

Genetic Diversity in *Escherichia coli*

Benyam Zenebe^{1*}, Tesfaye Sisay², Gurja Belay¹

¹Department of Microbial, Cellular and Molecular Biology, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia; ²Department of Biotechnology, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia

ABSTRACT

E. coli are genetically diverse bacteria which can live in different habitats. The *E. coli* genome has two components: The core genome which contains essential genes and is shared among all strains of the species; and the accessory genome which is highly variable among strains encompassing many mobile genetic elements. Despite the extensive dynamics in the genomic content over evolutionary period, the genome size of *E. coli* seems to have been conserved due to processes of continuous gene gain and gene loss. Mobile genetic elements give strains of *E. coli* better adaptability to different environments such as the ability to resist antibiotics and to escape immune responses and establish infection in various hosts. Mobile genetic elements have been detected in both commensal and pathogenic strains. Horizontal gene transfer in the form of transformation, transduction, and conjugation plays a big role in the genetic diversification, evolution, and adaptation of various *E. coli* strains besides genome mutation events which vary in rates among wild type and domestic strains with domestic strains displaying increased mutability due to reduced methyl directed mismatch repair activity.

E. coli genetic diversity has been studied from the stand point of traits such as: Antibiotic resistance, repeat sequences, insertion sequences, bacteriophage susceptibility, DNA hybridization characteristics, plasmids, restriction fragment polymorphisms and different molecular techniques have been devised to study the genetic diversity of *E. coli* strains based on such traits. *E. coli* bacteria are highly genetically diverse as indicated by the recovery of genetically diverse strains from different habitats by different researchers. Some multi drug resistant strains have attained a worldwide distribution. Host, environment or season related genetic diversity or relatedness is observed among phylogeny groups and strains of *E. coli* although different studies on the host/environment based genetic structure of *E. coli* do not come up with similar results.

Keywords: *E. coli*; Genetic diversity; Population genetics; Genome variation; Gene transfer

INTRODUCTION

Diversity is the degree of uniqueness or variability that is displayed among individuals or classes of a population/s. Diversity can also be defined as the condition of being different, variety, and a point of difference. Studying the diversity of different populations can be helpful for scientists to discriminate population classes according to their defined characteristics. Besides, studying diversity is also important to treat, manipulate,

assess, and deal with different classes or individuals of a given population according to their specified characteristics [1].

A wide range of definitions of genetic diversity are employed by scientists in different evolutionary and ecological studies. Some of the definitions of genetic diversity commonly used are: Allelic diversity (the average number and relative frequency of alleles per locus); allelic richness (the average number of alleles per locus); genotypic richness (number of genotypes within a

Correspondence to: Benyam Zenebe, Department of Microbial, Cellular and Molecular Biology, College of Natural and Computational Sciences, Addis Ababa University, Addis Ababa, Ethiopia, Tel: +251911627679; E-mail: benyam.zenebe@aaau.edu.et

Received: 29-Jul-2023, Manuscript No. AMOA-23-25865; **Editor assigned:** 31-Jul-2023, PreQC No. AMOA-23-25865 (PQ); **Reviewed:** 14-Aug-2023, QC No. AMOA-23-25865; **Revised:** 14-Jun-2024, Manuscript No. AMOA-23-25865 (R); **Published:** 21-Jun-2024, DOI: 10.35248/2471-9315.24.10.311

Citation: Zenebe B, Sisay T, Belay G (2024) Genetic Diversity in *Escherichia coli*. Appl Microbiol Open Access. 10:311.

Copyright: © 2024 Zenebe B, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

population); heterozygosity (average proportion of loci that carry two different alleles at a single locus within an individual); mutational diversity and effective population size (a measure of nucleotide diversity that provides a combined measure of effective population size and mutation rate); nucleotide diversity (average number of nucleotide differences per site between two random individuals selected from a population); percentage of polymorphic loci (percentage of loci that are polymorphic).

Genetic diversity is an ingredient of biodiversity and is a prerequisite for evolution by natural selection, *i.e.* genotypes may vary in ecologically important ways there by contributing to the fitness of individuals of a population. Molecular markers are the most common tools by which genetic diversity of allelic states is assayed in natural populations. Genetic diversity can affect a wide range of populations, communities and ecosystems of organisms. A population's genetic diversity is influenced by mutation, natural selection, migration and genetic drift [2,3].

Genetic diversity is also affected by the size of a population. A reduction in population size in relation to a random event may lead to a reduction in genetic variability due to genetic drift. This decrease in genetic variability may further lead to reduced fitness. This reduced fitness may sometimes lead to extreme loss of individuals of a population which is termed genetic bottle neck. Genetic diversity may be imputed into a population either through mutation or gene flow from a neighboring population, *i.e.*, migration. Apart from genetic drift, genetic variability may decrease due to natural selection in inbreeding populations. In this case, genetic drift tends to be passive while natural selection tends to be an active ingredient of the reduction in a population's genetic variability [4].

Classification of microorganisms according to their species levels is helpful in studying their genetic diversity. A species is a group of organisms that can be engaged in reproduction. As organisms reproduce and genes replicate, changes occur in populations resulting in speciation. A clear demarcation between species of a genus is helpful for various biological applications. Plants and animals can be classified according to their species based on their morphological features. Classification based on morphological features alone may not be convenient for the classification of microorganisms because microbes have simple morphology. Advances in molecular techniques have resulted in clues to the tremendous genetic diversity present within the microorganisms such as bacteria and fungi. Increasing knowledge of the properties of DNA with the development of molecular biological techniques has resulted in the idea that bacteria might best be classified by comparing their genomes. This may be helpful in formulating ecological and evolutionary theories by identifying the DNA sequence clusters most likely corresponding to the fundamental units of bacterial genetic diversity [5-7].

LITERATURE REVIEW

Bacteria generally display a wide range of intra species genetic variation which may be presented not only in the form of sequence variation as in the case of other species but also the presence or absence of whole genes or clusters of genes. It is

thus important to have the sequence of all bacterial DNA to have a species genome in bacteria unlike eukaryotes in which case an individual organism's genome sequence may provide with the vast majority of the genes for the species in which the organism is found. Bacterial genome variation is also observed in the fact that bacteria carry plasmids and bacteriophages and that these plasmids and phages are present in some strains but are absent in other strains. Furthermore, there are groups of genes present in the chromosomes of some strains but not in others. Many of these groups of genes may be pathogenicity islands. Bacterial strains may also vary in genome sequences that code for enzymes, *e.g.*, those involved in metabolism. Species of bacteria can receive genes from other closely related species through horizontal gene transfer; this adds to the genetic variation of different strains of bacterial species. Events of horizontal gene transfer can result in differences in adaptation characteristics to different ecologies among species of bacteria. Antimicrobial resistance genes and plasmids which confer pathogenic character can be transferred from one species of bacteria in to another. These transferred genetic substances give bacteria better adaptive characters so that they can effectively escape host defense mechanisms. Horizontal gene transfer thus increases bacterial genetic diversification. The organization of bacterial genes into operons may reflect an evolutionary history of horizontal genetic transfer where by selfish operons may present themselves as a package of functionally coupled genes (genes involved in the same biological function up on integration into various genomes). Horizontal gene transfer in this case is usually limited to operons with simple, nonessential cellular functions. Although the rate of recombination in bacteria is rare, recombination can add up to result acquisition of genetic characters for altered adaptation from other organisms due to the enormous population sizes of bacteria. Besides horizontal gene transfer, mutation is another mechanism of genetic diversification in bacteria [8,9].

There is size and DNA composition variation among the genomes of prokaryotes. The degree of genetic variation observed in prokaryotes is by far higher than that observed in that of eukaryotes.

E. coli has numerous commensal and pathogenic variants which are highly variable in terms of gene content and which may have undergone various episodes of gene rearrangements including gene acquisition and loss. Hence it can be employed as a model organism in the study of evolutionary processes of prokaryotes. *E. coli* lives in diverse forms in nature. *E. coli* genetic diversity has been studied from the standpoint of traits such as antibiotic resistance, susceptibility to bacteriophage, DNA hybridization, restriction fragment length polymorphisms, DNA sequences, insertion sequences, plasmids, etc. There is high linkage disequilibrium in *E. coli* populations resulting in highly nonrandom combinations of alleles at different loci. Reproduction is largely asexual, with insufficient genetic recombination. Plasmids and other extra chromosomal genetic elements represent a special category because of their ability to transfer horizontally among clones of the species. Phylogenetically, *E. coli* has recently been classified into eight groups: A, B1, B2, C, D, E and F. There are key genomic islands that determine the resulting genetic and phenotypic differences

among variants of the species of *E. coli*. Pathogenic strains, for example, contain genomic regions with virulence genes that code for different traits like colonization, invasion, secretion of toxic compounds, etc. Current knowledge of genotypic and phenotypic diversity in the species *E. coli* is based almost entirely on strains recovered from humans or zoo animals. Observed genetic diversity of *E. coli* exhibits both host, taxonomic and environmental components. The genetic structure of *E. coli* is shaped by combined host and environmental factors; genomic determinants involved in characters like virulence of the bacteria reflect the species' adaptability to different habitats [10-13].

The species of *E. coli* have served as good models for the study of evolution in prokaryotes. It is, therefore, of advantage to review literatures written on the genetic diversity of *E. coli* as this can add to the accumulation of academic knowledge on the diversification and speciation characteristics of prokaryotes in general and of bacteria in particular. Research works published by various researchers were browsed from google scholar and pub med data bases for this review. An attempt was made to include peer reviewed, online available, open access articles. Meta-analysis reviews, systematic reviews and articles with contents of only an abstract were excluded.

***Escherichia coli* genome dynamics, intra species variation and adaptation**

The genomes of microbes display different degree of variability among different strains. The most stable microbial genomes identified so far are those of the genus *Buchnera aphidicola*. They are believed to have undergone no rearrangements or horizontal gene transfers in their genomes during the past many years despite high nucleotide substitution rates. The *E. coli* genome in contrast, is among those bacterial species which show high genome variability with high ratio of rearrangements to substitutions, *i.e.*, there is higher fixation rate for mutations that alter genomic structures relative to the effect of nucleotide substitutions. One reason for the pronounced evolutionary dynamics of *E. coli* and other related bacterial species may be a result of high recombination frequencies at the open replication forks of their genomes that lead to symmetric translocation and inversion events. Insertion Sequence (IS) elements may be important mediators of such gene rearrangements by offering multiple sequences among which recombination can be initiated. Insertion deletion rearrangements are other factors that account for increased genome flexibility. Generally, *E. coli* and other related bacterial genomes show high structural and functional variability, and the degree of genome flexibility is dependent on the content of repeated and mobile sequences such as IS elements, plasmids and phages. Although most analyses have focused on pathogenic organisms, the increased variation in the genome content of *E. coli* and related bacterial genomes is not to be left to pathogenic species only [14].

The *E. coli* genome consists of two parts: A core genome (a mosaic-like structure) which encodes essential cellular functions and flexible genomic parts which are interspersed variable regions. The core genome of a species is phylogenetically relatively stable. It contains essential genes which are shared by all strains of a species those genes whose functions are always

needed. Genes in this region are not subject to transfer among strains except under rare conditions. If such genes are subject to transfer, they are replaced by a functioning resident copy of a foreign gene with the same activity, usually an ortholog. Genes that belong to the flexible genomic regions code for factors involved in fitness and adaptation of the organism to different environments.

Pathogenic as well as nonpathogenic strains carry mobile and accessory genetic elements such as plasmids, bacteriophages, genomic islands and others, which code for adaptive traits of the bacteria. Genome dynamics processes like gene transfer, genome reduction, rearrangements and point mutations contribute to the adaptation of the bacteria into different environments. Genetic elements which are found in the flexible region are involved in such genome dynamics processes. There is high degree of flexibility in the *E. coli* genome as is seen in the fact that the genomes of many strains do share only a certain percent of their coding capacity, *i.e.*, only a certain portion of the coding region of the *E. coli* genome is shared among different strains.

The genome segment surrounding the replication terminus displays the most variation in chromosome size and organization of the *E. coli* genome most probably due to increased levels of recombination in this region. Horizontal gene transfer adds to genetic variation in genomic content among different strains of *E. coli*; the contribution of horizontal gene transfer is, in fact, greater than all forms of mutations. Horizontal gene transfer can take place between different organisms through transformation (uptake of DNA from the environment), transduction (packaging and transport of bacterial DNA by viruses), and conjugation (bacterial mating). Much of the horizontally transferred DNA is part of the flexible gene pool. Virulence genes of pathogenic strains, for example, are found on mobile genetic elements such as genomic islands (dynamic, ancient integrative elements in bacterial evolution which are important sources of genomic diversification and adaptation), bacteriophages, plasmids or transposons. Virulence genes can thus be transferred from one *E. coli* bacteria to another or can be acquired from other species of bacteria and contribute to the development of adaptive characters of different pathogenic *E. coli* strains [15-17].

Horizontal gene transfer may produce mosaic chromosomes comprised of genes which differ in ancestries and duration in the genome. Horizontally transferred genes in *E. coli* chromosomes mostly get inserted next to tRNA loci. This is a site where lysogenic phages get inserted; bacteriophages thus serve as vehicles for the introduction of horizontally transferred genes in these sites. Horizontal gene transfer has contributed to the emergence of pathogenic strains and new adaptive traits in *E. coli*.

There is variation in base composition of the *E. coli* genome in which variation in base composition occurs due to biases in mutation rates in the four nucleotide bases: A, T, C and G. This phenomenon serves as a basis for identifying horizontally transferred genes in the flexible region of the *E. coli* genome. Horizontally transferred genes initially have the base composition of the donor genome. After they are transferred to a recipient genome, the genes are subject to mutational

processes which take place in the recipient genome; the acquired gene sequences will thus undergo substitutions and finally will reflect the DNA composition of the recipient genome. This is called 'gene amelioration'. This process of adjustment of the acquired gene sequences to the base composition of the recipient genome is a function of the relative rate of G/C to A/T mutations. Based on substitution rates estimated for *E. coli* and the mutational bias of this species, it is possible to predict the amount of time required for a transferred gene to fully resemble native DNA because the nucleotide composition at each position of horizontally transferred genes ameliorates at a characteristic rate.

Variation in the rate of amelioration at each codon position gives a unique property to horizontally transferred genes and allows scientists to estimate the amount of time a transferred gene has been residing in a genome. Recently transferred genes show the patterns of nucleotide composition which resemble that of the donor genome, and fully ameliorated genes show the nucleotide compositions which resemble that of the recipient *E. coli* genome. According to amelioration analysis, much of the recently acquired *E. coli* DNA includes IS elements, remnants of prophages, and other sequences. On the other hand, the rate of horizontal gene transfer can be estimated by analyzing the amount of horizontally transferred DNA and the time period the transferred gene has been residing in the recipient genome. Although chromosome size of *E. coli* isolates do vary, the *E. coli* genome has not shown a steady increase in size over the evolutionary periods because chromosome lengths are conserved in *E. coli*. This implies that, on an evolutionary timescale, increases in genome size due to the acquisition of horizontally transferred sequences are compensated by equivalent losses of DNA through deletion. These processes of gene gain and gene loss result in an extremely dynamic genome in which substantial amounts of DNA are introduced into and/or deleted from chromosomes which bring about change in adaptation characters of the species.

E. coli have long been used as a model for genetic research; it is thus the ideal microorganism for studies at the genomic level. Mutation is a biological force that derives variation; therefore assessment of mutation rates of a species is important to understand the ecological and evolutionary characteristics of the species. Mutations may occur due to errors in DNA replication. There is generally low rate of spontaneous mutation in the genome of wild type *E. coli* species due to an enhanced accuracy of DNA replication and many enzymatic activities which control and repair DNA among which is Methyl directed Mismatch Repair (MMR) which controls newly replicated DNA. In this enzymatic activity, mismatched bases, if any, are detected and enzymes are recruited to destroy the new DNA strand with mismatched bases; re-polymerization is then undertaken using the old DNA strand as a template. Spontaneous mutation rate in wild type *E. coli* is thus lower than other microbes which do not undergo MMR. Although MMR is found in all domains of life, some bacterial strains isolated from natural environments often do not undergo this process in their genomes. In this case, MMR deficient strains are likely to be dominant after long term evolution. Loss of MMR increases mutation rate in *E. coli* by favoring base pair substitution mutations. Certain domestic *E.*

coli strains may have lost their MMR capacity over evolutionary periods and thus exhibit higher genome mutability. Base pair substitution mutations usually occur in coding regions of the strains than in non-coding regions. This shows that MMR pathway mainly prevents errors of replication in the coding regions of the wild type *E. coli* genome rendering coding regions less susceptible to mutation than non-coding regions; mutations occurring in non-coding regions may be usually neutral and contribute less to the adaptability of the species. There are many factors that contribute to continuous adaptation as far as mutation is concerned: The amount of time required for a beneficial mutation to be fixed in a population, selective advantage of the beneficial mutation, the population size, and random genetic drift. As discussed above, horizontal gene transfer is another biological phenomenon that leads to the genetic diversity and altered adaptability of *E. coli*. Horizontal gene transfer makes the tree of life hypothesis of evolution of less effect in that genes are not only acquired vertically from parent to offspring; they can also be acquired horizontally from a neighboring organism.

Mutation and recombination of vertically inherited genes can give novel characters to offspring generation resulting in species diversification which may be much pronounced over long periods of time but horizontally acquired genes can better add to species diversity by integrating entirely new genes in to a species genome there by creating more novel characters with in short periods of time. The rate of horizontal gene transfer in *E. coli* and other bacteria depends on: Availability of environmental DNA; rate of integration of foreign DNA into recipient genome by conjugation, transformation, and transduction; degree of successful integration of foreign DNA into recipient genome; probability of fixation of the new allele; and degree of the benefit resulted from the expression of the integrated gene. In prokaryotes like *E. coli*, horizontal gene transfer has greater potential of effecting species genetic diversification than mutation. Mutations can promote adaptability of a species by modifying existing functions to some degree there by contributing to a gradual ecological expansion of a species while horizontal gene transfer renders a species effectively adaptive to a new habitat by integrating new DNA into the species genome there by giving the species genome completely new adaptive character the introduction of foreign DNA is more effective in creating better adapted traits than genomic rearrangements [18,19].

Variants of the species of *E. coli* are adapted to various host organisms; *E. coli* lives as a commensal in the gut of animals and humans, as a pathogen causing a wide range of infections, and as model organism for various biological applications. Among the many isolates, only the strains K-12, B, C, and W are commonly used in the laboratory. These strains and their derivatives are designated as risk group 1 organisms in biological safety guidelines. K-12 strain is the reference strain which was used in the sequencing of the first *E. coli* genome. *E. coli* W has the largest chromosome of all the sequenced standard laboratory strains having 4,901 Kbps. Comparatively its genome is similar in size to the commensal *E. coli* strain SE11 which has a chromosome size of 4,888 Kbps but it is smaller than most sequenced pathogenic strains. 89% of the chromosome of K-12

strains is comprised of 4,764 genes with 82 non-coding RNA genes and has 4288 protein coding genes. These standard laboratory strains of *E. coli* belong to phylogenetic group A; commensal strains are mainly found in phylogroups A or B1 while pathogenic strains belong to remaining phylogroups.

The conserved core genome of *E. coli* is shared with closely related genera such as *Shigella* and *Salmonella* suggesting that these organisms share common ancestral relation. The accessory genomic contents have different G+C content from the core genome. They are a result of lateral gene transfer and they give the bacteria adaptive characters to their respective environments such as the ability to cause pathogenicity in different mammals or to produce some enzymes involved in cellular activities required for survival in altered habitats. The ecological and genetic diversity of *E. coli* strains is a prerequisite for the use of *E. coli* as a model organism to study processes of bacterial genome evolution. The genome contents of commensal, pathogenic and laboratory strains of *E. coli* are heterogeneous. There is sequence alteration in the vicinity of tRNA genes due to the integration of foreign DNA fragments. The insertion of genomic islands or pathogenicity islands and plasmids into tRNA loci has been described in many pathogenic and nonpathogenic bacteria. Insertion sequences also play important roles in helping strains of *E. coli* to evolve and adapt to new environments. Generally, the gene content variation and degree of genomic alteration among both pathogenic and commensal *E. coli* isolates is very high. This may be indicative of frequent acquisition of foreign DNA by horizontal gene transfer and processes of gene gain and deletion during the evolution of the genome of the bacteria [20,21].

Methods employed to study genetic diversity of *Escherichia coli*

Information about the intra species genetic variation within bacteria helps to identify and classify different strains as well as to examine evolutionary divergence and population dynamics of different species. Various methods exist for observing genetic variation in *E. coli* strains. In this section we will discuss selected examples of molecular methods used to study the genetic diversity of different *E. coli* sub types in different habitats.

Analysis of Simple Sequence Repeats (SSRs)

Simple sequence repeats are tracts of repetitive DNA in which certain DNA motifs variable in length are repeated many times. They may occur at numerous locations within an organism's genome. They have a higher mutation rate than other areas of DNA and this high mutation rate leads to high genetic diversity or polymorphism to this area of the genome. They represent hyper mutable loci of the genome which is thus subject to reversible changes in length [22,23].

Several such repeats of DNA sequences are found in the *E. coli* genome. Mutations resulting in these repeats of DNA may result in their expansion or contraction. Such mutations can be introduced during various cellular processes affecting the DNA, including replication, recombination and different repair mechanisms. There are different classes of these tandem repeats

such as long SSRs, short SSRs, mono nucleotide SSR, dinucleotide SSR, etc. Gur-Arie and his colleagues screened an entire genome of *E. coli* K-12 strains for the presence, locations, and composition of simple sequence repeats. They tested their observations against the null hypotheses that simple sequence repeats are randomly distributed among coding and non-coding regions of the specified *E. coli* strain. They wanted to show that simple sequence repeats are unevenly distributed among coding and non-coding regions and that these repeats are polymorphic among *E. coli* strains and thus can be used to characterize the different strains of *E. coli*. Computer based genome-wide screening of the DNA sequence of *Escherichia coli* strain K-12 was employed to assess the distribution of these simple sequence repeats across the genome. Their results confirmed that simple sequence repeats in *E. coli* are numerous and diverse in terms of length and repeat number and showed that they are widely distributed throughout the genome a total of 235,495 simple sequence repeats were found which were distributed evenly throughout the genome. To assess the likelihood that simple sequence repeats might affect gene expression resulting in diversification, the positions of all simple sequence repeats in the genome with regard to open reading frames were examined. A DNA sequence immediately upstream of an open reading frame contains regulatory elements that play important roles in controlling gene expression.

The presence and variation of simple sequence repeats in upstream regulatory elements might affect the expression of open reading frames in such a way that it can turn either on or off gene expression, or the effect can be quantitative wise affecting the amount of genes to be expressed. In eukaryotes simple sequence repeats are located in non-coding regions of the genome that have little or no effect on gene expression. Because of the compact nature of prokaryotic genomes, intergenic regions are usually short, and simple sequence repeats in non-coding regions are found upstream open reading frames where their variation can affect gene expression. In the *E. coli* strain under the aforementioned study, many of the simple sequence repeats were found to be localized up to 200 bps from the ATG, the first codon of open reading frames. DNA sequence analyses showed that a number of mono-nucleotide simple sequence repeats on non-coding regions upstream open reading frames were polymorphic, exhibiting two to four alleles for each simple sequence repeat number. Screening for polymorphisms among strains of *E. coli* was conducted at 14 arbitrarily chosen loci containing simple sequence repeats. DNA at chosen loci was amplified by PCR using primers flanking the particular SSR locus. Repeat number polymorphism at the *ycgW* locus was observed as differential mobility of radioactively labeled amplification products through polyacrylamide gels demonstrating hyper variable single-locus DNA fingerprint bands distinguishing among *E. coli* strains; four of the examined SSRs exhibited length polymorphism among strains of *E. coli*. All four polymorphic sites shared similar characteristics in that they involved mono-nucleotide SSRs in non-coding regions upstream open reading frames. Simple sequence repeat polymorphism in *E. coli* is therefore important for genome wide variation assessment, strain identification, etc. [24].

Analysis of Repetitive Extragenic Palindromic (REP) sequences

Repetitive palindrome sequences are major components of the bacterial genome. They include inverted repeats and they can be found singly or in adjacent copies in the genome. These repetitive sequences show polymorphism across different strains and have been used to compare the genomes of *E. coli* strains in studies of environmental sources of fecal contamination. One approach to detect the source of fecal contamination is to compare genetic profiles of *E. coli* strains isolated from contaminated water with strains collected from suspected sources.

A general assumption can be made such that a host-specific genetic structure exists across the *E. coli* population [25]. McLellan, et al. characterized *E. coli* from different host sources of fecal pollution: Sewage treatment from residential areas, expected to be of human sources; birds from lake beaches; and cattle from farm sites within watershed. The genetic diversity of strains from the different host sources was characterized by using DNA fingerprints generated by REP-PCR with REP primers. *E. coli* isolates that were collected from a single bird fecal sample gave identical REP fingerprint patterns for majority of isolates analyzed with few exceptions. Similar results were obtained with isolates obtained from human and cattle samples. In contrast, a high degree of genetic variation was found among fingerprint patterns of *E. coli* isolates collected from different bird and human fecal samples. The cattle isolates demonstrated the highest degree of genetic similarity among strains; identical strains occurred across different farm sites as well as within single farms. REP-PCR finger prints generated from the different hosts (bird, cattle, human) were generally diverse in this study.

DISCUSSION

In another study, the polymorphism of the Clustered Regularly Inter-Spaced Short Palindromic Repeat (CRISPR) regions of Enterohemorrhagic *Escherichia coli* (EHEC) was assessed with an aim to design PCR assays for the different EHEC serotypes. CRISPRs contain tandem sequences containing direct repeats of 21 to 47 bps separated by spacers of similar size. This study focused on the CRISPR loci of EHEC strains associated with the most frequent clinical cases with the aim to identify conserved CRISPR motifs associated with EHEC strains of the main 7 serotypes. Sequencing the CRISPR loci of various EHEC strains has shown the genetic diversity of its sequences derived from EHEC O157:H7, O26:H11, O145:H28, O103:H2, O111:H8, O121:H19, and O45:H2 strains. The CRISPR PCR assays developed in the stated study strongly correlated with both O:H serotypes and the presence of EHEC virulence factors (*stx* and *eae* genes). In general, the CRISPR locus of enterohemorrhagic *E. coli* exhibits polymorphism, the different sub strains EHEC can thus be identified through specifically designed primers by PCR [26].

Multi Locus Sequence Typing (MLST)

This is a technique used for the sequencing of internal fragments of multiple housekeeping genes of microbial genomes. MLST is used to characterize strains by their unique allelic profiles. It has been used to assign *E. coli* strains into different phylogenetic groups. *E. coli* consist of a number of distinct phylogeny-groups and strains of the different phylogeny-groups vary in their host preferences, phenotypic characteristics and the tendency to cause disease. Therefore we can learn a lot by assigning strains of *E. coli* into one of the known phylogeny-groups. A method of assigning *E. coli* strains into different groups depending on the presence/absence of two genes *chuA* and *yjaA* and a DNA fragment TSPE4.C2 was developed by Clermont and his colleagues in the year 2000. Validation of this method was based on few strains which were limited to human and human related hosts. No further study has been undertaken to validate the study in a wide range of strains from variety of hosts. Multi-locus sequence typing data for isolates of *E. coli* was used to determine the frequency with which *E. coli* strains can be appropriately assigned to phylogeny-groups established by Clermont and his colleagues. Strains from different host ranges were used. Population assignment algorithms such as BAPS (Bayesian Analysis of Population Structure) and neighbor joining were used to assign strains to phylogeny-groups based on the multi-locus sequence typing data. The analyses revealed that 85%-90% of *E. coli* strains can be assigned to a phylogeny-group and that 80%-85% of the phylogeny-group memberships assigned using the Clermont method are correct [27].

Similarly, multi locus sequence typing was used to investigate the genetic relatedness among shigatoxin producing *E. coli* (STEC) strains isolated from animal food supply and from humans diagnosed with gastroenteritis. Seven housekeeping genes described for STEC were amplified by PCR, sequenced, and analyzed by MLST. Isolates were characterized by allele composition and Sequence Type (ST); and the diversity association within and among different sources of the strains was examined. *E. coli* O157:H7 occurred at a higher rate in slaughterhouse and retail samples than at farm or in humans in this study. The most common shigatoxin in this study was shigatoxin ST171, a shigatoxin common to enterotoxigenic *E. coli* and atypical enteropathogenic.

From this MLST analysis, it was possible to conclude that enterohemorrhagic *E. coli* pathogenic strains are present along the supply chain at different levels and with varying relatedness [28].

Restriction Fragment Length Polymorphism (RFLP) analysis

Restriction fragment length as the name indicates is a measure of the difference in the length of homologous DNA sequence fragments which can be detected after digestion of homologous DNA with specific restriction endonucleases. Resulting restriction fragments can be separated according to their length by gel electrophoresis. An RFLP probe, a labeled DNA sequence hybridizes with one or more fragments of the digested DNA sample after they are separated by gel electrophoresis, thus

revealing a unique band pattern characteristic to a specific genotype at a specific locus [29].

The RFLP analysis method has been used for the study of plasmids such as those resistance plasmids harboring drug resistance genes. Sun, et al. analyzed the multi-drug resistance plasmid PXZ isolated from *E. coli* strains resistant to aminoglycosides. The objective of this study was to characterize the multi-drug resistant plasmid which mediates resistance to beta-lactams, fosfomycin and aminoglycosides, from *E. coli* isolates and to detect the prevalence of similar resistance plasmids in 224 *E. coli* isolates by designing a multiplex PCR assay. The plasmid was digested in to different fragments with restriction enzymes EcoRI and Sall, and sequenced. The plasmid was digested into five fragments by EcoRI and six fragments by Sall. The finding was a novel multi drug resistant plasmid, PXZ, harboring four resistance genes (*rmtB*, *fosA3*, *blaTEM-1* and *blaCTX-M*). Primers were designed according to resultant digests and the prevalence of similar resistance plasmids was then assessed by multiplex PCR and by RFLP analysis among different *E. coli* strains. Among the 224 *E. coli* isolates, 17 contained a plasmid with the MDR encoding region of pXZ, which showed high-level resistance to amino glycosides.

Random Amplified Polymorphic DNA (RAPD) markers

RAPD analysis is a type of polymerase chain reaction in which the DNA segments amplified are random. There is no need to know the DNA sequence of the target gene. This technique requires small amounts of DNA without the need for sequencing or characterization of the genome of the species to be studied as the primers will bind somewhere in the target DNA sequence. A primer which binds to many different loci is used to amplify random sequences from a complex DNA template in the genome on the assumption that a given DNA sequence which is complementary to the primer will occur in the genome. The amplified fragment generated by PCR depends on the length and size of both the primer and the target genome. These amplified products may be separated on agarose gels and visualized by ethidium bromide staining [30].

This technique was used in the analysis of the epidemiological relationship of Extended Spectrum Beta (β) Lactamase (ESBL) producing pathogenic *E. coli* strains isolated from children by Amaya and his colleagues in 2011. These were strains which exhibited resistance to the third generation cephalosporins (cefotaxime, ceftriaxone and ceftazidime). Thirteen *E. coli* isolates producing ESBLs were selected for RAPD analysis. The analysis revealed that these *E. coli* isolates could be separated into five clones based on the analysis. Resistance to ceftazidime and/or ceftriaxone and a pattern of multi-drug resistance was related to CTX-M-5- or CTX-M-15-producing *E. coli* isolates. One of the clones contained one diarrheagenic and one non-diarrheagenic isolates producing the beta lactamase enzyme CTXM-5 from children with diarrhea, which could indicate the transfer of resistance genes from non-pathogenic to pathogenic bacteria. In a similar study by the same author, RAPD analysis of isolated *E. coli* strains was done to evaluate the epidemiological relationships between *E. coli* isolates producing ESBL from the

hospital sewage water samples and from well-water samples. This analysis was carried out in 22 *E. coli* isolates from the well water samples and 17 from the hospital sewage water samples. RAPD analysis revealed that 17 resistant *E. coli* isolates from the hospital sewage water sample could be separated into 11 clones and that the isolates from the well-water samples could be separated into five clones. *E. coli* producing ESBL and harboring *blaCTX-M* genes were detected in one of the hospital sewage samples and in 26% of the resistant isolates from the well-water samples.

The *blaCTX-M-9* group was more prevalent in *E. coli* isolates from the hospital sewage samples and the *blaCTX-M-1* group was more prevalent in the well-water samples. The analysis did not show any clonal similarity between the ESBL-producing *E. coli* isolates from the wells and the hospital sewage water samples. Therefore, RAPD markers can be used for the assessment of clonal similarity and difference among different *E. coli* strains.

Detection of known virulence genes from pathogenic *Escherichia coli* by PCR

Escherichia coli are groups of diverse species comprising both commensal and pathogenic strains. Pathogenic strains can cause infections of the mammalian intestine resulting in diarrhea or extra intestinal infections. The primary habitat of these bacteria is the gut of vertebrates, where they live as a commensals and their secondary habitat is water and sediment [31]. Based on virulence and colonization factors encoded in the chromosomal or plasmids, pathogenic *E. coli* bacteria have been grouped in to different strains including: Enteroinvasive *E. coli* strains (EIEC) which possess invasion genes like *virF* and *ipaH*; Enteropathogenic *E. coli* strains (EPEC) that possess a pathogenicity island in their genome called Locus of Enterocyte Effacement (LEE) which encodes for the target gene *intimin*, designated as *eae* (there are classes of Enteropathogenic *E. coli* strain group called typical enteropathogenic *E. coli* which contain EPEC Adherence Factor (EAF) plasmid which codes for the target gene bundle forming pillus (*bfp*) in addition to LEE); Enterotoxigenic *E. coli* (ETEC) that possess heat labile and heat stable enterotoxin genes (*LT*, *ST*); shiga toxin producing *E. coli* also called verotoxin producing *E. coli* that possess shiga toxin genes (*stx1* and *stx2*); Enteroaggregative *E. coli* (EAEC) which produce pic and other toxins. Different molecular primers have been designed according to the target genes of pathogenic strains. The diversity of pathogenic strains in different animal and human hosts can be analyzed by amplification of the different strain associated target genes by PCR.

Host and environment based genetic structure of *Escherichia coli*

Escherichia coli are among the best characterized microorganisms. They are ecologically and genetically diverse organisms and live in many different habitats both in the environment and in human and animal hosts. *E. coli* are mainly harmless bacteria however they frequently exhibit significant pathogenicity towards humans and animals [32].

Because of the extensive genetic diversity the *E. coli* group displays, it is difficult to put a generalized picture of the environment and host related genetic structure of these bacteria. The bacteria show remarkable genetic diversity from host to host and from environment to environment and even within the same host and environment. Here, it has been tried to compile genetic structure data of different *E. coli* populations obtained from different published researches conducted by different researchers in different places. According to Tenaillon, et al., genetic diversity of *E. coli* exhibits host, taxonomic and environmental components. Genetic diversity is variable among human host populations with the highest genetic diversity observed in populations living in tropical areas. Domesticated animals have lower strain diversity than wild animals. Strains of phylogeny group A and B2 are predominant in humans while a predominance of B1 strains followed by A and B2 strains is observed in animals. However, this variation in the prevalence of phylogeny groups does not guarantee the existence of host-specific strains. Only a few strains seem to be host specific. Host characteristics such as diet, gut morphology and body mass, and domestication seem to be important predictors of the distribution of the phylogeny groups. Domesticated animals have a decreased proportion of B2 strains than wild animals and increased proportion of A strains. Similarly, large changes in the prevalence of *E. coli* groups are found among different human populations. Commensal strains isolated from Europe (France and Croatia) in the 1980's and from Africa (Mali and Benin), Asia (Pakistan), and South America (French Guiana, Colombia and Bolivia) belong mainly to the A and B1 phylogenetic groups. Conversely, strains isolated from Europe (France and Sweden) in the 2000's and from North America (USA), Japan and Australia belong mainly to the B2 group. A study made in Australia among pathogenic *E. coli* strains recovered from human hosts with extra intestinal infections evaluated whether isolates within the same infection site were genetically related.

The isolates exhibited extensive genetic diversity in that human hosts were infected either by several distinct *E. coli* clones or by members of a single clone. Isolates also showed varied resistance patterns to anti-bacterials. This study concluded that strains isolated from a specific infection site were genetically diverse with host and environment related factors contributing to the fitness and adaptability of *E. coli* clones in human hosts during infections. The courses of adaptability of strains in the human host were thought to be similar to *in vitro* experimental evolutionary processes [33].

Furthermore, 395 non diarrheagenic isolates (270 from children with diarrhea and 125 from children without diarrhea) and 332 diarrheagenic *E. coli* isolates (241 from children with diarrhea and 91 from children without diarrhea) were analyzed in a study made in Central America, Nicaragua to assess the antimicrobial resistance patterns of *E. coli* in children. Among individual *E. coli* categories, enteroaggregative *E. coli* isolates from children with and without diarrhea exhibited significantly higher levels of resistance to ampicillin and trimethoprim-sulfamethoxazole compared to the other *E. coli* categories due to resistance genes that produce the beta lactamase enzymes CTX-M-5- or CTX-M-15.

Similarly, another study was made in France by Clermont and his colleagues on a panel of *E. coli* strains to assess the genetic background of human and animal pathogenic isolates. Phylogenetic analysis from multiple locus sequence typing data of the pathogenic isolates showed that pathogenic animal and human strains are genetically related as exhibited by their sharing of many common virulence genes. Few *E. coli* strains isolated from animals exhibited specificity to certain adhesins in this study. This study concluded that human and animal pathogenic strains share common genetic backgrounds. This study also analyzed the phylogenetic relationship between strains isolated from intestinal and extra intestinal infection sites and it showed that most intestinal infections are caused by B2 phylogeny groups while most extra intestinal infections are caused by phylogeny groups A, B1 and E.

A related study was done in South America, The Guiana on the genetic structure of human and animal *E. coli* populations. This study demonstrated that most animal isolates majorly belonged to B2 phylogenetic group while B2 isolates were rare among human isolates.

Animal isolates exhibited unique and infrequent sequence types with many virulence genes but the human isolates exhibited few virulence genes. High housekeeping gene nucleotide diversity per site was also observed among the *E. coli* strains in this study.

Petty, et al., conducted a study on a globally disseminated multi-drug resistant *E. coli* clone designated *Escherichia coli* sequence type 131 (ST131) that is associated with human urinary tract and bloodstream infections. A genome sequencing of a large collection of *E. coli* ST131 strains isolated from six geographical regions across the world (Australia, Canada, India, Spain, United Kingdom, New Zealand) revealed that *E. coli* ST131 strains are distinct from other extra intestinal pathogenic and that they may have probably evolved from a single progenitor strain. Three closely related *E. coli* ST131 sub lineages were identified which did not exhibit geographical or temporal genetic variation. Lineages displayed single-nucleotide variations with each other due to recombination in their regions adjacent to mobile genetic elements. Four different variants of the Extended Spectrum β -lactamase (CTX-M ESBL) resistance gene were identified in these strains which may have led to the evolution and worldwide dispersal of the strain. The study concluded that the rapid emergence and spread of these multi-drug resistant strains may be related to high virulence gene content, the possession of species specific fimbriae (FimH30 allele), and the production of Extended Spectrum B-Lactamase (ESBL) in addition to mobile genetic elements and associated recombination events.

A study made in Canada, British Columbia to assess the temporal genetic diversity of *E. coli* strains isolated from forests with watershed showed that the genetic diversity of *E. coli* isolates varied among seasons over the years exhibiting relatively higher genetic diversity during fall. Majority of REP-PCR fingerprints in this study showed a tendency to cluster according to year, season, and month indicating that the diversity and population structure of *E. coli* strains fluctuate according to time scale due to presence of diverse host sources with varying behaviors over time in the watershed [34].

Another study was made in Denmark on the genetic diversity of *E. coli* populations isolated from nursery pigs located at different pigpens (pig houses) in the same farm. The aim of the study was to investigate how the genetic diversity of *E. coli* in fecal samples differs between pigs located in the same pigpen or in different pigpens in the same farm.

REP-PCR was used to assess genetic diversity. A high genetic diversity of *E. coli* strains was observed at the pigpen level and between pigpens among isolated strains the REP-PCR DNA fingerprint analysis showed extensive genetic diversity among *E. coli* strains isolated from different pigs regardless of whether they shared the same pigpen or not [35].

Moreover, a study was made in Qatar on the genetic relatedness of Enterohemorrhagic *E. coli* strains isolated from animal meat taken from different food supply chains and in human patients with gastroenteritis. Seven housekeeping genes of enterohaemorrhagic *E. coli* varieties were amplified by PCR, sequenced, and analyzed by multi locus sequence typing. Isolates were characterized by allele composition, sequence type and assessed for epidemiological relationship within and among different sources. The study showed that the strains isolated from food supply chains were genetically diverse but the strains isolated from human patients were more genetically diverse probably due to the poly-clonal diversity in the human macrobiotic environment.

In A cross-sectional study conducted in Bahirdar, Ethiopia in 422 children less than five years of age with diarrhea from December 2011 to February 2012, the overall isolation rate of *E. coli* was 48.3% and isolates belonging to enteropathogenic *E. coli* were the most commonly isolated strains [36-47].

CONCLUSION

Genetic diversity is a measure of the variation of organisms at the genetic level; it may involve whole chromosomes, some segment of DNA, or specific nucleotide bases. Genetic diversity is affected by factors like mutation, natural selection, genetic drift, migration, and population size. Prokaryotes can better be classified in to different species based on their genomic features because they display extensive genetic variation compared to Eukaryotes. *E. coli* and other related bacteria possess extra chromosomal genetic materials like plasmids which can be transferred from one bacterium to another or from the environment into bacterial genomes through horizontal gene transfer. Horizontal gene transfer promotes bacterial genetic diversity by helping bacteria to evolve into new adapted strains. *E. coli* is one of the best studied bacterial species and has been used as a model organism in different genetic studies.

E. coli bacteria live in a wide range of environments: As commensals in the gut of animals and humans, freely in the environment, and as pathogens causing many human and animal diseases. *E. coli* genomes contain core genome containing essential genes and accessory genomic content which is involved in adaptive characters. It is the accessory genome which is highly variable from strain to strain and contains mobile genetic elements such as plasmids, insertion sequences, bacteriophages, etc., which are the agents of horizontal gene

transfer. It is possible to estimate the age horizontally transferred genes through gene amelioration studies. Horizontal gene transfer contributes more to the genetic diversity of *E. coli* than mutation. Horizontal gene transfer can take place in *E. coli* in three different mechanisms: Transduction, transformation, and conjugation. Domestic *E. coli* isolates often undergo high mutation rates in their genomes due to a reduced or deficient methyl mismatch repair activity. Different molecular techniques have been developed to measure the genetic diversity of *E. coli* strains isolated from different habitats including the analysis of simple sequence repeats, repetitive palindrome sequences, multi locus analysis, restriction fragment length polymorphism, randomly amplified polymorphic DNA analysis etc. *E. coli* bacteria are generally genetically diverse showing tremendous diversity from host to host and from environment to environment; isolates obtained from the same host and environment or even the same specific site exhibit significant genetic diversity. Host and environment genetic diversity of these bacteria doesn't seem to show clear species or strain specific pattern.

REFERENCES

- Peterson L. The definition of diversity: Two views. A more specific definition. Managing multiculturalism and diversity in the Library. 1st Edition. Routledge. 2014;17-26.
- Reed DH, Frankham R. Correlation between fitness and genetic diversity. *Conserv Biol.* 2003;17(1):230-237.
- Hughes AR, Inouye BD, Johnson MT, Underwood N, Vellend M. Ecological consequences of genetic diversity. *Ecol Lett.* 2008;11(6):609-623.
- Amos W, Harwood J. Factors affecting levels of genetic diversity in natural populations. *Philos Trans R Soc Lond B Biol Sci.* 1998;353(1366):177-186.
- Amann R, Snaidr J, Wagner M, Ludwig W, Schleifer KH. *In situ* visualization of high genetic diversity in a natural microbial community. *J Bacteriol.* 1996;178(12):3496-3500.
- Cohan FM, Perry EB. A systematics for discovering the fundamental units of bacterial diversity. *Curr Biol.* 2007;17(10):R373-R386.
- Sharma R, Munns K, Alexander T, Entz T, Mirzaagha P, Yanke LJ, et al. Diversity and distribution of commensal fecal *Escherichia coli* bacteria in beef cattle administered selected sub-therapeutic antimicrobials in a feedlot setting. *Appl Environ Microbiol.* 2008;74(20):6178-6186.
- Lawrence JG. Gene transfer, speciation, and the evolution of bacterial genomes. *Curr Opin Microbiol.* 1999;2(5):519-523.
- Lan R, Reeves PR. Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends Microbiol.* 2000;8(9):396-401.
- Hartl DL, Dykhuizen DE. The population genetics of *Escherichia coli*. *Annu Rev Genet.* 1984;18(1):31-68.
- Dobrindt U, Chowdary MG, Krumbholz G, Hacker J. Genome dynamics and its impact on evolution of *Escherichia coli*. *Med Microbiol Immunol.* 2010;199(3):145-154.
- Tenaillon O, Skurnik D, Picard B, Denamur E. The population genetics of commensal *Escherichia coli*. *Nat Rev Microbiol.* 2010;8(3):207-217.
- Lescat M, Clermont O, Woerther PL, Glodt J, Dion S, Skurnik D, et al. Commensal *Escherichia coli* strains in guiana reveal a high genetic diversity with host-dependant population structure. *Environ Microbiol Rep.* 2013;5(1):49-57.

14. Mira A, Klasson L, Andersson SG. Microbial genome evolution: Sources of variability. *Curr Opin Microbiol.* 2002;5(5):506-512.
15. Lawrence JG, Ochman H. Molecular archaeology of the *Escherichia coli* genome. *Proc Natl Acad Sci USA.* 1998;95(16):9413-9417.
16. Hacker J, Carniel E. Ecological fitness, genomic islands and bacterial pathogenicity. *EMBO Rep.* 2001;2(5):376-381.
17. Boyd EF, Almagro-Moreno S, Parent MA. Genomic islands are dynamic, ancient integrative elements in bacterial evolution. *Trends Microbiol.* 2009;17(2):47-53.
18. Elena SF, Lenski RE. Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nat Rev Genet.* 2003;4(6):457-469.
19. Gogarten JP, Townsend JP. Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol.* 2005;3(9):679-687.
20. Blattner FR, Plunkett III G, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. *Science.* 1997;277(5331):1453-1462.
21. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, et al. The genome sequence of *E. coli* W (ATCC 9637): Comparative genome analysis and an improved genome-scale reconstruction of *E. coli*. *BMC genomics.* 2011;12(1):1-20.
22. Turnpenny PD, Ellard S, Cleaver R. Emery's elements of medical genetics e-book. Health Elsevier sci. 2020.
23. Mrazek J, Guo X, Shah A. Simple sequence repeats in prokaryotic genomes. *Proc Natl Acad Sci USA.* 2007;104(20):8472-8477.
24. Gur-Arie R, Cohen CJ, Eitan Y, Shelef L, Hallerman EM, Kashi Y. Simple sequence repeats in *Escherichia coli*: Abundance, distribution, composition, and polymorphism. *Genome Res.* 2000;10(1):62-71.
25. Gordon DM, Bauer S, Johnson JR. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. *Microbiol.* 2002;148(5):1513-1522.
26. Delannoy S, Beutin L, Fach P. Use of clustered regularly interspaced short palindromic repeat sequence polymorphisms for specific detection of enterohemorrhagic *Escherichia coli* strains of serotypes O26: H11, O45: H2, O103: H2, O111: H8, O121: H19, O145: H28, and O157: H7 by real-time PCR. *J Clin Microbiol.* 2012;50(12): 4035-4040.
27. Gordon DM, Clermont O, Tolley H, Denamur E. Assigning *Escherichia coli* strains to phylogenetic groups: Multi-locus sequence typing versus the PCR triplex method. *Environ Microbiol.* 2008;10(10): 2484-2496.
28. Palanisamy S, Chang Y, Scaria J, Filho RAP, Peters KE, Doiphode SH, et al. Genetic relatedness among shiga toxin-producing *Escherichia coli* isolated along the animal food supply chain and in gastroenteritis cases in Qatar using multilocus sequence typing. *Foodborne Pathog Dis.* 2017;14(6):318-325.
29. Avise JC. Molecular markers, natural history and evolution. Springer Science and Business Media. 2012.
30. Kumar NS, Gurusubramanian G. Random Amplified Polymorphic DNA (RAPD) markers and its applications. *Sci Vis.* 2011;11(3):116-124.
31. Clermont O, Olier M, Hoede C, Diancourt L, Brisse S, Keroudean M, et al. Animal and human pathogenic *Escherichia coli* strains share common genetic backgrounds. *Infect Genet Evol.* 2011;11(3):654-662.
32. Chandra M, Cheng P, Rondeau G, Porwollik S, McClelland M. A single step multiplex PCR for identification of six diarrheagenic *E. coli* pathotypes and *Salmonella*. *Int J Med Microbiol.* 2013;303(4): 210-216.
33. Levert M, Zamfir O, Clermont O, Bouvet O, Lespinats S, Hipeaux MC, et al. Molecular and evolutionary bases of within-patient genotypic and phenotypic diversity in *Escherichia coli* extraintestinal infections. *PLoS Pathogens.* 2010;6(9):e1001125.
34. Chandran A, Mazumder A. Pathogenic potential, genetic diversity, and population structure of *Escherichia coli* strains isolated from a forest-dominated watershed (Comox lake) in British Columbia, Canada. *Appl Environ Microbiol.* 2015;81(5):1788-1798.
35. Herrero-Fresno A, Ahmed S, Hansen MH, Denwood M, Zachariassen C, Olsen JE. Genotype variation and genetic relationship among *Escherichia coli* from nursery pigs located in different pens in the same farm. *BMC Microbiol.* 2017;17(1):1-10.
36. Adugna A, Kibret M, Abera B, Nibret E, Adal M. Antibigram of *E. coli* serotypes isolated from children aged under five with acute diarrhea in Bahir Dar town. *Afr Health Sci.* 2015;15(2):656-664.
37. Boyd EF, Hartl DL. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J Bacteriol.* 1998;180(5):1159-1165.
38. Rugeles LC, Bai J, Martinez AJ, Vanegas MC, Gomez-Duarte OG. Molecular characterization of diarrheagenic *Escherichia coli* strains from stools samples and food products in Colombia. *Int J Food Microbiol.* 2010;138(3):282-286.
39. Amaya E, Reyes D, Vilchez S, Paniagua M, Mollby R, Nord CE, et al. Antibiotic resistance patterns of intestinal *Escherichia coli* isolates from Nicaraguan children. *J Med Microbiol.* 2011;60(2): 216-222.
40. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep.* 2013;5(1):58-65.
41. Amaya E, Reyes D, Paniagua M, Calderon S, Rashid MU, Colque P, Kuhn I, et al. Antibiotic resistance patterns of *Escherichia coli* isolates from different aquatic environmental sources in Leon, Nicaragua. *Clin Microbiol Infect.* 2012;18(9):E347-E354.
42. Clermont O, Christenson JK, Denamur E, Gordon DM. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo groups. *Environ Microbiol Rep.* 2013;5(1):58-65.
43. Dobrindt U, Agerer F, Michaelis K, Janka A, Buchrieser C, Samuelson M, et al. Analysis of genome plasticity in pathogenic and commensal *Escherichia coli* isolates by use of DNA arrays. *J Bacteriol.* 2003;185(6):1831-1840.
44. McLellan SL. Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. *Appl Environ Microbiol.* 2004;70(8): 4658-4665.
45. Petty NK, Ben Zakour NL, Stanton-Cook M, Skippington E, Totsika M, Forde BM, et al. Global dissemination of a multidrug resistant *Escherichia coli* clone. *Proc Natl Acad Sci USA.* 2014;111(15): 5694-5699.
46. Sun H, Li S, Xie Z, Yang F, Sun Y, Zhu Y, et al. A novel multidrug resistance plasmid isolated from an *Escherichia coli* strain resistant to aminoglycosides. *J Antimicrob Chemother.* 2012;67(7):1635-1638.
47. van Elsas JD, Semenov AV, Costa R, Trevors JT. Survival of *Escherichia coli* in the environment: Fundamental and public health aspects. *ISME J.* 2011;5(2):173-183.