Journal of Food Protection, Vol. 68, No. 6, 2005, Pages 1271–1284

Chips and SNPs, Bugs and Thugs: A Molecular Sleuthing Perspective

THOMAS A. CEBULA,* SCOTT A. JACKSON, ERIC W. BROWN, BISWENDU GOSWAMI, AND J. EUGENE LECLERC

Division of Molecular Biology (HFS-025), Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, Laurel, Maryland 20708, USA

MS 04-705: Received 10 May 2004/Accepted 9 July 2004

ABSTRACT

Recent events both here and abroad have focused attention on the need for ensuring a safe and secure food supply. Although much has been written about the potential of particular select agents in bioterrorism, we must consider seriously the more mundane pathogens, especially those that have been implicated previously in foodborne outbreaks of human disease, as possible agents of bioterrorism. Given their evolutionary history, the enteric pathogens are more diverse than agents such as *Bacillus anthracis, Francisella tularensis,* or *Yersinia pestis.* This greater diversity, however, is a double-edged sword; although diversity affords the opportunity for unequivocal identification of an organism without the need for whole-genome sequencing, the same diversity can confound definitive forensic identification if boundaries are not well defined. Here, we discuss molecular approaches used for the identification of *Salmonella enterica, Escherichia coli,* and *Shigella* spp. and viral pathogens and discuss the utility of these approaches to the field of microbial molecular forensics.

Several years ago, Nobel Laureate Joshua Lederberg quipped that because of the countless numbers on their side, we could at best expect to stay but even in a war with our microbial adversaries. He went on to say that it would take our ingenuity and technological advances to ensure that we peacefully coexist within a world dominated by the evolving microbe. Today, his words resonate even more loudly considering that a human adversary, the bioterrorist, is poised to thwart our attempts at détente with the microbial world.

With the events of 11 September 2001, and the incidents involving anthrax-laced letters, it is evident that our food supply could become a target for terrorist activities. Much attention, therefore, has been focused on particular microbiological agents as the most severe threats to public health. Improved methods for identifying *Bacillus anthracis, Franciscella tularensis, Yersinia pestis, Brucella* spp., and *Burkholderia* spp. are actively being developed, and the application of these methods for detection of these agents in foods is actively being assessed. The more conventional and established foodborne pathogens also could be used as etiological agents in bioterrorist acts. Proper tools for molecular discrimination also are needed for rapid detection, identification, and tracing of foodborne bacteria such as *Escherichia coli, Shigella* spp., and *Salmonella enterica* and the enteric viruses that have frequently contaminated the food supply.

The evolution and diversity of organisms are the bases of the genetic approaches for identifying them; mutations and recombinational exchanges ultimately define the unique fingerprint and ''personality'' of an individual strain. The more time a bacterial species has had to evolve, the more diversity might be expected among individual strains of that species. For example, genetic methods for strain identification have been difficult to develop for *B. anthracis* because its genome and those of its siblings *Bacillus cereus* and *Bacillus thuringiensis* are so similar *(80).* Although whole genome sequencing of *B. anthracis* isolates has revealed polymorphic sites *(82),* characterization of variable-number tandem repeats (VNTRs) is a more rapid means for discriminating strains *(51).* Such methods have also been developed for strain typing of *Y. pestis (53)* and *F. tularensis (34).* The lack of extensive diversity in these genomes is an indication that these pathogens are relatively young as measured on an evolutionary time frame.

In contrast, over 100 million years of evolution have elapsed since *E. coli* and *Salmonella* diverged *(69).* During this time, multiple polymorphisms have accumulated, making it easier to develop methods for discriminating strains, without the need for total genomic sequencing. Here, we describe molecular approaches that will make strain discrimination among enteric pathogens achievable. Although the technologies will continue to evolve for greater speed and larger sample sizes, the methods based in cladistic analysis of genomic diversity should help establish and maintain the field of microbial molecular forensics. Such methodologies are essential for rapidly detecting and identifying microbial contaminants and for determining whether the contamination is accidental or deliberate.

METHODS USED IN THE IDENTIFICATION AND DIFFERENTIATION OF BACTERIAL STRAINS

Historical perspectives. The identification and typing of bacteria have relied upon phenotypic or biotypic methods. Many methods are based on the differences in bacterial

* Author for correspondence. Tel: 301-827-8281; Fax: 301-827-8260; E-mail: tcebula@cfsan.fda.gov.

morphology, physiology, nutrient utilization, and serology *(11, 23, 99).* For example, the Gram stain was developed as a method for discriminating one bacterium from another. Although they are relatively rapid, classical approaches to bacterial typing have many limitations. In most cases, typing results can differ within or between individual laboratories, and variations can be due to subtle differences in culture conditions. Both the relative dearth of phenotypic characters and the chance that such characters are due to convergent evolution may confound the correct typing of microbial agents associated with a disease outbreak.

In general, the problem of accurate strain identification (i.e., species assignment) has been largely circumvented by the numerous molecular and phenotypic approaches now available. Many of these approaches are more than adequate for species identification. These diagnostics are less useful for discrimination of subspecies. Nonetheless, several serological and molecular genetic methods have been developed that in many cases may be useful for subtype identification of bacterial strains.

One of the most commonly used phenotypic tools for differentiating bacteria is serotyping *(70).* This approach sorts strains into distinct classes based on biochemical variations that accrue among bacterial cell surface antigens. Serotyping has been a mainstay in the bacteriologist's arsenal and will likely continue to play a key role in distinguishing strains. However, the inability of other phenotypic methods to recapitulate strain origins and their distribution patterns has led to development of molecular genetic methods for strain identification and differentiation.

RFLP-based methods of strain typing. Molecular methods for strain differentiation exploit the nucleotide differences that accrue among strains over evolutionary time. These nucleotide differences may be uncovered using specific restriction endonucleases, which cleave the DNA of various strains at different places creating DNA fragments of different lengths. One such method, ribotyping, makes use of restriction fragment length polymorphisms (RFLPs) that emerge within the intergenic spacer and flanking regions of the rRNA genes of bacteria *(88).* Although this technique is applicable to all bacteria, its resolving power is somewhat limited because of the highly conserved nature of the bacterial rRNA operon and the limited number of rRNA operons in many species of bacteria *(99).* Thus, ribotyping and its automated successor riboprinting *(74)* can provide unique genotypic information about a bacterium at the species level, yet many closely related bacteria yield identical banding patterns. Subspecies resolution is not possible using RFLPs within rDNA sequences alone.

The most widely used method of RFLP-based bacterial typing is pulsed-field gel electrophoresis (PFGE) *(15).* This method uses restriction enzymes that cleave the entire bacterial genome into 30 to 50 very large fragments. The fragments of DNA are then separated using gel electrophoresis with a pulsing electrical current capable of separating these large segments. Although the technique has been exploited widely in the typing of foodborne bacteria associated with outbreaks *(4),* PFGE has limitations. Restriction of bacterial genomes, which are two to five million base pairs in size, into only tens of subgenomic fragments clearly does not provide the resolution necessary for surveying the full breadth of mutational and recombination diversity that is now widely known to exist among genomes of the same species. Moreover, electromorphs revealed by PFGE need not be related genetically, i.e., they can be derived from distinctly different regions of the genome. Thus, PFGE similarity coefficients offer poor measures of relatedness, particularly with single-enzyme PFGE systems *(25).* PFGE technology is not well suited for screening large numbers of strains and, the resulting data are not easily standardized between and among individual laboratories.

A recently introduced variation on the RFLP and PFGE methods for detecting mutations and polymorphisms exploits the ability of the mismatch-specific endonuclease CEL I to cleave heteroduplexes formed from the DNAs of different bacterial strains *(89).* This method can be used as a genomewide mutation scan because it is applicable to large regions of the chromosome that are amenable to analysis by PFGE.

Amplification-based fingerprinting methods for strain discrimination. Several methods now exist that allow for amplification-based fingerprinting of bacterial genomes. Some of these methods rely on the annealing of degenerate oligonucleotide primer sequences to anonymous regions of DNA in the bacterial genome. Genetic diversity between strains then allows for differential primer annealing between them, resulting in a series of strain-specific size differences among amplicons. Two examples of this technology are AFLP (amplification fragment length polymorphism) and RAPD (randomly amplified polymorphic DNA) analyses. In AFLP analysis, bacterial DNA is enzymatically restricted, yielding DNA fragments of various lengths *(94).* Double-stranded adapter oligonucleotides are ligated to the ends of the digested DNA fragments, and adapter-specific primers containing selective 3′ nucleotides are then used to amplify a subset of fragments from the total pool of restricted fragments.

RAPD-PCR has been used to type closely related genomes *(61, 98)* and to map regions of plant genomes and resolve fungal races *(24).* RAPD analysis uses a short single random primer of 8 to 10 bp to amplify segments of the bacterial genomes that share sequences complementary to the primer. Strains that deviate by only several nucleotide substitutions can be differentiated by the presence or absence of specific amplicons. Because of its high level of sensitivity, the method has been particularly useful in the tracking of nosocomial outbreaks and in other epidemiological investigations where there is a need to resolve highly homologous strains *(61).* One drawback of RAPD analysis is how readily the amplicon patterns are altered by minor perturbations. Several factors are known to influence RAPD-generated banding patterns, including template quality, source of thermostable (e.g., *Taq*) polymerase, and the thermal cycling protocols employed *(46).*

Other amplification-based methods rely on specific well-characterized iterations of short sequences scattered

throughout the bacterial genome. Repetitive PCR fingerprint analysis makes use of particular conserved families of nucleotide repeat elements (e.g., REP, ERIC, and BOX) dispersed throughout the bacterial chromosome *(100).* Primers are used to target these specific motifs and amplify the chromosomal regions between them. The amplicon pattern then is a function of the physical location of the repeat element on the bacterial chromosome.

Another novel molecular strain typing system exploits the variation in length found among regions of short sequence DNA repeat (SSR) elements in the prokaryotic genome *(92).* These repeated sequences are contained in a single locus, a VNTR locus, and analysis reveals length variability between individual strains *(65).* Changes in copy number of repeat sequences among single-locus VNTRs can be distinguished using flanking PCR primers. The polymorphisms among several different VNTR loci should provide a useful means to discriminate among strains. This approach, termed multiple-locus VNTR analysis *(51),* has been used with some success in distinguishing *B. anthracis* strains *(51).* Because variation among VNTR loci seems to occur at the individual level and VNTR detection is relatively rapid and simple, this approach is expected to play a prominent role in strategies for detection and identification of individual microbial strains.

Although the various molecular techniques may be useful for the differentiation of enteric bacterial strains *(27),* it is unclear whether the resultant relationships that emerge from many of these DNA repeat band-sharing analyses are evolutionarily meaningful. The current distributions of many of the genomic elements of enteric bacteria likely arose by horizontal gene transfer. Because horizontal transfer obscures the evolutionary line of descent, phylogenetic assignment of ancestry and/or familial relationship for these elements among strains based on cladistic analysis is largely impossible.

SNPs and the use of DNA sequence differences to resolve strains. The most accurate means for strain identification and discrimination is direct examination of DNA sequences. Random nucleotide mutation, horizontal gene transfer, and/or intragenic recombination events often give rise to substitution differences in the genomes of closely related bacterial strains *(97).* Single nucleotide polymorphisms (SNPs), whether nonsynonymous (codon altering) or synonymous (producing no difference in the coded protein), provide potential targets for strain identification. Strain genotypes that are built upon SNP variation are highly amenable to evolutionary reconstruction and can be readily analyzed in a phylogenetic and population genetic context to (i) assign unknown strains into well-characterized clusters (clades), (ii) reveal closely related siblings of a particular strain, and (iii) examine the prevalence of a specific allele in a population of closely related strains. State-of-the-art methods include the utilization of comprehensive SNP collections for the comparison of whole genomic sequences among closely related bacterial strains *(42, 82)* and SNP analysis directly from genomic DNA by flow cytometry of microspheres *(81).*

DNA sequence analysis and the targeting of SNPs offer several advantages over many of the fingerprinting methods. Because each individual nucleotide is a useful genetic character, the cumulative differences in two or more sequences provide a larger number of discriminators that can be used to genotype and distinguish bacterial strains, thus allowing a more determinant and evolutionarily reliable identification. Genotypic results based on nucleotide variation are unambiguous and do not rely on secondary measurements such as amplicon length or size variations of restriction fragments; two amplicons or fragments, although identical in size, may not be chromosomally syntenic or even homologous.

These characteristics of the SNP analysis make this approach valuable in the legal arena for the microbial forensic community. This technique is well-established in human forensic casework and largely unimpeachable as a molecular determinant for forensic identification *(71).* Because most of the PCR approaches based on amplicon size have yet to be challenged in a legal setting, it is unclear whether those data would be consistently allowed in legal deliberations.

The utility of SNP analysis was highlighted recently with the characterization of a *B. anthracis* strain associated with the anthrax letters in the autumn of 2001 *(82).* SNP analysis also has been used successfully to identify and resolve clinically relevant species and strains of *Mycobacterium tuberculosis (43), Helicobacter pylori (33), E. coli (95),* and *S. enterica (31).*

For both *E. coli* and *S. enterica,* where substantial sequence variation exists between individuals, SNPs should be very informative and could be used to establish a sequence-based strain identification system. Recently, we have had success in identifying SNPs from the *E. coli* and *Shigella dysenteriae* type 1 *mutS* gene to subdifferentiate pathogenic strains into several smaller clades (Fig. 1), some of which comprise only two individuals *(12).* Current DNA-based methodologies for detection of enteric microbes have been aimed at discrimination at the genus, species, or at best the serovar or serotype level. Although important PCR methods exist that can be used to distinguish among enteropathogenic, enterohemorrhagic, enterotoxigenic, and enteroaggregative *E. coli (20, 58, 85, 95),* none of these methods are capable of distinguishing individual strains from within these groups. Restriction endonuclease–based strategies for discrimination of pathogenic *E. coli, Salmonella,* and *Shigella* are likewise lacking. Although PFGE has been used in molecular epidemiology studies, this method lacks the resolution and discriminatory power to distinguish individual strains. Its utility for forensic and legal purposes, therefore, has been questioned.

CLADISTIC ANALYSIS AND THE CULLING OF SIGNATURE NUCLEOTIDE SUBSTITUTIONS

Synapomorphies. Of particular interest is the development and application of bioinformatics and computational methodologies that will rapidly recognize SNPs and identify a strain or group of strains containing a specific SNP. Ideally, each SNP should be evolutionarily informa-

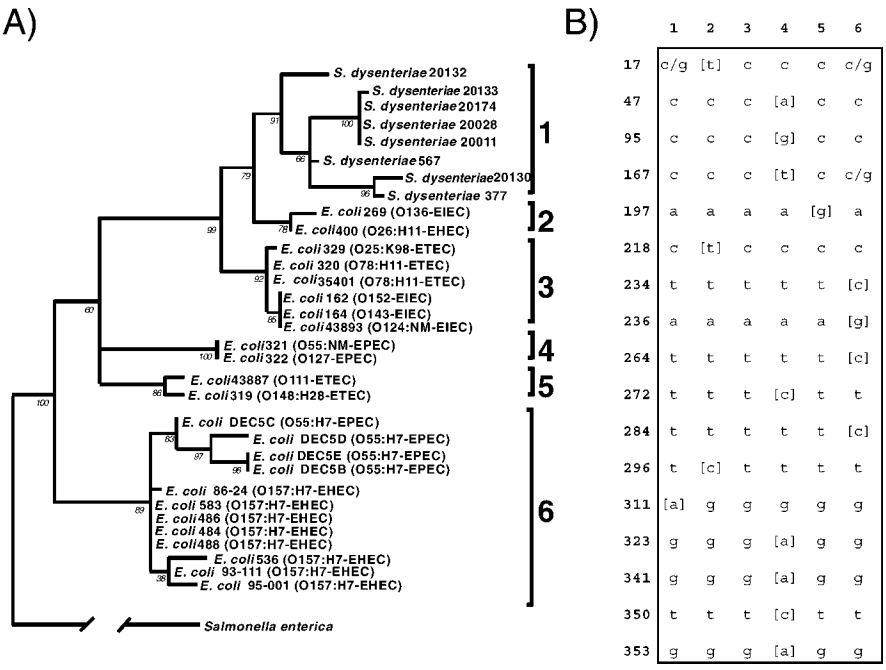

FIGURE 1. mutS *clades of pathogenic* E. coli *and* Shigella *and their defining signature synapomorphic nucleotide substitutions. (A) Cladogram of* mutS *nucleotide sequences from pathogenic* E. coli *and* S. dysenteriae *type 1 strains. Strains are organized into six distinct clades (1 through 6) that are designated by brackets to the right of the tree. Measures of clade confidence are reported below each node as bootstrap values. (B) List of signature synapomorphies (SNPs) defining each of the strain clades. The SNP defining a particular clade is bracketed. Clade numbers are listed across the top, and the nucleotide position in the alignment is listed at the left. This figure demonstrate the utility of cladistics in the identification of SNPs. (Adapted from Brown et al. (14).)*

tive, i.e., capable of defining strains in a phylogenetic context. Synapomorphies are shared derived characters (e.g., nucleotide changes) that, when found in common between two or more strains, serve as a genetic indication of the relatedness of the strains harboring them *(38).* These types of data are highly amenable to evolutionary studies of strain relatedness. Phylogenetic trees (cladograms) are built hierarchically, combining the synapomorphies as derived from the cladistic analysis of nucleotide sequence data to define the various groups (clades) of strains (see Fig. 2).

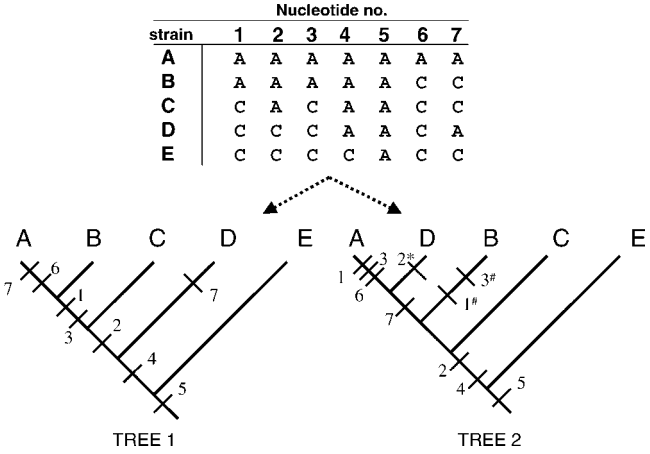

FIGURE 2. *Cladistic (maximum parsimony) concept. The nucleotide sequence matrix at the top consists of seven characters (nucleotide positions) for five hypothetical strains (A through E). These characters have been transformed into two competing phylogenetic trees (cladograms). The tree to the left (tree 1) has a length of eight steps (i.e., eight character changes). The tree to the right (tree 2) has a length of ten steps. Hence, the eight-step tree is more parsimonious and would be the preferred phylogenetic solution. Specific steps and their positions on the tree are indicated by bars. The numbers beside each bar denote specific characters in the nucleotide matrix. Parallel character changes that were mapped onto the tree twice are indicated by #, and reversed character changes are denoted by asterisks.*

Cladistic analysis. Cladistics, also known as maximum parsimony, is a method of phylogenetic analysis that can be used to develop testable hypotheses of natural relationships among bacterial strains *(38).* These hypotheses of strain relatedness can be tested mathematically and used for evolutionary discrimination of strains. Clades are defined by synapomorphies so that organisms within a clade are more closely related to each other than to members outside of this group. The clades and their synapomorphies can be illustrated by a cladogram showing the most parsimonious tree (Fig. 2). The approach follows the idea of Occam's Razor, i.e., the preferred solution is the simplest when compared with all other possible competing theories (the most parsimonious solution). Three basic premises underpin cladistic analyses for bacterial strains: (i) any group of strains is related by descent from a common ancestor, (ii) changes in characters occur over time (i.e., DNA mutates), and (iii) divergent evolution occurs between strains.

Cladistic methodology differs from other phylogenetic algorithms (i.e., nearest neighbor and likelihood approaches) in that it uses the nascent sequence data as characters and does not invoke an a priori model of gene evolution. In cladistic phylogenies, character gains and losses throughout evolution are placed on a tree, allowing for robust comparisons between distinct genes or populations to ascertain the congruence or evolutionary compatibility of various hypotheses of relationships. Because the data used for this type of cladistic analysis are the DNA sequences themselves, the approach is amenable to the concatenation of multiple gene sequences into a single data matrix *(101).* Because a multilocus sequence analysis can be constructed for an unknown strain, the addition of gene sequences to the analysis, providing additional SNPs that define groups,

allows for a substantially more refined resolution of strains. As an example, our laboratory recently used the combined sequences of three housekeeping genes (*mdh*, *icd*, and *gapA*) to fully resolve the *S. enterica* strains in the *Salmonella* Reference Collection into a tree reflective of *Salmonella* subspecies relationships *(12).*

Several programs are now available that augment cladistic analysis. WINCLADA *(67)* is a Windows-based cladistics interface that when coupled with NONA, its parsimony search engine *(39),* is a rapid and powerful analytical tool. This package is particularly useful for the analysis of larger data sets containing numerous strains and multiple gene sequences. PAUP* *(90)* supports cladistic analysis in association with detailed statistical and postanalysis tree confidence measures (e.g., bootstrap, jackknife, *g* statistic, and *F* test). The program also has other cladistic-based methods associated with it, such as the incongruence length difference (ILD) test *(35),* that test for character compatibility and evolutionary congruence among strains. PHYLIP *(36)* offers a comprehensive set of phylogenetic analysis programs, including MAXPARS, a parsimony algorithm for tree construction.

Cladistic analysis for detection of aberrant gene evolution. Cladistics is a valuable method for the study of enteric bacterial evolution. It is particularly useful in cases of enteric gene evolution where a phylogeny of DNA sequences acquired laterally will be incongruent with the phylogenies of stable housekeeping genes or with the phylogeny of the whole chromosome *(29).* Deviation from a common evolutionary pattern can serve as a signal for horizontal gene transfer and usually indicate a role for recombination of a particular sequence (Fig. 3). Our laboratory has employed cladistics as an independent and confirmatory method in detecting recombination among *mutS* alleles in pathogenic populations of *E. coli* and *Salmonella (11–14).* Cladograms can also reveal a lack of diversity in specific loci, as for the *E. coli polA* gene, where recombinational shuffling of *polA* alleles has led to a near uniform sequence in the nucleotide binding domain. The underrepresentation of third-position substitutions in this sequence suggests that a preferred *polA* allele is being transferred and maintained among feral *E. coli* strains *(14, 72).*

An additional application of cladistic methodology extends beyond the examination of individual alleles. Recently, we mapped segmental polymorphisms in the *mutS-rpoS* genomic region onto a cladogram composed of various enteric species. The most parsimonious solution was a pattern of evolution for this region that included genomic expansion by horizontal transfer events and subgenomic rearrangements of DNA from diverse origins *(54).*

Cladistics also can be used to detect incongruence within a single sequence. Intragenic fragments that retain unique histories decoupled from that of adjacent sequences can be readily identified. Specifically, the ILD test is capable of measuring the evolutionary discordance that exists within a single gene sequence *(35).* For example, using the ILD test and the sliding-window approach, the *mutS* gene from *Salmonella* was identified as an evolutionary mosaic

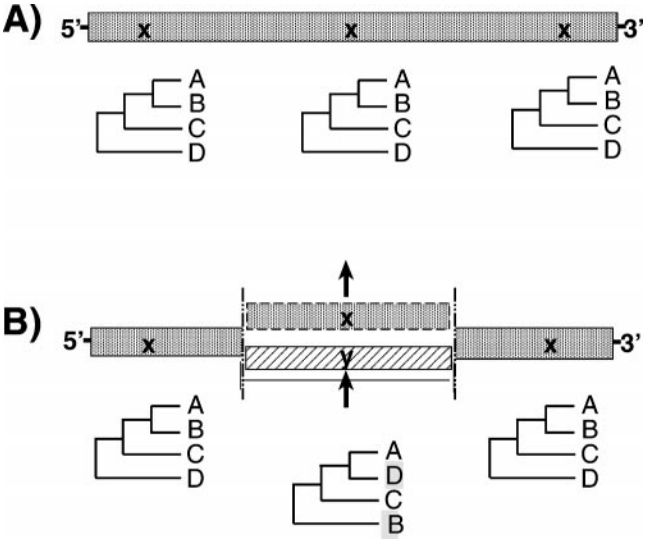

FIGURE 3. *Use of phylogenetic analysis to detect mosaic gene structure. (A) A gene (x) that has been evolving intact on the bacterial chromosome will display phylogenetic concordance when distinct intragenic sections of the sequence are compared in a phylogenetic context. Trees constructed for taxa A, B, C, and D across three separate regions of gene x are in perfect concordance. (B) The insertion of a foreign sequence (y) in place of a portion of gene x often will be detectable through phylogenetic analysis. In the region of transposed DNA, strains D and B are in discordance with their phylogenetic positions in the other regions of gene x. ILD testing can aid in delimiting the breakpoints of foreign DNA that has been inserted into a gene (12).*

composed of a recombined patch of DNA surrounded by two evolutionarily stable flanking sequences *(12).* Whether screening DNA sequences for intergenic or intragenic horizontal exchange, such discriminatory power is powerful for deducing the foreign origins of gene sequences may be important from a forensic perspective. The lack of diversity between certain strains may be pivotal in assigning forensic origins and potential reservoirs from which specific bacterial threat agents emanate.

Cladistic applications should greatly augment subtyping data as they become available for individual enteric strains. Examination of the diversity or its lack frames these data in an evolutionary context while allowing for assignments regarding siblings, ancestry, and gene stability. Forensic investigations can benefit from the ability to use cladistic methods to detect horizontal exchanges of gene sequences, either naturally occurring or manmade.

SNPs and biomarkers: a one-two forensic punch. SNPs are the forensic building blocks from which an evidentiary conclusion regarding strain identification must be made, and cladistic analysis can be used to identify and characterize those SNPs. Such applications of this technique has produced major advances in the forensic analysis of human mitochondrial sequences. The compilation of the Scientific Working Group on DNA Analysis Methods mitochondrial sequence data matrices in a phylogenetic context has allowed for the rapid identification of SNPs that can be used to define specific individuals or clades composed of a small number of individuals *(3).* Once identified

from a cladogram, the frequency of an SNP can be calculated and its relative importance in a population assessed. Such information ultimately will determine the degree to which a particular SNP can be relied upon in evidentiary proceedings.

Unlike many of the bacteria traditionally associated with biological warfare, the enteric foodborne pathogens have broader niches and consequently richer genetic diversity. In coping with a survive-or-die dictum over evolutionary time, the foodborne pathogens have encountered and overcome the manifold stresses induced by factors such as oxidative damage, pH changes, and nutrient deprivation. Thus, these pathogens have many unique phenotypic and genotypic variations that parse and distinguish the individual from otherwise closely related strains. Even when SNP analysis has resolved clades to only a limited number of foodborne isolates (e.g., 1 to 5%), secondary biomarkers can be utilized in the final delineation of a strain. Given the prevalence of antibiotic resistance among outbreak and other feral foodborne enteric strains *(45),* antibiotic resistance phenotypic patterns could be useful for resolving intractable clades or could be further exploited at the sequence level. For example, because widely different virulence factor profiles exist among pathogenic populations of *E. coli (50),* differences in these profiles could be extremely useful for developing individual strain identification strategies. Likewise, subtle metabolic differences can now be exploited using automated systems to discriminate strains. Phenotype microarray systems (e.g., Biolog, Hayward, Calif.) are now available for screening of 2,000 metabolic substrates. Given the vast number of genes associated with metabolic function within the cell, such a system holds promise for differentiating closely related strains based on alterations of one or more of these pathways *(10).* The horizontally derived segmental genomic differences that differentiate many pathogen populations should not be overlooked as potential targets for the development of unique and rapid strain identification systems. This approach seems especially promising for pathogenic *E. coli,* where numerous segmental differences have been identified between and among individual strains *(54, 73).*

Confounding effects and practical uses of bacterial mutators. Given the need to establish a solid evidentiary base for SNP analysis within foodborne pathogen populations, it is important not to overlook the potential impact of mechanisms that can accelerate bacterial evolution. Understanding these effects is critical to the application of SNPs to forensic analysis. Bacterial mutators (cells capable of enhancing the bacterial mutation rate) can invade and propagate in populations of normal bacteria *(87).* Such mutators, which constitute more than 1 in 100 isolates of *E. coli* and *Salmonella (55),* can accelerate mutation rates 100- to 1,000-fold. In addition to the dramatic effect that mutators have on the basal mutation rate of the bacterial genome, methyl-directed mismatch repair–deficient mutators may also play important roles in enhancing homologous recombination between and among bacterial strains and species *(16, 17, 55, 79).* Estimates indicate that mutators can accelerate bacterial adaptation from acquired DNA of up to 30% divergence *(91).* Although the effects of mutators on recoverable mutations are minimal in continuous cultures (>20,000 generations) of *E. coli (56),* it is unclear what effect mutators have on the genomes of progeny bacteria that persist within an infected host. Selection pressures in vivo could have more determinant effects on mutator biology and alter the interplay between mutators and the populations in which they reside.

Mutators are a double-edged sword for the forensic community. Mutators might confound accurate interpretation of the cladistic relatedness and DNA sequence identity among strains, but they also may delimit the upper limits of diversifying mutations that occur as a function of in vivo growth and selection, providing a worst-case scenario. A typical bacterium falls well short of accruing the levels of nucleotide diversity that can collect within the mutator over the same time period. Thus, mutator studies afford the forensic scientist unique opportunities for evaluating whether evidence linking an individual strain with an intentional outbreak is incontrovertible.

DETECTION OF FOODBORNE VIRAL PATHOGENS

Within the context of accelerated evolution, the RNA viruses, in particular the two most common groups of foodborne viral pathogens, hepatitis A virus (HAV) and Norwalk-like viruses (NLVs), are the yardstick against which broad genomic diversity can be measured. Detection of HAV or NLVs by infectivity assays is currently not possible because of the lack of a cell culture host for human NLVs and the slow and noncytopathic replication of wild-type HAV strains. Detection of these viruses normally utilizes reverse transcription (RT) coupled to PCR *(5, 41, 49).* Identification of genetic variants of a known pathogen such as HAV is possible by RT-PCR but is predicated upon post-PCR operations such as DNA sequencing or single-strand conformation polymorphism *(40).* Another drawback in using PCR or any other nucleic acid sequence–based detection of viral pathogens in food or water is the apparent lack of correlation between the PCR signal and the infectivity of the virus *(57, 68, 86).* However, a methodology has been proposed to overcome this deficiency *(7),* and criteria have been established to discriminate between infectious and inactivated HAV, a necessary requisite for making a nucleic acid–based detection method relevant to public health concerns.

For this method, a set of RT-PCR primers is positioned to detect damage in template RNA induced by common chemical or physical agents (e.g., heat or UV irradiation), rendering it unsuitable for RT. Priming with oligo$(dT)_{15}$ and amplification of virus-specific sequence is then initiated using a collection of primers directed to the 5′ end of the viral genome. Any damage to the viral genomic RNA due to heat or UV treatment results in a truncated cDNA, which cannot be amplified by primers targeting the 5′ end. The correlation between residual infectivity following UV treatment and the intensity of PCR signal is excellent, with no PCR signal detectable for inactivated virus. These experi-

ments indicate that any detection of an RT-PCR signal in suspect samples is cause for public health concern.

HIGH-THROUGHPUT GENOTYPING: METHODS AVAILABLE

Microarray-based genome composition analysis. Recent advances in technology are allowing the categorization of bacteria based on their genetic makeup rather than their phenotypic markers. This genotype information provides a deeper insight into the evolutionary relationships between species that are phenotypically indistinguishable. DNA microarray technology is increasing in popularity as a technique for investigating genetic relationships among closely related organisms.

Microarray-based genome composition analysis is typically performed by hybridization of labeled genomic DNA from the organism being investigated to an array of DNA probes representing all or most of the genes present in a reference organism. These arrays have typically been constructed by PCR amplification of entire open reading frames that have been annotated from an organism's sequenced genome and therefore can provide a comparison between organisms at the resolution of a single gene. A number of organisms have been analyzed using this method, including *Salmonella (18, 75–77), H. pylori (9, 48, 84), Campylobacter (28),* mycobacteria *(6), Staphylococcus aureus (37),* obligate endosymbionts of the tsetse fly *(1, 2), Shewanella oneidensis (63), Pseudomonas* species *(22),* and *Vibrio cholerae (30).*

Decoding of the complete genome sequence of the foodborne pathogen *S. enterica* serovar Typhimurium strain LT2 *(62)* allowed researchers to construct an LT2 microarray of PCR-amplified open reading frames that represented over 97% of the annotated coding sequences from this organism. With this array, the genetic content of the entire *Salmonella* clade was surveyed with respect to the *Salmonella* Typhimurium LT2 genome. Results from these analyses were in agreement with the phylogenetic relationships predicted from sequence data available from housekeeping or invasion genes from each of these *Salmonella* serovars.

Although microarray-based genome composition assays are very powerful for obtaining information on the genome composition of many strains of a given species, this technique can be used to categorize genes only as either present or divergent, based on the amount of signal indicating hybridization to the reference strain *(52).* Any genes unique to the strain under investigation will go unnoticed using this method. Furthermore, the resolution provided by this technique is limited to a single gene, and therefore this assay cannot be used to discriminate between identical strains that differ by only an SNP.

Resequencing with high-density microarrays. The most accurate method for strain identification and discrimination remains determination of the nucleotide sequence of the DNA itself. Oligonucleotide microarray-based hybridization analysis is a powerful technique that provides a rapid and cost-effective analysis of all possible mutations and sequence variations in genomic DNA. One application of this technology, known as resequencing, is a powerful tool for searching large sections of a microbial genome for the presence of SNPs that could serve as markers for further cladistic analyses. The increasing availability of high-density oligonucleotide arrays should allows for the resequencing of tens of thousands of nucleotide positions in parallel on a single chip *(19, 59).* Most chip-based sequence analyses have been carried out with arrays designed to evaluate specific sequences and have been produced most often by using a photolithographic method developed by Affymetrix, Inc. (Santa Clara, Calif.).

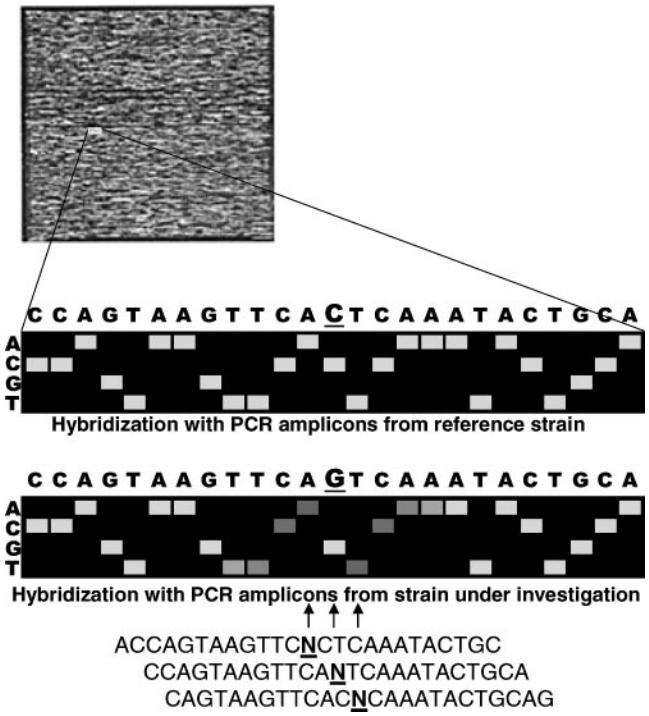

FIGURE 4. *Resequencing with DNA microarrays. High-density oligonucleotide-based microarray chip (Affymetrix) carrying ~10^5 25-nucleotide probe sequences. Genomic regions being investigated are first amplified by PCR in the presence of a fluorescently tagged dNTP. Long PCR amplicons are fragmented and denatured to maximize hybridization efficiency. The maximum hybridization signal will be observed for perfectly matched duplexes. SNPs are identified by a loss of hybridization signal resulting from the single mismatch. The exact position and identity of the SNP is determined by the increased hybridization signal observed from the single probe that forms a perfect match (59).*

In a typical assay (Fig. 4) designed to search both strands of a target DNA sequence (of length *N*) for all possible single nucleotide substitutions, an array consisting of 8*N* probes, typically 20 to 25 nucleotides in length, would be required. For each position within each strand, four probes are designed to differ at only one position and represent all possible substitutions at this position (G, A, T, or C). These probes are designed such that the location of the base under investigation is in the center position of the potential target-probe hybrid. This design allows for the best discrimination for hybridization specificity *(60).* With this technique, therefore, an analysis of all possible nucleotide substitutions at every position on both strands of a 30-kb DNA target would require 240,000 probes; standard

photolithographic techniques routinely yield 500,000 to 700,000 probes in a small area (1 to 1.5 cm^2). To maximize the hybridization efficiency between target and probe, target DNA amplification is designed to yield many small (approximately 200 bp) amplicons or several long PCR amplicons that are then fragmented prior to hybridization.

Polymerase-based single base extension methodologies. Single base extension (SBE), or minisequencing, derives from the ability of DNA polymerase to extend a DNA strand from 5′ to 3′ in a template-directed manner. Based on this simple principle, DNA polymerase-based SBE assays have been designed to allow SNP genotyping. This technique has been applied in many diverse genotyping assays, many of which have become available commercially. SBE assays are performed initially in a manner similar to that described for resequencing assays. The allele of interest is first amplified by PCR, and this amplicon is then hybridized to a complementary oligonucleotide probe that ends immediately adjacent to the position of the SNP being investigated. The probe is then extended by a single nucleotide by adding DNA polymerase and dideoxy nucleoside triphosphates (ddNTPs). The missing 3′-hydroxyl groups cause these ddNTPs to act as terminators and therefore prevent the primer from extending more that one nucleotide. The $N + 1$ oligonucleotides can then be analyzed by several methods that can determine the identity of the incorporated nucleotide and therefore reveal the presence or absence of an SNP at that position (Fig. 5).

Various methods are available for determining the identity of the extended oligonucleotide, including visualization with a confocal scanner (microarray scanning) and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry *(44).* For visualization via confocal scanning, the oligonucleotide probes can be arrayed onto a glass slide at high density (500,000 to 700,000 unique probes per slide). In this, the SBE reactions would be carried out in the presence of fluorescently tagged ddNTPs such that each of the four ddNTPs added (G, A, T, or C) will fluoresce at a different wavelength and therefore emit light that is characteristic of the nucleotide incorporated. This array-based SBE technique is very robust, has a high success rate, and allows screening for many different SNPs in parallel.

For mass spectrometry-based SBE detection, oligonucleotide probes are not arrayed onto glass slides but rather used in 96-well or 384-well plates that allow high-throughput robotic automation. These probes typically are modified at their 5′ end to include a streptavidin moiety. Following PCR amplification of the allele of interest, the amplicon is denatured and annealed to the oligonucleotide probe, and the enzymatic SBE reaction is carried out. The ddNTPs used for these techniques need not be tagged with fluorophores because mass spectrometry detection is based on the difference in the molecular weight of the N + 1 extended probe. The slight differences in the molecular weights of ddATP, ddTTP, ddCTP, and ddGTP can be quickly (fractions of a second) and easily detected by mass spectrometry. Another benefit of mass spectrometry detection is its

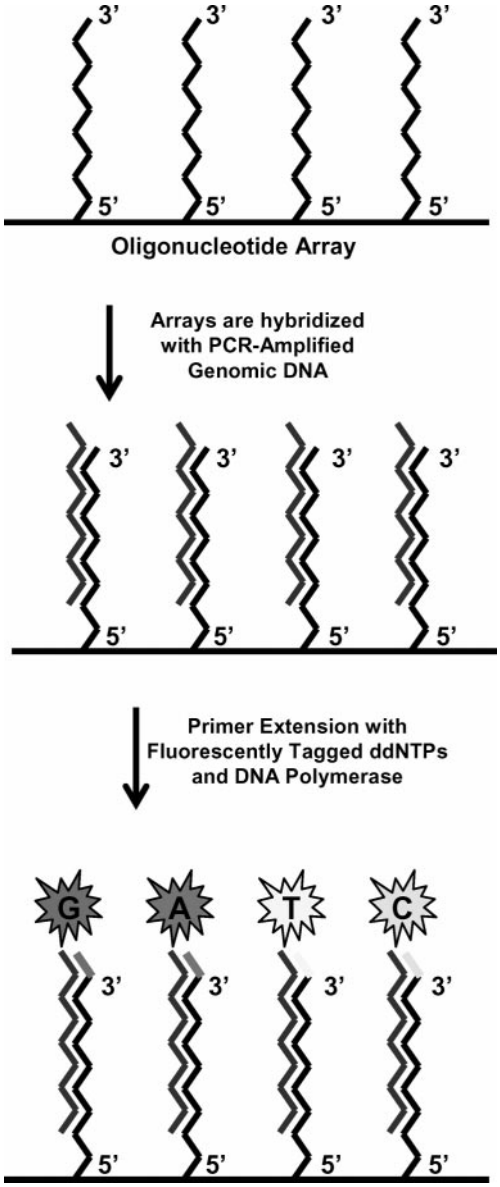

FIGURE 5. *Single-base extension assay. Oligonucleotide probes are hybridized with PCR-amplified denatured DNA. These perfect match probes end immediately adjacent to the position of the SNP being investigated. The DNA polymerase–based single base extension is carried out in the presence of ddNTPs. Array-based SBE assays utilize fluorescently tagged ddNTPs that allow for the subsequent identification of the incorporated nucleotide using a microarray scanner. Single base extensions using unlabeled ddNTPs can be analyzed via MALDI-TOF mass spectrometry to identify the SNP of interest.*

quantification. Among the technologies tested for allele frequency estimation, mass spectrometry detection appears to be one of the most sensitive. Although potentially capable of delivering 10,000 to 50,000 genotypes per day at minimal cost, initial set-up fees are significant and have limited the widespread use of this technique.

Sequencing by synthesis: pyrosequencing. Pyrosequencing is a fairly new technique for investigating short stretches of DNA by synchronized primer extension. Marketed by Pyrosequencing AB (Uppsala, Sweden), this clev-

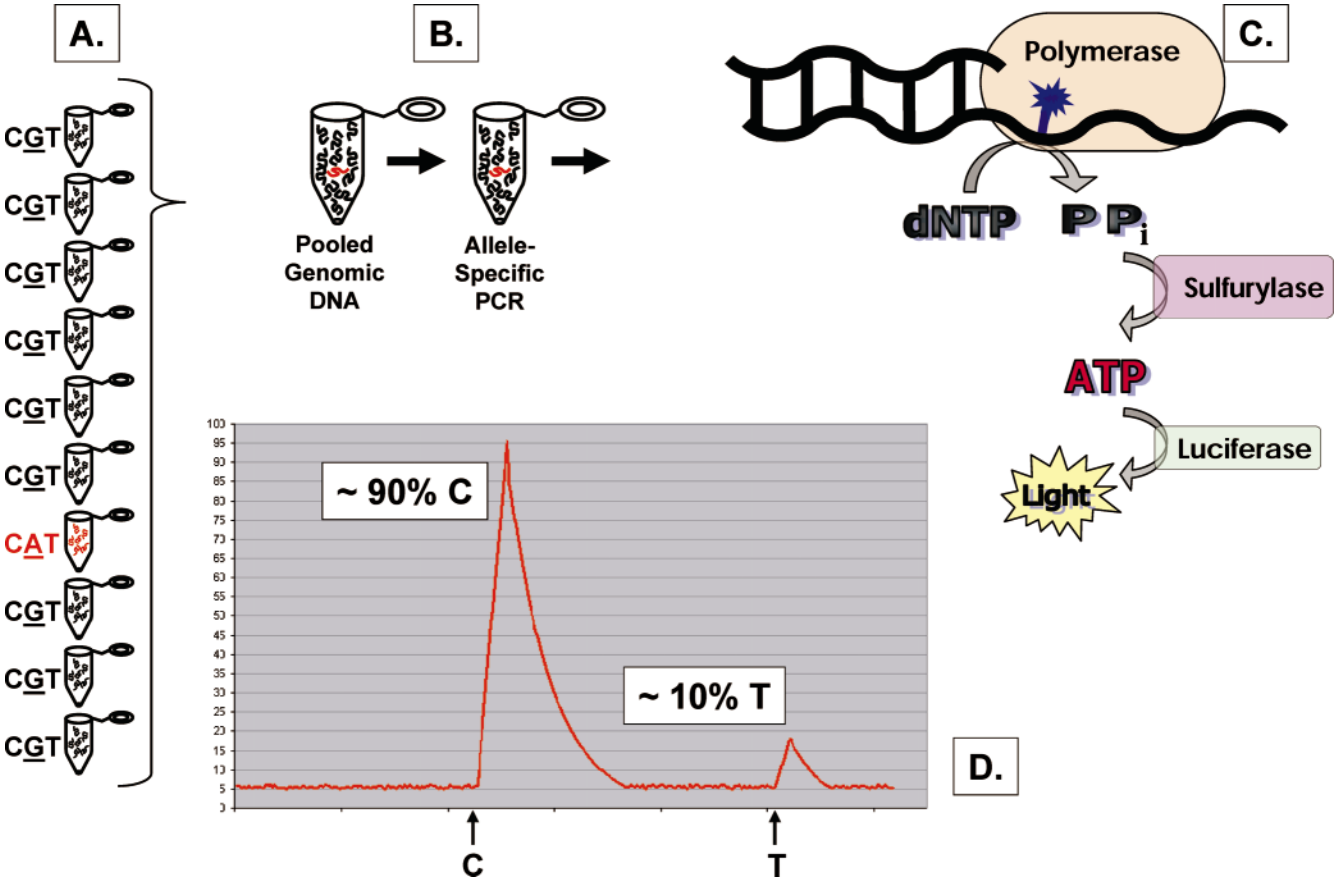

FIGURE 6. *Estimation of allele frequency via pyrosequencing. SNP analysis can be performed from pooled sample populations. (A) Individual isolates of a particular strain of interest are cultured, and genomic DNA is extracted. In this example, 1 of 10 samples carries a G-A single nucleotide polymorphism within a particular allele. (B) Equal amounts of genomic DNA from each sample are pooled into a single tube and amplified by PCR using primers specific for the allele of interest. (C) Following PCR, pyrosequencing is performed on the pooled PCR sample. Pyrosequencing utilizes the pyrophosphate (PP_i) produced as a by-product of DNA synthesis to generate a quantifiable chemiluminescent signal. (D) A pyrogram displays the quantity of light emitted following the sequential addition of dCTP and dTTP to the pyrosequencing reaction. The relationship between the peak height and number of nucleotides incorporated at any position allows accurate determination of the allele frequency at any position.*

erly designed method consists of a careful balance of enzyme activities that use pyrophosphate (a normal by-product of DNA synthesis) to generate ATP, which is a substrate for the enzyme luciferase and results in a quantifiable bioluminescent signal *(83)*.

As for other methods described, the allele(s) of interest is first amplified by PCR. The amplicons are denatured, and a sequencing primer is annealed to the single-strand PCR-amplified DNA template. This DNA substrate is then added to a reaction cocktail consisting of the enzymes DNA polymerase, ATP sulfurylase, apyrase, and luciferase. The primer extension reaction is initiated by adding the first of four dNTPs to the reaction. The DNA polymerase incorporates the deoxyribonucleotide into the DNA strand if it is complementary to the PCR-amplified template strand. Each incorporation event results in the release of an equimolar amount of pyrophosphate. ATP sulfhydrase quickly and quantitatively converts this pyrophosphate into ATP in the presence of adenosine 5′ phosphosulfate. This ATP is then used by luciferase to convert luciferin to oxyluciferin and, in so doing, generates stoichiometric amounts of visible light. A charged couple device camera detects the light produced by this reaction, which is recorded as a peak in a pyrogram (see Fig. 6). Unincorporated dNTPs are quickly degraded by apyrase, quickly extinguishing the light, and the reaction mixture is regenerated. The next dNTP is now ready to be added, and the series of reactions is repeated for each dNTP addition.

For bacterial typing and strain identification, pyrosequencing has already proved itself a powerful tool with many advantages over other methods. Primers can be designed to flank regions known to contain polymorphisms and real time sequence data can be acquired from many reactions performed in parallel in as little as 10 min. Pyrosequencing has been used to group 24 naturally isolated strains of different *Listeria monocytogenes* serovars by sequence analysis of a short region of the *inlB* gene covering the two variable positions *(78)*.

Data acquired with this method reflect underlying allele frequencies of SNPs in a very accurate and reproducible way. The stoichiometric relationship between peak height (displayed on the pyrogram) and the number of nucleotide

incorporations allows allele frequencies in a large population to be determined. These allele frequencies can be determined from pools of PCR products and from pooled populations of genomic DNA samples, therefore requiring only a single PCR amplification step followed by a single sequencing reaction for each region of interest. Allele frequencies as low as 2% can be accurately detected using this technique *(42, 66).*

MICROARRAY-BASED DETECTION OF FOODBORNE PATHOGENS

Bacterial pathogens. The application of DNA microarrays for the detection and identification of bacterial pathogens in the food supply is in its infancy, but useful technologies have been developed. These oligonucleotide-based microarrays provide a means for detecting and discriminating among strains of bacteria based on the presence of short DNA sequences that are unique for each organism. Short oligonucleotides (20 to 70 nucleotides) arrayed onto glass slides can potentially probe for the presence or absence of multiple unique gene sequences within a single bacterial strain. Furthermore, the density at which these probes can be arrayed allows simultaneously screening for many unique gene sequences from many different species in parallel.

Two general approaches using this technique have been developed and applied by several laboratories to detect many prominent human intestinal bacteria. In one approach, a single universal gene is probed for the presence of polymorphic regions that can be used to identify different species. The most widely used example of this approach is the rDNA genes found in all organisms. Historically, the 16S rDNA gene sequences have been used to identify and classify different species of bacteria. The presence of highly conserved regions within the 16S rDNA molecule allows amplification of this gene in many organisms using a single universal primer pair. A fluorescently labeled 16S rDNA amplicon can then be hybridized to an oligonucleotide array that contains multiple probes that uniquely identify variable regions within the rDNA strand that differ in bacteria of different species. This method has been used to detect 20 different human intestinal bacteria from fecal samples *(96)*: *Bacteroides thetaiotaomicron, Bacteroides vulgatus, Bacteroides fragilis, Bacteroides distasonis, Clostridium clostridiiforme, Clostridium leptum, Fusobacterium prausnitzii, Peptostreptococcus productus, Ruminococcus obeum, Ruminococcus bromii, Ruminococcus callidus, Ruminococcus albus, Bifidobacterium longum, Bifidobacterium adolescentis, Bifidobacterium infantis, Eubacterium biforme, Eubacterium aerofaciens, Lactobacillus acidophilus, E. coli,* and *Enterococcus faecium.* Two universal primers were able to amplify the entire 16S rDNA molecule from all 20 bacterial species tested, and three oligonucleotide probes specific for 16S rDNA sequences from each bacterial species were sufficient for successful identification. This technique is useful for detecting multiple types of bacteria with minimal preparation. With the availability of high-density oligonucleotide microarrays, a single array could be used to detect thousands of different bacterial species based on their rDNA sequences.

In a second approach, arrays of oligonucleotides can be generated for detecting genes encoding bacterial antigenic determinants and virulence factors from different bacterial strains. This method could potentially replace many of the current multiplex PCR-based assays designed to amplify regions unique to an individual serovar *(21).* For the detection of *E. coli* O157:H7 for example, several of the common target genes for multiplex PCR are the conserved regions of *slt-I, slt-II,* and *eaeA (64, 102),* which mediate the adherence of the organism to host cells. However, because these regions are not unique to *E. coli* O157:H7, several more specific target genes have been used, including *rfbE* and *fliC,* which are involved in the biosynthesis of the O157 and H7 antigens, respectively *(8, 26, 47).* As an alternative to multiplex PCR, genomic DNA purified from an individual serotype or from pools of different serotypes could be fluorescently labeled and hybridized to a microarray that contains one or more oligonucleotide probes complementary to these unique regions. Oligonucleotide microarrays containing probe sequences complementary to DNA regions encoding serotype-specific antigens from many different serotypes could be constructed to allow for the detection of many different strains in parallel.

Viral pathogens. Microarray-based detection of viral foodborne pathogens has lagged behind that for bacterial pathogens. As with bacterial pathogens, microarray-based detection of viruses when developed as a multicomponent system along with methods for the concentration and amplification of multiple viral genomes in a single sample will significantly enhance our ability to detect and respond to emergencies involving major viral diseases. To accomplish these aims, we are developing a set of RT-PCR primers for the global and sequence-independent amplification of all nucleic acid molecules (including viral nucleic acids) in a sample and then for detection of specific virus strains based on oligoprobe hybridization in a microarray. Such amplification is made possible by adapting either anti-sense RNA (aRNA) amplification technology *(93)* (Ambion, Inc., Austin, Tex.) or PCR-based amplification strategies. Both processes involve synthesis of cDNA using T7 oligo$(dT)_{15}$ for the RT step, followed by conversion of the cDNA to double-stranded form using DNA polymerase. For the aRNA approach, double-stranded cDNA synthesis is carried out using a combination of RNase H to create nicks in the RNA-cDNA hybrid and is followed by nick translation with DNA polymerase. Multiple copies of RNA are then synthesized using T7 RNA polymerase and a mixture of ribonucleoside triphosphates containing one or more labeled nucleotides. For the PCR-based approach, double-stranded cDNA synthesis is conducted using tagged random hexamers as the primer for second strand synthesis. PCR amplification is then performed using T7 oligo$(dT)_{15}$ and T7N6 as primers. By the incorporation of biotin-labeled (deoxy)nucleotides during the amplification step, the resulting mixture would be a highly representative population of biotin-labeled RNA or DNA. Specific sequences (e.g.,

HAV or NLV) are detected by hybridization of the RNA or DNA to an oligonucleotide filter array containing immobilized complementary sequences, and the biotin label is detected with a streptavidin conjugate, further increasing sensitivity. If successful, the procedure could be adapted to glass microarrays and to detection of hybridized probe by fluorescence. This regimen would achieve identification at the level of virus strain, as currently envisioned for bacterial pathogens.

FOOD SECURITY AND FOOD SAFETY JOINED

In the aftermath of 11 September 2001, food security measures have proceeded in parallel with food safety procedures to ensure that our nation's food supply remains the safest and most wholesome worldwide. The possibility of intentional contamination of our food supply makes it necessary to include a microbial forensic component in our existing system. Forensics would allow prompt pathogen identification at the strain level, thus permitting proper attribution of the contaminating agent. Although recent efforts have emphasized the select agents that may be introduced intentionally into foods, foodborne pathogens are just as likely to be used as agents of bioterrorism. As pointed out in a *Science* News Focus, others share this thought even at a time when the lists of bioterrorist weapons do not include agents such as *E. coli* and *Salmonella (32).*

Although the goal of forensic discrimination of *E. coli, Salmonella,* and the other typical foodborne pathogens seems daunting, the task may be more readily achievable than it appears. We are dealing with versatile enteric organisms, each carrying unique sets of phenotypic characters, traits that help define distinct niches. These phenotypic properties have been exploited in the food safety arena for years. Genotypic signatures that have accrued over millions of years of evolution are now accessible by a variety of molecular techniques. Those techniques are likewise being exploited in genotyping protocols. The combinatorial power of these phenotypic and genotypic methods should allow the unique personalities of individual strains to be revealed. The analysis of these traits, both phenotypic and molecular, within a cladistic framework would elucidate to the unique fingerprints of microbes that could be used for purposes of bioterrorism. The database obtained would serve the needs of the microbial forensic community should a suspect strain have to be identified within statistical certainty. The database will also be helpful for public health laboratories and other food safety entities in their continual surveillance to ensure a safe food supply.

REFERENCES

1. Akman, L., and S. Aksoy. 2001. A novel application of gene arrays: *Escherichia coli* array provides insight into the biology of the obligate endosymbiont of tsetse flies. *Proc. Natl. Acad. Sci. USA* 98: 7546–7551.
2. Akman, L., R. V. Rio, C. B. Beard, and S. Aksoy. 2001. Genome size determination and coding capacity of *Sodalis glossinidius,* an enteric symbiont of tsetse flies, as revealed by hybridization to *Escherichia coli* gene arrays. *J. Bacteriol.* 183:4517–4525.
3. Allard, M. W., K. Miller, M. Wilson, K. Monson, and B. Budowle. 2002. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. *J. Forensic Sci.* 47:1215–1223.
4. Armstrong, G. L., J. Hollingsworth, and J. G. Morris, Jr. 1996. Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model entry of a new pathogen into the food supply of the developed world. *Epidemiol. Rev.* 18:29–51.
5. Atmar, R. L., T. G. Metcalf, F. H. Neill, and M. K. Estes. 1993. Detection of enteric viruses in oysters by using the polymerase chain reaction. *Appl. Environ. Microbiol.* 59:631–635.
6. Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* 284: 1520–1523.
7. Bhattacharya, S. S., M. Kulka, K. A. Lampel, T. A. Cebula, and B. B. Goswami. 2004. Use of reverse transcriptase and PCR to discriminate between infectious and non-infectious hepatitis A virus. *J. Virol. Meth.* 116:181–187.
8. Bilge, S. S., J. C. Vary, Jr., S. F. Dowell, and P. I. Tarr. 1996. Role of the *Escherichia coli* O157:H7 O side chain in adherence and analysis of an *rfb* locus. *Infect. Immun.* 64:4795–4801.
9. Bjorkholm, B., A. Lundin, A. Sillen, K. Guillemin, N. Salama, C. Rubio, J. I. Gordon, P. Falk, and L. Engstrand. 2001. Comparison of genetic divergence and fitness between two subclones of *Helicobacter pylori. Infect. Immun.* 69:7832–7838.
10. Bochner, B. R. 2003. New technologies to assess genotype-phenotype relationships. *Nat. Rev. Genet.* 4:309–314.
11. Brown, E. W. 2001. Molecular differentiation of bacterial strains, p. 29–66. *In* M. Carrington and A. R. Hoelzel (ed.), Molecular epidemiology: a practical approach. Oxford University Press, Oxford.
12. Brown, E. W., M. L. Kotewicz, and T. A. Cebula. 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. *Mol. Phylogenet. Evol.* 24:102–120.
13. Brown, E. W., M. L. Kotewicz, J. E. LeClerc, and T. A. Cebula. 2001. Three R's of bacterial evolution: how replication, repair, and recombination frame the origin of species. *Environ. Mol. Mutagen.* 38:248–260.
14. Brown, E. W., J. E. LeClerc, B. Li, W. L. Payne, and T. A. Cebula. 2001. Phylogenetic evidence for horizontal transfer of *mutS* alleles among naturally occurring *Escherichia coli* strains. *J. Bacteriol.* 183:1631–1644.
15. Burns, D. N., R. J. Wallace, Jr., M. E. Schultz, Y. S. Zhang, S. Q. Zubairi, Y. Pang, C. L. Gilbert, B. A. Brown, E. S. Noel, and F. M. Gordin. 1991. Nosocomial outbreak of respiratory tract colonization with *Mycobacterium fortuitum*: demonstration of the usefulness of pulsed-field-gel-electrophoresis in an epidemiologic investigation. *Am. Rev. Respir. Dis.* 144:1153–1159.
16. Cebula, T. A., and J. E. LeClerc. 1997. To be a mutator, or how pathogenic and commensal bacteria can evolve rapidly. *Trends Microbiol.* 5:428–429.
17. Cebula, T. A., and J. E. LeClerc. 2000. DNA repair and mutators: effects on antigenic variation and virulence of bacterial pathogens, p. 143–159. *In* K. A. Brogden, J. A. Roth, T. B. Stanton, C. Bolin, F. Minion, and M. J. Wannemuehler (ed.), Virulence mechanisms of bacterial pathogens, 3rd ed. ASM Press, Washington, D.C.
18. Chan, K., S. Baker, C. C. Kim, C. S. Detweiler, G. Dougan, and S. Falkow. 2003. Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar Typhimurium DNA microarray. *J. Bacteriol.* 185:553–563.
19. Chee, M., R. Yang, E. Hubbell, A. Berno, X. C. Huang, D. Stern, J. Winkler, D. J. Lockhart, M. S. Morris, and S. P. Fodor. 1996. Accessing genetic information with high-density DNA arrays. *Science* 274:610–614.
20. Chen, J., and M. W. Griffiths. 2001. Detection of *Salmonella* and simultaneous detection of *Salmonella* and Shiga-like toxin–producing *Escherichia coli* using the magnetic capture hybridization polymerase chain reaction. *Lett. Appl. Microbiol.* 32:7–11.
21. Chizhikov, V., A. Rasooly, K. Chumakov, and D. D. Levy. 2001.

Microarray analysis of microbial virulence factors. *Appl. Environ. Microbiol.* 67:3258–3263.
22. Cho, J. C., and J. M. Tiedje. 2001. Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl. Environ. Microbiol.* 67:3677–3682.
23. Crichton, P. B., and D. C. Old. 1980. Differentiation of strains of *Escherichia coli*: multiple typing approach. *J. Clin. Microbiol.* 11: 635–640.
24. Crowhurst, R. N., B. T. Hawthorne, E. H. Rikkerink, and M. D. Templeton. 1991. Differentiation of *Fusarium solani* f. sp. *Cucurbitae* races 1 and 2 by random amplification of polymorphic DNA. *Curr. Genet.* 20:391–396.
25. Davis, M. A., D. D. Hancock, T. E. Besser, and D. R. Call. 2003. Evaluation of pulsed-field-gel-electrophoresis as a tool for determining the degree of genetic relatedness between strains of *Escherichia coli* O157:H7. *J. Clin. Microbiol.* 41:1843–1849.
26. Desmarchelier, P. M., S. S. Bilge, N. Fegan, L. Mills, J. C. Vary, Jr., and P. I. Tarr. 1998. A PCR specific for *Escherichia coli* O157 based on the *rfb* locus encoding O157 lipopolysaccharide. *J. Clin. Microbiol.* 36:1801–1804.
27. Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Appl. Environ. Microbiol.* 66:2572–2577.
28. Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren. 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.* 11:1706–1715.
29. Dykhuizen, D. E., and L. Green. 1991. Recombination in *Escherichia coli* and the definition of bacterial species. *J. Bacteriol.* 173: 7257–7268.
30. Dziejman, M., E. Balon, D. Boyd, C. M. Fraser, J. F. Heidelberg, and J. J. Mekalanos. 2002. Comparative genomic analysis of *Vibrio cholerae*: genes that correlate with cholera endemic and pandemic disease. *Proc. Natl. Acad. Sci. USA* 99:1556–1561.
31. Elder, R. O., G. E. Duhame, M. R. Matheison, E. D. Erickson, C. J. Gebhart, and R. D. Oberst. 1997. Multiplex polymerase chain reaction for simultaneous detection of *Lawsonia intracellularis*, *Serpulina hyodysenteriae*, and salmonellae in porcine intestinal specimens. *J. Vet. Diagn. Invest.* 9:281–286.
32. Enserink, M. 2003. FBI's top scientist takes the lead in forensic biology. *Science* 300:41–43.
33. Falush, D., T. Wirth, B. Linz, J. K. Pritchard, M. Stephens, M. Kidd, M. J. Blaser, D. Y. Graham, S. Vacher, G. I. Perez-Perez, Y. Yamaoka, F. Megraud, K. Otto, U. Reichard, E. Katzowitsch, X. Wang, M. Achtman, and S. Suerbaum. 2003. Traces of human migrations in *Helicobacter pylori* populations. *Science* 299:1582–1585.
34. Farlow, J., K. L. Smith, J. Wong, M. Abrams, M. Lytle, and P. Keim. 2001. *Francisella tularensis* strain typing using multiple-locus, variable-number tandem repeat analysis. *J. Clin. Microbiol.* 39:3186–3192.
35. Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. *Cladistics* 10:315–319.
36. Felsenstein, J. 1993. PHYLIP (phylogeny inference package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
37. Fitzgerald, J. R., D. E. Sturdevant, S. M. Mackie, S. R. Gill, and J. M. Musser. 2001. Evolutionary genomics of *Staphylococcus aureus*: insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. *Proc. Natl. Acad. Sci. USA* 98: 8821–8826.
38. Forey, P. L., C. J. Humphries, I. Kitching, R. W. Scotland, D. J. Siebert, and D. Williams. 1992. Cladistics: a practical course in systematics. Clarendon Press, Oxford.
39. Goloboff, P. 1994. Nona: a tree search program. Available at: www.cladistics.org. Accessed February 2005.
40. Goswami, B. B., W. Burkhardt, and T. A. Cebula. 1997. Identification of genetic variants of hepatitis A virus. *J. Virol. Methods* 65:95–103.
41. Goswami, B. B., W. H. Koch, and T. A. Cebula. 1993. Detection of hepatitis A virus in *Mercenaria mercenaria* by coupled reverse transcription and polymerase chain reaction. *Appl. Environ. Microbiol.* 59:2765–2770.
42. Gruber, J. D., P. B. Colligan, and J. K. Wolford. 2002. Estimation of single nucleotide polymorphism allele frequency in DNA pools by using pyrosequencing. *Hum. Genet.* 110:395–401.
43. Gutacker, M. M., L. C. Smoot, C. A. Lux Migliaccio, S. M. Ricklefs, S. Hua, D. V. Cousins, E. A. Graviss, E. Shashkina, B. N. Kreiswirth, and J. M. Musser. 2002. Genome-wide analysis of synonymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* 162:1533–1543.
44. Haff, L. A., and I. P. Smirnov. 1997. Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res.* 7:378–388.
45. Hartman, A. B., I. I. Essiet, D. W. Isenbarger, and L. E. Lindler. 2003. Epidemiology of tetracycline resistance determinants in *Shigella* spp. and enteroinvasive *Escherichia coli*: characterization and dissemination of tet(A)-1. *J. Clin. Microbiol.* 41:1023–1032.
46. Hilton, A. C., D. Mortiboy, J. G. Banks, and C. W. Penn. 1997. RAPD analysis of environmental, food, and clinical isolates of *Campylobacter* spp. *FEMS Immunol. Med. Microbiol.* 18:119–124.
47. Hu, Y., Q. Zhang, and J. C. Meitzler. 1999. Rapid and sensitive detection of *Escherichia coli* O157:H7 in bovine feces by a multiplex PCR. *J. Appl. Microbiol.* 87:867–876.
48. Israel, D. A., N. Salama, U. Krishna, U. M. Rieger, J. C. Atherton, S. Falkow, and R. M. Peek, Jr. 2001. *Helicobacter pylori* genetic diversity within the gastric niche of a single human host. *Proc. Natl. Acad. Sci. USA* 98:14625–14630.
49. Jiang, X., J. Wang, D. Y. Graham, and M. K. Estes. 1992. Detection of Norwalk virus in stools using the polymerase chain reaction. *J. Clin. Microbiol.* 30:2529–2534.
50. Johnson, J. R., T. T. O'Bryan, M. Kuskowski, and J. N. Maslow. 2001. Ongoing horizontal and vertical transmission of virulence genes and *papA* alleles among *Escherichia coli* blood isolates from patients with diverse-source bacteremia. *Infect. Immun.* 69:5363–5374.
51. Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka, P. J. Jackson, and M. E. Hugh-Jones. 2000. Multiple-locus, variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J. Bacteriol.* 182:2928–2936.
52. Kim, C. C., E. A. Joyce, K. Chan, and S. Falkow. 2001. Improved analytical methods for microarray-based genome-composition analysis. *Genome Biol.* 3:r0065.1–r0065.17.
53. Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, R. Okinaka, J. Wong, and P. Keim. 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J. Clin. Microbiol.* 39:3179–3185.
54. Kotewicz, M. L., E. W. Brown, J. E. LeClerc, and T. A. Cebula. 2003. Genomic variability among enteric pathogens: the case of the *mutS-rpoS* region. *Trends Microbiol.* 11:2–6.
55. LeClerc, J. E., B. Li, W. L. Payne, and T. A. Cebula. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 274:1208–1211.
56. Lenski, R. E., C. L. Winkworth, and M. A. Riley. 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J. Mol. Evol.* 56:498–508.
57. Lewis, G. D., S. L. Molloy, G. E. Greening, and J. Dawson. 2000. Influence of environmental factors on virus detection by RT-PCR and cell culture. *J. Appl. Microbiol.* 88:633–640.
58. Li, B., W. H. Koch, and T. A. Cebula. 1997. Detection and characterization of the *fimA* gene of *Escherichia coli* O157:H7. *Mol. Cell. Probes* 11:397–406.
59. Lindblad-Toh, K., E. Winchester, M. J. Daly, D. G. Wang, J. N. Hirschhorn, J. P. Laviolette, K. Ardlie, D. E. Reich, E. Robinson, P. Sklar, N. Shah, D. Thomas, J. B. Fan, T. Gingeras, J. Warrington,

N. Patil, T. J. Hudson, and E. S. Lander. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat. Genet.* 24:381–386.
60. Lipshutz, R. J., S. P. Fodor, T. R. Gingeras, and D. J. Lockhart. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21: 20–24.
61. Mazurier, S., A. van de Giessen, K. Heuvelman, and K. Wernars. 1992. RAPD analysis of *Campylobacter* isolates: DNA fingerprinting without the need to purify DNA. *Lett. Appl. Microbiol.* 14:260–262.
62. McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413:852–856.
63. Murray, A. E., D. Lies, G. Li, K. Nealson, J. Zhou, and J. M. Tiedje. 2001. DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. *Proc. Natl. Acad. Sci. USA* 98:9853–9858.
64. Nagano, I., M. Kunishima, Y. Itoh, Z. Wu, and Y. Takahashi. 1998. Detection of verotoxin-producing *Escherichia coli* O157:H7 by multiplex polymerase chain reaction. *Microbiol. Immunol.* 42:371–376.
65. Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, M. Holm, C. Culver, E. Martin, M. Fujimoto, E. Hoff, E. Kumlin, and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 235:1616–1622.
66. Neve, B., P. Froguel, L. Corset, E. Vaillant, V. Vatin, and P. Boutin. 2002. Rapid SNP allele frequency determination in genomic DNA pools by pyrosequencing. *BioTechniques* 32:1138–1142.
67. Nixon, K. 1999. The parsimony ratchet: a new method for rapid parsimony analysis. *Cladistics* 15:407–414.
68. Nuanualsuwan, S., and D. A. Cliver. 2002. Pretreatment to avoid positive RT-PCR results with inactivated viruses. *J. Virol. Methods* 104:217–225.
69. Ochman, H., and A. C. Wilson. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Biol.* 26:74–86.
70. Orskov, I., F. Orskov, B. Jann, and K. Jann. 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. *Bacteriol. Rev.* 41:667–710.
71. Parsons, W., T. J. Parsons, R. Scheithauer, and M. M. Holland. 1998. Population data for 101 Austrian Caucasian mitochondrial DNA D-loop sequences: amplification of mtDNA sequence analysis for a forensic case. *Int. J. Legal Med.* 111:124–132.
72. Patel, P. H., and L. A. Loeb. 2000. DNA polymerase active site is highly mutable: evolutionary consequences. *Proc. Natl. Acad. Sci. USA* 97:5095–5100.
73. Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. T. Dimalanta, K. D. Potamousis, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533.
74. Pfaller, M. A., C. Wendt, R. J. Hollis, R. P. Wenzel, S. J. Fritschel, J. Neubauer, and L. A. Herwaldt. 1996. Comparative evaluation of an automated ribotyping system versus pulsed-field-gel-electrophoresis for epidemiological typing of clinical isolates of *Escherichia coli* and *Pseudomonas aeruginosa* from patients with recurrent gram-negative bacteremia. *Diagn. Microbiol. Infect. Dis.* 25:1–8.
75. Porwollik, S., J. Frye, L. D. Florea, F. Blackmer, and M. McClelland. 2003. A non-redundant microarray of genes for two related bacteria. *Nucleic Acids Res.* 31:1869–1876.
76. Porwollik, S., R. M. Wong, and M. McClelland. 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* 99:8956–8961.
77. Porwollik, S., R. M. Wong, S. H. Sims, R. M. Schaaper, D. M. DeMarini, and M. McClelland. 2001. The Δ*uvrB* mutations in the Ames strains of *Salmonella* span 15 to 119 genes. *Mutat. Res.* 483: 1–11. (Erratum, 484:107–110.)
78. Pyrosequencing AB. 2002. Detection and typing of bacteria: *Listeria monocytogenes*. Available at: www.pyrosequencing.com. Accessed February 2005.
79. Radman, M., I. Matic, and F. Taddei. 1999. Evolution of evolvability. *Ann. N.Y. Acad. Sci.* 870:146–155.
80. Radnedge, L., P. G. Agron, K. K. Hill, P. J. Jackson, L. O. Ticknor, P. Keim, and G. L. Andersen. 2003. Genome differences that distinguish *Bacillus anthracis* from *Bacillus cereus* and *Bacillus thuringiensis*. *Appl. Environ. Microbiol.* 69:2755–2764.
81. Rao, K. V., P. Wilkins Stevens, J. G. Hall, V. Lyamichev, B. P. Neri, and D. M. Kelso. 2003. Genotyping single nucleotide polymorphisms directly from genomic DNA by invasive cleavage restriction on microspheres. *Nucleic Acids Res.* 31:e66.
82. Read, T. D., S. L. Salzberg, M. Pop, M. Shumway, L. Umayam, L. Jiang, E. Holtzapple, J. D. Busch, K. L. Smith, J. M. Schupp, D. Solomon, P. Keim, and C. M. Fraser. 2002. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 296:2028–2033.
83. Ronaghi, M., M. Uhlen, and P. Nyren. 1998. A sequencing method based on real-time pyrophosphate. *Science* 281:363–365.
84. Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow. 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. USA* 97:14668–14673.
85. Sharma, V. K., and S. A. Carlson. 2000. Simultaneous detection of *Salmonella* strains and *Escherichia coli* O157:H7 with fluorogenic PCR and single-enrichment-broth culture. *Appl. Environ. Microbiol.* 66:5472–5476.
86. Slomka, M. J., and H. Appleton. 1998. Feline calicivirus as a model system for heat inactivation studies of small round structured viruses in shellfish. *Epidemiol. Infect.* 121:401–407.
87. Sniegowski, P. D., P. J. Gerrish, and R. E. Lenski. 1997. Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* 387:703–705.
88. Snipes, K. P., D. C. Hirsh, R. W. Kasten, L. M. Hansen, D. W. Hird, T. E. Carpenter, and R. H. McCapes. 1989. Use of an rRNA probe and restriction endonuclease analysis to fingerprint *Pasteurella multocida* isolated from turkeys and wildlife. *J. Clin. Microbiol.* 27: 1847–1853.
89. Sokurenko, E. V., V. Tchesnokova, A. T. Yeung, C. A. Oleykowski, E. Trintchina, K. T. Hughes, R. A. Rashid, J. M. Brint, S. L. Moseley, and S. Lory. 2001. Detection of simple mutations and polymorphisms in large genomic regions. *Nucleic Acids Res.* 29:e111.
90. Swofford, D. L. 1999. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.03B. Program and documentation. Smithsonian Institution, Washington, D.C.
91. Townsend, J. P., K. M. Nielson, D. S. Fisher, and D. L. Hartl. 2003. Horizontal acquisition of divergent chromosomal DNA in bacteria. Effects of mutator phenotypes. *Genetics* 164:13–21.
92. van Belkum, A., S. Scherer, L. van Alphen, and H. Verbrugh. 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* 622:275–293.
93. van Gelder, R. N., M. E. von Zastrow, A. Yool, W. C. Dement, J. D. Barchas, and J. H. Eberwine. 1990. Amplified RNA synthesized from limited quantities of heterogenous cDNA. *Proc. Natl. Acad. Sci. USA* 87:1663–1667.
94. Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407–4414.
95. Wang, G., C. G. Clark, and F. G. Rodgers. 2002. Detection in *Escherichia coli* of the genes encoding the major virulence factors, the genes defining the O157:H7 serotype, and components of the type 2 Shiga toxin family by multiplex PCR. *J. Clin. Microbiol.* 40:3613–3619.
96. Wang, R.-F., M. L. Beggs, L. H. Robertson, and C. E. Cerniglia. 2002. Design and evaluation of oligonucleotide-microarray method

for the detection of human intestinal bacteria in fecal samples. *FEMS Microbiol. Lett.* 213:175–182.
97. Weissman, S. J., S. L. Moseley, D. E. Dykhuizen, and E. V. Sokurenko. 2003. Enterobacterial adhesins and the case for studying SNPs in bacteria. *Trends Microbiol.* 11:115–117.
98. Welsh, J., and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res.* 18:7213–7218.
99. Woese, C. 1987. Bacterial evolution. *Microbiol. Rev.* 51:221–271.
100. Woods, C. R., J. Versalovic, T. Koeuth, and J. R. Lupski. 1993. Whole-cell repetitive element sequence–based polymerase chain reaction allows rapid assessment of clonal relationships of bacterial isolates. *J. Clin. Microbiol.* 31:1927–1931.
101. Xiong, J., W. M. Fischer, K. Inoue, M. Nakahara, and C. E. Bauer. 2000. Molecular evidence for the early evolution of photosynthesis. *Science* 289:1703–1705.
102. Yu, J., and J. B. Kaper. 1992. Cloning and characterization of the *eae* gene of enterohaemorrhagic *Escherichia coli* O157:H7. *Mol. Microbiol.* 6:411–417.