International Congress Series 1261 (2004) 56–58

www.ics-elsevier.com

Genetic considerations for interpreting molecular microbial forensic evidence

B. Budowle[a,*], R. Chakraborty[b]

[a] *FBI Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA*
[b] *Center for Genome Information, Department of Environmental Health, University of Cincinnati, Cincinnati, OH 45267, USA*

Abstract. Genetic analyses of microbial evidence will be employed to assist in attribution of perpetrators of bioterrorism and biocrimes. There are some similarities and differences between human forensic and microbial forensic DNA analysis practices to consider. These population genetic and statistical interpretation issues center on the different genetic make-up, different inheritance mechanisms, different regulation mechanisms, and lineage-based analyses. In some cases, a quantitative assessment of the results of analysis may be possible; in other cases, it may be more appropriate to provide only a qualitative statement. Published by Elsevier B.V.

Keywords: Microbial forensics; Population genetics; Statistics; Bioterrorism; Biocrime

1. Introduction

Microbial forensic identity testing is used to determine the source of microbial samples or their toxins that have been used in acts of bioterrorism or in the commission of biocrimes. One approach to obtain microbial signatures is by assaying genetic polymorphisms between microbial strains or substrains to infer the origin, relationship, or transmission route of a particular isolate.

2. Interpretation

When the interpretation is ''inclusion or match or same lineage,'' it is desirable to place some significance or weight on these results. There are similarities between human forensic and microbial forensic DNA analysis practices, which include the use of population databases, the use of qualitative conclusions of test results, and the application of QA/QC practices. However, there are several major differences, particularly in the context of data interpretation. Some obvious differences in approach are: database size and composition, statistical interpretation methods, and confidence in outcome of an interpre-

* Corresponding auhtor. Tel.: +1-703-632-8386; fax: +1-703-632-7817.
E-mail address: bruce.budowle@verizon.net (B. Budowle).

0531-5131/ Published by Elsevier B.V.
doi:10.1016/S0531-5131(03)01833-8

tation. Most bacteria and viruses are haploid, so diploid-based statistical approaches would not be appropriate and do not apply. Moreover, most microbes reproduce predominately asexually and, thus, are clonal in nature, making the identification by genetic means of the sole source of an isolate or a laboratory-prepared pathogen often impossible. Therefore, the majority of microbial forensic genetic evidence may only infer common lineage instead of identity.

Population genetics issues to consider are origin and maintenance of microbial diversity. These are dependent on the mode of reproduction (i.e., asexual, although sexual reproduction also must be considered), inter-genic/site variation of mutation rate, recent natural vs. laboratory origin, experimental and natural stress conditions, and horizontal gene transfer. The effects of natural selection together with mutation and recombination will contribute to variation and need to be considered as well.

Since diversity levels in a number of microbial species will not reach that of human populations, microbial forensics will need lineage-based models in which population structure issues are somewhat different than for human samples. But, approaches for statistical analysis do exist that need to be validated with empirical genomic data from model organisms. The extent of mismatch with comparison to a database of samples may be needed to provide quantitative support of the match/non-match evidence. Analysis will likely be based on coalescence theory incorporating population size fluctuation over generations, inter-site variation of mutation rate, recombination, horizontal gene transfer, and conservation of sites (based on function).

Alternatively, some cases may not require any quantitative assessment and a qualitative statement may suffice [1–4]. In other scenarios, a quantitative evaluation can be made, but the bounds on the estimate may be large, owing to the sparsity of data. To increase confidence in quantitative estimates, a better understanding of microbial genome dynamics is necessary. Data on genetic variation and relatedness within and between species can assist in distinguishing between two strains that appear phenotypically identical. Analysis of pathogenicity domains, virulence genes and antibiotic resistance genes, location of recombination sites, development of effective horizontal gene transfer and vertical gene transfer statistical models, general and specific genome site stability studies, near neighbor analyses, and diversity studies [5–8] will enhance capabilities to quantitatively assess the significance of lineage-based comparisons.

Although we provide a list of further studies, this does not suggest that inferences for attribution cannot be made. Substantial data on genetic markers and population diversity have been accumulated for some potential bioweapon species that include: smallpox, *BA*, *Fancisella tularensis*, *Yersinia pestis*, *Burkholderia pseudomallei*, *Burkholderia mallei*, *Escherichia coli*, *Salmonella* ssp., *Shigella* ssp., *Brucella* ssp., and Foot and Mouth Disease Virus. A unifying statistical framework should be considered based on using lineage-based models where the extent of match/non-match evidence is assessed under the hypotheses of a particular sample(s) belonging to one group (or lineage) of putative samples or not. Such statistical models, based on recent developments in phylogenetic analyses of molecular evolution, are under development.

Another important distinction of the human DNA forensic practices from that of microbial forensics is worth noting. In the human forensic context, DNA markers generally chosen for forensic analyses are devoid of functions. In contrast, terrorism-

related microbes generally will contain virulence factors. Hence, a reliable and quick method of detecting signatures of virulence could be an important aspect of genomic work in microbial forensics. Molecular tools for detection of signatures of functionally active sites of the genome are now available to apply them to the emerging genomic data on biothreat agents (see, e.g., Ref. [9]).

3. Conclusion

In conclusion, comparative genome sequencing already is providing details of genome organization and genetic variation [10,11]. The number of markers is increasing and the distribution of polymorphisms is being established in representative sets of strains. Model organisms are being studied for estimating inter-site variation of mutation rate, recombination, horizontal gene transfer, and conservation of sites (based on function) [6]. Analyses of these data, based on comparative genomic principles, should improve the capabilities and confidence in providing quantitative assessments of microbial forensic genetic evidence.

Acknowledgements

This is publication number 03-20 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

[1] P. Keim, L.B. Price, A.M. Klevytska, K.L. Smith, J.M. Schupp, R. Okinaka, P.J. Jackson, M.E. Hugh-Jones, Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*, J. Bacteriol. 182 (2000) 2928–2936.

[2] P. Keim, K.L. Smith, C. Keys, H. Takahashi, T. Kurata, A. Kaufmann, Molecular investigation of the Aum Shinrikyo anthrax release in Kameido, Japan. J. Clin. Microbiol. 39 (2001) 4566–4567.

[3] C.A. Cummings, D.A. Relman, Microbial forensics—cross-examining pathogens, Science 296 (2002) 1976–1978.

[4] R.S. Lanciotti, et al., Origin of the West Nile virus responsible for an outbreak of encephalitis in the northeastern United States, Science 286 (1999) 2333–2337.

[5] J.W. Drake, B. Charlesworth, D. Charlesworth, J.F. Crow, Rates of spontaneous mutation, Genetics 148 (1998) 1667–1686.

[6] J.E. LeClerc, B.G. Li, W.L. Payne, T.A. Cebula, High mutation frequencies among *Escherichia coli* and *Salmonella pathogens*, Science 274 (1996) 1208–1211.

[7] E.R. Moxon, P.B. Rainey, M.A. Nowak, R.E. Lenski, Adaptive evolution of highly mutable loci in pathogenic bacteria, Curr. Biol. 4 (1994) 24–33.

[8] P.D. Sniegowski, P.J. Gerrish, R.E. Lenski, Evolution of high mutation rates in experimental populations of *Escherichia coli*, Nature 387 (1997) 703–705.

[9] W.H. Li, M. Tanimura, C.C. Luo, S. Datta, L. Chan, The apolipoprotein multigene family: biosynthesis, structure, structure–function relationships, and evolution, J. Lipid Res. 29 (1988) 245–271.

[10] T.D. Read, et al., Comparative genome sequencing for discovery of novel polymorphisms in *Bacillus anthracis*, Science 296 (2002) 2028–2033.

[11] T.D. Read, et al., The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria, Nature 423 (2003) 81–86.