This is to certify that the
dissertation entitled

POPULATION GENETIC ANALYSES OF *SALMONELLA ENTERICA* SEROVAR ENTERITIDIS AND *ESCHERICHIA COLI* FROM BIOTIC AND ABIOTIC SOURCES

presented by

SETH TAYLOR WALK

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Department of Zoology and Program in Ecology, Evolutionary Biology, and Behavior

Mark Kuhn, Chair, Zoology
Major Professor's Signature

5/8/2007
Date

MSU is an affirmative-action, equal-opportunity employer

POPULATION GENETIC ANALYSES OF *SALMONELLA ENTERICA* SEROVAR ENTERITIDIS AND *ESCHERICHIA COLI* FROM BIOTIC AND ABIOTIC SOURCES

By

Seth Taylor Walk

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Zoology and Program in Ecology, Evolutionary Biology, and Behavior

2007

ABSTRACT

POPULATION GENETIC ANALYSES OF *SALMONELLA ENTERICA* SEROVAR ENTERITIDIS AND *ESCHERICHIA COLI* FROM BIOTIC AND ABIOTIC SOURCES

By

Seth Taylor Walk

Humans and animals host a myriad of microbes including enteric bacteria of the family *Enterobacteriaceae*. While these organisms are related, they represent a range of evolutionary strategies and interactions with the host gastrointestinal tract. Clinically important enteric species, including *Salmonella enterica* and *Escherichia coli,* have been thoroughly characterized by phenotypic, genetic, and epidemiologic methods. Accordingly, these organisms serve as models for studying the genetic and phenotypic diversity of this bacterial family. Although a few strains have been well characterized, less information is available about the natural history of these organisms and, specifically, how populations experience and adapt to selective evolutionary pressures. The goal of this research is to present a novel perspective on this topic through population genetic analyses. These methods, when applied to longitudinal samples of strains, quantify the abundance and distribution of alleles, genotypes, and phylogenetic groups in evolving populations. The results of each individual research area add to the current understanding of microbial ecology and evolution by incorporating information about populations from natural (biotic and abiotic) sources.

DEDICATION

This work is dedicated to my parents and family for their unwavering commitment to my intellectual pursuits, their support during times of financial inadequacies, and, above all, their love.

Also, this work is dedicated to Thomas Whittam, a leader, mentor, role model, and friend. His dedication to my career and the careers of all students under his advisement speaks loudly of his outstanding character and scientific professionalism.

ACKNOWLEDGEMENTS

I wish to thank my mentor, Thomas S. Whittam, for his gracious advisement and faith in my abilities. I thank my committee members, Elizabeth Alm, Vincent Young, Richard Lenski, and Edward Walker for their guidance and suggestions. Also, I thank Lisa Craft and Pat Resler for their secretarial support.

The members of the Microbial Evolution Laboratory have been instrumental to the completion of this research by providing valuable insight, experience, and encouragement. I wish to thank Teresa Bergholz, David Lacher, Adam Nelson, Lindsey Ouellette, Galeb Abu-Ali, Hans Steinsland, Priya Kailasan Vanaja, Amber Cody Springman, James Riordan, Shannon Manning, Anthony Heidt, Julie Cunningham, Jaclyn Middleton, Janice Mladonicky, Lukas Wick, Weihong Qi, Lisa Calhoun, Shanda Birkland, and Christopher Moeller.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Images in this dissertation are presented in color.

CHAPTER 1

OVERVIEW

Enteric bacteria are members of the microbial milieu found in the gastrointestinal (GI) tract of humans and animals. These organisms belong to the bacterial family *Enterobacteriaceae* and represent >1000 different species (47). Most organisms in this family have yet to be cultured and are known to exist based solely on comparative genomic analyses (95, 166). The majority of what is known about the biology of this family has been inferred from analyses using representative strains that are easy to culture and amenable to laboratory growth and manipulation. In addition, studies of enterics tend to focus on organisms that either cause disease or have similar ecological requirements as pathogens so that discoveries have a direct impact on public health (107).

Species within the *Enterobacteriaceae* are ecologically diverse. For example, the enteric bacterium *Escherichia coli* is widely distributed among vertebrate hosts (56), is one of the most commonly cultured enteric bacteria from mammals (57), and often does not cause disease (174). Likewise, *Salmonella enterica* has a wide host range and can be carried asymptomatically by hosts (147). However, certain strains within each of these species are remarkably cell-type adapted, wherein they replicate at the expense of the host (46). Such a wide diversity of host-bacterial interaction and the rapidity with which they can be cultured are two useful characteristics that make species, such as *E. coli* and *S. enterica*, excellent model organisms for studying the natural history of enteric bacteria (147) and they are the subjects of this dissertation.

***S. enterica* serovar Enteritidis.** The species designation, *S. enterica*, contains >2,500 serotypes, or combinations of two antigens; the lipopolysaccharide (O antigen) and flagellin proteins (H antigen) (29). *S. enterica* serovar Enteritidis

(SE) is one of these serotypes and, like most *Salmonella*, has the ability to invade epithelial cells that line the gastrointestinal tract of warm-blooded vertebrates. Human infection often results in massive fluid secretion and diarrhea, fever, bacteremia, and an estimated 16,000 hospitalization per year in the United States (86, 147). During the 1990's, there was a human pandemic of Enteritidis and it soon became the most commonly isolated *Salmonella* serovar in the United States (3). Since then, the U.S. incidence of all types of salmonellosis has decreased, while the incidence of Enteritidis has increased (3). It remains a serious health problem in the U.S. and other countries around the world.

Most information about the natural history of SE comes from epidemiologic investigations of outbreaks and sporadic cases of disease. For example, human infections in the U.S. are most often associated with shell eggs (142) and the consumption of contaminated chicken products (86). Examination of poultry rearing facilities has shown that wild animals, such as birds and mice, significantly influence the persistence of SE infection in chickens (39). There is also evidence that this serotype can be transmitted vertically in chickens, as invading cells enter the reproductive tract of the hen and into the developing egg (72). These observations have led to a hypothesis that SE is maintained in nature by zoonotic infection, is spread to chickens via the ingestion of contaminated fecal material, and is transmitted to humans by chickens via undercooked eggs. The work presented in the second chapter of this dissertation addresses this hypothesis through the description of a population genetic based analysis of an assembled collection of Enteritidis strains from these different hosts.

Antibiotic use provides experimental opportunity. Quantifying the influence of evolutionary forces, like selection, in natural populations of bacteria often requires the interpretation of many complex and interacting variables. One approach to these types of analyses has been to evolve strains in the laboratory under conditions where variables can be controlled (43). Another approach relies on comparisons between populations that experience selection and those that do not. The latter approach has the advantage of incorporating stochastic variation into the experimental design. However, certain effects are difficult to control and may result in a variable effect size that is too subtle to observe under natural conditions.

Selection caused by the use of antibiotics is an anthropomorphic evolutionary force, the impacts of which are not well understood. Specifically, therapeutic and sub-therapeutic doses of drugs are common in clinical and agricultural settings, but there are few data with which to assess these influences. The third chapter of this dissertation takes advantage of a collection of *E. coli* from conventional farms that commonly use antibiotics and organic farms, where antibiotic use is rare (136). This type of study represents a natural experiment, which can be used to assess the influence of selection on bacterial populations in nature.

Secondary habitat *E. coli* are understudied. Enteric bacteria, such as *E. coli*, circulate between two habitats (137). The first habitat, called the primary habitat, is the GI tract of humans and animals. The secondary habitat is broadly defined as the environment (water, sediment, and soil). Much of the information concerning the evolutionary ecology of *E. coli* has been collected from host-associated (primary habitat) strains. Perhaps this was a historical oversight, but more

likely, it was the result of a bias (either funding or human interest or both) toward strains of clinical and laboratory importance. Regardless, population dynamic characterizations of *E. coli* in the secondary habitat are understudied, but important. For example, cell counts of this species have long been used by the Environmental Protection Agency as an indicator of fecal pollution (167). In addition, enteric bacteria are often spread via the fecal-oral route and it has been estimated that half of all *E. coli* cells are currently in transit between host GI tracts, in the secondary habitat (137). This leads to a number of questions: What, if anything, are they doing there? Do certain microenvironments outside the host promote growth and reproduction? How does selection operate on these populations? The results presented in the fourth chapter of this dissertation build on current *E. coli* evolutionary biology by adding a population genetic perspective with relevance to dynamics in a particular secondary habitat – the freshwater beach environment.

A novel evolutionary perspective. As discussed briefly in the previous section, an interesting hypothesis is that certain aspects concerning the natural history of *E. coli* have been biased by popular strains and strain collections. If this is true, strains from under-represented sources will provide new insights into how this species has evolved. The fifth and final chapter of this dissertation attempts to show how *E. coli* strains from a novel environment suggest a new evolutionary history for this species.

CHAPTER 2

THE CLONAL STRUCTURE OF *SALMONELLA ENTERICA* SEROVAR ENTERITIDIS ISOLATED FROM MICE, CHICKENS, AND HUMANS

INTRODUCTION

Salmonellae are common gastrointestinal pathogens of humans and animals. In the last three decades, *Salmonella enterica* serotype Enteritidis (SE) has emerged to become a major food-borne pathogen (142, 168). Investigations of both outbreak and sporadic cases of infection identified eggs and chickens as vehicles for the majority of cases (66, 158), which indicated that egg-laying chicken flocks are a major reservoir for SE. This zoonotic serotype led to a pandemic of the disease in human populations in Europe and the United States (132). In 1987, 9 of 21 (43%), countries reported SE as their most common serotype, and 8 of 9 (89%), were European countries. Several epidemiological investigations have incriminated eggs in human outbreaks of SE in the United States and Europe (66, 158). In the United States, SE ranks second to *Salmonella enterica* serotype Typhimurium as the cause of thousands of cases of salmonellosis each year (6, 105).

Reservoirs and transmission. As with many foodborne pathogens, SE is of zoonotic origin and has various routes of transmission into humans. It is most readily spread between animals via the fecal-oral route (i.e. through ingestion of contaminated fecal material). Strains often circulates in wild animals and, in particular, have often been recovered from rodent populations associated with chicken farms (39, 67). As is also seen with infections caused by its relative, *Salmonella enterica* serotype Typhimurium, SE is often highly pathogenic to laboratory mice, causing systematic infection and mortality (165). However, the feral domestic mouse (*Mus musculus*) is a frequent carrier that survives infection. In a

study of captured mice from hen houses on farms, approximately 20% of the population had spleens positive for SE (61). It has been suggested that physiologically diverse subpopulations of SE are propagated in mouse populations, and some have an enhanced ability to infect hens (59).

Transmission within chickens. There are no easily observable disease symptoms in poultry when SE colonizes the avian intestinal tract, which may in part be due to the ability of the pathogen to actively mitigate signs of illness in the hen (119). SE contaminates eggs either on the shell surface through contact with fecal material or by contamination of the internal contents as a result of infection of the reproductive tissue of laying hens, resulting in transovarian transmission (72, 108). The pathogen has been cultured from the albumen (white) of the egg, the vitelline (yolk) membrane, and the yolk (52, 72). Contamination of the vitelline membrane and yolk follows adherence to and invasion of the granulosa cells of the preovulatory follicles (161, 162), whereas the oviduct and fallopian tube are infected by bacteremic spread (40) and may particularly involve colonization of tubular cells (40, 119). While contamination of the preovulatory follicles may be facilitated by expression of fimbriae, the absence of fimbriae appears to facilitate bacteremia and translocation of the bacterium to the oviduct (41). Thus, access to the internal contents of the egg may involve 2 fundamentally different anatomical routes, namely the albumen or the yolk and its associated membranes. Both routes may contribute to the outbreak potential of the organism. However, only oviduct transmission differentiates the biology of SE from that of the classical pathology associated with *S. enterica* serovar Pullorum, which is a serious poultry pathogen that generates substantial ovarian

pathology but does not cause illness in humans (150). The production of eggs with *Salmonella*-infected contents not only serves as the primary vehicle for human disease, but it can also perpetuate the bacteria through vertical inheritance via infected chicks hatched from contaminated eggs (72, 161). This route of transmission may contribute to additional outbreaks of SE that follow consumption of meat from broiler chickens that were initially infected in hatcheries.

Transmission to Humans. Transmission of SE to humans occurs primarily through contaminated food, most often via uncooked eggs, egg-containing foods, and poultry. The Centers for Disease Control and Prevention (CDC) has reported that 80% of 371 outbreaks of SE in the United States between 1985 and 1999 were egg-associated (120). SE-contaminated eggs are estimated to have accounted for approximately 180 000 illnesses in the United States in 2000 (142). Non-egg-containing foods have also been associated with SE infections. One notable example is an outbreak of salmonellosis in North America in 2001 that was linked to contaminated almonds (30). Subsequent investigation found that SE phage type (PT) 30 was detected in raw almonds collected from multiple sources, including environmental swabs of orchards and their associated processing equipment, which suggests that, under some environmental conditions, there is diffuse contamination (76).

Factors in the emergence of SE. A variety of factors have been implicated in the dramatic rise in the incidence of SE in the past 4 decades, including changes in farm practices, eradication of competing *Salmonella* strains, and virulence evolution of the pathogen. Farm practices clearly contribute to the incidence of *Salmonella*.

For example, molting strategies can influence bird susceptibility to infection (51, 70). Rodents that inhabit farm environments are a principal source of infection for flocks (38). It has been suggested that the use of rodenticides containing certain strains of SE for vermin control may have contributed to the initial spread of this organism (116). However, the SE isolates from rodenticides used belong to PT6a, while the majority of human SE epidemics were attributable to PT4 in England and PT8 and 13a in the United States (163).

Eradication of *S. enterica* serovars Pullorum and Gallinarum from chicken flocks during the 1960s has also been suggested as an important emergence factor. In essence, some argue that an open niche resulted from the eradication, thereby fostering the spread of SE (11). This hypothesis is based on the observation that all 3 avian serovars have the O9 surface antigen as the immunodominant epitope. In addition, multilocus enzyme electrophoresis (MLEE) and sequencing of the flagellin gene (*fliC*) indicate that the nonmotile Pullorum/Gallinarum complex is monophyletic and shares a recent ancestor with SE (96). There is some evidence that *S. enterica* serovar Typhi displaces SE in human populations, because India is one of the few reporting countries that have a high incidence of Typhi and a low incidence of SE (60). In addition, research has shown that both SE and Typhi are capable of producing a type of capsule that is associated with production of high-molecular mass lipopolysaccharide (LPS; (60, 125)).

A third factor in SE emergence, the evolution of virulence evolution, garners support from several types of observations. There is evidence for substantial variation in virulence among different SE strains. For example, Solano and

colleagues discovered that SE strains with high virulence in a chicken model produced filaments and aggregates in vitro and were, thus, phenotypically distinguishable from low-virulence strains (155). Dominant PTs of SE isolated in the United Kingdom (PT4) and in the United States (PT13a) show substantial within-PT variation that results in some strains having enhanced durability characteristics as well as different abilities to contaminate eggs and to grow to high cell density (34, 60, 73). Strains of SE also vary in the production of a type of capsule associated with production of high-molecular mass LPS. For example, isolates from eggs and avian reproductive tissue are more likely to produce high-molecular mass LPS than isolates from avian intestines or rodent samples, suggesting that this special LPS is critical for infection of the reproductive tract (59).

To begin to address how SE has evolved and adapted to its multiple hosts and modes of transmission, we have assessed the clonal relatedness of isolates obtained from various sources over the past 25 years. Clonal relatedness among isolates was determined by multilocus analysis of conserved genes with housekeeping functions. We found that most isolates of SE collected over the past few decades have the same multilocus genotype, suggesting that the strains involved in the human epidemic mark a single widespread clone. In addition, we found evidence that closely related, but genetically distinct clonal variants may be shifting to new hosts in wild animal populations.

EXPERIMENTAL PROCEDURES

SE isolates. A collection of SE isolates spanning the observed emergence of this pathogen (1978-2004), including isolates from non-human sources (164) to isolates that were recently collected from human and non-human sources are listed in Table 1. All isolates were confirmed biochemically and serologically at the National Veterinary Service Laboratories (Ames, IA) and were stored at –80°C in tryptic soy broth containing 15% glycerol. Three Enteritidis strains from the *Salmonella* reference collection B (SARB; (22)) were obtained from E. Fidelma Boyd (National University of Ireland - Cork) and served as genetic control strains.

Multilocus enzyme electrophoresis (MLEE). Enzyme extraction, gel electrophoresis, and specific enzyme staining was carried out as described in Selander et *al.* (144). Briefly, lysates (whole cell enzymes) were extracted from overnight nutrient broth cultures and frozen at -80°C. Lysate samples were individually loaded and electrophoresed under non-denaturing conditions in a buffered starch gel matrix at the appropriate concentration for the particular enzymes being stained. Gel slices were incubated in enzyme specific staining solutions and fixed for analysis.

The mobilities of 18 housekeeping enzymes were recoded for 196 isolates spanning the years and sources of the collection (Table 1). Mobilities were scored relative to previously characterized *Salmonella* strains from the SARB collection. The enzymes used in this study were: ADH (alcohol dehydrogenase), THD (threonine dehydrogenase), SKD (shikimate dehydrogenase), G6P (glucose-6-phosphate), MPI (mannose phosphate isomerase), GLUD (glutamate dehydrogenase), MDH (malate dehydrogenase), NSP (nucleoside phosphorylase), PEP (peptidase),

Table 1. Sources of *Salmonella enterica* serovar Enteritidis by years of isolation.

Source	1978-1987	1990-1992	1994-1999	2000-2004	Total
Environment	0	109	195	0	304
Mice	0	10	1	0	11
Chickens	28	136	325	0	489
Eggs	0	42	63	0	105
Humans	0	121	202	24	347
Others	2	2	4	9	17
Total	30	420	790	33	1273

GOT (glutamic oxalacetic transaminase), CAK (carbamylate kinase), AK (adenosine kinase), MPD (mannitol-1-phosphate dehydrogenase), PGD (6-phosphogluconate dehydrogenase), PGI (phosphoglucose isomerase), IDH (isocitrate dehydrogenase), ACO (aconitase), LDH (lactate dehydrogenase). Electrophoretic mobility variants (electromorphs) were assigned scores in side-by side comparisons relative the SARB standards. Electromorphs were equated with alleles at the corresponding enzyme locus based on the assumption that each mobility difference reflects at least one amino acid replacement in the protein. Electrophoretic types (ETs) were assigned to isolates with indistinguishable allele profiles across all enzyme loci studied. Allele frequencies and ETs were used to estimate population genetic parameters as described in Selander et *al.* (144).

To make quantitative estimates about the genetic relationships among isolates, we equate electromorphs with alleles at an enzyme locus and electrophoretic types (ETs) with multilocus genotypes. We assume that isolates with the same ET owe their genotypic similarity to recent descent from a common ancestral cell; that is, they are members of a naturally occurring bacterial clone.

Dendrograms of genetic relatedness and eBURST analysis. By comparing the multilocus profiles of sampled populations, we assessed the genetic relationships among ETs using a distance-based neighbor joining algorithm for dendrogram construction and an eBURST analysis of allelic profiles (using the eBURST algorithm) described in Feil et *al.* (49).

RESULTS

Genetic diversity and clonal analysis. Characterization of 196 SE strains by MLEE of 18 housekeeping enzymes resolved an average of 2.7 alleles per locus and distinguished 26 distinct electrophoretic types (ETs). Fourteen of the 18 loci were polymorphic with the number of alleles per locus ranging from 1 to 6 for NSP (Table 4). The average single-locus genetic diversity was 0.051 (± 0.012) across 196 isolates and 0.246 (± 0.051) among the 26 ETs. The difference in these two measures of diversity reflects the fact most SE strains belong to a single ET. Of the 26 ETs identified, 19 (73%) were recovered only once. There were 7 ETs that were represented by more than one isolate. The most common genotype, ET-3, accounted for 151 (77%) of the 196 Enteritidis strains (Table 2). The most common ET also included the SARB 16 reference strain. These MLEE results are consistent with the previous findings of Boyd et *al.* (22), and indicate that most Enteritidis strains, marked by ET-3, belong to a single widespread clone.

The dendrogram of the overall genetic relatedness among the ETs shows that ET-3 falls near the node of a cluster of 13 ETs at the top of the diagram (Figure 1, cluster A). This cluster of ETs represents the most common genotypes of SE circulating in various sources. The lower half of Figure 1 includes 13 divergent ETs only one of which was isolated more than once. We suspect that these represent recombinant genotypes that are rare in populations and that have evolved by horizontal gene transfer.

To further assess genetic relatedness between the ETs of SE, we used the allele profiles to perform an eBURST. The analysis focuses on discerning the

Table 2. Multilocus allele profiles defining 26 ETs of 196 isolates of Salmonella enterica serovar Enteritidis. Profiles are defined by electromorphs (mobility variants) resolved at 14 polymorphic loci. Four of the 18 loci were monomorphic (PGI, AK, GOT, and ADH). ETs are presented in the order of the dendrogram in Figure 2. The most common ET (ET-3) is in italics and alleles that differ from the common ET are marked in bold.

Tree		No. of	Electromorphs (allele) of enzyme locus[b]													
order	ET	isolates	IDH	ACO	LDH	PGD	M1P	NSP	MDH	TDH	SKD	G6P	PEP	GLU	CAK	MPI
1	1	1	3	3	4	3	3	3	3	3	0.5	3	**0.25**	**5**	**2**	3
2	16	1	3	3	4	3	3	3	3	3	0.5	3	**0.5**	3	**2**	3
3	26	2	3	3	0	3	3	3	3	3	0.5	3	3	3	**2**	3
4	21	1	3	3	4	3	3	3	3	3	0.5	3	0	3	**4**	3
5	20	1	3	3	**5**	3	3	3	3	3	0.5	3	3	3	3	3
6	*3*	*151*	*3*	*3*	*4*	*3*	*3*	*3*	*3*	*3*	*0.5*	*3*	*3*	*3*	*3*	*3*
7	22	15	**2**	3	4	3	3	3	3	3	0.5	3	3	3	3	3
8	10	2	3	**4**	4	3	3	3	3	3	0.5	3	3	3	3	3
9	24	1	3	3	4	3	3	3	3	**2**	0.5	3	3	3	3	3
10	11	1	3	**4**	4	3	3	3	3	**4**	0.5	3	3	3	3	3
11	12	1	3	3	4	3	3	3	3	**4**	0.5	3	3	3	3	3
12	19	1	3	3	4	3	3	**4**	3	3	**1**	3	3	3	3	3
13	17	3	3	3	**3**	**4**	**2**	3	3	3	0.5	3	3	3	3	3
14	25	2	3	3	4	3	3	**2.5**	3	3	0.5	3	3	3	3	3
15	7	1	3	**2**	4	**2**	3	**4**	3	3	0.5	3	3	3	3	3
16	4	1	3	3	4	3	3	**4**	**5**	3	0	3	**4**	3	3	3
17	8	1	3	3	4	**2**	3	**4**	**5**	3	0	3	3	3	3	3

Table 2 (continued).

Tree order	ET	No. of isolates	Electromorphs (allele) of enzyme locus[b]													
			IDH	ACO	LDH	PGD	M1P	NSP	MDH	TDH	SKD	G6P	PEP	GLU	CAK	MPI
18	6	1	3	3	4	**2**	3	**4**	3	3	**1**	3	3	3	3	3
19	23	1	3	3	4	**2**	3	**6**	3	3	**0.75**	3	3	3	3	3
20	5	1	3	3	4	**2**	3	**6**	3	3	**1**	3	3	3	3	3
21	9	1	3	3	4	**2**	3	**6**	3	3	**1**	3	**4**	3	3	3
22	18	2	3	3	4	3	3	**5**	3	3	**1**	3	3	3	3	3
23	2	1	3	**4**	4	3	**0.5**	**9**	3	3	**1**	3	3	3	3	**2**
24	14	1	3	**2**	4	3	3	**4**	3	3	0	3	**4**	3	3	**2**
25	13	1	**5**	**2**	4	3	**2**	**4**	**5**	3	**1**	**4**	**4**	3	**4**	**5**
26	15	1	3	**2**	4	3	**4**	**4**	**5**	3	**1**	3	3	3	3	**2**

[b] 0 designates a null allele, i.e. no detectable enzyme activity.

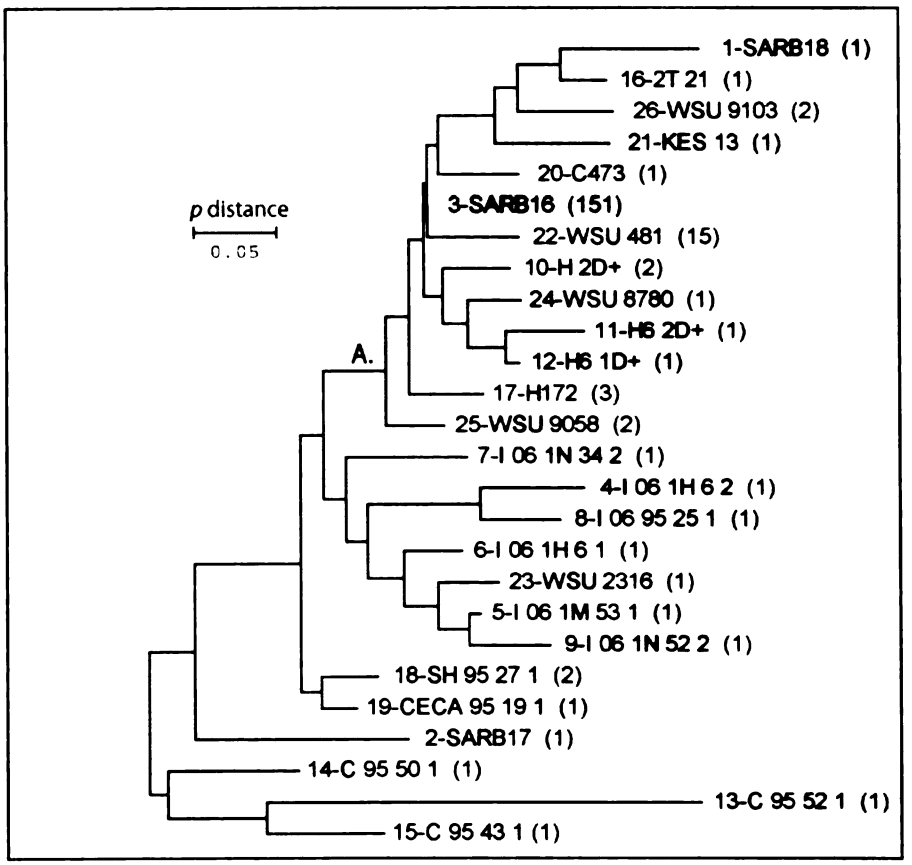


Figure 1. Dendrogram of the genetic relationships of 26ETs of SE inferred from electrophoretic allelic variation detected at 18 enzyme loci. The dendrogram was constructed using a neighbor-joining algorithm and proportional distance (p-distance) between ETs. The common ET-3 is highlighted in red. The number of isolates of each ET is given in parentheses.

relationships of multilocus genotypes in which both mutation and recombination can be a source of genotypic divergence (48). Application of the eBURST analysis identified ET-3 as the founding genotype based on the number of single locus variants connected this electrophoretic type (Figure 2). The BURST diagram shows that the 12 ETs of cluster A are closely related to ET-3 as single or double locus variants. The eBURST analysis indicates that ET-3 represents the founding clonal genotype of SE that has diversified with the epidemic spread of the clone.

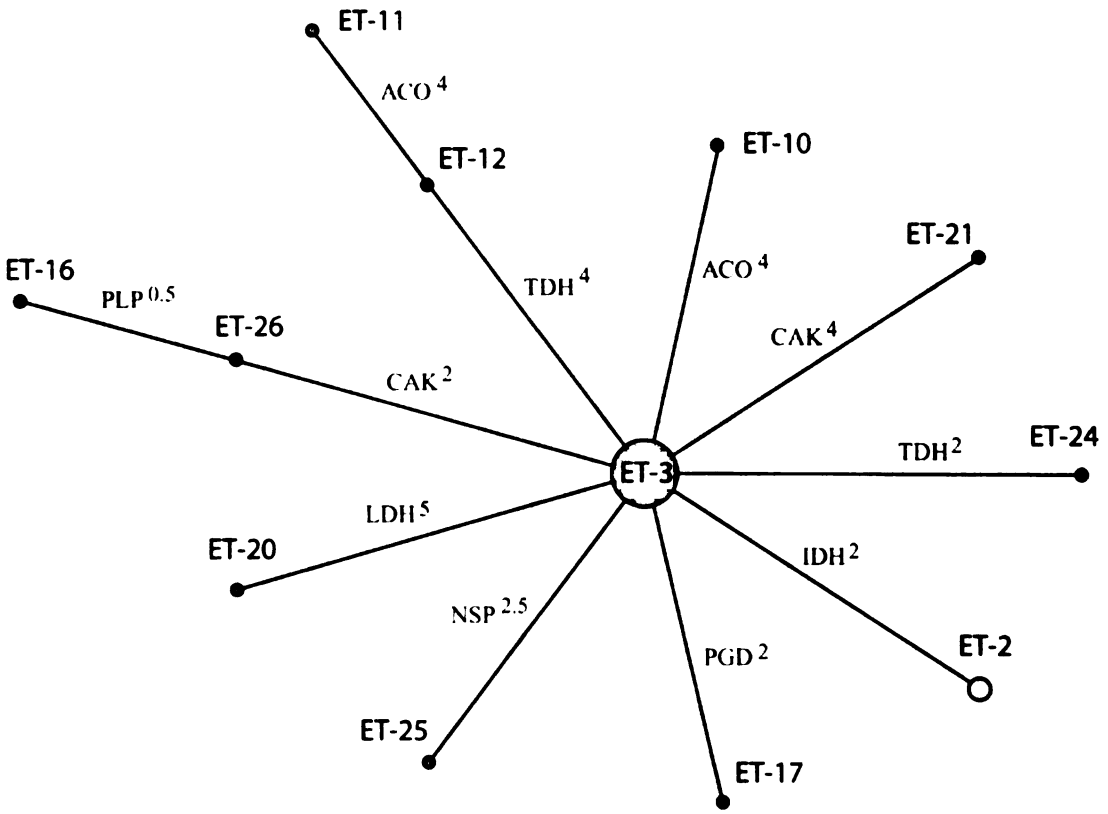


Figure 2. eBURST diagram of the main clone complex of SE. ET-3 is the founder genotype (center) that is connected to related ETs that are single locus variants (SLVs). The inferred allele substitutions, which create the SLVs, are given above the lines. For example, ET-2, the second most commonly recovered clone, is a SLV that differs by IDH^2 from ET-3.

DISCUSSION

This genetic characterization demonstrates that most SE strains have an identical multilocus enzyme genotype (ET-3) regardless of PT, geographic origin, or time of isolation that spanned over 2 decades (1978 to 2004). These data are consistent with the hypothesis that these strains are members of a single widespread clone (Table 1). The ET-3 strains include 7 different PTs, with a predominance of PT28 (42%), PT8 (28%), and PT13a (18%). PT4, the most observed type among recent European outbreaks, was also found among ET-3 strains, but at a relatively low frequency (~4%). While this genotype has been observed previously (22), it is noteworthy that 23 of the 26 ETs found in this study are not represented among the SE strains of the SARB collection.

An eBURST analysis identified ET-3 as the founding genotype for a group of clonally related, but genetically distinct strains. These new ETs are defined either by novel alleles or new combinations of preexisting alleles. This analysis suggests that ET-3 marks the ancestral clone that founded these genotypic clusters, but subsequent nucleotide change (either single base pair mutations or recombination of horizontally transferred alleles) has generated new variants. The identification of a number of more distantly related SE types suggests that these strains have radiated even further through similar mutational events or through the spread of the SE-distinct LPS antigen region into divergent backgrounds. In any case, these distant variants are distinctly different from the ET-3 cluster, are rare in this collection, and do not appear to be commonly associated with human disease.

SE strains of 3 major PTs (8, 13a, and 28) were distributed among 24 genotypes that were carried by a range of animal hosts, including chickens, cattle, horses, and wild animals. These data suggest that SE strains have evolved the ability to cross a number of host barriers, a research area that warrants further study. SE strains from clinical cases of human gastroenteritis belong to the same bacterial clone (ET-3) as strains from eggs, mice, and chickens. Our interpretation is that the ET-3 clone has evolved via successful adaptation to the avian host, which serve as a reservoir for pathogenic strains. While this clone may not cause overt disease in mammals, the ability of some strains to cause severe gastroenteritis in humans could be due to the emergence of highly pathogenic variants. If this is the case, strains from chickens and human disease may be differentiated by phenotypic virulence characteristic in cell culture.

ACKNOWLEDGMENTS

This project is funded in part with federal funds from NIAID, NIH, and the Department of Health and Human Services, under Contract No. N01-AI-30058. This study was part of a published paper (Saeed, A.M., Walk, S.T., Arshad, M., Whittam, T.S. 2006. Clonal structure and variation in virulence of *Salmonella* Enteritidis isolated from mice, chickens, and humans. Journal of AOAC International 89(2): 504-511.).

CHAPTER 3

THE INFLUENCE OF ANTIBIOTIC SELECTION ON THE POPULATION GENETIC COMPOSITION OF *ESCHERICHIA COLI* FROM CONVENTIONAL AND ORGANIC DAIRY FARMS

INTRODUCTION

Escherichia coli is an indicator species for a variety of anthropogenic effects on microbial populations, such as the emergence and spread of antibiotic resistance in agriculture (5, 13-15, 18, 19, 23, 42, 84, 127, 138, 139, 149, 152). Although most strains are commensal bacteria and nonpathogenic to humans and animals, there are well-recognized pathogenic strains that can cause a variety of human and zoonotic diseases. In addition, some commensal populations are known to carry a high level of antibiotic resistance (10, 117, 118). Such resistant populations pose a public and veterinary health risk because of the potential transfer of genetic resistance determinants to pathogens. In addition, certain virulence factors may be mobilized on genetic elements and transferred to normally commensal, but antibiotic resistant strains via horizontal exchange (128, 179, 182).

During antibiotic selection in the laboratory, resistance-conferring mutations often have a measurable deleterious effect (i.e. a resistance cost) due to a reduction in function of genes where resistance mutations arise. In order to maintain a competitive advantage over other members of the population, it is hypothesized that deleterious effects on fitness are compensated by changes elsewhere in the genome (93, 99, 100, 129, 134, 141). The occurrence of such compensatory fitness mutations makes it difficult to determine whether the abundance and distribution of resistant strains is a result of direct selection on the original mutation that caused resistance, selection on compensatory changes, or other ecological factors that limit population diversity (environmental selection,

bottlenecks, genetic drift, etc.). In addition, there is sound evidence that antibiotic use increases the abundance of resistant phenotypes (136), but it is not clear if the cessation of antibiotic use will decrease abundance after compensatory changes have occurred (8, 94).

Antibiotic use in dairy cattle provides a useful opportunity to assess the role of natural selection in bacterial populations for several reasons: the source of the antibiotic selective pressure is known and the dosage is often recorded; the common genetic determinants for certain resistant phenotypes have been characterized and high throughput assays are available for their identification; hypotheses generated under laboratory conditions can be tested *in vivo* by comparing bacteria from farms that regularly use antibiotics (conventional) and bacteria from farms that rarely use antibiotics (organic) (136); and a number of studies have previously characterized resistance dynamics on both farm types and have identified variables that significantly influence the abundance of resistant phenotypes (5, 13, 15, 42, 127, 136, 138, 139).

The purpose of the present study was to assess the influence of antibiotic selection on the genetic composition of *E. coli* populations from conventional and organic dairy farms. First, we used a PCR based assay (33) to quantify the abundance and distribution of 4 phylogenetic groups in populations cultured during a longitudinal sampling of cattle from matched conventional and organic dairy farms in Wisconsin (136). We then assessed the pattern of statistical dependence among farm types (conventional vs. organic), cattle ages (cows vs. calves), bacterial phenotypes (resistant vs. sensitive), and bacterial genetic

composition (ECOR groups A, B1, B2, and D) using hierarchical log-linear modeling.

EXPERIMENTAL PROCEDURES

***E. coli* strain collection.** A total of 678 *E. coli* strains (367 random sensitive and 311 resistant strains) were assembled from a collection of 1,121 strains of a longitudinal sampling of 10 randomly selected cows and calves from a matched set of 30 conventional and 30 organic dairy farms in Wisconsin (78). Briefly, a cluster of organic farms was selected and the geographically closest conventional farm was selected for purposes of comparison so as to minimize the effects of distance (cline effects). All organic farms were certified by a USDA-accredited certification agency as not having treated adult cows for at least 3 years (mean, 8 years; range, 3-15 years) prior to this study. More information about these farms is available at Sato et *al.* (136).

In the original study (136), fecal samples were taken from 5 lactating cows and 5 calves (< 6 months of age) at each of two visits (once in March and once in September) and conducted with aseptic technique. Laboratory isolation was begun within 72 hours and a single *E. coli* colony was isolated from each fecal sample so as to exclude any single farm or within-animal bias. All isolates were confirmed by standard biochemical assays. Minimum inhibitory concentrations (MICs) of 17 antibiotics were determined for each strain as recommended by the NCCLS (109) using a commercially available semiautomatic microbroth dilution test (Sensititre, Trek Diagnostic Systems Inc, Cleveland, OH) and appropriate quality control strains. These antibiotics included ampicillin, amoxicillin-clavulanic acid, cephalothin, cefoxitin, ceftiofur, ceftriaxone, streptomycin, kanamycin, gentamicin, apramycin, amikacin, tetracycline, sulfamethoxazole,

trimethoprim-sulfamethoxazole, nalidixic acid, and ciprofloxacin). Resistant phenotypes were confirmed by the presence of overnight growth on LB broth (Lennox; Becton, Dickinson, and Company, Sparks, MD) agar containing antibiotic at the NCCLS cut-off concentrations. More details about the strain collection and isolation procedures can be found at Sato et *al.* (136).

ECOR phylogrouping by multiplex PCR. Strains were grouped into 1 of 4 phylogenetic lineages (A, B1, B2, or D) based on methods adapted from Clermont et al. (33). Genomic DNA was isolated from 2 mL of overnight culture in LB broth (Lennox; Becton, Dickinson, and Company, Sparks, MD) using the Puregene DNA isolation kit (Gentra Systems Inc., Minneapolis, MN.). DNA preparations were quantified with a NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE.), diluted to a final concentration of 100 ng/μL, and stored at 4°C. Genomic DNA preparations were tested using primers targeting a 650 bp region of the conserved housekeeping gene, *mdh* (see www.shigatox.net/stec/mlst-new/index.html for primer sequences and reaction conditions), and AmpliTaq Gold ® DNA polymerase (Applied Biosystems). This protocol has produced a positive amplicon in strains representing the genotypic diversity of the species as well as *E. coli*'s most recent common ancestor, *Escherichia albertii*. Genomic DNA was re-isolated if the assay was negative. Strains that were negative for duplicate, independent genomic isolations were considered to be members of species other than *E. coli* and excluded from further analysis. Representative ECOR (*E. coli* reference collection) strains were used as template controls for a duplex PCR targeting the

genes *chuA* and *yjaA*. We found that the following duplex conditions yielded higher PCR specificity with AmpliTaq Gold ® than the published triplex: denaturation at 94°C for 10 minutes; 35 cycles of 92°C for 1 minute, 59°C for 1 minute, 72°C for 30 seconds; and a final elongation at 72°C for 5 minutes. A separate PCR was run with primers targeting the TspE4.C2 anonymous DNA locus using published conditions (33).

Resistance loci and class I integron PCR. Ampicillin resistant (amp^R) and tetracycline resistant (tet^R) strains were screened for the presence of 6 previously identified resistance loci. A multiplex PCR was used to detect the presence of bla_{TEM}, bla_{SHV}, and bla_{OXA-1} in amp^R strains as reported by Colom et *al.* (35). Fragments of the *tetA*, *tetB*, and *tetC* genes were targeted in tet^R strains using the primers and conditions published by Boerlin et *al.* (19). Three primer sets were used to determine the presence of class 1 integrons in the resistant strains. Primers sets targeting the class 1 integrase locus, *intI1*, the conserved region cassette region A and B, and the quarternary ammonium compound resistance gene, *qacEΔ1* are given along with the reaction condition in Lindstedt et *al.* (97). Integron presence was defined as amplification of all 3 loci.

Statistical analyses. Strains were categorized for analysis as follows: F = *farm type* (conventional vs. organic), A = *cattle age* (calf vs. cow), D = *resistant phenotype* (resistant vs. sensitive) or *drug susceptibility level* (high, medium, vs. low), and E = *ECOR group* (A, B1, B2, and D), and recorded in the cells of contingency tables. Hierarchical log-linear modeling with nested effects was used to assess dependent associations using the CATMOD procedure and SAS

statistical software (SAS Institute, Cary, NC). Non-significant, higher-order interactions were removed until the most parsimonious model was found based on the likelihood ratio chi-square statistic for testing goodness of fit (G^2). Non-significant G^2 values indicated that the fit model was not significantly different from the saturated model. Odds ratios were calculated based on parameter estimates from the most parsimonious models. Higher-order (three-way) interactions for multi-drug resistant phenotypes were visualized in mosaic plots, which were obtained online at http://euclid.psych.yorku.ca/cgi/mosaics. The original plots were redrawn and shaded with respect to the significant ($\alpha = 0.05$) associations from the SAS analysis.

RESULTS

Overall abundance of *E. coli* phylogroups. Strains belonging to all 4 ECOR phylogroups were identified (Figure 3) among the 678 *E. coli* strains from calves and cows on dairy farms. The relative phylogroup composition of these bacterial populations was used to compare different patterns of antibiotic use. The populations analyzed here represent the natural variation between farm type (conventional vs. organic), cattle age (calf vs. cow), and resistance phenotype (resistant vs. sensitive). It is clear that phylogroup abundance was not evenly distributed among the different types of dairy farms (Figure 3). The most abundant phylogenetic groups were B1 (58.3%) and A (27.4%), whereas groups D (11.5%) and B2 (2.8%) were rare. B2 strains were not sampled at each variable level (no resistant B2 genotypes were found on organic farms), so these data were combined with group D strain data (B2D) for statistical analyses.

Genetic composition of sensitive and resistant *E. coli* populations. Our initial goal was to test for dependent associations among three nominal variables (F = farm type, A = age of cattle, and E = ECOR phylogroup, E) by analyzing the number of strains in these categories. The tests for associations in the sensitive population (susceptible to 17 antimicrobials) by log-linear modeling of the 376 sensitive strains revealed no significant interactions with ECOR phylogrouping (Tables 3a and 3b). In other words, the distribution of phylogroups in sensitive *E. coli* sampled from calves and cows on conventional and organic dairy farms was similar and not significantly different. A significant negative association was found between conventional farms and the number of sensitive calf strains (i.e.

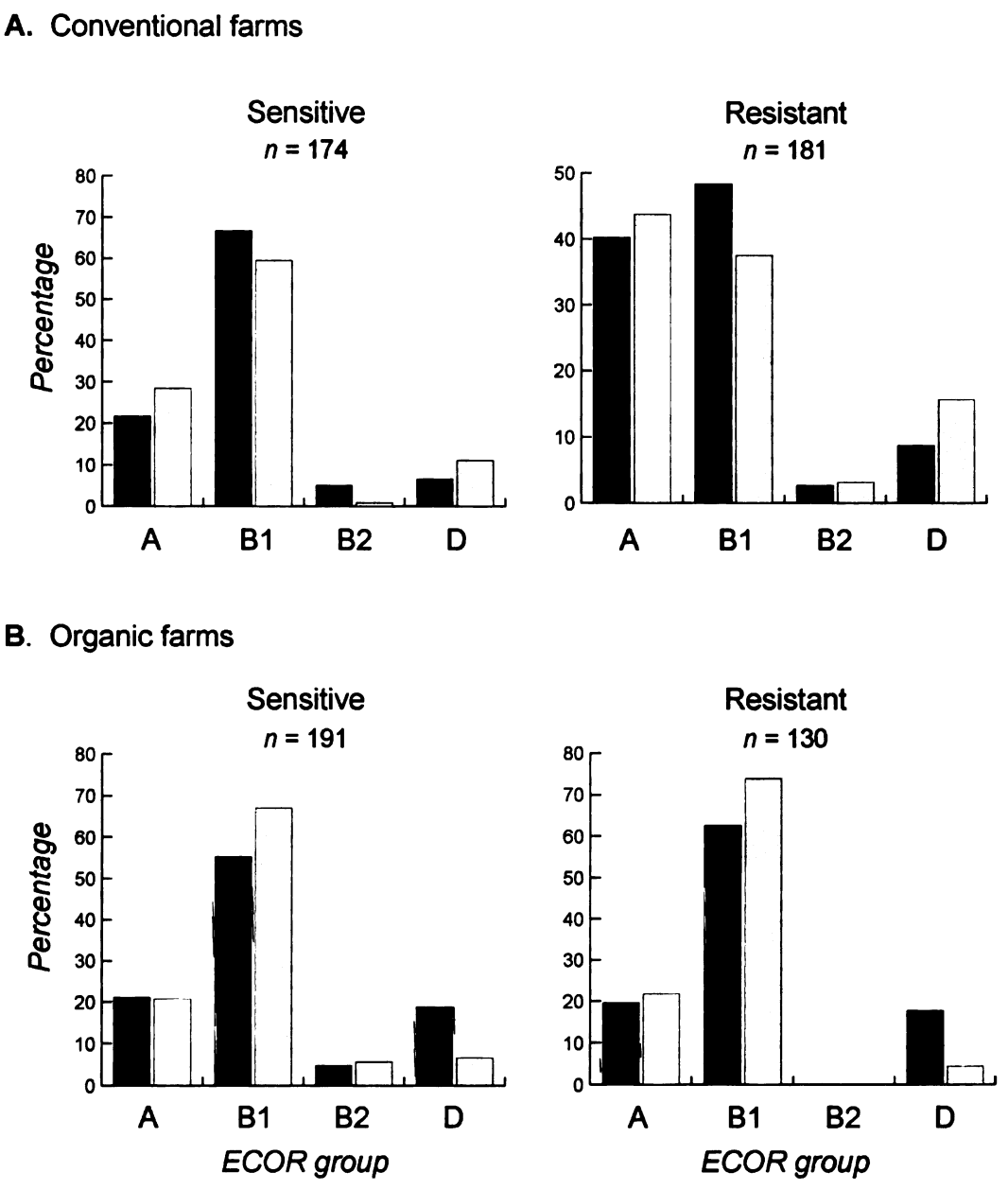


Figure 3. Histogram plots of ECOR phylogroups for sensitive and resistant E. coli populations from conventional (A) and organic (B) farms (black bars = calf strains, gray bars = cow strains).

Table 3. The best-fit models explaining the frequency of sensitive and resistant *E. coli* in conventional and organic diary farm. The analysis is based on testing hierarchical log-linear models with nested effects in parentheses. Nominal categorical variables are designated as A = animal age (calf or cow), E = ECOR group (A, B1, B2, D), and F = farm type (conventional or organic). μ designates the overall main effect. **A.** Likelihood ratio chi-square statistic was used to test for goodness of fit (compared to the saturated model). **B.** Significant interactions for the sensitive and resistant populations.

A. Population model	Final model	G^2 (df, Pr > χ^2)
Sensitive	μ + A + E + F(A=calf)	8.85 (7, 0.26)
Resistant	μ + A + E + F(E=ECOR A)	5.32 (7, 0.62)

B. Population model	Significant interactions	χ^2 (df, Pr > χ^2)
Sensitive	F(A = calf)	4.27 (1, 0.04)
Resistant	F(E = ECOR A)	21.1 (1, < 0.0001)

the F(A = calf) interaction in Table 1b). This result was expected because the abundance of resistant strains was higher in calves on conventional farms than calves on organic farms. Despite this discrepancy in abundance, however, these data indicate that sensitive strains of the four phylogroups were circulating at similar frequencies on both farm types in young and adult animals.

A similar analysis was applied to the 311 resistant strains and revealed a significant association between farm type and ECOR phylogrouping (Tables 1a and 1b). Based on parameter estimates, the odds of recovering resistant *E. coli* of the phylogroup A were significantly greater on conventional farms than on organic farms ($df = 1$, $\chi^2 = 21.1$, $Pr > \chi^2 < 0.0001$). This overabundance of phylogroup A strains was not seen in the sensitive population or the resistant population from organic farms. In addition, there were no significant farm-phylogroup (i.e. FE) interactions when the sensitive populations from both farms and the resistant population from organic farms were analyzed together (model not shown). These data suggest that resistance determinants on conventional farms were linked to the genetic backgrounds of phylogroup A and that these strains increased in frequency as a result of antibiotic use. Interestingly, animal age was not associated with the distribution of phylogroups in the resistant population, suggesting that similar phylogroups circulate at similar frequencies in young and adult dairy cattle.

The influence of multi-drug resistance on genetic composition. Strains were categorized according to their level of drug susceptibility (D) as defined by the number of resistant antimicrobial phenotypes (low = 1 – 2, medium = 3 – 4,

and high = 5 or more). A log-linear model fit to the data according to farm type (F), multi-drug resistance level (D), and ECOR phylogrouping (E) revealed significant heterogeneity in the association between these variables (Tables 4a and 4b) including the presence of a significant three-way (FDE) interaction. In other words, the best-fit model to these data included all three variables. The model was simplified slightly by reparameterizing and nesting the variables (FDE models I, II, and III in Table 4), which allowed non-significant levels to be removed.

To illustrate the complexity of the interactions affecting bacterial multi-drug resistance, we summarized the components of the *E. coli* populations using mosaic plots of three different parameterizations of the FDE log-linear model (Figure 4). Odds ratios were estimated for significant interactions with respect to a fixed or nested factor. For example, when the effect of multi-drug resistance was nested, or fixed, a significant two-way interaction between farm type (F) and phylogroup (E) was found and can be seen by comparing the size of the shaded box to the size of the non-shaded boxes for a given level of drug susceptibility (D). When the low multi-drug resistance level is considered, it is clear that the shaded box representing group B1 strains on organic farms is larger than the one for conventional farms. The opposite is true for phylogroup A or B2D strains (larger boxes for the conventional farm category); hence a significant (positive) interaction is represented by the shaded organic B1 box ($df = 2$, $\chi^2 = 6.3$, $Pr > \chi^2 = 0.044$). The odds of isolating phylogroup A strains with medium multi-drug resistance were significantly higher on conventional farms ($df = 2$, $\chi^2 = 8.3$, $Pr >$

Table 4. Log-linear modeling of significant associations between farm type (F), multi-drug resistance (D), and ECOR phylogrouping (E). A. Three separate parameterizations of the FDE model are given to show statistical dependence as a function of nested effects. B. Significant interactions (3-way) with E.

A. FDE model	Final model	G^2 (df, Pr > χ^2)
I	μ + F + D + E + FE + DE + FD(E = ECOR B1)	5.84 (4, 0.21)
II	μ + F + D + E + FE + DE + DE(F = Organic)	1.79 (4, 0.78)
III	μ + F + D + E + FE + DE + FE(D = High) + FE(D = Medium) + FE(D = Low)	Saturated

B. FDE model	Significant interactions with E	χ^2(df, Pr > χ^2)
I	FD(E = ECOR B1)	9.55 (2, 0.008)
II	DE(F = Organic)	18.64 (4, < 0.001)
III	FE(D = High)	6.25 (2, 0.044)
III	FE(D = Medium)	8.29 (2, 0.016)
III	FE(D = Low)	10.34 (2, 0.006)

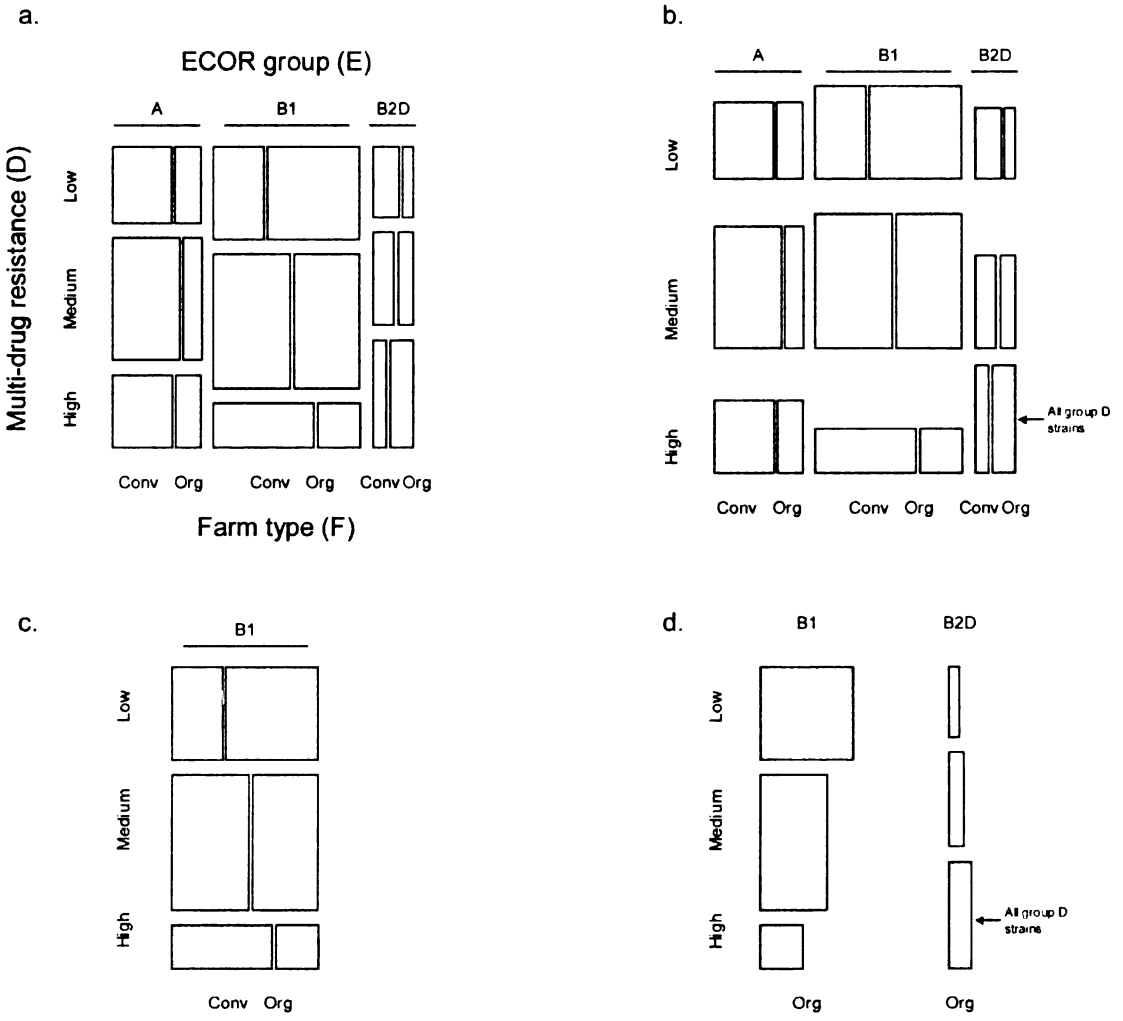


Figure 4. Mosaic plot of the dependent associations between farm type (F), multi-drug resistance (D), and ECOR phylogrouping (E). Shaded boxes mark significant odds ratio estimates (positive odds only). a. Overall mosaic plot for FDE. b. FE interactions at fixed levels of D. c. FD interactions at fixed levels of E. d. DE interactions at fixed levels of F.

$\chi^2 = 0.016$), and the odds of isolating highly resistant, phylogroup B2D strains were significantly higher on organic farms ($df = 2$, $\chi^2 = 10.3$, $Pr > \chi^2 = 0.006$). As mentioned above, there were no resistant phylogroup B2 strains isolated from organic farms, so the shaded B2D box on organic farms represents group D strains only. Phylogroup specific interactions were also found when model effects were fixed for E ($df = 2$, $\chi^2 = 9.6$, $Pr > \chi^2 = 0.008$) and F ($df = 4$, $\chi^2 = 18.64$, $Pr > \chi^2 < 0.001$). These data suggest that conventional farms are associated with medium and highly resistant phylogroup A and B1 strains, whereas, in contrast, organic farms with virtually no antibiotic use are associated with low and highly resistant phylogroup B1 and D strains.

The association between cattle age (A), multi-drug resistance (D), and ECOR phylogrouping (E) was also found to be heterogeneous. The high abundance of resistant calf strains and limited overall resistance on organic farms resulted in sampling zeros for 3 of the 9 possible categories in cows (no medium resistant ECOR B2D strains, highly resistant phylogroup A strains, or highly resistant phylogroup B1 strains were sampled). After correcting for sampling zeros in the resistant cows categories, the three-way interaction term (ADE) was not significant in the model. These data suggest that cattle age influences the abundance of multi-drug resistant strains, but does not influence the genetic composition of this population.

Tetracycline and ampicillin resistance determinants. Of the 311 resistant strains analyzed, 129 (41.5%) were ampicillin resistant (amp^R), 281 (90.4%) were tetracycline resistant (tet^R), and 112 (36.0%) were resistant to both

drugs. Based on PCR screening for 3 common *E. coli* β-lactamase genes (bla_{TEM}, bla_{SHV}, and bla_{OXA-1}), 119 (92.2%) amp^R strains carried the bla_{TEM} locus, while the remaining 10 (7.8%) strains did not produce an amplicon for any of the targeted loci. Similarly for 3 genes known to confer *E. coli* tetracycline resistance (*tetA*, *tetB*, and *tetC*), 268 (95.4%) tet^R strains carried at least one of these loci, while 13 (4.5%) did not. The *tetB* and *tetA* genes were the most abundant (64.8% and 28.1%), while the *tetC* gene was rarely sampled (4.6%).

Genetic composition and resistance determinants. We created 4 datasets according to the 4 genetic determinants present in the resistant population. Data from the sensitive population analyzed above were added to each to create a two level factor (G) for log-linear modeling. The factor "G" categorized strains that carried a resistance gene (bla_{TEM}, *tetA*, *tetB*, *tetC*) or did not (sensitive). Due to the low occurrence in the sample, data for $tetC^+$ strains (n = 7) were pooled with data for strains that were negative for all 3 loci (n = 13) and called *tetC*/other. Data for strains that were negative for the 3 β-lactamase loci (n = 10) were omitted. Log-linear models were fit to each of the 4 datasets to test for associations between F, G, and E (Tables 5a and 5b).

Interactions between the resistance loci and ECOR phylogroups were not dependent on farm type (no FGE interactions). The genetic composition of the sensitive population was not significantly different from the *tetA*, *tetB*, or *tetC*/other populations (models not shown). The only significant GE association was found in the bla_{TEM}^+ population (Table 5b), where odds of sampling the bla_{TEM} locus on conventional farms were significantly associated with ECOR

Table 5. Log-linear modeling of farm type (F), resistance gene (G), and ECOR phylogrouping (E). A. Resistant populations were defined by the determinant they carried. Notice that GE was a significant term in the bla_{TEM} population only. B. Significant GE (two-way) interaction for the bla_{TEM} population.

A. Resistance gene model	Final model	G^2 (df, Pr > c^2)
tetA	$\mu + F + G + E + FE + FG$	7.32 (4, 0.12)
tetB	$\mu + F + G + E + FE + FG$	4.10 (4, 0.39)
tetC/other	$\mu + F + G + E$	11.24 (7, 0.13)
bla_{TEM}	$\mu + F + E + FE + FG +$ GE	3.81 (2, 0.15)
B. Resistance gene model	**Significant interactions with E**	**χ^2 (df, Pr > c^2)**
bla_{TEM}	GE	20.84 (2, < 0.0001)

phylogroup A (df = 1, χ^2 = 5.0, Pr > χ^2 = 0.025). These data suggest that the genetic composition of resistant *E. coli* populations on dairy farms is dependent on individual resistance determinants.

Genetic composition and class I integrons. All amp^R and tet^R strains were screened for the presence of class I integrons based on the presence of 3 loci (*intI1*, *qacEΔ1*, and the conserved cassette region). Of the total 298 amp^R and/or tet^R strains, 59 (19.8%) carried a class I integron. We created a 3 level factor called "integron populations" that was comprised of resistant, integron positive strains (int$^+$); resistant, integron negative strains (int$^-$); and sensitive, integron negative strains (sensitive). Log-linear models were then used to test for significant associations between farm type (F), ECOR phylogrouping (E), and integron populations (I).

There was no significant phylogroup-integron (EI) interaction with farm type (no FEI interaction). However, the distribution of ECOR phylogroups was dependent on integron presence in these populations (df = 4, χ^2 = 12.4, Pr > χ^2 = 0.015). The int$^-$ and sensitive populations were compositionally the same (B1 > A > B2D), but phylogroups in the int$^+$ population were evenly sampled. This analysis suggests that the int$^+$ population had significantly more group A and B2D strains than the other (int$^-$ and sensitive) populations.

DISCUSSION

In this study, we examined the dynamics of antibiotic selection on conventional and organic dairy farms by comparing the genetic composition of resistant and sensitive *E. coli* populations. The overall abundance of 4 phylogenetic groups was not similar, suggesting that phylogroup B1 strains colonize at a higher abundance and, therefore, have a higher relative fitness in dairy cattle. Although phylogroup B1 strains are common in a variety of host species (56), they were not the numerically dominant group in healthy swine (32) or humans (112).

The rate of compositional change in conventional-resistant populations. A key finding of this study is that there is an overabundance of resistant phylogroup A strains on conventional diary farms compared to phylogroup A strains on organic farms where antibiotic use has been limited. Based on two observations, we are confident that this overabundance has been a consequence of antibiotic use and not some other conventional management practice. First, conventional-sensitive and organic-sensitive populations are nearly identical in genetic composition and are not statistically different suggesting that these *E. coli* populations experience similar selective pressures in both agricultural environments. Second, the composition of the organic-resistant population was not significantly different from the sensitive populations. These observations also suggest the possibility that the conventional-resistant population will evolve to that of the sensitive, organic-resistant populations if antibiotic use was stopped. Given that organic farms in this study were certified as having not

used antibiotics for at least 3 years (mean = 8 years, range = 3 – 15 years), we estimate that, if antibiotic selective pressure to be removed, it will take at least this long for the compositional transition.

Evidence for clonal resistance dynamics. These data say little about the acquisition of resistance determinant by sensitive strains. However, we feel the data can adequately describe dynamics after resistance is conferred. For example, we expected to find a significant difference between the genetic composition of resistant and sensitive populations if a resistant clone swept to high frequency during drug use on conventional farms. We had the same expectation if clonal interference was operating between resistant clones of the same phylogroup. The significant association between conventional farms, antibiotic resistance, and phylogroup A strains supports this expectation. Further characterization is needed, however, to differentiate between the spread of one or multiple closely related clones.

Compositional similarity between sensitive populations and the organic-resistant population suggests that there is an optimal genetic composition (OGC) for the farms in this study. An overabundance of phylogroup A strains was significantly associated with the bla_{TEM} locus in the amp^R population and the presence of class I integrons in the overall resistant population. These data suggest that bla_{TEM} and class I integrons were linked to phylogroup A strains during selection on conventional farms and resulted in a departure from OGC. If this interpretation is correct, we predict that a more discriminate genetic

characterization of amp^R strains from conventional farms will reveal less genetic diversity in group A strains compared to phylogroups B1 and B2D.

Evidence for hitchhiking of resistance loci. In contrast to the amp^R population, there was no evidence supporting an underlying clonal model for the dynamics in the tet^R population. Populations carrying tet^R determinants (*tetA*, *tetB*, *tetC*/other) were at OGC on both farm types. This observation is difficult to explain if antibiotic selection and clonal spread were occurring on a single farm type. One explanation is that the organic farms received an occasional flux of tet^R strains from conventional farms and the migration was sufficient to maintain the observed similarity. However, this explanation seems unlikely because the occasional flux would likely bring amp^R strains from conventional farms as well, which in turn would ameliorate the differences discussed above. If tet^R loci were linked to other compensatory, beneficial mutations, then the composition of these populations might appear similar regardless of antibiotic use.

Several lines of evidence support the hypothesis for the role of hitchhiking or compensatory mutations in tet^R antibiotic resistance spread. Bartoloni et *al.* initially described a resistant *E. coli* population from humans living in a remote Guarani Indian community in Bolivia (9). Individuals of the village had little contact with outsiders, no veterinary or agricultural antibiotic use, relied on rainwater for survival, and had limited available health care (every 3 months). Yet, tetracycline resistance was found in 64% (69 of 108) of the individuals tested. Pallecchi et *al.* recently characterized the underlying genetic determinants and ECOR phylogroups for 113 resistant strains of the original collection (117).

The authors found that of the 103 tet^R strains analyzed, 52 carried *tetA* and 51 carried *tetB*. These loci were distributed among all 4 *E. coli* phylogroups (same procedure used in this study) and were found on all 5 conjugative plasmids identified in this study. The abundance and distribution of tet^R strains in this remote community supports the hypothesis that naturally occurring tet^R determinants circulate in hosts for reasons other than selection by drug use.

Support the hitchhiking hypothesis for tet^R loci is also consistent with the description of a "calf-adapted" *E. coli* population that was multiply resistant to streptomycin, sulfadiazine, and tetracycline (SSuT) (84). Almost all strains (49 of 50) analyzed shared a ~140 kb plasmid, the same resistance loci (*strA*, *sul2*, and *tetB*), and were genetically diverse by pulsed-field gel electrophoresis. Khachatryan and colleagues showed that on average the SSuT population out-competed sensitive strains *in vitro* and in neonatal calves (82). They also showed that the resistance loci themselves do not influence this selective advantage (83). Their main conclusion was that the combination of *strA*, *sul2*, and *tetB* in the original resistant population had hitchhiked with some other fitness–conferring locus.

Age effects on genetic composition. Sato et *al.* showed that the resistant strains examined here were most prevalent in calves on conventional farms (136). A similar positive association has been reported in other studies of pre-weaned calves and adult cattle (13, 15, 84). However, cattle age had little influence on the distribution of phylogenetic groups in either the sensitive or resistant populations of this study. These data suggest that the abundance of resistant strains decrease

as cattle get older, but the genetic composition of this population remains stable. Other analyses of human strains showed a significant association between host age and genetic composition, but the time reported for such change may be longer than the average life span of dairy cattle (58).

Multi-drug resistant influences on genetic composition. We found a rather complicated interaction between farm type, multi-drug resistance, and ECOR phylogrouping (Figure 4). Significant associations depended on the way our log-linear model was parameterized. However, all 3 possible parameterizations resulted in a significant association between low multi-drug resistance, group B1 strains, and organic farms. These data suggest an inverse relationship between multi-drug resistance and fitness for group B1 strains on organic farms. Since phylogroup B1 strains were the numerically dominant group overall, this result should be encouraging for those seeking to reduce the amount of multi-drug resistant strains in dairy cattle through limited antibiotic usage.

Two of the parameterizations showed an association with high multi-drug resistance, group D strains, and organic farms. This result is important because a number of human pathogens, including the strain most associated with human enterohemorrhagic colitis, O157:H7, belong to this group (according to the PCR method used here). However, we are cautious to base generalizations on this analysis because 1.) we did not design our sampling study to directly address this question and 2.) the abundance of strains used for these comparisons were low. For example, the FE(D = high) association (Table 2) between highly resistant phylogroup D strains and organic farms becomes non-significant if 2 fewer strains

were sampled on organic farms and 2 additional strains were sampled on conventional farms. Similarly, we are cautious about the association between high multi-drug resistant phylogroup B1 strains and medium multi-drug resistant phylogroup A strains on conventional farms because the significance of the association depends on the model parameterization.

Conclusion. The genetic composition for the conventional-sensitive, organic-sensitive, and organic-resistant *E. coli* populations was the same, suggesting an optimal genetic composition (OGC) for the farms in this study. The conventional-resistant population had an overabundance of amp^R, group A strains, that could be explained by linked loci (bla_{TEM} and class I integrons) during a selective sweep or clonal interference among closely related strains. Given the amount of time since organic farms had abandoned conventional practices, the rate of compositional change was estimated to be between 3 and 15 years (mean = 8 years). In contrast to the amp^R population, the tet^R populations analyzed here showed no clonal dynamics and appeared to be at the OGC. These data add support to the previously founded hypothesis that the abundance and distribution of tet^R determinants are weakly influenced by antibiotic use. We found that cattle age had little influence on the genetic composition of the resistant or sensitive populations. Finally, phylogroup B1 strains with low multi-drug resistance were significantly associated with organic farms, suggesting that these dairy farming practices have a proportionately large, negative effect on the prevalence of multi-drug resistant strains.

ACKNOWLEDGMENTS

This work was supported in part by MRU matching funds from the College of Veterinary Medicine and Graduate College of Michigan State University.

CHAPTER 4

GENETIC DIVERSITY AND POPULATION STRUCTURE OF *ESCHERICHIA COLI* ISOLATED FROM FRESHWATER BEACHES

INTRODUCTION

Escherichia coli reaches high density in the gastrointestinal tract of humans (~10^6 cells per gram of colonic content) and other warm-blooded animals and are inevitably excreted to the external environment where hosts live (137). These two habitats, called the primary (within host) and secondary (outside the host), represent distinct ecosystems and differ in both the number and heterogeneity of harmful stimuli. Secondary habitat stimuli such as UV radiation, temperature, and predation have been shown to decrease the density of individual strains to undetectable levels under controlled conditions (7, 81, 153). Based on such evidence, it is often concluded that secondary habitats do not actively support *E. coli* growth and therefore have little influence on the adaptive evolution of the species (181). This conclusion underlies both the use of *E. coli* as an indicator organism of fecal contamination and water pollution as well as its use as a tool for bacterial source tracking (64, 167).

One implication of this hypothesis is that natural selection in the primary habitat is a dominant influence on the genetic structure of populations sampled from the secondary habitat (54). However, there are strikingly few data with which to assess the contribution of selection and other genetic processes in the secondary habitat on the variability and organization of natural *E. coli* populations. Two observations suggest that adaptive evolution in the secondary habitat can substantially influence population genetic structure of the *E. coli* species as a whole; i.) First, the population size of *E. coli* in the secondary habitat may be very large, as it is estimated that half of all living cells are presently

outside of a host (137). ii.) Second, data from multiple studies in both tropical and temperate regions suggest that this organism can replicate and reach high densities under favorable conditions outside of mammalian hosts (2, 4, 16, 156, 171) and in the absence of regular fecal input (92, 123).

To address questions about the fundamental evolutionary processes operating in *E. coli* populations, two principal methods utilizing population genetic approaches have been utilized. Multilocus enzyme electrophoresis (MLEE) has been applied for more than 25 years to collections of *E. coli* isolates obtained from a variety of human sources and other animals (53, 106, 115, 145, 157, 174). MLEE of protein polymorphisms has uncovered abundant allelic variation and extensive multilocus linkage disequilibrium (176), that is, populations or collections of pathogenic strains where certain genotypes are numerically dominant and geographically widespread (37, 55, 101, 179). Recently, a worldwide analysis based on multilocus sequence typing (MLST) revealed extensive allelic variation and homologous recombination, with accelerated rates of evolution in pathogenic lineages of *E. coli* (182). However, there have been relatively few studies of the genetic variability and population structure of native *E. coli* isolated from environmental sources (55, 173).

In this paper, we characterize the phenotypic and genetic diversity of 190 *E. coli* strains isolated from six freshwater beaches along Lake Huron and the St. Clair River in Michigan (2). We use standard biochemical assays to establish and characterize individual phenotypic profiles (biotypes). Population genetic analyses from MLEE of 18 housekeeping loci are compared to those obtained

using MLST of 7 conserved genes. We assess the abundance and distribution of biotypes, multilocus genotypes, and the accuracy of a commonly used PCR-based phylogrouping technique (33). We also analyze recombination rates and linkage disequilibrium, which suggests that despite extensive recombination in nature, natural selection is favoring certain *E. coli* genotypes, particularly those of the B1 phylogroup, in the secondary habitat.

EXPERIMENTAL PROCEDURES

Beach site description and strain isolation. A detailed description of the sampled beach sites and isolation procedures can be found elsewhere (i.e. air and water temperatures, soil composition, and moisture content were monitored throughout the study period) (2). Briefly, six sites along Lake Huron and the St. Clair River of Michigan were sampled from August 2001 to March 2003 with the majority of samples taken during the summer months of 2002. For clarity, sites were labeled here using the following scheme: Conger Lighthouse Beach, 1; Holland Road Beach, 2; Lakeport State Park, 3; Lakeport State Campground, 4; Marine City Beach, 5; Chrysler Park Beach, 6. Samples were taken from sand cores (< 20cm) in the wave-wash zone and the overlying water column. The sites were chosen to represent a variety of fecal input sources (agricultural, industrial, and human septic runoff as well as wildlife inputs).

Single colonies of presumed *E. coli* strains were isolated and enumerated from the water column of the beach sites based on a protocol published by the U.S. Environmental Protection Agency (167). Strains associated with sand were isolated in a similar manner from 9 x 20 cm sand cores that were sliced into 5cm sections (1 – 5 cm, 6 – 10 cm, 11 – 15 cm, 16 – 20 cm). Sediments were agitated for ~1 minute to suspend cells and then processed as water samples. Approximately 35 strains per site (a total of 205) were randomly selected and characterized.

Species delimitation. Tests of 21 standard biochemical reactions (see API20E in biotyping section below) revealed that some of the 205 isolates

originally selected for analysis had characteristics discordant with the traditional (biochemical) *E. coli* species definition. Nine strains were positively identified as species other than *E. coli* and were excluded from further analysis. To define which of the remaining 196 isolates should be included as representative, we compared the biochemical-based classification with the genetic relatedness of isolates. Clustering of isolates based on MLEE and MLST (dendrograms not shown) identified two distinct groups of strains. Group 1 was the largest (190 isolates) and contained strains with excellent, acceptable, low, and unknown API20E profiles. Group 2 comprised 6 strains that clustered distinctly apart from Group 1 and also had excellent, acceptable, and unknown profiles. To ensure these strains were not typical, but divergent, *E. coli*, we attempted to verify their genus and species based on 16S rRNA operon sequencing. Published universal primers (8F and 1492R) targeting the 16S rRNA gene and appropriate PCR conditions were used to amplify a 1484 bp product as described by Schmidt *et al.* (90, 140). Sequencing was performed in both directions of the locus (2X coverage). Genus and species identity was determined based sequence similarity using the nucleotide BLAST algorithm (blastn) on the National Center for Biotechnology Information (NCBI) website (www.ncbi.nlm.nih.gov/BLAST/). Two isolates had 99% similarity to the distantly related, *E. fergusonii*. However, four strains had 99% 16S rRNA sequence similarity to *E. coli* (3 to *E. coli* and 1 to *Shigella boydii*). It was unclear if these isolates were representatives of a more distantly related *Escherichia* lineage or perhaps represented species "hybrids." In either case, we excluded all Group 2 isolates from further analyses.

Phenotypic profiling (biotyping). A phenotypic profile was generated for each strain using 20 reactions of the API20E bacterial identification system (BioMerieux, Inc.) and a filter paper oxidase test (Becton, Dickinson and Company). Positive (1) or negative (0) results were recorded for the following reactions: beta-galactosidase (ONPG), arginine dihydrolase (ADH), lysine decarboxylase (LDC), ornithine decarboxylase (ODC), citrate utilization (CIT), H2S production (H2S), urease (URE), tryptophane deaminase (TDA), indole production (IND), acetoin production (VP), gelatinase (GEL), and fermentation/oxidation of glucose (GLU), mannitol (MAN), inositol (INO), soribitol (SOR), rhamnose (RHA), sucrose (SAC), melibiose (MEL), amygdalin (AMY), and arabinose (ARA). Only the informative reactions (positive or negative for at least two strains) were used to analyze phenotypic diversity.

Multilocus enzyme electrophoresis (MLEE). Enzyme extraction, gel electrophoresis, and specific enzyme staining were carried out as described in Selander *et al.* (144). Briefly, lysates (whole cell enzymes) were extracted via centrifugation from overnight nutrient broth cultures and frozen at -80°C. Of the 190 *E. coli* strains, 5 cultures failed to pellet when centrifuged and did not produce sufficient concentration of enzyme lysates for further MLEE analysis. These cultures produced an excess exopolysaccharide or other capsular matrix that inhibited sedimentation of cells under the centrifugation conditions. Lysate for the remaining 185 samples were individually loaded and electrophoresed under non-denaturing conditions in a buffered starch gel matrix at the appropriate concentration for the particular enzymes being stained. Gel slices were incubated

in enzyme specific staining solutions and fixed for analysis. The mobility of 18 housekeeping enzymes was recoded for all strains analyzed. Mobilities were scored relative to previously published strains from our database. The enzymes characterized in this study were: ADH (alcohol dehydrogenase), THD (threonine dehydrogenase), SKD (shikimate dehydrogenase), G6P (glucose-6-phosphate), MPI (mannose phosphate isomerase), GLUD (glutamate dehydrogenase), MDH (malate dehydrogenase), NSP (nucleoside phosphorylase), PEP (peptidase), GOT (glutamic oxalacetic transaminase), CAK (carbamylate kinase), AK (adenosine kinase), M1P (mannitol-1-phosphate dehydrogenase), PGD (6-phosphogluconate dehydrogenase), PGI (phosphoglucose isomerase), IDH (isocitrate dehydrogenase), ACO (aconitase), G3P (glyceraldehyde-phosphate dehydrogenase).

Electrophoretic mobility variants (electromorphs) were assigned scores in side-by side comparisons relative to the published strains. Electromorphs were equated with alleles at the corresponding enzyme locus based on the assumption that each electromorph represents at least one amino acid replacement in the protein. The combination of 18 electromorphs (alleles) was used to generate a multilocus genotype called an electrophoretic type (ET). The same ET was assigned to strains with indistinguishable allele profiles. We assumed that strains with the same ET owe their genotypic similarity to recent descent from a common ancestral cell; that is, they are members of a naturally occurring bacterial clone.

Allele frequencies and the number of strains were used to estimate population genetic parameters as described in Selander *et al.* (144). Computer programs were used to calculate parameter estimates and can be found at (www.foodsafe.msu.edu/whittam/programs/index.html). We assessed genetic differences between beach sites by partitioning the total genetic diversity as described previously (180). The coefficient of genetic differentiation among groups (G_{ST}) was calculated for each locus using the program ETdiv.

DNA isolation and multilocus sequence typing (MLST). Genomic DNA was isolated from 2 ml of overnight culture in LB broth, Lennox (Becton, Dickinson, and Company, Sparks, MD) using a Puregene DNA isolation kit (Gentra Systems Inc., Minneapolis, MN.). DNA preparations were quantified with a NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE.), diluted to a final concentration of 100 ng/μL, and stored at 4°C.

Primers and sequencing methodology used for MLST were chosen and carried out as part of a system described in detail elsewhere (www.shigatox.net/cgi-bin/mlst7/index). Briefly, the internal fragments of the following 7 housekeeping genes were obtained for analyses: *aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh*, and *uidA*. The program SeqMan II (DNASTAR, Inc., Madison, WI) was used to edit and align the sequences. Sequences were concatenated and uploaded into MEGA3 software (91) for descriptive statistics. For further analysis, we assumed that each unique gene sequence represents an allele and each unique concatenated sequence represents a multilocus genotype or sequence

type (ST). Descriptive statistic for MLST data including the number of variable sites and codons, parsimoniously informative nucleotide sites, and estimates of d_S and d_N were done using MEGA3.1 (modified Nei-Gojobori model; Jukes-Cantor). G_{ST} was calculated as described in the MLEE procedures above for each of the 7 sequenced loci.

Estimation of diversity. Nonparametric diversity estimates were generated based on STs sampled per site and the Chao1 algorithm (classic formula) in the program EstimateS (version 7.5.0, The University of Connecticut, Storrs, CT) (36). We chose to use this algorithm to estimate genotypic diversity because it is well suited for a small number of samples from diverse habitats (20) and it has a closed-form solution for variance estimation (31). As more sites are included in the analysis, the confidence intervals around the diversity estimate tend to decrease, allowing for direct comparisons between studies (71). The estimate calculated here was based on the total number of observed STs, the number of STs sampled once (singletons), and the number of STs sampled twice (doubletons).

ECOR phylogrouping by multiplex PCR. All strains were grouped into 1 of 4 phylogenetic groups based on methods adapted from Clermont *et al.* (33). To help prevent false negative results, the same genomic DNA used for MLST analysis was used as template for a duplex PCR targeting the genes *chuA* and *yjaA* (duplex conditions: denaturation at 94°C for 10 minutes; 35 cycles of 92°C for 1 minute, 59°C for 1 minute, 72°C for 30 seconds; and a final elongation at 72°C for 5 minutes). A separate PCR was run with primers targeting the TspE4.C2

anonymous DNA region using published conditions. An *R* X *C* test of independence was done using the G-test to test for significant association between phylogroups and sampling sites (154). This test is a nonparametric method used to determine if variables are significantly associated, and the null hypothesis considered was that the frequency of a phylogroup is independent of sampling site.

Phylogenetic analysis. Two phylogenetic analyses were applied to the sequence data for 130 sequence types (STs) determined by MLST. A neighbor-joining tree was constructed using the Kimura 2-parameter model of nucleotide substitution with the MEGA3 software (91) and the inferred phylogenies were each tested with 1000 bootstrap replications. Phylogenetic network analysis was conducted with the SplitsTree 4 (74) program using the neighbor-net algorithm (26) and untransformed distances (*p* distance).

Linkage disequilibrium and recombination. We used allele profiles and the program MultiLocus (version 1.2, Imperial College, available at www.bio.ic.ac.uk/evolve/software/multilocus) to test for significant linkage disequilibrium. Two statistics used (I_A and $rbar_d$) are detailed by the authors of MultiLocus (1). Briefly, the index of association, I_A, measures linkage disequilibrium and is used to describe nonrandom associations of alleles at loci (24, 65, 102). The $rbar_d$ calculation is a standardized measure of I_A that corrects for the number of loci used in the analysis. Both statistics should equal 0 if there is no linkage between alleles at different loci. After calculating each statistic the program generates p-values based on comparisons between the observed and

expected (panmictic) datasets where alleles are randomly shuffled among strains at loci. We used 1000 randomizations for all p-value calculations. The ϕ_w recombination test (25) as implemented by SplitsTree 4 was used to distinguish recurrent mutation from recombination in generating genotypic diversity. The ratio of recombination to mutation events was inferred by grouping sequence types into BURST groups (49) and the counting method for groups with founding sequence types (50).

RESULTS

Biotype diversity. Of the 21 biochemical reactions analyzed, 4 were positive (ONPG, GLU, MAN, and ARA) for all strains, 6 were negative (H2S, URE, TDA, GEL, INO, and OX) across the 190 *E. coli* strains, and 1 was positive (CIT) in only one strain and was, therefore, not informative. The remaining 10 assays were variable with some strains either positive or negative for each test. The results of the 10 variable tests were used to generate a strain- specific biotype (Table 6). There were a total of 76 distinct biotypes among the 190 *E. coli* strains and a similar number of biotypes were sampled across the six beach sites (average = 12.6, range = 10 – 14). One profile (Biotype 1) was recovered 86 times and accounted for 45.3% of biotypic variation (Table 7). Biotype 1 was the most frequently isolated biotype at all sites (average = 14.3, range = 11 – 19), and the four most common profiles (Biotypes 1 – 4) accounted for 70% of total biotypic variation.

Levels of protein polymorphism by MLEE analysis. To measure the extent and organization of genetic diversity among environmental strains of *E. coli*, we used the methods of MLEE to resolve protein polymorphisms at 18 enzyme encoding genes (144). Among the 185 strains analyzed (see *MLEE* section in Experimental procedures), the number of electromorphs (alleles) resolved per locus ranged from 3 (NSP and CAK) to 13 (PGD) with an average of 6.4 alleles per locus (Table 8). The average single-locus genetic diversity (h = the probability that two randomly sampled strains have a different allele at a locus)

Table 6. The number of bacterial isolates recovered at 6 sampling sites along Lake Huron and the St. Clair River of Michigan. The number of biotypes and genotypes (ETs and STs) are given for all confirmed *E. coli* isolates.

Site[a]	total	Biotypes	ETs[b]	STs
1	32	13	26	30
2	31	13	24	24
3	32	13	26	26
4	34	10	26	28
5	33	13	28	30
6	28	14	26	27
Total	190	76	143	130

[a]Site names are listed in Experimental procedures as published in Alm et *al.* (2003).

[b]Five isolates were not analyzed by MLEE (See *MLEE* section in Experimental procedures).

Table 7. Phenotypic profile, number, and overall percentage of the four most common biotypes at each sampling site and nine unknown biochemical biotypes.

Biotype	Profile[a]	API call (% of call)	Site A	B	C	D	E	F	Overall percentage
1	0111011110	*E. coli* (99.6)	16	11	13	19	16	11	45.3
2	0101011010	*E. coli* (99.8)	4	2	5	4	1	4	10.5
3	0111011010	*E. coli* (99.9)	3	1	0	5	2	3	7.4
4	0101011110	*E. coli* (99.9)	3	1	2	1	4	2	6.8
5	0011111000	Unknown	0	1	0	0	0	0	0.5
6	0101001111	Unknown	1	0	0	0	0	0	0.5
7	0110111010	Unknown	0	0	1	0	0	0	0.5
8	0111011110	Unknown	0	0	0	0	1	0	0.5
9	0111101110	Unknown	0	0	0	1	0	0	0.5
10	0111111010	Unknown	0	0	0	1	0	0	0.5
11	0111110110	Unknown	0	1	0	0	0	0	0.5
12	0111011110	Unknown	0	0	0	0	0	2	1.0
13	1111011110	Unknown	0	0	2	0	0	0	1.0

[a]Profile is based on 10 informative (variable) biochemical reactions. Each digit represents a positive (1) or negative (0) reaction for ADH, LDC, ODC, IND, VP, SOR, RHA, SAC, MEL, and AMY, respectively.

Table 8. Electrophoretic variation of 18 housekeeping enzymes analyzed by MLEE.

Enzyme locus	No. of alleles[a]	Single locus heterozygosity (h)[b]	Frequency of the most common allele[c]	G_{ST}[d]
PGI	8	0.587	0.541	0.000
IDH	6	0.342	0.795	0.000
ACO	8	0.373	0.784	0.000
G3P	5	0.510	0.584	0.015
PGD	13	0.544	0.665	0.000
M1P	4	0.455	0.697	0.034
NSP	3	0.141	0.924	0.000
MDH	4	0.187	0.897	0.000
TDH	5	0.283	0.843	0.002
SkD	11	0.577	0.632	0.023
G6P	6	0.256	0.860	0.023
PEP	7	0.403	0.762	0.000
GLU	5	0.053	0.973	0.000
CAK	3	0.103	0.946	0.000
AK	4	0.200	0.892	0.014
GOT	5	0.161	0.914	0.000

Table 8 (continued).

Enzyme locus	No. of alleles[a]	Single locus heterozygosity (h)[b]	Frequency of the most common allele[c]	G_{ST}[d]
ADH	6	0.523	0.649	0.036
MPI	9	0.726	0.449	0.023

[a]Average = 6.2

[b]Average single locus heterozygosity (H) = 0.357

[c]Product of allele frequencies =

[d]Average GST = 0.011

among the 185 strains was 0.357 (± 0.046; SE). The allelic variation was organized into 143 distinct multilocus genotypes or electrophoretic types (ETs).

Across the six beach sites, the genetic diversity was relatively uniform with similar numbers of ETs recovered at each site (average = 26 ETs per site, range 24 to 28). There were 132 unique (sampled only once) and 11 common (repeatedly sampled) ETs. The most common ET (ET-1) was isolated 30 times (16% of all of the Group 1 isolates) from all 6 beach sites, over a 12 month sampling period. This common genotype represents the modal ET, that is, the combination of the most common allele at every locus (Table 3).

Is there evidence for genetic differentiation or subdivision of the *E. coli* populations from different sampling sites? To assess this, we partitioned the total genetic diversity per locus into within and between site components (110). On average, the within-site component of diversity (H_S = 0.356) accounted for nearly all of the total genetic diversity (H_T = 0.360). The average coefficient of genetic differentiation (G_{ST}) across the 18 loci was 0.011 and ranged from <0.0001 to 0.036 (Table 3), indicating that ~1% of the total diversity is accounted for by differences between sites; thus, there is no evidence for local population subdivision or the differentiation of allele frequencies across beach sites.

DNA sequence polymorphisms revealed by MLST analysis. In contrast to MLEE, which assigns alleles indirectly via the electrophoretic mobility of their gene products, alleles at multiple housekeeping loci can be defined directly by DNA sequencing via multilocus sequence typing (MLST). We sequenced the internal ~500 bp on both strands of 7 housekeeping genes (*aspC*, *clpX*, *fadD*,

icdA, *lysP*, *mdh*, and *uidA*) and assembled consensus sequences for the genes in every strain. Sequence comparisons among the 190 *E. coli* strains uncovered substantial DNA polymorphism with an average of 61.3 variable nucleotide sites (range = 40 to 100) and 8.1 variable codons (range = 2 to 25) per locus. We resolved a similar number of alleles at each of 7 housekeeping loci (average = 41.6) ranging from 30 alleles for *lysP* to 49 for *uidA* (Table 9).

An advantage of MLST data is that one can assess the nature of selection on allelic variation by comparing rates of nonsynonymous (amino acid changing) nucleotide substitutions to synonymous (non-amino acid changing) substitutions (110). There was a wide range of variation observed in the percentage of nonsynonymous nucleotide substitutions per nonsynonymous site (d_N x 100, Table 9) with respect to the percentage of synonymous substitutions per synonymous site (d_S x 100). This variation resulted in d_S to d_N ratios that ranged across loci by 2 orders of magnitude (16.2 for *uidA* to 1484.2 for *clpX*). The low percentage of nonsynonymous substitutions relative to synonymous substitutions, especially for *clpX* and *lysP*, suggests that natural selection at the molecular level is strongly negative and acting to limit the amount of amino acid polymorphism at housekeeping genes in this environment.

The number of distinct combinations of alleles at 7 MLST loci resolved 130 different multilocus genotypes or sequence types (STs); the number of STs found for 7 MLST loci is comparable to, but slightly less than the number of ETs resolved for 18 MLEE loci. Among the 130 STs, there are 103 singletons (STs isolated only once) and 27 common STs (isolated more than once). The

Table 9. Nucleotide variation within seven housekeeping genes analyzed by MLST.

Locus	Length	Variable sites (%)	Variable codons (%)	No. of alleles	d_S x 100 (mean ± SE)	d_Nx100 (mean ± SE)	d_S / d_N
aspC	513	40 (7.8)	8 (4.7)	32	3.26 (0.85)	0.04 (0.02)	79
clpX	567	69 (12.2)	2 (1.1)	45	8.00 (1.30)	0.01 (<0.01)	1484
fadD	492	100 (20.3)	10 (6.1)	48	6.96 (1.29)	0.04 (0.01)	164
icdA	567	59 (10.4)	5 (2.7)	44	6.11 (1.10)	0.03 (0.03)	183
lysP	477	44 (9.2)	2 (1.3)	30	1.61 (0.45)	<0.01 (<0.01)	260
mdh	549	46 (8.4)	5 (2.7)	43	3.162 (0.78)	0.113 (0.09)	28
uidA	588	71 (12.1)	25 (12.8)	49	4.136 (0.84)	0.255 (0.09)	16
All strains	3753	429 (11.4)	57 (4.6)	130[a]	4.727 (0.38)	0.074 (0.02)	64

[a]Number of sequence types (STs)

distribution of the 12 most common STs (sampled 3 or more times) was similar to the distribution of ET-1 from MLEE analysis. For example, one of these common STs was independently isolated 7 times, from 5 of 6 sites, in the water column, at all 5 depths in the sand cores, and at separate times over the 35 months of the collecting times. These observations suggest that recovery of isolates of the same ST is a result of the repeated isolation of a widespread genotype that is in high frequency in the beach habitat and is not an artifact of the sampling protocol.

For comparison to MLEE, we used the allelic variation determined by MLST to assess the extent of genetic differentiation of the bacterial populations across the sampling localities. The calculated G_{ST} for the MLST alleles ranged from 0.003 to 0.023 with an average of 0.012, a value almost identical to that calculated by the MLEE analysis. Based on the similarity of G_{ST} values calculated by both the MLEE and MLST analyses, it is likely that the *E. coli* samples from the different beach sites are genetically indistinguishable and thus represent random samples from a single, diverse and well mixed bacterial population.

If the population genetic diversity is relatively uniform across sites, as suggested by G_{ST}, then estimates of allelic and genotypic diversity based on the observed dataset should be very similar to estimates obtained when alleles or STs are randomly assigned to sites. We tested this hypothesis by estimating the total number of STs across beach sites based on the observed (open circles) dataset and a randomized (black circles) dataset (Figure 5). There was no significant difference between estimates made using the Chao 1 algorithm (36), although the

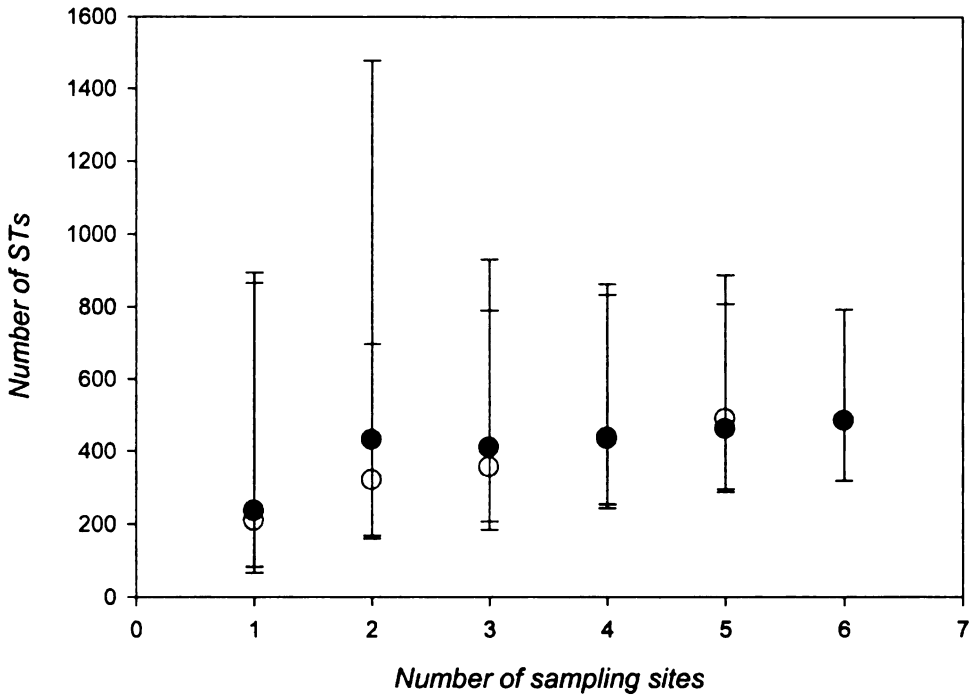


Figure 5. Chao1 diversity (36) estimates for the total number multilocus genotypes (STs) based on MLST analysis. Average estimates are marked by circles and 95% confidence intervals are designated by brackets. Estimates based on the observed dataset (open circles) are not significantly different from estimates based on a random site assignment (black circles).

95% confidence intervals were large. This finding agrees with the small G_{ST} values, and indicates that there is virtually no difference in the diversity of STs across sampling sites.

Abundance and distribution of phylogroups. Phylogenetic studies have shown that commensal and pathogenic *E. coli* strains can be subdivided into 4 major phylogroups referred to as the ECOR A, B1, B2, and D groups (68, 145). These groups appear to differ in important aspects of their population biology. For example, many commensal strains belong to groups A and B1 whereas *E. coli* associated with human extraintestinal infections belong to groups B2 and D (58, 111-113). There are, however, exceptions as certain human populations have relatively high counts of naturally occurring B2 and D strains in their fecal microbial flora (45).

To determine the distribution of ECOR phylogroups among the *E. coli* from the beach environment, we classified all 190 strains into 1 of 4 phylogenetic groups based on PCR methods (33). Group B1 strains were isolated more often than any other group across 5 of 6 sites and 56% (106/190) of all strains were classified into this group (Figure 6). Group B2 strains were infrequently recovered across all 6 sites and only 12 strains (6%) were classified into this group. Group A and Group D strains were isolated at intermediate frequencies (23% and 15%, respectively). Overall, the distribution of groups was similar across the sites (Figure 6), and the frequency of phylogroup isolation was statistically independent of beach sites (G-test of independence, $\chi^2_{0.05,15} = 25$, $G_{adj.} = 5.9$). At site 6, there were more Group A strains than B1, but the difference was

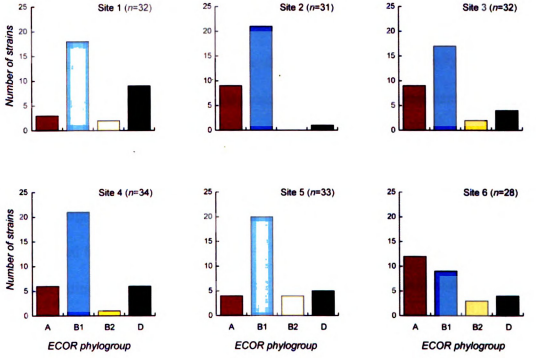


Figure 6. Distribution of *E. coli* phylogroups across the 6 sampling sites. The frequency of each phylogroup (A, B1, B2, and D) was independent of site (i.e. there was no difference in the overall distribution of the phylogroups from site to site).

not significant. These data are consistent with the previous inferences drawn from the G_{ST} values and Chao 1 estimates that the genetic composition of the *E. coli* samples is virtually the same and across the various beach sites.

Recombination and linkage disequilibrium. To test statistically for recombination, we used the ϕ_w test, which has been shown to discriminate between recurrent mutation and recombination in a variety of circumstances (25). Application to each of the 7 MLST loci found that 4 out of the 7 loci show evidence of significant recombination in generating allelic variation (Table 10). We also tested for the role of recombination in generating multilocus genotypes by concatenating the sequences of the 130 STs. There were 279 informative sites and the ϕ_w test found statistically significant evidence of recombination (p <0.001) overall including both the within-gene recombination, which generates new alleles, and assortative recombination, which shuffles existing alleles into new genotypic combinations.

What is the relative contribution of recombination and mutation to genotypic variation in the natural environment? To address this question, we used the ratio of number of recombination events, R, to the number of mutation events, M, based on organizing STs into distinct BURST groups (49, 50). There were 83 STs that could be classified into BURST groups and for which an ancestral or founder genotype could be inferred. Comparison of the allele profiles to that of the founder genotype of each group identified a total of 30 genetic events. Twenty-one of the 25 recombination events (R) are accounted for by changes at the 3 loci, all of which had significant p values in the ϕ_w test (Table 10). Overall

Table 10. Recombination analysis based on sequence comparisons. There is significant evidence for recombination (*p* values for ϕ_w recombination test) was detected for 4 loci among all isolate (*n*=190) and across sequence types (STs, *n*=130). However, there was no significant recombination detected for strains assigned to BURST groups (*n*=83). Also given are the number of recombination (R) and mutation (M) events inferred within BURST groups in which an ancestral sequence type was predicted.

	All isolates	STs only	BURST groups	R	M
aspC	0.166	0.142	0.084	3	0
clpX	<0.001	<0.001	0.118	4	1
fadD	<0.001	<0.001	0.121	11	1
icdA	<0.001	<0.001	0.132	6	0
lysP	0.116	0.116	1.000	0	1
uidA	<0.001	<0.001	0.258	1	2

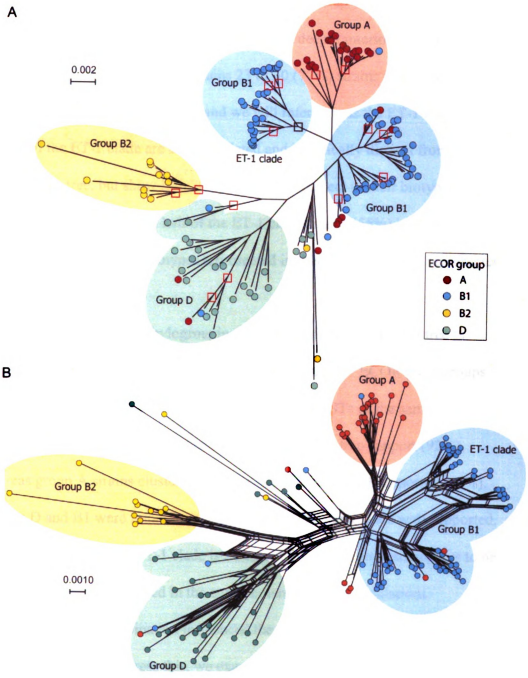
the R:M ratio was 5:1 which indicates that the rate of recombination is ~5 times the rate of mutation in generating new STs within the BURST groups.

We also analyzed sequences for nonrandom associations between alleles at the MLST loci. Two test statistics, I_A and $rbar_d$, were used to compare the observed statistical associations to those expected from a population at linkage equilibrium (i.e. undergoing extensive recombination). Both statistics show significant departures from the expected value (test statistic of zero) when all strains were analyzed together. Further analyses with the strains subdivided based on STs, beach sites, or dendrogram clustering (clusters with bootstrap support > 90) found that all statistics for that subdivided data were significant (data not shown) indicating the presence of strong and substantial multilocus linkage disequilibrium. Thus, there is significant linkage disequilibrium in this natural population despite the fact the recombination is generating new genotypes at a rate 5 times the point mutation rate. Therefore, the rate of recombination is not sufficient to randomize alleles in genotypes and remove linkage disequilibrium regardless of potential benefits of ongoing recombination in the natural habitat.

Phylogenetic analysis of genetic diversity. To compare the PCR-based classification of phylogroups to the level of sequence divergence as measured by MLST, we constructed a neighbor-joining (NJ) dendrogram showing the genetic relatedness among the 130 *E. coli* STs (Figure 7A). There were 15 highly supported bifurcating nodes identified that ranged in bootstrap confidence values from 90% to 100% (boxes in Figure 7A). As would be expected from MLST analysis where a small number of loci are analyzed, almost all highly supported

Figure 7. Phylogenetic relationships of freshwater beach *E. coli* based on MLST analysis and ECOR groupings. A. Phylogenetic tree based on the neighbor-joining algorithm and genetic distances for 130 sequence types (STs). There are 15 nodes with high bootstrap values (≥ 90%) from 1000 replicates that are marked with boxes (black box is the only interior node with high support). Circles on tips of branches represent STs and colors represent results from phylogrouping by multiplex PCR. B. Phylogenetic network based on the neighbor-net algorithm. Parallelograms denote branches with conflicting phylogenetic signals as a result of recombination or recurrent mutation.

Figure 7.


A
Group A
0.002
Group B1
Group B2
ET-1 clade
Group B1
Group D
ECOR group
A
B1
B2
D
B
Group A
ET-1 clade
Group B2
Group B1
0.0010
Group D

nodes contain a small number of STs (2 to 10) and are located at the tips of the clusters. However, one highly supported node (bootstrap value = 91%) comprised 27 STs representing 44 strains and is located at a deeper interior branch (black box in Figure 7A). This clade contains 23 of 30 (77%) strains identified as the ET-1 genotype by MLEE analysis and we will refer to it as the ET-1 clade. Strains of the ET-1 clade are closely related and genetically distinct from other strains in the tree, but show little uniqueness with respect to their biotypic profiles. For example, over half of the ET-1 clade strains (55%) had identical biotypes. However, all 8 biotypes represented in this clade were also represented in other clades.

The results of the phylogrouping PCR were plotted on the NJ tree to determine the correspondence between ST clusters and the ECOR phylogroups (Figure 7A). There is some evidence for clustering of STs of the same phylogroup. Group B2 strains clustered with high bootstrap support (92%), whereas group A strains clustered together, but were not well supported (50%). Groups D and B1 were found in multiple clusters and some were well supported. It is noteworthy that the ET-1 clade is a subset of the B1 group and appears to be numerically over-represented in the environmental *E. coli* populations.

How has recombination influenced the genetic relationships among sequence types? To address this, we examined a phylogenetic network (Figure 7B), based on Splitstree analysis (74). This analysis does not force the sequence data into a bifurcating tree and allows for numerous parallel paths indicative of the presence of phylogenetic incompatibilities in the divergence of STs. Such

incompatibilities could arise from recombination or recurrent mutation in the MLST loci. Interesting, despite the abundant recombination, the four main ECOR phylogroups are separated and relatively intact (Figure 7B). It also appears that the effect of recombination varies among phylogroups with genotypes of the B1 group showing the most extensive amount of recombination.

Phylogroup transition events. We compared the position of strains in the phylogenetic tree with the ECOR phylogrouping to detect potential gene loss and acquisition events among the three discriminating loci (*chuA*, *yjaA*, and TspE4.C2). A phylogroup transition matrix was made based on the possible transitions between the phylogroups A, B1, B2, and D (Table 11A). As can be seen in Figure 7, some STs cluster between these 4 major groups and, perhaps, represent previously discovered hybrid lineages that have a mixture of phylogenetic signals (182). If so, we would not expect the phylogrouping method used here to correctly classify these strains. However, we can answer the question, which transition events are more likely to occur in the freshwater beach environment? Of the 10 possible kinds of transitions (gene losses/acquisitions) between the groups, only five were seen. A total of 14, 3, and 2 false positive PCR results were discovered for certain strains of group A, B1, and B2, respectively, based on their position in the tree (Table 11B). There were statistically more transitions explained by gene loss as strains classified as phylogroup A by PCR clustered with strains of B1 (11 of 43 strains) or D (3 of 43 strains) by MLST. There were 2 transitions identified among 12 group B2 strains

Table 11. Possible gene loss/acquisition events (A.) and phylogrouping results (B.). False positive phylogrouping results were highest for gene loss events.

A. Gene loss/acquisition events

Possible Events	
1	loss of *chuA*
2	gain of *chuA*
3	loss of *yjaA*
4	gain of *yjaA*
5	loss of TspE4.C2
6	gain of TspE4.C2

B. Phylogrouping results

Strains of phylogroup	Clustered w/ other strains of	# of gene losses	# of gene acquisitions	total (%)[a]
A	A	0	0	29/43 (67)
A	B1	1	0	11/43 (26)
A	D	1	0	3/43 (7)
B1	B1	0	0	103/106 (97)
B1	D	1	1	2/106 (2)
B1	A	0	1	1/106 (1)
B2	B2	0	0	10/12 (83)

Table 11 (continued).

Strains of phylogroup	Clustered w/ other strains of	# of gene losses	# of gene acquisitions	total (%)[a]
B2	D	0	1	2/12 (17)
D	D	0	0	29/29 (100)

[a]For the rows where strains of a given phylogroup clustered with other strains of the same phylogroup, the values represent true positive results. For the rows where strains of a given phylogroup did not cluster with other strains of the phylogroup, the values represent false positive results.

(17%) and 3 transitions among 106 strains of group B1 (3%). All group D strains identified by PCR clustered together by NJ.

DISCUSSION

Genetic diversity of environmental *E. coli*. Although the presence of *E. coli* in natural waters has long been used as an indicator of fecal pollution, there is a growing body of data suggesting that there exists a specialized subset of *E. coli* that can reproduce and persist in secondary environments outside animal hosts in both tropical (16, 131, 156) and temperate climates (27, 55, 77, 103, 171). The purpose of the current study was to characterize the genetic variation and population structure of *E. coli* recovered from the water and sand in temperate freshwater beaches. We found that the average allelic diversity ($h = 0.357 \pm 0.046$) as measured by MLEE among the beach isolates is within the range (0.34-0.54) reported for natural populations of *E. coli* isolated from humans and a variety of other sources (145, 174). The allelic diversity in the beach habitat is ~1.5 times greater than that found among *E. coli* isolated from the secondary habitat represented by septic tanks in two households based on comparable methods (55).

Although several studies have revealed extensive genotypic diversity in *E. coli* populations from environmental sources (27, 103), it is difficult to directly compare measures of diversity based on different molecular methods. DNA fingerprinting techniques, such as pulsed-field gel electrophoresis (PFGE) of digested genomic DNA (104) and amplicon profiles produced by repetitive element-PCR (rep-PCR) (80), have been widely applied for microbial source tracking (160). These techniques simultaneously detect many classes of genomic change including point mutations, insertions, duplications, and deletions of DNA

(89). As a consequence, DNA fingerprints can change rapidly through movement of horizontally transmitted genetic elements (17).

In contrast to DNA fingerprinting methods, sequence based techniques such as MLST, provide precise information on the nucleotide differences underlying allelic and genotypic variation (98). MLST analysis of 7 housekeeping genes revealed an average 40 alleles per locus among the environmental strains, a level of allelic diversity less than that seen previously for the *uidA* gene among hundreds of *E. coli* strains recovered from a wide variety of environmental sources and wild animal species (126). The lower average allelic diversity observed here may be a result of sampling or the effect of negative selection in the beach environment, especially as observed in two of the MLST genes (*clpX* and *lysP*). Overall, the genotypic diversity was extensive: a total of 130 distinct sequence types (STs) were resolved with 103 STs recovered only once in the sampling. Despite this diversity, several of the genotypes were repeatedly recovered at multiple sites and sampling times, suggesting that natural selection is favoring certain genotypes.

Lack of geographic differentiation among sites. The freshwater beach habitat of Lake Huron and the St. Clair River in Michigan is an open system with many possible inputs of fecal contamination and a variety of habitats and microhabitats in which *E. coli* could survive, reproduce, and locally adapt. We therefore anticipated that the *E. coli* samples recovered at different sites would be genetically distinct, reflecting both the variety of indigenous inputs as well as the local environmental pressures. In contrast to this view, however, several lines of

evidence indicate a lack of geographic subdivision of the population. First, allele frequencies are relatively uniform across sites as reflected in small between-site component of total genetic diversity ($G_{ST} < 0.02$ for MLEE and MLST). Second, similar mixtures of biotypes and genotypes (ETs, and STs) were recovered from all sites. Third, there was no significant difference in diversity estimates, based on the Chao1 estimators, from the observed and expected datasets randomized across sites. Finally, the frequency of the 4 major phylogroups was relatively uniform across samples and statistically independent of site. Based on these findings, the population genetic analyses support the hypothesis that the bacterial isolates from the different beach sites represent samples from a single, well mixed *E. coli* population.

Linkage disequilibrium and recombination in nature. One of the main observations reported here is the presence of extensive linkage disequilibrium in the natural population of *E. coli* inhabiting the beach environment. Linkage disequilibrium, the statistical situation in which two or more alleles are found together in haplotypes more frequently than expected (48), can arise and be maintained by a variety of population genetic processes. For example, a simple situation occurs in asexually reproducing organisms where genotypes evolve as distinct genetic lineages or clones within the population (174). Under these circumstances, clones that increase in frequency in the population by clonal expansions or selective sweeps (63, 85), can predominant and drive specific multilocus genotypes to high frequencies, and account for high levels of strong and persistent linkage disequilibrium. Clonal genotypes can also be over-

represented in bacterial species associated with human disease, particularly in samples of isolates from clinical cases, which may artificially or temporally generate patterns of linkage disequilibrium. For example, among strains of *Listeria monocytogenes* examined by MLEE, one ET marked a clone that was common in epidemics in the 1980s (122). Widespread clonal genotypes were also found in bacteria such as the nontyphoidal serovars of *Salmonella* (12, 143) and Shiga toxin-producing *E. coli* O157 (177, 178) for which many of the strains were obtained from clinical sources at different times and places. The linkage disequilibrium often seen in these studies could be a reflection of the biased sampling of special genotypes in clinical collections or a consequence of epidemic spread of pathogenic clones (102).

Biased sampling or epidemic spread alone, however, cannot account for the linkage disequilibrium and population genetic structure seen in the environmental *E. coli*. The study here is based on the characterization of *E. coli* isolated from a randomly sampled set of beach locations (2) with the frequency of recovery reflecting the actual abundance of genotypes in the natural habitat. This population-based sampling contrasts the situation often encountered in studies based on historical collections of clinical strains or strain collections assembled from various researchers. In such cases, linkage disequilibrium may be artificially created by biased sampling or may not be accurately estimated. Here, however, the linkage disequilibrium found by both the MLEE and MLST analyses reflects allele frequencies in the population and identified several multilocus genotypes (11 ETs and 17 STs) that were numerous and widespread among the

sampling sites. There are significant associations of alleles in environmental genotypes, as seen for example in the I_A and $rbar_d$ values. The linkage disequilibrium exists despite the clear evidence for recombination at the sequence level generating new alleles at loci and assorting existing genes into new multilocus combinations. These data provide compelling evidence for complex linkage disequilibrium persisting at multiple levels from individual phylogenetic clusters to the entire population.

Our working hypothesis is that natural selection favors certain genotypes in the environment, particularly the ECOR B1 group, which generates linkage disequilibrium despite frequent recombination. The clearest example of such an adaptive clone is ET-1, which was found at a frequency of 30 out of 185 strains, and was recovered at multiple sites in the environment over the course of the study. This clone falls into the B1 phylogroup and appears to be diversifying at the sequence level by recombination based on the MLST data. Thus, it seems likely that recombination in nature is occurring between similar lineages more frequently than between distantly related ones. This idea is consistent with the findings of Wirth et *al.* who show evidence for extensive recombination between strains of the 4 ECOR groups, particularly between members of groups A and B1 (182).

Based on eBURST analysis, we estimate that recombination generates new alleles 5 times faster than mutation (R:M ratio of 5 to1), which is less than the theoretical threshold of between 10 and 20 recombination to mutation events required to keep a natural population at linkage equilibrium (102). Assuming that

recombination exchanges fragments that are approximately the size of the loci we sequenced (~500 bp), we can also express an R:M ratio in terms of generating new nucleotide sites. This is accomplished by first expressing recombination and mutation in terms of changes per site. For recombination, we count up the number of pair wise nucleotide differences in recombinant alleles between members of clonal groups and divide by the length of the sequence. Likewise for mutation, it is a single base pair change divided by the length of the locus. The per site R:M ratio is then the ratio of these two numbers. If each recombinant allele differed from the alleles of other clonal group members by a single nucleotide, then the R:M ratio would approach unity. On the other hand, if recombinant alleles differed by more than one nucleotide, then the influence of recombination would be greater than mutation and the ratio would increase. For the clonal groups found in the freshwater beach habitat, recombination is generating novel nucleotide sites 8 times faster than mutation. While it appears that recombination is effective at creating genetic diversity on a per site basis, this value is less than the estimated ~50:1 ratio reported for within the ECOR A group and similar to the 5:1 estimate for all ECOR groups (62).

Accuracy of the phylogrouping PCR technique. Clermont et *al.* originally described a decision tree based on a triplex PCR that groups strains into one of the four phylogenetic lineages of the *E. coli* reference (ECOR) collection (33). This technique has been applied extensively to characterize diverse collections of strains including commensal and pathogenic *E. coli* from humans (133) and a variety of animal hosts (32, 56, 151). We have shown here that the

false positive rate across the four groups is ~10%, which is one misclassification per 10 strains analyzed. It is noteworthy that the distribution of false positives is not even across the groups. For example, all group D strains clustered together and thus appear to be correctly classified, while the false positive rate for group A strains was 33%. These findings suggest that the diagnostic absence of a PCR amplicon, presumably resulting from gene deletion or mutation within the PCR primer sites, leads to the misclassification of group A strains. It is clear that a more discriminatory PCR assay is needed to distinguish strains of group A.

The ET-1 clade. There are few studies of secondary habitat *E. coli* populations with which to compare the frequency of the B1 phylogroup and, indirectly, the importance of the ET-1 clade. However, numerous studies have used ECOR phylogrouping to assess the genetic composition of strains from different hosts and host populations (44, 45, 56, 58, 111-113, 135). Strains of group B1 were infrequently recovered from various human populations (45, 58, 113), and were the most frequently sampled from certain types of wild animals (56). For example, Gordon et *al.* reported that B1 strains were most frequently isolated from ectotherms (fish, frogs, and reptiles) living in or near water, suggesting that the occurrence of B1 strains in these hosts was associated with the abundance of B1 strains in the secondary habitat (56). Bacteria of the B1 phylogroup were also identified in the characterization of natural *E. coli* bloom events in an Australian freshwater lake where fecal contamination was unlikely to account for abundance (123). Finally, 70% of strains from river and surface water around Munich, Germany were classified as either phylogroup B1 or A, based on

the presence of the *chuA* gene and the Clermont et al. PCR primers (69). These phylogrouping studies are in agreement with our hypothesis that the ET-1 clade represents a closely related group of secondary habitat adapted *E. coli*. If our interpretation of the ET-1 clade is correct, then approximately 23% of strains are adapted to the freshwater beach environment. In addition, the utility of these strains for use in water quality assessment or microbial source tracking will be suspect.

Conclusions. Colonization and persistence of certain *E. coli* genotypes in humans and other animals have been well documented (28, 79). When excreted outside the host, the average strain, however, probably does not persist and dies off at a rapid rate (172, 181). This results in a net-negative growth rate under some secondary habitat conditions (137). Consequently, adaptation to primary habitat stimuli is considered to dominate evolutionary processes in the *E. coli* species as a whole (137). However, certain groups of *E. coli* have been shown to colonize and persist autochthonously (without fecal input) in secondary habitats (121, 123, 131), and recent data suggest that certain genotypes are adapted to these conditions (4, 55, 103, 173). Ishii and colleagues (77) have proposed the term "naturalized" strains to refer the persistent *E. coli* genotypes that comprise the autochthonous members of the microbial community in the environment. The findings reported here add support to the emerging model that naturalized *E. coli* strains are a significant component of the environmental coliform microbiota. In addition, the persistent genotypes are concentrated in a distinct clade, the B1

phylogroup, whose members may possess special traits that allow them to survive and reproduce in the environment under temperate conditions.

ACKNOWLEDGMENTS

We would like to acknowledge those who participated in this study from Central Michigan University: Janice Burke, Erica Francis, Erin Hagan, AJ Matthews, Anne Spain, and Erin Zacharias. This work was supported in part by MRU matching funds from the College of Veterinary Medicine and Graduate College at Michigan State University.

CHAPTER 5

A REVISED MOLECULAR PHYLOGENY FOR *ESCHERICHIA COLI*

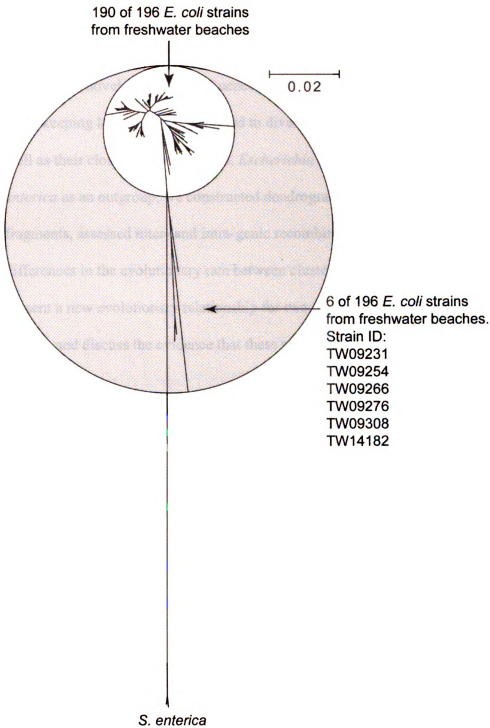
INTRODUCTION

The species *Escherichia coli* belongs to the bacterial family, *Enterobacteriaceae*, and is known to circulate within and between humans and animals (21). The host gastrointestinal tract is referred to as the primary habitat for *E. coli*. However, it is estimated that half of all strains are presently outside of hosts in the secondary habitat, or the environment (137). The consensus view is that *E. coli* cannot survive in the environment (21, 181), although recent evidence suggests that autochthonous strains do exist (2, 4, 16, 54, 92, 123, 156, 171). There are few data with which to assess the natural history of *E. coli* in the secondary habitat because most studies are concerned with commensal strains from the primary habitat (28, 45, 53, 55, 56, 58, 145, 146) or strains that cause disease (128, 148, 175, 179, 180).

In a recent population genetic characterization of strains from a secondary habitat (169), we discovered 6 strains that were highly divergent in nucleotide sequence using multilocus sequence typing (Figure 8). These strains were originally isolated from freshwater beaches along Lake Huron and the St. Clair River in Michigan and were phenotypically identified as *E. coli* using standard biochemical tests (2, 169). It appears that such strains represent novel genetic diversity that may be used to make evolutionary inferences about the species as a whole. For example, it has been hypothesized that a similarly divergent lineage of *E. coli* represents the breadth species-level diversity that existed some 10 – 15 million years ago (182). An alternate or additional hypothesis is that multiple lineages of

Figure 8. Divergent *E. coli* strains isolated from freshwater beaches in Michigan. Neighbor-joining tree is based on concatenated sequences from 7 housekeeping loci (*aspC*, *clpX*, *fadD*, *icdA*, *lysP*, *mdh*, and *uidA*). The white circle contains represents the typical genetic diversity of *E. coli* and the grey circle represents that of the divergent strains.

Figure 8.


190 of 196 E. coli strains
from freshwater beaches
0.02
6 of 196 E. coli strains
from freshwater beaches.
Strain ID:
TW09231
TW09254
TW09266
TW09276
TW09308
TW14182
S. enterica

biochemically similar, but genetically distinct bacteria have evolved, so that multiple species currently exist within what is now considered to be *E. coli*.

The purpose of this study was to assemble a collection of genetically divergent *E. coli* and present a novel phylogenetic context for these strains. We developed a novel multilocus sequence typing (MLST) protocol based on 24 housekeeping loci that can be applied to divergent and typical strains of *E. coli* as well as their closest known relative, *Escherichia albertii*. Using *Salmonella enterica* as an outgroup, we constructed dendrograms based on individual gene fragments, assessed inter- and intra-genic recombination, and tested for differences in the evolutionary rate between clusters of strains. Finally, we present a new evolutionary relationship for two distinct bacterial lineages within *E. coli*, and discuss the evidence that these strains are prevalent in the secondary habitat.

EXPERIMENTAL PROCEDURES

Strain collection and sequence source. All *E. coli* and *E. albertii* strains included in this study (Table 12) are part of a strain repository maintained at Michigan State University. More information on each isolate can be found online at www.shigatox.net. The *E. coli* strains were assembled from 5 sources; 1). Six strains (TW09231, TW09254, TW09266, TW09276, TW09308, and TW14182) were isolated between 2001 and 2003 in Michigan from freshwater beaches along Lake Huron and the St. Clair River as part of a previously described study (2); 2). One strain (TW11588) was isolated from soil in the Puerto Rican rainforest by Dr. Gary A. Toranzos (Department of Biology, University of Puerto Rico – Rio Piedras). This strain was sampled as part of a collection of strains that are unassociated with human activity. It was isolated on February 26, 2003 at a depth of 6-10 cm; 3). Five strains (TW14263, TW14264, TW14265, TW14266, and TW14267) were isolated by Jeffrey L. Ram (Department of Physiology, Wayne State University) in 2005 as part of a source tracking experiment in the Great Lakes region. Four strains were isolated from the environment (water), and one was isolated from a raccoon; 4). Two previously published (182) strains (TW14351 and TW14352) were obtained from Mark Achtman (Department of Molecular Biology, The Max Planck Institute) and were part of a previously published study (182). One strain was isolated from a parrot and one strain was isolated from a dog; 5). Four strains (TW11930, TW11966, TW12018, and TW14421) were included from an ongoing cohort study of enterotoxigenic *E. coli*

(ETEC) infection in infants from Africa and were supplied by Hans Steinsland (Microbial

Table 12. *E. coli*, *E. albertii*, and *S. enterica* strains information.

Species	Strain ID	Habitat	Locale	Strain source[a]	Reference
E. coli	K-12 (MG1655)	Human	No data	N/A	Blattner et *al.* 1997
	S. flexneri 2457T	Human	Japan	N/A	Wei et *al.* 2003
	EHEC (EDL933)	Human	MI, USA	N/A	Perna et *al.* 2001
	EAEC (042)	Human	Lima, Peru	N/A	Nataro et *al.* 1995
	EPEC (e2348/69)	Human	England	N/A	Cravioto et *al.* 1979
	UPEC (CFT073)	Human	MD, USA	N/A	Welch et *al.* 2002
	RL325/96	Dog	No data	M. Achtman	Wirth et *al.* 2006
	Z205	Parrot	No data	M. Achtman	Wirth et *al.* 2006
	TW09231	Water	MI, USA	E. W. Alm	This study
	TW09254	Water	MI, USA	E. W. Alm	This study
	TW09266	Water	MI, USA	E. W. Alm	This study
	TW09276	Water	MI, USA	E. W. Alm	This study
	TW09308	Water	MI, USA	E. W. Alm	This study
	TW11588	Soil	Puerto Rico	G. A. Toranzos-Soria	This study
	TW11930	Human	Guinea Bissau	H. Steinsland	This study
	TW11966	Human	Guinea Bissau	H. Steinsland	This study
	TW12018	Human	Guinea Bissau	H. Steinsland	This study
	TW14182	Water	MI, USA	E. W. Alm	This study
	TW14263	Raccoon	MI, USA	J. L. Ram	This study
	TW14264	Water	MI, USA	J. L. Ram	This study
	TW14265	Water	MI, USA	J. L. Ram	This study
	TW14266	Water	MI, USA	J. L. Ram	This study
	TW14267	Water	MI, USA	J. L. Ram	This study
	TW14421	Human	Guinea Bissau	H. Steinsland	This study

Table 12 (continued).

Species	Strain ID	Habitat	Locale	Strain source[a]	Reference
E. albertii	9194	Human	Bangladesh	J. M. Janda	Hyma et *al.* 2005
	19982	Human	Bangladesh	J. M. Janda	Hyma et *al.* 2005
	C-425	Human	No data	ATCC 12032	Hyma et *al.* 2005
	K-1	Human	Bangladesh	K. A. Talukder	Hyma et *al.* 2005
S. enterica	Typhi TY2	Human	No data	N/A	Deng et *al.* 2003
	Typhimurium LT2	Human	No data	N/A	McClelland et *al.* 2001

[a]N/A indicates online genome data for *coli*BASE.

Evolution Laboratory, Michigan State University). The *E. albertii* strains (9194, 19982, C-425, and K-1) all came from Bangladesh and were part of a previously published study (75). Type strains for *E. coli* and *S. enterica* (K-12 MG1655, *Shigella flexneri* 2457T, EDL933, 042, e2348/69, CFT073, *S. enterica* serovar Typhi TY2, and *S. enterica* serovar Typhimurium LT2) were analyzed *in silico* by downloading sequences from the comparative genomics database, *coli*BASE (colibase.bham.ac.uk).

Confirmation of *E. coli*. The strains in this study satisfied 4 biochemical criteria. 1). A phenotypic profile was generated for each strain using 20 reactions of the API20E bacterial identification system (BioMerieux, Inc.). Positive (1) or negative (0) results were recorded for the following reactions: beta-galactosidase (ONPG), arginine dihydrolase (ADH), lysine decarboxylase (LDC), ornithine decarboxylase (ODC), citrate utilization (CIT), H2S production (H2S), urease (URE), tryptophane deaminase (TDA), indole production (IND), acetoin production (VP), gelatinase (GEL), and fermentation/oxidation of glucose (GLU), mannitol (MAN), inositol (INO), soribitol (SOR), rhamnose (RHA), sucrose (SAC), melibiose (MEL), amygdalin (AMY), and arabinose (ARA). 2). Strains were screened for cytochrome c oxidase activity using a filter paper test. Single colonies from overnight growth on LB agar plates were streaked onto filter paper containing the reagent *N,N,N',N'*-Tetramethyl-*p*-phenylenediamine (Oxidase test kit, Becton-Dickinson and Company). 3). Strains were tested for *β*-glucuronidase activity by streaking single colonies onto Luria-Bertani (LB) agar plates containing the flourogenic compound 4-methylumbelliferyl-*β*-D-glucuronide (MUG, Sigma Aldrich). 4). Strains were tested

for glutamate decarboxylase (GAD) activity using the protocol described by Rice et *al.* (130). All strains were positively identified as *E. coli* by API20E, negative for oxidase activity, positive for MUG activity, and positive for GAD activity.

DNA isolation and multilocus sequence typing (MLST). Genomic DNA was isolated from 2 ml of overnight culture in LB broth, Lennox (Becton, Dickinson, and Company, Sparks, MD) using a Puregene DNA isolation kit (Gentra Systems Inc., Minneapolis, MN.). DNA preparations were quantified with a NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE.), diluted to a final concentration of 100 ng/μL, and stored at 4°C.

Sequencing methodology used for MLST was carried out as part of a system described in detail elsewhere (www.shigatox.net/cgi-bin/mlst7/index). Briefly, we analyzed the internal fragments of 24 housekeeping genes (see Table 13). We selected 20 genes from two publicly available protocols (www.shigatox.net (124) and www.mlst.net), 1 gene (*metG*) from a previous *E. coli* MLST study (170), and 3 genes (*kdsA*, *torC*, and *yjdB*) from a recent analysis of *E. coli* MLST gene candidates (87). The program SeqMan II (DNASTAR, Inc., Madison, WI) was used to edit and align the sequences.

Descriptive statistics, Tajima's test, and recombination analysis. Sequences were concatenated and uploaded into MEGA3 software (91). Descriptive statistics for MLST data including the number of variable sites, parsimoniously informative nucleotide sites, and estimates of d_S and d_N were done using MEGA3.1 (modified Nei-Gojobori model; Jukes-Cantor). Tests for

Table 13. *E. coli* MLST primers for 24 housekeeping loci. The primers listed here were used to amplify typical and divergent *E. coli* and *E. albertii* strains.

Locus	Size (bp)	Primer sequences (5'-3')[a]	position from start of locus[b]	Size of product	Annealing temperature	Primer reference[c]
adk	645	F – ATTCTGCTTGGCGCTCCGGG	10	584	54	Achtman
		R – CCGTCAACTTTCGCGTATTT	594			Achtman
arcA	717	F – GACAGATGGCGCGGAAATGC	99	552	58	EcMLST
		R – TCCGGCGTAGATTCGAAATG	651			EcMLST
aroE	819	F – GGGGGCGTTTAAATCCTTCA	-69[b]	759	58	This study
		R – GCCTCGCTGCTCACACCA	690			This study
aspC	1191	F – GTTTCGTGCCGATGAACGTC	57	594	58	EcMLST
		R – AAACCCTGGTAAGCGAAGTC	651			EcMLST
clpX	1275	F – CTGGCGGTCGCGGTATACAA	262	672	58	EcMLST
		R – GACAACCGGCAGACGACCAA	934			EcMLST

Table 13 (continued).

Locus	Size (bp)	Primer sequences (5'-3')[a]	position from start of locus[b]	Size of product	Annealing temperature	Primer reference[c]
cyaA	2547	F – CTCGTCCGTAGGGCAAAGTT	312	571	58	EcMLST
		R – AATCTCGCCGTCGTGCAAAC	883			EcMLST
dnaG	1746	F – CGCTGAACCCAATCGTCT	765	696	58	This study
		R – TCTCTGAATAAGCCAAGTCCA	1461			This study
fadD	1686	F – GCTGCCGCTGTATCACATTT	768	580	58	EcMLST
		R – GCGCAGGAATCCTTCTTCAT	1348			EcMLST
fumC	1404	F – TCACAGGTCGCCAGCGCTTC[a]	10	806	54	Achtman
		R – GTACGCAGCGAAAAAGATTC[a]	816			Achtman
grpE	594	F – CCCGGAAGAAATTATCATGG	39	488	58	EcMLST
		R – TCTGCATAATGCCCAGTACG	527			EcMLST

Table 13 (continued).

Locus	Size (bp)	Primer sequences (5'-3')[a]	position from start of locus[b]	Size of product	Annealing temperature	Primer reference[c]
gyrB	2415	F – GACGGGCGCGGCATTCC	220	689	58	This study
		R – CTGTAGCCTTCTTTGTCCA	909			This study
icdA	1251	F1 – GCAACGTGGTGGCAGAC	-49[b]	541	58	This study
		R1 – TTCGATACCCGCATAAAT	492			This study
		F2 – CTGCGCCAGGAACTGGATCT	371	669	58	EcMLST
		R2 – ACCGTGGGTGGCTTCAAACA	1020			EcMLST
kdsA	855	F – AAAAAGTGGTTAGCATTGG	8	502	58	This study
		R – GCACCGCGATCGCAAAGAAT	510			This study
lysP	1470	F – CTTACGCCGTGAATTAAAGG	36	628	58	EcMLST
		R – GGTTCCCTGGAAAGAGAAGC	664			EcMLST

Table 13 (continued).

Locus	Size (bp)	Primer sequences (5'-3')[a]	position from start of locus[b]	Size of product	Annealing temperature	Primer reference[c]
mdh	939	F – GTCGATCTGAGCCATATCCCTAC	130	650	58	EcMLST
		R – TACTGACCGTCGCCTTCAAC	780			EcMLST
metG	2034	F – CACATCCAGGCTGATGTCTG	85	573	58	Johnson
		R – CATTTTATTTGCCACCTGCTC	658			This study
mtlD	1149	F – GCAGGTAATATCGGTCGTGG	22	658	58	EcMLST
		R – CGAGGTACGCGGTTATAGCAT	680			EcMLST
mutS	2562	F – GGCCTATACCCTGAACTACA	1683	596	58	EcMLST
		R – GCATAAAGGCAATGGTGTC	2279			EcMLST
purA	1299	F – CGCGCTGATGAAAGAGATGA	234	817	54	Achtman
		R – CATACGGTAAGCCACGCAGA	1051			Achtman

Table 13 (continued).

Locus	Size (bp)	Primer sequences (5'-3')[a]	position from start of locus[b]	Size of product	Annealing temperature	Primer reference[c]
recA	1062	F – CGCATTCGCTTTACCCTGACC[a]	185	734	58	Achtman
		R – TCGTCGAAATCTACGGACCGGA[a]	919			Achtman
rpoS	993	F – CGCCGGATGATCGAGAGTAA	274	618	58	EcMLST
		R – GAGGCCAATTTCACGACCTA	892			EcMLST
torC	1173	F – TGAATGGGCGCGAATGAAAGA	375	630	58	This study
		R – GCGCCGTGGCACTGGTTACA	1005			This study
uidA	1812	F – CATTACGGCAAAGTGTGGGTCAAT	277	728	58	EcMLST
		R – TCAGCGTAAGGGTAATGCGAGGTA	1005			This study
yjdB	1674	F – CATTCAGCGAGCGGGCATCA	966	667	58	This study
		R – AGCCTGGTAATACTTCGTCTCAAC	1633			This study

Table 13 (continued).

[a]These primers are listed here as they appear in the reference, but they appear to be in the opposite orientation.

[b]The forward primer for *aroE* is located upstream in *yrdC*. The forward primer for *icdA* is also located upstream, but in the intergenic region between *ymfC* and *icdA*. Negative values indicate upstream position relative to the start of the locus.

[c]Achtman and EcMLST refer to publicly available online MLST databases at www.mlst.net and www.shigatox.net (124), respectively. The Johnson reference can be found at Weissman et *al*. 2006.

unequal rates of molecular evolution were performed using the Tajima's test in Mega3.1. This test has been used elsewhere (43) to display differences in the rate of evolution compared to the null hypothesis that neutral mutations accrue through time at a uniform rate (molecular clock).

The approach we used to assess the effects of recombination was called Genetic Algorithm for Recombination Detection (GARD) (88). This analysis screens for evidence of phylogenetic incongruence and identifies the number and location of breakpoints in aligned sequences. There are basically 3 parts to the analysis: 1). Upload a file containing aligned sequences, 2). Select an appropriate model of nucleotide evolution, and 3). Run the analysis. GARD is publicly available and the analysis can be executed on clusters of computers through the website www.datamonkey.org.

RESULTS

New primers for MLST of 24 housekeeping loci. The strains included in this analysis were assembled to represent the genetic diversity between *E. coli* and *E. albertii* (Table 12). To maximize efficiency and cost, we developed an MLST protocol that utilizes one PCR primer pair per locus for PCR amplification and sequencing. Published primers for *E. coli* MLST analysis were initially used for PCR amplification and resulted in adequate sequence for 16 of 21 (76%) loci (Table 13). New primer pairs were designed and tested for the 5 loci that failed to work under published conditions. In addition, 3 other primer pairs were designed for candidate *E. coli* MLST loci (87). These new loci were selected to maximize sequence coverage around the *E. coli* chromosome (Figure 9). Primers were optimized so that annealing occurred at the same temperature (58°C), which meant that multiple loci could be amplified in the same PCR themocycling reaction. As expected, the *uidA* locus was absent in *E. albertii* and *S. enterica* strains, but was present in all *E. coli* lineages.

MLST analysis. Sequencing of 24 housekeeping loci resulted in 12,191 bp of nucleotide sequence, encoding for approximately 3,861 codons. There were 2,703 variable sites, of which 2,453 were parsimoniously informative (occurred in more than one strain). Pairwise strain comparisons of the concatenated sequence resulted in clusters of five well-supported groups in a neighbor-joining (NJ) dendrogram (Figure 10A). As expected, all strains of defined taxa (*S. enterica*, *E. albertii*, and *E. coli*) clustered together. However, two additional groups of strains were found. We labeled them as environmental group I and environmental

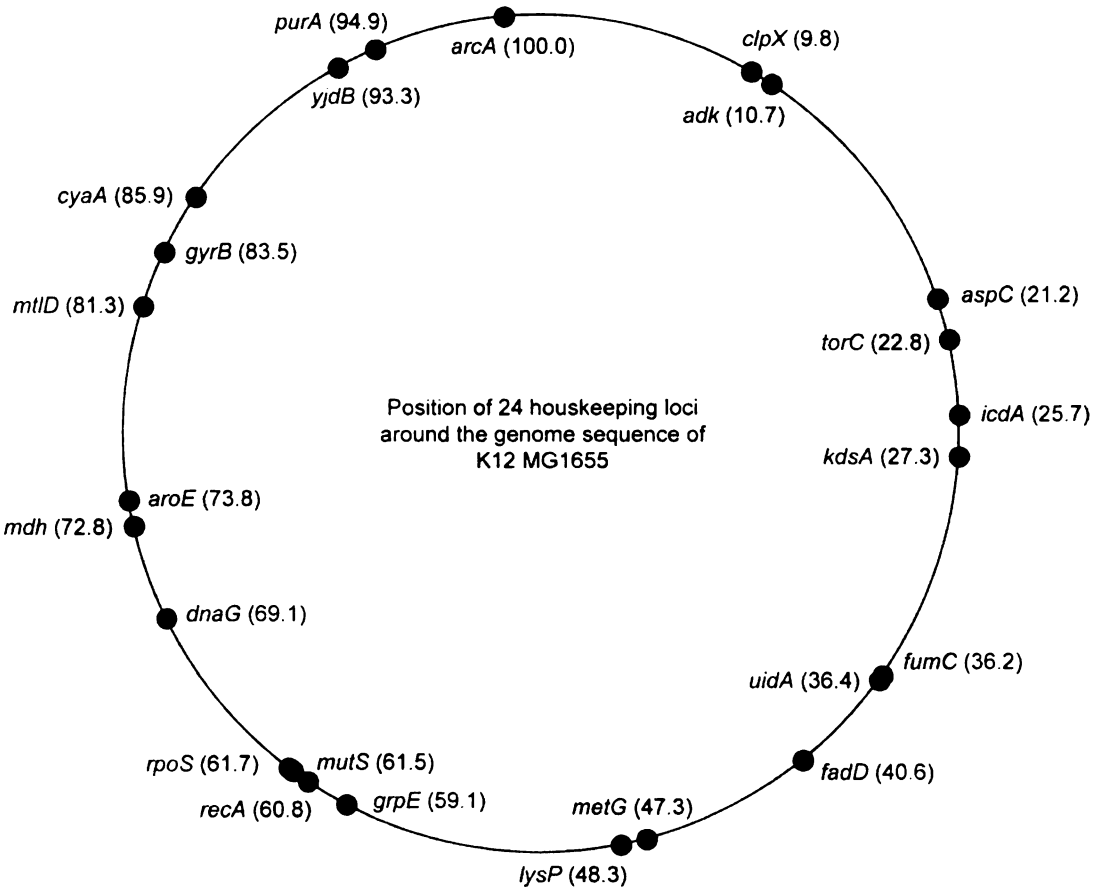
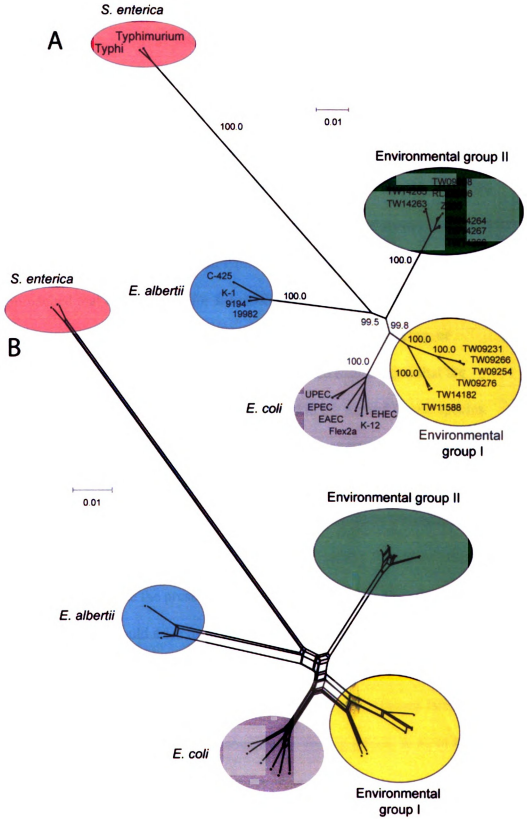


Figure 9. Position of 24 housekeeping loci around the *E. coli* K-12 MG1655 chromosome.

Figure 10. Neighbor-joining tree (panel A) and Neighbor-Net (panel B) of *E. coli*, Environmental group I, Environmental group II, *E. albertii*, and *S. enterica* based on 24 housekeeping loci (see Table 13).

Figure 10.

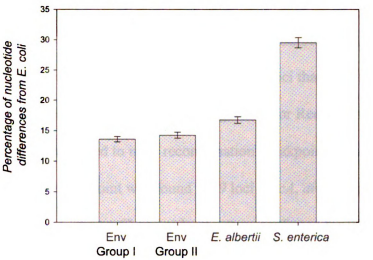

A
S. enterica
Typhimurium
Typhi
100.0
0.01
Environmental group II
TW14263
100.0
S. enterica
B
E. albertii
C-425
K-1
9194
19982
100.0
99.5
99.8
100.0
100.0
TW09231
TW09266
TW09254
TW09276
TW14182
TW11588
100.0
E. coli
UPEC
EPEC
EAEC
Flex2a
EHEC
K-12
Environmental group I
0.01
Environmental group II
E. albertii
E. coli
Environmental group I

group II because they contained the 6 divergent freshwater beach strains from Michigan (Figure 8).

The environmental groups also contained 4 additional strains from Michigan water and one strain from pristine (non-human associated) soil in a remote area of a Puerto Rican jungle. In addition, two previously reported (182) *E. coli* strains (one from a dog and one from a parrot) and one strain from a raccoon in Michigan clustered with environmental group II strains. Pairwise group comparisons revealed that both of the environmental groups had a similar level of sequence divergence from typical *E. coli*. The extent of this divergence approached that of *E. albertii* for silent, or synonymous, substitutions (Figure 11A) and was identical to that of *E. albertii* for amino acid altering, or nonsynonymous, substitutions (Figure 11B). These data suggest that the environmental groups may be distinct taxa within the named *E. coli* species.

Recombination analysis. To visualize the overall contribution of recombination between these groups, we constructed a phylogenetic network (Neighbor-Net) (Figure 10B) using SplitsTree analysis (74). This analysis does not force sequence data into a bifurcating tree and allows for numerous parallel paths, which indicate the presence of phylogenetic incompatibilities. Such incompatibilities could arise from recombination or recurrent mutation in the loci. It is interesting that despite the influence of recombination, the five groups defined in the NJ tree remain distinct and relatively intact (Figure 10B). It also appears that the effect of recombination varies among groups as more alternate

A. Synonymous substitutions



B. Nonsynonymous substitutions

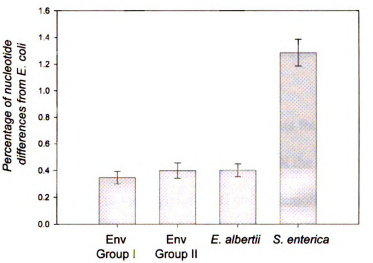


Figure 11. Pairwise sequence divergence between *E. coli* and environmental group I, environmental group II, *E. albertii*, and *S. enterica*. Environmental groups appear to be as divergent from *E. coli* as *E. albertii* at synonymous (panel A) and nonsynonymous (panel B) sites.

paths can be seen within environmental group I and between this group and *E. coli*.

To further assess the influence of recombination, we characterized the number of complete loci and fragments of loci that were either shared or clustered with divergent groups. Genetic Algorithm for Recombination Detection (88), or GARD, was used to infer recombination breakpoints within the 24 loci (Table 14). No breakpoint was found for 9 loci (*arcA*, *aroE*, *aspC*, *dnaG*, *fumC*, *kdsA*, *mdh*, *purA*, and *torC*) and at least one was inferred for the remaining 15, resulting in 39 non-recombinant fragments. For each fragment, we constructed a neighbor-joining dendrogram to assess the extent of recombination between groups. Evidence for recombination was found if a non-monophyletic relationship was found for the groups (strains of each group did not cluster together on a single branch).

For the 9 loci where no breakpoint was found, 4 (44%) produced monophyletic relationships for all groups. Of the remaining 30 fragments, only 7 (23%) produced monophyletic relationships among the groups, suggesting that recombination of gene fragments (intra-genic) occurred more often than the transfer of entire loci. Three groups were monophyletic at all or nearly all fragments. The *S. enterica* group was always monophyletic (39 out of 39 fragments); strains of *E. albertii* clustered together with 36 of the 39 (92%) fragments; and environmental group II was monophyletic for 35 of the 39 (90%) fragments. In contrast, strains of environmental group I and *E. coli* were rarely monophyletic (18 of 39 and 15 of 39, respectively). These data suggest that

Table 14. Results of single locus analyses. Recombination (intra-genic) breakpoints were identified using GARD analysis. Inter-genic recombination was defined by loci with no breakpoints that do not exhibit a monophyletic relationship.

GARD fragment	Locus	Position of breakpoint	Significant Tajima's test	Monophyletic relationship[a]
1	*arcA*	None	No	Yes
2	*aroE*	None	No	No
3	*aspC*	None	No	No
4	*dnaG*	None	No	No
5	*fumC*	None	No	Yes
6	*kdsA*	None	No	Yes
7	*mdh*	None	No	Yes
8	*purA*	None	Yes	No
9	*torC*	None	No	No
10	*adk*	1-394	No	No
11	*adk*	395-536	No	No
12	*clpX*	1-240	No	No
13	*clpX*	241-567	No	Yes
14	*cyaA*	1-294	No	No
15	*cyaA*	295-498	No	No
16	*fadD*	1-186	No	No
17	*fadD*	187-492	No	No
18	*grpE*	1-277	No	No
19	*grpE*	278-417	No	No
20	*gyrB*	1-270	Yes	No
21	*gyrB*	271-460	No	No
22	*icdA*	1-472	No	No
23	*icdA*	473-826	No	No
24	*lysP*	1-223	Yes	Yes
25	*lysP*	224-477	No	Yes
26	*metG*	1-193	No	No
27	*metG*	194-406	No	No
28	*metG*	407-588	No	No
29	*mtlD*	1-363	Yes	Yes
30	*mtlD*	364-540	Yes	Yes
31	*mutS*	1-240	Yes	No
32	*mutS*	241-393	No	No
33	*mutS*	394-507	No	No
34	*recA*	1-181	No	No

Table 14 (continued).

GARD fragment	Locus	Position of breakpoint	Significant Tajima's test	Monophyletic relationship[a]
35	*recA*	182-510	Yes	No
36	*rpoS*	1-360	No	No
37	*rpoS*	361-585	No	Yes
38	*yjdB*	1-145	No	No
39	*yjdB*	146-430	No	Yes

[a]Monophyletic relationship indicates that strains of all five groups labeled in.

environmental group I is as recombinationally isolated from *E. coli* as it is from *E. albertii*, while environmental group II and *E. coli* are sharing loci much more frequently.

Tajima's test for group rates of evolution. Here, we wanted to test the hypothesis that groups were not evolving at a similar rate (that of the molecular clock). We used Tajima's test (χ^2 test, df = 1, $p < 0.050$) to examine the nucleotide sequence from a representative strain of each group. The analysis was done on the third codon positions only, to minimize effects of selection. For each fragment defined by the GARD analysis, we conducted pairwise tests between *E. coli* strain K-12 and one strain from *E. albertii* (strain 9194), environmental group I (strain TW09231), and environmental group II (TW09308) using *S. enterica* (strain Typhimurium LT2) as an outgroup. Of the 39 fragments, 7 (18%) showed unequal rates of evolution (Table 14). The rate of the environmental group II lineage was unequal to *E. coli* at all 7 fragments. It had more unique differences than K-12 (accelerated rate of evolution) at 3 fragments and fewer unique differences than K-12 (slower rate of evolution) at the other 4 fragments. According to 2 fragments, environmental group I evolved slower than *E. coli*, and only one fragment suggested that *E. albertii* experienced accelerated evolution compared to *E. coli*. These data suggest a fairly uniform rate of evolution for these groups.

Phylogenetic construction and divergence time. In order to construct the most robust phylogeny possible, we used only the loci that satisfied two criteria: 1) they show no evidence of between-group recombination

(monophyletic relationships) and 2) they must satisfy the molecular clock hypothesis of uniform neutral substitution (non-significant Tajima's test). Of the total 39 GARD fragments, 8 satisfied both criteria (fragments 1, 5, 6, 7, 13, 25, 37, and 39) and were used to generate another phylogenetic network (Figure 12). Adherence to these 2 criteria removed most of the phylogenetic incompatibility between the groups. This analysis clearly shows that *E. coli* is more related to the environmental groups than to *E. albertii*. In addition, the data suggest that *E. coli*, environmental group I, and environmental group II diverged from one another in a short period of time. This divergence occurred at such a rate that it is not clear which of these groups is ancestral. This observation is in contrast to the relationship inferred by the NJ tree (Figure 10A), where it appears that environmental group II is ancestral to environmental group I and *E. coli*.

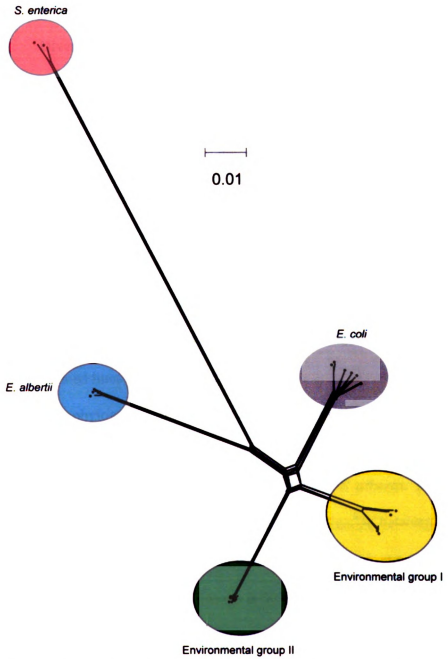


Figure 12. Neighbor-Net of *E. coli*, Environmental group I, Environmental group II, *E. albertii*, and *S. enterica* based on 8 GARD fragments (fragments 1, 5, 6, 7, 13, 25, 37, and 39). These fragments were selected based on criteria presented in Table 14 (a monophyletic relationship for the groups and non-significant Tajima's Test).

DISCUSSION

Divergent bacterial lineages within *E. coli*. All strains analyzed in this study were confirmed by standard biochemical tests. However, some strains are remarkably divergent in their nucleotide sequence when compared to typical *E. coli* strains that have been completely sequenced. The analysis presented here shows that divergent *E. coli* strains represent two distinct evolutionary lineages. Based on synonymous and nonsynonymous nucleotide substitutions in conserved housekeeping genes, we have shown that one of these lineages, environmental group II, is as divergent from *E. coli* as *E. coli* is from its recent ancestor *E. albertii*. In addition, this group appears to be recombinationally isolated from closely related lineages (*E. albertii* and *E. coli*).

The phylogenetic relationship between environmental group I and *E. coli* is not as clear. Based on nucleotide substitutions, environmental group I is as divergent from *E. coli* as environmental group II and *E. albertii*. However, there does not appear to be a similar limitation on recombination between *E. coli* and environmental group I as there is between *E. coli* and other groups. This suggests that the environmental group I lineage is diverging away from *E. coli*, but has not yet accrued enough genetic variation to be recombinationally isolated.

An interesting hypothesis generated by our analysis is that the environmental groups are actually unique species. Two "gold standards" for describing a bacterial species are similarity based on DNA-DNA hybridization and a molecular characterization of 16S rRNA gene (159). While it is not the purpose of this study to propose a formal new species designation, we feel that it

is noteworthy that we have sequenced 1,217 bp of the 16S rRNA gene in all strains. Based on preliminary analyses, the 16S rRNA gene for *E. coli*, environmental group I, environmental group II, and *E. albertii* are no more than 98.7% similar. We recommend that lineages with this degree of similarity in the 16S rRNA gene undergo further analysis by DNA-DNA hybridization (159). Regardless, these data further support our interpretation that environmental groups I and II are as divergent from *E. coli* as *E. coli* is from its close relative *E. albertii*.

The natural habitat of divergent *E. coli* strains. The current understanding of the evolutionary ecology of *E. coli* appears to be biased by analyzing strains from hosts. While the first divergent lineage of *E. coli* to be described came from host samples (182), these strains (2 strains of the same genotype) were rare in the overall collection of 462 strains. While it is enticing to describe this observation as a frequency (2/462 or 0.4%), it is worth noting that the collection was not a random sample of strains from hosts. For example, included in this collection were 72 strains from the *E. coli* reference collection (ECOR) that represent much of the known diversity of the species (114). These strains were selected from 2,600 strains that, among other criteria, represent the breadth of genetic diversity across host species and geographic distribution. In addition to the ECOR strains, the collection also included 15 strains representing the breadth of genetic diversity found in a collection of 1,844 strains from humans and their septic tanks (55). Based then on these numbers, a more accurate estimate of how frequent this divergent *E. coli* lineage is in humans and animals is

on the order of 2 in 4,819 or 0.04%. However, this value is still inflated with respect to other population genetic characterizations of *E. coli* isolated from human and animal hosts (28, 53, 55, 56, 58, 180).

The most diverse *E. coli* collection to be analyzed by similar population genetic methods was perhaps a worldwide sampling of 202 strains from mammals and birds (157). ECOR strains were included in the analysis and based on their position in a NJ dendrogram it appears that this collection still does not account for the divergent lineages discussed here, although some strains do cluster outside the diversity of ECOR. Regardless, we observed 2 divergent lineages among 6 strains of a randomized collection (196 strains) from freshwater beaches in Michigan (169). Although it has not been addressed directly, both the literature and our recent observations (169) suggest that divergent *E. coli* strains are rare in humans and animals, but abundant in freshwater beaches in Michigan (~0.04% vs. ~3%).

Further evidence that divergent *E. coli* strains are rare in human hosts comes from an ongoing analysis of a collection of 715 enterotoxigenic *E. coli* (ETEC) strains. This collection comes from a cohort sampling of infants from Guinea Bissau in Africa (Table 12). Based on MLST analysis, 3 strains in this collection (TW11930, TW12018, and TW14421) are indistinguishable at 7 housekeeping loci from 3 strains of environmental group I (TW09231, TW09266, and TW09254). A fourth strain (TW11966) was also identified as environmental group I based on clustering in a NJ tree. These 4 pathogens carry certain plasmid encoded virulence factors that are common among ETEC. However,

environmental group I is not strictly an ETEC group because the strains analyzed in this study do not carry the common virulence factors. If this collection is representative of the normal *E. coli* flora in humans, which is questionable, then the frequency of divergent strains is 4 in 715 or 0.6%, an estimate that is still an order of magnitude less than that observed in freshwater beaches. Our interpretation of these data is that divergent *E. coli* lineages have an adaptive advantage in certain environments outside the host where they are found more frequently.

Conclusions. Divergent bacterial lineages exist within the biochemically-defined species, *E. coli*, and they may represent novel species. Similarly divergent strains are rarely sampled from human and animal hosts. This difference in abundance may be a result of adaptive evolution because they appear to be much more prevalent in the secondary habitat. Finally, the divergent lineages presented here represent excellent opportunities for comparative genomic studies, as they can be used to test the rates of evolutionary processes, such as the limits of gene flow, between closely related, but distinct bacterial taxa.

ACKNOWLEDGMENTS

We would like to acknowledge Konstantinos Konstantinidis for his supplying data on potential MLST loci for this study. This work was supported in part by MRU matching funds from the College of Veterinary Medicine and Graduate College at Michigan State University.

REFERENCES

1. **Agapow, P., and A. Burt.** 2000. MultiLocus, 1.2 ed. Dept. of Biology, Imperial College, Silwood Park, Ascot, Berks.

2. **Alm, E., J. Burke, and A. Spain.** 2003. Fecal indicator bacteria are abundant in wet sand at freshwater beaches. Water Res **37:**3978-82.

3. **Altekruse, S. F., N. Bauer, A. Chanlongbutra, R. DeSagun, A. Naugle, W. Schlosser, R. Umholtz, and P. White.** 2006. *Salmonella enteritidis* in broiler chickens, United States, 2000-2005. Emerg Infect Dis **12:**1848-52.

4. **Anderson, K. L., J. E. Whitlock, and V. J. Harwood.** 2005. Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Appl Environ Microbiol **71:**3041-8.

5. **Anderson, M. A., J. E. Whitlock, and V. J. Harwood.** 2006. Diversity and distribution of *Escherichia coli* genotypes and antibiotic resistance phenotypes in feces of humans, cattle, and horses. Appl Environ Microbiol **72:**6914-22.

6. **Angulo, F. J., and D. L. Swerdlow.** 1998. *Salmonella enteritidis* infections in the United States. J Am Vet Med Assoc **213:**1729-31.

7. **Arana, I., A. Irizar, C. Seco, A. Muela, A. Fernandez-Astorga, and I. Barcina.** 2003. gfp-Tagged cells as a useful tool to study the survival of *Escherichia coli* in the presence of the river microbial community. Microb Ecol **45:**29-38.

8. **Barbosa, T. M., and S. B. Levy.** 2000. The impact of antibiotic use on resistance development and persistence. Drug Resist Updat **3:**303-311.

9. **Bartoloni, A., F. Bartalesi, A. Mantella, E. Dell'Amico, M. Roselli, M. Strohmeyer, H. G. Barahona, V. P. Barron, F. Paradisi, and G. M. Rossolini.** 2004. High prevalence of acquired antimicrobial resistance unrelated to heavy antimicrobial consumption. J Infect Dis **189:**1291-4.

10. **Bartoloni, A., L. Pallecchi, M. Benedetti, C. Fernandez, Y. Vallejos, E. Guzman, A. L. Villagran, A. Mantella, C. Lucchetti, F. Bartalesi, M. Strohmeyer, A. Bechini, H. Gamboa, H. Rodriguez, T. Falkenberg, G. Kronvall, E. Gotuzzo, F. Paradisi, and G. M. Rossolini.** 2006. Multidrug-resistant commensal *Escherichia coli* in children, Peru and Bolivia. Emerg Infect Dis **12:**907-13.

11. **Baumler, A. J., B. M. Hargis, and R. M. Tsolis.** 2000. Tracing the origins of *Salmonella* outbreaks. Science **287:**50-2.

12. **Beltran, P., J. M. Musser, R. Helmuth, J. J. F. III, W. M. Frerichs, I. K. Wachsmuth, K. Ferris, A. C. McWhorter, J. G. Wells, A. Cravioto, and R. K. Selander.** 1988. Toward a population genetic analysis of *Salmonella*: Genetic diversity and relationships among strains of serotypes *S. cholerasuis, S. derby, S. dublin, S. enteritidis, S. heidelberg, S. infantis, S. newport*, and *S. typhimurium*. Proceedings of the National Academy of Sciences USA **85:**7753-7757.

13. **Berge, A. C., E. R. Atwill, and W. M. Sischo.** 2005. Animal and farm influences on the dynamics of antibiotic resistance in faecal *Escherichia coli* in young dairy calves. Prev Vet Med **69:**25-38.

14. **Berge, A. C., P. Lindeque, D. A. Moore, and W. M. Sischo.** 2005. A clinical trial evaluating prophylactic and therapeutic antibiotic use on health and performance of preweaned calves. J Dairy Sci **88:**2166-77.

15. **Berge, A. C., D. A. Moore, and W. M. Sischo.** 2006. Field trial evaluating the influence of prophylactic and therapeutic antimicrobial administration on antimicrobial resistance of fecal *Escherichia coli* in dairy calves. Appl Environ Microbiol **72:**3872-8.

16. **Bermudez, M., and T. C. Hazen.** 1988. Phenotypic and genotypic comparison of *Escherichia coli* from pristine tropical waters. Appl Environ Microbiol **54:**979-83.

17. **Bielaszewska, M., R. Prager, W. Zhang, A. W. Friedrich, A. Mellmann, H. Tschape, and H. Karch.** 2006. Chromosomal dynamism in progeny of outbreak-related sorbitol-fermenting enterohemorrhagic Escherichia coli O157:NM. Appl Environ Microbiol **72:**1900-9.

18. **Bjorkman, J., I. Nagaev, O. G. Berg, D. Hughes, and D. I. Andersson.** 2000. Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. Science **287:**1479-82.

19. **Boerlin, P., R. Travis, C. L. Gyles, R. Reid-Smith, N. Janecko, H. Lim, V. Nicholson, S. A. McEwen, R. Friendship, and M. Archambault.** 2005. Antimicrobial resistance and virulence genes of *Escherichia coli* isolates from swine in Ontario. Appl Environ Microbiol **71:**6753-61.

20. **Bohannan, B. J., and J. Hughes.** 2003. New approaches to analyzing microbial biodiversity data. Curr Opin Microbiol **6:**282-7.

21. **Bopp, C. A., F. W. Brenner, J. G. Wells, and N. A. Strockbine.** 1999. *Escherichia, Shigella*, and *Salmonella*, p. 459-474. *In* P. R. Murray, Baron,

E.J., Jorgensen, J.H., Pfaller, M.A., Yolken, R.H. (ed.), Manual of Clinical Microbiology, 7th ed. American Society for Microbiology, Washington, DC.

22. **Boyd, E. F., F. S. Wang, P. Beltran, S. A. Plock, K. Nelson, and R. K. Selander.** 1993. *Salmonella* reference collection B (SARB): strains of 37 serovars of subspecies I. J Gen Microbiol **139 Pt 6:**1125-32.

23. **Brinas, L., M. A. Moreno, T. Teshager, M. Zarazaga, Y. Saenz, C. Porrero, L. Dominguez, and C. Torres.** 2003. Beta-lactamase characterization in *Escherichia coli* isolates with diminished susceptibility or resistance to extended-spectrum cephalosporins recovered from sick animals in Spain. Microb Drug Resist **9:**201-9.

24. **Brown, A. H. D., M. W. Feldman, and E. Nevo.** 1980. Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics **96:**523-536.

25. **Bruen, T. C., H. Philippe, and D. Bryant.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172:**2665-81.

26. **Bryant, D., and V. Moulton.** 2004. Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol **21:**255-65.

27. **Byappanahalli, M. N., R. L. Whitman, D. A. Shively, M. J. Sadowsky, and S. Ishii.** 2006. Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed. Environ Microbiol **8:**504-13.

28. **Caugant, D. A., B. R. Levin, and R. K. Selander.** 1981. Genetic diversity and temporal variation in the *E. coli* population of a human host. Genetics **98:**467-90.

29. **CDC.** 2003. *Salmonella* surveillance summary, 2002. US Department of Health and Human Survices.

30. **Chan, E. S., J. Aramini, B. Ciebin, D. Middleton, R. Ahmed, M. Howes, I. Brophy, I. Mentis, F. Jamieson, F. Rodgers, M. Nazarowec-White, S. C. Pichette, J. Farrar, M. Gutierrez, W. J. Weis, L. Lior, A. Ellis, and S. Isaacs.** 2002. Natural or raw almonds and an outbreak of a rare phage type of *Salmonella enteritidis* infection. Can Commun Dis Rep **28:**97-9.

31. **Chao, A.** 1987. Estimating the population size for capture-recapture data with unequal catchability. Biometrics **43:**783-91.

32. **Chapman, T. A., X. Y. Wu, I. Barchia, K. A. Bettelheim, S. Driesen, D. Trott, M. Wilson, and J. J. Chin.** 2006. Comparison of virulence gene

profiles of *Escherichia coli* strains isolated from healthy and diarrheic swine. Appl Environ Microbiol **72:**4782-95.

33. **Clermont, O., S. Bonacorsi, and E. Bingen.** 2000. Rapid and simple determination of the *Escherichia coli* phylogenetic group. Appl Environ Microbiol **66:**4555-8.

34. **Cogan, T. A., and T. J. Humphrey.** 2003. The rise and fall of *Salmonella Enteritidis* in the UK. J Appl Microbiol **94 Suppl:**114S-119S.

35. **Colom, K., J. Perez, R. Alonso, A. Fernandez-Aranguiz, E. Larino, and R. Cisterna.** 2003. Simple and reliable multiplex PCR assay for detection of blaTEM, bla(SHV) and blaOXA-1 genes in Enterobacteriaceae. FEMS Microbiol Lett **223:**147-51.

36. **Colwell, R.** 2005. EstimateS: Statistical estimation on species richness and shared species from samples, 7.5 ed.

37. **Czeczulin, J. R., T. S. Whittam, I. R. Henderson, F. Navarro-Garcia, and J. P. Nataro.** 1999. Phylogenetic analysis of enteroaggregative and diffusely adherent *Escherichia coli*. Infect Immun **67:**2692-9.

38. **Davies, R., and M. Breslin.** 2001. Environmental contamination and detection of *Salmonella enterica* serovar enteritidis in laying flocks. Vet Rec **149:**699-704.

39. **Davies, R. H., and C. Wray.** 1995. Mice as carriers of *Salmonella enteritidis* on persistently infected poultry units. Vet Rec **137:**337-41.

40. **De Buck, J., F. Pasmans, F. Van Immerseel, F. Haesebrouck, and R. Ducatelle.** 2004. Tubular glands of the isthmus are the predominant colonization site of Salmonella enteritidis in the upper oviduct of laying hens. Poult Sci **83:**352-8.

41. **De Buck, J., F. Van Immerseel, F. Haesebrouck, and R. Ducatelle.** 2004. Effect of type 1 fimbriae of Salmonella enterica serotype Enteritidis on bacteraemia and reproductive tract infection in laying hens. Avian Pathol **33:**314-20.

42. **Donaldson, S. C., B. A. Straley, N. V. Hegde, A. A. Sawant, C. DebRoy, and B. M. Jayarao.** 2006. Molecular epidemiology of ceftiofur-resistant *Escherichia coli* isolates from dairy calves. Appl Environ Microbiol **72:**3940-8.

43. **Elena, S. F., and R. E. Lenski.** 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat Rev Genet **4:**457-69.

44. **Escobar-Paramo, P., O. Clermont, A. B. Blanc-Potard, H. Bui, C. Le Bouguenec, and E. Denamur.** 2004. A specific genetic background is required for acquisition and expression of virulence factors in *Escherichia coli*. Mol Biol Evol **21:**1085-94.

45. **Escobar-Paramo, P., K. Grenet, A. Le Menac'h, L. Rode, E. Salgado, C. Amorin, S. Gouriou, B. Picard, M. C. Rahimy, A. Andremont, E. Denamur, and R. Ruimy.** 2004. Large-scale population structure of human commensal *Escherichia coli* isolates. Appl Environ Microbiol **70:**5698-700.

46. **Falkow, S.** 1996. The evolution of pathogenicity in *Escherichia coli*, *Shigella*, and *Salmonella*, p. 2723-2729. *In* F. C. Neidhardt, Ingraham, J.L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umarger, H.E. (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed, vol. 2. ASM Press, Washington, D.C.

47. **Farmer, J. J. I.** 2003. *Enterobacteriaceae*: Introduction and Identification, p. 636-653. *In* P. R. Murray, Baron, E.J., Jorgensen, J.H., Pfaller, M.A., Yolken, R.H. (ed.), Manual of Clinical Microbiology, 8th ed, vol. 1. ASM Press, Washington, DC.

48. **Feil, E. J.** 2004. Small change: keeping pace with microevolution. Nat Rev Microbiol **2:**483-95.

49. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol **186:**1518-30.

50. **Feil, E. J., and B. G. Spratt.** 2001. Recombination and the population structures of bacterial pathogens. Annu Rev Microbiol **55:**561-90.

51. **Garber, L., M. Smeltzer, P. Fedorka-Cray, S. Ladely, and K. Ferris.** 2003. *Salmonella enterica* serotype enteritidis in table egg layer house environments and in mice in U.S. layer houses and associated risk factors. Avian Dis **47:**134-42.

52. **Gast, R. K., and P. S. Holt.** 2001. Multiplication in egg yolk and survival in egg albumen of Salmonella enterica serotype Enteritidis strains of phage types 4, 8, 13a, and 14b. J Food Prot **64:**865-8.

53. **Gordon, D. M.** 1997. The genetic structure of *Escherichia coli* populations in feral house mice. Microbiology **143 (Pt 6):**2039-46.

54. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiology **147:**1079-85.

55. **Gordon, D. M., S. Bauer, and J. R. Johnson.** 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiology **148:**1513-22.

56. **Gordon, D. M., and A. Cowling.** 2003. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. Microbiology **149:**3575-86.

57. **Gordon, D. M., and F. FitzGibbon.** 1999. The distribution of enteric bacteria from Australian mammals: host and geographical effects. Microbiology **145 (Pt 10):**2663-71.

58. **Gordon, D. M., S. E. Stern, and P. J. Collignon.** 2005. Influence of the age and sex of human hosts on the distribution of *Escherichia coli* ECOR groups and virulence traits. Microbiology **151:**15-23.

59. **Guard-Bouldin, J., R. K. Gast, T. J. Humphrey, D. J. Henzler, C. Morales, and K. Coles.** 2004. Subpopulation characteristics of egg-contaminating *Salmonella enterica* serovar Enteritidis as defined by the lipopolysaccharide O chain. Appl Environ Microbiol **70:**2756-63.

60. **Guard-Petter, J.** 2001. The chicken, the egg and *Salmonella enteritidis*. Environ Microbiol **3:**421-30.

61. **Guard-Petter, J., D. J. Henzler, M. M. Rahman, and R. W. Carlson.** 1997. On-farm monitoring of mouse-invasive *Salmonella enterica* serovar enteritidis and a model for its association with the production of contaminated eggs. Appl Environ Microbiol **63:**1588-93.

62. **Guttman, D. S., and D. E. Dykhuizen.** 1994. Clonal divergence in Escherichia coli as a result of recombination, not mutation. Science **266:**1380-3.

63. **Guttman, D. S., and D. E. Dykhuizen.** 1994. Detecting selective sweeps in naturally occurring *Escherichia coli*. Genetics **138:**993-1003.

64. **Hamilton, M. J., T. Yan, and M. J. Sadowsky.** 2006. Development of goose- and duck-specific DNA markers to determine sources of *Escherichia coli* in waterways. Appl Environ Microbiol **72:**4012-9.

65. **Haubold, B., M. Travisano, P. B. Rainey, and R. R. Hudson.** 1998. Detecting linkage disequilibrium in bacterial populations. Genetics **150:**1341-8.

66. **Hedberg, C. W., M. J. David, K. E. White, K. L. MacDonald, and M. T. Osterholm.** 1993. Role of egg consumption in sporadic *Salmonella enteritidis* and *Salmonella typhimurium* infections in Minnesota. J Infect Dis **167:**107-11.

67. **Henzler, D. J., and H. M. Opitz.** 1992. The role of mice in the epizootiology of *Salmonella enteritidis* infection on chicken layer farms. Avian Dis **36:**625-31.

68. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli.* J Bacteriol **172:**6175-81.

69. **Hoffmann, H., M. W. Hornef, S. Schubert, and A. Roggenkamp.** 2001. Distribution of the outer membrane haem receptor protein ChuA in environmental and human isolates of Escherichia coli. Int J Med Microbiol **291:**227-30.

70. **Holt, P. S.** 1995. Horizontal transmission of Salmonella enteritidis in molted and unmolted laying chickens. Avian Dis **39:**239-49.

71. **Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. Bohannan.** 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Appl Environ Microbiol **67:**4399-406.

72. **Humphrey, T. J.** 1994. Contamination of egg shell and contents with *Salmonella enteritidis*: a review. Int J Food Microbiol **21:**31-40.

73. **Humphrey, T. J., E. Slater, K. McAlpine, R. J. Rowbury, and R. J. Gilbert.** 1995. Salmonella enteritidis phage type 4 isolates more tolerant of heat, acid, or hydrogen peroxide also survive longer on surfaces. Appl Environ Microbiol **61:**3161-4.

74. **Huson, D. H., and D. Bryant.** 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **23:**254-67.

75. **Hyma, K. E., D. W. Lacher, A. M. Nelson, A. C. Bumbaugh, J. M. Janda, N. A. Strockbine, V. B. Young, and T. S. Whittam.** 2005. Evolutionary genetics of a new pathogenic *Escherichia* species: *Escherichia albertii* and related *Shigella boydii* strains. J Bacteriol **187:**619-28.

76. **Isaacs, S., J. Aramini, B. Ciebin, J. A. Farrar, R. Ahmed, D. Middleton, A. U. Chandran, L. J. Harris, M. Howes, E. Chan, A. S. Pichette, K. Campbell, A. Gupta, L. Y. Lior, M. Pearce, C. Clark, F. Rodgers, F. Jamieson, I. Brophy, and A. Ellis.** 2005. An international outbreak of salmonellosis associated with raw almonds contaminated with a rare phage type of *Salmonella enteritidis*. J Food Prot **68:**191-8.

77. **Ishii, S., W. B. Ksoll, R. E. Hicks, and M. J. Sadowsky.** 2006. Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds. Appl Environ Microbiol **72:**612-21.

78. **Jacobs, J. L., T. L. Carroll, and G. W. Sundin.** 2005. The role of pigmentation, ultraviolet radiation tolerance, and leaf colonization strategies in the epiphytic survival of phyllosphere bacteria. Microb Ecol **49:**104-13.

79. **Jenkins, M. B., P. G. Hartel, T. J. Olexa, and J. A. Stuedemann.** 2003. Putative temporal variability of *Escherichia coli* ribotypes from yearling steers. J Environ Qual **32:**305-9.

80. **Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky.** 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. Appl Environ Microbiol **70:**4478-85.

81. **Jones, T., C. O. Gill, and L. M. McMullen.** 2004. The behaviour of log phase *Escherichia coli* at temperatures that fluctuate about the minimum for growth. Lett Appl Microbiol **39:**296-300.

82. **Khachatryan, A. R., T. E. Besser, D. D. Hancock, and D. R. Call.** 2006. Use of a nonmedicated dietary supplement correlates with increased prevalence of streptomycin-sulfa-tetracycline-resistant *Escherichia coli* on a dairy farm. Appl Environ Microbiol **72:**4583-8.

83. **Khachatryan, A. R., D. D. Hancock, T. E. Besser, and D. R. Call.** 2006. Antimicrobial drug resistance genes do not convey a secondary fitness advantage to calf-adapted *Escherichia coli*. Appl Environ Microbiol **72:**443-8.

84. **Khachatryan, A. R., D. D. Hancock, T. E. Besser, and D. R. Call.** 2004. Role of calf-adapted *Escherichia coli* in maintenance of antimicrobial drug resistance in dairy calves. Appl Environ Microbiol **70:**752-7.

85. **Kilmartin, D., D. Morris, C. O'Hare, G. Corbett-Feeney, and M. Cormican.** 2005. Clonal expansion may account for high levels of quinolone resistance in *Salmonella enterica* serovar enteritidis. Appl Environ Microbiol **71:**2587-91.

86. **Kimura, A. C., V. Reddy, R. Marcus, P. R. Cieslak, J. C. Mohle-Boetani, H. D. Kassenborg, S. D. Segler, F. P. Hardnett, T. Barrett, and D. L. Swerdlow.** 2004. Chicken consumption is a newly identified risk factor for sporadic *Salmonella enterica* serotype Enteritidis infections in the United States: a case-control study in FoodNet sites. Clin Infect Dis **38 Suppl 3:**S244-52.

87. **Konstantinidis, K. T., A. Ramette, and J. M. Tiedje.** 2006. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. Appl Environ Microbiol **72:**7286-93.

88. **Kosakovsky Pond, S. L., D. Posada, M. B. Gravenor, C. H. Woelk, and S. D. Frost.** 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics **22:**3096-8.

89. **Kudva, I. T., P. S. Evans, N. T. Perna, T. J. Barrett, F. M. Ausubel, F. R. Blattner, and S. B. Calderwood.** 2002. Strains of *Escherichia coli* O157:H7 differ primarily by insertions or deletions, not single-nucleotide polymorphisms. Journal of Bacteriology **184:**1873-1879.

90. **Kuehl, C. J., H. D. Wood, T. L. Marsh, T. M. Schmidt, and V. B. Young.** 2005. Colonization of the cecal mucosa by *Helicobacter hepaticus* impacts the diversity of the indigenous microbiota. Infect Immun **73:**6952-61.

91. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform **5:**150-63.

92. **Lasalde, C., R. Rodriguez, and G. A. Toranzos.** 2005. Statistical analyses: possible reasons for unreliability of source tracking efforts. Appl Environ Microbiol **71:**4690-5.

93. **Levin, B. R., and C. T. Bergstrom.** 2000. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. Proc Natl Acad Sci U S A **97:**6981-5.

94. **Levy, S. B.** 2002. The 2000 Garrod lecture. Factors impacting on the problem of antibiotic resistance. J Antimicrob Chemother **49:**25-30.

95. **Ley, R. E., F. Backhed, P. Turnbaugh, C. A. Lozupone, R. D. Knight, and J. I. Gordon.** 2005. Obesity alters gut microbial ecology. Proc Natl Acad Sci U S A **102:**11070-5.

96. **Li, J., N. H. Smith, K. Nelson, P. B. Crichton, D. C. Old, T. S. Whittam, and R. K. Selander.** 1993. Evolutionary origin and radiation of the avian-adapted non-motile salmonellae. J Med Microbiol **38:**129-39.

97. **Lindstedt, B. A., E. Heir, I. Nygard, and G. Kapperud.** 2003. Characterization of class I integrons in clinical strains of *Salmonella enterica* subsp. enterica serovars Typhimurium and Enteritidis from Norwegian hospitals. J Med Microbiol **52:**141-9.

98. **Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt.** 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A **95:**3140-5.

99. **Maisnier-Patin, S., and D. I. Andersson.** 2004. Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. Res Microbiol **155:**360-9.

100. **Maisnier-Patin, S., O. G. Berg, L. Liljas, and D. I. Andersson.** 2002. Compensatory adaptation to the deleterious effect of antibiotic resistance in *Salmonella typhimurium*. Mol Microbiol **46:**355-66.

101. **Maslow, J. N., T. S. Whittam, C. F. Gilks, R. A. Wilson, M. E. Mulligan, K. S. Adams, and R. D. Arbeit.** 1995. Clonal relationships among bloodstream isolates of *Escherichia coli*. Infect Immun **63:**2409-17.

102. **Maynard Smith, J., N. H. Smith, M. O'Rourke, and B. G. Spratt.** 1993. How clonal are bacteria? Proc Natl Acad Sci U S A **90:**4384-8.

103. **McLellan, S. L.** 2004. Genetic diversity of *Escherichia coli* isolated from urban rivers and beach water. Appl Environ Microbiol **70:**4658-65.

104. **McLellan, S. L., A. D. Daniels, and A. K. Salmore.** 2001. Clonal populations of thermotolerant *Enterobacteriaceae* in recreational water and their potential interference with fecal *Escherichia coli* counts. Appl Environ Microbiol **67:**4934-8.

105. **Mead, P. S., L. Slutsker, V. Dietz, L. F. McCaig, J. S. Bresee, C. Shapiro, P. M. Griffin, and R. V. Tauxe.** 1999. Food-related illness and death in the United States. Emerg Infect Dis **5:**607-25.

106. **Milkman, R.** 1973. Electrophoretic variation in *Escherichia coli* from natural sources. Science **182:**1024-6.

107. **Miller, W. A., M. A. Miller, I. A. Gardner, E. R. Atwill, B. A. Byrne, S. Jang, M. Harris, J. Ames, D. Jessup, D. Paradies, K. Worcester, A. Melli, and P. A. Conrad.** 2006. *Salmonella* spp., *Vibrio* spp., *Clostridium perfringens*, and *Plesiomonas shigelloides* in marine and freshwater invertebrates from coastal California ecosystems. Microb Ecol **52:**198-206.

108. **Nakamura, M., N. Nagamine, M. Norimatsu, S. Suzuki, K. Ohishi, M. Kijima, Y. Tamura, and S. Sato.** 1993. The ability of *Salmonella enteritidis* isolated from chicks imported from England to cause transovarian infection. J Vet Med Sci **55:**135-6.

109. **NCCLS.** 2002. Performance standard for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals. Approved standard., 2nd ed, Wayne, PA.

110. **Nei, M., and S. Kumar.** 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY.

111. **Nowrouzian, F., B. Hesselmar, R. Saalman, I. L. Strannegard, N. Aberg, A. E. Wold, and I. Adlerberth.** 2003. *Escherichia coli* in infants' intestinal microflora: colonization rate, strain turnover, and virulence gene carriage. Pediatr Res **54:**8-14.

112. **Nowrouzian, F. L., I. Adlerberth, and A. E. Wold.** 2006. Enhanced persistence in the colonic microbiota of *Escherichia coli* strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect **8:**834-40.

113. **Nowrouzian, F. L., A. E. Wold, and I. Adlerberth.** 2005. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. J Infect Dis **191:**1078-83.

114. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J Bacteriol **157:**690-3.

115. **Ochman, H., T. S. Whittam, D. A. Caugant, and R. K. Selander.** 1983. Enzyme polymorphism and genetic population strucutre in *Escherichia coli* and *Shigella*. Journal of General Microbiology **129:**2715-2726.

116. **Painter, J. A., K. Molbak, J. Sonne-Hansen, T. Barrett, J. G. Wells, and R. V. Tauxe.** 2004. *Salmonella*-based rodenticides and public health. Emerg Infect Dis **10:**985-7.

117. **Pallecchi, L., C. Lucchetti, A. Bartoloni, F. Bartalesi, A. Mantella, H. Gamboa, A. Carattoli, F. Paradisi, and G. M. Rossolini.** 2007. Population structure and resistance genes in antibiotic-resistant bacteria from a remote

community with minimal antibiotic exposure. Antimicrob Agents Chemother. **51**:1179-84.

118. **Pallecchi, L., M. Malossi, A. Mantella, E. Gotuzzo, C. Trigoso, A. Bartoloni, F. Paradisi, G. Kronvall, and G. M. Rossolini.** 2004. Detection of CTX-M-type beta-lactamase genes in fecal *Escherichia coli* isolates from healthy children in Bolivia and Peru. Antimicrob Agents Chemother **48:**4556-61.

119. **Parker, C. T., B. Harmon, and J. Guard-Petter.** 2002. Mitigation of avian reproductive tract function by Salmonella enteritidis producing high-molecular-mass lipopolysaccharide. Environ Microbiol **4:**538-45.

120. **Patrick, M. E., P. M. Adcock, T. M. Gomez, S. F. Altekruse, B. H. Holland, R. V. Tauxe, and D. L. Swerdlow.** 2004. *Salmonella enteritidis* infections, United States, 1985-1999. Emerg Infect Dis **10:**1-7.

121. **Perez-Rosas, N., and T. C. Hazen.** 1989. *In situ* survival of *Vibrio cholerae* and *Escherichia coli* in a tropical rain forest watershed. Appl Environ Microbiol **55:**495-9.

122. **Piffaretti, J. C., H. Kressebuch, M. Aeschbacher, J. Bille, E. Bannerman, J. M. Musser, R. K. Selander, and J. Rocourt.** 1989. Genetic characterization of clones of the bacterium Listeria monocytogenes causing epidemic disease. Proc Natl Acad Sci U S A **86:**3818-22.

123. **Power, M. L., J. Littlefield-Wyer, D. M. Gordon, D. A. Veal, and M. B. Slade.** 2005. Phenotypic and genotypic characterization of encapsulated *Escherichia coli* isolated from blooms in two Australian lakes. Environ Microbiol **7:**631-40.

124. **Qi, W., D. W. Lacher, A. C. Bumbaugh, K. E. Hyma, L. M. Ouellette, T. M. Large, C. L. Tarr, and T. S. Whittam.** 2004. Presented at the Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference CSB2004, Los Alamitos, CA.

125. **Rahman, M. M., J. Guard-Petter, and R. W. Carlson.** 1997. A virulent isolate of Salmonella enteritidis produces a Salmonella typhi-like lipopolysaccharide. J Bacteriol **179:**2126-31.

126. **Ram, J. L., R. P. Ritchie, J. Fang, F. S. Gonzales, and J. P. Selegean.** 2004. Sequence-based source tracking of *Escherichia coli* based on genetic diversity of beta-glucuronidase. J Environ Qual **33:**1024-32.

127. **Raymond, M. J., R. D. Wohrle, and D. R. Call.** 2006. Assessment and promotion of judicious antibiotic use on dairy farms in Washington State. J Dairy Sci **89:**3228-40.

128. **Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam.** 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. Nature **406:**64-7.

129. **Reynolds, M. G.** 2000. Compensatory evolution in rifampin-resistant *Escherichia coli*. Genetics **156:**1471-81.

130. **Rice, E. W., C. H. Johnson, M. E. Dunnigan, and D. J. Reasoner.** 1993. Rapid glutamate decarboxylase assay for detection of *Escherichia coli*. Appl Environ Microbiol **59:**4347-9.

131. **Rivera, S. C., T. C. Hazen, and G. A. Toranzos.** 1988. Isolation of fecal coliforms from pristine sites in a tropical rain forest. Appl Environ Microbiol **54:**513-7.

132. **Rodrigue, D. C., R. V. Tauxe, and B. Rowe.** 1990. International increase in *Salmonella enteritidis*: a new pandemic? Epidemiol Infect **105:**21-7.

133. **Sabate, M., E. Moreno, T. Perez, A. Andreu, and G. Prats.** 2006. Pathogenicity island markers in commensal and uropathogenic *Escherichia coli* isolates. Clin Microbiol Infect **12:**880-6.

134. **Sander, P., B. Springer, T. Prammananan, A. Sturmfels, M. Kappler, M. Pletschette, and E. C. Bottger.** 2002. Fitness cost of chromosomal drug resistance-conferring mutations. Antimicrob Agents Chemother **46:**1204-11.

135. **Sannes, M. R., M. A. Kuskowski, K. Owens, A. Gajewski, and J. R. Johnson.** 2004. Virulence factor profiles and phylogenetic background of *Escherichia coli* isolates from veterans with bacteremia and uninfected control subjects. J Infect Dis **190:**2121-8.

136. **Sato, K., P. C. Bartlett, and M. A. Saeed.** 2005. Antimicrobial susceptibility of *Escherichia coli* isolates from dairy farms using organic versus conventional production methods. J Am Vet Med Assoc **226:**589-94.

137. **Savageau, M. A.** 1983. *Escherichia coli* habitats, cell types, and molecular mechanisms of gene control. The American Naturalist **122:**732-744.

138. **Sawant, A. A., N. V. Hegde, B. A. Straley, S. C. Donaldson, B. C. Love, S. J. Knabel, and B. M. Jayarao.** 2007. Antimicrobial-resistant enteric bacteria from dairy cattle. Appl Environ Microbiol **73:**156-63.

139. **Sawant, A. A., L. M. Sordillo, and B. M. Jayarao.** 2005. A survey on antibiotic usage in dairy herds in Pennsylvania. J Dairy Sci **88:**2991-9.

140. **Schmidt, T. M., and D. A. Relman.** 1994. Phylogenetic identification of uncultured pathogens using ribosomal RNA sequences. Methods Enzymol **235:**205-22.

141. **Schrag, S. J., V. Perrot, and B. R. Levin.** 1997. Adaptation to the fitness costs of antibiotic resistance in *Escherichia coli*. Proc Biol Sci **264:**1287-91.

142. **Schroeder, C. M., A. L. Naugle, W. D. Schlosser, A. T. Hogue, F. J. Angulo, J. S. Rose, E. D. Ebel, W. T. Disney, K. G. Holt, and D. P. Goldman.** 2005. Estimate of illnesses from *Salmonella enteritidis* in eggs, United States, 2000. Emerg Infect Dis **11:**113-5.

143. **Selander, R. K., P. Beltran, N. H. Smith, R. Helmuth, F. A. Rubin, D. J. Kopecko, K. Ferris, B. D. Tall, A. Cravioto, and J. M. Musser.** 1990. Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. Infection and Immunity **58:**2262-2275.

144. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl Environ Microbiol **51:**873-84.

145. **Selander, R. K., Caugant, D.A., and Whittam, T.S.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*., p. 1625-1648. *In* F. C. Neidhardt, Ingraham, J.L., Magasanik, B., Schaechter, M., and Umarger, H.E. (ed.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology. American Society for Microbiology, Washington, DC.

146. **Selander, R. K., and B. R. Levin.** 1980. Genetic diversity and structure in *Escherichia coli* populations. Science **210:**545-7.

147. **Selander, R. K., Li, J., Nelson, K.** 1996. Evolutionary genetics of *Salmonella enterica*, p. 2691-2707. *In* F. C. Neidhardt, Ingraham, J.L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umarger, H.E. (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed, vol. 2. ASM Press, Washington, D.C.

148. **Selander, R. K., J. M. Musser, D. A. Caugant, M. N. Gilmour, and T. S. Whittam.** 1987. Population genetics of pathogenic bacteria. Microb Pathog **3:**1-7.

149. **Sherley, M., D. M. Gordon, and P. J. Collignon.** 2004. Evolution of multi-resistance plasmids in Australian clinical isolates of *Escherichia coli*. Microbiology **150:**1539-46.

150. **Shivaprasad, H. L.** 2000. Fowl typhoid and pullorum disease. Rev Sci Tech **19:**405-24.

151. **Simpson, K. W., B. Dogan, M. Rishniw, R. E. Goldstein, S. Klaessig, P. L. McDonough, A. J. German, R. M. Yates, D. G. Russell, S. E. Johnson, D. E. Berg, J. Harel, G. Bruant, S. P. McDonough, and Y. H. Schukken.** 2006. Adherent and invasive *Escherichia coli* is associated with granulomatous colitis in boxer dogs. Infect Immun **74:**4778-92.

152. **Singh, R., C. M. Schroeder, J. Meng, D. G. White, P. F. McDermott, D. D. Wagner, H. Yang, S. Simjee, C. Debroy, R. D. Walker, and S. Zhao.** 2005. Identification of antimicrobial resistance and class 1 integrons in Shiga toxin-producing *Escherichia coli* recovered from humans and food animals. J Antimicrob Chemother **56:**216-9.

153. **Sinton, L. W., C. H. Hall, P. A. Lynch, and R. J. Davies-Colley.** 2002. Sunlight inactivation of fecal indicator bacteria and bacteriophages from waste stabilization pond effluent in fresh and saline waters. Appl Environ Microbiol **68:**1122-31.

154. **Sokal, R. R., and F. J. Rohlf.** 1995. Biometry, 3rd ed. W. H. Freeman and Company, New York.

155. **Solano, C., B. Sesma, M. Alvarez, T. J. Humphrey, C. J. Thorns, and C. Gamazo.** 1998. Discrimination of strains of Salmonella enteritidis with differing levels of virulence by an in vitro glass adherence test. J Clin Microbiol **36:**674-8.

156. **Solo-Gabriele, H. M., M. A. Wolfert, T. R. Desmarais, and C. J. Palmer.** 2000. Sources of *Escherichia coli* in a coastal subtropical environment. Appl Environ Microbiol **66:**230-7.

157. **Souza, V., M. Rocha, A. Valera, and L. E. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Appl Environ Microbiol **65:**3373-85.

158. **St Louis, M. E., D. L. Morse, M. E. Potter, T. M. DeMelfi, J. J. Guzewich, R. V. Tauxe, and P. A. Blake.** 1988. The emergence of grade A eggs as a major source of *Salmonella enteritidis* infections. New implications for the control of salmonellosis. Jama **259:**2103-7.

159. **Stackebrandt, E., and J. Ebers.** 2006. Taxonomic properties revisited: tarnished gold standards. Microbiology Today:6-9.

160. **Stoeckel, D. M., M. V. Mathes, K. E. Hyer, C. Hagedorn, H. Kator, J. Lukasik, T. L. O'Brien, T. W. Fenger, M. Samadpour, K. M. Strickler, and B. A. Wiggins.** 2004. Comparison of seven protocols to identify fecal contamination sources using Escherichia coli. Environ Sci Technol **38:**6109-17.

161. **Thiagarajan, D., A. M. Saeed, and E. K. Asem.** 1994. Mechanism of transovarian transmission of *Salmonella enteritidis* in laying hens. Poult Sci **73:**89-98.

162. **Thiagarajan, D., M. Saeed, J. Turek, and E. Asem.** 1996. In vitro attachment and invasion of chicken ovarian granulosa cells by *Salmonella enteritidis* phage type 8. Infect Immun **64:**5015-21.

163. **Threlfall, E. J., A. M. Ridley, L. R. Ward, and B. Rowe.** 1996. Assessment of health risk from Salmonella-based rodenticides. Lancet **348:**616-7.

164. **Truchanowicz, J., E. Burek, and Z. Gorzelak.** 1970. [Clinical observations on the course of Salmonella enteritidis infections in children]. Przegl Epidemiol **24:**101-6.

165. **Tsolis, R. M., L. G. Adams, T. A. Ficht, and A. J. Baumler.** 1999. Contribution of *Salmonella typhimurium* virulence factors to diarrheal disease in calves. Infect Immun **67:**4879-85.

166. **Turnbaugh, P. J., R. E. Ley, M. A. Mahowald, V. Magrini, E. R. Mardis, and J. I. Gordon.** 2006. An obesity-associated gut microbiome with increased capacity for energy harvest. Nature **444:**1027-31.

167. **USEPA.** 2000. Improved enumeration methods for the recreational water quality indicators: *Enterococci* and *Escherichia coli*. USEPA.

168. **Velge, P., A. Cloeckaert, and P. Barrow.** 2005. Emergence of *Salmonella* epidemics: the problems related to *Salmonella enterica* serotype Enteritidis and multiple antibiotic resistance in other major serotypes. Vet Res **36:**267-88.

169. **Walk, S. T., E. W. Alm, L. M. Calhoun, J. M. Mladonicky, and T. S. Whittam.** 2007. Genetic diverstiy and population structure of *Escherichia coli* isolated from freshwater beaches. Environ Microbiol **In Press**.

170. **Weissman, S. J., S. Chattopadhyay, P. Aprikian, M. Obata-Yasuoka, Y. Yarova-Yarovaya, A. Stapleton, W. Ba-Thein, D. Dykhuizen, J. R.**

Johnson, and E. V. Sokurenko. 2006. Clonal analysis reveals high rate of structural mutations in fimbrial adhesins of extraintestinal pathogenic *Escherichia coli*. Mol Microbiol **59:**975-88.

171. **Whitman, R. L., and M. B. Nevers.** 2003. Foreshore sand as a source of *Escherichia coli* in nearshore water of a Lake Michigan beach. Appl Environ Microbiol **69:**5555-62.

172. **Whitman, R. L., M. B. Nevers, G. C. Korinek, and M. N. Byappanahalli.** 2004. Solar and temporal effects on *Escherichia coli* concentration at a Lake Michigan swimming beach. Appl Environ Microbiol **70:**4276-85.

173. **Whittam, T. S.** 1989. Clonal dynamics of *Escherichia coli* in its natural habitat. Antonie Van Leeuwenhoek **55:**23-32.

174. **Whittam, T. S.** 1996. Genetic variation and evolutionary processes in natural populations of *Escherichia coli*, p. 2708-2720. *In* F. C. Neidhardt, Ingraham, J.L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umarger, H.E. (ed.), *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology, 2nd ed, vol. 2. ASM Press, Washington, D.C.

175. **Whittam, T. S., and A. C. Bumbaugh.** 2002. Inferences from whole-genome sequences of bacterial pathogens. Curr Opin Genet Dev **12:**719-25.

176. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. Proc Natl Acad Sci U S A **80:**1751-5.

177. **Whittam, T. S., I. K. Wachsmuth, and R. A. Wilson.** 1988. Genetic evidence of clonal descent of *Escherichia coli* O157:H7 associated with hemorrhagic colitis and hemolytic uremic syndrome. Journal of Infectious Diseases **157:**1124-1133.

178. **Whittam, T. S., and R. A. Wilson.** 1988. Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. Infection and Immunity **56:**2467-2473.

179. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect Immun **61:**1619-29.

180. **Whittam, T. S., M. L. Wolfe, and R. A. Wilson.** 1989. Genetic relationships among *Escherichia coli* isolates causing urinary tract infections in humans and animals. Epidemiol Infect **102:**37-46.

181. **Winfield, M. D., and E. A. Groisman.** 2003. Role of nonhost environments in the lifestyles of *Salmonella* and *Escherichia coli*. Appl Environ Microbiol **69:**3687-94.

182. **Wirth, T., D. Falush, R. Lan, F. Colles, P. Mensa, L. H. Wieler, H. Karch, P. R. Reeves, M. C. Maiden, H. Ochman, and M. Achtman.** 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol **60:**1136-51.