

## **Population Geneticists Discover Bacteria and their Genetic/Molecular Epidemiology**

**Bruce R. Levin**  
**Department of Biology**  
**Emory University**  
**Atlanta, Georgia 30322**  
**blevin@emory.edu**

In this personal, opinionated, occasionally nostalgic and doubtless egoistic rant I consider the origins and early history of the enterprise to which Tom Whittam made his awesome contributions. While we all know what those contributions are, we almost certainly differ in what we call this enterprise; Population Genetics, Genetic Epidemiology, Evolutionary Genetics, Molecular Epidemiology, Molecular Evolution, Molecular Phylogeny and Molecular Forensics. (We have to be molecular, “who is not” (15).) Tom’s research includes projects that fit all of these descriptions and more. By all standards, the breadth as well the importance, utility and quality of Tom’s contributions are impressive.

Tom was not the first p,q-trained and inclined population and evolutionary geneticist to work with bacteria, but when he began to do so in the early 1980s he was a member of a small minority. As Carl Bergstrom and I once quipped, “The genetical theory of evolution was developed by sexually reproducing eukaryotes for sexually reproducing eukaryotes” (13). Indeed, shortly before I published my first research with bacteria (11), Theodosius Dobzhansky admonished me when I told him I gave up *Drosophila* to work on *E. coli*, “Zat is not an organism”<sup>1</sup>. Among the enlightened

---

<sup>1</sup> Although at the time, few card-carrying population geneticists and evolutionary biologist showed much interest in bacteria when they were not ill with some infection, a study with bacteria was responsible for the one of the most significant contributions to the genetical theory of evolution. The first evidence for the pre-adaptive nature of mutations, that “Natural selection is the editor, rather than the composer, of the genetic message” (10) were experiments with *E. coli* (14). From his remark to me, which I later learned to be no more than a good-natured tease, it may appear that Dobzhansky was unaware of the evolutionary implications of Luria and Delbruck’s experiments and other studies of the genetics of bacteria. That was not the case; see pages 87 and 88 in (8) It was certainly not the case for Jim Crow, who was not only very attuned these studies with bacteria, but his student Jack Bennett provided the first evidence for pre-adaptive mutation in eukaryotes (3,7). There was also an *E. coli* study by Kim Atwood and colleagues (2), which in a contemporary perspective would be seen as particularly classy experimental population genetics.

minority working with these ‘non-organisms’, Tom was one of the first to apply population genetic procedures and theory to the genetic epidemiology of bacteria from clinical and other natural sources.

The Neutral Gene Hypothesis and the Clone Concept: In my interpretation, the immediate antecedent of Tom’s research with bacteria as well as that of Bob Selander, Dominique Caugant and Howard Ochman (eukaryote-trained biologists all) was the neutral gene controversy, that dominated population genetics of the Disco era, in particular an article by Roger Milkman (17). Using a collection of 829 isolates of *E. coli* from diverse sources and cellulose acetate electrophoresis, Roger estimated the frequency of different electrophoretic variants for five housekeeping proteins. He interpreted the results of his study as evidence against the neutral gene hypothesis. He argued that for a species with as large a genetically effective population size as he assumed *E. coli* must have, in accord with the neutral gene hypothesis (1) the effective number of alleles would be vastly greater than that he estimated, and (2) the distribution of motility classes (alleles) would be different than that he observed. Central to and implicit in his interpretation was the assumption that *E. coli* was in linkage equilibrium; variation among isolates from the same individual was great enough to “suggest that recombination occurs regularly within hosts”.

Was Milkman’s 1973 study of enzyme variation in *E. coli* the definitive, “no excuse”, test of the neutral gene hypothesis he proposed it would be (18)? It may well have seemed that way to the pan-selectionist<sup>2</sup>, and even the most astute neutralists like Motoo Kimura didn’t counter it<sup>3</sup>. It wasn’t until 1980 that an “excuse” for Milkman’s study was provided in a report by Bob Selander and myself (25).

The origins of that paper were a talk I gave a talk at the University of Rochester. As I recall, the subject my presentation was the population dynamics of bacteriophage and plasmids. After completing my talk, Bob Selander - who was already been well known for his extensive and first-rate enzyme electrophoretic studies of genetic variation, evolution and the genetic structure of natural populations of a number of species of

---

<sup>2</sup> Although I never really knew whether it was rhetorical posturing or sociological, at the time there were indeed pan-selectionists, despite the fact that the neutral gene and selection hypotheses were not mutually exclusive for the genome at large or any given gene. Needless to say, nobody was misguided enough to champion pan-neutrality, although some may have been accused of doing just that.

<sup>3</sup> In 1980 I received the first of my letters and beautiful Christmas cards from Motoo Kimura. While I can’t say for sure, I always attributed that to the publication of Bob and my 1980 Science paper, or Motoo learning about our results and interpretation before from a super-able mathematical population geneticist, the late Takeo Maruyama, who had visited our lab at about that time. Indeed, one can almost sense a tone of relief in a fine article Maruyama and Kimura wrote in 1980 about the genetic structure and effective population size of asexual haploids (16).

eukaryotes - asked if he could come to my lab at UMass Amherst to learn how to culture *E. coli*. Bob's two days in our lab was the start of a collaboration that included a number of visits to each other's labs over the next three or so years and five co-authored articles. On one of those visits, I met Tom for the first time. He had just arrived in Rochester, I assume expecting to study genetic variation in mice (see Bob Selander's and Howard Ochman's chapters).

At the time, both Bob and I had technical and theoretical issues with Roger's 1973 report, which we addressed in our 1980 "excuse" paper (25). In that study, we (really Bob and his assistants) used starch gel electrophoresis, which had greater resolving power than the cellulose acetate method Roger employed and examined 20 rather than 5 enzymes. Most importantly, we then recorded the electrophoretic profiles separately for each clone, rather than for each enzyme, as did Roger. Of the 109 clones of *E. coli* we examined, 90 were included in Roger's study and the remaining 19 were clones I isolated from the feces of 16 infants and 2 adults in Massachusetts. In that study we also examined the enzyme electrophoretic profile for the three primary laboratory strains of *E. coli*, single clones of *E. coli* B and C and 24 clones of *E. coli* K-12.

Our estimate of the genic diversity of *E. coli* from this sample,  $h=0.47$ , exceeded that we calculated from Roger's data by about factor of two. More importantly, our data clearly showed that the genetic structure of *E. coli* was very different than that of the over-sexed, complete linkage-equilibrium eukaryotes assumed in the models upon which Roger's interpretations were based. Despite the relatively large amount of variability we observed for the 20 enzymes examined, there were only 98 distinct electrophoretic types in the 109 wild strains. Isolates with identical allozymes for all 20 proteins appeared in strains from two pairs of infants from Massachusetts and adults in Iowa, and a Giraffe from a zoo in Iowa and a feral sheep in California. Moreover, a clone isolated from an infant in Massachusetts was identical at 20 enzymes to those expressed by *E. coli* K-12, although the strains were quite different in the plasmids they carried.

Although Bob Selander was hardly adverse to the use of expletives, words like "eureka" weren't part of his repertoire. I, however, very distinctly recall a eureka feeling as we sat in my study in Amherst as Bob passed a ruler through a printout of the enzyme data and saw the identity and near identity of the allozyme profiles of independent isolates. In our second collaborative article (6), we provided evidence that was inconsistent with Roger Milkman's assertion of recombination occurring regularly between *E. coli* in mammalian hosts, at least not within me as the human host of this study.

Not really news: From a population genetics perspective the result of our *E. coli* enzyme variation study and its interpretation had sufficient 'man bites dog' appeal to appear in SCIENCE (a condition for publication then as well as now). On the other hand, neither the considerable variability nor the clonal population genetic structure we reported would have been surprising to microbiologists studying the epidemiology of bacteria from clinical and natural sources. This was certainly so for the late Fritz and Ida Ørskov from the Serum Institute in Copenhagen. For some 40 years they and their predecessors used serotyping to study the epidemiology of *E. coli*. (9, 20, 24). Although there were 164 O

antigens, ~100 K antigens and 56 H antigens (more than 900,000 possible combinations), isolates of the same O:K:H serotype appeared commonly from geographically and temporally different sources, and specific O:K:H serotypes were associated with different symptomatic infections.

Similar observations were made for a number of different species of bacteria with serological as well as other phenotypic markers, like phage resistance patterns (phage typing), repertoires of fermentation capabilities (biotyping), and the distribution of plasmids carried (plasmid typing). While direct sequencing of DNA was not yet in their toolbox, by the early 1980s investigators studying the genetic epidemiology of bacteria were also beginning to use various kinds of restriction endonuclease cutting procedures for these epidemiological studies. For a superb perspective on genetic epidemiology before population geneticists entered the biz, see the summary of a “Workshop on the Clone Concept in the Epidemiology, Taxonomy, and Evolution of the Enterobacteriaceae and other Bacteria” (22). This 1982 workshop that Fritz and Ida Ørskov organized at the NIH Fogarty Center was, I believe, Tom’s debut to the world of microbial genetic epidemiology. Tom gave a superb talk at that meeting presenting the results of his first bacterial population genetics article (26).

The Ørskov’s were also well aware that recombination was a rare event in *E. coli*. They had done a study of the frequency and rate of homologous gene recombination with 199 different O:K:H *E. coli* serotypes as donors and an *E. coli* K12 recipient (21). Were they population and evolutionary geneticists and not as genteel as they were, they would have probably set Roger straight for his assertion about *E. coli*’s proclivity for sex<sup>4</sup>.

### **Population Genetics becomes Genetic/Molecular Epidemiology**

In retrospect, I believe the single most important consequence of our early 1980s studies of genetic variation in *E. coli* was not direct genetic evidence for the clonal structure of *E. coli*, which the enzyme data provided, but rather the introduction of population genetics theory and approaches to the genetic epidemiology of bacteria. Perhaps as significant was the recruitment of p,q -trained population geneticists to this enterprise. Too freaking much, would you believe evolutionary biologists could do something of value for the world beyond the precious realm of academe?

---

<sup>4</sup> Joshua Lederberg also knew how rare recombination is in *E. coli*. Had he worked with any of the other then prominent laboratory strains like *E. coli* B or C, or almost any wild strain he would have gotten negative results in his 1946 experiments testing for recombination in this bacteria. The strain he used, *E. coli* K-12, was an odd-ball. It bore a plasmid, F, which was permanently de-repressed for conjugative pili synthesis and a chromosome with IS sequences homologous to those on the plasmid which enabled this conjugation-encoding accessory element to integrate into the chromosome by Rec-mediated recombination. I once asked Josh, how many other strains would he have looked at if he got negative results with K-12. He told me, “one”. Lederberg and Tatum’s preparation to do these experiments, generating the amino acid and fermentation negative mutants, was a considerable task at the time. What a combination: serendipity and a prepared, brilliant mind!

Although our 1980 Science paper may have made Motoo Kimura feel better about *E. coli* and the neutral gene hypothesis, neither this report nor subsequent studies on the population genetics of bacteria resolved the controversy, which seems to have appropriately faded away. What we did add to the genetic and now the molecular epidemiology of bacteria were explanations for why bacteria retain a clonal structure in the face of recurrent mutation and recombination (12, 16). And, thanks to Tom and others population geneticists provided the theory and tools to analyze and interpret the genetic/molecular epidemiological data. Multilocus enzyme electrophoresis (MLEE), the main population genetics tool of the 70s and early 80s (a bit of an art form with superb practitioners like Bob and Dominique Caugant as well as Howard Ochman and Tom), also had a considerable virtue over the purely phenotypic methods microbiologists were using epidemiological studies at the time. It made it possible to quantify the genetic (phylogenetic, if you prefer) relationships and distances between clones and within and between populations. To a large extent, the trees grown in the MLEE days of the last century have retained their shape into this century and I expect will beyond. What MLEE data and MLST data failed to detect was the considerable variation in size of the genomes of *E. coli* and other bacteria (see Howard Ochman's chapter) and the fundamental role of horizontal gene transfer as a source of variation for evolution in bacteria. Also missed by these multi- but not-that-many- locus typing procedures were the movements of IS elements and other more rapid changes occurring in bacterial genomes, which are important for forensic considerations and probably adaptive evolution as well.

What a good meeting can do: As part of my preparation for writing this essay, I reread Fritz and Ida Ørskov's summary of their 1982 Clone Concept workshop, (22), a literary excursion that convinced me of the seminal role that good, interdisciplinary meetings can play. Save for the forestry (trees), the anlagen of much of what is now the molecular epidemiological of bacteria were displayed at that meeting. In my interpretation, directly and indirectly, that workshop had a great deal of influence on the population and evolutionary geneticists whose work later became central to our understanding of the genetic epidemiology and population and evolutionary genetics of bacteria. I know Bob Selander, who was at the meeting, was listening from a comment he made to a speaker, "if you fuckin' people really did genetics rather than just made believe you were ....". While I don't know how much influence that workshop had on the awesome genetic/molecular epidemiological studies Bob and his students and collaborators did in the years that immediately followed, I am sure it had some, see for example(1, 5, 19). From the conversations he and I had at the workshop and after and by this collaboration criterion (4, 23, 27). I know this was also the case for Tom. I also believed that meeting convinced the more enlightened molecular biologists and epidemiologically- and clinically- oriented microbiologists of the considerable value of the technology and theory of academic population genetics, if not the well-mannered nature of its practitioners, (check out Mark Achtman's chapter).

## **In Memoriam:**

In addition to dedicating this chapter to the memory of Tom Whittam, I also dedicated it to the memories of Fritz and Ida Ørskov; all super scientists and kind, generous and open-minded people whom we greatly miss.

## **ACKNOWLEDGEMENTS**

I wish to thank Amy Kirby, Omar Cornejo and Amoolya Singh for reading this manuscript and for making helpful suggestions and acknowledge my appreciation for their considerable diplomacy in that enterprise. I also want express my considerable gratitude and admiration for Jim (James F.) Crow for insightful, always, historical information. This “rant” was prepared whilst funded by an NIH Grant GM091875.

## **REFERENCES**

1. **Achtman, M., M. Heuzenroeder, B. Kusecek, H. Ochman, D. Caugant, R. K. Selander, V. Vaisanen-Rhen, T. K. Korhonen, S. Stuart, F. Orskov, and et al.** 1986. Clonal analysis of *Escherichia coli* O2:K1 isolated from diseased humans and animals. *Infect Immun* **51**:268-76.
2. **Atwood, K. C., L. K. Schneider, and F. J. Ryan.** 1951. Periodic Selection in *Escherichia coli*. *Proc. Nat. Acad. Sci. USA* **37**:146-155.
3. **Bennett, J.** 1960. A comparison of selective methods and a test of the pre-adaptation hypothesis. *Heredity* **15**:65-77.
4. **Beutin, L., I. Orskov, F. Orskov, S. Zimmermann, J. Prada, H. Gelderblom, R. Stephan, and T. S. Whittam.** 1990. Clonal diversity and virulence factors in strains of *Escherichia coli* of the classic enteropathogenic serogroup O114. *J Infect Dis* **162**:1329-34.
5. **Caugant, D. A., B. R. Levin, I. Orskov, F. Orskov, C. Svanborg Eden, and R. K. Selander.** 1985. Genetic diversity in relation to serotype in *Escherichia coli*. *Infect Immun* **49**:407-13.
6. **Caugant, D. A., B. R. Levin, and R. K. Selander.** 1981. Genetic diversity and temporal variation in the *E. coli* population of a human host. *Genetics* **98**:467-90.
7. **Crow, J. F.** 1957. The genetics of insecticide resistance to chemicals *Annual Review of Entomology* **2**: 227-246. *Annual Review of Entomology* **2**:227-246.
8. **Dobzhansky, T.** 1951. *Genetics and the Origin of the Species*, 3rd ed, vol. Columbia University Press, New York.
9. **Evans, D. J., Jr., D. G. Evans, H. L. DuPont, F. Orskov, and I. Orskov.** 1977. Patterns of loss of enterotoxigenicity by *Escherichia coli* isolated from adults with diarrhea: suggestive evidence for an interrelationship with serotype. *Infection & Immunity* **17**:105-11.

10. **King, J. L., and T. H. Jukes.** 1969. Non-Darwinian evolution. *Science* **164**:788-98.
11. **Levin, B. R.** 1972. Coexistence of two asexual strains on a single resource. *Science* **175**:1272-1274.
12. **Levin, B. R.** 1981. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. *Genetics* **99**:1-23.
13. **Levin, B. R., and C. T. Bergstrom.** 2000. Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc Natl Acad Sci U S A* **97**:6981-5.
14. **Luria, S. E., and M. Delbruck.** 1943. Mutations of Bacteria from Virus Sensitivity to Virus Resistance. *Genetics* **28**:491-511.
15. **Lwoff, A.** 1966. The Prophage and I, p. 88-99. *In* J. Cairns, G. S. Stent, and J. D. Watson (ed.), *Phage and the Origins of Molecular Biology*. Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring Harbor.
16. **Maruyama, T., and M. Kimura.** 1980. Genetic variability and effective population size when local extinction and recolonization of subpopulations are frequent. *Proc Natl Acad Sci U S A* **77**:6710-4.
17. **Milkman, R.** 1973. Electrophoretic variation in *Escherichia coli* from natural sources. *Science* **182**:1024-6.
18. **Milkman, R.** 1972. How much room is left for non-Darwinian Evolution. *Brookhave Symp. Biol.* **23**:217-229.
19. **Musser, J. M., D. M. Granoff, P. E. Pattison, and R. K. Selander.** 1985. A population genetic framework for the study of invasive diseases caused by serotype b strains of *Haemophilus influenzae*. *Proc Natl Acad Sci U S A* **82**:5078-82.
20. **Orskov, F., I. Orskob, D. J. Evans, R. B. Sack, D. A. Sack, and T. Wadstrom.** 1976. Special *Escherichia coli* O:K:H serotypes among enterotoxogenic strains from diarrhoea in adults and children,. *Medical Microbiology and Immunology* **162**::173-180.
21. **Orskov, F., and I. Orskov.** 1961. The fertility of *Escherichia coli* antigen test strains in crosses with K12. *Acta Path. et Microbiologica Scand* **51**:280-290.
22. **Orskov, F., and I. Orskov.** 1983. From the national institutes of health. Summary of a workshop on the clone concept in the epidemiology, taxonomy, and evolution of the enterobacteriaceae and other bacteria. *Journal of Infectious Diseases* **148**:346-57.
23. **Orskov, F., T. S. Whittam, A. Cravioto, and I. Orskov.** 1990. Clonal relationships among classic enteropathogenic *Escherichia coli* (EPEC) belong to different O groups. *J Infect Dis* **162**:76-81.
24. **Orskov, I., and F. Orskov.** 1977. Special O:K:H serotypes among enterotoxigenic *E. coli* strains from diarrhea in adults and children. Occurrence of the CF (colonization factor) antigen and of hemagglutinating abilities. *Medical Microbiology & Immunology* **163**:99-110.
25. **Selander, R. K., and B. R. Levin.** 1980. Genetic diversity and structure in *Escherichia coli* populations. *Science* **210**:545-7.

26. **Whittam, T. S., H. Ochman, and R. K. Selander.** 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **80**:1751-5.
27. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. *Infect Immun* **61**:1619-29.