιο *iron(III) enterobactin esterase (fes)* αποδείχθηκε επίσης ότι αποτελεί μέρος της διαδικασίας βιοσύνθεσης του salmocheline siderophore καθώς η διαγραφή του στο στέλεχος αναφοράς APEC  $\chi$ 7122 μείωσε τα επίπεδα της πρωτεΐνης αλλά δεν την εξάλειψε εντελώς (Caza et al., 2015). Σε αυτό το πλαίσιο, ένα επιπλέον σημαντικό γονίδιο, το elongation factor Tu2 (*tufB*), έχει επίσης αποδειχθεί ότι παίζει σημαντικό ρόλο στα στελέχη ExPEC (στα οποία τα APEC είναι υποομάδα) που προκαλούν ασθένειες σε ανθρώπους και ζώα (κυρίως θηλαστικά) (Zhuge, Sun, et al., 2019).

Το δίκτυο αλληλεπίδρασης πρωτεΐνης-πρωτεΐνης υποδηλώνει ότι η πρόσληψη πρόσθετων θρεπτικών συστατικών παίζει ρόλο στην παθογένεια των APEC: α) πρόσληψη βιταμίνης B12, που σχετίζεται με την πρωτεΐνη import ATP-binding (*btuD*) που δεσμεύει τη βιταμίνη B12 και β) acetate assimilation, που σχετίζεται με το γονίδιο formate acetyltransferase (*pf1B*). Είναι ενδιαφέρον ότι μια πρόσφατη μελέτη έδειξε ότι η διαγραφή τριών γονιδίων (*acs*, *yjcH* και *actP*) που σχετίζονται με το γονίδιο *pf1B* σε εξαιρετικά παθογόνα APEC, είχε ως αποτέλεσμα την εξάλειψη του συστήματος αφομοίωσης οξικού (acetate assimilation system) προκαλώντας έτσι μείωση της κυτταροτοξικότητας των APEC (Zhuge, Sun, et al., 2019). Ως εκ τούτου, η μελέτη πρότεινε ότι το σύστημα αυτό διαδραματίζει κρίσιμο ρόλο στην παθογένεια των APEC. Συλλογικά, τα αποτελέσματα από αυτήν την ανάλυση υποδηλώνουν ότι η πρόσληψη βασικών θρεπτικών συστατικών όπως ο σίδηρος, το οξικό (acetate) και η βιταμίνη B12 είναι κρίσιμης σημασίας για την λοιμογόνο δράση των APEC.

Η δεύτερη κατηγορία σημαντικών γονιδίων που προέκυψαν από την ανάλυση του πρωτεϊνικού δικτύου ήταν γονίδια που συμμετείχαν στην άμυνα από το ανοσοποιητικό σύστημα του ξενιστή. Αυτή η κατηγορία περιλάμβανε το γονίδιο *ompT* protease (*ompT*), το serum survival

gene (iss) και το SDR family oxidoreductase (hlyF). Το γονίδιο ompT υπάρχει σε πλασμίδια ColV και συμμετέχει στην αποικοδόμηση των πρωτεϊνών του ξενιστή και των πεπτιδίων με αντιμικροβιακή λειτουργία (Johnson et al., 2006; Kukkonen & Korhonen, 2004). Το γονίδιο ανήκει στους προαναφερθέντες ελάχιστους προγνωστικούς παράγοντες APEC που προτάθηκαν από τους Johnson et al. (2008) και έχει αποδειχθεί πειραματικά ότι ρυθμίζει την προσκόλληση, την εισβολή, τον αποικισμό και τον πολλαπλασιασμό των APEC στον ξενιστή (Hejair et al., 2017). Στο ίδιο σύμπλεγμα δικτύου βρέθηκε το γονίδιο iss (γνωστό και ως borD), το οποίο επίσης υπάρχει σε πλασμίδια ColV και εκφράζεται στην εξωτερική μεμβράνη των βακτηριακών κυττάρων (Johnson et al., 2006; Lynne et al., 2007). Αυτό το γονίδιο είναι επίσης ένας από τους γνωστούς ελάχιστους προγνωστικούς παράγοντες και πειραματικά αποδεδειγμένο ότι ρυθμίζει την ανάπτυξη και την επιβίωση των APEC στον ορό του ξενιστή, ωστόσο η ακριβής βιολογική λειτουργία και ο μηχανισμός δράσης του παραμένουν αδιευκρίνιστα (Huja et al., 2015; Johnson et al., 2008). Επιπλέον, στα στελέχη ExPEC (η ευρύτερη ομάδα των APEC), αποδείχθηκε ότι το γονίδιο iss εμπλέκεται στη βιοσύνθεση του group 5 capsule, μιας πρωτεΐνης που προστατεύει τα APEC από τις πρωτεΐνες συμπληρώματος του ανοσοποιητικού συστήματος του ξενιστή (Biran et al., 2021). Ένα άλλο γονίδιο που είναι ελάχιστος προγνωστικός παράγοντας APEC και επίσης βρίσκεται σε πλασμίδια ColV, είναι το γονίδιο hlyF (Johnson et al., 2008; Lynne et al., 2007). Πρόσφατα βρέθηκε ότι ρυθμίζει τη βιοσύνθεση των κυστιδίων της εξωτερικής μεμβράνης (outer membrane vesicles), τα οποία προάγουν τη διαφυγή από το σύμπλεγμα αυτοφαγώματος-λυσώματος (autophagosome-lysosome complex) του ανοσοποιητικού συστήματος, παίζοντας έτσι έναν κρίσιμο ρόλο στην παθογένεια των APEC (David et al., 2022; Murase et al., 2016). Τέλος, το γονίδιο traQ έχει περιγραφεί να εκφράζεται σημαντικά περισσότερο σε στελέχη APEC που επιβιώνουν στον ορό του αίματος πουλερικών, αν και η βιολογική της λειτουργία είναι ακόμα άγνωστη. Συλλογικά, αυτά τα ευρήματα δείχνουν ότι πολλαπλοί αμυντικοί μηχανισμοί των APEC προάγουν την παθογένεια ρυθμίζοντας την αποφυγή των στελεχών από το ανοσοποιητικό σύστημα του ξενιστή.

Η τρίτη κατηγορία περιλάμβανε γονίδια που σχετίζονται ή/και συνεπιλέγονται (co-selected) με τον φαινότυπο της μικροβιακής αντοχής. Εδώ, το SNP στο γονίδιο (gyrA) ήταν σημαντικά εμπλουτισμένο στα APEC. Αξίζει να σημειωθεί ότι αυτό το γονίδιο κωδικοποιεί μια πρωτεΐνη που σχετίζεται ισχυρά με την μικροβιακή αντοχή στις φθοριοκινολόνες (Awad et al., 2016; Kim et al., 2021; Ozawa et al., 2008; Papouskova et al., 2020; Thomrongsuwannakij et al., 2020, 2022; Vounba et al., 2018, 2019; Yoon et al., 2020). Επιπλέον, το outer membrane porin

(ompD, γνωστό και ως nmpC) πιθανώς συνεπιλέγεται με μηχανισμούς μικροβιακής αντοχής. Σε αυτό το πλαίσιο, μια πρόσφατη μελέτη έδειξε αυξημένη έκφραση του γονιδίου σε στελέχη APEC που είχαν το γονίδιο rfs διαγραμμένο, και ως εκ τούτου προτάθηκε η συμμετοχή του ompD στην μικροβιακή αντοχή (Hu et al., 2022).

Μια τέταρτη κατηγορία περιλάμβανε σημαντικά γονίδια τα οποία πιθανότατα επιλέχθηκαν από κοινού στα APEC επειδή γειτνιάζουν με άλλα σημαντικά γονίδια με καθορισμένους (γνωστούς) ρόλους. Οι υποθετικές τρανσποζάσες (group\_2364, group\_180, and group\_6516) βρέθηκαν σε αυτήν την κατηγορία και ενδέχεται να ρυθμίζουν την οριζόντια μεταφορά σημαντικών γονιδίων που βρίσκονται στις υπόλοιπες κατηγορίες. Επιπλέον, το γονίδιο που κωδικοποιεί μια υποθετική πρωτεΐνη (group\_6989), η οποία είναι ομόλογη της πρωτεΐνης *Salmonella enterica* (MIG-14) και ρυθμίζει την αντίσταση στα πεπτίδια με αντιμικροβιακή δράση (Brodsky et al., 2002). Αυτό το γονίδιο πιθανότατα μεταφέρεται μαζί με το hlyF καθώς είναι τοποθετημένα κοντά στα γονιδιώματα αναφοράς APEC (O1 και O78).

Είναι ενδιαφέρον ότι τρία γονίδια από το σύνολο των γονιδίων 21 που συνδέονται σημαντικά με τον φαινότυπο APEC, δεν έχουν περιγραφεί προηγουμένως για συμμετοχή στην λοιμογόνο δράση. Το μη χαρακτηρισμένο γονίδιο που κωδικοποιεί το multi-pass membrane protein-encoding gene (yciC), είχε τον υψηλότερο επιπολασμό στα APEC (Πίνακας 2) και βρέθηκε επίσης ως σημαντικός προγνωστικός παράγοντας στον ταξινομητή δέντρου απόφασης (Εικόνα 5). Αυτά τα ευρήματα υποδηλώνουν ότι το γονίδιο μπορεί να παίζει έναν κρίσιμο αλλά ακόμα άγνωστο ρόλο στην παθογένεια των APEC. Τέλος, το putative membrane lipoprotein (ydhA) έχει έναν υποτομέα (subdomain) που εμπλέκεται στην αναστολή της λυσοζύμης, επομένως δεν αποκλείεται ένας πιθανός ρόλος του στην αποφυγή από το ανοσοποιητικό σύστημα του ξενιστή (Callewaert et al., 2008). Ωστόσο, σε μια σχετική μελέτη δεν αναφέρθηκαν αλλαγές φαινοτύπου σε στελέχη *E. coli* με διαγραμμένα γονίδια ydhA (Lam & Winkler, 1992).

Και άλλες μελέτες επικεντρώθηκαν στον εντοπισμό γονιδίων που εμπλέκονται στην παθογένεια των APEC χρησιμοποιώντας ανάλυση mGWAS. Μια πρόσφατη μελέτη-ορόσημο είναι αυτή των Mageiros et al. (2021), η οποία διεξήγαγε mGWAS σε ~570 στελέχη *E. coli* από πτηνά (αυτά περιλαμβάνονται και στη βάση δεδομένων αυτής της διατριβής). Η ανάλυσή τους κατέληξε σε μια λίστα με 140 γονίδια που σχετίζονται σημαντικά με τα APEC, τα οποία συμμετέχουν σε βιολογικές διαδικασίες μεταβολισμού, τοξικότητας και βιοσύνθεσης πολυσακχαριτών. Αξίζει να σημειωθεί ότι μόνο το γονίδιο hlyF ήταν κοινό μεταξύ της λίστας των σημαντικών γονιδίων της παρούσας μελέτης και αυτής των Mageiros et al. (2021). Μια

εξήγηση για αυτή την ασυμφωνία μπορεί να είναι το γεγονός ότι οι Mageiros et al. διεξήγαγαν ξεχωριστές αναλύσεις mGWAS για κάθε φυλογενετική ομάδα *E. coli*, επομένως η στατιστική ισχύς της ανάλυσής τους μειώθηκε δραστικά (San et al., 2020). Μια επιπλέον μελέτη mGWAS που έγινε μετά την ολοκλήρωση της συλλογής δεδομένων αυτής της διατριβής, πραγματοποιήθηκε σε ~950 γονιδιώματα APEC και AFEC από γαλοπούλες (ινδόρνιθες), αν και συμπεριλήφθηκαν επίσης στελέχη από άλλες πηγές (π.χ. περιβάλλον) (Johnson et al., 2022). Η ανάλυση αυτής της μελέτης ανέδειξε 431 σημαντικά γονίδια, αριθμός σημαντικά μεγαλύτερος σε σύγκριση με την ανάλυσή της παρούσας, που μπορεί να εξηγείται από διαφορές στα κριτήρια φιλτραρίσματος των Johnson et al. (2022). Αξίζει να σημειωθεί ότι από τα 14 σημαντικά γονίδια του mGWAS που βασίζεται σε γονίδια, τα οκτώ ταυτοποιήθηκαν επίσης από τους Johnson et al. (2022) (δηλαδή τα *yciC*, *iroE*, *iroN*, *iroB*, *btuD*, *iss*, *ompT* και *hlyF*). Αυτό το εύρημα επικυρώνει περαιτέρω τα σημαντικά γονίδια που εντοπίστηκαν σε αυτή τη διατριβή. Επιπλέον, ο κοινός κατάλογος σημαντικών γονιδίων περιλαμβάνει το *yciC*, το ακόμα αχαρακτήριστο γονίδιο στο πλαίσιο της παθογένειας, που είχε υψηλό επιπολασμό στα APEC. Στόχος μελλοντικών μελετών θα πρέπει να είναι ο περαιτέρω χαρακτηρισμός και η πειραματική λειτουργική ανάλυση αυτού του γονιδίου.

Εκτός από την περιγραφή των υποκείμενων γενετικών μηχανισμών της παθογονικότητας των APEC, εφαρμόστηκε μια μέθοδος μηχανικής μάθησης για τη δημιουργία ενός ταξινομητή ικανού να διακρίνει τα APEC και AFEC. Το δέντρο αποφάσεων που εκπαιδεύτηκε με το πανγονιδίωμα των *E. coli* από πτηνά, ανέδειξε 12 γονίδια (*yciC*, *rbsK*, *umuD*, *tufB*, *manC1*, *pir*, *ampC*, *csqR*, *narY*, *silP*, *bla*, *emrE*) ως σημαντικές μεταβλητές. Μαζί με ένα σύνολο κανόνων απόφασης, ο ταξινομητής μπόρεσε να ταξινομήσει τα APEC και AFEC με ακρίβεια 79.5%. Μετά από περαιτέρω λειτουργικό χαρακτηρισμό αυτών των γονιδίων, ο οποίος δεν αποτέλεσε στόχο αυτής της διατριβής, αυτό το σύνολο γονιδίων θα μπορούσε να χρησιμοποιηθεί για τη δημιουργία ενός νέου πρωτοκόλλου PCR ή RT-PCR για τη γρήγορη εργαστηριακή διάκριση των στελεχών *E. coli* σε APEC και AFEC. Είναι σημαντικό να σημειωθεί ότι οι Mageiros et al. επίσης δημιούργησαν έναν ταξινομητή (τυχαίο δάσος) για να διακρίνει τα παθογόνα από τα μη-παθογόνα στελέχη και ο αλγόριθμος μηχανικής μάθησης είχε ακρίβεια 76.8%. Συνολικά, τα ευρήματα αυτής της διατριβής και των Mageiros et al. προτείνουν ότι οι προσεγγίσεις μηχανικής μάθησης έχουν τη δυνατότητα να εντοπίσουν γονίδια που συνδέονται σημαντικά με βακτηριακούς φαινοτύπους και έτσι να συνεπικουρήσουν τις αναλύσεις mGWAS. Ωστόσο, είναι σημαντικό να αναφερθεί ότι ο τομέας του mGWAS με τη χρήση μηχανικής μάθησης είναι αρκετά καινούριος (Jia et al.,

2021), αν και αναμένεται επέκταση τέτοιων προσεγγίσεων καθώς τα υψηλής ανάλυσης βακτηριακά γονιδιωματικά δεδομένα συνεχίζουν να αυξάνονται.

Συνοπτικά, αυτή η διατριβή προτείνει ότι οι κύριες βιολογικές διεργασίες που εμπλέκονται στην παθογένεια του *E. coli* των πτηνών είναι α) η πρόσληψη διαφόρων θρεπτικών συστατικών από τα κύτταρα του ξενιστή και β) η αποφυγή της ανοσοαπόκρισης του ξενιστή. Προτάθηκε επίσης ένα σύνολο γονιδίων και συνοδευτικών κανόνων για τη γρήγορη διάκριση των στελεχών AFEC και APEC που μετά από περαιτέρω έρευνα και επικύρωση μπορούν να χρησιμοποιηθούν για την διάκριση των AFEC και APEC *in silico* ή στο εργαστήριο με τεχνικές PCR/RT-PCR.

## Παραρτήματα

### Παράρτημα Α: Κώδικας

```

library(tidyverse)
library(roperators)

# Load metadata
Metadata = readxl::read_xlsx("avian-ecoli-metadata.xlsx")

# ----- PHYLOGROUPS ----- #

# Plot distribution of pathotype metadata by phylogroup
metadata %>% ggplot2::ggplot() + geom_bar(aes(x = Phylogroup, fill =
Pathotype)) + theme_minimal()

# Perform chi-square test of phylogroup vs Pathotype
metadata %>% dplyr::filter(Phylogroup != "cladeI") %>%
infer::chisq_test(Phylogroup~Pathotype)

# A tibble: 1 × 3
#   statistic chisq_df p_value
#   <dbl>     <int>   <dbl>
# 1         1       378.     7 1.04e-77

# Perform chi-square post-hoc test to see which phylogroups show
significant deviations of Pathotype from 50/50
table(metadata$Phylogroup, metadata$Pathotype) %>%
chisq.posthoc.test::chisq.posthoc.test()

# Dimension      Value      AFEC      APEC
# 1              A Residuals 11.024657 -11.024657
# 2              A p values 0.000000 0.000000
# 3              B1 Residuals 7.510682 -7.510682
# 4              B1 p values 0.000000 0.000000
# 5              B2 Residuals -8.941236 8.941236
# 6              B2 p values 0.000000 0.000000
# 7              C Residuals -8.174030 8.174030
# 8              C p values 0.000000 0.000000
# 9              cladeI Residuals 0.630945 -0.630945
# 10             cladeI p values 1.000000 1.000000
# 11             D Residuals 2.070584 -2.070584
# 12             D p values 0.691159 0.691159
# 13             E Residuals 2.520126 -2.520126
# 14             E p values 0.211163 0.211163
# 15             F Residuals 4.531845 -4.531845

```

```
# 16      F  p values  0.000105    0.000105
# 17      G  Residuals -9.926235    9.926235
# 18      G  p values  0.000000    0.000000
# ----- GENOME LENGTHS ----- #

# Compare genome length of APEC vs AFEC for all isolates
metadata %>% dplyr::filter(Phylogroup != "cladeI") %>%
  rstatix::wilcox_test(`Genome Length` ~ Pathotype)
metadata %>% dplyr::filter(Phylogroup != "cladeI") %>%
  dplyr::group_by(Pathotype) %>% dplyr::summarize_at(vars(`Genome
  Length`), skimr::skim)

# Compare genome length of APEC vs AFEC separately by phylogroup
metadata %>% dplyr::filter(Phylogroup != "cladeI") %>%
  dplyr::group_by(Phylogroup) %>% rstatix::wilcox_test(`Genome Length` ~
  Pathotype)
metadata %>% dplyr::filter(Phylogroup != "cladeI") %>%
  dplyr::group_by(Phylogroup, Pathotype) %>%
  dplyr::summarize_at(vars(`Genome Length`), skimr::skim)

# ----- GENE NUMBERS ----- #

# ----- SEQUENCE TYPES ----- #

# Select top sequence types with highest percentage of systemic
summary_ST_pathotype = metadata %>%
  dplyr::filter(ST != "-") %>%
  dplyr::group_by(ST) %>%
  dplyr::filter(n() >= 10) %>%
  dplyr::select(Phylogroup, ST, Pathotype) %>%
  dplyr::group_by(ST) %>%
  dplyr::summarize(Pathotype, across()) %>%
  dplyr::ungroup() %>%
  dplyr::count(Phylogroup, ST, Pathotype) %>%
  tidyr::pivot_wider(names_from = Pathotype, values_from = n) %>%
  replace(is.na(.), 0) %>%
  dplyr::mutate(`AFEC (%)` = AFEC/(AFEC+APEC)*100, `APEC (%)` =
  APEC/(AFEC+APEC)*100)

# Sort summary table once by percentage of AFEC and a second time by
percentage of APEC
summary_ST_pathotype %>% arrange(desc(`AFEC (%)`)) %>% as.data.frame()
%>% head(10) %>% writexl::write_xlsx("top10-ST-AFEC.xlsx")
summary_ST_pathotype %>% arrange(desc(`APEC (%)`)) %>% as.data.frame()
%>% head(10) %>% writexl::write_xlsx("top10-ST-APEC.xlsx")
```

Κώδικας στη γλώσσα προγραμματισμού R για τη δημιουργία περιγραφικών στατιστικών/γραφημάτων και τη διεξαγωγή στατιστικών δοκιμών (δοκιμή Chi-Square, δοκιμή Mann-Whitney U).

```
### Boruta feature selection ###

```{r}
## Import libraries and set options

library(randomForest)
library(caret)
library(e1071)
library(Boruta)
library(tidyverse)
setwd("G:/Other computers/Huawei/APEC-database_LOCAL/RF_classifier")
options("expressions" = 5e5)
set.seed(1234)
```

```{r}
## Load data

#data = read.csv("./rf_data.tsv",sep = "\t", stringsAsFactors = TRUE)
load("./rf-apec.RData")
#str(data)
```

```{r}
## shuffle the data and check for NAs

shuffle_index <- sample(1:nrow(data))
#head(shuffle_index)
data.shuffled <- data[shuffle_index, ]
#head(data)
unique (unlist (lapply (data.shuffled, function (x) which (is.na (x))))))
```

```{r}
## feature selection with Boruta (takes ~1.5h)

## drop genes that are present only in afec or apec ()
data.boruta = data[, sapply(data, function(col) length(unique(col))) >
1]

boruta.raw <- Boruta(binary ~ ., data = data.boruta, doTrace = 2,
maxRuns = 1000)
boruta.raw

## fix tentative genes (n=150)
boruta.fixed <- TentativeRoughFix(boruta.raw)
boruta.fixed
```

```{r}
## Keep only important features in the dataset

keep_cols = getSelectedAttributes(boruta.fixed)
names.use <- names(data.shuffled) [(names(data.shuffled) %in% keep_cols)]
data.model = data[, c("binary",names.use)]
```

```
### Train the decision tree classifier ###

# Load metadata with genes presence/absence
cat("Loading metadata.", "\n")
metadata = readr::read_tsv("avian-ecoli-metadata.tsv", show_col_types =
F)

# Extract genes selected by the Boruta feature selection algorithm
boruta_genes = readxl::read_xlsx("../geneids-ranked-by-importance-by-
boruta.xlsx")
metadata = metadata %>% dplyr::select(1:15, all_of(boruta_genes$GeneID))

# Encode APEC pathotype as a binary variable
metadata = metadata %>% mutate(APEC = ifelse(Pathotype == "APEC", 1, 0),
.after = Pathotype) %>% mutate(Pathotype = as.factor(Pathotype))

# Change gene and SNPs presence variables to categorical
metadata = metadata %>% mutate(across(17:last_col(), as.factor))
metadata = metadata %>% dplyr::select(Pathotype, 17:last_col())

# Shuffle metadata
cat("Creating training and test set.", "\n")
set.seed(123)
metadata_shuffle = metadata %>% dplyr::sample_n(2015)

# Create training and test set
sample = caTools::sample.split(metadata_shuffle$Pathotype, SplitRatio =
0.8)
train_set = subset(metadata_shuffle, sample == T)
test_set = subset(metadata_shuffle, sample == F)

# Create and plot decision tree
cat("Creating decision tree.", "\n")
par(mfrow = c(1,2))
decision_tree = rpart::rpart(Pathotype~., data = train_set, cp = 0.01)
rpart.plot::rpart.plot(decision_tree, type = 5, extra = 0)

# Plot ROC curve
pred_probs = predict(decision_tree, test_set)[, "APEC"]
pred = ROCR::prediction(pred_probs, test_set$Pathotype)
perf = ROCR::performance(pred, "acc")
roc = ROCR::performance(pred, "tpr", "fpr")
plot(roc, colorize = T, lwd = 2)
abline(a = 0, b = 1)

# Compute accuracy and optimal cutoff
max_ind = which.max(slot(perf, "y.values")[[1]])
acc = slot(perf, "y.values")[[1]][max_ind]
cutoff = slot(perf, "x.values")[[1]][max_ind]
print(c(accuracy= acc, cutoff = cutoff))

# Compute recall and precision
test_set = test_set %>% dplyr::mutate(Predicted_Pathotype =
ifelse(pred_probs >= cutoff, "APEC", "AFEC"), .after = Pathotype)
```

Κώδικας στη γλώσσα προγραμματισμού R για την εκτέλεση επιλογής χαρακτηριστικών με τον αλγόριθμο Boruta, την εκπαίδευση ενός ταξινομητή δέντρου αποφάσεων, τον υπολογισμό του πίνακα σύγκρισης και των μετρήσεων απόδοσης και, τέλος, τον υπολογισμό της καμπύλης ROC.

## **Παράρτημα Β: Συμπληρωματικός Πίνακας 1**

Ο πίνακας κατατέθηκε ως ξεχωριστό αρχείο λόγω του μεγέθους του.

## Βιβλιογραφικές αναφορές

- Abraham, S., O’Dea, M., Sahibzada, S., Hewson, K., Pavic, A., Veltman, T., Abraham, R., Harris, T., Trott, D. J., & Jordan, D. (2019). *Escherichia coli* and *Salmonella* spp. isolated from Australian meat chickens remain susceptible to critically important antimicrobial agents. *PLOS ONE*, *14*(10), e0224281. <https://doi.org/10.1371/journal.pone.0224281>
- Abram, K., Udaondo, Z., Bleker, C., Wanchai, V., Wassenaar, T. M., & Ussery, D. W. (2019). What can we learn from over 100,000 *Escherichia coli* genomes? *BioRxiv*, *July*, 708131. <https://doi.org/10.1101/708131>
- Ahmed, S., Das, T., Islam, M. Z., Herrero-Fresno, A., Biswas, P. K., & Olsen, J. E. (2020). High prevalence of mcr-1-encoded colistin resistance in commensal *Escherichia coli* from broiler chicken in Bangladesh. *Scientific Reports*, *10*(1), 18637. <https://doi.org/10.1038/s41598-020-75608-2>
- Alba, P., Leekitcharoenphon, P., Franco, A., Feltrin, F., Ianzano, A., Caprioli, A., Stravino, F., Hendriksen, R. S., Bortolaia, V., & Battisti, A. (2018). Molecular Epidemiology of mcr-Encoded Colistin Resistance in Enterobacteriaceae From Food-Producing Animals in Italy Revealed Through the EU Harmonized Antimicrobial Resistance Monitoring. *Frontiers in Microbiology*, *9*. <https://doi.org/10.3389/fmicb.2018.01217>
- Allocati, N., Masulli, M., Alexeyev, M. F., & Di Ilio, C. (2013). *Escherichia coli* in Europe: An overview. *International Journal of Environmental Research and Public Health*, *10*(12), 6235–6254. <https://doi.org/10.3390/ijerph10126235>
- Antão, E.-M., Ewers, C., Gürlebeck, D., Preisinger, R., Homeier, T., Li, G., & Wieler, L. H. (2009). Signature-Tagged Mutagenesis in a Chicken Infection Model Leads to the Identification of a Novel Avian Pathogenic *Escherichia coli* Fimbrial Adhesin. *PLoS ONE*, *4*(11), e7796. <https://doi.org/10.1371/journal.pone.0007796>
- Apostolakos, I., Feudi, C., Eichhorn, I., Palmieri, N., Fasolato, L., Schwarz, S., & Piccirillo, A. (2020). High-resolution characterisation of ESBL/pAmpC-producing *Escherichia coli* isolated from the broiler production pyramid. *Scientific Reports*, *10*(1), 11123. <https://doi.org/10.1038/s41598-020-68036-9>

- Apostolakos, I., Laconi, A., Mughini-Gras, L., Yapicier, Ö. Ş., & Piccirillo, A. (2021). Occurrence of Colibacillosis in Broilers and Its Relationship With Avian Pathogenic *Escherichia coli* (APEC) Population Structure and Molecular Characteristics. *Frontiers in Veterinary Science*, 8(September), 1–13. <https://doi.org/10.3389/fvets.2021.737720>
- Awad, A., Arafat, N., & Elhadidy, M. (2016). Genetic elements associated with antimicrobial resistance among avian pathogenic *Escherichia coli*. *Annals of Clinical Microbiology and Antimicrobials*, 15(1), 59. <https://doi.org/10.1186/s12941-016-0174-9>
- Azam, M., Mohsin, M., Johnson, T. J., Smith, E. A., Johnson, A., Umair, M., Saleemi, M. K., & Sajjad-ur-Rahman. (2020). Genomic landscape of multi-drug resistant avian pathogenic *Escherichia coli* recovered from broilers. *Veterinary Microbiology*, 247(January), 108766. <https://doi.org/10.1016/j.vetmic.2020.108766>
- Beghain, J., Bridier-Nahmias, A., Le Nagard, H., Denamur, E., & Clermont, O. (2018). ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia* genus strain phylotyping. *Microbial Genomics*, 4(7). <https://doi.org/10.1099/mgen.0.000192>
- Biran, D., Sura, T., Otto, A., Yair, Y., Becher, D., & Ron, E. Z. (2021). Surviving Serum: the *Escherichia coli* iss Gene of Extraintestinal Pathogenic *E. coli* Is Required for the Synthesis of Group 4 Capsule. *Infection and Immunity*, 89(10). <https://doi.org/10.1128/IAI.00316-21>
- Brodsky, I. E., Ernst, R. K., Miller, S. I., & Falkow, S. (2002). mig - 14 Is a *Salmonella* Gene That Plays a Role in Bacterial Resistance to Antimicrobial Peptides. *Journal of Bacteriology*, 184(12), 3203–3213. <https://doi.org/10.1128/JB.184.12.3203-3213.2002>
- Callewaert, L., Aertsen, A., Deckers, D., Vanoirbeek, K. G. A., Vanderkelen, L., Van Herreweghe, J. M., Masschalck, B., Nakimbugwe, D., Robben, J., & Michiels, C. W. (2008). A New Family of Lysozyme Inhibitors Contributing to Lysozyme Tolerance in Gram-Negative Bacteria. *PLoS Pathogens*, 4(3), e1000019. <https://doi.org/10.1371/journal.ppat.1000019>
- Caza, M., Garénaux, A., Lépine, F., & Dozois, C. M. (2015). Catechol siderophore esterases Fes, IroD and IroE are required for salmochelins secretion following utilization, but only IroD contributes to virulence of extra-intestinal pathogenic *Escherichia coli*. *Molecular*

*Microbiology*, 97(4), 717–732. <https://doi.org/10.1111/mmi.13059>

- Caza, M., Lépine, F., Milot, S., & Dozois, C. M. (2008). Specific Roles of the iroBCDEN Genes in Virulence of an Avian Pathogenic *Escherichia coli* O78 Strain and in Production of Salmochelins. *Infection and Immunity*, 76(8), 3539–3549. <https://doi.org/10.1128/IAI.00455-08>
- Chen, L., Wang, L., Yassin, A. K., Zhang, J., Gong, J., Qi, K., Ganta, R. R., Zhang, Y., Yang, Y., Han, X., & Wang, C. (2018). Genetic characterization of extraintestinal *Escherichia coli* isolates from chicken, cow and swine. *AMB Express*, 8(1), 117. <https://doi.org/10.1186/s13568-018-0646-8>
- Chen, X., Liu, W., Li, H., Yan, S., Jiang, F., Cai, W., & Li, G. (2021). Whole genome sequencing analysis of avian pathogenic *Escherichia coli* from China. *Veterinary Microbiology*, 259, 109158. <https://doi.org/10.1016/j.vetmic.2021.109158>
- Clermont, O., Christenson, J. K., Denamur, E., & Gordon, D. M. (2013). The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environmental Microbiology Reports*, 5(1), 58–65. <https://doi.org/10.1111/1758-2229.12019>
- Collingwood, C., Kemmett, K., Williams, N., & Wigley, P. (2014). Is the concept of avian pathogenic *Escherichia coli* as a single pathotype fundamentally flawed? *Frontiers in Veterinary Science*, 1(OCT), 1–4. <https://doi.org/10.3389/fvets.2014.00005>
- Collins, C., & Didelot, X. (2018). A phylogenetic method to perform genome-wide association studies in microbes that accounts for population structure and recombination. *PLOS Computational Biology*, 14(2), e1005958. <https://doi.org/10.1371/journal.pcbi.1005958>
- Cordoni, G., Woodward, M. J., Wu, H., Alanazi, M., Wallis, T., & La Ragione, R. M. (2016). Comparative genomics of European avian pathogenic *E. Coli* (APEC). *BMC Genomics*, 17(1), 960. <https://doi.org/10.1186/s12864-016-3289-7>
- Cummins, M. L., Reid, C. J., Roy Chowdhury, P., Bushell, R. N., Esbert, N., Tivendale, K. A., Noormohammadi, A. H., Islam, S., Marendia, M. S., Browning, G. F., Markham, P. F., & Djordjevic, S. P. (2019). Whole genome sequence analysis of Australian avian pathogenic *Escherichia coli* that carry the class 1 integrase gene. *Microbial Genomics*, 5(2).

<https://doi.org/10.1099/mgen.0.000250>

David, L., Taieb, F., Pénary, M., Bordignon, P.-J., Planès, R., Bagayoko, S., Duplan-Eche, V., Meunier, E., & Oswald, E. (2022). Outer membrane vesicles produced by pathogenic strains of *Escherichia coli* block autophagic flux and exacerbate inflammasome activation. *Autophagy*, *18*(12), 2913–2925. <https://doi.org/10.1080/15548627.2022.2054040>

Davis, J. J., Wattam, A. R., Aziz, R. K., Brettin, T., Butler, R., Butler, R. M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E. M., Gabbard, J. L., Gerdes, S., Guard, A., Kenyon, R. W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E. K., ... Stevens, R. (2019). The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Research*, *48*(D1), D606–D612. <https://doi.org/10.1093/nar/gkz943>

de Been, M., Lanza, V. F., de Toro, M., Scharringa, J., Dohmen, W., Du, Y., Hu, J., Lei, Y., Li, N., Tooming-Klunderud, A., Heederik, D. J. J., Fluit, A. C., Bonten, M. J. M., Willems, R. J. L., de la Cruz, F., & van Schaik, W. (2014). Dissemination of Cephalosporin Resistance Genes between *Escherichia coli* Strains from Farm Animals and Humans by Specific Plasmid Lineages. *PLoS Genetics*, *10*(12), e1004776. <https://doi.org/10.1371/journal.pgen.1004776>

de Oliveira, A. L., Newman, D. M., Sato, Y., Noel, A., Rauk, B., Nolan, L. K., Barbieri, N. L., & Logue, C. M. (2020). Characterization of Avian Pathogenic *Escherichia coli* (APEC) Associated With Turkey Cellulitis in Iowa. *Frontiers in Veterinary Science*, *7*(July), 1–16. <https://doi.org/10.3389/fvets.2020.00380>

Denamur, E., Clermont, O., Bonacorsi, S., & Gordon, D. (2021). The population genetics of pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, *19*(1), 37–54. <https://doi.org/10.1038/s41579-020-0416-x>

Didelot, X., & Wilson, D. J. (2015). ClonalFrameML: Efficient Inference of Recombination in Whole Bacterial Genomes. *PLOS Computational Biology*, *11*(2), e1004041. <https://doi.org/10.1371/journal.pcbi.1004041>

Dozois, C. M., Daigle, F., & Curtiss, R. (2003). Identification of pathogen-specific and conserved genes expressed in vivo by an avian pathogenic *Escherichia coli* strain. *Proceedings of the*

*National Academy of Sciences*, 100(1), 247–252.  
<https://doi.org/10.1073/pnas.232686799>

Dziva, F., Hauser, H., Connor, T. R., van Diemen, P. M., Prescott, G., Langridge, G. C., Eckert, S., Chaudhuri, R. R., Ewers, C., Mellata, M., Mukhopadhyay, S., Curtiss, R., Dougan, G., Wieler, L. H., Thomson, N. R., Pickard, D. J., & Stevens, M. P. (2013). Sequencing and functional annotation of avian pathogenic *Escherichia coli* serogroup O78 strains reveal the evolution of *E. coli* lineages pathogenic for poultry via distinct mechanisms. *Infection and Immunity*, 81(3), 838–849. <https://doi.org/10.1128/IAI.00585-12>

Dziva, F., & Stevens, M. P. (2008). Colibacillosis in poultry: unravelling the molecular basis of virulence of avian pathogenic *Escherichia coli* in their natural hosts. *Avian Pathology*, 37(4), 355–366. <https://doi.org/10.1080/03079450802216652>

EFSA/ECDC. (2016). European Food Safety Authority and European Centre for Disease Prevention and Control. The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2014. *EFSA Journal*, 14(2), 4380. <https://doi.org/10.2903/j.efsa.2016.4380>

EFSA/ECDC. (2019). European Food Safety Authority and European Centre for Disease Prevention and Control. The European Union summary report on antimicrobial resistance in zoonotic and indicator bacteria from humans, animals and food in 2017. *EFSA Journal*, 17(2), 5598. <https://doi.org/10.2903/j.efsa.2013.3196>

Ewers, C., Antão, E. M., Diehl, I., Philipp, H. C., & Wieler, L. H. (2009). Intestine and environment of the chicken as reservoirs for extraintestinal pathogenic *Escherichia coli* strains with zoonotic potential. *Applied and Environmental Microbiology*, 75(1), 184–192. <https://doi.org/10.1128/AEM.01324-08>

Ewers, C., de Jong, A., Prenger-Berninghoff, E., El Garch, F., Leidner, U., Tiwari, S. K., & Semmler, T. (2021). Genomic Diversity and Virulence Potential of ESBL- and AmpC-β-Lactamase-Producing *Escherichia coli* Strains From Healthy Food Animals Across Europe. *Frontiers in Microbiology*, 12. <https://doi.org/10.3389/fmicb.2021.626774>

Ewers, C., Göttig, S., Bülte, M., Fiedler, S., Tietgen, M., Leidner, U., Heydel, C., Bauerfeind, R., & Semmler, T. (2016). Genome Sequence of Avian *Escherichia coli* Strain IHIT25637, an

Extraintestinal Pathogenic *E. coli* Strain of ST131 Encoding Colistin Resistance Determinant MCR-1. *Genome Announcements*, 4(5).  
<https://doi.org/10.1128/genomeA.00863-16>

Ewers, C., Janssen, T., Kiessling, S., Philipp, H.-C., & Wieler, L. H. (2005). Rapid detection of virulence-associated genes in avian pathogenic *Escherichia coli* by multiplex polymerase chain reaction. *Avian Diseases*, 49(2), 269–273. <https://doi.org/10.1637/7293-102604R>

Falgenhauer, L., Imirzalioglu, C., Oppong, K., Akenten, C. W., Hogan, B., Krumkamp, R., Poppert, S., Levermann, V., Schwengers, O., Sarpong, N., Owusu-Dabo, E., May, J., & Eibach, D. (2019). Detection and Characterization of ESBL-Producing *Escherichia coli* From Humans and Poultry in Ghana. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.03358>

Flament-Simon, S.-C., de Toro, M., Chuprikova, L., Blanco, M., Moreno-González, J., Salas, M., Blanco, J., & Redrejo-Rodríguez, M. (2020). High diversity and variability of pipolins among a wide range of pathogenic *Escherichia coli* strains. *Scientific Reports*, 10(1), 12452. <https://doi.org/10.1038/s41598-020-69356-6>

Galardini, M., Clermont, O., Baron, A., Busby, B., Dion, S., Schubert, S., Beltrao, P., & Denamur, E. (2020). Major role of iron uptake systems in the intrinsic extra-intestinal virulence of the genus *Escherichia* revealed by a genome-wide association study. *PLOS Genetics*, 16(10), e1009065. <https://doi.org/10.1371/journal.pgen.1009065>

Gao, J., Duan, X., Li, X., Cao, H., Wang, Y., & Zheng, S. J. (2018). Emerging of a highly pathogenic and multi-drug resistant strain of *Escherichia coli* causing an outbreak of colibacillosis in chickens. *Infection, Genetics and Evolution*, 65, 392–398. <https://doi.org/10.1016/j.meegid.2018.08.026>

Gao, Q., Wang, X., Xu, H., Xu, Y., Ling, J., Zhang, D., Gao, S., & Liu, X. (2012). Roles of iron acquisition systems in virulence of extraintestinal pathogenic *Escherichia coli*: salmochelin and aerobactin contribute more to virulence than heme in a chicken infection model. *BMC Microbiology*, 12(1), 143. <https://doi.org/10.1186/1471-2180-12-143>

Giufré, M., Accogli, M., Graziani, C., Busani, L., & Cerquetti, M. (2015). Whole-Genome

- Sequences of Multidrug-Resistant *Escherichia coli* Strains Sharing the Same Sequence Type (ST410) and Isolated from Human and Avian Sources in Italy. *Genome Announcements*, 3(4). <https://doi.org/10.1128/genomeA.00757-15>
- Hafez, H. M., & Attia, Y. A. (2020). Challenges to the Poultry Industry: Current Perspectives and Strategic Future After the COVID-19 Outbreak. *Frontiers in Veterinary Science*, 7. <https://doi.org/10.3389/fvets.2020.00516>
- Hejair, H. M. A., Ma, J., Zhu, Y., Sun, M., Dong, W., Zhang, Y., Pan, Z., Zhang, W., & Yao, H. (2017). Role of outer membrane protein T in pathogenicity of avian pathogenic *Escherichia coli*. *Research in Veterinary Science*, 115, 109–116. <https://doi.org/10.1016/j.rvsc.2017.01.026>
- Hu, J., Che, C., Jiang, W., Chen, Z., Tu, J., Han, X., & Qi, K. (2022). Avian Pathogenic *Escherichia coli* through Pfs Affects the Tran-Scriptation of Membrane Proteins to Resist  $\beta$ -Lactam Antibiotics. *Veterinary Sciences*, 9(3), 98. <https://doi.org/10.3390/vetsci9030098>
- Huja, S., Oren, Y., Trost, E., Brzuszkiewicz, E., Biran, D., Blom, J., Goesmann, A., Gottschalk, G., Hacker, J., Ron, E. Z., & Dobrindt, U. (2015). Genomic avenue to avian colisepticemia. *MBio*, 6(1), 1–13. <https://doi.org/10.1128/mBio.01681-14>
- Jia, X., Yang, L., Li, C., Xu, Y., Yang, Q., & Chen, F. (2021). Combining comparative genomic analysis with machine learning reveals some promising diagnostic markers to identify five common pathogenic non-tuberculous mycobacteria. *Microbial Biotechnology*, 14(4), 1539–1549. <https://doi.org/10.1111/1751-7915.13815>
- Johnson, T. J., Miller, E. A., Flores-Figueroa, C., Munoz-Aguayo, J., Cardona, C., Fransen, K., Lighty, M., Gonder, E., Nezworski, J., Haag, A., Behl, M., Kromm, M., Wileman, B., Studniski, M., & Singer, R. S. (2022). Refining the definition of the avian pathogenic *Escherichia coli* (APEC) pathotype through inclusion of high-risk clonal groups. *Poultry Science*, 101(10), 102009. <https://doi.org/10.1016/j.psj.2022.102009>
- Johnson, T. J., Siek, K. E., Johnson, S. J., & Nolan, L. K. (2006). DNA Sequence of a ColV Plasmid and Prevalence of Selected Plasmid-Encoded Virulence Genes among Avian *Escherichia coli* Strains. *Journal of Bacteriology*, 188(2), 745–758. <https://doi.org/10.1128/JB.188.2.745-758.2006>

- Johnson, T. J., Wannemuehler, Y., Doetkott, C., Johnson, S. J., Rosenberger, S. C., & Nolan, L. K. (2008). Identification of minimal predictors of avian pathogenic *Escherichia coli* virulence for use as a rapid diagnostic tool. *Journal of Clinical Microbiology*, *46*(12), 3987–3996. <https://doi.org/10.1128/JCM.00816-08>
- Jørgensen, S. L., Kudirkiene, E., Li, L., Christensen, J. P., Olsen, J. E., Nolan, L., & Olsen, R. H. (2017). Chromosomal features of *Escherichia coli* serotype O2:K2, an avian pathogenic *E. coli*. *Standards in Genomic Sciences*, *12*(1), 33. <https://doi.org/10.1186/s40793-017-0245-3>
- Kaper, J. B., Nataro, J. P., & Mobley, H. L. T. (2004). Pathogenic *Escherichia coli*. *Nature Reviews Microbiology*, *2*(2), 123–140. <https://doi.org/10.1038/nrmicro818>
- Kaspersen, H., Sekse, C., Zeyl Fiskebeck, E., Slettemeås, J. S., Simm, R., Norström, M., Urdahl, A. M., & Lagesen, K. (2020). Dissemination of Quinolone-Resistant *Escherichia coli* in the Norwegian Broiler and Pig Production Chains and Possible Persistence in the Broiler Production Environment. *Applied and Environmental Microbiology*, *86*(7). <https://doi.org/10.1128/AEM.02769-19>
- Kathayat, D., Lokesh, D., Ranjit, S., & Rajashekara, G. (2021). Avian Pathogenic *Escherichia coli* (APEC): An Overview of Virulence and Pathogenesis Factors, Zoonotic Potential, and Control Strategies. *Pathogens*, *10*(4), 467. <https://doi.org/10.3390/pathogens10040467>
- Kemmett, K., Williams, N. J., Chaloner, G., Humphrey, S., Wigley, P., & Humphrey, T. (2014). The contribution of systemic *Escherichia coli* infection to the early mortalities of commercial broiler chickens. *Avian Pathology*, *43*(1), 37–42. <https://doi.org/10.1080/03079457.2013.866213>
- Kim, J.-H., Lee, H.-J., Jeong, O.-M., Kim, D.-W., Jeong, J.-Y., Kwon, Y.-K., & Kang, M.-S. (2021). High prevalence and variable fitness of fluoroquinolone-resistant avian pathogenic *Escherichia coli* isolated from chickens in Korea. *Avian Pathology*, *50*(2), 151–160. <https://doi.org/10.1080/03079457.2020.1855322>
- Kołosut, J., Borówka, P., Marciniak, B., Wójcik, E., Wojtasik, A., Strapagiel, D., & Dastyk, J. (2017). In silico analysis of virulence associated genes in genomes of *Escherichia coli* strains causing colibacillosis in poultry. *Journal of Veterinary Research*, *61*(4), 421–426.

<https://doi.org/10.1515/jvetres-2017-0051>

- Kukkonen, M., & Korhonen, T. K. (2004). The ompT family of enterobacterial surface proteases/adhesins: from housekeeping in *Escherichia coli* to systemic spread of *Yersinia pestis*. *International Journal of Medical Microbiology*, 294(1), 7–14. <https://doi.org/10.1016/j.ijmm.2004.01.003>
- Lam, H. M., & Winkler, M. E. (1992). Characterization of the complex pdxH-tyrS operon of *Escherichia coli* K-12 and pleiotropic phenotypes caused by pdxH insertion mutations. *Journal of Bacteriology*, 174(19), 6033–6045. <https://doi.org/10.1128/jb.174.19.6033-6045.1992>
- Landman, W. J. M., Heuvelink, a, & van Eck, J. H. H. (2013). Reproduction of the *Escherichia coli* peritonitis syndrome in laying hens. *Avian Pathology : Journal of the W.V.P.A*, 42(2), 157–162. <https://doi.org/10.1080/03079457.2013.775694>
- Landman, W. J. M., & van Eck, J. H. H. (2015). The incidence and economic impact of the *Escherichia coli* peritonitis syndrome in Dutch poultry farming. *Avian Pathology*, 44(5), 370–378. <https://doi.org/10.1080/03079457.2015.1060584>
- Li, J., Bi, Z., Ma, S., Chen, B., Cai, C., He, J., Schwarz, S., Sun, C., Zhou, Y., Yin, J., Hulth, A., Wang, Y., Shen, Z., Wang, S., Wu, C., Nilsson, L. E., Walsh, T. R., Börjesson, S., Shen, J., ... Wang, Y. (2019). Inter-host Transmission of Carbapenemase-Producing *Escherichia coli* among Humans and Backyard Animals. *Environmental Health Perspectives*, 127(10), 107009. <https://doi.org/10.1289/EHP5251>
- Li, T., Castañeda, C. D., Arick, M. A., Hsu, C.-Y., Kiess, A. S., & Zhang, L. (2020). Complete genome sequence of multidrug-resistant avian pathogenic *Escherichia coli* strain APEC-O2-MS1170. *Journal of Global Antimicrobial Resistance*, 23, 401–403. <https://doi.org/10.1016/j.jgar.2020.11.009>
- Liu, Z., Xiao, X., Li, Y., Liu, Y., Li, R., & Wang, Z. (2019). Emergence of IncX3 Plasmid-Harboring blaNDM-5 Dominated by *Escherichia coli* ST48 in a Goose Farm in Jiangsu, China. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.02002>
- Lozica, L., Repar, J., & Gottstein, Ž. (2021). Longitudinal study on the effect of autogenous vaccine application on the sequence type and virulence profiles of *Escherichia coli* in

- broiler breeder flocks. *Veterinary Microbiology*, 259, 109159.  
<https://doi.org/10.1016/j.vetmic.2021.109159>
- Lynne, A. M., Skyberg, J. A., Logue, C. M., & Nolan, L. K. (2007). Detection of Iss and Bor on the surface of *Escherichia coli*. *Journal of Applied Microbiology*, 102(3), 660–666.  
<https://doi.org/10.1111/j.1365-2672.2006.03133.x>
- Maciuca, I. E., Cummins, M. L., Cozma, A. P., Rimbu, C. M., Guguianu, E., Panzaru, C., Licker, M., Szekely, E., Flonta, M., Djordjevic, S. P., & Timofte, D. (2019). Genetic Features of mcr-1 Mediated Colistin Resistance in CMY-2-Producing *Escherichia coli* From Romanian Poultry. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.02267>
- Mageiros, L., Méric, G., Bayliss, S. C., Pensar, J., Pascoe, B., Mourkas, E., Calland, J. K., Yahara, K., Murray, S., Wilkinson, T. S., Williams, L. K., Hitchings, M. D., Porter, J., Kemmett, K., Feil, E. J., Jolley, K. A., Williams, N. J., Corander, J., & Sheppard, S. K. (2021). Genome evolution and the emergence of pathogenicity in avian *Escherichia coli*. *Nature Communications*, 12(1). <https://doi.org/10.1038/s41467-021-20988-w>
- Maluta, R. P., Nicholson, B., Logue, C. M., Nolan, L. K., Rojas, T. C. G., & Dias da Silveira, W. (2016). Complete Genomic Sequence of an Avian Pathogenic *Escherichia coli* Strain of Serotype O7:HNT. *Genome Announcements*, 4(1).  
<https://doi.org/10.1128/genomeA.01611-15>
- Manges, A. R. (2016). *Escherichia coli* and urinary tract infections: the role of poultry-meat. *Clinical Microbiology and Infection*, 22(2), 122–129.  
<https://doi.org/10.1016/j.cmi.2015.11.010>
- Mellata, M. (2013a). Human and avian extraintestinal pathogenic *Escherichia coli*: infections, zoonotic risks, and antibiotic resistance trends. *Foodborne Pathogens and Disease*, 10(11), 916–932. <https://doi.org/10.1089/fpd.2013.1533>
- Mellata, M. (2013b). Human and avian extraintestinal pathogenic *Escherichia coli*: Infections, zoonotic risks, and antibiotic resistance trends. *Foodborne Pathogens and Disease*, 10(11), 916–932. <https://doi.org/10.1089/fpd.2013.1533>
- Mo, S. S., Telke, A. A., Osei, K. O., Sekse, C., Slettemeås, J. S., Urdahl, A. M., Ilag, H. K., Leangapichart, T., & Sunde, M. (2020). blaCTX-M-1/IncI1-Iy Plasmids Circulating in

- Escherichia coli* From Norwegian Broiler Production Are Related, but Distinguishable. *Frontiers in Microbiology*, 11(March), 1–12. <https://doi.org/10.3389/fmicb.2020.00333>
- Murase, K., Martin, P., Porcheron, G., Houle, S., Helloin, E., Pénary, M., Nougayrède, J.-P., Dozois, C. M., Hayashi, T., & Oswald, E. (2016). HlyF Produced by Extraintestinal Pathogenic *Escherichia coli* Is a Virulence Factor That Regulates Outer Membrane Vesicle Biogenesis. *Journal of Infectious Diseases*, 213(5), 856–865. <https://doi.org/10.1093/infdis/jiv506>
- Nicholson, B. A., Wannemuehler, Y. M., Logue, C. M., Li, G., & Nolan, L. K. (2016). Complete Genome Sequence of the Avian-Pathogenic *Escherichia coli* Strain APEC O18. *Genome Announcements*, 4(6). <https://doi.org/10.1128/genomeA.01213-16>
- Nielsen, D. W., Mangiamale, P., Ricker, N., Barbieri, N. L., Allen, H. K., Nolan, L. K., & Logue, C. M. (2018). Complete Genome Sequence of Avian Pathogenic *Escherichia coli* Strain APEC O2-211. *Microbiology Resource Announcements*, 7(12). <https://doi.org/10.1128/MRA.01046-18>
- Nolan, L. K., Vaillancourt, J. P., Barbieri, N. L., & Logue, C. M. (2019). Colibacillosis. In *Diseases of Poultry* (pp. 770–830). Wiley. <https://doi.org/10.1002/9781119371199.ch18>
- OECD-FAO. (2017). *OECD-FAO Agricultural Outlook 2017-2026*. OECD. [https://doi.org/10.1787/agr\\_outlook-2017-en](https://doi.org/10.1787/agr_outlook-2017-en)
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>
- Ozawa, M., Harada, K., Kojima, A., Asai, T., & Sameshima, T. (2008). Antimicrobial Susceptibilities, Serogroups, and Molecular Characterization of Avian Pathogenic *Escherichia coli* Isolates in Japan. *Avian Diseases*, 52(3), 392–397. <https://doi.org/10.1637/8193-120907-Reg>
- Papouskova, A., Masarikova, M., Valcek, A., Senk, D., Cejkova, D., Jahodarova, E., & Cizek, A. (2020). Genomic analysis of *Escherichia coli* strains isolated from diseased chicken in the Czech Republic. *BMC Veterinary Research*, 16(1), 189. <https://doi.org/10.1186/s12917-020-02407-2>

- 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 Research*, 25(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Poulsen, L. L., Bisgaard, M., Jørgensen, S. L., Dideriksen, T., Pedersen, J. R., & Christensen, H. (2018). Characterization of *Escherichia coli* causing cellulitis in broilers. *Veterinary Microbiology*, 225, 72–78. <https://doi.org/10.1016/j.vetmic.2018.09.011>
- Poulsen, L. L., Kudirkiene, E., Jørgensen, S. L., Djordjevic, S. P., Cummins, M. L., Christensen, J. P., Christensen, H., Bisgaard, M., & Thøfner, I. (2020). Whole genome sequence comparison of avian pathogenic *Escherichia coli* from acute and chronic salpingitis of egg laying hens. *BMC Veterinary Research*, 16(1), 148. <https://doi.org/10.1186/s12917-020-02369-5>
- Power, R. A., Parkhill, J., & de Oliveira, T. (2017). Microbial genome-wide association studies: lessons from human GWAS. *Nature Reviews Genetics*, 18(1), 41–50. <https://doi.org/10.1038/nrg.2016.132>
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, 26(7), 1641–1650. <https://doi.org/10.1093/molbev/msp077>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Rafique, M., Potter, R. F., Ferreiro, A., Wallace, M. A., Rahim, A., Ali Malik, A., Siddique, N., Abbas, M. A., D'Souza, A. W., Burnham, C.-A. D., Ali, N., & Dantas, G. (2020). Genomic Characterization of Antibiotic Resistant *Escherichia coli* Isolated From Domestic Chickens in Pakistan. *Frontiers in Microbiology*, 10. <https://doi.org/10.3389/fmicb.2019.03052>
- Rojas, T. C. G., Maluta, R. P., Parizzi, L. P., Koenigkan, L. V., Yang, J., Yu, J., Pereira, G. A. G., & Dias da Silveira, W. (2013). Genome Sequences of Avian Pathogenic *Escherichia coli* Strains Isolated from Brazilian Commercial Poultry. *Genome Announcements*, 1(2). <https://doi.org/10.1128/genomeA.00110-13>
- Rojas, T. C. G., Parizzi, L. P., Tiba, M. R., Chen, L., Pereira, G. A. G., Sangal, V., Yang, J., Yu, J., &

- Dias da Silveira, W. (2012). Draft Genome of a Brazilian Avian-Pathogenic *Escherichia coli* Strain and In Silico Characterization of Virulence-Related Genes. *Journal of Bacteriology*, *194*(11), 3023–3023. <https://doi.org/10.1128/JB.00394-12>
- Ronco, T., Stegger, M., Andersen, P. S., Pedersen, K., Li, L., Thøfner, I. C. N., & Olsen, R. H. (2016). Draft Genome Sequences of Two Avian Pathogenic *Escherichia coli* Strains of Clinical Importance, E44 and E51. *Genome Announcements*, *4*(4). <https://doi.org/10.1128/genomeA.00768-16>
- Ronco, T., Stegger, M., Olsen, R. H., Sekse, C., Nordstoga, A. B., Pohjanvirta, T., Lilje, B., Lyhs, U., Andersen, P. S., & Pedersen, K. (2017). Spread of avian pathogenic *Escherichia coli* ST117 O78:H4 in Nordic broiler production. *BMC Genomics*, *18*(1), 13. <https://doi.org/10.1186/s12864-016-3415-6>
- Roschanski, N., Fischer, J., Falgenhauer, L., Pietsch, M., Guenther, S., Kreienbrock, L., Chakraborty, T., Pfeifer, Y., Guerra, B., & Roesler, U. H. (2018). Retrospective Analysis of Bacterial Cultures Sampled in German Chicken-Fattening Farms During the Years 2011–2012 Revealed Additional VIM-1 Carbapenemase-Producing *Escherichia coli* and a Serologically Rough *Salmonella enterica* Serovar Infantis. *Frontiers in Microbiology*, *9*. <https://doi.org/10.3389/fmicb.2018.00538>
- San, J. E., Baichoo, S., Kanzi, A., Moosa, Y., Lessells, R., Fonseca, V., Mogaka, J., Power, R., & de Oliveira, T. (2020). Current Affairs of Microbial Genome-Wide Association Studies: Approaches, Bottlenecks and Analytical Pitfalls. *Frontiers in Microbiology*, *10*. <https://doi.org/10.3389/fmicb.2019.03119>
- Sarowska, J., Futoma-Koloch, B., Jama-Kmiecik, A., Frej-Madrzak, M., Ksiazczyk, M., Bugla-Ploskonska, G., & Choroszy-Krol, I. (2019). Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: Recent reports. *Gut Pathogens*, *11*(1), 1–16. <https://doi.org/10.1186/s13099-019-0290-0>
- Schouler, C., Schaeffer, B., Brée, A., Mora, A., Dahbi, G., Biet, F., Oswald, E., Mainil, J., Blanco, J., & Moulin-Schouleur, M. (2012). Diagnostic strategy for identifying avian pathogenic *Escherichia coli* based on four patterns of virulence genes. *Journal of Clinical*

*Microbiology*, 50(5), 1673–1678. <https://doi.org/10.1128/JCM.05057-11>

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>

Singer, R. S. (2015). Urinary tract infections attributed to diverse ExPEC strains in food animals: Evidence and data gaps. *Frontiers in Microbiology*, 6(FEB), 1–9. <https://doi.org/10.3389/fmicb.2015.00028>

Sul, J. H., Martin, L. S., & Eskin, E. (2018). Population structure in genetic studies: Confounding factors and mixed models. *PLOS Genetics*, 14(12), e1007309. <https://doi.org/10.1371/journal.pgen.1007309>

Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, 49(D1), D605–D612. <https://doi.org/10.1093/nar/gkaa1074>

Thomrongsuwannakij, T., Blackall, P. J., Djordjevic, S. P., Cummins, M. L., & Chansiripornchai, N. (2020). A comparison of virulence genes, antimicrobial resistance profiles and genetic diversity of avian pathogenic *Escherichia coli* (APEC) isolates from broilers and broiler breeders in Thailand and Australia. *Avian Pathology*, 49(5), 457–466. <https://doi.org/10.1080/03079457.2020.1764493>

Thomrongsuwannakij, T., Narinthorn, R., Mahawan, T., & Blackall, P. J. (2022). Molecular and phenotypic characterization of avian pathogenic *Escherichia coli* isolated from commercial broilers and native chickens. *Poultry Science*, 101(1), 101527. <https://doi.org/10.1016/j.psj.2021.101527>

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J. A., Gladstone, R. A., Lo, S., Beaudoin, C., Floto, R. A., Frost, S. D. W., Corander, J., Bentley, S. D., & Parkhill, J. (2020). Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biology*, 21(1), 180. <https://doi.org/10.1186/s13059-020-02090-4>

Treangen, T. J., Ondov, B. D., Koren, S., & Phillippy, A. M. (2014). The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial

- genomes. *Genome Biology*, 15(11), 524. <https://doi.org/10.1186/s13059-014-0524-x>
- Tuntufye, H. N., Lebeer, S., Gwakisa, P. S., & Goddeeris, B. M. (2012). Identification of Avian Pathogenic *Escherichia coli* Genes That Are Induced In Vivo during Infection in Chickens. *Applied and Environmental Microbiology*, 78(9), 3343–3351. <https://doi.org/10.1128/AEM.07677-11>
- Vounba, P., Arsenault, J., Bada-Alambédji, R., & Fairbrother, J. M. (2019). Prevalence of antimicrobial resistance and potential pathogenicity, and possible spread of third generation cephalosporin resistance, in *Escherichia coli* isolated from healthy chicken farms in the region of Dakar, Senegal. *PLOS ONE*, 14(3), e0214304. <https://doi.org/10.1371/journal.pone.0214304>
- Vounba, P., Kane, Y., Ndiaye, C., Arsenault, J., Fairbrother, J. M., & Bada Alambédji, R. (2018). Molecular Characterization of *Escherichia coli* Isolated from Chickens with Colibacillosis in Senegal. *Foodborne Pathogens and Disease*, 15(8), 517–525. <https://doi.org/10.1089/fpd.2017.2394>
- Wang, X., Wei, L., Wang, B., Zhang, R., Liu, C., Bi, D., Chen, H., & Tan, C. (2016). Complete genome sequence and characterization of avian pathogenic *Escherichia coli* field isolate ACN001. *Standards in Genomic Sciences*, 11(1), 13. <https://doi.org/10.1186/s40793-015-0126-6>
- Wang, Y., Zhang, R., Li, J., Wu, Z., Yin, W., Schwarz, S., Tyrrell, J. M., Zheng, Y., Wang, S., Shen, Z., Liu, Z., Liu, J., Lei, L., Li, M., Zhang, Q., Wu, C., Zhang, Q., Wu, Y., Walsh, T. R., & Shen, J. (2017). Comprehensive resistome analysis reveals the prevalence of NDM and MCR-1 in Chinese poultry production. *Nature Microbiology*, 2(4), 16260. <https://doi.org/10.1038/nmicrobiol.2016.260>
- Wang, Z., Zheng, X., Guo, G., Dong, Y., Xu, Z., Wei, X., Han, X., Liu, Y., & Zhang, W. (2022). Combining pangenome analysis to identify potential cross-protective antigens against avian pathogenic *Escherichia coli*. *Avian Pathology*, 51(1), 66–75. <https://doi.org/10.1080/03079457.2021.2005240>
- Weinert, L. A., & Welch, J. J. (2017). Why Might Bacterial Pathogens Have Small Genomes? *Trends in Ecology & Evolution*, 32(12), 936–947.

<https://doi.org/10.1016/j.tree.2017.09.006>

- Wirth, T., Falush, D., Lan, R., Colles, F., Mensa, P., Wieler, L. H., Karch, H., Reeves, P. R., Maiden, M. C. J., Ochman, H., & Achtman, M. (2006). Sex and virulence in *Escherichia coli*: An evolutionary perspective. *Molecular Microbiology*, *60*(5), 1136–1151. <https://doi.org/10.1111/j.1365-2958.2006.05172.x>
- Wu, C., Wang, Y., Shi, X., Wang, S., Ren, H., Shen, Z., Wang, Y., Lin, J., & Wang, S. (2018). Rapid rise of the ESBL and mcr-1 genes in *Escherichia coli* of chicken origin in China, 2008–2014. *Emerging Microbes & Infections*, *7*(1), 1–10. <https://doi.org/10.1038/s41426-018-0033-1>
- Yin, D., Cheng, B., Yang, K., Xue, M., Lin, Y., Li, Z., Song, X., Shao, Y., Tu, J., Li, P., & Qi, K. (2021). Complete Genetic Analysis of Plasmids Carrying mcr-1 and Other Resistance Genes in Avian Pathogenic *Escherichia coli* Isolates from Diseased Chickens in Anhui Province in China. *MSphere*, *6*(2). <https://doi.org/10.1128/mSphere.01135-20>
- Yoon, M. Y., Kim, Y. Bin, Ha, J. S., Seo, K. W., Noh, E. B., Son, S. H., & Lee, Y. J. (2020). Molecular characteristics of fluoroquinolone-resistant avian pathogenic *Escherichia coli* isolated from broiler chickens. *Poultry Science*, *99*(7), 3628–3636. <https://doi.org/10.1016/j.psj.2020.03.029>
- Zajac, M., Sztromwasser, P., Bortolaia, V., Leekitcharoenphon, P., Cavaco, L. M., Ziętek-Barszcz, A., Hendriksen, R. S., & Wasyl, D. (2019). Occurrence and Characterization of mcr-1-Positive *Escherichia coli* Isolated From Food-Producing Animals in Poland, 2011–2016. *Frontiers in Microbiology*, *10*(August), 1–14. <https://doi.org/10.3389/fmicb.2019.01753>
- Zhu Ge, X., Jiang, J., Pan, Z., Hu, L., Wang, S., Wang, H., Leung, F. C., Dai, J., & Fan, H. (2014). Comparative Genomic Analysis Shows That Avian Pathogenic *Escherichia coli* Isolate IMT5155 (O2:K1:H5; ST Complex 95, ST140) Shares Close Relationship with ST95 APEC O1:K1 and Human ExPEC O18:K1 Strains. *PLoS ONE*, *9*(11), e112048. <https://doi.org/10.1371/journal.pone.0112048>
- Zhuge, X., Jiang, M., Tang, F., Sun, Y., Ji, Y., Xue, F., Ren, J., Zhu, W., & Dai, J. (2019). Avian-source mcr-1-positive *Escherichia coli* is phylogenetically diverse and shares virulence

characteristics with *E. coli* causing human extra-intestinal infections. *Veterinary Microbiology*, 239, 108483. <https://doi.org/10.1016/j.vetmic.2019.108483>

Zhuge, X., Sun, Y., Jiang, M., Wang, J., Tang, F., Xue, F., Ren, J., Zhu, W., & Dai, J. (2019). Acetate metabolic requirement of avian pathogenic *Escherichia coli* promotes its intracellular proliferation within macrophage. *Veterinary Research*, 50(1), 31. <https://doi.org/10.1186/s13567-019-0650-2>

Υπεύθυνη Δήλωση Συγγραφέα:

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.