Perspectives

Anecdotal, Historical And Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

Recombination and Population Structure in *Escherichia coli*

Roger Milkman

Department of Biological Sciences, The University of Iowa, Iowa City, Iowa

A major focus of population genetics, and now of molecular evolution, is the study of gene lineages. Population genetic parameters usually apply to small chromosomal regions rather than to the genome as a whole, although genes do not always descend independently of their surroundings. The genome of *Escherichia coli* strikes me as an unusually favorable theater in which to observe the lineages of genes and their relationships with the evolutionary processes that operate at the population level.

I should like to trace the developing understanding of *E. coli*'s population genetics in terms of the coexistence of clonality and recombination, the recognition of restriction as important to recombination, and the emerging features of its genomic structure.

When ATWOOD, SCHNEIDER and RYAN (1951) set out to describe *E. coli* population structure, they explicitly excluded recombination, assuming *E. coli* to be a "non-sexual" species, at least outside of the lab. This simplifying assumption laid the groundwork for a clonal selection model, in which a rare, broadly favorable mutation led its entire (haploid) genome to prominence: since linkage was absolute, the entire genome hitchhiked to a high frequency as the new allele responded to selection. This *periodic selection model*, based on a series of experiments with asexual laboratory populations, contradicted my naive view of *E. coli* as having ordinary (read "Drosophila") population genetics but unburdened by diploidy, when I proposed *E. coli* as an organism with which to test the neutral hypothesis of electrophoretic variation in proteins (MILKMAN 1972, 1975; see also MILKMAN 1985). I was not convinced that the model applied to natural populations, but ATWOOD and many others (BRUCE LEVIN and ALLAN WILSON, personal communications; KOCH 1974; KUBITSCHKE 1974) felt that recombination in natural populations of *E. coli* was negligible in this context. Clonality and the frequency of recombination thus became the crucial issues. Addressing these issues and their evolutionary significance required a good look at *E. coli* from nature, and this proved more fruitful than the intended test of neutrality.

With five enzyme loci and a collection of 829 newly isolated strains of diverse natural origins, I found far less allelic electrophoretic variation than the panmictic model predicted, and inferred that the electromorphs were not neutral (MILKMAN 1973, 1975). In a study of 20 electrophoretically variable loci in some 100 of these strains and a few others, however, there were four cases of identical pairs (highly improbable by chance alone!), which clearly argued for a clonal structure of the species (SELANDER and LEVIN 1980; LEVIN 1981). My neutrality test was invalid, and neutral variation was soon addressed and abundantly demonstrated in DNA and elsewhere (MILKMAN 1985; MAYNARD SMITH 1996). Meanwhile, further support for a basic clonal structure of *E. coli* followed from comparative sequencing in the *tryptophan* (*trp*) operon, where 1-kilobase (kb) sequences from K12 (the geneticists' lab strain) and 12 of our new strains, which had diverse enzyme mobility and thermostability at other loci, were classified as follows: three were identical to K12, five were different from K12 by one nucleotide (each unique), three differed from K12 by the same set of 10 nucleotides, and one differed by 44 (MILKMAN and CRAWFORD 1983). But HARTL and DYKHUIZEN (1984) were quick to conjecture, "Conceivably, each small region of the chromosome in a group of strains could have a unique phylogeny, reflecting identity by descent, but the phylogeny might differ according to the region examined." This mosaic structure would be due, of course to recombination in the past, but at a level not remotely approaching that in a panmictic species like *Drosophila melanogaster*. In a panmictic species, individuals undergo random mating and recombination every generation.

Confirmation of the importance of recombination in nature followed in 1986, when DYKHUIZEN and GREEN (1986, 1991) sequenced 770 nucleotides in the *gnd* (6-phosphogluconate dehydrogenase) gene in nine of the 13 strains referred to above, plus the LT strain of *Salmonella typhimurium* (now part of *S. enterica*). The *gnd* gene was already known to be unusually variable (MILKMAN 1973; SELANDER and LEVIN 1980). The sequences were indeed highly varied: some strains that were identical

in *trp* were as different from one another as they were from *S. typhimurium* LT. But the striking feature was a tree entirely different from the *trp* tree. A regional mosaic of phylogenies within the genome must reflect recombination in the recent history of *E. coli*, specifically after the formation of the existing group of clones. A small, excited group met to discuss the implied genomic structure. What should we call the segments? DAN HARTL suggested isoancestral segments, which persisted until the shorter (not necessarily better) clonal segments (MILKMAN and STOLTZFUS 1988). Clearly, it was time to begin a systematic comparative molecular analysis of the *E. coli* chromosome. This would benefit greatly from the recent establishment of the ECOR (*E. coli* Reference) collection of 72 strains of diverse natural origin by OCHMAN and SELANDER (1984).

First, a comparison of restriction patterns in a large number of ECOR strains in 15 different sets of 1.5-kb PCR fragments showed a regional mosaic of similarities among strains (MILKMAN and MCKANE BRIDGES 1990). Nevertheless, certain groupings were prominent, suggesting that while some of the clonal segments were (remnants of) recombinational replacements, others were directly descended from the most recent ancestor of the clone. These were now called part of the clonal frames. Nine of the sets of PCR fragments were contiguous in and near the *trp* operon [28 min (BERLYN *et al.* 1996)], and the rest were scattered on both sides, leaving about half the chromosome (46 to 93 min) unsurveyed at that time.

The extensive multilocus enzyme electrophoresis (MLEE) studies of the 1980s (CAUGANT *et al.* 1981; WHITTAM *et al.* 1983) were of immense value in documenting extensive genetic variation and providing a basis for a tree-like phenogram defining similarity-based groupings among the ECOR strains (SELANDER *et al.* 1987; HERZER *et al.* 1990) as well as other strains, especially pathogens (SELANDER *et al.* 1996; WHITTAM 1996). The phenogram compiles degrees of difference but does not (in contrast to a true phylogenetic tree) imply genome-wide uniformity of relationships. (see also AVISE 1989.) The resolution of differences at a given locus by enzyme electrophoresis was lower than that indicated by restriction fragment length polymorphism (RFLP), but the effort required was far less for MLEE than for the five to eight digests of PCR fragments. Until more recently, the map locations of many of the enzyme loci were not known.

The restriction survey led to comparative sequencing of K12 and 36 ECOR strains, first for 4.4 kb in the *trp* operon (MILKMAN and MCKANE BRIDGES 1993), and eventually for a total of 12.7 kb (MILKMAN and MCKANE 1995; Figure 4 in MILKMAN 1996). The sequence comparisons now revealed discontinuities on a small scale: numerous clonal segments were on the order of 1 kb in length. Interestingly, despite the mosaic similarity patterns, the grouping that emerged from the sequenc-

ing closely parallels that of the overall MLEE phenogram of the ECOR strains; this confirms the prominence of DNA descended from common clonal ancestors, the clonal frames. Nevertheless, distinct highly localized phylogenies were obvious. These two patterns are easily reconciled by envisioning the descent of a chromosome in a growing clone. The DNA inherited from the clonal ancestor initially is the clonal frame. Initially, the frame is 100% of the genome, but individual genomes are punctuated progressively by small discrete recombinant replacements from outside the clone. These segments, which often appear in numerous related strains, are inferred to be related by descent and in this sense can still be called clonal segments, too. After some time, as the present situation reflects, the cells descended from an original clonal ancestor—a clonal Eve—still share possession of a large proportion of the ancestral DNA, but these residual clonal frames differ from one another in extent. It now made sense to refer to these descendants, not as a true (genomically uniform) clone, but as a meroclone—a “partial clone” (MILKMAN and MCKANE 1995; MILKMAN 1996, 1997). While we are unlikely to calculate back to the time that Eve Coli lived (unless rRNA will serve), we can in principle use the fragmentary clonal frames, each shared by some high proportion of the meroclone, to envision the most recent common ancestor of each meroclone. Clearly, this situation is different from that in sexual eukaryotes.

The three major groups of ECOR strains that emerged from MLEE studies reflected sets of specific correlated alleles at a large number of loci—that is, alleles in linkage disequilibrium. And alleles that follow the overall correlation pattern must be part of the respective clonal frames of the major groups. Similarly, agreement between local restriction site patterns or sequence patterns on one hand and the overall MLEE grouping on the other identifies them with the clonal frames. Disagreements reveal the extent of exogenous clonal segments, which may locally exceed in extent the segments of the clonal frame.

Meanwhile, LEVIN (1986) had presented theoretical and experimental evidence supporting the possibility that bacteriophages in natural populations might impose frequency-dependent selection on *E. coli* favoring rare restriction-modification types, and that this could help maintain clonal diversity. Further, PRICE and BICKLE (1986) proposed that “DNA restriction and modification systems serve to accelerate evolution by stimulating recombination . . . essentially by cleaving the incoming DNA and providing double-stranded DNA ends that are highly recombinogenic.” This view, foreshadowed or shared by others (BOYER 1964; PIT-TARD 1964; ARBER 1965; ARBER and MORSE 1965; DUBOSE *et al.* 1988), led to an interest in reconciling the mosaic patterns (particularly those seen in sequence comparisons) with recombinational events, and possi-

bly restriction. The extensive degree of polymorphism at the *hsd* restriction-modification locus has been documented in detail by MURRAY's group (DANIEL *et al.* 1988; SHARP *et al.* 1992; BARCUS and MURRAY 1995; BARCUS *et al.* 1995).

We began with the analysis of phage P1 transductants from ECOR 47 into a K12 *trpA* strain, in order to see the extent to which the DNA entering the recipient cell was incorporated into its chromosome (MCKANE and MILKMAN 1995). P1 transducing particles contain about 80 kb of donor DNA, and we anticipated that after entry into the cell this DNA might be cut by restriction endonucleases and (before or after cutting) shortened by exonucleases. Discrimination between donor and recipient DNA is possible at high resolution because, for example, ECOR 47 and K12 sequences differ by several percent of their nucleotides. With primers for a string of contiguous 1.5-kb PCR fragments (often distinguishable between ECOR 47 and K12 by several restriction sites with a single enzyme), we found donor fragment sizes averaging about 9 kb (much smaller than 80 kb, and much larger than the 1-kb size range of the observed clonal segments). And seven of the 18 transductants had more than one donor fragment. Was this due to restriction? JOHN ROTH (personal communication) had suggested following this cross with a backcross to the K12 recipient on the grounds that, aside from the ECOR 47 DNA in the *trp* region, the transductants would be identical to K12, notably in restriction-modification properties. (For the same reason that bacteria don't cut up their own chromosomes, crosses within a given strain, or between strains with identical R-M systems, are not subject to restriction.) Indeed, the results of two backcrosses, involving a transductant with a 25- or a 20-kb fragment of donor DNA, were dramatically different from those of the original cross: in the two sets of 15 backtransductants, 28 were identical to the donor transductant, and two were shortened but not split. This experimental paradigm has been followed frequently in subsequent transduction and conjugation experiments, always with comparable results. LISE RALEIGH (1987; KELLEHER and RALEIGH 1994) also deleted the known restriction genes from our K12 strain; as a recipient, this version fragmented and shortened the donor DNA far less than ordinary K12 (MILKMAN 1997). And to address the possibility that DNA mismatch plays a major role in the patterns (note that the mismatch is far more extensive in the initial cross than in the backcross), reciprocal conjugations between ECOR 47 and K12 (whose extent of mismatch would be identical) were found to be quite disparate.

Last summer in Woods Hole, I learned from ED AD-ELBERG that a postdoc in his lab, HERBERT BOYER (1964), had done some backcrosses, too. BOYER wanted to transfer arabinose genes from *E. coli* B into K12; efforts to introduce the K12 Hfr conjugation system into B worked poorly. BOYER wondered whether bacte-

ria might restrict bacterial DNA the way they did viral DNA. His backcross results were superficially similar to ours in that the restriction effects were overcome, and further crosses were successful—for a different reason. K12 Hfr × B progeny were crossed back to the K12 Hfr donor, and it turned out that in some of the initial progeny the K12 restriction-modification genes had been transferred to B. BOYER correctly localized the R-M genes to near the threonine locus, concluded that strain B could be made in this way to behave like K12, and explained the observations in terms of restriction and modification. Finally, in interrupted mating experiments, few markers were successfully transferred until 30 minutes had passed, after which time the linkage of the markers was greatly reduced from normal. He concluded that short pieces of DNA would be digested completely by exonucleases, but longer pieces overloaded the system, and some of them lasted long enough to be incorporated into the recipient chromosome. They had usually been cut into pieces by restriction endonucleases, however, and linkage was therefore reduced. Restriction polymorphism and its role in recombination had to be rediscovered when population genetics arrived in *E. coli*, and, how does that go? Phylogeny recapitulated ontogeny.

In *E. coli* the male sex factors can be carried on a free ("F") plasmid, in which case only they are transmitted in conjugation. But if the plasmid is integrated into the chromosome, the origin of transfer splits the male factors, so that some of them lead a chromosomal strand into the recipient cell, resulting in the high-frequency recombination of chromosomal genes for which the Hfr strains are named.

Chromosomal recombination in *E. coli* (FIRTH *et al.* 1996; LOW 1996; MASTERS 1996; WEISBERG 1996) can be seen as an (actually very infrequent!) two-step process. In transduction or Hfr conjugation, a donor cell (1) transmits a stretch of DNA to one or more recipient cells, after which (2) some of the DNA is incorporated into the recipient chromosome. Its size is influenced by more than the mode of transmission (transduction or conjugation); once it enters the recipient cell, it is subject to the action of restriction endonucleases, which are highly polymorphic in the species, as well as exonuclease activity. This often results in a set of small DNA fragments, which can be incorporated independently into one of one-to-several incipient chromosomes, since chromosome replication is continuous in the life of a growing bacterial cell. This permits the segregation of a small variety of recombinants, whose replacements may often be small enough to avoid containing the grab-bag of changes that might be brought in by a large stretch of DNA. After any phenotypic lag, the new recombinants are promptly exposed to natural selection. Conjugative plasmids are transmitted and are either propagated as free plasmids or, with a very wide range of probabilities, integrated into the chromosome.

Nonconjugative plasmids may be transmitted by transformation (HANAHAN and BLOOM 1996) and integrated or not.

Our reciprocal conjugations and other crosses became possible after PETER KUEMPEL provided us with an F plasmid that contains the Broca 7 (LOW 1996) Hfr origin of transfer and establishes a dynamic equilibrium between free and integrated states (the latter always at the Broca 7 region near 31 min). This plasmid has been useful in several ways: it can be conjugated into various natural isolates (e.g., ECOR strains) and laboratory marker strains, converting them to Hfrs. Indeed, some cells may carry, and transmit, both an integrated origin of transfer with a long stretch of chromosome and also a free plasmid with no chromosomal genes but all the male factors. In such a case, some transconjugants are ready (after integration) to act as Hfrs in a subsequent cross (HEINEMANN *et al.* 1996; ANKENBAUER 1997). Because of the dynamic equilibrium, these Hfrs produce fewer transconjugants than the standard fixed Hfr strains, but relatively few are needed for the intensive analysis of closely linked PCR fragments.

At this point it might appear that the periodic selection model of *E. coli* population structure and dynamics needs only a little random recombinational speckling, spread uniformly over the originally clonal genomes, but this is not entirely true. There are in fact two dramatic exceptions, and they reflect the localized interplay of powerful frequency-dependent selection, recombination, and random genetic drift. PETER REEVES and co-workers (BASTIN *et al.* 1993; REEVES 1993; LIU and REEVES 1994; LAN and REEVES 1996; STEVENSON *et al.* 1996) have described and analyzed the molecular genetics of the polymorphic O antigen, a complex lipopolysaccharide, in *Salmonella* and in *E. coli*. There are literally hundreds of different O antigens, each determined by a set of sugar synthases and transferases, whose genes have been assembled in a small region of the chromosome. The sets of such genes are not entirely homologous: the antigenic variation results, not from amino acid substitutions, or small deletions/insertions, but from the acquisition of novel genes, evidently by lateral transfer. Apparently the hosts recognize the bacterial O antigens and exercise some restraint over the growth of the cells bearing them. A novel antigen would then have a powerful advantage—as long as it is in very low frequency. Presumably this initial advantage is repeated as each new serotype colonizes host after host, in place after place. Eventually the frequency in the species rises to the level where the O antigen complex is essentially neutral, but new imports evidently continue to arrive and spread.

The pertinent interaction of random genetic drift and powerful, frequency-dependent selection occurs at the first appearance in a species or a population of a novel recombinant. Perhaps the most interesting absolute allele frequency is 1, because a mere ordinary selec-

tive advantage does not improve the probability of persistence of a single allele very much. For natural selection to gain the upper hand over drift, the probability in haploids may be about equal to the value of the selection coefficient (difference in fitness) s . If $s = 0.0001$, then this probability may be about 0.0002. [The interpretation of s is not the same as for classical diploid organisms, and the growth and multiplication dynamics of *E. coli* are not clearly isomorphic with a process characterized by random mating and a Poisson distribution of progeny number, upon which the calculation is based (CROW and KIMURA 1970, pp. 425–426).] Thus, short of drug resistance genes, there are very few individual recombinational replacements that have a good chance of becoming established. New O antigen gene complexes seem to have what it takes. In addition, the “JUMPstart” sequences in the O antigen gene complex (HOBBS and REEVES 1994) may increase the probability of the incorporation of foreign DNA there, synergizing with the subsequent action of natural selection.

In any region, the frequency of effective recombinational replacement, including horizontal transfer from phylogenetically distant (but presumably physically close) sources, is the product of the frequency of all replacement times its probability of persisting. Thus, the O antigen gene complex near 45 min may be one of the few regions of the genome that can capitalize on horizontal transfer effectively enough to demonstrate via polymorphism that it occurs at evolutionarily potentially important rates. Halfway around the chromosome, at 98 min, a region containing several restriction loci, including the highly polymorphic *hsd* restriction-modification complex, is the only presently known region with the same property. In both cases, nonhomologous replacements are flanked by regions of unusually high homologous variation stretching a few minutes on each side. The existence of these hypervariable regions (one of which is *gnd*) seems to be due to the occasional importation of valuable novelties in small packages and their redistribution among various strains by the usual processes of transduction and conjugation, by which more extensive, essentially neutral, homologous-but-not-identical substitutions flank the critical novelty (SELANDER *et al.* 1996). For a much broader set of historical inferences of horizontal transfer, not specifying frequency-dependent selection, see LAWRENCE and ROTH (1996) and LAWRENCE and OCHMAN (1997).

The present version of the periodic selection model of *E. coli* population structure/dynamics goes as follows: rare, broadly favorable mutations (*motivating alleles*) occur that carry their respective genomes to high frequencies in the species—these events are called *clonal sweeps*, or *selective sweeps*, a term applied to chromosomal regions in *Drosophila* and elsewhere as well. No genetic variability is present in a newborn clone. The clones do not reach fixation, owing either to environmental heterogeneity or to the acquisition of one another's

motivating alleles. The new clones gradually become randomly speckled with recombinational replacements from other clones. They are thus no longer clones (the genome does not have a single phylogeny applicable uniformly over its entire length), yet they share a high proportion of their common ancestral DNA (*clonal frames*), so they can be called meroclonal (*partial clones*).

Finally, the two hypervariable regions resist the clonal sweep by virtue of their novel antigens or restriction systems, which confer their own great advantages, but which lose their advantages as they rise in frequency, so that they themselves can never function as motivating alleles. The two regions are *bastions of polymorphism*.

Now the recent completion of the *E. coli* K12 MG1655 genome sequence (BLATTNER *et al.* 1997) has brought three great benefits to studies of genomic function and variation in the species. For one thing, all the DNA in this strain has been described as an ordered set of sequences presently or potentially described further in terms of functionally specific and other properties such as GC content and Codon Adaptation Index (SHARP and LI 1986). These constitute the complete set of genetic factors responsible, together with its epigenetic history, for the form and function of the organism. Second, in addition to intragenomic comparisons to reveal paralogy, genes may be compared with genes, and genomes with genomes, from other *E. coli* isolates and from other species. And finally, the genome sequence, determined with a high standard of care, is a reference by which to verify individually collected sequences, whose error frequency is presumably greater. Comparisons with genome-standard sequences can often reveal disparities that do not follow the pattern of evolutionary divergence and can thus be retested.

The population structure of *E. coli* also expresses a genomic character. Local sequence variation among the ECOR strains is an example: the tightest group of 10 ECOR strains, plus K12, show no or essentially no RFLP variation over most parts of the genome; but near the two bastions of polymorphism, variation is high and declines with distance (minutes). Thus, one can screen the entire genome for local discontinuities of this type. In *E. coli*, clonal sweeps are major events, and bastions of polymorphism are major structures—not something to be discovered by chance, but inescapable features of the species genome.

LITERATURE CITED

- ANKENBAUER, R. G., 1997 Reassessing forty years of genetics doctrine: retrotransfer and conjugation. *Genetics* **145**: 543–549.
- ARBER, W., 1965 Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.* **19**: 365–378.
- ARBER, W., and M. L. MORSE, 1965 Host specificity of DNA produced by *Escherichia coli*. VI. Effects on bacterial conjugation. *Genetics* **51**: 137–148.
- ATWOOD, K. C., L. K. SCHNEIDER and F. J. RYAN, 1951 Selective mechanisms in bacteria. *Cold Spring Harbor Symp. Quant. Biol.* **16**: 345–355.
- AVISE, J., 1989 Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution* **43**: 1192–1208.
- BARCUS, V. A., and N. E. MURRAY, 1995 Barriers to recombination: restriction, pp. 31–58 in *Population Genetics of Bacteria*, edited by R. BISHOP. Cambridge University Press, Cambridge, UK.
- BARCUS, V. A., J. B. TITHERADGE and N. E. MURRAY, 1995 The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*. *Genetics* **140**: 1187–1197.
- BASTIN, D. A., G. STEVENSON, P. K. BROWN, A. HAASE and P. R. REEVES, 1993 Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol. Microbiol.* **7**: 725–734.
- BERLYN, M. K. B., K. B. LOW, K. E. RUDD, and M. SINGER, 1996 Linkage map of *Escherichia coli* K12, edition 9, pp. 1715–1902 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- BLATTNER, F., G. PLUNKETT III, C. BLOCH, N. T. PERNA, M. RILEY *et al.*, 1997 The complete genome sequence of *Escherichia coli*. *Science* (in press).
- BOYER, H., 1964 Genetic control of restriction and modification in *Escherichia coli*. *J. Bacteriol.* **88**: 1652–1660.
- CAUGANT, D., B. R. LEVIN and R. K. SELANDER, 1981 Genetic diversity and temporal variation in the *E. coli* population of a human host. *Genetics* **98**: 467–490.
- CROW, J. F., and M. KIMURA, 1970 *An Introduction to Population Genetics Theory*. Harper and Row, New York.
- DANIEL, A. S., F. V. FULLER-PACE, D. M. LEGGE and N. E. MURRAY, 1988 Distribution and diversity of *hsd* genes in *E. coli* and other enteric bacteria. *J. Bacteriol.* **170**: 1775–1782.
- DUBOSE, R. F., D. E. DYKHUIZEN and D. L. HARTL, 1988 Genetic exchange among natural isolates of bacteria: recombination within the *phoA* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **85**: 7036–7040.
- DYKHUIZEN, D. E., and L. GREEN, 1986 DNA sequence variation, DNA phylogeny, and recombination. *Genetics* **113**: s71.
- DYKHUIZEN, D. E., and L. GREEN, 1991 Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* **173**: 7257–7268.
- FIRTH, N., K. IPPEN-IHLER and R. A. SKURRAY, 1996 Structure and function of the F factor and mechanism of conjugation, pp. 2377–2401 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- HANAHAN, D., and F. R. BLOOM, 1996 Mechanisms of DNA transformation, pp. 2449–2459 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- HARTL, D. L., and D. DYKHUIZEN, 1984 The population genetics of *Escherichia coli*. *Annu. Rev. Genet.* **18**: 31–68.
- HEINEMANN, J. A., H. E. SCOTT and M. WILLIAMS, 1996 Doing the conjugative two-step: evidence of recipient autonomy in retrotransfer. *Genetics* **143**: 1425–1435.
- HERZER, P. J., S. INOUE, M. INOUE and T. WHITTAM, 1990 Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J. Bacteriol.* **172**: 6175–6181.
- HOBBS, M., and P. R. REEVES, 1994 The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. *Mol. Microbiol.* **12**: 855–856.
- KELLEHER, J. E., and E. A. RALEIGH, 1994 Response to UV damage by four *Escherichia coli* K-12 restriction systems. *J. Bacteriol.* **176**: 5888–5896.
- KOCH, A. L., 1974 The pertinence of the periodic selection phenomenon to prokaryotic evolution. *Genetics* **77**: 127–142.
- KUBITSCHKE, H. E., 1974 Operation of selection pressure on microbial populations, pp. 105–130 in *Evolution in the Microbial World*, edited by M. J. CARLILE and J. J. SKEHEL. Cambridge University Press, Cambridge.
- LAN, R., and P. R. REEVES, 1996 Gene transfer is a major factor in bacterial evolution. *Mol. Biol. Evol.* **13**: 47–55.
- LAWRENCE, J. G., and H. OCHMAN, 1997 Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44**: 383–397.
- LAWRENCE, J. G., and J. R. ROTH, 1996 Selfish operons: horizontal

- transfer may drive the evolution of gene clusters. *Genetics* **143**: 1843–1860.
- LEVIN, B. R., 1981 Periodic selection, infectious gene exchange, and the genetic structure of *E. coli* populations. *Genetics* **99**: 1–23.
- LEVIN, B. R., 1986 Restriction-modification immunity and the maintenance of genetic diversity in bacterial populations, pp. 669–688 in *Evolutionary Processes and Theory*, edited by S. KARLIN and E. NEVO. Academic Press, New York.
- LIU, D., and P. R. REEVES, 1994 Presence of different O antigen forms in three isolates of one clone of *Escherichia coli*. *Genetics* **138**: 6–10.
- LOW, K. B., 1996 Hfr strains of *Escherichia coli* K-12, pp. 2402–2405 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- MASTERS, M., 1996 Generalized transduction, pp. 2421–2441 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- MAYNARD SMITH, J., 1996 Population genetics: an introduction, pp. 2685–2690 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- McKANE, M., and R. MILKMAN, 1995 Transduction, restriction and recombination patterns in *Escherichia coli*. *Genetics* **139**: 35–43.
- MILKMAN, R., 1972 How much room is left for Non-Darwinian evolution? pp. 217–229 in *Evolution of Genetic Systems*, edited by H. H. SMITH. Gordon and Breach, New York.
- MILKMAN, R., 1973 Electrophoretic variation in *Escherichia coli* from natural sources. *Science* **182**: 1024–1026.
- MILKMAN, R., 1975 Allozyme variation in *E. coli* of diverse natural origins, pp. 273–285 in *Isozymes. IV. Genetics and Evolution*, edited by C. L. MARKERT. Academic Press, New York.
- MILKMAN, R., 1985 Two elements of a unified theory of population genetics and molecular evolution, pp. 65–83 in *Population Genetics and Molecular Evolution*, edited by T. OHTA and K. AOKI. Japan Scientific Societies Press, Tokyo.
- MILKMAN, R., 1996 Recombinational exchange among clonal populations, pp. 2663–2684 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- MILKMAN, R., 1997 Recombination and sequence variation in *E. coli*, pp. 177–189 in *Ecology of Pathogenic Bacteria, Molecular and Evolutionary Aspects*, edited by B. A. M. VAN DER ZEIJST, W. P. M. HOEKSTRA, J. D. A. VAN EMBDEN and A. J. W. VAN ALPHEN. North-Holland, Amsterdam.
- MILKMAN, R., and I. P. CRAWFORD, 1983 Clustered third-base substitutions among wild strains of *Escherichia coli*. *Science* **221**: 378–380.
- MILKMAN, R., and M. McKANE, 1995 DNA sequence variation and recombination in *E. coli*, pp. 127–142 in *Population Genetics of Bacteria*, edited by S. BAUMBERG, J. P. W. YOUNG, E. M. H. WELLINGTON and J. R. SAUNDERS. Cambridge University Press, Cambridge, UK.
- MILKMAN, R., and M. McKANE BRIDGES, 1990 Molecular evolution of the *E. coli* chromosome. III. Clonal frames. *Genetics* **126**: 505–517.
- MILKMAN, R., and M. McKANE BRIDGES, 1993 Molecular evolution of the *E. coli* chromosome. IV. Sequence comparisons. *Genetics* **133**: 455–468.
- MILKMAN, R., and A. STOLTZFUS, 1988 Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. *Genetics* **120**: 359–366.
- OCHMAN, H., and R. K. SELANDER, 1984 Standard reference strains of *E. coli* from natural populations. *J. Bacteriol.* **157**: 690–693.
- PITTARD, J., 1964 Effect of phage-controlled restriction on genetic linkage in bacterial crosses. *J. Bacteriol.* **87**: 1256–1257.
- PRICE, C., and T. A. BICKLE, 1986 A possible role for DNA restriction in bacterial evolution. *Microbiol. Sci.* **3**: 296–299.
- RALEIGH, E., 1987 Restriction and modification *in vivo* by *E. coli* K12. *Methods Enzymol.* **152**: 130141.
- REEVES, P., 1993 Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *Trends Genet.* **9**: 17–22.
- SELANDER, R. K., and B. R. LEVIN, 1980 Genetic diversity and structure in *Escherichia coli* populations. *Science* **210**: 545–547.
- SELANDER, R. K., D. A. CAUGANT and T. S. WHITTAM, 1987 Genetic structure and variation in natural populations of *Escherichia coli*, pp. 1625–1648 in *Escherichia coli and Salmonella typhimurium Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- SELANDER, R. K., J. LI and K. NELSON, 1996 Evolutionary genetics of *Salmonella enterica*, pp. 2691–2707 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- SHARP, P., and W.-H. LI, 1986 An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**: 28–38.
- SHARP, P., J. E. KELLEHER, A. S. DANIEL, G. M. COWAN and N. E. MURRAY, 1992 Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Natl. Acad. Sci. USA* **89**: 9836–9840.
- STEVENSON, G., K. ANDRIANOPOULOS, M. W. HOBBS and P. R. REEVES, 1996 Organization of the *Escherichia coli* K12 gene cluster responsible for the extracellular polysaccharide colanic acid. *J. Bacteriol.* **178**: 4885–4893.
- WEISBERG, R. A., 1996 Specialized transduction, pp. 2442–2448 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- WHITTAM, T. S., 1996 Genetic variation and evolutionary processes in natural populations of *ESCHERICHIA COLI*, pp. 2708–2720 in *Escherichia coli and Salmonella Cellular and Molecular Biology*, edited by F. C. NEIDHARDT. American Society for Microbiology, Washington, D.C.
- WHITTAM, T. S., H. OCHMAN and R. K. SELANDER, 1983 Multilocus genetic structure in natural populations of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **80**: 1751–1755.