BIOFILM FORMATION OF *ESCHERICHIA COLI* FROM SURFACE SOILS IS INFLUENCED BY VARIATION IN CELL ENVELOPE, IRON METABOLISM, AND ATTACHMENT FACTOR GENES

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Morgan Lynn Petersen

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Program:
Microbiology

June 2018

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Biofilm formation of *Escherichia coli* from surface soils is influenced by variation in cell envelope, iron metabolism, and attachment factor genes

By

Morgan Lynn Petersen

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Peter Bergholz
Chair

Dr. Birgit Prüß

Dr. Ryan Hunter

Approved:

7/2/2018
Date

Dr. John McEvoy
Department Chair

ABSTRACT

Biofilm formation may increase survival and persistence of *Escherichia coli* in the highly variable conditions of soil environments, though it remains unknown the extent variation in biofilm formation affects survival. We asked what genetic traits influence biofilm formation in phylogroup D *E. coli* isolates from surface soils, and are they associated with the soil environment? Biofilm density was analyzed and compared with soil environment characteristics. Isolates produced more biofilm per unit growth at 15°C than 37°C. Biofilm formation was greater in soil isolates than fecal isolates and in soils with moisture and higher calcium and pH levels. A GWAS analysis found variants involved in cell envelope formation and structure were associated with biofilm formed at 37°C, and stress response and iron acquisition variants were associated with biofilm formed at 15°C. Motility variants were associated with a negative effect on biofilm formed and adhesion variants associated with a positive effect.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. Peter Bergholz for advising and for giving me an opportunity to work in his laboratory. I would like to acknowledge Kaycie Schmidt for helping me with the work in the laboratory. I would also like to thank my supervisory committee, Dr. Birgit Prüß and Dr. Ryan Hunter, for guidance through my master's studies.

I'd like to acknowledge the funding sources that allowed me to perform this research and for giving me an opportunity to study at North Dakota State University: National Science Foundation award no. DEB-1453397 to P.W.B. and ND-EPSCoR.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA .....Analysis of variance

BLASTP.....Basic local alignment search tool protein

CAA .....Casamino acids

CD-HIT .....Cluster database at high identity with tolerance

CFU.....Colony forming units

CV .....Crystal violet

DNA.....Deoxyribonucleic acid

EHEC .....Enterohemorrhagic *Escherichia coli*

EPEC.....Enteropathogenic *Escherichia coli*

EPS.....Extracellular polymeric substances

ETEC.....Enterotoxigenic *Escherichia coli*

FDR.....False discovery rate

GATK .....Genome analysis tool kit

GDMM.....Glucose defined minimal media

GWAS.....Genome-wide association study

iTOL.....Interactive tree of life

LB .....Lysogeny broth

MCL.....Markov cluster algorithm

MLST .....Multi-locus sequence typing

OD.....Optical density

PBS .....Phosphate-buffered saline

qPCR.....Quantitative polymerase chain reaction

RAST .....Rapid annotation using subsystem technology

RAxML.....Randomized axelerated maximum likelihood

RNA .....Ribonucleic acid

SNP .....Single nucleotide polymorphism

SPANC.....Self-preservation and nutrient competence

TCA.....Tricarboxylic acid

UTI.....Urinary tract infection

LIST OF APPENDIX TABLES

CHAPTER 1: LITERATURE REVIEW

Genetic diversity of *E. coli* and its effect on survival

Escherichia coli is a highly genetically diverse species, in part due to its large global population size (approximately 10^{20}) (1). It is known to have fast mutation rates, especially in the accessory genomes, which can be beneficial when exposed to stress or variable environments (2–5). The core genome consists of genes shared across all *E. coli,* including genes involved in metabolic versatility, environmental hardiness, house-keeping functions, and genes used as phylogenetic signatures to distinguish between the eight evolutionarily distinct *E. coli* phylogroups: A, B1, B2, C, D, E, F, and Clade I (1, 6–8). The accessory genome contains genes not shared across all *E. coli* members and includes strain specific genes, such as antibiotic resistance or some virulence factors. High variability exists in the core and accessory genomes of *E. coli*, with approximately 2000 genes found in the *E. coli* core genome, and 4700 genes found in the average *E. coli* genome. Overall, the *E. coli* pan-genome contains approximately 12000 - 24000 genes, depending on the study, and is influenced by changes in the genetic structure of the bacteria, including gene loss or acquisition, mutations, horizontal gene transfer, and adaptive responses (1, 8–11).

E. coli is thought to mainly inhabit the intestines of warm-blooded animals, but it can be dispersed into extrahost environments, such as soils, through means such as deposition of fecal matter. Thus, presence of *E. coli* has often been an indicator of fecal contamination (12, 13). However, there is evidence of environmental strains of *E. coli* that have come to colonize and persist in these habitats (12, 14). This includes *E. coli* from clades that have diverged to become entirely environmentally-adapted and have no association with the mammal-host environment yet are phenotypically indistinguishable from other *E. coli* (15). As such, *E. coli* can come to

form stable populations in extrahost environments. That survival is dependent on *E. coli's* ability to adapt to the conditions of these environments.

The variability of soil environments may lead to an increased chance for evolutionary rescue, which occurs when a declining population undergoes an adaptive change that increases survival and lowers the chance of extinction (16). The general pattern of evolutionary rescue first begins with a population declining due to exposure to a new environmental stress. Once the population size declines below the stochastic threshold, or the critical population size where the chance of extinction due to environmental variability increases, the population will either go extinct or an adaptive change will allow the population to recover and survive (16). The chance of evolutionary rescue occurring varies. For example, initial population size is an important determinant for evolutionary rescue. Greater population sizes increase the chance of evolutionary rescue occurring due to the higher level of genetic diversity and increased time it would take for all members of the population to die off, and thus more time for evolutionary rescue to possibly occur (17, 18). For example, in a study analyzing evolutionary rescue of *Pseudomonas fluorescens* populations when exposed to varying levels of streptomycin, it was found that the larger populations were more likely to undergo evolutionary rescue, with more than 50% of the large populations showing evolutionary rescue compared to 6.4% of the small populations (19). A similar result was found in a study analyzing *Saccharomyces cerevisiae* under salt stress, where at least 526 individuals was required for a 50% chance of adaptation to occur (20). Immigration of new individuals into a population would also increase the chance for evolutionary rescue as new individuals would add to the population and increase the genetic diversity. This was shown in a study of the influence of immigration rates of *Pseudomonas aeruginosa* on evolution of rifampicin resistance, with higher immigration rates leading to an

increase in resistance (21). The level of genetic diversity in a population impacts evolutionary rescue as greater genetic diversity would increase the chance that an adaptive allele was present among the population that would allow for survival (16). An example of this was shown in a study analyzing evolutionary rescue in *Chlamydomonas reinhardtii* exposed to salt stress. Populations with greater genetic diversity were more likely to undergo evolutionary rescue and declined slower than populations with lower genetic diversity (22). This was also shown in rifampicin resistance in *P. aeruginosa* and salt stress resistance in *S. cerevisiae* (20, 21). Since evolutionary rescue leads to increases in adaptive traits, it may play a role in survival of *E. coli* in variable environments such as soils.

Recombination events also play a role in the genetic diversity of *E. coli.* As a clonal species, mutations influence the genetic diversity of *E. coli* and can confer an adaptive advantage if mutations lead to a trait that increases the ability of *E. coli* to survive under certain conditions (23, 24). Those mutations can be passed to other *E. coli* strains through recombination events. In general, the rate of recombination in *E. coli* is low (*r/m* of about 1, depending on the study) (24). Two studies using MLST analysis to assess variation in *E. coli* determined an *r/m* rate of 0.32-2.14 among 432 strains and an *r/m* rate of 0.70 among 44 strains, with both studies examining only 7 genes (25, 26). Three studies using whole genome sequencing analyzing at all shared genes in *E. coli* determined *r/m* rates of 0.90 for 20 strains, 1.02 for 27 strains, and 0.92 for 19 strains (24, 27, 28). Rates of recombination also vary by *E. coli* phylogroup and regions in the genome (24). For example, *E. coli* of the same phylogroup are more likely to have recombination events than *E. coli* of different phylogroups, possibly due to the distinct ecological overlap of bacteria in phylogroups (28). Intergroup recombination varies based on phylogroups as well. For example, phylogroup B2 was found to have the lowest intergroup recombination frequencies and

phylogroup B1 was found to have the highest, as well as recombination rates between phylogroups A and B1 found to be higher than recombination rates between other groups (7, 28). Some hotspots of recombination in the *E. coli* genome include the *rfb* operon, which codes for O-antigen synthesis, and the *fim* operon, which is involved in adhesion to host cells and in host interactions and virulence (28–30). Recombination can influence pathogenicity if genes associated with virulence hitchhike during recombination events, leading to more strains expressing virulence factors. These virulence factors include genes found on pathogenicity-associated islands like the P fimbrial structure subunit gene that aids in attachment (6, 31, 32). Thus, these events are important for the genetic diversity and survival of *E. coli*, however genetic differences in phylogroups are also important.

E. coli phylogroups

The eight *E. coli* phylogroups can be distinguished comparing SNPs among alleles of various previously defined housekeeping genes using techniques including MLST, genotyping qPCR, comparative hybridization, etc. (33–35). These phylogroups vary in the habitats they persist in or are most commonly found. For example, in a study analyzing the distribution of phylogroups from *E. coli* isolated from surface waters, it was found that phylogroup B1 and A strains were the most prevalent (36). For distribution in different animal species, B1 strains were common among geese, ducks, deer, sheep, cows, dogs, horses, and pigs, phylogroup A dominated among chickens, phylogroup B2 was most prevalent in beaver and goats, and phylogroup D was most common in turkeys (36). A similar result was found in population structures of *E. coli* in fecal depositions, where phylogroup B1 was found to be the dominant phylogroup among the fecal and soil isolates (3). B1 and D phylogroup *E. coli* isolates were found to persist better on the surfaces of plant leaves than any other phylogroup (37). Phylogroup

A has often been isolated from humans and is known to be a human-associated phylogroup, though B1 is the predominant subpopulation in humans (1). Thus, the presence of phylogroups vary based on the habitat.

Phylogroups also vary in their association with disease. For example, phylogroup D *E. coli* has been associated with urinary tract infections (UTI) and some bacteraemia infections (36, 38–40). Phylogroup B2 is known as a subpopulation in humans and is often associated with disease, as well as phylogroups E and F. B2 is involved in diseases such as EPEC, EHEC, and ETEC infections, and is the predominant cause of UTIs (1, 38). However, phylogroup A is also known as a subpopulation in humans, but is less associated with virulence (1, 38). Virulence factors differ between the phylogroups. For example, in a study analyzing *E. coli* isolated from chicken, the prevalence of virulence factors, such as *eaeA, iss, iucD, Tsh, etc.,* was more frequent among *E. coli* phylogroups B2 and D (16%) compared to phylogroups A and B1 (10%) (41). In an analysis of the distribution of phylogroups and virulence factors in *E. coli* isolates from diarrheic cattle, the number of virulence factors was significantly higher in phylogroup B1 than phylogroups A and D (42). Virulence factors associated with type III secretion system effectors, toxins, hemolysin, adhesins (FanA, Fim41, Efa1, ToxB), and the plasmid-encoded catalase KatP were only associated with phylogroup B1. Conversely, virulence factors associated with adhesins (IpfA, PrfB/PapB, Iha, F17, H), bacteriocin, and siderophore receptors were mainly found in phylogroups A and D. Other studies found virulence factors to be highest among phylogroups B1 and E from diarrhea in dogs (43) and a higher number of virulence genes in isolates from phylogroup B2 in turkeys with airsacculitis (44). Thus, in general, virulence factors and the number of virulence factors differs between phylogroups, as well as the source of isolation.

Virulence of *E. coli* involved in UTIs is influenced by attachment and the ability to form biofilms, with greater biofilm formation increasing resistance to antibiotics and prolonging infections (45–47). As a phylogroup associated with UTIs, biofilm formation in phylogroup D impacts its virulence. Soil habitats, with the high variability in the environmental conditions, select for traits that increase persistence (48, 49). Biofilm formation is a trait that can increase persistence, thus soil environments may select for greater biofilm formers. It has been shown that the genetic structure of phylogroup D *E. coli* is influenced more by soil environments than any other phylogroup (3), which could possibly give them an advantage to survival and persistence in soil habitats. This may lead to a selection for greater biofilm formers of phylogroup D *E. coli* in soil environments. Such selection of greater biofilm formation can be detrimental to human health if the selection occurs in UTI-causing isolates that may later infect the human population. Thus, it is of importance to understand the conditions of soil environments that increase persistence in *E. coli*.

Environmental stressors in soil

Soil habitats pose a separate set of stresses to *E. coli* compared to the gastrointestinal environment. For example, one of the main stresses *E. coli* can face in the gastrointestinal environment is the high acidity of the stomach. However, pH in soils is not consistent among locations and ranges in acidity and alkalinity. Temperature may also fluctuate, exposing bacteria to low temperature stress in contrast to the consistent warmer temperatures of the intestines. Other factors that impose stresses on *E. coli* are 1) varying levels of oxygen, which may lead to oxidative stress; 2) generally low levels of simple carbon sources, making energy acquisition difficult; 3) varying levels of osmolarity, which can lead to desiccation stress; and 4) possible exposure to toxins (8, 50, 51). Along with variation in nutrients and resources, *E. coli* can also be

exposed to predators, such as the amoeba *Dictyostelium discoideum* (52, 53). This creates a highly variable environment in which stressful conditions can influence the survival of bacteria present. Therefore, bacteria must have mechanisms in place to protect themselves from stresses and predation as well as compete with other microbes for resources. A mechanism of interest is biofilm formation.

Biofilm formation as a survival/resistance trait in *E. coli*

A biofilm is a protective structure produced by *E. coli* when they grow in close proximity to each other while forming a matrix of extracellular polymeric substances (EPS). Biofilms form a barrier that protects *E. coli* by preventing predators, antimicrobial agents, or other bacteria from penetrating into the biofilms and can buffer against environmental changes (54). Formation starts with a microbe entering a new area, either through dispersal or motility, and initially attaching itself using flagella through a physiochemical or electrostatic interaction with a free surface. The bacteria further their attachment using fimbriae and curli in either a specific or non-specific interaction with the surface. Attachment to abiotic surfaces are usually non-specific but attachment to biotic surfaces often require a specific receptor-ligand binding interaction (55). The *E. coli* then proliferate to form microcolonies and produce and secrete EPS, the composition of which varies among species and is produced for the maturation of biofilms. EPS is mostly made up of polysaccharides and proteins but can also contain lipids and nucleic acids. For *E. coli*, a main component of EPS is usually colanic acid, which is made up of repeating subunits of simple sugars with O-acetyl and pyruvyl side chains (56, 57). Another main component of EPS in *E. coli* is cellulose, a polysaccharide of linked D-glucose molecules that forms early in biofilm formation (58, 59). Cellulose is paired with curli production to strengthen biofilm formation and attachment (60, 61). It is thought that cellulose production decreases flagella activity by the

binding of cellulose fibers around flagella and increasing attachment (62). Maturation of the biofilm is also influenced by quorum sensing. Once a certain density is reached, the mature biofilm releases a secondary messenger molecule that signals the dispersal of bacteria to swim to a new location to form another biofilm (59).

Regulation of *E. coli* biofilm formation is complex and accomplished by regulating genes for motility and attachment. Motility is involved in the dispersal of bacteria to new locations where they can form new biofilms, an important first step in the biofilm formation process, though the need for motility to form biofilms is dependent on the environment (63, 64). It was found that some *E. coli* strains with greater motility were able to form denser biofilms. However, cells transition from a motile to a sessile lifestyle once biofilm formation starts, and a change in motility to attachment factors takes place (65).

Motility is dependent on the presence and activity of flagella. The direction the bacteria moves is dependent on the 'run and tumble' motion of the flagella and can change based on chemotaxis signals the bacteria receive. The run motion consists of the left-handed helical filaments of flagella bundling together and rotating in a counterclockwise direction to propel the *E. coli* forward. The tumble motion is when one or more of the flagella motors switch rotational direction, leading to the *E. coli* to turn or 'tumble' (66). Flagellar production is controlled by the master regulator *flhDC* (67, 68). It is thought that flagella aid in the attachment to abiotic surfaces by their ability to penetrate crevices and other subsurface features that are not accessible to the bacterial bodies (69). Attachment factors used in the adherence of bacteria to surfaces for biofilms to form include curli, type IV pili, and type I fimbriae (70–72). These have rod-like structures and aid in the attachment of *E. coli* to surfaces through physiochemical interactions with the surface. Curli are fimbrial structures made up of functional amyloids involved in initial

and permanent attachment (55). Production of curli in *E. coli* is controlled by the regulator *csgD*, which also plays a role in cellulose production (71). Type IV pili are long thin filaments made up of pilin protein. Type I fimbriae are essential for permanent attachment to surfaces during biofilm formation and are attributed to virulence and increased persistence of *E. coli* in urinary tract infections (59, 72–74).

Both motility and attachment factors in *E. coli* are influenced by the secondary messenger molecule cyclic-di-GMP. Cyclic-di-GMP is produced through the activity of diguanylate cyclases and is degraded by phosphodiesterases (75). It plays a role in biofilm formation by upregulating *csgD*, which leads to increased curli and cellulose production, and upregulating *ycgR*, a gene that inhibits flagellar production (75–77). Increased expression of the flagellar master regulator *flhDC* upregulates phosphodiesterase activity and decreases c-di-GMP production, leading to a decrease in attachment factors (76–78). Thus, an increase in motility factors leads to a decrease in attachment factors and *vice versa*, leading to a switch from a motile lifestyle to a sessile lifestyle during biofilm formation, of which c-di-GMP plays a key role. Overall, genes involved in motility can have a negative impact on genes involved in attachment, which are required for greater biofilm formation.

Biofilm formation is also influenced by some environmental factors, such as temperature and some chemical components, due to their impact on motility or attachment factors. For example, lower temperatures have been shown to increase expression of *csgA* in *E. coli*, a gene involved in the structure of curli during production, *mlrA,* which promotes *csgD* expression, and *yaiC*, a diguanylate cyclase that plays a role in EPS production (71, 79). It is possible that this may have an increase in biofilm formation of bacteria, though lower growth rates at lower temperatures may inhibit biofilms. This has also not been demonstrated in culture, thus the exact

effect on biofilm formation is yet to be quantified. In some Gram-negative bacteria, including several strains of *Xylella fastidiosa* and *Pseudomonas fluorescens,* greater concentrations of calcium and magnesium have been shown to increase biofilm formation (80, 81). It is possible that calcium and magnesium's positive charge as ions helps the extracellular matrix of biofilms to bind to negatively charged surfaces and aid in attachment (80, 81).

Due to the variability of factors such as temperature, nutrients, chemical components, etc. in soil environments, and the influence they may have on biofilm formation, it is important to consider the impact of these factors when analyzing biofilm formation as a survival trait for *E. coli* in soil environments. Such environmental variability may select differing traits among bacteria. However, genetic variation also plays a role in traits that affect biofilm formed and should be analyzed when determining the factors that affect biofilm formation in an environment.

Microbial GWAS and genotype-phenotype associations

Understanding the origins of phenotypic variation and the role it plays in the survival and persistence of an organism in an environment requires evaluating the genomic variability that leads to such differences in phenotypes. This can be done by determining genotype-phenotype associations, as found through a genome-wide association study (GWAS) (82, 83). GWAS starts with obtaining the genomic sequences of the organisms in question and annotating them, usually using an online service such as RAST, which utilizes a collection of known subsystems to compare against and annotate genomes (84). Further annotation of genomes usually involves identifying the core metabolic genes as well as genes that belong to the accessory genome. This can be done using a pan-genome analysis through software such as Roary. Roary takes an annotated assembly and converts it to protein sequences with partial sequences filtered out using

CD-HIT. An all-against-all BLASTP is run against the filtered protein sequences with a defined percentage sequence identity and are then clustered using MCL. Homologous groups determined to contain paralogs are separated into groups of true orthologs (85). Identification of single nucleotide polymorphisms (SNPs) and small insertion/deletion events (indels) can then be identified by comparing sequences against a reference. Programs that can be used to do this include GATK and kSNP3, which use different methods to call SNPs. GATK works by calling variants from genomes that have been mapped to a reference genome and then filtering out those variants that are not suitable for the data analysis, i.e. do not have reliable coverage, are at too high or too low frequencies, etc., in order to avoid false positives (86). kSNP3 works by identifying SNPs based on k-mer analysis. A central base k-mer frequency distribution is used to filter out k-mers that would result in allele conflicts. K-mers are then compared across genomes to find SNP loci (87).

Once variants have been called, they can be compared with a trait of interest to determine genotype-phenotype associations. The phenotypes of all isolates in the analysis must be measured under the same conditions before this can take place. Association analyses are dependent on the data to be analyzed. For normally distributed data, a Student's *t*-test or ANOVA can be used, and their respective non-parametric counterparts for non-normally distributed data (Mann-Whitney U and Kruskal-Wallis tests, respectively) (88). If the data is continuous, a Pearson's chi-squared test and Kendall tau rank correlation coefficient or Spearman's rank correlation coefficient is used for normally distributed and non-normally distributed data, respectively (88). If the number of genotypic variables is much larger than the number of isolates or strains, it is more suitable to use a machine learning algorithm to determine the importance of variants since there is a higher chance of fitting the function for individual

genomes too closely to the set data points used to set up the conceptual model for predicting outcomes, a process called overfitting or overtraining the data, and ends up decreasing its predictive power (88, 89). Machine learning methods, such as random forest, work by using data to 'train' the algorithm to predict and analyze the importance of genotypic variables (90). That importance can then be used to filter out variables that do not directly contribute to accurately predicting the phenotype.

In the past, GWAS has been used to determine the correlation of genetic variants with specific traits in individuals. Due to the effect these genetic variants have on genotypes and their subsequent phenotypes, they are important to identify when determining the roles genetics plays in specific traits in bacteria. An example would be how mutations lead to changes in the genome and bacteria traits. Mutations can either change the amino acid sequence (non-synonymous) or not change the amino acid sequence (synonymous). Synonymous mutations change codons and subsequent tRNA used for peptide formation, which can impact gene expression due to differences in concentrations of different tRNAs in the cell. Mutations can be either beneficial, harmful, or have no overall effect. Such genetic changes can lead to changes in the phenotype, possibly increasing adaptability through those traits. An example of this can be seen in a study where GWAS was used to determine a mutation of the *rpoB* gene in *Staphylococcus aureus* that increased vancomycin resistance (91). Another example includes a study that analyzed *Streptococcus pneumoniae* beta-lactam antibiotic resistance, where it was determined through GWAS several SNPs and indels that conferred beta-lactam non-susceptibility. These included genes involved in peptidoglycan biosynthesis, penicillin binding proteins, and cell wall biogenesis (92). Other genetic variants include indels, which can lead to frameshifts in the genetic code. Such events can change large swaths of the amino acid sequence, possibly leading

to a change in the expression of the gene, such as a loss of functionality. Gene acquisition/loss events can also have an impact on subsequent phenotypes. For example, in a study on host specificity in *Campylobacter sp.*, it was found that host adaptation was influenced by the gain/loss of the *panBCD* gene that encode vitamin B_5 biosynthesis (93). Other examples where GWAS has been utilized as a means of determining the effects of genetic variants on phenotypes include a study that determined antibiotic resistance traits in *Mycobacterium tuberculosis*, as well as studies analyzing associations in other species, like artemisinin-drug resistance in *Plasmodium falciparum* and heart-disease in humans (83, 94, 95).

Associations between genotypes and phenotypes of interest can be determined if the phenotypes can be adequately characterized. This includes antibiotic resistance, association with disease, and survival traits such as biofilm formation. For biofilms, it could be possible that genetic variations in traits that are directly related to biofilm structure and regulation may have the greatest impact on differences in phenotypes, such as those involved in adhesion and motility. In a GWAS study on *Campylobacter jejuni* biofilm formation, genes that were found to have a robust association with biofilm formation included genes involved in adhesion and motility, as well as glycosylation, capsule production, and oxidative stress (96).

In each of these examples of GWAS studies, the analysis was based on categorical phenotypes. Analysis can be conducted with continuous phenotypes as well, through the use of a logistic regression analysis to predict the importance of variants and their effect on the phenotype of interest (97). Using a GWAS analysis, the genetic variants that play the greatest role in biofilm formation among *E. coli* could be analyzed, as this has yet to be determined.

Rationale

Due to the role biofilms may have on the persistence of *E. coli* in soil and the possibility of reservoirs for virulence and antibiotic resistance, it is important to understand the mechanisms that take place in the formation of biofilms in soil environments. Since genetic variation plays a significant role in the variation of traits that may increase or decrease biofilm formation, and thus survival and persistence of *E. coli*, it is important to determine the genetic variability between *E. coli* isolates and the differences in phenotypes that comes from that variability. Thus, the main question we wanted to address was, "What is the variability that exists in biofilm formation of phylogroup D *E. coli* isolates from surface soils in terms of genotypes and their associated phenotypes and how does this variability affect biofilm formation in these isolates?" To do this, we will measure biofilm density of a collection of phylogroup D *E. coli* isolates that have been sampled from soils and conduct a GWAS analysis to determine what genetic variants play the greatest role in biofilm formation. Also, since specific environmental conditions may influence biofilm formation, we will analyze any effect the conditions of the soil environment from which the isolates were sampled may play in influencing biofilm formation. One specific condition we are most interested in is the lower temperature of soil environments and how this may influence biofilm formation and subsequent survival. Thus, we ask what is the difference in biofilm density at optimal and low temperatures and what genetic variants play the greatest role in biofilm formation at low temperatures?

CHAPTER 2: BIOFILM FORMATION OF *ESCHERICHIA COLI* FROM SURFACE SOILS IS INFLUENCED BY VARIATION IN CELL ENVELOPE, IRON METABOLISM, AND ATTACHMENT FACTOR GENES

Introduction

It has been suggested that extrahost environments select for persistence in *E. coli*, and the variability of these environments may be a key factor in the biodiversity of the species (49, 98, 99). The survival of *E. coli* in extrahost environments, such as surface soils, can be influenced by the ability to form biofilms. Biofilms are produced in response to stressful conditions, such as those found in soil environments, including low temperatures, extreme pH levels, and to protect against predation, toxins, and aid in competition against other microbes (3, 54, 64, 100, 101). Biofilms have been shown to increase survival in extrahost environments, as seen in increased persistence of *E. coli* O157:H7 in sand with the formation of biofilms (102) and increased resistance to bacteriophages and amoeba predation in open environments, or environments that are influenced by its surroundings and factors moving into and out of the system (103).

E. coli phylogroups (A, B1, B2, C, D, E, F, and Clade I) (7, 33) vary by the habitats from where they are commonly isolated, such as soils (phylogroup B1) (3), surfaces of plants (phylogroups B1 and D) (37), surface water (phylogroup A) (36), and human hosts (phylogroups B2 and A) (1). Variation in biofilm production is associated with the source of isolation. For example, a study on biofilm formation from animal-associated and plant-associated *E. coli* isolates found that the plant-associated isolates produced greater biofilms than animal-associated isolates (37). Thus, extrahost environments select for traits that increase biofilm formation in *E. coli*. This is important since soil environments can become reservoirs for transmission of virulence between strains (8). Virulent *E. coli* strains from phylogroup D have been commonly

found among urinary tract infections as well as bacteraemia infections, making it a phylogroup of interest (39, 40). Soil environments have also shown to have the greatest influence on the genetic structure of phylogroup D *E. coli* than any other phylogroup (3), which could possibly give them an advantage to surviving in soils.

Although it has been well established that *E. coli* produces biofilms to survive abiotic and biotic environmental stressors, it remains unknown the extent to which variation in biofilm formation among isolates affects survival in soil environments. We sought to answer the research question of what genetic traits increase or decrease biofilm formation in phylogroup D *E. coli* isolates from surface soils, due to the biodiversity that soil environments select for, and is biofilm formation associated with the soil environment? Also, what effect does variability in soil environments have on biofilm production? To answer this question, we conducted environmental and genome-wide association studies (GWAS) with the following objectives: to determine the trait diversity in biofilm formation among phylogroup D *E. coli* isolates, to compare biofilm phenotypes with soil environmental conditions, and to determine what genetic variants played the greatest role in biofilm formation among the isolates.

Results

Biofilm formation of phylogroup D soil isolates.

A crystal violet assay was used to determine the variation in biofilm formation of each of the phylogroup D *E. coli* soil isolates at the optimal growth temperature of 37°C. Median biofilm density (OD_{600}) was 0.4197, with a minimum value of 0.03843 and maximum of 2.013 and a coefficient of variation of 65.64% among replicates, indicating a higher level of trait diversity among isolates (**Fig 1)**. Since these isolates show such variability in biofilm density, it is possible that there is a presence of adaptive variants that are playing a role in increasing biofilm formation

and subsequently leading to a possible increase in the chance of survival for *E. coli* in the soil environment.

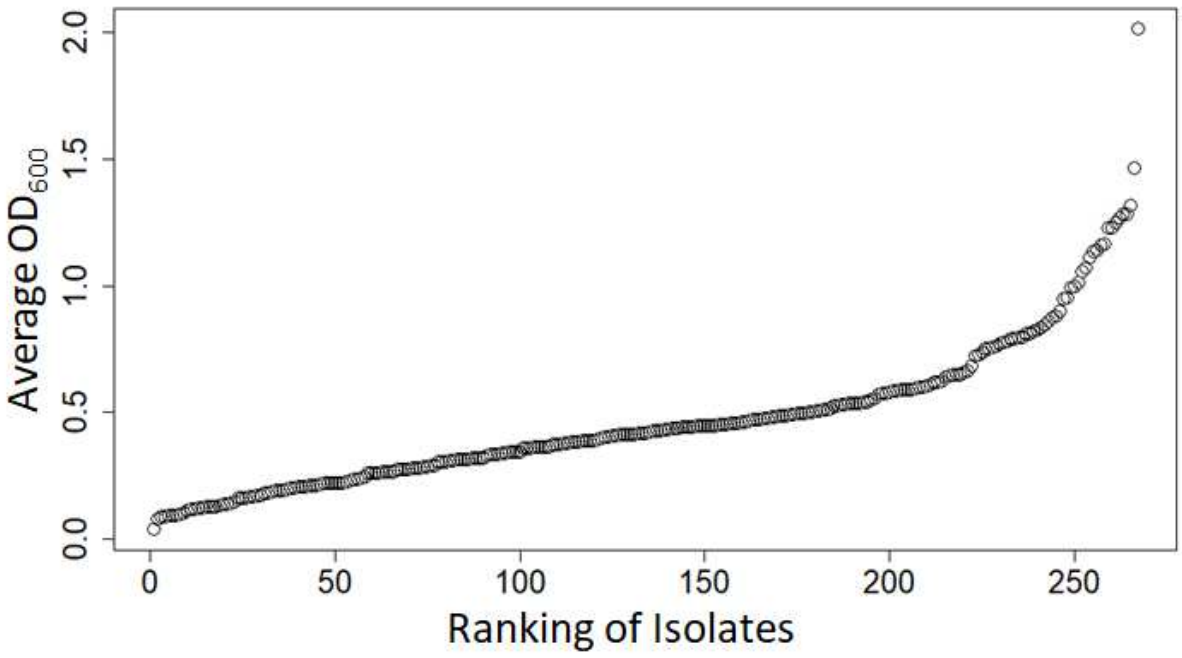

Figure 1: Rank plot of average biofilm density of isolates grown at 37°C. Biofilm density varied among isolates, indicating the possible presence of adaptive variants that are playing a role in increasing biofilm formation.

It was hypothesized that biofilm formation at 15°C, which was the average soil temperature of the environment from which the *E. coli* samples were collected, would be greater than biofilm formation at 37°C. A crystal violet assay was used to measure biofilm density of the *E. coli* isolates at 15°C. Median biofilm density (OD_{600}) was 0.3333, with a minimum of 0.0363 and maximum of 2.8768 and a coefficient of variance among replicates of 119.34%, again indicating a higher level of trait diversity among isolates. Biofilm density at 37°C and 15°C was compared with each other to determine under which conditions biofilm formation was greater. Homogeneity of variance was measured using a Levene's Test, with variance being greater at 15°C (p-value = 1.056e-06), suggesting that 15°C is a more stressful environment than 37°C. A

paired Mann-Whitney-Wilcoxon Test was used for the comparison of biofilm density at 15°C and 37°C. Isolates grown at 15°C showed no significant difference in median biofilm density than isolates grown at 37°C (p-value = 0.574). This suggested that 15°C does not affect biofilm density in comparison with 37°C.

Biofilm density was normalized for growth OD_{600} to analyze the effect growth had on biofilm density. Isolates were grown for the same amount of time as for the CV assay, and OD_{600} measured to determine growth. A paired Mann-Whitney-Wilcoxon test was used to compare median growth at 37°C and 15°C, with 37°C having about 44% higher median growth (p-value < 2.2e-16). A paired Mann-Whitney-Wilcoxon Test was used for the comparison of the ratio of biofilm density to growth at 15°C and 37°C, with the median ratio about 21% greater at 15°C than at 37°C (p-value =1.378e-07) (**Fig 2**). Therefore, growth is less at 15°C compared to 37°C, but the ratio of biofilm to growth is greater. Thus, isolates are growing less at 15°C but producing more biofilm per unit growth, indicating that 15°C is selecting for greater biofilm density per unit growth among our isolates, possibly due to the stress of low temperature.

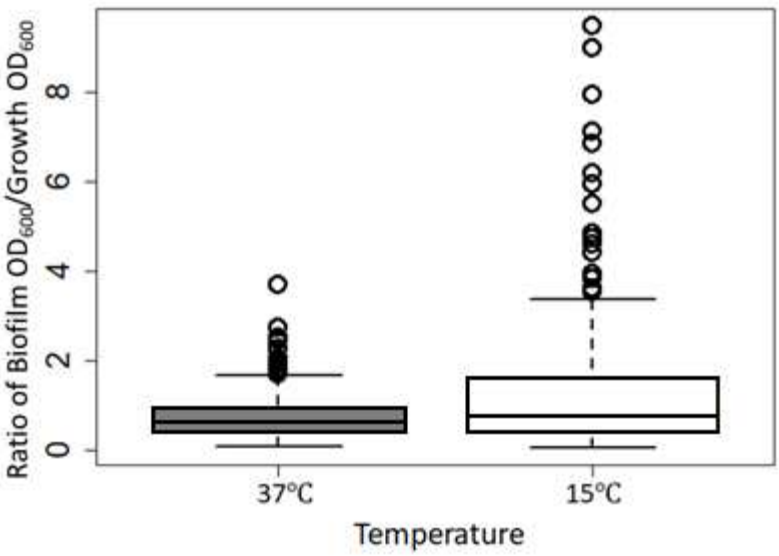

Figure 2: Comparison of ratio of biofilm density to growth between phylogroup D *E. coli* isolates from soil at 37°C and 15°C. Biofilm density was found to be not significantly different at 37°C and 15°C (p-value = 0.574) with growth about 44% greater at 37°C (p-value < 2.2e-16), thus comparative biofilm density to growth was about 21% greater at 15°C (p-value =1.378e-07). This indicates that isolates were growing less at 15°C but producing more biofilm per unit growth.

Viable plate counts were performed on 48 randomly selected isolates from the biofilms formed at both 37°C and 15°C. The number of colony forming units (CFU) per mL was log transformed before comparing the average $\log_{10}$ CFU mL^{-1} at 37°C and 15°C using a paired t-test. There was no significant difference in the average $\log_{10}$ CFU mL^{-1} of isolates at 37°C and 15°C (p-value = 0.8999) (**Fig 3**). Equal viable cell counts but less growth at 15°C suggests that cells in culture at 15°C may be smaller in size than at 37°C.

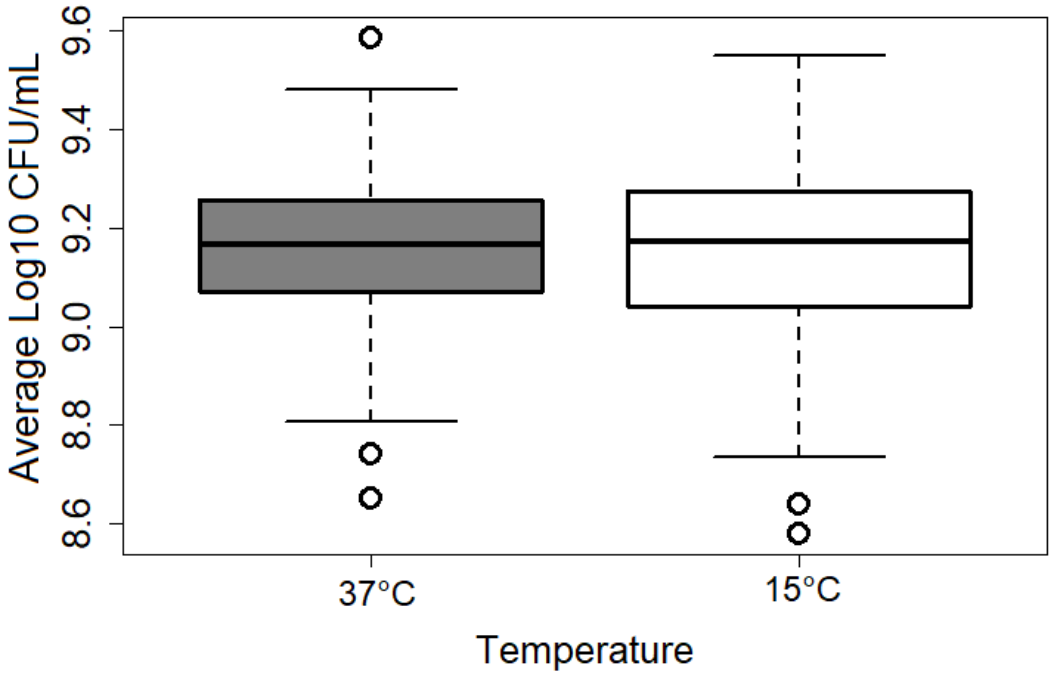

Figure 3: Comparison of average $\log_{10}$ CFU mL^{-1} between phylogroup D *E. coli* isolates at 37°C and 15°C. Viable plate counts found no significant difference in the average $\log_{10}$ CFU mL^{-1} of isolates at 37°C and 15°C (p-value = 0.8999). Equal viable cell counts but less growth at 15°C suggests that cells in culture at 15°C may be smaller in size than at 37°C.

Associations between biofilm density phenotypes and soil environment characteristics.

Soil environment characteristics were compared with biofilm density to determine any possible effects the environment may have had on biofilms. Of the different soil characteristics analyzed, only soil moisture, pH, and calcium levels were significantly associated with biofilm density. Soil moisture, with levels categorized as dry, moist, or saturated, was found to tend to increase as biofilm density increased (**Fig 4**). A Chi-square test showed that when comparing soil moisture with biofilm density groups, there was a significant difference in the biofilm density groups among the dry soils and the moist and saturated soils (p-value = 0.02446), with biofilm density groups tending to be greater, respectively. Average biofilm density of isolates from the moist and saturated soils was about 19% greater than average biofilm density of isolates from the

dry soils. This shows that biofilm formation is lower in isolates from dry soils, suggesting that moist soils select for greater biofilm formation.

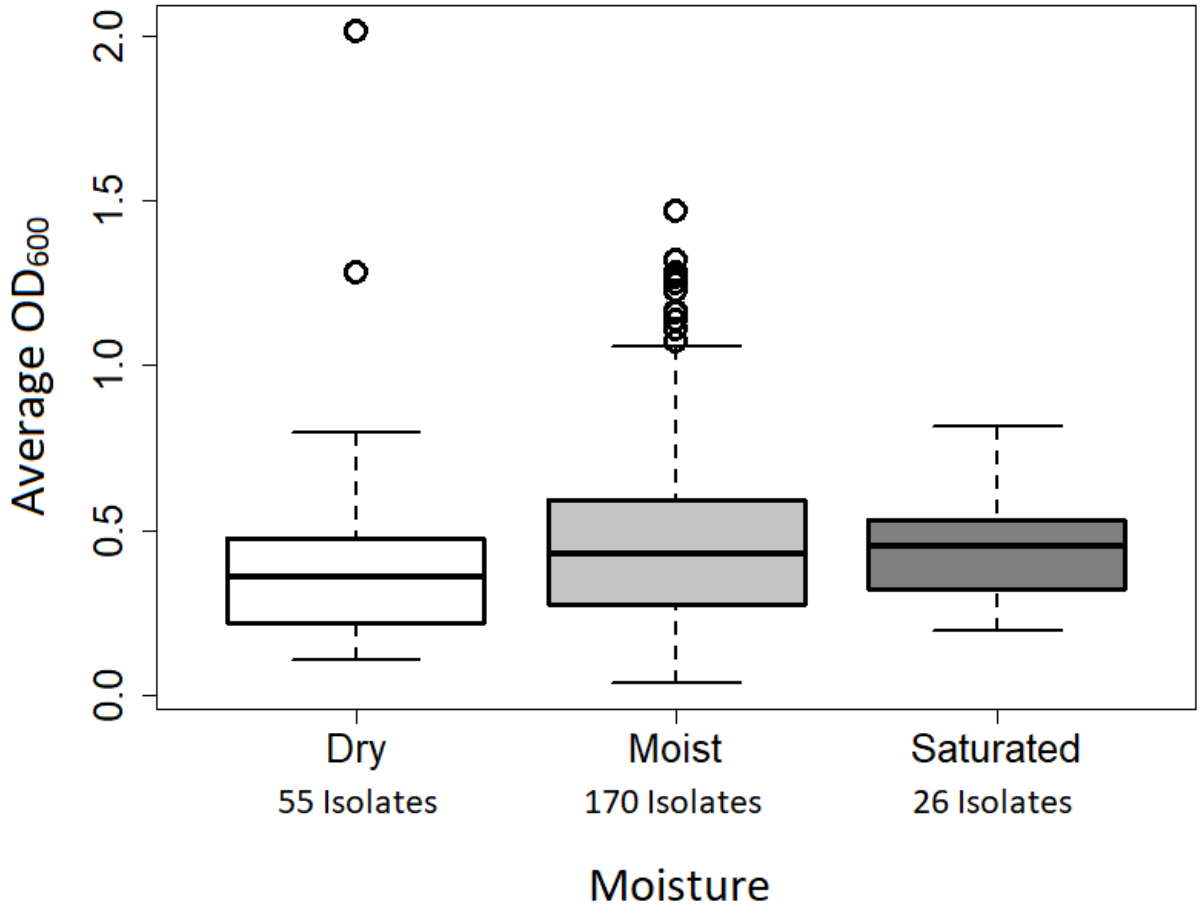

Figure 4: Comparison of biofilm densities of phylogroup D *E. coli* isolates among various soil moisture levels. Soil moisture was found to tend to increase as biofilm density increased. When comparing soil moisture with biofilm density groups, there was a significant difference in the biofilm density groups among the dry soils and the moist and saturated soils (p-value = 0.02446), with biofilm density groups tending to be greater, respectively. The average biofilm density of isolates from the moist and saturated soils was about 19% greater than average biofilm density of isolates from the dry soils. This suggests that biofilm formation is lower in isolates from dry soils.

A cumulative logistic regression was used to determine the change in frequency of biofilm density groups when compared to calcium and pH levels of the soils from which the isolates were sampled. It was found that, as pH levels increased, the frequency of higher biofilm

density groups, and subsequently greater biofilm formers, increased as well (p-value < 0.05) (**Fig 5A**). This was found with soil calcium as well (p-value < 0.05) (**Fig 5B**). Soil pH and calcium levels were moderately correlated with each other (Pearson correlation test, r=0.42, p-value < 0.05) (**Fig 5C**). These results suggested that soils with higher pH and calcium levels select for greater biofilm formation.

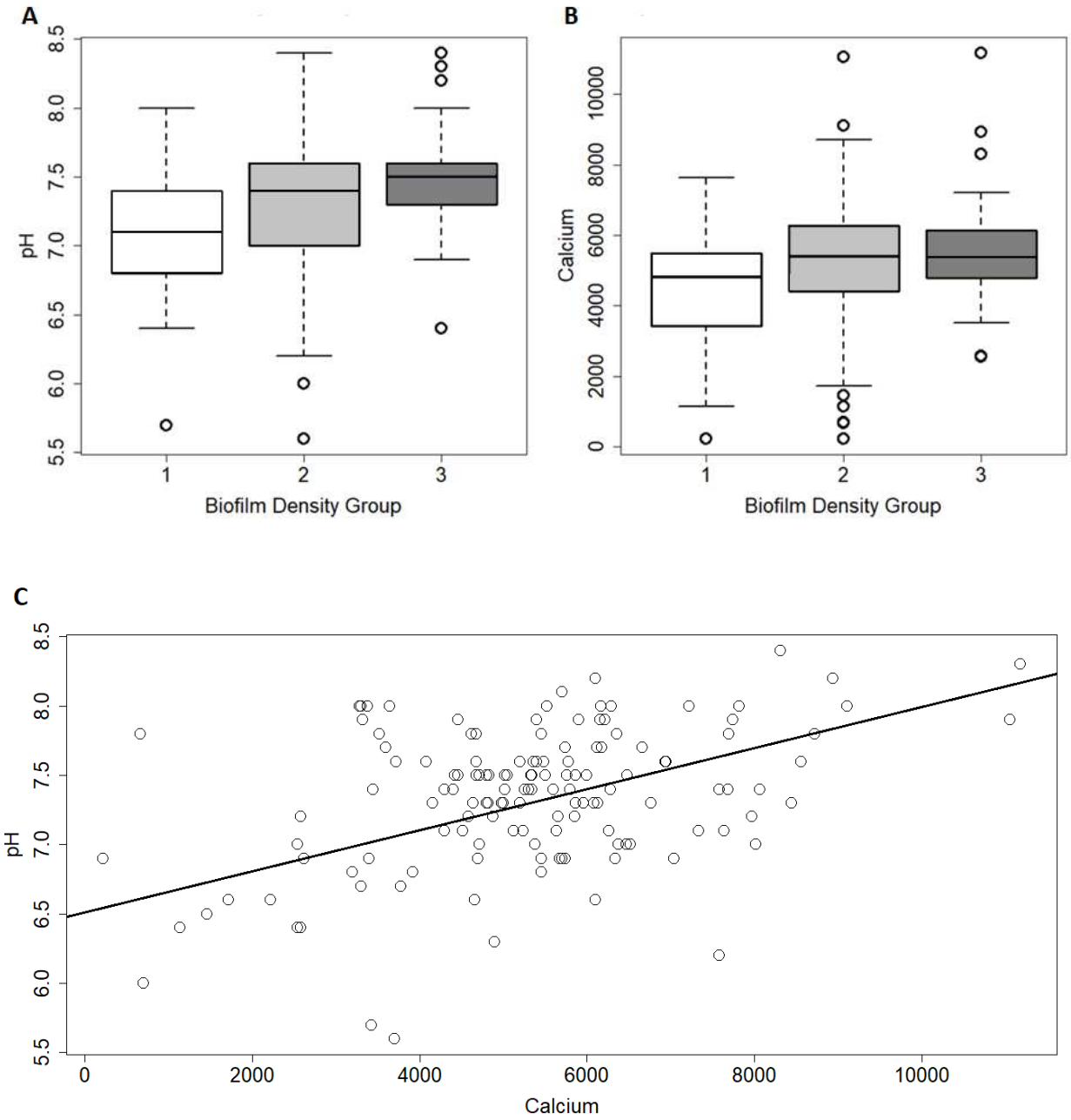

Figure 5: Comparison of soil pH and calcium level to biofilm density group of phylogroup D *E. coli* isolates. A) As pH levels increased, the frequency of higher biofilm density groups, and subsequently greater biofilm formers, increased as well (p-value < 0.05). B) As calcium levels increased, the frequency of higher biofilm density groups, and subsequently greater biofilm formers, increased as well (p-value < 0.05). C) Soil pH and calcium levels were moderately correlated with each other (Pearson correlation test, r=0.42), which may suggest that only one factor is ultimately leading to the increase in biofilm formation.

Different environments are known to have an impact on biofilm formation, so the biofilm densities of phylogroup D *E. coli* isolates from fecal and soil samples were compared (**Fig 6**). A CV assay was conducted on 96 randomly chosen fecal isolates and 82 randomly chosen soil isolates at 37°C and the OD_{600} values compared. We found that isolates taken from soil samples were on average 18% greater than fecal isolates (t-test, p-value = 0.04302). This suggests that soil environments may be selecting for greater biofilm formation. However, when biofilm density of the soil and fecal samples were compared at 15°C, there was no significant difference (t-test, p-value = 0.337). So, while the soil isolates may have an adaptive advantage with biofilm formation over the fecal isolates, it is only at 37°C.

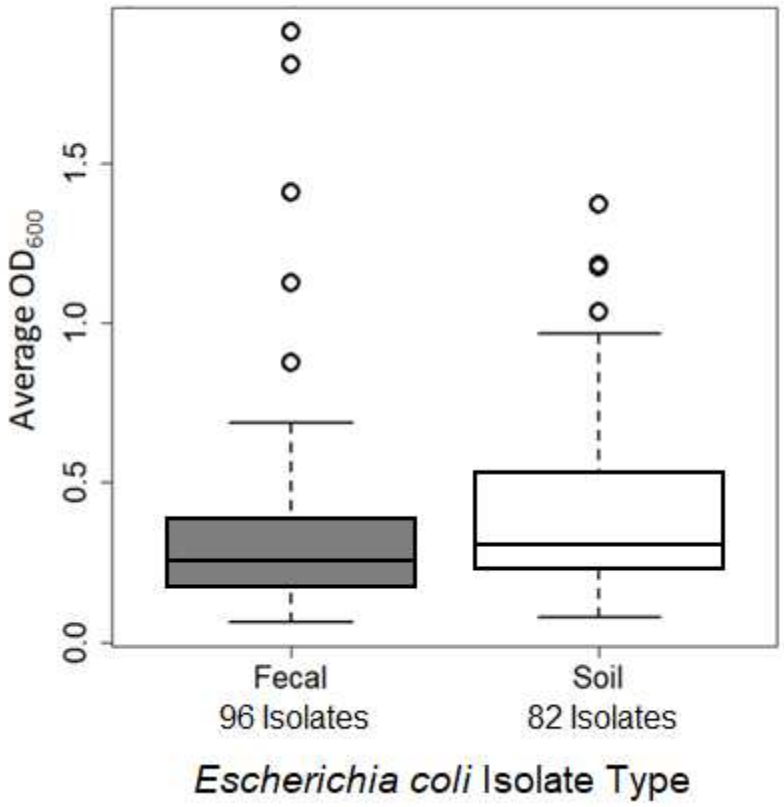

Figure 6: Comparison of biofilm density between phylogroup D *E. coli* isolates from fecal and soil samples. Isolates taken from soil samples were on average 18% greater than fecal isolates (p-value = 0.04302), suggesting that soil environments may be selecting for greater biofilm formation.

GWAS.

Differences in genetic variants conferring a specific trait play a role in variation of phenotypes and may aid or be detrimental to bacteria in an environment. A genome-wide association study was conducted to determine which genetic variants among the phylogroup D *E. coli* genomes had the greatest effect on biofilm formation. Variants were first determined using GATK and kSNP3. After comparing agreements in SNP identification among the two programs and filtering out redundant or extremely rare variants that would be prone to promote false

positives, GATK and kSNP3 agreed on 10,753 SNPs from our genomes. A Jaccard Similarity test was used to cluster variants that were similarly present/absent across the genomes into groups termed 'mosaics'. A total of 1,316 mosaics were identified among our SNPs. Logistic regression was used to determine the significance of SNPs and mosaics on their effect on biofilm formation at 15°C and 37°C. The phylogeny on core genome alignment using ClonalFrameML identified a total of seven different phylogroup D clades, or a group of genetically similar isolates sharing a common early divergence point, among the 277 *E. coli* genomes analyzed (**Fig 7**). This was used in the logistic regression to correct for phylogenetic distance between isolates. After filtering out variants based on FDR q-values of < 0.05, a total of 91 and 577 significant variants were found for biofilm formation at 37°C and 15°C, respectively. Of those variants, a total of 13 mosaics were found to be significant for 37°C and 43 mosaics for 15°C.

Biofilm formation of the isolates varied between clades. The greatest variation existed with clades 1 and 3. Clade 1 had the greatest proportion of low biofilm formers at 37°C, with 26% of the isolates categorized as low biofilm formers. Clade 3 had the greatest proportion of high biofilm formers at 37°C, with 38% of the isolates categorized as high biofilm formers.

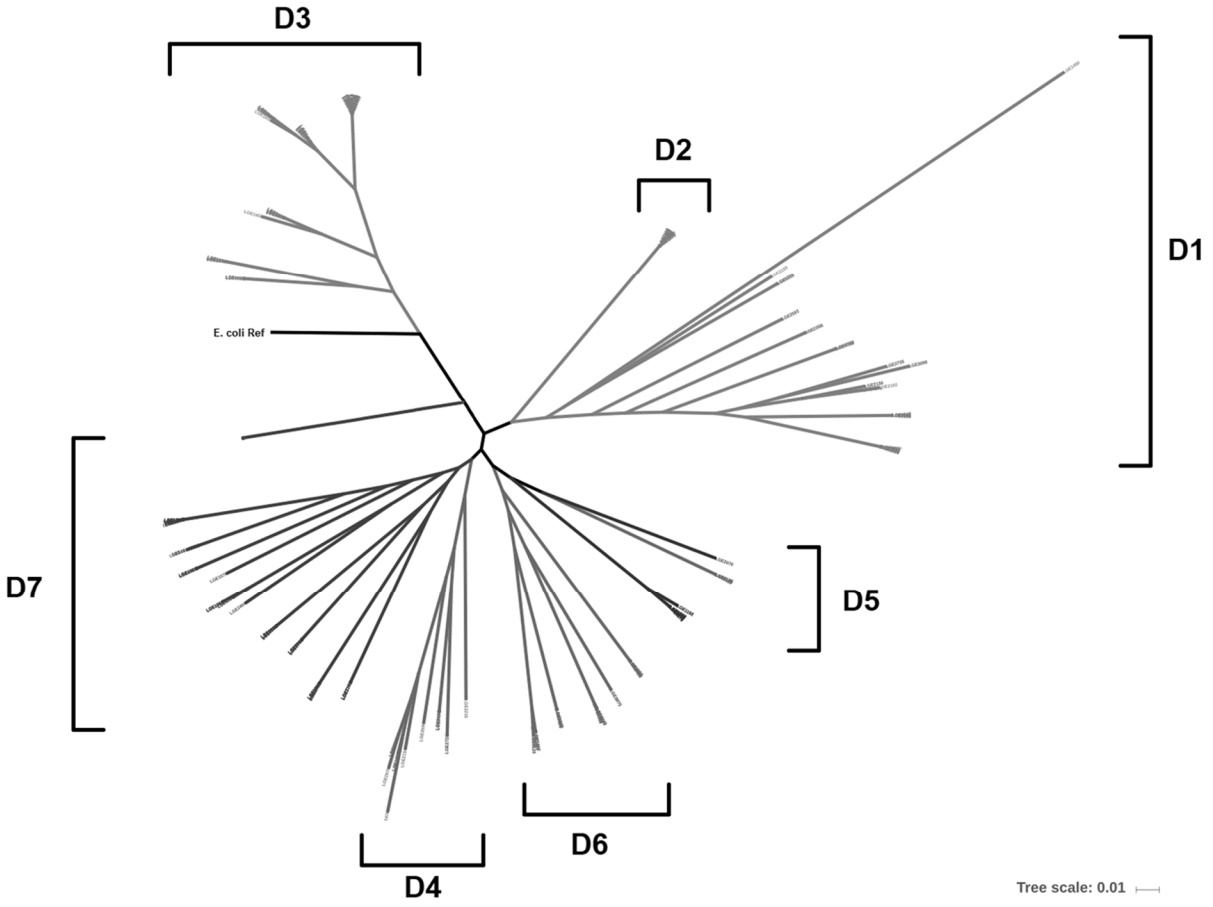

Figure 7: Phylogeny of the phylogroup D *E. coli* genomes based on core genome relatedness. Of the phylogroup D *E. coli,* a total of seven different clades were identified. Phylogeny tree was generated using RAxML and the figure generated using iTOL.

The presence or absence of genetic variants that were associated with biofilm formation differed between clades, especially clades 1 and 3. For variants associated with an effect at 37°C, clade 1 had more variants that were associated with a negative effect on biofilm density than a positive effect (**Fig 8**). Clade 3 had more variants that were associated with a positive effect on biofilm density than a negative effect (Fig 8). For variants associated with an effect on biofilm density at 15°C, clade 3 could be separated into two different subgroups (**Fig 9**). One group consisted of isolates that had most of the variants associated with a negative effect on biofilm

density and generally had lower biofilm phenotypes. The other group had most of the variants associated with a positive effect on biofilm density and generally had higher biofilm phenotypes.

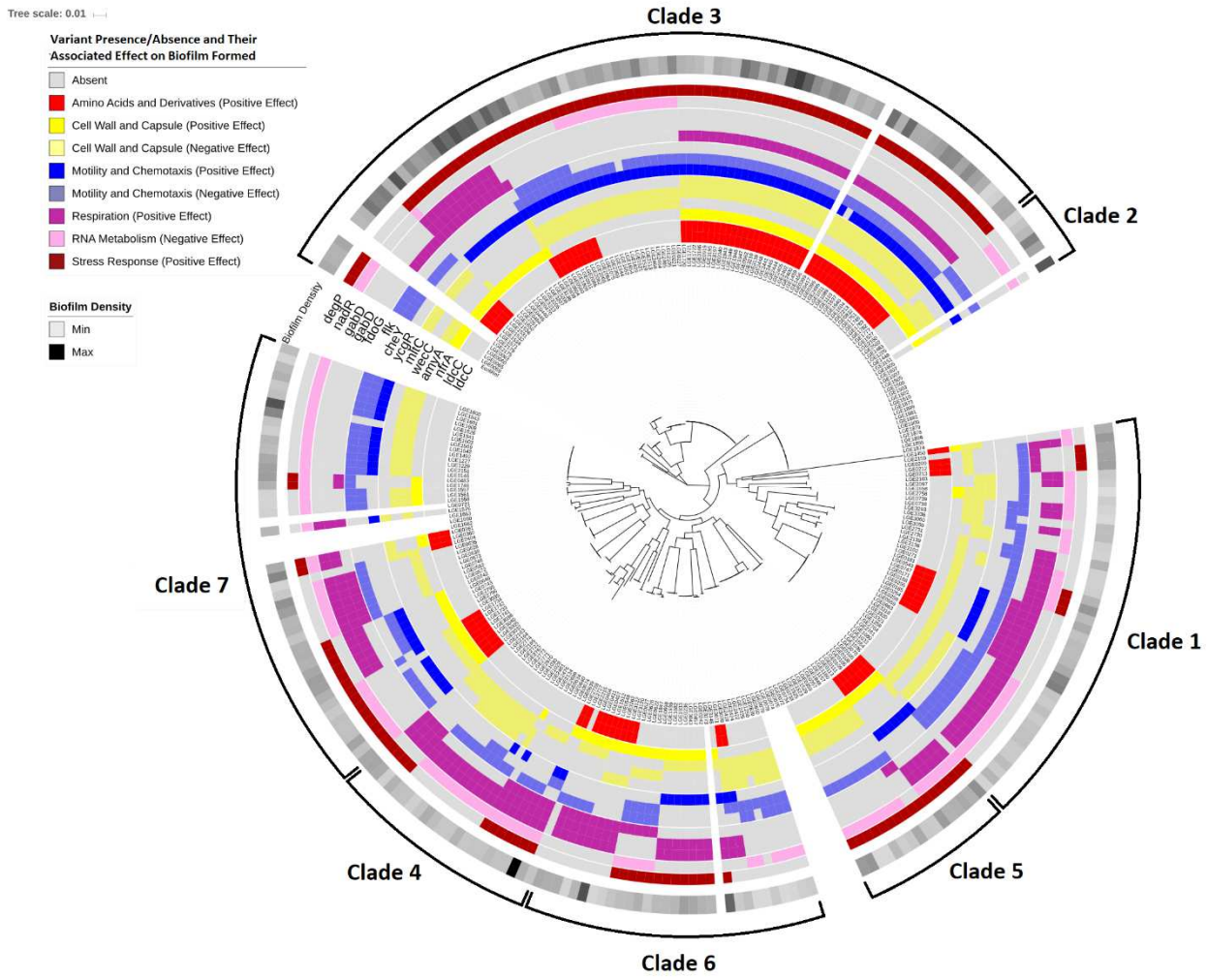

Figure 8: Difference in presence/absence of variants of possible importance for biofilm formation at 37°C among clades. Core gene phylogeny generated by RAxML of the phylogroup D isolates is shown in the tree in the middle. The presence/absence of variants and their associated function and effect on biofilm formed is shown for all clades. The color scale on the outermost ring represents the effect of the variant on biofilm density, with darker color representing a greater effect. Clade 1 had the greatest proportion of low biofilm formers at 37°C, with 26% of the isolates being categorized as low biofilm formers. Clade 3 had the greatest proportion of high biofilm formers at 37°C, with 38% of the isolates being categorized as high biofilm formers. The figure was generated using iTOL.

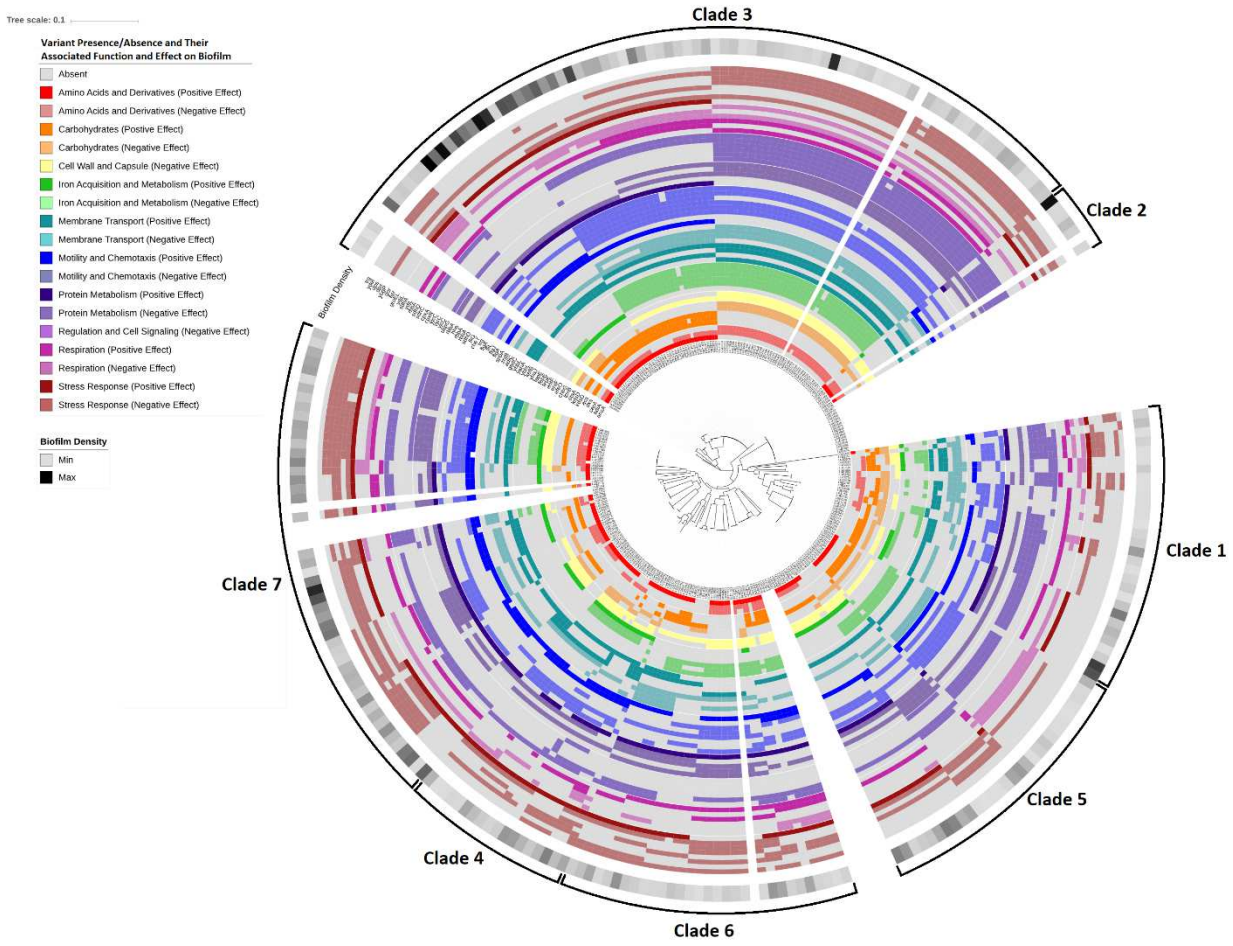

Figure 9: Presence/absence of variants of possible importance for biofilm formation at 15°C among clades. Core gene phylogeny generated by RAxML of the phylogroup D isolates is shown in the tree in the middle. The presence/absence of variants and their associated function and effect on biofilm formed is shown for all clades. The color scale on the outermost ring represents the effect of the variant on biofilm density, with darker color representing a greater effect. For variants associated with an effect on biofilm density at 15°C, clade 3 could be separated into two different subgroups. One group consisted of isolates that had most of the variants associated with a negative effect on biofilm density and generally had lower biofilm phenotypes. The other group had most of the variants associated with a positive effect on biofilm density and generally had higher biofilm phenotypes. The figure was generated using iTOL.

A total of 46 variants were statistically significant for having an association with biofilm density at 37°C (**Table A1**). Of these variants, 2 were involved in cell division, 3 involved in DNA repair, 9 involved in cell envelope formation and structure, 14 involved in biosynthesis of various metabolites, 3 involved in motility, 5 involved in replication, transcription, or translation,

1 involved in respiration, 3 involved in stress response, and 6 involved in transportation. The most important variants are shown in Table 1.

Table 1: List of variants that are most important for biofilm formation at 37°C.

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
nfrA	Bacteriophage N4 receptor, outer membrane protein	695019	Synonymous	Central metabolism	0.222	Membrane/Cell Wall
amyA	Alpha-amylase	2240378	Synonymous	May reduce biofilm formation by degrading EPS	-0.313	Membrane/Cell Wall
wecC	UDP-N-acetyl-D-mannosamine dehydrogenase	4458415	Synonymous	Membrane biogenesis	-0.314	Membrane/Cell Wall
mltC	Membrane-bound lytic murein transglycosylase C	3440095	Synonymous	Membrane biogenesis	-0.358	Membrane/Cell Wall
ldcC	Lysine decarboxylase 2	213544	Missense	Helps regulate periplasmic pH	0.288	Metabolite Biosynthesis
ldcC	Lysine decarboxylase 2	213555	Synonymous	Helps regulate periplasmic pH	0.288	Metabolite Biosynthesis
gabD	Succinate-semialdehyde dehydrogenase (NADP)	3095058	Synonymous	Upregulation of succinate dehydrogenase leads to increase in biofilm formation	0.243	Metabolite Biosynthesis
gabD	Succinate-semialdehyde dehydrogenase (NADP)	3095001	Synonymous	Upregulation of succinate dehydrogenase leads to increase in biofilm formation	0.243	Metabolite Biosynthesis
ycgR	Molecular brake that regulates flagellar motility in response to c-di-GMP	1492833	Intergenic Region ycgR-ymgE	Decreases motility	0.300	Motility
cheY	Chemotaxis binding protein	2201461	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.248	Motility
flk	Putative flagella assembly protein	2737533	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.257	Motility
nadR	NadR DNA-binding transcriptional repressor and NMN adenylyltransferase	5184339	Synonymous	May affect expression	-0.328	Replication/ Transcription/ Translation
fdoG	Formate dehydrogenase-O, alpha subunit	4575196	Synonymous	Energy metabolism	0.471	Respiration
degP	Serine protease Do	185468	Synonymous	Increases biofilm formation through upregulation of Cpx signal transduction pathway	0.200	Stress Response

For cell envelope formation and structure, enzymes involved in the degradation of extracellular polymeric substances (EPS), such as amylase encoded by *amyA*, have been shown to inhibit or disrupt biofilms, leading to a decrease in biofilm formation (104, 105). Some variants, such as *degP* and *gabD*, are indirectly associated with biofilm formation via the RpoS regulon. Overall, the results suggested that cell envelope formation and structure and central metabolism had the greatest effects on biofilm density at 37°C.

For association with biofilm density at 15°C, a total of 204 variants were statistically significant (**Table A2**). Of these variants, 7 were involved in biofilm formation, 7 involved in cell division, 6 involved in DNA repair, 6 involved in iron metabolism, 24 involved in cell envelope formation and structure, 49 involved in biosynthesis of various metabolites, 8 involved in motility, 1 involved in recombination, 16 involved in replication, transcription, or translation, 18 involved in respiration, 33 involved in stress response, 5 involved in the TCA cycle, and 24 involved in transportation. The most important variants are listed in Table 2.

Table 2: List of variants that are most important for biofilm formation at 15°C.

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
ybgP	Putative fimbrial chaperone	849172	Synonymous	Involved in fimbrial attachment	0.202	Biofilm Formation
yahA	Phosphodiesterase, c-di-GMP-specific	386338	Synonymous	May decrease biofilm formation by decreasing c-di-GMP	-0.125	Biofilm Formation
yjcC	Putative c-di-GMP-specific phosphodiesterase	4769269	Missense	Overexpression reduces biofilm formation	-0.131	Biofilm Formation
yjcC	Putative c-di-GMP-specific phosphodiesterase	4769251	Missense	Overexpression reduces biofilm formation	-0.131	Biofilm Formation
yjcC	Putative c-di-GMP-specific phosphodiesterase	4769357	Synonymous	Overexpression reduces biofilm formation	-0.131	Biofilm Formation
hha	Haemolysin expression modulating protein	538736	Intergenic Region maa-hha	Decreases biofilm formation by inhibiting fimbrial genes	-0.192	Biofilm Formation
efeO	Iron uptake system component	1257333	Synonymous	Involved in iron uptake and homeostasis	0.159	Iron Homeostasis
entF	Enterobactin synthase component F	730487	Missense	Involved in iron uptake and homeostasis	-0.080	Iron Homeostasis
entF	Enterobactin synthase component F	729500	Missense	Involved in iron uptake and homeostasis	-0.090	Iron Homeostasis
fepE	Ferric enterobactin (enterochelin) transport	733649	Synonymous	Involved in iron uptake and homeostasis	-0.107	Iron Homeostasis
fepE	Ferric enterobactin (enterochelin) transport	733424	Synonymous	Involved in iron uptake and homeostasis	-0.115	Iron Homeostasis
fepE	Ferric enterobactin (enterochelin) transport	734065	Missense	Involved in iron uptake and homeostasis	-0.118	Iron Homeostasis
chbG	Chito-oligosaccharide mono-deacetylase	2050963	Synonymous	Involved in chitin and glycan degradation, by use of hydrolase	-0.433	Membrane/Cell Wall
dxs	1-Deoxyxylulose-5-phosphate synthase	497472	Synonymous	Central metabolism	0.313	Metabolite biosynthesis
ansA	Asparaginase I	2083908	Intergenic Region sppA-ansA	Central metabolism	0.284	Metabolite biosynthesis
sppA	Protease IV	2083908	Intergenic Region sppA-ansA	Central metabolism	0.284	Metabolite Biosynthesis
acs	Acetyl-CoA synthetase (AMP-forming)	4779477	Synonymous	Central metabolism	0.269	Metabolite biosynthesis
kdsD	D-Arabinose 5-phosphate isomerase	3795168	Synonymous	Membrane biogenesis	0.206	Metabolite biosynthesis
ydjL	Putative oxidoreductase, Zn-dependent and NAD(P)-binding	2094898	Intergenic Region ydjL-yeaC	Central metabolism	-0.111	Metabolite biosynthesis

Table 2: List of variants that are most important for biofilm formation at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
adiA	Biodegradative arginine decarboxylase	4826014	Synonymous	Plays role in regulating intracellular pH	-0.142	Metabolite biosynthesis
kdsD	D-Arabinose 5-phosphate isomerase	3794898	Synonymous	Membrane biogenesis	-0.219	Metabolite biosynthesis
otsA	Trehalose-6-phosphate synthase	2213757	Synonymous	Central metabolism	-0.256	Metabolite biosynthesis
lpxB	Lipid A disaccharide synthase	206549	Synonymous	Membrane biogenesis	-0.322	Metabolite Biosynthesis
carA	Carbamoyl phosphate synthetase	34921	Synonymous	Pyrimidine synthesis	-0.328	Metabolite biosynthesis
ydjL	Putative oxidoreductase, Zn-dependent and NAD(P)-binding	2093916	Synonymous	Central metabolism	-0.360	Metabolite biosynthesis
trg	Methyl accepting chemotaxis protein - ribose/galactose/glucose sensing	1686707	Missense	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	0.196	Motility
flgA	Flagellar biosynthesis; assembly of basal-body periplasmic P ring	1295248	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	0.179	Motility
flgJ	Peptidoglycan hydrolase	1302736	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.110	Motility
fliL	Flagellar biosynthesis	2253464	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.123	Motility
flgA	Flagellar biosynthesis; assembly of basal-body periplasmic P ring	1295200	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.124	Motility
trg	Methyl accepting chemotaxis protein - ribose/galactose/glucose sensing	1686176	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.173	Motility
cheY	Chemotaxis binding protein	2201446	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.176	Motility
fliG	Flagellar motor switch protein FliG	2249437	Synonymous	Upregulation of motility has negative effect on biofilm formation after start of maturation of biofilm	-0.191	Motility
rbbA	Ribosome-associated ATPase	4074691	Synonymous	May play a role in regulating genes linked to membrane, protein synthesis and energy metabolism	-0.188	Replication/ Transcription/ Translation
rbbA	Ribosome-associated ATPase	4074681	Synonymous	May play a role in regulating genes linked to membrane, protein synthesis and energy metabolism	-0.188	Replication/ Transcription/ Translation
rbbA	Ribosome-associated ATPase	4076557	Synonymous	May play a role in regulating genes linked to membrane, protein synthesis and energy metabolism	-0.273	Replication/ Transcription/ Translation
groEL	Heat shock protein	4856787	Synonymous	Protects against oxidative stresses and essential for growth	0.258	Stress Response
cpxR	Transcriptional regulatory protein	4590896	Synonymous	Induced during cell envelope stress, important for biofilm formation, but induction by stress may indicate conditions that decrease biofilms	-0.157	Stress Response

Table 2: List of variants that are most important for biofilm formation at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
asr	Acid shock protein	1904826	Intergenic Region ynfM-asr	Involved in acid resistance, but may not be beneficial at low temperatures	-0.159	Stress Response
gor	Glutathione reductase (NADPH)	4094001	Synonymous	Protects against oxidative deleterious reactions	-0.160	Stress Response
ycfR	Multiple stress resistance protein	1335677	Synonymous	Involved in general stress response	-0.161	Stress Response
dps	Stationary phase nucleoid protein that sequesters Iron and protects DNA from damage	994563	Synonymous	Protects DNA from oxidative damage based on sequestration of iron ions	-0.166	Stress Response
dps	Stationary phase nucleoid protein that sequesters Iron and protects DNA from damage	994857	Synonymous	Protects DNA from oxidative damage based on sequestration of iron ions	-0.167	Stress Response
cpxA	Sensor histidine kinase	4589982	Synonymous	Induced during cell envelope stress, important for biofilm formation, but induction by stress may indicate conditions that decrease biofilms	-0.189	Stress Response
ydeI	Stress response protein	1833499	Intergenic Region ydeH-ydeI	Involved in general stress response	-0.213	Stress Response
uspG	Universal stress protein UP12	756448	Synonymous	Involved in general stress response	-0.214	Stress Response
adiC	AdiC arginine:agmatine antiporter	4821637	Synonymous	Involved in acid resistance, but may not be beneficial at low temperatures	-0.219	Stress Response
sdhD	Succinate dehydrogenase membrane protein	855159	Synonymous	Central metabolism	0.335	TCA cycle
sdhD	Succinate dehydrogenase membrane protein	855321	Synonymous	Central metabolism	0.254	TCA cycle
sdhA	Succinate dehydrogenase flavoprotein	856595	Synonymous	Central metabolism	0.195	TCA cycle
sdhC	Succinate dehydrogenase membrane protein	854792	Missense	Central metabolism	-0.279	TCA Cycle
ybbL	Putative transporter subunit: ATP-binding component of ABC superfamily protein	603519	Synonymous	Transportation of metabolites	0.282	Transportation
kefB	K : H antiporter KefB	3910378	Synonymous	Transportation of metabolites	0.237	Transportation
ybaJ	YbaL CPA2 transporter	560016	Synonymous	Transportation of metabolites	0.182	Transportation
ybbL	Putative transporter subunit: ATP-binding component of ABC superfamily protein	603243	Synonymous	Transportation of metabolites	-0.162	Transportation
gadC	Glutamic acid:4-aminobutyrate antiporter	1767540	Synonymous	Transportation of metabolites	-0.212	Transportation
zntB	Zinc transport protein	1646137	Synonymous	Low zinc levels can inhibit biofilm formation	-0.386	Transportation

The results suggest that iron metabolism and stress response play a greater role in biofilm formation at 15°C than at 37°C. It is also shown that variants involved in motility or attachment were of importance for biofilm formation at both 37°C and 15°C (**Fig 10**). Variants involved in motility tended to have a negative effect on biofilm density while variants involved in attachment

tended to have a positive effect on biofilm density. This suggests that attachment factors are advantageous for biofilm formation and that motility may not be as required under these conditions to form greater biofilms.

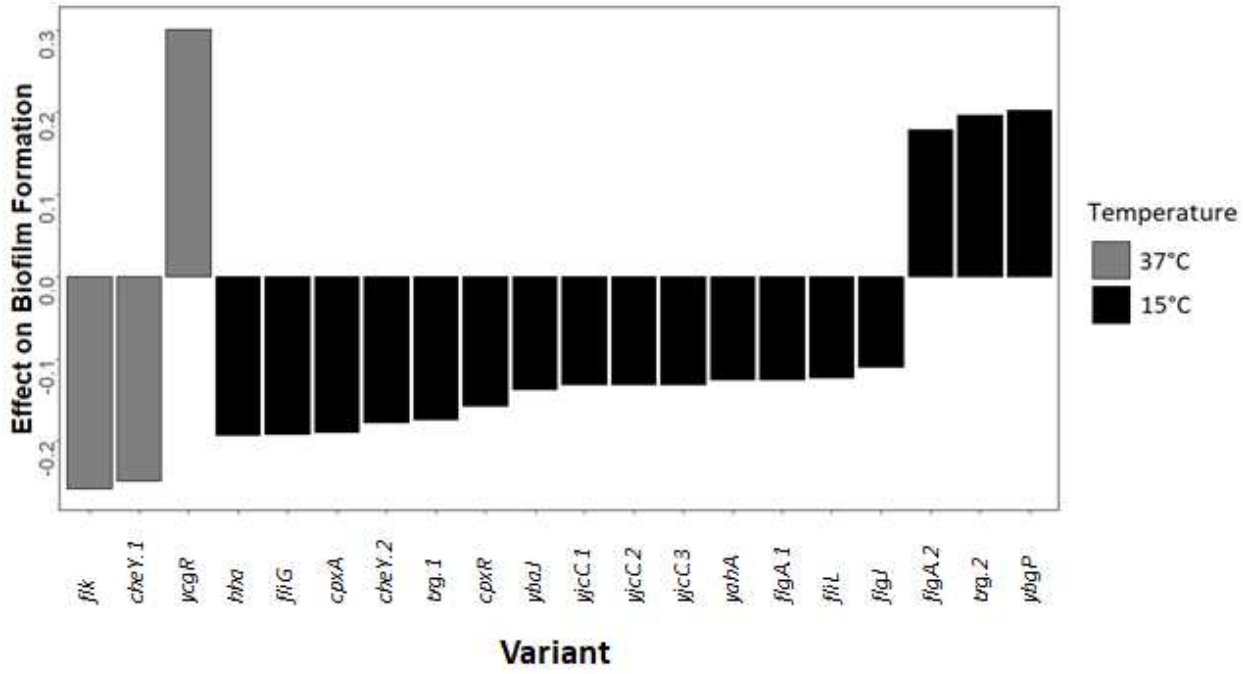

Figure 10: Variants that effect motility or attachment factors and their effect on biofilm formation at 37°C and 15°C. A number of important variants were involved in motility or attachment at both 37°C and 15°C. Variants that had an effect on biofilm density at 37°C are colored in gray and variants that had an effect at 15°C are colored in black. Variants involved in promoting motility had a negative effect on biofilm formed and variants involved in attachment factors, regulation of biofilm formation, or inhibition of motility had a positive effect on biofilm formed.

Discussion

Some soil environments may select for enhanced biofilm formation.

The biphasic lifestyle of *E. coli* makes understanding the survival and persistence in soil environments of importance when studying ecology of *E. coli*. Soil environments compared to the host environment are more variable in stresses present, leading to a pressure on *E. coli* to diversify. When analyzing the survival and persistence of bacteria in the soil environment, it is essential to consider the variability in soil characteristics that may influence biofilm formation.

Our results showed that of the environmental conditions of the soils from which the isolates were sampled, only pH, calcium, and moisture showed any significant association with density of biofilms in culture. Past studies have attributed an increase in biofilm formation to higher calcium levels among some Gram-negative bacteria (80, 106–108). It is thought that calcium-binding proteins play a role in bacterial adhesion to a surface and that calcium can act as an ionic cross-bridging molecule for negatively charged polysaccharides (80, 106). Calcium is also thought to possibly be used as a signaling molecule for gene expression during biofilm-associated growth and regulation of channels and transporters (107). Along with this, it is possible that the association with increased biofilm formation and higher pH levels may be a product of calcium levels.

Higher moisture levels have also shown to increase biofilm formation in *E. coli.* Increased moisture levels have been attributed to longer persistence and survival of *E. coli* in soils, due to the *E. coli* meeting their water requirements (109). An increase in motility, which plays an important role in the first stages of biofilm formation, has been attributed to an increase in moisture (63). Our results fit what has been observed in the past, but it is unclear the exact mechanism that is selecting for increased biofilm formation with greater moisture. Perhaps greater moisture is selecting for bacteria with greater motility, possibly increasing their ability to spread and colonize more surfaces, where biofilm formation is then beneficial.

Biofilm formation may be influenced by stress from lower temperatures in soils. This stems from the balance between nutrient utilization and stress tolerance, or SPANC balance (self-preservation and nutrient competence). *E. coli* with greater stress tolerance had lower utilization of nutrients and *vice versa*, as controlled by regulation of RpoS (110). Enhanced

nutrient utilization can lead to an increase in growth (110, 111), thus possibly increasing biofilm formation.

Our results indicated that biofilm formed per unit growth was greater at 15°C than at 37°C. Past studies have shown that low temperatures may increase gene expression of genes involved in attachment of *E. coli*, such as *csgA* which codes for curli production, and *crl* which enhances expression of the *csgBA* operon and is known to be more stable at lower temperatures (79, 112). Interestingly, the equal number of viable CFU at 37°C suggests that cells grown at 15°C may have less volume than those grown at 37°C. This change in cell shape at low temperature has been observed before, with *E. coli* grown at 22°C compared to 42°C having cells that were shorter and thicker, with slightly less cell volume (113).

A past study showed that growing *E. coli* in glucose defined minimal media alone does not produce visible biofilm as measured using a CV assay, and required the addition of casamino acids (72). Indeed, this was shown in our isolates as well, with an addition of 0.5% CAA added to the GDMM to produce visible biofilms. It is possible that the CAA provided additional amino acids for use by *E. coli* for biofilm formation, thus possibly decreasing the energy expended by *E. coli* during central metabolism to produce the required amino acids and increasing biofilm formed. The large number of variants involved in central metabolism influencing biofilm formed at both 37°C and 15°C, for example variants involved in regulation and production of products in the TCA cycle, suggests that mechanisms influencing central metabolism are important for impacting biofilm formed.

Variants involved in cell envelope formation and structure were associated with control of biofilm formed at 37°C due to influencing adhesion ability.

Hydrophobicity of the cell surface is known to influence adhesion ability and increase attachment to surfaces, especially surfaces that are hydrophobic (114–116). Different polysaccharides and membrane proteins can influence this hydrophobicity (117–119). Our results indicate that variants involved in cell envelope structure and biosynthesis are associated with some of the greatest effects on biofilm density at 37°C.

Stress response also plays a role in regulating the cell envelope, such as with the CpxA-CpxR two-component signal transduction system that controls genes involved in the folding and degradation of proteins during cell envelope formation (120). This includes up-regulation of *degP*, which serves as a chaperone to eliminate misfolded outer-membrane proteins (121) and whose mutants are known to lead to a decrease in biofilm formation (122). Our results show that *degP* was associated with a positive effect on biofilm density at 37°C.

Another function that may influence cell envelope hydrophobicity, and thus biofilms, includes cell division. Higher levels of cell division and growth rate have shown a decrease in surface hydrophobicity (123), thus decreasing adhesion ability. Our results show that variants involved in cell division were associated with a negative effect on biofilm density at 37°C. However, this is more likely a growth effect than an adhesion effect due to cell division being the cause of the effect.

Variants involved in stress response and iron acquisition were associated with biofilm formed at 15°C due to mutations and effects of oxidative stress.

E. coli has several stress responses that take place to protect the bacteria from harmful conditions, such as highly acidic or alkaline pH, temperature stress, oxidative stress, etc. (49,

124–126). Surprisingly, most identified variants involved in stress response were associated with a negative effect on biofilm density at 15°C, which may be due to the SPANC balance between stress response and nutrient utilization. These stress variants were involved in acid resistance, oxidative stress response, or general stress response. Past studies have shown that the transcription of a number of acid-resistance genes are down-regulated during biofilm formation, but the exact mechanism is not currently well understood (127–130). For example, the signaling molecule AI-2, a quorum sensing molecule that increases biofilm formation (130), has been shown to repress *gadABC* (131), which are genes involved in protection against highly acidic conditions (132). Our results show that variants in *gadC* was associated with a decrease in biofilm formed at 15°C. Some other oxidative and general stress response genes have been shown to be linked with acid resistance as well. This includes *ycfR*, a multiple stress resistance protein that may decrease biofilm formation by repressing cell-cell interactions and cell surface interactions as well as decreasing indole synthesis (133).

Some oxidative and general stress responses are linked with iron acquisition, including *dps* and *uspG* (134, 135). Iron is an important molecule utilized by bacteria in metabolic pathways including the tricarboxylic acid cycle, respiration, DNA synthesis, and synthesis of metabolites (136). However, the use of iron imposes a set of challenges. Iron is poorly soluble, especially at higher pH, thus making uptake more difficult. It is also toxic at high concentrations within the cell, leading to hydroxyl radical formation that can lead to DNA damage (137). Our results indicate that variants involved in iron acquisition and homeostasis are mostly associated with a negative effect on biofilm density at 15°C. Three of the five enterobactin iron transport variants, *entF* and *fepE*, are missense mutations. While it is unknown the exact effect of these mutations, a change in the functions of these genes may have taken place, leading to a possible

decrease in cellular iron content or the possibility of damage by redox radicals. Previous studies have shown that enterobactin is used in oxidative stress response and mutations in iron storage proteins lead to iron deficiency and growth impairment in *E. coli* (138–140).

A number of variants involved in protecting against oxidative stress or repairing damage done by potential reactive oxygen species were associated with a negative effect on biofilm formed. This also suggests that oxidative stress may be affecting biofilm formation at 15°C, which is sensible due to increased solubility of O_2 as temperature declines. This negative effect could be due to reactive oxygen species damaging DNA, proteins, and lipids within the cell (141–143). The energy required to repair and protect against this damage, as well as the damage itself, could interfere with other cellular processes, such as biofilm formation. However, without a direct measurement, it can only be speculated that oxidative stress is present and that it is having this effect on forming biofilms.

Variants involved in motility were associated with a negative effect on biofilm formed at 37°C and 15°C due to switch to sessile lifestyle during biofilm formation.

Biofilm formation is a multi-step process that is influenced by motility and attachment to surfaces. Motility is involved in the dispersal of bacteria to new locations where they can form new biofilms and is important in the first stages of biofilm formation (65). However, after initial attachment, bacteria switch from a motile lifestyle to a sessile lifestyle, with a downregulation of motility genes and an upregulation of attachment genes (78, 144). Our results showed that variants involved in flagellar production were associated with a negative effect on biofilm density. A number of c-di-GMP phosphodiesterases, which degrades cyclic diguanylate, a secondary messenger involved in the regulation of forming biofilms by inhibiting motility (77), were also shown to be associated with a negative effect on biofilm density. Variants that coded

for attachment factors, like fimbrial structures, or that were associated with a decrease in flagellar movement or production, such as the flagellar brake gene *ycgR*, were associated with a positive effect on biofilm density.

Phylogroup D clades have evolved to be distinct in their ability to produce biofilms and the presence/absence of variants associated with an effect on biofilm formation present among the clades.

E. coli is a species whose genetic diversity is influenced not only by mutation and recombination events, but also through adaptation to differing environments and their conditions (4, 5, 23). Past studies have shown that variability exists in the genomes and in the adaptive traits of *E. coli* taken from different environments, including isolates taken from different animal hosts (36) and isolates taken from plant-surfaces compared to isolates taken from animal hosts (3, 37). Due to the variability of the soil conditions from which the isolates were sampled and the natural diversity of *E. coli* as a species, it was expected that genetic diversity would be present among our isolates. Indeed, our phylogroup D *E. coli* showed diversity in variants involved in biofilm formation across isolates and between different clades. Such variability suggests the possible adaptation of traits to differing environments, with traits being shared among closely related individuals, such as the isolates within clades. This leads to differences in biofilm phenotypes between clades, possibly giving some clades an advantage to surviving in the soil environments compared to other clades.

Biofilm formation of E. coli in soil environments is a public health concern.

We've established that *E. coli* from extrahost environments produce greater biofilms than *E. coli* from host-associated environments, as shown by our soil isolates producing greater biofilms than our fecal isolates at 37°C. Past studies have shown similar results, with plant and

soil-associated isolates producing greater biofilms than animal-associated isolates (37). Due to the greater variability in stresses of extrahost environments compared to the intestinal environment, biofilm formation would better serve as a survival mechanism in soils than in the intestine, thus leading to extrahost environments selecting for greater biofilm formation (37, 102).

The issue of increasing antibiotic resistance among pathogens makes it important to understand where and how antibiotic resistance traits are spread among bacteria. Soil and other extrahost environments, with the highly variable conditions, have been proposed to be a source of most of the biodiversity of *E. coli*, as well as possible reservoirs for virulence and antibiotic resistance genes (99, 145). This is especially true of *E. coli* found in environments near farms with antibiotic-treated livestock, as the use of antibiotics in food production has been linked to an increase in drug-resistant pathogens, due to run-off of fecal matter into soils and waterways selecting for resistance in *E. coli* present (145, 146). Such genes can then be passed on to other *E. coli* through horizontal gene transfer. Contamination of crops, waterways, etc. by *E. coli* may then lead to resistant strains infecting the human population and causing disease (146–148).

Biofilm formation also plays a role in virulence in *E. coli*, as it can be an integral part of *E. coli* infections, such as in UTIs (45, 47, 149). Increased biofilm formation has been attributed to increased resistance to antibiotics, due to the protective matrix reducing penetration of antibiotics and prolonging an infection (46). Some virulence traits in *E. coli* are also associated with biofilm formation, such as the P fimbrial structure subunit gene, which is associated with virulence and increased attachment (40, 73, 74). Thus, biofilm formation not only increases survival in extrahost environments, but also may increase virulence and antibiotic resistance among *E. coli* infecting the human population.

Materials and methods

E. coli isolate sampling.

A total of 265 phylogroup D *E. coli* isolates were cultivated from surface soil samples near the Buffalo River in Minnesota and North Dakota (150). Environmental and soil properties for each sample were measured according to standard methods (151). This included soil moisture, soil consolidation, soil texture, landcover type, vegetation, nitrate, magnesium, calcium, pH, organic matter, soluble salts, iron, copper, sodium, and phosphorous.

Measurement of biofilm density by crystal violet assay.

A study showed that glucose defined minimal media (GDMM) alone did not produce visible biofilms, but the addition of casamino acids (CAA), which are known to promote biofilm formation, to GDMM at a concentration of 0.5% CAA produced visible biofilms with crystal violet staining (72). We found similar results when testing biofilm formation with and without CAA. The average OD_{600} values for the GDMM and GDMM + 0.5% CAA were approximately 0.101 and 0.569, respectively. The coefficient of variance of the GDMM was 43.933 and of the GDMM + 0.5% CAA was 35.953. Thus, GDMM + 0.5% CAA was used as the medium for the biofilm assay.

A crystal (CV) assay of biofilm density was performed (72). CV biofilm assays are known for being difficult to reproduce (152, 153), thus six replicates of four phylogroup D culture collections were assayed in a randomized block design with six replicates per block. The experiment was repeated over a period of four weeks to obtain data for six replicates of each *E. coli* isolate, as described below.

Isolates were first inoculated in 200 µL LB broth in 96-well plates using freezer stock cultures that had been grown on LB agar plates. The plates were incubated at 37°C until

stationary phase was reached (up to 16 h). A 1% inoculum was transferred from LB broth into 198 µL GDMM broth containing 0.1% glucose (w/v) and 0.5% CAA in new 96 well plates. These cultures were incubated at 37°C until stationary phase was reached. This was repeated three more times. The transfers were done to allow the *E. coli* to acclimate to the glucose limited environment. After the final transfer into GDMM + 0.5% CAA, cultures were incubated at 37°C for 48 h to allow biofilms to form.

After 48 h, the media and suspended planktonic cells in each well of the 96-well plates was discarded using a pipette, being careful not to disturb the biofilm. Each well was then washed with 200 µL 1X PBS a total of three times to remove loosely bound biomass. Assay plates were allowed to dry for approximately 1 h at room temperature after the final wash. The washed biofilms were stained using 200 µL of a 0.1% CV solution and incubated at room temperature for 15 min. Excess CV was then aspirated and each well was washed three times with 200 µL 1X PBS and the PBS discarded. The plates were then incubated at room temperature to dry completely. After drying, 200 µL of fresh ethanol:acetone (80:20) solution was added to each well to extract the CV from the stained biofilm. The plates were incubated for 15 minutes at room temperature and then 150 µL of the dye extract was transferred into new 96-well plates. Optical density at 600 nm was measured using a BioTek® Synergy H1 Hybrid Reader spectrophotometer. Wells containing only 150 µL of pure 80:20 ethanol:acetone solution were used as blanks.

It has been found that *E. coli* phylogroups are influenced by niche-specific selective pressures and variation in *E. coli* can be impacted by the environmental stresses in primary and secondary habitats (37). Using the CV assay previously described, biofilm densities of 96 phylogroup D *E. coli* isolates extracted from fecal samples were used as a comparison to the

phylogroup D soil *E. coli* isolates. Twelve replicates were measured, and the average biofilm density was compared to the average biofilm density of the soil isolates. To minimize any variation due to block effects of conducting the experiments on different days, the biofilm densities of 96 randomly chosen fecal isolates and 82 randomly chosen soil isolates were measured at the same time using the CV assay. Three replicates of each isolate were measured and the average biofilm densities between the fecal and soil isolates were calculated and compared.

Comparison of biofilm formation at 37°C and 15°C.

Lower temperatures have been suggested to influence biofilm formation by increasing gene expression of genes involved in biofilm formation (71, 79). A comparison of biofilm density at 15°C (the average temperature of the soil from which the isolates were sampled) to 37°C was done to determine the effect lower temperatures had on biofilm formation. The CV assay as previously described was used, with the exception of the isolates being incubated at 15°C for 120 h instead of 37°C for 48 h to allow for biofilms to form. This time frame was chosen based on a previous paper that implemented similar methods (49). Nine replicates of each phylogroup D soil isolate was measured and the average biofilm density compared to the biofilm density at 37°C.

A comparison of fecal and soil phylogroup D *E. coli* isolates at 15°C was also conducted, using the previously described methods for the comparison at 37°C with the exception of the isolates being incubated at 15°C for 120 h instead of 37°C for 48 h to allow for biofilms to form. Three replicates of each isolate were measured and the average biofilm densities between the fecal and soil isolates were calculated and compared.

A growth assay was done to determine the effect that growth had on biofilm formation at 37°C and 15°C. The assay was done by following the procedure for the CV assay for four replicates of each soil phylogroup D *E. coli* isolate at both temperatures. Instead of removing the excess media and staining the biofilms with CV, the plates were covered with optically clear film and the OD_{600} of the cultures was measured. Plates were shaken for 15 seconds to resuspend cells before the OD measurement. The ratio of biofilm density to growth of the isolates was then calculated by taking the OD_{600} of biofilm density and dividing it by the OD_{600} of growth.

A total of 48 randomly chosen isolates were used to determine the number of living, culturable cells after the designated period of growth for 37°C and 15°C. Isolates were chosen based on the summary statistics: 12 isolates with growth OD_{600} values less than the 1st quartile at both 37°C and 15°C, 12 isolates with growth OD_{600} values greater than the 1st quartile but less than the 2nd quartile, 12 isolates with growth OD_{600} values greater than the 2nd quartile but less than the 3rd, and 12 isolates with growth OD_{600} values greater than the 3rd quartile. Isolates were grown using the same procedure for the CV assay and growth assay at 37°C and 15°C and spread plated on LB agar plates. Cultures were diluted 10^{-6} in 1X PBS before plating (0.1 mL per plate). Three replicates of each isolate were made and the average CFU/mL determined. Average CFU/mL values were $\log_{10}$ transformed before comparison with biofilm density and average growth.

***Motility assay*.**

Motility for each isolate was measured using motility plates containing 1% tryptone, 0.5% NaCl, and 0.3% agar. A previously described method was used in which the isolates were spotted onto the motility plates and incubated in a humid environment at 25°C and 34°C (154). For the next 6 to 8 h, the expanding motility zones were observed, and the motility of the isolates

categorized using a series of control strains. The non-motile *E. coli* K-12 strain MC1000*flhD::kn* (155) was used as a negative control (category: non-motile) and the strain AJW678 (156) as a positive control (category: motile). The highly motile *E. coli* K-12 strain MC1000 was used as an added control (category: highly motile) and spotted into the center of the motility plates to avoid interference with any less motile isolates. Any isolates that had motility greater than MC1000*flhD::kn* but less motility than AJW678 were categorized as slightly motile, or 'fuzzy' motile.

Statistical analysis.

RStudio-1.0.136 with R-3.3.2 was used to conduct statistical analyses on the data collected. A linear regression analysis was conducted comparing average biofilm density among experimental blocks and culture collection plates. The estimates calculated from the model was used to determine and correct any block effects in the data. Spearman rank correlation tests were performed on each replicate of each plate to determine the strength of rank-order correlations between replicates, and this permitted assessment of reproducibility among replicates of the same plate.

Using the block-corrected OD_{600} data, the averages of all replicates for each isolate were calculated. Using a rank plot and summary statistics, biofilm density values were categorized into three different biofilm density groups to distinguish between similar biofilm phenotypes. Group 1 consisted of 37 isolates (14% of all isolates) that had OD_{600} values less than the 1st quartile value of ~ 0.2 and were categorized as weak biofilm formers. Group 2 consisted of 183 isolates (69% of all isolates) that had OD_{600} values greater than the 1st quartile value of ~ 0.2 but less than the 3rd quartile value of ~ 0.7 and were categorized as medium biofilm formers. Group

3 consisted of 45 isolates (17% of all isolates) that had OD_{600} values greater than the 3rd quartile value of ~ 0.7 and were categorized as strong biofilm formers.

For continuous soil data, Spearman rank correlation tests were conducted on average OD_{600} values of the isolates with corresponding soil properties in the soils of origin. ANOVA tests (specifically Kruskal-Wallis tests for non-parametric data) were conducted using models that compared categorical environmental data and average OD_{600} values. For categorical soil data, chi-square tests were conducted on environmental data and biofilm density groups. Fisher's Exact test was used when the sample groups were small. A Bonferroni correction was conducted on each comparison in the Fisher's Exact test to decrease the likelihood of a Type 1 error. Of the Bonferroni corrected p-values, five were found to correct from significant to not significant: the comparison between biofilm density groups and non-motile isolates at 34°C, soil moisture and non-motile isolates at 34°C, motility at 34°C and soil moisture that was categorized as standing water, biofilm density groups and standing water, and landcover and standing water. Only soil environmental conditions that were found to be significantly associated with biofilm density were reported in the results.

GWAS of biofilm formation.

Genome sequences of each of the isolates was obtained by performing Genome QC using FastQC 0.11.2 (157) on fastq.gz files from Macrogen Clinical Labs (Seoul, South Korea). Genomes were trimmed using Trimmomatic 0.32 (158) and assembled *de novo* using VelvetOptimiser 2.2.5 (159). Prokka 1.12 (160) was used for annotation and a pan-genome analysis was conducted using ROARY 3.8.2 (85). *E. coli* strain UMN026 was used as a reference to the isolates since it is a well categorized phylogroup D strain. RAxML 7.3 (161) was used to create a phylogeny on core genome alignment. GATK 3.3.0 (86) and kSNP3.0 (87) was

used to determine SNPs and indels among the genomes. SNP calls were filtered out based on redundancy or rarity (if they were present in 90% or absent in 90% of all genomes). A Jaccard Similarity was used to group SNPs that were present together in at least 90% of genomes into mosaics. A total of 22,706 genes were present in the pan-genome of the 277 phylogroup D *E. coli* isolates analyzed. A total of 2,797 genes were present in at least 98% of the genomes and classified as core genes with a 95% sequence identity cutoff. A total of 28,118 gene clusters were determined using ROARY. A presence/absence matrix of variants in each isolate was generated and a procedural logistic regression was run with the matrix and biofilm density data for 37°C and 15°C with each of the seven genetic eigenvectors from previously defined phylogroup D clades as predictors. Variants were filtered by calculated FDR *q*-values (<0.05) to determine which variants were significant in effecting biofilm formation.

Variants significant in effecting biofilm formation were determined to be of more or less importance by first categorizing their functions and biological processes. Variants that were determined to be of greater importance included: variants associated with motility or attachment factors, variants associated with a comparatively high negative/positive effect on biofilm formed, multiple variants of the same gene, multiple variants involved in similar functions and biological processes, and variants of genes that have been associated with an effect on biofilm formation. Importance of variants was also determined through a gene enrichment analysis using PANTHER (162) and a cellular overview analysis of the variants associated with an effect on biofilm formed using Ecocyc Pathway Tools Enrichment Analysis (163). Variants associated with cellular growth or a comparatively low negative/positive effect on biofilm formed were determined to be of less importance when analyzing the effect on biofilm formation.

REFERENCES

1. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol 8:nrmicro2298.
2. Yan H, Zhou W. 2004. Allelic variations in gene expression. Curr Opin Oncol 16:39–43.
3. Bergholz PW, Noar JD, Buckley DH. 2011. Environmental patterns are imposed on the population structure of *Escherichia coli* after fecal deposition. Appl Environ Microbiol 77:211–219.
4. Lyautey E, Lu Z, Lapen DR, Wilkes G, Scott A, Berkers T, Edge TA, Topp E. 2010. Distribution and Diversity of *Escherichia coli* Populations in the South Nation River Drainage Basin, Eastern Ontario, Canada. Appl Environ Microbiol 76:1486–1496.
5. Park C, Zhang J. 2012. High Expression Hampers Horizontal Gene Transfer. Genome Biol Evol 4:523–532.
6. van Elsas JD, Chiurazzi M, Mallon CA, Elhottova D, Kristufek V, Salles JF. 2012. Microbial diversity determines the invasion of soil by a bacterial pathogen. Proc Natl Acad Sci U S A 109:1159–1164.
7. Turrientes M-C, González-Alba J-M, del Campo R, Baquero M-R, Cantón R, Baquero F, Galán JC. 2014. Recombination blurs phylogenetic groups routine assignment in *Escherichia coli*: setting the record straight. PloS One 9:e105395.
8. van Elsas JD, Semenov AV, Costa R, Trevors JT. 2011. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. ISME J 5:173–183.
9. Lukjancenko O, Wassenaar TM, Ussery DW. 2010. Comparison of 61 Sequenced *Escherichia coli* Genomes. Microb Ecol 60:708–720.

10. Kaas RS, Friis C, Ussery DW, Aarestrup FM. 2012. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. BMC Genomics 13:577.

11. Rasko DA, Rosovitz MJ, Myers GSA, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J. 2008. The Pangenome Structure of *Escherichia coli*: Comparative Genomic Analysis of *E. coli* Commensal and Pathogenic Isolates. J Bacteriol 190:6881–6893.

12. Wheeler Alm E, Burke J, Spain A. 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Water Res 37:3978–3982.

13. Alm EW, Burke J, Hagan E. 2006. Persistence and Potential Growth of the Fecal Indicator Bacteria, *Escherichia coli*, in Shoreline Sand at Lake Huron. J Gt Lakes Res 32:401–405.

14. Gordon DM, Bauer S, Johnson JR. 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiology 148:1513–1522.

15. Luo C, Walk ST, Gordon DM, Feldgarden M, Tiedje JM, Konstantinidis KT. 2011. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. Proc Natl Acad Sci 108:7200–7205.

16. Carlson SM, Cunningham CJ, Westley PAH. 2014. Evolutionary rescue in a changing world. Trends Ecol Evol 29:521–530.

17. Bell G. 2013. Evolutionary rescue and the limits of adaptation. Philos Trans R Soc B Biol Sci 368.

18. Gomulkiewicz R, Holt RD. When Does Evolution by Natural Selection Prevent Extinction? Evolution 49:201–207.

19. Ramsayer J, Kaltz O, Hochberg ME. Evolutionary rescue in populations of *Pseudomonas fluorescens* across an antibiotic gradient. Evol Appl 6:608–616.

20. Bell G, Gonzalez A. Evolutionary rescue can prevent extinction following environmental change. Ecol Lett 12:942–948.

21. Perron GG, Gonzalez A, Buckling A. The rate of environmental change drives adaptation to an antibiotic sink. J Evol Biol 21:1724–1731.

22. Lachapelle J, Bell G. Evolutionary Rescue of Sexual and Asexual Populations in a Deteriorating Environment. Evolution 66:3508–3518.

23. Adato O, Ninyo N, Gophna U, Snir S. 2015. Detecting Horizontal Gene Transfer between Closely Related Taxa. PLoS Comput Biol 11.

24. Bobay L-M, Traverse CC, Ochman H. 2015. Impermanence of bacterial clones. Proc Natl Acad Sci 112:8893–8900.

25. Mau B, Glasner JD, Darling AE, Perna NT. 2006. Genome-wide detection and analysis of homologous recombination among sequenced strains of *Escherichia coli*. Genome Biol 7:R44.

26. Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. ISME J 3:199–208.

27. Touchon M, Hoede C, Tenaillon O, Barbe V, Baeriswyl S, Bidet P, Bingen E, Bonacorsi S, Bouchier C, Bouvet O, Calteau A, Chiapello H, Clermont O, Cruveiller S, Danchin A, Diard M, Dossat C, Karoui ME, Frapy E, Garry L, Ghigo JM, Gilles AM, Johnson J, Bouguénec CL, Lescat M, Mangenot S, Martinez-Jéhanne V, Matic I, Nassif X, Oztas S, Petit MA, Pichon C, Rouy Z, Ruf CS, Schneider D, Tourret J, Vacherie B, Vallenet D, Médigue C, Rocha EPC, Denamur E. 2009. Organised Genome Dynamics in the

Escherichia coli Species Results in Highly Diverse Adaptive Paths. PLOS Genet 5:e1000344.

28. Didelot X, Méric G, Falush D, Darling AE. 2012. Impact of homologous and non-homologous recombination in the genomic evolution of *Escherichia coli*. BMC Genomics 13:256.
29. Milkman R, Jaeger E, McBride RD. 2003. Molecular evolution of the *Escherichia coli* chromosome. VI. Two regions of high effective recombination. Genetics 163:475–483.
30. Weissman SJ, Chattopadhyay S, Aprikian P, Obata-Yasuoka M, Yarova-Yarovaya Y, Stapleton A, Ba-Thein W, Dykhuizen D, Johnson JR, Sokurenko EV. 2006. Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. Mol Microbiol 59:975–988.
31. Boyd EF, Hartl DL. 1998. Chromosomal Regions Specific to Pathogenic Isolates of *Escherichia coli* Have a Phylogenetically Clustered Distribution. J Bacteriol 180:1159–1165.
32. Johnson JR, Delavari P, Kuskowski M, Stell AL. 2001. Phylogenetic Distribution of Extraintestinal Virulence-Associated Traits in *Escherichia coli*. J Infect Dis 183:78–88.
33. Clermont O, Bonacorsi S, Bingen E. 2000. Rapid and Simple Determination of the *Escherichia coli* Phylogenetic Group. Appl Environ Microbiol 66:4555–4558.
34. Clermont O, Gordon D, Denamur E. 2015. Guide to the various phylogenetic classification schemes for *Escherichia coli* and the correspondence among schemes. Microbiol Read Engl 161:980–988.

35. Gordon DM, Clermont O, Tolley H, Denamur E. 2008. Assigning *Escherichia coli* strains to phylogenetic groups: multi-locus sequence typing versus the PCR triplex method. Environ Microbiol 10:2484–2496.

36. Johnson JR, Johnston BD, Delavari P, Thuras P, Clabots C, Sadowsky MJ. 2017. Phylogenetic Background and Virulence-Associated Traits of *Escherichia coli* Obtained from Surface Waters and Diverse Animals in Minnesota and Wisconsin. Appl Environ Microbiol.

37. Méric G, Kemsley EK, Falush D, Saggers EJ, Lucchini S. 2013. Phylogenetic distribution of traits associated with plant colonization in *Escherichia coli*. Environ Microbiol 15:487–501.

38. Picard B, Garcia JS, Gouriou S, Duriez P, Brahimi N, Bingen E, Elion J, Denamur E. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. Infect Immun 67:546–553.

39. Salipante SJ, Roach DJ, Kitzman JO, Snyder MW, Stackhouse B, Butler-Wu SM, Lee C, Cookson BT, Shendure J. 2015. Large-scale genomic sequencing of extraintestinal pathogenic *Escherichia coli* strains. Genome Res 25:119–128.

40. Rijavec M, Müller-Premru M, Zakotnik B, Zgur-Bertok D. 2008. Virulence factors and biofilm production among *Escherichia coli* strains causing bacteraemia of urinary tract origin. J Med Microbiol 57:1329–1334.

41. Pavlickova S, Dolezalova M, Holko I. 2015. Resistance and virulence factors of *Escherichia coli* isolated from chicken. J Environ Sci Health Part B 50:417–421.

42. Valat C, Auvray F, Forest K, Métayer V, Gay E, Garam CP de, Madec J-Y, Haenni M. 2012. Phylogenetic Grouping and Virulence Potential of Extended-Spectrum-β-

Lactamase-Producing *Escherichia coli* Strains in Cattle. Appl Environ Microbiol 78:4677–4682.

43. Coura FM, Diniz AN, Junior O, Augusto C, Lage AP, Lobato FCF, Heinemann MB, Silva ROS, Coura FM, Diniz AN, Junior O, Augusto C, Lage AP, Lobato FCF, Heinemann MB, Silva ROS. 2018. Detection of virulence genes and the phylogenetic groups of *Escherichia coli* isolated from dogs in Brazil. Ciênc Rural 48.

44. P V Cunha M, Xavier de Oliveira M, Caroline Vilela de Oliveira M, Silva K, Ribeiro Gomes C, Moreno A, Knöbl T. 2014. Virulence Profiles, Phylogenetic Background, and Antibiotic Resistance of *Escherichia coli* Isolated from Turkeys with Airsacculitis. Sci World J 1–8.

45. Ejrnæs K. 2011. Bacterial characteristics of importance for recurrent urinary tract infections caused by *Escherichia coli*. Dan Med Bull 58:B4187.

46. Blango MG, Mulvey MA. 2010. Persistence of Uropathogenic *Escherichia coli* in the Face of Multiple Antibiotics. Antimicrob Agents Chemother 54:1855–1863.

47. Soto SM, Smithson A, Horcajada JP, Martinez JA, Mensa JP, Vila J. 2006. Implication of biofilm formation in the persistence of urinary tract infection caused by uropathogenic *Escherichia coli*. Clin Microbiol Infect 12:1034–1036.

48. Brennan FP, O'Flaherty V, Kramers G, Grant J, Richards KG. 2010. Long-Term Persistence and Leaching of *Escherichia coli* in Temperate Maritime Soils. Appl Environ Microbiol 76:1449.

49. Somorin Y, Abram F, Brennan F, O'Byrne C. 2016. The General Stress Response Is Conserved in Long-Term Soil-Persistent Strains of *Escherichia coli*. Appl Environ Microbiol 82:4628–4640.

50. Moynihan EL, Richards KG, Ritz K, Tyrrel SF, Brennan FP. 2013. Impact of Soil Type, Biology and Temperature on the Survival of Non-Toxigenic *Escherichia coli* O157. Biol Environ Proc R Ir Acad 113B:41–46.

51. Lang NL, Smith SR. 2007. Influence of soil type, moisture content and biosolids application on the fate of *Escherichia coli* in agricultural soil under controlled laboratory conditions. J Appl Microbiol 103:2122–2131.

52. Adiba S, Nizak C, Baalen M van, Denamur E, Depaulis F. 2010. From Grazing Resistance to Pathogenesis: The Coincidental Evolution of Virulence Factors. PLOS ONE 5:e11882.

53. DiSalvo S, Brock DA, smith jeff, Queller DC, Strassmann JE. 2014. In the social amoeba *Dictyostelium discoideum*, density, not farming status, determines predatory success on unpalatable *Escherichia coli*. BMC Microbiol 14:328.

54. Nadell CD, Drescher K, Foster KR. 2016. Spatial structure, cooperation and competition in biofilms. Nat Rev Microbiol 14:589–600.

55. Berne C, Ducret A, Hardy GG, Brun YV. 2015. Adhesins involved in attachment to abiotic surfaces by Gram-negative bacteria. Microbiol Spectr 3.

56. Danese PN, Pratt LA, Kolter R. 2000. Exopolysaccharide Production Is Required for Development of *Escherichia coli* K-12 Biofilm Architecture. J Bacteriol 182:3593–3596.

57. Whitfield C. 2006. Biosynthesis and Assembly of Capsular Polysaccharides in *Escherichia coli*. Annu Rev Biochem 75:39–68.

58. Hobley L, Harkins C, MacPhee CE, Stanley-Wall NR. 2015. Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes. FEMS Microbiol Rev 39:649–669.

59. Vogeleer P, Tremblay YDN, Mafu AA, Jacques M, Harel J. 2014. Life on the outside: role of biofilms in environmental persistence of Shiga-toxin producing *Escherichia coli*. Front Microbiol 5.

60. Gualdi L, Tagliabue L, Bertagnoli S, Ieranò T, De Castro C, Landini P. 2008. Cellulose modulates biofilm formation by counteracting curli-mediated colonization of solid surfaces in *Escherichia coli*. Microbiol Read Engl 154:2017–2024.

61. Re SD, Ghigo J-M. 2006. A CsgD-Independent Pathway for Cellulose Production and Biofilm Formation in *Escherichia coli*. J Bacteriol 188:3073–3087.

62. Serra DO, Richter AM, Hengge R. 2013. Cellulose as an Architectural Element in Spatially Structured *Escherichia coli* Biofilms. J Bacteriol 195:5540–5554.

63. Verstraeten N, Braeken K, Debkumari B, Fauvart M, Fransaer J, Vermant J, Michiels J. Living on a surface: swarming and biofilm formation. Trends Microbiol 16:496–506.

64. Beloin C, Roux A, Ghigo J-M. 2008. *Escherichia coli* biofilms. Curr Top Microbiol Immunol 322:249–289.

65. Guttenplan SB, Kearns DB. 2013. Regulation of flagellar motility during biofilm formation. FEMS Microbiol Rev 37:849–871.

66. Kumar MS, Philominathan P. 2009. The physics of flagellar motion of *E. coli* during chemotaxis. Biophys Rev 2:13–20.

67. Kalir S, McClure J, Pabbaraju K, Southward C, Ronen M, Leibler S, Surette MG, Alon U. 2001. Ordering Genes in a Flagella Pathway by Analysis of Expression Kinetics from Living Bacteria. Science 292:2080–2083.

68. Wang S, Fleming RT, Westbrook EM, Matsumura P, McKay DB. 2006. Structure of the *Escherichia coli* FlhDC Complex, a Prokaryotic Heteromeric Regulator of Transcription. J Mol Biol 355:798–808.

69. Friedlander RS, Vlamakis H, Kim P, Khan M, Kolter R, Aizenberg J. 2013. Bacterial flagella explore microscale hummocks and hollows to increase adhesion. Proc Natl Acad Sci 110:5624–5629.

70. Prigent-Combaret C, Brombacher E, Vidal O, Ambert A, Lejeune P, Landini P, Dorel C. 2001. Complex regulatory network controls initial adhesion and biofilm formation in *Escherichia coli* via regulation of the csgD gene. J Bacteriol 183:7213–7223.

71. Barnhart MM, Chapman MR. 2006. Curli Biogenesis and Function. Annu Rev Microbiol 60:131–147.

72. Pratt LA, Kolter R. 1998. Genetic analysis of *Escherichia coli* biofilm formation: roles of flagella, motility, chemotaxis and type I pili. Mol Microbiol 30:285–293.

73. Connell I, Agace W, Klemm P, Schembri M, Mărild S, Svanborg C. 1996. Type 1 fimbrial expression enhances *Escherichia coli* virulence for the urinary tract. Proc Natl Acad Sci 93:9827–9832.

74. Johnson JR. 1991. Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev 4:80–128.

75. Hengge R. 2009. Principles of c-di-GMP signalling in bacteria. Nat Rev Microbiol 7:263–273.

76. Lindenberg S, Klauck G, Pesavento C, Klauck E, Hengge R. 2013. The EAL domain protein YciR acts as a trigger enzyme in a c□di□GMP signalling cascade in *E. coli* biofilm control. EMBO J 32:2001–2014.

77. Paul K, Nieto V, Carlquist WC, Blair DF, Harshey RM. 2010. The c-di-GMP binding protein YcgR controls flagellar motor direction and speed to affect chemotaxis by a "backstop brake" mechanism. Mol Cell 38:128–139.

78. Pesavento C, Becker G, Sommerfeldt N, Possling A, Tschowri N, Mehlis A, Hengge R. 2008. Inverse regulatory coordination of motility and curli-mediated adhesion in *Escherichia coli*. Genes Dev 22:2434–2446.

79. White-Ziegler CA, Um S, Pérez NM, Berns AL, Malhowski AJ, Young S. 2008. Low temperature (23 °C) increases expression of biofilm-, cold-shock- and RpoS-dependent genes in *Escherichia coli* K-12. Microbiology 154:148–166.

80. Cruz LF, Cobine PA, De La Fuente L. 2012. Calcium Increases *Xylella fastidiosa* Surface Attachment, Biofilm Formation, and Twitching Motility. Appl Environ Microbiol 78:1321–1331.

81. Song B, Leff LG. 2006. Influence of magnesium ions on biofilm formation by *Pseudomonas fluorescens*. Microbiol Res 161:355–361.

82. Collins C, Didelot X. 2017. A Phylogenetic Method To Perform Genome-Wide Association Studies In Microbes That Accounts For Population Structure And Recombination. bioRxiv 140798.

83. Power RA, Parkhill J, Oliveira T de. 2016. Microbial genome-wide association studies: lessons from human GWAS. Nat Rev Genet 18:nrg.2016.132.

84. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O,

Vonstein V, Wilke A, Zagnitko O. 2008. The RAST Server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75.

85. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinforma Oxf Engl 31:3691–3693.

86. McKenna AH, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, Depristo M. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res gr.107524.110.

87. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31:2877–2878.

88. Dutilh BE, Backus L, Edwards RA, Wels M, Bayjanov JR, Hijum V, A.f.t S. 2013. Explaining microbial phenotypes on a genomic scale: GWAS for microbes. Brief Funct Genomics 12:366–380.

89. Zhang M, Zhang D, Wells MT. 2008. Variable selection for large p small n regression models with incomplete data: Mapping QTL with epistases. BMC Bioinformatics 9:251.

90. Breiman L. 2001. Random Forests. Mach Learn 45:5–32.

91. Alam MT, Petit RA, Crispell EK, Thornton TA, Conneely KN, Jiang Y, Satola SW, Read TD. 2014. Dissecting Vancomycin-Intermediate Resistance in *Staphylococcus aureus* Using Genome-Wide Association. Genome Biol Evol 6:1174–1185.

92. Chewapreecha C, Marttinen P, Croucher NJ, Salter SJ, Harris SR, Mather AE, Hanage WP, Goldblatt D, Nosten FH, Turner C, Turner P, Bentley SD, Parkhill J. 2014.

Comprehensive Identification of Single Nucleotide Polymorphisms Associated with Beta-lactam Resistance within Pneumococcal Mosaic Genes. PLOS Genet 10:e1004547.

93. Sheppard SK, Didelot X, Meric G, Torralbo A, Jolley KA, Kelly DJ, Bentley SD, Maiden MCJ, Parkhill J, Falush D. 2013. Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in *Campylobacter*. Proc Natl Acad Sci 110:11923–11927.

94. Farhat MR, Shapiro BJ, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR, Gardy JL, Johnston JC, Rodrigues M, Tang PKC, Kato-Maeda M, Borowsky ML, Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin EJ, Lander ES, Sabeti PC, Murray M. 2013. Genomic Analysis Identifies Targets of Convergent Positive Selection in Drug Resistant *Mycobacterium tuberculosis*. Nat Genet 45:1183–1189.

95. Miotto O, Amato R, Ashley EA, MacInnis B, Almagro-Garcia J, Amaratunga C, Lim P, Mead D, Oyola SO, Dhorda M, Imwong M, Woodrow C, Manske M, Stalker J, Drury E, Campino S, Amenga-Etego L, Thanh T-NN, Tran HT, Ringwald P, Bethell D, Nosten F, Phyo AP, Pukrittayakamee S, Chotivanich K, Chuor CM, Nguon C, Suon S, Sreng S, Newton PN, Mayxay M, Khanthavong M, Hongvanthong B, Htut Y, Han KT, Kyaw MP, Faiz MA, Fanello CI, Onyamboko M, Mokuolu OA, Jacob CG, Takala-Harrison S, Plowe CV, Day NP, Dondorp AM, Spencer CCA, McVean G, Fairhurst RM, White NJ, Kwiatkowski DP. 2015. Genetic architecture of artemisinin-resistant *Plasmodium falciparum*. Nat Genet 47:226–234.

96. Pascoe B, Méric G, Murray S, Yahara K, Mageiros L, Bowen R, Jones NH, Jeeves RE, Lappin-Scott HM, Asakura H, Sheppard SK. 2015. Enhanced biofilm formation and

multi-host transmission evolve from divergent genetic backgrounds in *Campylobacter jejuni*. Environ Microbiol 17:4779–4789.

97. Hastie T, Tibshirani R, Friedman J. 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition, 2nd ed. Springer-Verlag, New York.

98. Brennan FP, Abram F, Chinalia FA, Richards KG, O'Flaherty V. 2010. Characterization of Environmentally Persistent *Escherichia coli* Isolates Leached from an Irish Soil. Appl Environ Microbiol 76:2175–2180.

99. Blount ZD. The unexhausted potential of *E. coli*. eLife 4.

100. Lee KWK, Periasamy S, Mukherjee M, Xie C, Kjelleberg S, Rice SA. 2014. Biofilm development and enhanced stress resistance of a model, mixed-species community biofilm. ISME J 8:894–907.

101. Oliveira NM, Martinez-Garcia E, Xavier J, Durham WM, Kolter R, Kim W, Foster KR. 2015. Biofilm Formation As a Response to Ecological Competition. PLOS Biol 13:e1002191.

102. Wang A, Lin B, Sleep BE, Liss SN. 2011. The Impact of Biofilm Growth on Transport of *Escherichia coli* O157:H7 in Sand. Ground Water 49:20–31.

103. J W Costerton, Z Lewandowski, D E Caldwell, D R Korber, Lappin-Scott and HM. 1995. Microbial Biofilms. Annu Rev Microbiol 49:711–745.

104. Kalpana BJ, Aarthy S, Pandian SK. 2012. Antibiofilm Activity of α-Amylase from *Bacillus subtilis* S8-18 Against Biofilm Forming Human Bacterial Pathogens. Appl Biochem Biotechnol 167:1778–1794.

105. Molobela IP, Cloete TE, Beukes M. 2010. Protease and amylase enzymes for biofilm removal and degradation of extracellular polymeric substances (EPS) produced by *Pseudomonas fluorescens* bacteria. Afr J Microbiol Res 4:1515–1524.

106. Patrauchan MA, Sarkisova S, Sauer K, Franklin MJ. 2005. Calcium influences cellular and extracellular product formation during biofilm-associated growth of a marine *Pseudoalteromonas sp.* Microbiology 151:2885–2897.

107. Holland IB, Jones HE, Campbell AK, Jacq A. 1999. An assessment of the role of intracellular free Ca2+ in *E. coli*. Biochimie 81:901–907.

108. Kierek K, Watnick PI. 2003. The *Vibrio cholerae* O139 O-antigen polysaccharide is essential for Ca2+-dependent biofilm development in sea water. Proc Natl Acad Sci 100:14357–14362.

109. Mubiru D, Coyne M, Grove J. 2000. Mortality of *Escherichia coli* O157:H7 in Two Soils with Different Physical and Chemical Properties. J Environ Qual 1821–1825.

110. King T, Ishihama A, Kori A, Ferenci T. 2004. A Regulatory Trade-Off as a Source of Strain Variation in the Species *Escherichia coli*. J Bacteriol 186:5614–5620.

111. Chen G, Patten CL, Schellhorn HE. 2004. Positive selection for loss of RpoS function in *Escherichia coli*. Mutat Res Mol Mech Mutagen 554:193–203.

112. Bougdour A, Lelong C, Geiselmann J. 2004. Crl, a low temperature-induced protein in *Escherichia coli* that binds directly to the stationary phase sigma subunit of RNA polymerase. J Biol Chem 279:19540–19550.

113. Trueba FJ, van Spronsen EA, Traas J, Woldringh CL. 1982. Effects of temperature on the size and shape of *Escherichia coli* cells. Arch Microbiol 131:235–240.

114. Bendinger B, Rijnaarts HHM, Altendorf K, Zehnder AJB. 1993. Physicochemical Cell Surface and Adhesive Properties of Coryneform Bacteria Related to the Presence and Chain Length of Mycolic Acids. Appl Environ Microbiol 59:3973–3977.

115. Fletcher M, Loeb GI. 1979. Influence of Substratum Characteristics on the Attachment of a Marine Pseudomonad to Solid Surfaces. Appl Environ Microbiol 37:67–72.

116. Pringle JH, Fletcher M. 1983. Influence of Substratum Wettability on Attachment of Freshwater Bacteria to Solid Surfaces. Appl Environ Microbiol 45:811–817.

117. Fletcher JH, Fletcher M, Ellwood DC. 1983. Selection of Attachment Mutants during the Continuous Culture of *Pseudomonas fluorescens* and Relationship between Attachment Ability and Surface Composition. Microbiology 129:2557–2569.

118. DeFlaun MF, Tanzer AS, McAteer AL, Marshall B, Levy SB. 1990. Development of an Adhesion Assay and Characterization of an Adhesion-Deficient Mutant of *Pseudomonas fluorescens*. Appl Environ Microbiol 56:112–119.

119. Williams V, Fletcher M. 1996. *Pseudomonas fluorescens* adhesion and transport through porous media are affected by lipopolysaccharide composition. Appl Environ Microbiol 62:100–104.

120. Pogliano J, Lynch AS, Belin D, Lin EC, Beckwith J. 1997. Regulation of *Escherichia coli* cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. Genes Dev 11:1169–1182.

121. Krojer T, Sawa J, Schäfer E, Saibil HR, Ehrmann M, Clausen T. 2008. Structural basis for the regulated protease and chaperone function of DegP. Nature 453:885–890.

122. Altindis E, Fu Y, J Mekalanos J. 2014. Proteomic analysis of *Vibrio cholerae* outer membrane vesicles. Proc Natl Acad Sci U S A 111.

123. Allison DG, Evans DJ, Brown MR, Gilbert P. 1990. Possible involvement of the division cycle in dispersal of *Escherichia coli* from biofilms. J Bacteriol 172:1667–1669.

124. Vidovic S, Mangalappalli-Illathu AK, Korber DR. 2011. Prolonged cold stress response of *Escherichia coli* O157 and the role of rpoS. Int J Food Microbiol 146:163–169.

125. Chiang SM, Schellhorn HE. 2012. Regulators of oxidative stress response genes in *Escherichia coli* and their functional conservation in bacteria. Arch Biochem Biophys 525:161–169.

126. Chung H j., Bang W, Drake M a. 2006. Stress Response of *Escherichia coli*. Compr Rev Food Sci Food Saf 5:52–64.

127. Herzberg M, Kaye IK, Peti W, Wood TK. 2006. YdgG (TqsA) Controls Biofilm Formation in *Escherichia coli* K-12 through Autoinducer 2 Transport. J Bacteriol 188:587–598.

128. Domka J, Lee J, Wood TK. 2006. YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K-12 biofilm formation by influencing cell signaling. Appl Environ Microbiol 72:2449–2459.

129. Hu M, Zhang C, Mu Y, Shen Q, Feng Y. 2010. Indole Affects Biofilm Formation in Bacteria. Indian J Microbiol 50:362–368.

130. González Barrios AF, Zuo R, Hashimoto Y, Yang L, Bentley WE, Wood TK. 2006. Autoinducer 2 Controls Biofilm Formation in *Escherichia coli* through a Novel Motility Quorum-Sensing Regulator (MqsR, B3022). J Bacteriol 188:305–316.

131. Ren D, Bedzyk LA, Ye RW, Thomas SM, Wood TK. 2004. Differential gene expression shows natural brominated furanones interfere with the autoinducer-2 bacterial signaling system of *Escherichia coli*. Biotechnol Bioeng 88:630–642.

132. Ma Z, Gong S, Richard H, Tucker DL, Conway T, Foster JW. 2003. GadE (YhiE) activates glutamate decarboxylase-dependent acid resistance in *Escherichia coli* K-12. Mol Microbiol 49:1309–1320.

133. Zhang X-S, García-Contreras R, Wood TK. 2007. YcfR (BhsA) Influences *Escherichia coli* Biofilm Formation through Stress Response and Surface Hydrophobicity. J Bacteriol 189:3051–3062.

134. Nair S, Finkel SE. 2004. Dps Protects Cells against Multiple Stresses during Stationary Phase. J Bacteriol 186:4192–4198.

135. Nachin L, Nannmark U, Nyström T. 2005. Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. J Bacteriol 187:6265–6272.

136. Andrews SC, Robinson AK, Rodríguez-Quiñones F. 2003. Bacterial iron homeostasis. FEMS Microbiol Rev 27:215–237.

137. Keyer K, Imlay JA. 1996. Superoxide accelerates DNA damage by elevating free-iron levels. Proc Natl Acad Sci 93:13635–13640.

138. Adler C, Corbalan NS, Peralta DR, Pomares MF, Cristóbal RE de, Vincent PA. 2014. The Alternative Role of Enterobactin as an Oxidative Stress Protector Allows *Escherichia coli* Colony Development. PLOS ONE 9:e84734.

139. Abdul-Tehrani H, Hudson AJ, Chang Y-S, Timms AR, Hawkins C, Williams JM, Harrison PM, Guest JR, Andrews SC. 1999. Ferritin Mutants of *Escherichia coli* Are Iron Deficient and Growth Impaired, and fur Mutants are Iron Deficient. J Bacteriol 181:1415–1428.

140. Wu Y, Outten FW. 2009. IscR Controls Iron-Dependent Biofilm Formation in *Escherichia coli* by Regulating Type I Fimbria Expression. J Bacteriol 191:1248–1257.

141. Tamarit J, Cabiscol E, Ros J. 1998. Identification of the major oxidatively damaged proteins in *Escherichia coli* cells exposed to oxidative stress. J Biol Chem 273:3027–3032.

142. Demple B, Halbrook J. 1983. Inducible repair of oxidative DNA damage in *Escherichia coli*. Nature 304:466–468.

143. González-Flecha B, Demple B. 1995. Metabolic sources of hydrogen peroxide in aerobically growing *Escherichia coli*. J Biol Chem 270:13681–13687.

144. Caiazza NC, Merritt JH, Brothers KM, O'Toole GA. 2007. Inverse Regulation of Biofilm Formation and Swarming Motility by *Pseudomonas aeruginosa* PA14. J Bacteriol 189:3603–3612.

145. Esiobu N, Armenta L, Ike J. 2002. Antibiotic resistance in soil and water environments. Int J Environ Health Res 12:133–144.

146. Holvoet K, Sampers I, Callens B, Dewulf J, Uyttendaele M. 2013. Moderate Prevalence of Antimicrobial Resistance in *Escherichia coli* Isolates from Lettuce, Irrigation Water, and Soil. Appl Environ Microbiol 79:6677–6683.

147. Kljujev I, Raicevic V, Andrews S, Jackson R, Lalevic B, Dorati F. 2018. Transmission of *E. coli* from Contaminated Irrigation Water and Soil to Plant Tissue.

148. Mootian G, Wu W-H, Matthews KR. 2009. Transfer of *Escherichia coli* O157:H7 from soil, water, and manure contaminated with low numbers of the pathogen to lettuce plants. J Food Prot 72:2308–2312.

149. Trautner BW, Darouiche RO. 2004. Role of biofilm in catheter-associated urinary tract infection. Am J Infect Control 32:177–183.

150. Dusek N, Hewitt AJ, Schmidt KN, Bergholz PW. 2018. Landscape scale factors affecting prevalence of *Escherichia coli* in surface soil include land cover type, edge interactions, and soil pH. Appl Environ Microbiol AEM.02714-17.

151. Soil Survey Staff. 2014. Soil Survey Field and Laboratory Methods ManualR. Burt and Soil Survey Staff. U.S. Department of Agriculture, Natural Resources Conservation Service.

152. Feoktistova M, Geserick P, Leverkus M. 2016. Crystal Violet Assay for Determining Viability of Cultured Cells. Cold Spring Harb Protoc 2016:pdb.prot087379.

153. Shukla SK, Rao TS. 2017. An Improved Crystal Violet Assay for Biofilm Quantification in 96-Well Microtitre Plate. bioRxiv 100214.

154. Wolfe AJ, Berg HC. 1989. Migration of bacteria in semisolid agar. Proc Natl Acad Sci 86:6973–6977.

155. Malakooti J. 1990. Molecular analysis of the *Escherichia coli* motor components and regulation of internal control elements of theflaA operon.

156. Wolfe AJ, Chang D-E, Walker JD, Seitz-Partridge JE, Vidaurri MD, Lange CF, Prüss BM, Henk MC, Larkin JC, Conway T. 2003. Evidence that acetyl phosphate functions as a global signal during biofilm development. Mol Microbiol 48:977–988.

157. Andrews S. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data.

158. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinforma Oxf Engl 30:2114–2120.

159. Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829.

160. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinforma Oxf Engl 30:2068–2069.

161. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690.

162. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD. 2017. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Res 45:D183–D189.

163. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M, Latendresse M, Muñiz-Rascado L, Ong Q, Paley S, Peralta-Gil M, Subhraveti P, Velázquez-Ramírez DA, Weaver D, Collado-Vides J, Paulsen I, Karp PD. 2017. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. Nucleic Acids Res 45:D543–D550.

APPENDIX

Table A1: List of genetic variants that were determined to have a significant association with biofilm density at 37°C.

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
mukE	Protein involved in chromosome partitioning	1163474	Synonymous	Positive genetic interaction with mltC, leading to decreased biofilm formation	-0.152	Cell Division
gidB	Ribosomal RNA small subunit methyltransferase G	4410071	Synonymous	Possibly inhibited by glucose, decreasing cell division and subsequent biofilm formation; may activate stringent response and down-regulate rRNA-synthesis	-0.178	Cell Division
mutM	Formamidopyrimidine DNA glycosylase	4269988	Missense	Mutation may lead to change in function	0.167	DNA Repair
mutS	MutHLS complex, methyl-directed mismatch repair	3164064	Synonymous	Possible signal to biofilms to disperse due to DNA damage, thus decreasing biofilm	-0.143	DNA Repair
mutS	MutHLS complex, methyl-directed mismatch repair	3164076	Synonymous	Possible signal to biofilms to disperse due to DNA damage, thus decreasing biofilm	-0.143	DNA Repair
nfrA	Bacteriophage N4 receptor, outer membrane protein	695019	Synonymous	Receptor for phage	0.222	Membrane/Cell Wall
nlpC	NlpC-putative lipoprotein hydrolase	2027384	Synonymous	Membrane biogenesis	0.196	Membrane/Cell Wall
rfaD	ADP-L-glycero-D-mannoheptose-6-epimerase	4256929	Synonymous	LPS core and outer membrane biogenesis	-0.159	Membrane/Cell Wall
rffM	UDP-N-acetyl-D-mannosaminuronic acid transferase	4466085	Synonymous	Membrane biogenesis	-0.163	Membrane/Cell Wall
mrcA	Peptidoglycan synthetase; penicillin-binding protein 1A	3950341	Synonymous	Cell wall biogenesis	-0.181	Membrane/Cell Wall
ycbB	L,D-transpeptidase YcbB	1169625	Synonymous	Cell wall biogenesis	-0.311	Membrane/Cell Wall
amyA	Alpha-amylase	2240378	Synonymous	May reduce biofilm formation by degrading EPS	-0.313	Membrane/Cell Wall
wecC	UDP-N-acetyl-D-mannosamine dehydrogenase	4458415	Synonymous	Membrane biogenesis	-0.314	Membrane/Cell Wall
mltC	Membrane-bound lytic murein transglycosylase C	3440095	Synonymous	Membrane biogenesis	-0.358	Membrane/Cell Wall
ldcC	Lysine decarboxylase 2	213544	Missense	Helps regulate pH, helping survival	0.288	Metabolite Biosynthesis
ldcC	Lysine decarboxylase 2	213555	Synonymous	Helps regulate pH, helping survival	0.288	Metabolite Biosynthesis
gabD	Succinate-semialdehyde dehydrogenase (NADP)	3095058	Synonymous	Upregulation of succinate dehydrogenase leads to increase in biofilm formation	0.243	Metabolite Biosynthesis
gabD	Succinate-semialdehyde dehydrogenase (NADP)	3095001	Synonymous	Upregulation of succinate dehydrogenase leads to increase in biofilm formation	0.243	Metabolite Biosynthesis
ribF	Bifunctional riboflavin kinase / FMN adenylyltransferase	25956	Synonymous	Riboflavin biosynthesis	0.180	Metabolite Biosynthesis

Table A1: List of genetic variants that were determined to have a significant association with biofilm density at 37°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
yebR	Free methionine-R-sulfoxide reductase	2149584	Missense	Helps maintain adhesive ability	0.162	Metabolite Biosynthesis
yebR	Free methionine-R-sulfoxide reductase	2149184	Synonymous	Helps maintain adhesive ability	0.148	Metabolite Biosynthesis
aldB	Acetaldehyde dehydrogenase	4219116	Intergenic Region aldB-yiaY	Central Metabolism	-0.104	Metabolite Biosynthesis
yiaY	Putative Fe-containing alcohol dehydrogenase	4219116	Intergenic Region aldB-yiaY	Central Metabolism	-0.104	Metabolite Biosynthesis
purT	Phosphoribosylglycinamide formyltransferase 2	2166500	Missense	Mutation may lead to decrease in EPS production as influenced by purine synthesis	-0.145	Metabolite Biosynthesis
yedQ	Putative diguanylate cyclase	2260936	Synonymous	Involved in regulation of cellulose production	-0.159	Metabolite Biosynthesis
torT	Periplasmic protein TorT	1229763	Missense	Involved in anaerobic respiration	-0.171	Metabolite Biosynthesis
hscA	Chaperone, member of Hsp70 protein family	2950164	Synonymous	Chaperone involved in the maturation of iron-sulfur cluster-containing proteins	-0.185	Metabolite Biosynthesis
yhiN	Putative oxidoreductase with FAD/NAD(P)-binding domain protein	4084856	Synonymous	Central Metabolism	-0.268	Metabolite Biosynthesis
ycgR	Molecular brake that regulates flagellar motility in response to c-di-GMP	1492833	Intergenic Region ycgR-ymgE	Decreases motility	0.300	Motility
cheY	Chemotaxis binding protein	2201461	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.248	Motility
flk	Putative flagella assembly protein	2737533	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.257	Motility
relA	GDP pyrophosphokinase / GTP pyrophosphokinase	3218856	Synonymous	Involved in stringent stress response and downregulates production of rRNA	0.154	Replication/ Transcription/ Translation
xseA	Exonuclease VII, large subunit	2914508	Synonymous	May affect expression and DNA repair	-0.166	Replication/ Transcription/ Translation
xseA	Exonuclease VII, large subunit	2914808	Synonymous	May affect expression and DNA repair	-0.238	Replication/ Transcription/ Translation
cysS	Cysteinyl-tRNA synthetase	635271	Synonymous	Involved in cysteine synthesis	-0.306	Replication/ Transcription/ Translation
nadR	NadR DNA-binding transcriptional repressor and NMN adenylyltransferase	5184339	Synonymous	May affect expression	-0.328	Replication/ Transcription/ Translation
fdoG	Formate dehydrogenase-O, alpha subunit	4575196	Synonymous	Energy metabolism	0.471	Respiration
degP	Serine protease Do	185468	Synonymous	Increases biofilm formation through upregulation of Cpx signal transduction pathway	0.200	Stress Response

Table A1: List of genetic variants that were determined to have a significant association with biofilm density at 37°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
rcsC	Sensor histidine kinase RcsC	2617541	Synonymous	Used to detect stress in outer membrane and peptidoglycan layer	-0.154	Stress Response
cbpA	DNA binding protein	1236872	Intergenic Region cbpA-yccE	Involved in general stress response	-0.169	Stress Response
tsgA	YhfC MFS transporter	3924410	Synonymous	Transportation of metabolites	-0.139	Transportation
mdtB	MdtABC-TolC multidrug efflux transport system - membrane subunit	2456662	Synonymous	Transcriptionally activated by CpxR	-0.159	Transportation
cysZ	Putative inner membrane protein	2819016	Synonymous	Provides sulfur for cysteine synthesis	-0.175	Transportation
hofC	Protein transport protein HofC	119127	Synonymous	Overexpression results in undetectable type IV pili	-0.182	Transportation
ytfN	Hypothetical protein	4932522	Synonymous	Transportation of metabolites	-0.205	Transportation
mdtB	MdtABC-TolC multidrug efflux transport system - membrane subunit	2456426	Synonymous	Transcriptionally activated by CpxR	-0.236	Transportation

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C.

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
ybgP	Putative fimbrial chaperone	849172	Synonymous	Involved in fimbrial attachment	0.202	Biofilm Formation
yahA	Phosphodiesterase, c-di-GMP-specific	386338	Synonymous	May decrease biofilm formation by decreasing c-di-GMP	-0.125	Biofilm Formation
yjcC	Putative c-di-GMP-specific phosphodiesterase	4769269	Missense	Overexpression reduces biofilm formation	-0.131	Biofilm Formation
yjcC	Putative c-di-GMP-specific phosphodiesterase	4769251	Missense	Overexpression reduces biofilm formation	-0.131	Biofilm Formation
yjcC	Putative c-di-GMP-specific phosphodiesterase	4769357	Synonymous	Overexpression reduces biofilm formation	-0.131	Biofilm Formation
ybaJ	Hha toxicity modulator	539581	Intergenic Region ybaJ-acrB	Regulates biofilm formation	-0.137	Biofilm Formation
hha	Haemolysin expression modulating protein	538736	Intergenic Region maa-hha	Decreases biofilm formation by inhibiting fimbrial genes	-0.192	Biofilm Formation
ftsE	Cell division protein ftsE	4035119	Synonymous	Involved in growth	0.258	Cell Division
ftsN	Essential cell division protein ftsN	4609858	Synonymous	Involved in growth	0.257	Cell Division
ftsN	Essential cell division protein ftsN	4609897	Synonymous	Involved in growth	0.199	Cell Division
ftsN	Essential cell division protein ftsN	4609588	Synonymous	Involved in growth	-0.118	Cell Division
nlpD	NlpD putative outer membrane lipoprotein	3172269	Synonymous	Involved in growth	-0.149	Cell Division
cedA	Cell division	2048478	Intergenic Region cedA-katE	Involved in growth	-0.158	Cell Division

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
nlpD	NlpD putative outer membrane lipoprotein	3172422	Synonymous	Involved in growth	-0.263	Cell Division
uvrB	DNA repair; excision nuclease subunit B	961353	Synonymous	Repairs oxidative DNA damage	-0.149	DNA Repair
uvrB	DNA repair; excision nuclease subunit B	961335	Synonymous	Repairs oxidative DNA damage	-0.149	DNA Repair
mutS	MutHLS complex, methyl-directed mismatch repair	3164742	Synonymous	DNA repair, possibly due to oxidative damage	-0.155	DNA Repair
uvrA	Excision nuclease subunit A	4766388	Synonymous	Repairs oxidative DNA damage	-0.168	DNA Repair
uvrB	DNA repair; excision nuclease subunit B	961662	Synonymous	Repairs oxidative DNA damage	-0.174	DNA Repair
uvrB	DNA repair; excision nuclease subunit B	961209	Synonymous	Repairs oxidative DNA damage	-0.188	DNA Repair
efeO	Iron uptake system component	1257333	Synonymous	Involved in iron uptake and homeostasis	0.159	Iron Homeostasis
entF	Enterobactin synthase component F	730487	Missense	Involved in iron uptake and homeostasis	-0.080	Iron Homeostasis
entF	Enterobactin synthase component F	729500	Missense	Involved in iron uptake and homeostasis	-0.090	Iron Homeostasis
fepE	Ferric enterobactin (enterochelin) transport	733649	Synonymous	Involved in iron uptake and homeostasis	-0.107	Iron Homeostasis
fepE	Ferric enterobactin (enterochelin) transport	733424	Synonymous	Involved in iron uptake and homeostasis	-0.115	Iron Homeostasis
fepE	Ferric enterobactin (enterochelin) transport	734065	Missense	Involved in iron uptake and homeostasis	-0.118	Iron Homeostasis
cls	Cardiolipin synthase	1560171	Synonymous	May protect against acidity	0.254	Membrane/Cell Wall
yhiM	Inner membrane protein with a role in acid resistance	4083737	Intergenic Region yhiM-yhiN	Membrane biogenesis	0.236	Membrane/Cell Wall
yhiM	Inner membrane protein with a role in acid resistance	4083536	Synonymous	Membrane biogenesis	0.236	Membrane/Cell Wall
kdsD	D-arabinose 5-phosphate isomerase	3795168	Synonymous	Membrane biogenesis	0.206	Membrane/Cell Wall
yciB	Putative inner membrane protein	1564487	Synonymous	Membrane biogenesis	0.203	Membrane/Cell Wall
murA	UDP-N-acetylglucosamine enolpyruvoyl transferase	3789409	Synonymous	Cell wall biogenesis	0.192	Membrane/Cell Wall
dacB	D-alanyl-D-alanine endopeptidase	3782434	Synonymous	Cell wall biogenesis	0.186	Membrane/Cell Wall
ydjX	Putative inner membrane protein	2068085	Synonymous	Membrane biogenesis	-0.101	Membrane/Cell Wall
chbG	Chito-oligosaccharide mono-deacetylase	2051314	Synonymous	Involved in chitin and glycan degradation, by use of hydrolase	-0.103	Membrane/Cell Wall
chbG	Chito-oligosaccharide mono-deacetylase	2051458	Synonymous	Involved in chitin and glycan degradation, by use of hydrolase	-0.103	Membrane/Cell Wall
chbG	Chito-oligosaccharide mono-deacetylase	2051416	Synonymous	Involved in chitin and glycan degradation, by use of hydrolase	-0.103	Membrane/Cell Wall

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
chbG	Chito-oligosaccharide mono-deacetylase	2051446	Synonymous	Involved in chitin and glycan degradation, by use of hydrolase	-0.103	Membrane/Cell Wall
chbG	Chito-oligosaccharide mono-deacetylase	2051388	Synonymous	Involved in chitin and glycan degradation, by use of hydrolase	-0.103	Membrane/Cell Wall
dacB	D-alanyl-D-alanine endopeptidase	3782962	Synonymous	Cell wall biogenesis	-0.122	Membrane/Cell Wall
yaiO	Outer membrane protein	438263	Synonymous	Membrane biogenesis	-0.124	Membrane/Cell Wall
igaA	Putative membrane protein	3953830	Synonymous	Membrane biogenesis	-0.128	Membrane/Cell Wall
wzxC	Lipopolysaccharide biosynthesis protein	2418333	Synonymous	Membrane biogenesis	-0.168	Membrane/Cell Wall
sohB	Putative inner membrane peptidase	1581581	Synonymous	Membrane biogenesis	-0.171	Membrane/Cell Wall
cls	Cardiolipin synthase	1559224	Synonymous	May protect against acidity	-0.211	Membrane/Cell Wall
sohB	Putative inner membrane peptidase	1581674	Synonymous	Membrane biogenesis	-0.213	Membrane/Cell Wall
kdsD	D-arabinose 5-phosphate isomerase	3794898	Synonymous	Membrane biogenesis	-0.219	Membrane/Cell Wall
cls	Cardiolipin synthase	1559203	Synonymous	May protect against acidity	-0.226	Membrane/Cell Wall
lpxB	Lipid A disaccharide synthase	206549	Synonymous	Membrane biogenesis	-0.322	Membrane/Cell Wall
chbG	Chito-oligosaccharide mono-deacetylase	2050963	Synonymous	Involved in chitin and glycan degradation, by use of hydrolase	-0.433	Membrane/Cell Wall
dxs	1-deoxyxylulose-5-phosphate synthase	497472	Synonymous	Central metabolism	0.313	Metabolite biosynthesis
ansA	Asparaginase I	2083908	Intergenic Region sppA-ansA	Central metabolism	0.284	Metabolite biosynthesis
sppA	Protease IV	2083908	Intergenic Region sppA-ansA	Central metabolism	0.284	Metabolite biosynthesis
citG	Triphosphoribosyl-dephospho-CoA synthase	760860	Synonymous	Central metabolism	0.272	Metabolite biosynthesis
acs	Acetyl-CoA synthetase (AMP-forming)	4779477	Synonymous	Central metabolism	0.269	Metabolite biosynthesis
ispG	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase	2933216	Synonymous	Central metabolism	0.263	Metabolite biosynthesis
yhiN	Putative oxidoreductase with FAD/NAD(P)-binding domain protein	4083737	Intergenic Region yhiM-yhiN	Central metabolism	0.236	Metabolite biosynthesis
cysI	Sulfite reductase, hemoprotein subunit	3193744	Synonymous	Involved in sulfur metabolism	0.225	Metabolite biosynthesis
purH	AICAR transformylase / IMP cyclohydrolase	4695722	Synonymous	Central metabolism; known to be cold induced	0.208	Metabolite biosynthesis
citF	Citrate lyase, citrate-ACP transferase alpha subunit	762094	Synonymous	Central metabolism	0.207	Metabolite biosynthesis
cysI	Sulfite reductase, hemoprotein subunit	3193276	Synonymous	Involved in sulfur metabolism	0.163	Metabolite biosynthesis

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
cysQ	Adenosine-3'(2'),5'-bisphosphate nucleotidase	4926283	Synonymous	Involved in sulfur metabolism	0.147	Metabolite biosynthesis
fadD	Fatty acyl-CoA synthetase	2122351	Synonymous	Central metabolism	0.147	Metabolite biosynthesis
yiaY	Putative Fe-containing alcohol dehydrogenase	4220066	Synonymous	Central metabolism	0.147	Metabolite biosynthesis
metB	O-succinylhomoserine lyase / O-succinylhomoserine(thiol)-lyase	4616007	Synonymous	Involved in methionine synthesis	-0.087	Metabolite biosynthesis
rpe	Ribulose-5-phosphate 3-epimerase	3940671	Synonymous	Central metabolism	-0.110	Metabolite biosynthesis
ydjL	Putative oxidoreductase, Zn-dependent and NAD(P)-binding	2094898	Intergenic Region ydjL-yeaC	Central metabolism	-0.111	Metabolite biosynthesis
citF	Citrate lyase, citrate-ACP transferase alpha subunit	762886	Synonymous	Central metabolism	-0.114	Metabolite biosynthesis
gltX	Glutamyl-trna synthetase	2811045	Synonymous	Expression reduces growth due to stringent response activated by ppGpp	-0.120	Metabolite biosynthesis
galT	Galactose-1-phosphate uridylyltransferase	897872	Synonymous	Central metabolism	-0.122	Metabolite biosynthesis
galT	Galactose-1-phosphate uridylyltransferase	897914	Synonymous	Central metabolism	-0.122	Metabolite biosynthesis
galT	Galactose-1-phosphate uridylyltransferase	897974	Synonymous	Central metabolism	-0.122	Metabolite biosynthesis
gltB	Glutamate synthase, large subunit	3810985	Synonymous	Central metabolism	-0.123	Metabolite biosynthesis
carA	Carbamoyl phosphate synthetase	34525	Synonymous	Pyrimidine synthesis	-0.131	Metabolite biosynthesis
allD	Ureidoglycolate dehydrogenase	626394	Synonymous	Use of allatoin as nitrogen source under aerobic conditions	-0.139	Metabolite biosynthesis
pepD	Peptidase D	327798	Synonymous	Central metabolism	-0.139	Metabolite biosynthesis
adiA	Biodegradative arginine decarboxylase	4826014	Synonymous	Plays role in regulating intracellular pH	-0.142	Metabolite biosynthesis
galK	Galactokinase	896867	Synonymous	Central metabolism	-0.149	Metabolite biosynthesis
fdhE	Formate dehydrogenase formation protein	4570167	Synonymous	Central metabolism	-0.149	Metabolite biosynthesis
glmS	Glutamate mutase sigma subunit	4400322	Synonymous	Central metabolism	-0.154	Metabolite biosynthesis
galK	Galactokinase	896894	Synonymous	Central metabolism	-0.158	Metabolite biosynthesis
galT	Galactose-1-phosphate uridylyltransferase	898097	Synonymous	Central metabolism	-0.162	Metabolite biosynthesis
dadA	D-amino acid dehydrogenase	1484637	Synonymous	Central metabolism	-0.173	Metabolite biosynthesis
adhE	Pyruvate formate-lyase deactivase [multifunctional]	1550718	Synonymous	Central metabolism	-0.177	Metabolite biosynthesis
mdoH	Membrane glycosyltransferase; synthesis of membrane-derived oligosaccharide (MDO)	1277788	Synonymous	Central metabolism	-0.181	Metabolite biosynthesis

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
puuA	Glutamate-putrescine ligase	1611275	Intergenic Region puuP-puuA	Central metabolism	-0.190	Metabolite biosynthesis
ackA	Acetate kinase	2713840	Synonymous	Catalyzes acetylation of CheY, leading to increase in signal strength for flagellar rotation	-0.194	Metabolite biosynthesis
metC	L-cysteine desulfhydrase / cystathionine-beta-lyase	3610632	Synonymous	Involved in methionine synthesis	-0.195	Metabolite biosynthesis
citG	Triphosphoribosyl-dephospho-CoA synthase	760981	Missense	Central metabolism	-0.216	Metabolite biosynthesis
ispG	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase	2932511	Synonymous	Central metabolism	-0.216	Metabolite biosynthesis
allD	Ureidoglycolate dehydrogenase	625938	Synonymous	Use of allatoin as nitrogen source under aerobic conditions	-0.218	Metabolite biosynthesis
glcF	Glycolate oxidase, predicted Iron-sulfur subunit	3574352	Synonymous	Central metabolism	-0.223	Metabolite biosynthesis
allC	Allantoate amidohydrolase monomer	625365	Synonymous	Use of allatoin as nitrogen source under aerobic conditions	-0.243	Metabolite biosynthesis
metH	Cobalamin-dependent methionine synthase	4715146	Synonymous	Involved in methionine synthesis	-0.251	Metabolite biosynthesis
metH	Cobalamin-dependent methionine synthase	4715119	Synonymous	Involved in methionine synthesis	-0.250	Metabolite biosynthesis
citG	Triphosphoribosyl-dephospho-CoA synthase	760836	Synonymous	Central metabolism	-0.253	Metabolite biosynthesis
otsA	Trehalose-6-phosphate synthase	2213757	Synonymous	Central metabolism	-0.256	Metabolite biosynthesis
carA	Carbamoyl phosphate synthetase	34921	Synonymous	Pyrimidine synthesis	-0.328	Metabolite biosynthesis
ydjL	Putative oxidoreductase, Zn-dependent and NAD(P)-binding	2093916	Synonymous	Central metabolism	-0.360	Metabolite biosynthesis
trg	Methyl accepting chemotaxis protein - ribose/galactose/glucose sensing	1686707	Missense	Upregulation of motility has negative effect on biofilm formation	0.196	Motility
flgA	Flagellar biosynthesis; assembly of basal-body periplasmic P ring	1295248	Synonymous	Upregulation of motility has negative effect on biofilm formation	0.179	Motility
flgJ	Peptidoglycan hydrolase	1302736	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.110	Motility
fliL	Flagellar biosynthesis	2253464	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.123	Motility
flgA	Flagellar biosynthesis; assembly of basal-body periplasmic P ring	1295200	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.124	Motility
trg	Methyl accepting chemotaxis protein - ribose/galactose/glucose sensing	1686176	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.173	Motility

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
cheY	Chemotaxis binding protein	2201446	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.176	Motility
fliG	Flagellar motor switch protein fliG	2249437	Synonymous	Upregulation of motility has negative effect on biofilm formation	-0.191	Motility
recB	Helicase/nuclease	3261910	Synonymous	Facilitaties stress-induced mutageneis; may have led to mutation-induced decrease in biofilm	-0.134	Recombination
rbbA	Ribosome-associated ATPase	4077112	Missense	May play a role in regulating genes linked to membrane, protein synthesis and energy metabolism	0.351	Replication/ Transcription/ Translation
rpoN	RNA polymerase, sigma 54 (sigma N) factor	3797968	Synonymous	May affect expression	-0.100	Replication/ Transcription/ Translation
rpoD	RNA polymerase, sigma 70 (sigma D) factor	3668476	Synonymous	May affect expression	-0.105	Replication/ Transcription/ Translation
greA	Transcription elongation factor greA	3781546	Synonymous	May affect expression	-0.106	Replication/ Transcription/ Translation
yehT	Transcriptional regulatory protein	2515198	Synonymous	May affect expression	-0.111	Replication/ Transcription/ Translation
metG	Methionyl-tRNA synthetase	2494144	Synonymous	Involved in methionine synthesis	-0.118	Replication/ Transcription/ Translation
dnaB	Primosome	4757341	Synonymous	Response to oxidative stress	-0.159	Replication/ Transcription/ Translation
dnaB	Primosome	4757380	Synonymous	Response to oxidative stress	-0.159	Replication/ Transcription/ Translation
gntR	GntR DNA-binding transcriptional repressor	4011564	Synonymous	May affect expression	-0.166	Replication/ Transcription/ Translation
dnaX	DNA polymerase III, gamma subunit	552128	Synonymous	May affect expression	-0.187	Replication/ Transcription/ Translation
gutM	GutM DNA-binding transcriptional activator	3134562	Synonymous	May affect expression	-0.188	Replication/ Transcription/ Translation
rbbA	Ribosome-associated ATPase	4074691	Synonymous	May play a role in regulating genes linked to membrane, protein synthesis and energy metabolism	-0.188	Replication/ Transcription/ Translation
rbbA	Ribosome-associated ATPase	4074681	Synonymous	May play a role in regulating genes linked to membrane, protein synthesis and energy metabolism	-0.188	Replication/ Transcription/ Translation
yehT	Transcriptional regulatory protein	2515309	Synonymous	May affect expression	-0.191	Replication/ Transcription/ Translation

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
mfd	Transcription-repair coupling factor	1338987	Missense	May affect expression	-0.221	Replication/ Transcription/ Translation
rbbA	Ribosome-associated ATPase	4076557	Synonymous	May play a role in regulating genes linked to membrane, protein synthesis and energy metabolism	-0.273	Replication/ Transcription/ Translation
sdhA	Succinate dehydrogenase flavoprotein	856595	Synonymous	Electron acceptor in aerobic respiration	0.195	Respiration
ygcR	Putative flavoprotein	3200478	Synonymous	Electron acceptor in aerobic respiration	0.185	Respiration
ygcR	Putative flavoprotein	3200490	Missense	Electron acceptor in aerobic respiration	0.185	Respiration
hcr	NADH oxidoreductase	1099739	Intergenic Region poxB-hcr	NADH Dehydrogenase	0.136	Respiration
poxB	Pyruvate oxidase monomer	1099739	Intergenic Region poxB-hcr	Electron acceptor in aerobic respiration	0.136	Respiration
nuoG	NADH:ubiquinone oxidoreductase, chain G	2698603	Synonymous	Electron acceptor in aerobic respiration	-0.125	Respiration
azoR	NADH:quinone oxidoreductase, FMN-dependent	1674817	Intergenic Region ynbD-azoR	Electron acceptor in aerobic respiration	-0.131	Respiration
yhaH	Putative cytochrome	3707348	Intergenic Region yhaH-yhaI	Electron acceptor in aerobic respiration	-0.145	Respiration
yeiQ	Putative dehydrogenase, NAD-dependent	2569826	Missense	NADH Dehydrogenase	-0.146	Respiration
ispB	Octaprenyl diphosphate synthase	3786715	Intergenic Region rplU-ispB	Involved in ubiquinone synthesis	-0.151	Respiration
ispB	Octaprenyl diphosphate synthase	3786909	Synonymous	Involved in ubiquinone synthesis	-0.151	Respiration
nuoF	NADH:ubiquinone oxidoreductase, chain F	2699981	Synonymous	Electron acceptor in aerobic respiration	-0.167	Respiration
ygfK	Putative oxidoreductase, Fe-S subunit	3341174	Synonymous	Electron acceptor in aerobic respiration	-0.179	Respiration
nuoG	NADH:ubiquinone oxidoreductase, chain G	2696926	Synonymous	Electron acceptor in aerobic respiration	-0.179	Respiration
nuoF	NADH:ubiquinone oxidoreductase, chain F	2700212	Synonymous	Electron acceptor in aerobic respiration	-0.207	Respiration
ygfK	Putative oxidoreductase, Fe-S subunit	3340250	Synonymous	Electron acceptor in aerobic respiration	-0.213	Respiration
yqcA	Putative flavoprotein	3229383	Synonymous	Electron acceptor in aerobic respiration	-0.327	Respiration
yqcA	Putative flavoprotein	3229365	Synonymous	Electron acceptor in aerobic respiration	-0.327	Respiration
groEL	Heat shock protein	4856787	Synonymous	Protects against oxidative stresses and essential for growth	0.258	Stress Response
htpG	Chaperone protein	554316	Synonymous	Stress response to DNA damage; induced during stationary phase	0.194	Stress Response

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
yedV	Putative sensory kinase in two-component regulatory system with yedW	2270828	Synonymous	Involved in repair of oxidized Met residues in bacterial cell envelope proteins through activation of mrsPQ operon	0.189	Stress Response
dsbB	Disulfide bond formation protein B	1479156	Synonymous	Involved in repair of oxidized Cys residues in extracytoplasmic proteins	0.147	Stress Response
ahpF	Alkyl hydroperoxide reductase	755808	Missense	Protects cell against DNA damage by alkyl hydroperoxides	0.140	Stress Response
ksgA	Ribosomal RNA small subunit methyltransferase A	56342	Synonymous	May play a role in protection of DNA against oxidative stress	0.112	Stress Response
ksgA	Ribosomal RNA small subunit methyltransferase A	56354	Synonymous	May play a role in protection of DNA against oxidative stress	0.112	Stress Response
btuE	Thioredoxin/glutathione peroxidase	2028345	Synonymous	Defense against oxidative stress conditions; possible reactive oxygen scavenger	-0.097	Stress Response
msrB	Peptide methionine sulphoxide reductase	2095292	Synonymous	Involved in protein repair due to oxidative stress; repairs oxidized methionine	-0.119	Stress Response
yfiQ	Protein lysine acetyltransferase	3014483	Synonymous	Involved in response to oxidative stress	-0.137	Stress Response
yfiQ	Protein lysine acetyltransferase	3014420	Synonymous	Involved in response to oxidative stress	-0.137	Stress Response
yfiQ	Protein lysine acetyltransferase	3014459	Synonymous	Involved in response to oxidative stress	-0.137	Stress Response
yfiQ	Protein lysine acetyltransferase	3014714	Synonymous	Involved in response to oxidative stress	-0.137	Stress Response
yfiQ	Protein lysine acetyltransferase	3014471	Synonymous	Involved in response to oxidative stress	-0.137	Stress Response
yfiQ	Protein lysine acetyltransferase	3014610	Missense	Involved in response to oxidative stress	-0.137	Stress Response
yfiQ	Protein lysine acetyltransferase	3014438	Synonymous	Involved in response to oxidative stress	-0.137	Stress Response
clpA	ATP-dependent Clp protease ATP-binding subunit	1112828	Synonymous	Degrades unfolded or abnormal proteins	-0.151	Stress Response
cpxR	Transcriptional regulatory protein	4590896	Synonymous	Induced during cell envelope stress, important for biofilm formation, but induction by stress may indicate conditions that decrease biofilms	-0.157	Stress Response
katE	Heme d synthase / hydroperoxidase	2048478	Intergenic Region cedA-katE	Protects against oxidative deleterious reactions	-0.158	Stress Response
asr	Acid shock protein	1904826	Intergenic Region ynfM-asr	Involved in acid resistance, but may not be beneficial at low temperatures	-0.159	Stress Response
gor	Glutathione reductase (NADPH)	4094001	Synonymous	Protects against oxidative deleterious reactions	-0.160	Stress Response
yjcQ	Multidrug resistance protein	4795030	Missense	Involved in general stress response	-0.161	Stress Response

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
ycfR	Multiple stress resistance protein	1335677	Synonymous	Involved in general stress response	-0.161	Stress Response
dps	Stationary phase nucleoid protein that sequesters Iron and protects DNA from damage	994563	Synonymous	Protects DNA from oxidative damage based on sequestration of iron ions	-0.166	Stress Response
dps	Stationary phase nucleoid protein that sequesters Iron and protects DNA from damage	994857	Synonymous	Protects DNA from oxidative damage based on sequestration of iron ions	-0.167	Stress Response
dsbA	Thiol:disulfide interchange protein dsbA - putative GTP-binding protein	4530611	Intergenic Region dsbA-yihF	Involved in repair of oxidized Cys residues in extracytoplasmic proteins	-0.180	Stress Response
cpxA	Sensor histidine kinase	4589982	Synonymous	Induced during cell envelope stress, important for biofilm formation, but induction by stress may indicate conditions that decrease biofilms	-0.189	Stress Response
gyrB	DNA gyrase, subunit B	4363390	Synonymous	Plays role in protection against oxidative damage	-0.209	Stress Response
ybhJ	Putative hydratase, aconitase	913157	Synonymous	Involved in release of free iron from aconitase, which exacerbates oxygen stress	-0.210	Stress Response
ydeI	Stress response protein	1833499	Intergenic Region ydeH-ydeI	Involved in general stress response	-0.213	Stress Response
uspG	Universal stress protein UP12	756448	Synonymous	Involved in general stress response	-0.214	Stress Response
adiC	AdiC arginine:agmatine antiporter	4821637	Synonymous	Involved in acid resistance, but may not be beneficial at low temperatures	-0.219	Stress Response
dsbG	Thiol:disulfide interchange protein dsbG	752978	Missense	Involved in repair of oxidized Cys residues in extracytoplasmic proteins	-0.251	Stress Response
sdhD	Succinate dehydrogenase membrane protein	855159	Synonymous	Central metabolism	0.335	TCA Cycle
sdhD	Succinate dehydrogenase membrane protein	855321	Synonymous	Central metabolism	0.254	TCA Cycle
fumC	Fumarase C monomer	1920563	Synonymous	Central metabolism	0.149	TCA Cycle
fumC	Fumarase C monomer	1920854	Synonymous	Central metabolism	-0.168	TCA Cycle
sdhC	Succinate dehydrogenase membrane protein	854792	Missense	Central metabolism	-0.279	TCA Cycle
ybbL	Putative transporter subunit: ATP-binding component of ABC superfamily protein	603519	Synonymous	Transportation of metabolites	0.282	Transportation
kefB	K : H antiporter kefB	3910378	Synonymous	Transportation of metabolites	0.237	Transportation
ybaL	YbaL CPA2 transporter	560016	Synonymous	Transportation of metabolites	0.182	Transportation
modF	Putative molybdenum transport ATP-binding protein	900729	Synonymous	Transportation of metabolites	0.105	Transportation
yhhJ	Putative transporter subunit: membrane component of ABC superfamily protein	4074125	Synonymous	Transportation of metabolites	-0.076	Transportation

Table A2: List of genetic variants that were determined to have a significant association with biofilm density at 15°C (continued).

Gene(s)	Annotation	Location	Class	Predicted Effect of Variant	Effect on Biofilm	Function
metI	L-methionine / D-methionine ABC transporter - membrane subunit	224370	Synonymous	Involved in methionine synthesis	-0.113	Transportation
metQ	L-methionine / D-methionine ABC transporter - periplasmic binding protein	223257	Synonymous	Involved in methionine synthesis	-0.126	Transportation
cynX	Cyanate transporter	418506	Synonymous	Transportation of metabolites	-0.131	Transportation
mgtA	Mg2 / Ni2 transporting ATPase	4956936	Synonymous	Transportation of metabolites	-0.161	Transportation
ybbL	Putative transporter subunit: ATP-binding component of ABC superfamily protein	603243	Synonymous	Transportation of metabolites	-0.162	Transportation
oppF	Murein tripeptide ABC transporter / peptide ABC transporter - putative ATP binding subunit	1557964	Synonymous	Transportation of metabolites	-0.164	Transportation
mgtA	Mg2 / Ni2 transporting ATPase	4958202	Intergenic Region mgtA-yjgF	Transportation of metabolites	-0.186	Transportation
acrB	AcrB RND-type permease	542382	Synonymous	May play a role in contact-dependent growth inhibition	-0.189	Transportation
mgtA	Mg2 / Ni2 transporting ATPase	4958245	Intergenic Region mgtA-yjgF	Transportation of metabolites	-0.199	Transportation
mgtA	Mg2 / Ni2 transporting ATPase	4957155	Synonymous	Transportation of metabolites	-0.207	Transportation
glnH	Glutamine ABC transporter	993836	Synonymous	Transportation of metabolites	-0.210	Transportation
gadC	Glutamic acid:4-aminobutyrate antiporter	1767540	Synonymous	Transportation of metabolites	-0.212	Transportation
oppF	Murein tripeptide ABC transporter / peptide ABC transporter - putative ATP binding subunit	1558379	Synonymous	Transportation of metabolites	-0.213	Transportation
zntA	Zinc, cadmium and lead efflux system	4041026	Synonymous	Transportation of metabolites	-0.219	Transportation
zntA	Zinc, cadmium and lead efflux system	4041110	Synonymous	Transportation of metabolites	-0.219	Transportation
oppD	Murein tripeptide ABC transporter / peptide ABC transporter - putative ATP binding subunit	1557410	Synonymous	Transportation of metabolites	-0.233	Transportation
ybbW	YbbW NCS1 Transporter	618399	Missense	Transportation of metabolites	-0.240	Transportation
yhhJ	Putative transporter subunit: membrane component of ABC superfamily protein	4074320	Synonymous	Transportation of metabolites	-0.253	Transportation
zntB	Zinc transport protein	1646137	Synonymous	Low zinc levels can inhibit biofilm formation	-0.386	Transportation