Microbial Forensics

EDITORS

Roger G. Breeze, BVMS, PhD, MRCVS
Centaur Science Group
Washington, DC

Bruce Budowle, PhD
Federal Bureau of Investigation
Laboratory Division
Quantico, Virginia

Steven E. Schutzer, MD
University of Medicine and Dentistry of New Jersey—New Jersey
Medical School
Department of Medicine
Newark, NJ

© 2005

ELSEVIER
ACADEMIC
PRESS

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

CHAPTER 16

Population Genetics of Bacteria in a Forensic Context

RICHARD E. LENSKI
Department of Microbiology and Molecular Genetics, Michigan State University, East Lansing, Michigan

PAUL KEIM
Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona

INTRODUCTION

One of the central goals of microbial forensics is to identify the source of a microorganism that has been used for terrorist or other illicit purposes. To achieve this goal, it is necessary to understand the extent of microbial diversity, the genetic processes that generate diversity, other evolutionary processes that shape the patterns of diversity, and the resulting genealogical relationships among diverse strains. The field of population genetics is concerned with precisely these issues and is therefore central to microbial forensics. Population genetics can be viewed as a subfield of evolutionary biology that focuses on genetic diversity within species, as opposed to differences between species and higher taxonomic groups, although this distinction is not always precise, especially in microbes.

An understanding of the population genetics of the human species already plays a major role in forensics. For example, by comparing the DNA "fingerprint" in a sample of tissue (e.g., blood) from a crime scene with the DNA from a suspect in that crime, one can say with a high degree of statistical certainty whether the suspect was the source of the forensic sample. Within an appropriate evidentiary context, a match in such a comparison provides powerful evidence of the suspect's involvement, whereas the absence of a match can exonerate an accused individual.

The power of DNA fingerprinting for forensics was not a forgone conclusion: instead there was intense debate for several years.[1] The eventual consensus that supported the utility of human DNA in forensics was reached after systematic collection of data on the extent of diversity at the genetic loci used in testing, as well as detailed analyses of how that genetic diversity was distributed among different populations. This research has led to a greater basic understanding of the population genetics of the human species, as well as to improved forensic methods.

Given the successful forensic application of population genetic analyses of human data, it is not surprising that similar approaches are being pursued to trace the source of microorganisms (including bacteria, viruses, and fungi) whose genetic material is present in forensic samples. In fact, such work also represents a direct extension to forensics of the approaches that are widely used by molecular epidemiologists to track the source of outbreaks of many pathogens.

There are important differences in population genetics between humans and microorganisms. The aims of this chapter are to explain these differences, illustrate ways in which useful forensic inferences might be drawn from microbial DNA sequences, and suggest avenues for future research. Because microorganisms are themselves extraordinarily diverse, we will focus on bacteria and, as appropriate, use data on *Bacillus anthracis* to illustrate certain calculations. To start, we will compare and contrast inferences that can be drawn from DNA-based forensic evidence derived from humans versus bacteria.

DNA FORENSICS OF HUMANS AND BACTERIA

All humans share the vast majority of their DNA sequences.[2–4] Nonetheless, with genomes containing billions of nucleotide base-pairs (bp), and with sexual reproduction scrambling the variants every generation, no two humans are identical throughout their genomes, with the rare exception of identical twins (equivalent to exact clones). With enough high-quality sequence data, therefore, the source of a human forensic sample can be attributed to a particular individual with certainty. Obtaining enough data to make a strong probabilistic argument is made easier because some regions of the human genome are hypervariable,[5] allowing analyses to focus on those regions rather than requiring whole-genome sequences. This individuality of the genetic signature gives rise to the metaphor of DNA fingerprinting.

In extending the use of DNA evidence from human samples to bacterial DNA forensics, one possible line of reasoning might be as follows. Most bacterial species harbor tremendous sequence diversity. With bacteria having much smaller genomes than humans, it also becomes feasible to obtain full genome sequences for forensic samples. Therefore, according to this view, it should be easier and more certain to trace back from a bacterial sample to its source than to do so in the case of human DNA forensics. Unfortunately, this reasoning is, at least in some cases, false owing to the greater potential for exact clones (equivalent to identical twins) in bacteria than in humans.

Bacteria reproduce asexually, hence the existing genetic diversity within a population or species is not scrambled every generation. Moreover, mutation rates in bacteria are generally not high enough to ensure that new mutations occur in every cell generation.[6,7] Thus, exact clones are much more prevalent in the bacterial realm than in the human population.

Furthermore, microbiological research laboratories often preserve exact or nearly exact clones, and may distribute them to other laboratories. Hence, the possibility of exact clones is especially important when considering pathogens that may derive, accidentally or deliberately, from a laboratory. By contrast, when tracing the source of a natural outbreak, exact clones are likely to be less problematic because the number of generations and accumulated mutations are usually much greater, the relevant sources are not deliberately stored somewhere as clones, and perfect attribution is often less critical.

CASE STUDY OF *BACILLUS ANTHRACIS*

In this section, we will present some quantitative considerations related to microbial forensics. We will use published research on *Bacillus anthracis* to illustrate several important issues. We do so because the anthrax terrorist attacks of October 2001 have received considerable attention and, thus, offer concrete data for discussion. However, we will gloss over certain complications of this particular example in order to emphasize more general issues that would probably apply to similar cases in the future.

Following the anthrax attacks, Read, et al.[8] sequenced and compared the complete genomes of two *B. anthracis* isolates. One isolate was a forensic sample taken from a victim in Florida, while the other was from the government research laboratory at Porton Down in the United Kingdom. The Porton Down isolate had been previously "cured" of the two extrachromosomal plasmids that encode virulence factors, but otherwise it was presumed to be representative of the Ames strain that had been widely used in anthrax research. These two isolates, as well as other potential sources of the Florida attack strain, are shown in Figure 16.1. These other sources include U.S. governmental laboratories and field isolates.

Read, et al. reported that "Only four differences were discovered between the main chromosomes of the Florida and Porton isolates . . . two of these are SNPs and two are short indels." SNPs refer to single-nucleotide polymorphisms, which in this context are typical point mutations that distinguish the two sequences. Indels are insertions or deletion mutations, which often occur in specific hypermutable regions. We will focus initially on the two point

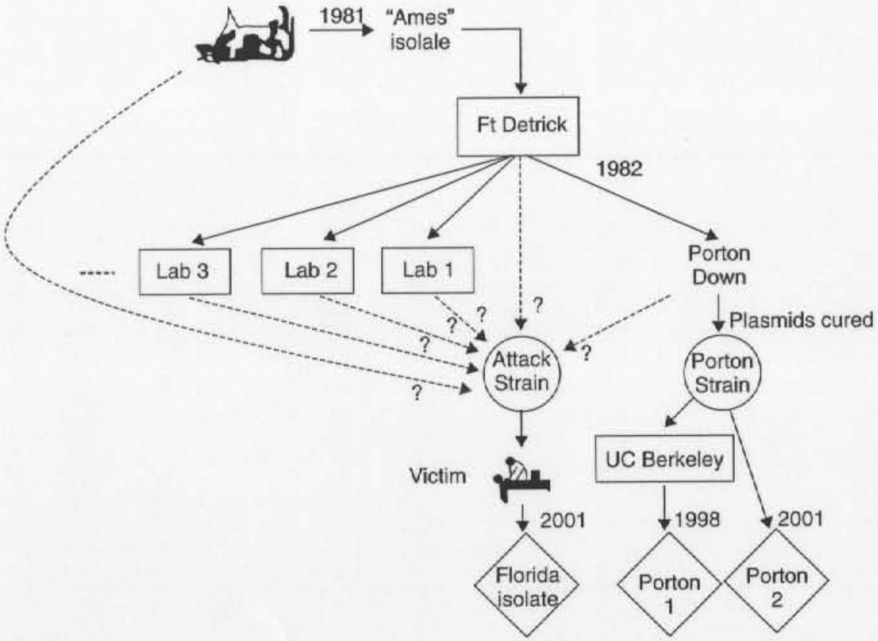

FIGURE 16.1 Possible derivations of the *Bacillus anthracis* isolated from the Florida anthrax victim in relation to several potential sources of the Ames strain. Chromosomes of the Florida isolate and the Porton Down laboratory strain have been sequenced. The Ames strain was originally isolated from a dead cow in Texas in 1981, then stored at Fort Detrick, and distributed to several other research laboratories (solid arrows), including Porton Down. The Porton Down strain was cured of its two virulence plasmids. (Reprinted with permission from ref. 8. Copyright 2002, American Association for the Advancement of Science.)

mutations that distinguish the Florida and Porton Down isolates. A complication is that the genome sequence of the Porton Down isolate was, in fact, based on DNA preparations from two substrains of the Porton Down strain, and these substrains were shown to have several mutational differences in their sequences. In our analysis, we will use only those differences that distinguish the Florida isolate from both Porton Down samples.

Given that there are two point mutations that distinguish the Florida and Porton Down isolates, we can ask several questions. Is that a surprisingly little difference, or is it a lot? What might these data tell us about how long ago the forensic isolate shared a common ancestor with the Ames laboratory strain? What might the data say about the relative likelihood of the forensic isolate coming from one source versus others?

To begin to answer these questions, we first need to understand the genomic mutation rate and the evidence concerning this rate in bacteria. The genomic mutation rate is simply the expected number of mutations per generation across the entire genome. To illustrate, consider *Escherichia coli*, which is the best studied bacterium. Our calculations assume functional DNA repair and ignore hypermutable sites, which constitute a small proportion of the genome. *E. coli* has a genome size of about 5×10^6 bp and a point mutation rate of about 5×10^{-10} per bp per generation[6] (see ref. 7 for a somewhat lower estimate). The product of these two quantities is the total genomic mutation rate, which in this case is approximately 2.5×10^{-3} point mutations per generation. The inverse of the genomic mutation rate is the expected number of generations until the first mutation occurs in a cell lineage (not in a population: see below), in this case about 400 generations. As it turns out, *B. anthracis* has a similar genome size,[9] and its mutation rate measured for a particular gene also is similar to *E. coli*.[10]

At first glance, the expected time of 400 generations until the first mutation may seem an inappropriately long period, because it ignores the fact that a bacterial population may contain millions of cells, such that many mutations can occur every generation. However, comparisons between genomic sequences are based on single representatives of each sample, not on entire populations. Without belaboring this point, the expected time that is relevant to our analysis will generally be somewhat longer, not shorter, than we estimated above.

In point of fact, we are most interested in a quantity called the genomic substitution rate. A substitution is any mutation that spreads throughout a population of interest. The details can get complicated quickly but, fortunately, there are some mathematical shortcuts. In the present context, we can treat the number of substitutions as the number of mutations that distinguish the two individuals whose genomes are under comparison. Neutral mutations have no effect on fitness; synonymous point mutations are often used as a proxy for neutral mutations. A robust result from theoretical population genetics is that the expected substitution rate of the class of neutral mutations is equal to their corresponding mutation rate.[11] Deleterious mutations—those which reduce a cell's survival or growth rate—have a substitution rate lower than the corresponding mutation rate, while beneficial mutations have a substitution rate above their mutation rate. Because many more mutations are deleterious than are beneficial, the genomic substitution rate is, in general, somewhat lower than the genomic mutation rate. Hence, as noted above, the expected time to the first substitution of a point mutation will be longer than 400 generations.

So what might we begin to conclude from the data? Given the two point mutations that distinguish the Florida and Porton Down isolates, the inferred time since their common ancestor is on the order of 800 cell generations (i.e., twice the expected time to the first mutation). Although not a huge number, it is to us surprisingly large in the context of standard lab practice, where one would expect working subcultures to be repeatedly restarted from a master culture (stored in a non-growing state, as spores or frozen vegetative cells), rather than by sustained propagation of subcultures. The inferred 800 generations would correspond to about 30 rounds of plating for single colonies (each colony representing some 25 cell divisions), and even more rounds if cells were propagated by serial dilution and transfer. Translating generations into chronological time would further depend on knowing how long cells might have been stored in a non-mutating state, for example as spores or in a freezer.

The inferred time since the common ancestor could be reduced, perhaps dramatically, if the Ames strain were defective in DNA repair (see refs. 7, 12 for the effects of loss of repair in *E. coli*) or if there was a history of mutagenesis. In terms of DNA repair, it appears that the Ames strain retains these functions given that mutation-rate estimates at specific loci are in line with estimates for *E. coli* that have normal DNA repair functions.[10] However, with respect to growth under mutagenic conditions, it could be relevant that curing the two virulence plasmids from the Porton Down strain involved treating cells with high temperature and an antibiotic.[9] In fact, depending on the timing of these treatments relative to the derivation of the two substrains of the Porton Down strain, these conditions may even explain the differences between the substrains as well as between the Florida isolate and the Porton Down strain.

Statistical uncertainties are also important with respect to inferring the time since two strains diverged from a common ancestor. Even if we accept the substitution rate as known, there remains the intrinsic error arising from the stochastic (random) occurrence of mutations. Using the Poisson distribution to reflect this intrinsic error, the probability of observing two or more mutations in anything fewer than 142 generations is below 5%. At the other end of the distribution, two strains are also unlikely to have substituted as few as two mutations in 1900 generations or longer ($p < 5\%$). Although these bounds are already large, they are the best that one can do given only two point-mutation differences, because they assume that all else is known precisely. The bounds become even larger with uncertainty in, for example, the mutation rate. Despite these wide bounds, one might still exclude certain scenarios. For the sake of illustration (leaving aside the facts that the Porton Down strain lacks the virulence plasmids, and was subjected to mutagenic treatments), these two point mutations make it unlikely that the forensic isolate came directly from the Porton Down strain or even that it was derived from that strain via fewer than several rounds of plating or subculturing.

Thus, after considering what is known about rates of mutation in bacteria and placing this information in an evolutionary context, the two point mutations separating the forensic isolate from Florida and the Ames laboratory strain appear to be more than might have been reasonably expected, not less, provided that the Porton Down strain is indeed representative of the Ames strain more generally. We will return to this proviso a bit later.

It should also be clear from what has been said that "either-or" inferences based on match versus no-match between a possible source and a forensic sample are less conclusive for bacteria than for humans, owing to the much greater possibility of exact clones (identical twins) in the microbial case and especially in the context of a deliberate attack using a strain taken from the laboratory.

Let us now shift gears, and consider these data from the perspective of establishing the most probable "line of descent" of a forensic isolate in relation to multiple potential sources. In this context, even one or a few distinguishing genetic substitutions could—in principle—provide compelling evidence to support or exclude certain scenarios. In the paragraphs that follow, we examine two such scenarios to illustrate how the data could be profitably analyzed. The first scenario is hypothetical and invokes imaginary data in order to make certain points clear. The second scenario accords with the relevant published data.

FIRST SCENARIO

Recall that the forensic isolate from Florida and the Porton Down version of the Ames strain differ by two point mutations. We can designate the genotype of the Porton Down isolate as AB and that of the Florida isolate as A′B′ to reflect these two mutations. Now suppose that isolates of the Ames strain from the four laboratories in Figure 16.1 are also characterized with respect to these mutations, along with an isolate from a hypothetical rogue laboratory discovered in an investigation. Suppose that the Ames isolates from Fort Detrick, Lab 1 and Lab 2 had the same AB genotype as the Porton Down strain, while the isolate from Lab 3 yielded the A′B genotype and the isolate from the rogue laboratory had the same A′B′ as the forensic isolate from Florida.

Under this hypothetical scenario, the identity of the isolate from the rogue laboratory with the forensic sample, coupled with the differences between these isolates and those from all other laboratories, would suggest that the likely source of the forensic isolate was the rogue laboratory. Moreover, the fact that the isolate from Lab 3 alone shared one of the two distinguishing point mutations would further suggest that the strain taken from the rogue lab was derived from Lab 3. Of course, DNA sequences of anthrax isolates would presumably not be the only line of evidence presented in a criminal case against the hypothetical operator of the rogue laboratory; the case would be further strengthened, for example, if the operator had previously worked in Lab 3.

At first glance, the finding that these two point mutations are shared by the forensic and rogue-lab isolates may not seem like much, given that the *B. anthracis* genome has about five million bp. However, when viewed against a background of genomic uniformity, even one or a few shared mutations provide compelling quantitative support for an association. To illustrate, we will make some simplifying assumptions and rough calculations. No doubt such assumptions could be relaxed, and the calculations refined, in any actual case.

If all mutational substitutions were equally likely, then the probability that some putative source would have independently substituted the exact same m mutations (and no others) as in a forensic sample is calculated as $p = 1/(3n)^m$, where n is the genome size in bp and the factor of 3 represents three alternative base-pairings at each genomic site. With a genome of $n = 5 \times 10^6$ bp and $m = 2$ mutations, this probability of chance convergence is $<10^{-14}$.

However, the refined probability estimate would be greater owing to variation among genome positions in substitution probabilities. Certain positions are more mutable than others,[13] and therefore coincident mutations become more probable than the calculation above suggests. Also, selection can sometimes lead to convergent substitutions even when the underlying mutation rates are the same.[14] To illustrate, imagine that 1/1000th of the genome (5×10^3 bp) is highly mutable, with all of these sites equally mutable and with unrestricted nucleotide substitution. In that case, the probability of chance convergence with $m = 2$ is still $<10^{-8}$. If substitutions at each of the highly mutable sites are restricted to one particular nucleotide, then the probability of convergence is higher (as $p = 1/n^m$), but still $<10^{-7}$. In any case, it is important to emphasize that a few, or even one, mutational matches could be quantitatively compelling, especially if the sequence data are supported by other evidence (such as the hypothetical connection between the rogue laboratory operator and a related source strain).

SECOND SCENARIO

In fact, when the mutations that distinguish the forensic isolate and the Porton Down strain were checked among isolates from the four other laboratories in Figure 16.1, it was found that these other isolates all shared the same genotype as the forensic sample.[8] Thus, it seems the Porton Down strain accumulated these distinguishing mutations, probably during the mutagenic treatments used to eliminate the virulence plasmids.[9]

More importantly, the forensic isolate appears to be identical in its genomic sequence to the Ames strain that was shared among several different laboratories. It is possible that complete genome sequences of isolates from one or more of these laboratories might exclude one of them as the proximate source. However, examination of hypermutable sites [variable-number tandem repeats (VNTRs)] in the genomes of these isolates does not reveal any exclusionary differences.[8] Thus, the genomic data per se lack the power to discriminate between these particular laboratories as potential sources of the Florida forensic isolate. Although it is not a focus of this chapter, we should mention that other types of evidence, such as trace elements reflecting bacterial growth conditions, might allow source discrimination even in the absence of distinguishing mutations. Of course, evidence concerning access to *Bacillus anthracis* by a suspect individual or group would also be quite relevant.

In Figure 16.1, one can also see that the Ames strain found in all of the laboratories was itself derived from a dead animal in 1981. This source also raises the possibility that the Florida isolate was not obtained from any of the research laboratories but, instead, was another isolate sampled from nature by the perpetrator. (In this case, the possibility of a natural infection was excluded by the circumstances of the deliberate attacks using the postal system.) The question then becomes: What is the probability that a fresh isolate taken from the wild would be an exact clone of the isolate sampled twenty years earlier? In bacteria that cannot form spores, one might estimate the number of elapsed generations and the expected number of substitutions that would have accumulated, and use these to calculate the likelihood of zero mutational substitutions. In fact, this likelihood would be the upper bound because it would assume that the same evolving lineage was chosen in both instances and thus ignores the extent of genetic diversity present at any moment. However, *B. anthracis* forms spores that can persist for a long time, making it difficult to estimate the number of elapsed generations and hence expected substitutions.

An alternative approach would be to estimate empirically the fraction of isolates from nature that are exact clones of the Ames strain. At first glance, this might seem a daunting task that would require genome sequences from hundreds of isolates. In fact, a shortcut exists by first scoring the hypermutable loci. For example, one could sample isolates from various geographic regions, including that where the Ames strain was sampled. All of these could be scored for the hypermutable genes, and any differences would exclude identity. Let us imagine, for the purpose of illustration, that none of 500 isolates from other geographical regions matched the Ames strain at all of these genes. In that case, the probability of a match from outside the region would be estimated as below 0.2%. Now imagine also that five isolates out of 100 sampled from the vicinity of the Ames strain matched that strain at all of the hypermutable loci. The probability of a match from within that region might first be estimated as 5%, but it must be emphasized that this estimate is likely to be an overestimate because it uses only a subset of the genome (i.e., the hypermutable loci) to establish differences. One might then sequence whole genomes of the five isolates that matched the Ames strain at all of the hypermutable genes to see whether any of them was truly an exact clone. Alternatively, various other genetic approaches less costly than whole-genome sequencing might be used to screen this subset of clones. Among our recommendations, we suggest efforts to develop methods for the rapid discovery of unknown sequence differences between closely related genomes. Also, any differences found in one natural isolate could be checked in the others, providing another potentially useful shortcut toward determining whether any of the natural isolates are truly exact clones of the Ames strain and the forensic isolate. None of the several natural isolates mentioned by Read, et al.[8] proved to be exact clones of the Ames strain, although distinguishing mutations were found in the virulence plasmids (but not on the chromosome).

SOME ISSUES NEEDING FURTHER ATTENTION

In this section, we briefly raise some questions that would benefit from further attention, either in general or in specific cases that might arise.

How similar are mutation rates across species? Among strains within a species? At the level of species, published data seem to indicate that point mutation rates are inversely correlated with genome size across microbes with DNA genomes ranging all the way from viruses through bacteria to yeast.[6,15] The product of these quantities, which equals the genomic mutation rate, is surprisingly constant by this analysis. It would be useful to have more data bearing on this pattern, especially as the explanation for the genome-level constancy remains unclear. In any case, these data support the general expectation that genome-wide rates of point mutation are often well below one, which allows the existence of exact clones, especially in laboratory settings where strains may be deliberately stored and then sent elsewhere.

Despite this apparent constancy in rates across such diverse taxa, there can be substantial variation in mutation rates among different strains of the same species. For example, there are strains of *E. coli* and many other bacterial and fungal species that have mutation rates elevated by 10, 100, or even 1000-fold owing to defects in mismatch repair and other pathways.[16] Such "mutator" strains can be fairly common, both in nature[17] and in the laboratory.[12] In these mutator strains, the potential for exact clones would be greatly reduced.

The apparent constancy of genomic point-mutation rates does not extend to viruses with RNA genomes, such as the virus that causes flu in humans. RNA-based viruses, including retroviruses (such as HIV, which causes AIDS), generally have much higher genomic mutation rates despite their small genome size.[18] Again, the potential for exact clones of these viruses would be very much reduced relative to the potential in bacteria with normal DNA repair functions.

How variable are mutation rates across sites within a given genome? On the one hand, variation in mutation rates will complicate estimation of the divergence times between isolates, as well as calculation of the probability of a spurious match between isolates that converged on the same mutations. On the other hand, hypermutable sites may provide opportunities to detect sequence differences that arise quickly (between closely related isolates), and these sites can be screened for a large set of isolates at much less cost than sequencing their entire genomes.

In the case of typical point mutations, the differences between sites are probably fairly minor in the greater scheme of things. However, there are certain sequences, including homopolymeric tracts (e.g., AAAAA . . .) and VNTRs (e.g., ATTATTATT . . .) that are much more mutable, with the repeated elements being added or lost to produce insertions and deletions.[19] Examination of multiple hypermutable loci increases the chance of detecting one or more informative mutational differences even over short time scales. The *B. anthracis* genome has several such regions[20] which are recognizable and can be avoided or subjected to particular scrutiny, depending on the appropriate context.

How are strains propagated? The expected number of substitutions that distinguish two isolates will depend on the number of generations between them as well as the relevant mutation rates. The realized number of substitutions will also depend on the fate of particular mutations, and the likelihood that certain kinds of mutation are substituted will in turn depend on how a population is grown.[14] If a population experiences repeated bottlenecks, such as when cells are plated as individual colonies, then deleterious mutations can accumulate. In large populations that do not experience any severe bottlenecks, natural selection will tend to eliminate deleterious mutations and substitute beneficial ones. Neutral mutations, such as those synonymous changes that do not alter a protein's structure, will not be affected in either case. In general, because more mutations are deleterious than beneficial, the genomic substitution rate will be somewhat lower than the genomic mutation rate. However, the discrepancy should not be more than a few-fold, because onequarter or so of all possible mutations in protein-encoding genes are synonymous and therefore should be neutral or nearly so.

A related issue is the history of mutagenesis, if any, during the derivation of strains and subcultures. We have already noted that the high-temperature and antibiotic treatments used to eliminate the virulence plasmids from the Porton Down isolate of *B. anthracis* are mutagenic, and may well account for the genetic differences between that isolate and the other Ames isolates. It is generally thought that mutations do not accumulate during the long-term storage of strains in freezers or as spores, because cells are not metabolically active. However, when vegetative cells are kept at higher temperatures, they continue to metabolize and may accumulate mutations, with the extent of this starvation-induced mutability being rather variable, at least among strains of *E. coli.*[21]

CONCLUSIONS

- In contrast to human DNA-based forensics, a perfect match between a bacterial forensic sample and some potential source is less definitive owing to exact clones that may exist. Such exact clones reflect the asexual mode of bacterial reproduction.
- Microbiology laboratories often store clones and exchange them with other laboratories. The opportunity for exact clones to confound forensic source-tracking is thus especially relevant in the context of forensic samples that may have come from a laboratory.
- If a forensic sample is genetically identical to samples from several labs, then the genome sequences lack the power to discriminate between them as potential sources. However, other types of evidence (e.g., trace elements) may exist that can allow discrimination and should be sought in any case.
- Even one or a few genetic differences between a forensic sample and some potential sources can exclude certain scenarios while focusing attention on others. The pattern of differences among potential sources might, for example, implicate one of them as an intermediate or final stage in the derivation of a forensic sample.
- The issues emphasized in this chapter are more relevant for tracking a deliberate attack than a natural epidemic. In the latter case, the number of generations and accumulated genetic differences are generally much greater, the potential sources are not deliberately stored as clones, and perfect attribution is often less critical.

RECOMMENDATIONS

In closing, we offer several recommendations relevant to the issues addressed in this chapter. Some of our recommendations are similar to ones put forth in a recent report of the American Academy of Microbiology,[22] which also included others beyond the scope of our chapter. Although our recommendations are cast with reference to improving the science of microbial forensics, we emphasize that fulfilling most of them would simultaneously enhance the investigation of natural epidemics.

- The scientific community should make plans to obtain whole-genome sequences for all significant pathogenic threats to humans, ideally in advance of any forensic investigation that might become necessary. Genome sequences should also be obtained for important pathogenic threats to those plant and animal species on which we depend for food and economic vitality.
- It would be most useful to have multiple whole-genome sequences for every pathogenic species that poses a significant threat. While a single sequence would allow the extent of differences with a subsequently sequenced forensic sample to be gauged, having several diverse sequences would enable a more powerful triangulation to indicate the pattern of relatedness among strains and the order in which genetic differences accumulated.
- Basic population-genetic research in support of microbial forensics is needed to estimate rates of mutation and genomic evolution, including differences between sites in the same genome, variation among diverse species, and under alternative culture conditions. Such research should include studies in experimental settings as well as comparative studies of natural populations.
- Basic molecular biological research is needed to develop methods for the rapid discovery of unknown sequence differences between closely related genomes. Various techniques for genotyping based on known differences are well established, but methods to find, for example, a single point-mutation difference between two genomes need more research.
- Attention must be paid to statistical issues when interpreting microbial forensic evidence, as has been the case with human DNA forensics. However, the precise nature and form of the calculations will be somewhat different from those used in human forensics, owing to differences in the underlying genetic processes and population structures.

ACKNOWLEDGMENTS

We thank Chris Borland, Bruce Budowle, Charles Cooke, David Rasko, Tim Read, and Steven Salzberg for helpful discussion and correspondence. R.E.L.'s research is supported primarily by the National Science Foundation (NSF), and P.K.'s research is supported by the NSF, National Institutes of Health, Department of Energy, and the Department of Justice.

REFERENCES

1. National Research Council. *The evaluation of forensic DNA evidence*. Washington, DC: National Academy Press; 1996.
2. Li WH, Sadler LA. Low nucleotide diversity in man. *Genetics* 1991;129:513–523.
3. Cavalli-Sforza LL, Menozzi P, Piazza A. *The history and geography of human genes*. Princeton, NJ: Princeton University Press; 1994.
4. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D. Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Gen* 2002;32:135–142.
5. Chakraborty R, Kidd KK. The utility of DNA typing in forensic work. *Science* 1991;254: 1735–1739.
6. Drake JW. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Nat Acad Sci USA* 1991;88:7160–7164.
7. Lenski RE, Winkworth CL, Riley MA. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *J Mol Evol* 2003;56:498–508.
8. Read TD, Salzberg SL, Pop M, Shumway M, Umayam L, Jiang L, Holtzapple E, Busch JD, Smith KL, Schupp JM, Solomon D, Keim P, Fraser CM. Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*. *Science* 2002;296:2028–2033.
9. Read TD, Peterson SN, Tourasse N, Baillie LW, Paulsen IT, Nelson KE, Tettelin H, Fouts DE, Eisen JA, Gill SR, Holtzapple EK, Okstad OA, Helgason E, Rilstone J, Wu M, Kolonay JF, Beanan MJ, Dodson RJ, Brinkac LM, Gwinn M, DeBoy RT, Madpu R, Daugherty SC, Durkin AS, Haft DH, Nelson WC, Peterson JD, Pop M, Khouri HM, Radune D, Benton JL, Mahamoud Y, Jiang LX, Hance IR, Weidman JF, Berry KJ, Plaut RD, Wolf AM, Watkins KL, Nierman WC, Hazen A, Cline R, Redmond C, Thwaite JE, White O, Salzberg SL, Thomason B, Friedlander AM, Koehler TM, Hanna PC, Kolsto AB, Fraser CM. The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 2003;423:81–86.
10. Vogler AJ, Busch JD, Percy-Fine S, Tipton-Hunton C, Smith KL, Keim P. Molecular analysis of rifampin resistance in *Bacillus anthracis* and *Bacillus cereus*. *Antimicrob Agents Chemother* 2002;46:511–513.
11. Kimura M. *The neutral theory of molecular evolution*. Cambridge: Cambridge University Press; 1983.
12. Sniegowski PD, Gerrish PJ, Lenski RE. Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature* 1997;387:703–705.
13. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 1978;274:775–780.
14. Elena SF, Lenski RE. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Gen* 2003;4:457–469.
15. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics* 1998;148:1667–1686.

16. Friedberg EC, Walker GC, Siede W. *DNA repair and mutagenesis*. Washington, DC: Am Soc Microbiol Press; 1995.
17. LeClerc JE, Li BG, Payne WL, Cebula TA. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* 1996;274:1208–1211.
18. Drake JW, Holland JJ. Mutation rates among RNA viruses. *Proc Nat Acad Sci USA* 1999;96: 13910–13913.
19. Moxon ER, Rainey PB, Nowak MA, Lenski RE. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Bio* 1994;4:24–33.
20. Keim P, Price LB, Klevytska AM, Smith KL, Schupp JM, Okinaka R, Jackson PJ, Hugh-Jones ME. Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* 2000;182:2928–2936.
21. Bjedov I, Tenaillon O, Gérard B, Souza V, Denamur E, Radman M, Taddei F, Matic I. Stress-induced mutagenesis in bacteria. *Science* 2003;300:1404–1409.
22. Keim P. Microbial forensics: A scientific assessment. Washington, DC: *Am Acad Microbiol*; 2003.