J Mol Evol (1995) 41:440–448

JOURNAL OF MOLECULAR EVOLUTION

Sex in *Escherichia coli* Does Not Disrupt the Clonal Structure of the Population: Evidence from Random Amplified Polymorphic DNA and Restriction-Fragment-Length Polymorphism

Patricia Desjardins,[1] Bertrand Picard,[2] Bernhard Kaltenböck,[3] Jacques Elion,[1] Erick Denamur[1]

[1] INSERM U 120, Hôpital Robert Debré, 48 Bd Sérurier, 75019 Paris, France
[2] Laboratoire de Microbiologie, Faculté de Médecine Xavier Bichat, 16 Rue H. Huchard, 75018 Paris, France
[3] Department of Pathobiology, College of Veterinary Medicine, Auburn University, Auburn, AL, USA

Received: 16 April 1994 / Accepted: 20 October 1994

Abstract. Analysis of the *Escherichia coli* population by multilocus enzyme electrophoresis (MLEE) has established its clonal organization, but there is increasing evidence that horizontal DNA transfer occurs in *E. coli*. We have assessed the genetic structure of the species *E. coli* and determined the extent to which recombination can affect the clonal structure of bacteria. A panel of 72 *E. coli* strains from the ECOR collection was characterized by random amplified polymorphic DNA (RAPD) and restriction-fragment-length polymorphism (RFLP) of the ribosomal RNA gene (*rrn*) regions. These strains have been characterized by MLEE and are assumed to reflect the range of genotypic variation in the species as a whole. Statistical analysis, including factorial analysis of correspondence (FAC) and hierarchical classifications, established that the data obtained with the three genetic markers are mutually corroborative, thus providing compelling evidence that horizontal transfer does not disrupt the clonal organization of the population. However, there is a gradient of correlation between the different classifications which ranges from the highly clonal structure of B2 group strains causing extraintestinal infections in humans to the less-stringent structure of B1 group strains that came mainly from nonprimate mammals. This group (B1) appears to be the framework from which the remaining non-A group strains have emerged. These results indicate that RAPD analysis is well suited to intraspecies characterization of *E. coli*. Lastly, treating the RAPD data by FAC allowed description of subgroup-specific DNA fragments which can be used, in a strategy comparable to positional cloning, to isolate virulence genes.

Key words: *Escherichia coli* — RAPD — RFLP — Clonal theory — Recombination

Correspondence to: E. Denamur

Introduction

The structure of *Escherichia coli* populations has been extensively studied over the past two decades. The concept of clonality (in which bacterial populations are a mixture of independently evolving clones without any significant horizontal recombination) was developed by Selander and Levin (Selander and Levin 1980) for *E. coli* at the end of the seventies and in the early eighties. This concept is based on data obtained by multilocus enzyme electrophoresis (MLEE), which indexes at the protein level reflecting the allelic variation of multiple chromosomal genes (Selander et al. 1987). The marked linkage disequilibrium between the frequently recovered alleles of one or a few multilocus genotypes on the one hand, and the global distribution of a single multilocus genotype over several decades and in unrelated hosts, on the other, suggest that the population structure is clonal and that the rate of recombination of DNA fragments is not high enough to randomize genomes or break up clonal associations.

However, there is a growing evidence from DNA se-

quencing that individual bacterial genes have a mosaic structure that could have arisen only by recombination (Maynard Smith et al. 1991). The evolutionary trees of *E. coli* obtained from comparing the sequences of several housekeeping genes (*gnd* [Bisercic et al. 1991; Dykhuizen and Green 1991], *trp* [Dykhuizen and Green 1991], *phoA* [Dykhuizen and Green 1991], and *gapA* [Nelson and Selander 1992]) are clearly not in agreement with dendrograms generated by MLEE. The relevance of these data needs to be discussed in terms of prokaryote evolution (Maynard Smith et al. 1991). It has also been recently pointed out that data from MLEE should be analyzed cautiously (Lenski 1993). Maynard Smith et al. (1993) critically analyzed MLEE data using a statistical test including subdivision of the data along biologically meaningful lines. They showed that bacterial populations are not invariably clonal, but that they occupy a spectrum of population structures ranging from that of the highly sexual *Neisseria gonorrhoeae* to the almost strictly clonal *Salmonella* (Maynard Smith et al. 1993). But the linkage disequilibrium in *E. coli* is very close to the theoretical maximum level, even when samples are taken from a restricted location (Souza et al. 1992), indicating the rare occurrence of recombination events in this organism.

Another way of assessing the structure of a population is to study other genetic markers, such as DNA polymorphism which are widespread in the genome, and to compare the deduced dendrograms to those derived from MLEE analysis. An absence of correlation is an argument for a high gene flow rate (Denamur et al. 1993), while a good correlation indicates the clonality of a population (Tibayrenc et al. 1993). Restriction-fragment-length polymorphisms (RFLPs) have been extensively used as genetic markers in population genetics since their discovery by Kan and Dozy in 1978 in the human β-globin gene (Kan and Dozy 1978). More recently, the analysis of random amplified polymorphic DNA (RAPD), based on the amplification of discrete loci with single, random-sequence, oligonucleotide primers at relaxed stringency, has become popular for generating genetic markers (Welsh et al. 1991; Welsh and McClelland 1991; Williams et al. 1990; Caetano-Anollés 1993). This method seems very promising but needs to be assessed in large-scale studies, since the possibility of artifactual outcomes has been raised (Parker et al. 1991).

The present study had two main goals: First, to validate the RAPD analysis on a large panel of strains previously characterized by another method. Second, to assess the genetic structure of *E. coli* populations by comparing three distinct genetic markers (MLEE, RAPD, and *rrn* RFLPs). To do this, we studied a set of 72 reference strains of *E. coli* isolated from a variety of hosts and geographical locations that are referred to as the ECOR collection (Ochman and Selander 1984). The strains selected from 2,600 *E. coli* isolates have been characterized by MLEE analysis and are believed to be representative of the range of genotypic variation in the species as a whole.

Materials and Methods

Bacterial Isolates. The set of 72 reference strains *E. coli* (ECOR) (Ochman and Selander 1984) was kindly provided by R.K. Selander. Details of the original sources of these strains are given by Ochman and Selander (Ochman and Selander 1984). The ECOR strains have been characterized for electrophoretically detected allelic variants of 38 enzymes (Goullet and Picard 1989; Selander et al. 1986a). Analysis of the genetic distance matrix by the neighbor-joining (NJ) method divided the strains into four main phylogenetic groups (A, B1, B2, D), plus one accessory group containing strains ECOR 37, 42, 31, and 43 (Herzer et al. 1990).

RAPD. The twenty 10-mer primer Kit A from Bioprobe Systems, Montreuil, France, plus six other primers (217δ, 8δ, 18δ, 32V3J1, 328δ, 338δ) of 15-20 bp, were used (Table 1). The conditions of the PCR reaction were as published (Cavé et al. 1994) and the PCR products were separated on 1.5% agarose gels containing ethidium bromide.

rrn *RFLPs.* Ribosomal DNA RFLPs (*rrn* RFLPs) were carried out as previously described using ribosomal 16+23S RNA from *E. coli* as a probe (Picard et al. 1991). Briefly, total *E. coli* DNA was prepared as in (Picard et al. 1991), digested with *Hin*dIII or *Eco*RI, and analyzed by Southern blotting with a chemiluminescence-labeled ribosomal probe (Bingen et al. 1992).

Data Analysis. The presence or absence of enzyme alleles or DNA bands was scored.

Factor Analysis of Correspondence (FAC). Seven FACs (Greenacre 1984) were first carried out using STAT.ITCF (STAT.ITCF 1988) with each of the endonuclease *Eco*RI and *Hin*dIII and each of the selected RAPD primer data. From these analyses, two to five rarely detected minor variables that do not form part of the structural, stable portion of the variance were discarded. A FAC was then performed with the ribosomal DNA data obtained with *Eco*RI and *Hin*dIII, and other FACs were run with the RAPD data resulting from several combinations of two, three, four, and five primers.

Phylogenetic Analysis. Cladistic analysis was performed with the phylogenetic computer program package PAUP (Swofford, D.L. 1993. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1, computer program distributed by the Illinois Natural History Survey, Champaign, IL). The extensive computation times involved and the large numbers of possible minimal trees prompted us to adopt an approximate search strategy. Random addition of taxa was used in heuristic search replicates with tree-bisection-reconnection branch swapping. No more than ten minimal trees (with respect to all previous replicates) were saved per replicate of the search. At least 100 minimal trees obtained in independent replicates for each of the enzyme, RAPD, and RFLP data sets were used to construct 50% majority-rule consensus trees. The percent frequencies of observed partitions are shown above the nodes. The tree lengths (evolutionary steps) and retention indices were 564 and 0.654 for the RAPD tree, and 150 and 0.753 for the RFLP tree.

Results

RAPD Data

The reproducibility and the stability of our RAPD analyses have been established (Cavé et al. 1994). The 26

Table 1. The RAPD primers tested and the results of the random amplification performed as described in the text on two genetically distinct *E. coli* strains (strain #6 [Cavé et al. 1994] and strain ECOR 71)[a]

Primer	5′ → 3′ Sequence	Number of pattterns	Number of DNA fragments for strain #6	Number of DNA fragments for strain ECOR 71	Number of polymorphic DNA fragments
338δ	ACCGGCCCCCGTACT	2	8	7	15
18δ	ACCTGGACCAAGGAGAAACT	2	4	5	7
328δ	CCTGTGACACCGTGGGGG	2	5	4	3
32V3J1	CAGGCCGGGCATAAGAAACT	2	5	8	6
8δ	TGTTCAGGGCCGGGG	2	5	7	4
217δ	GCCCCCAGGGGCACAGT	2	8	7	6
A-01	CAGGCCCTTC	1	1	1	0
A-02	TGCCGAGCTG	2	3	3	2
A-03	AGTCAGCCAC	2	4	5	3
A-04	AATCGGGCTG	2	3	3	4
A-05	AGGGGTCTTG	2	3	5	6
A-06	GGTCCCTGAC	2	6	4	3
A-07	GAAACGGGTG	2	7	4	7
A-08	GTGACGTAGG	/	/	/	/
A-09	GGGTAACGCC	2	6	6	2
A-10	GTGATCGCAG	2	6	7	9
A-11	CAATCGCCGT	2	5	7	4
A-12	TCGGCGATAG	2	3	4	1
A-13	CAGCACCCAC	/	/	/	/
A-14	TCTGTGCTGG	2	2	2	4
A-15	TTCCGAACCC	2	3	2	5
A-16	AGCCAGCGAA	2	5	6	8
A-17	GACCGCTTGT	/	/	/	/
A-18	AGGTGACCGT	2	2	4	6
A-19	CAAACGTCGG	2	4	6	4
A-20	GTTGCGATCC	2	8	6	5

[a] The DNA fragments that were present in one strain and not in the other are considered as polymorphics

primers were tested on two distinct *E. coli* strains that had been genotyped by esterase electrophoresis and *rrn* RFLPs (strain #6 [Cavé et al. 1994] isolated from a child suffering from meningitis and strain ECOR 71) and yielded up to eight DNA fragments per strain. We compared the electrophoretic patterns taking into account the presence or absence of the DNA bands, regardless of their relative intensity (which ranged from faint to strong). The relative intensity of individual bands was highly reproducible, as has been reported by others (Wang et al. 1993), and it may well be that polymorphism is underestimated by scoring only the presence or absence of a band without taking into account its intensity. The DNA fragments that were present in one strain but not in the others generated a polymorphism from zero (primer A-01) to 15 (primer 338δ) (Table 1). Based on these data, we selected five primers (338δ, 18δ, 328δ, 32V3J1, 8δ) which generated a large number of DNA fragments per strain (4–8) with 3–15 polymorphic fragments. These primers generated 65, 45, 41, 62, and 43 patterns among the 72 ECOR strains in kb ranges of 0.5–3.5, 0.3–1.3, 0.4–2.0, 0.5–2.0, and 0.5–2.0, respectively. An example of the patterns obtained with primers 338δ and 8δ is shown in Fig. 1. All the strains can be differentiated when all the primers are used.

rrn *RFLP Data*

Digestion with *Eco*RI gave five to eleven 2.7–23-kb rDNA fragments per strain. A total of 23 distinct fragments were obtained for the ECOR strains. The combination of these fragments generates 43 patterns among the 72 strains studied. Digestion with *Hin*dIII gave five to nine 3.7–22.5-kb rDNA fragments per strain and 17 distinct fragments were obtained for the ECOR strains. The combination of these fragments generates 49 patterns among the 72 strains studied. The combination of *Eco*RI and *Hin*dIII digests distinguished 64 ribotypes among the ECOR strains. The ribotype is the combined *Eco*RI and *Hin*dIII patterns.

Statistical Analyses

DNA fragments were numbered from the highest to the lowest molecular weight preceded by a capital letter indicating the primer (338δ = A, 18δ = B, 328δ = C, 32V3J1 = D, and 8δ = E) for RAPD data or the endonuclease (*Eco*RI = R and *Hin*dIII = H) for RFLP data.

A FAC was carried out with the RAPD data resulting from the combination of the five primers. Projections of strains in the plane F_1, F_2 (Fig. 2A), which accounted for

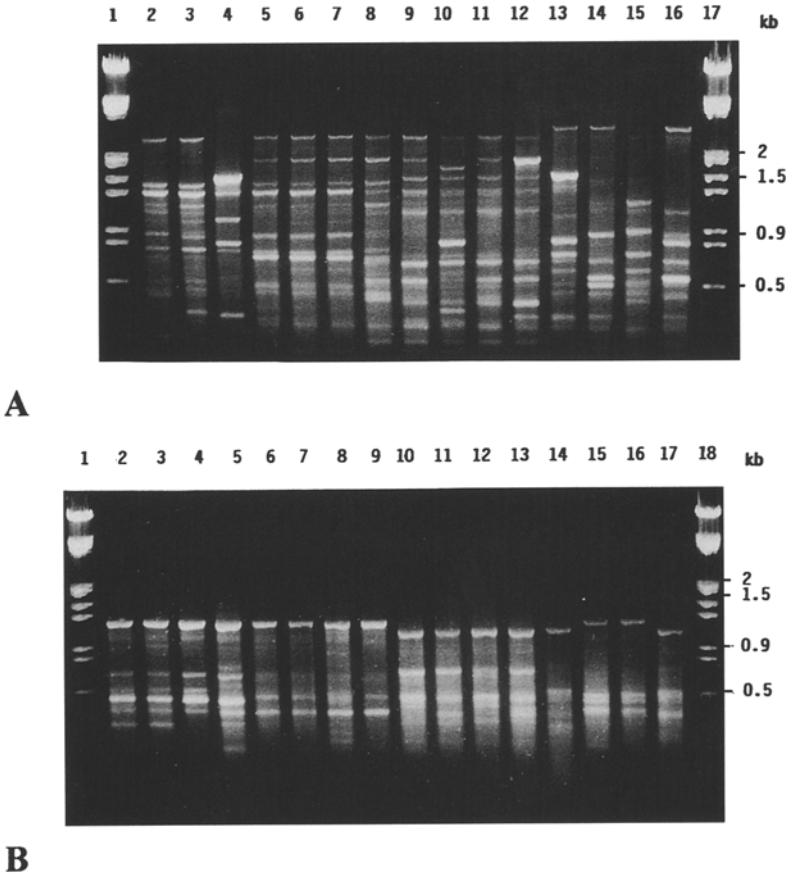

Fig. 1. RAPD gel patterns for *E. coli* ECOR strains. **A** Primer 338δ. *Lanes 1 and 17:* DNA molecular weight marker III (Boehringer, Mannheim, Germany), *lane 2:* ECOR 6, *lane 3:* ECOR 7, *lane 4:* ECOR 14, *lane 5:* ECOR 30, *lane 6:* ECOR 32, *lane 7:* ECOR 33, *lane 8:* ECOR 45, *lane 9:* ECOR 52, *lane 10:* ECOR 53, *lane 11:* ECOR 54, *lane 12:* ECOR 57, *lane 13:* ECOR 36, *lane 14:* ECOR 47, *lane 15:* ECOR 48, *lane 16:* ECOR 40. **B** Primer 18δ. *Lanes 1 and 18:* DNA molecular weight marker III (Boehringer), *lane 2:* ECOR 1, *lane 3:* ECOR 3, *lane 4:* ECOR 13, *lane 5:* ECOR 14, *lane 6:* ECOR 32, *lane 7:* ECOR 33, *lane 8:* ECOR 27, *lane 9:* ECOR 26, *lane 10:* ECOR 53, *lane 11:* ECOR 54, *lane 12:* ECOR 59, *lane 13:* ECOR 57, *lane 14:* ECOR 36, *lane 15:* ECOR 47, *lane 16:* ECOR 48, *lane 17:* ECOR 40. Strains ECOR 1, 3, 6, 7, and 14; ECOR 26, 27, 30, 32, 33, and 45; ECOR 52, 53, 54, 57, and 59; and ECOR 36, 40, 47, and 48 belong, respectively, to groups A, B1, B2, and D, as defined by Herzer et al. (1990).

32.3% of the total variance, indicated that the first axis separated the B2 group by its positive values from the A and B1 groups by its negative values, whereas group D strains were differentiated by the positive values of the second axis. Projections of the DNA fragments in the planes indicated the principal features providing the group identification of ECOR strains by RAPD. Thus, DNA fragments A20, B3, B13, D15, D17, and E2 were projected with the B2 group. Projections of strains in the plane F_1, F_3 (Fig. 2B), which accounted for 30.4% of the total variance, revealed that the third axis separated the group A strains by its positive values from the B1 strains by its negative values. Similarity between strains was used to draw a dendrogram (Fig. 3). Several phylogenetic groups were distributed into distinct lineages. The strains of groups A and B2 were the most clearly differentiated. Group A strains were well separated from the remaining ECOR strains. This group contained two main subgroups, corresponding roughly to the groups reported by Herzer et al. (1990). Group B2 strains were well clustered, at another level, with an intragroup organization similar to that obtained with MLEE data (Herzer et al. 1990). Among the remaining strains, the D and accessory-group strains had a higher level of heterogeneity, in agreement with MLEE analysis (Herzer et al. 1990), whereas group B1 strains appeared to be very heterogeneous and divided into several lineages. Strains ECOR 58 and 67, which belong to group B1, were unambiguously classified with the strains of group B2 by FAC and by the cladistic calculations.

A FAC was conduced using the *Eco*RI and *Hin*dIII rDNA data. Projections of strains in the plane F1, F2 (Fig. 4), which accounted for 44.7% of the total variance, indicated that the B2 group was separated by its positive values on the first axis, whereas group A had negative values on the two axes. The second axis differentiated the group D by its positive values. Projections of rDNA fragments in the plane indicated the essential characters identifying the group B2 strains by rDNA polymorphism: H1, H7, and H13 after *Hin*dIII digestion and R1, R4, and R14 after *Eco*RI digestion. The dendrogram resulting from the RFLP rDNA polymorphism data (Fig. 5) separated the lineage corresponding to phylogenetic group B2. The strains of group D were separated into two lineages, whereas groups A and B1 were not clearly classified.

Discussion

Starch gel electrophoresis of soluble proteins has been the workhorse technique of population genetics for nearly 30 years, but it is clear that it underestimated polymorphism. For example, the deduced amino acid sequences encoded by the *gnd* gene for 6-phosphogluconate dehydrogenase showed much more variation than

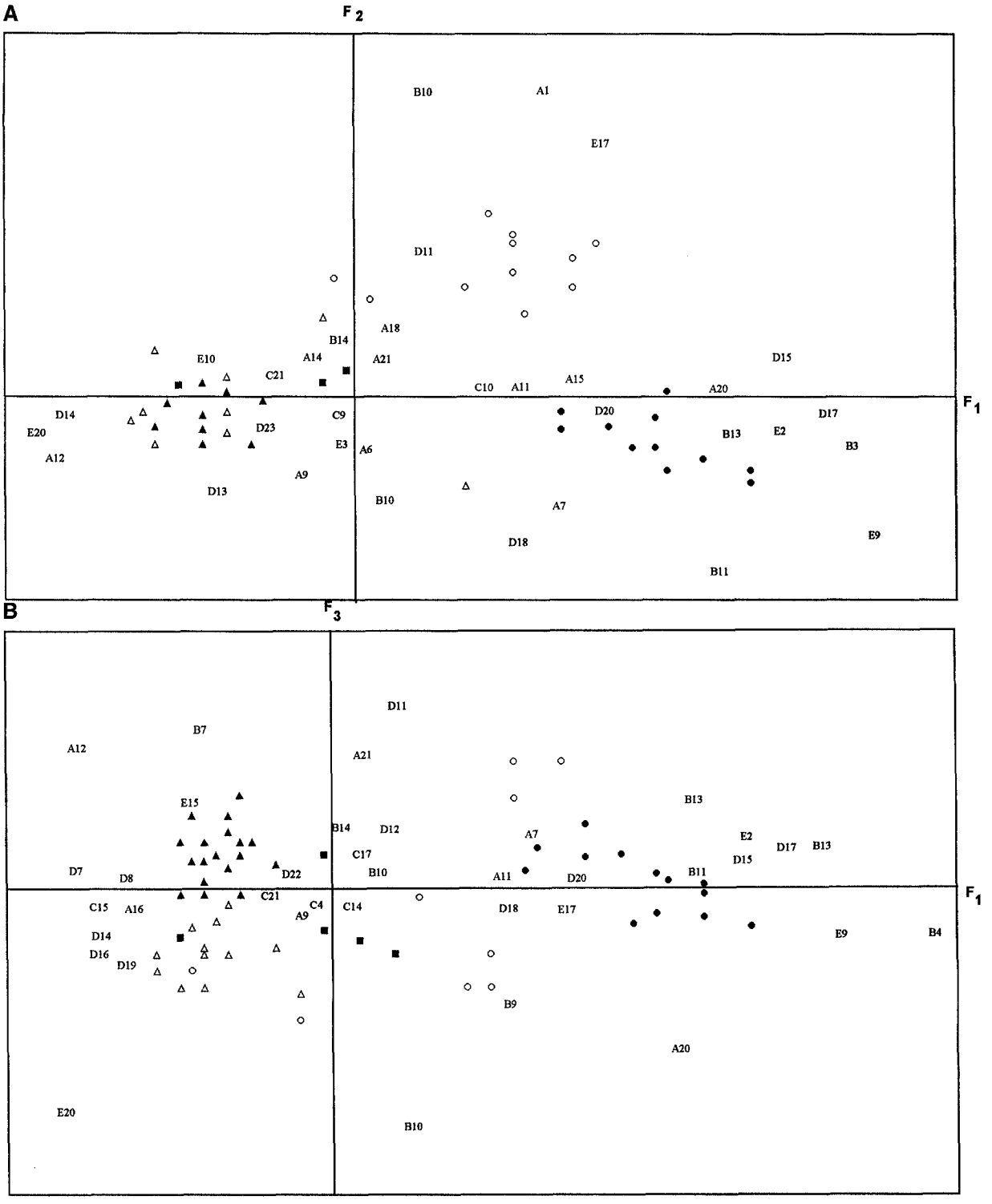

Fig. 2. Correspondence analysis of the 72 ECOR strains performed on the RAPD data. A plot of RAPD pattern score obtained in group A strains (▲), in group B1 strains (△), in group B2 strains (●), in group D strains (○) and in accessory group strains (■) and a plot score for DNA fragments obtained with primers A, B, C, D, and E for factors F_1 and F_2 **(A)** and for factors F_1 and F_3 **(B)** are shown. For clarity, when several strains are projected on the same point in the plane, only one is indicated. Based on the data of the three factors (F_1, F_2, and F_3), strains from groups A, B1, B2, and D (originally identified by MLEE) clamp together in the diagrams. This shows that a classification based on RAPD agrees with that using MLEE.

had been found by MLEE, and the net charge calculated did not correlate well with electrophoretic mobility (Bisercic et al. 1991). The working hypothesis of most population-genetic applications has also been that the majority of enzyme polymorphisms are selectively neutral, or nearly so. Several lines of statistical and experimental evidence seem to support this hypothesis in *E. coli* and other bacteria (Selander et al. 1987). However,

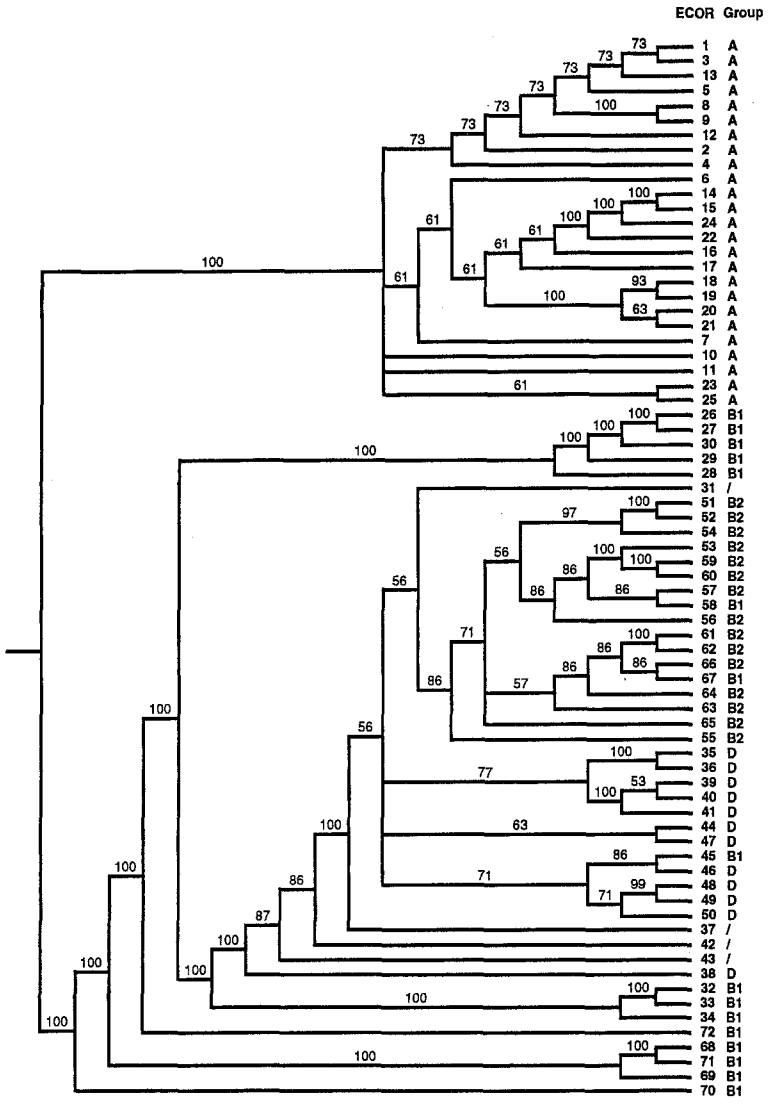

Fig. 3. Dendrogram prepared from the distance obtained from the cladistic analysis based on the RAPD data. The number of ECOR strains and their phylogenetic groups (Herzer et al. 1990) are indicated. The percent frequencies of observed partitions are shown above the nodes.

it has been shown in oysters that balancing selection can counter the influence of genetic drift (Karl and Avise 1992). Of course, MLEE still provides a rapid analysis of a large number of individuals and is an excellent method for assessing the genetic structure of a population. With the development of the PCR and the automation of DNA sequencing, determination of the nucleotide sequences of the genes themselves has become easier, and comparisons of sequences for a gene from several strains are now available as phylogenetic tools. The more striking observation resulting from these DNA sequence data is that some bacterial genes have a mosaic structure due to localized recombination (Maynard Smith et al. 1991; Brunham et al. 1993, Achtman 1994). This is particularly true for genes encoding highly antigenic cell-surface proteins, or those mediating the synthesis of polysaccharides, as well as genes implicated in the resistance to antibiotics. In such cases, the recombinant strain has clearly acquired a tremendous selective advantage that allows it to escape the defense of the host, or to be more resistant to antibiotics. But the sequencing of housekeeping genes, in which polymorphism can be considered as neutral, is needed to assess the relevance of recombination events in the genetic structure of bacteria. Several housekeeping genes, some of them conserved between distinct species, have now been sequenced and recombination events are found to occur at variable rates (Bisercic et al. 1991; Dykhuizen and Green 1991; Nelson and Selander 1992). In *E. coli,* the *gnd* locus seems to have a high rate of recombination, but several authors argue that this is because it is tightly linked to the *rfb* locus, which codes for O antigen (Dykhuizen and Green 1991). But drawing conclusions on the structure of a population based upon the study of a single gene can be misleading. Data from one gene give the phylogeny of a gene, not of a species. Several genes must be analyzed to assess the genetic structure of a species (Dykhuizen et al. 1993). However, this approach is still expensive and tedious for analysis of a large number of strains.

An alternative way to reveal the genetic structure of a population is to study several DNA genetic markers and to compare the results with MLEE data. We have used this approach to clarify the extent to which recombination can obscure phylogenetic relationships that depend

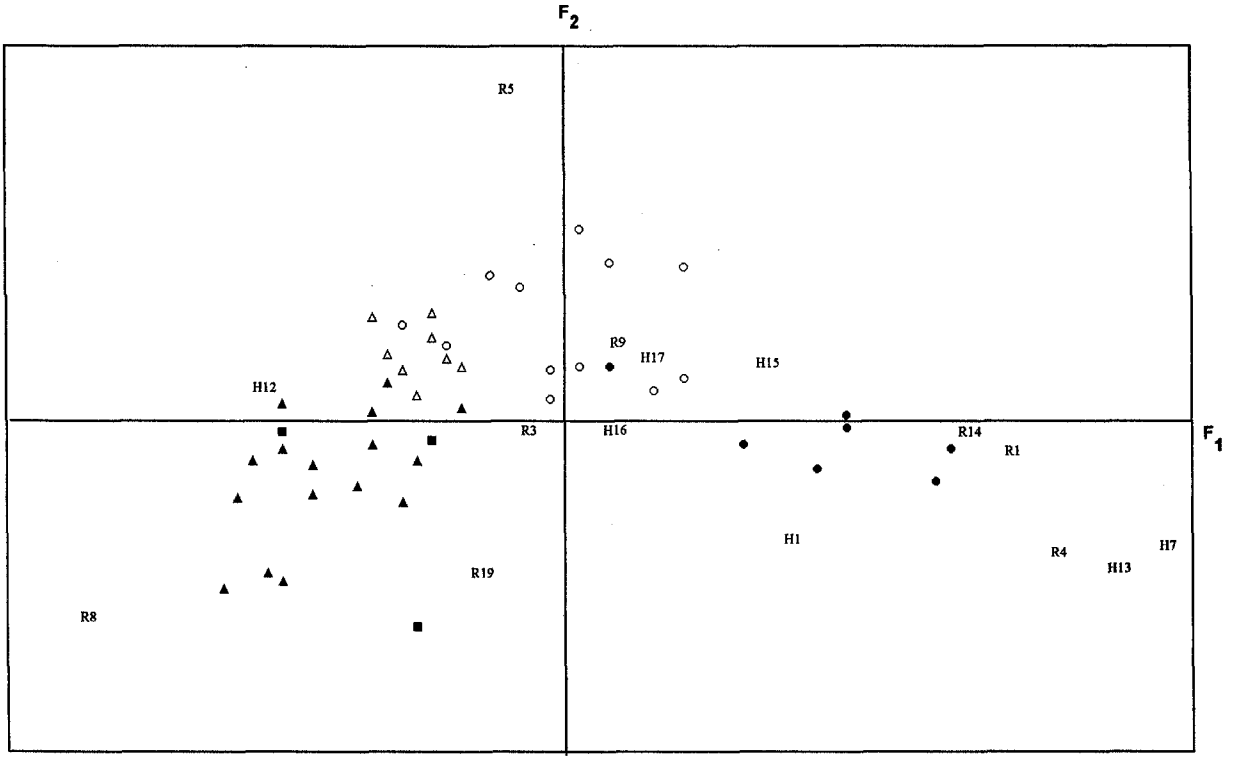

Fig. 4. Correspondence analysis of the 72 ECOR strains performed on the rDNA data obtained by digestion with *Hin*dIII and *Eco*RI. A plot of rDNA pattern score obtained in the five groups of strains (symbols as in legend to Fig. 2) and a plot of score of *Hin*dIII rDNA fragment (H) and of *Eco*RI rDNA fragment (R) for factor F_1 and F_2 are shown. For clarity, when several strains are projected on the same point in the plane, only one is indicated.

on mutational divergence in *E. coli.* Clearly, there is a good agreement between enzyme polymorphism and RAPD, and to a lesser extent with RFLP data, thus providing compelling evidence that horizontal transfer does not disrupt the clonal organization of the population. These results support theoretical estimations that point mutations are about five times more common than localized recombination events in *E. coli,* and that recombination must be two to four times more frequent than point mutation to generate a clear panmixic population (Maynard Smith et al. 1993). This study identifies major groups as well as some subgroups described by analysis of the MLEE data (Herzer et al. 1990) within the ECOR population. Phylogenetic analysis of RAPD data consistently separates *E. coli* group A strains from all other strains, as was reported from MLEE and biotype data (Selander et al. 1987). Ochman and Selander used 11 enzyme loci to distinguish group A strains by the positive value of the first axis of principal component analysis (Ochman and Selander 1984). These strains comprise K-12 and K-12-like strains, mostly isolated from humans. This must correspond to a deep and/or ancient divergence of these A strains during evolution. Among the remaining strains, group D and accessory group strains are well clustered, in agreement with the MLEE data. Group B1 strains are divided by PAUP into several unrelated groups which do not correlate with the 38 enzyme MLEE data. Ochman and Selander (Ochman and Selander 1984) found that these strains were distributed within both groups II and III. Three strains (ECOR 58, 66 and 67) that were considered to be borderline between B1 and B2 groups by MLEE data (Selander et al. 1987; Ochman and Selander 1984; Herzer et al. 1990) are clearly clustered within the B2 strains by the RAPD data. However, strains ECOR 66 and 67 have a B_2-type carboxylesterase B, while strain ECOR 58 has a B_1-type carboxylesterase B (Goullet and Picard 1989). Thus, B1 group strains seem to be heterogeneous in that they can have characters belonging to distinct lineages. These strains may have tolerated a higher recombination rate that somewhat obscured their clonal organization. It is also possible that the apparent absence of correlation between the classifications of the B1 strains obtained with the different genetic markers could result from higher polymorphism with a larger number of distinct clones. This polymorphism could correspond to the large ecological diversity of the *E. coli* species. Indeed, many of the B1 strains were originally isolated from nonprimate mammals. B1 group strains appear to form the framework from which the remaining non-A group strains have emerged.

In contrast, group B2 strains have a strictly clonal organization. These strains seem to have evolved more recently from the remaining *E. coli* than the other groups. Interestingly, *rrn* RFLP data distinguished the B2 strains from the remaining ECOR strains, and subclassified

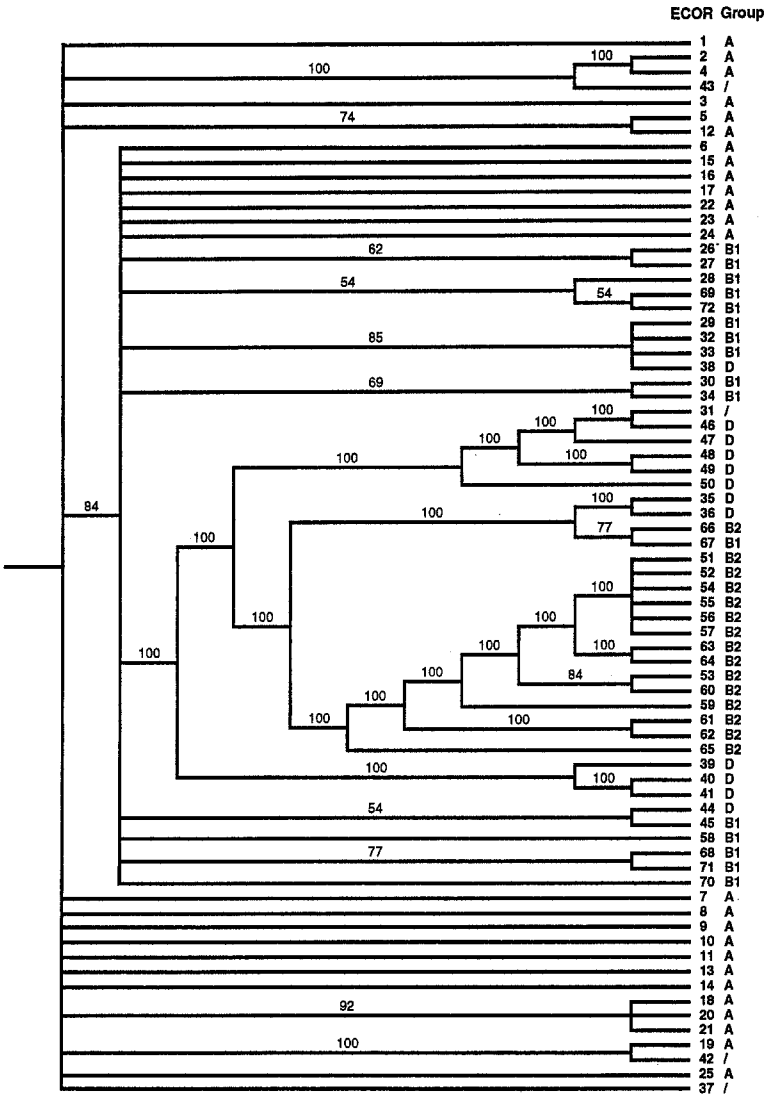

Fig. 5. Dendrogram prepared from the distance obtained from the cladistic analysis based on the rDNA data after digestion by *Eco*RI and *Hin*dIII. The number of ECOR strains and their phylogenetic group (Herzer et al. 1990) are indicated. The percent frequencies of observed partitions are shown above the nodes.

them in agreement with the MLEE data, as previously reported (Picard et al. 1993). However, the *rrn* RFLP fails to clearly identify the phylogenetic groups within the non-B2 strains. B2 strains have a B_2-type carboxylesterase, and this type of strain has been shown to cause extraintestinal infections in humans (Goullet and Picard 1986). Most of the B2 strains in the panel of ECOR were obtained from humans and other primates. Careful analysis of the available DNA sequences of housekeeping genes for B2 strains (excluding strain ECOR 67, which can be considered as a borderline strain) shows a strong conservation among them, as compared to other ECOR strains (Nelson and Selander 1992), indicating a highly clonal population structure. These pathogenic clones, in which the various isolates are indistinguishable by MLEE, biotyping, and serotyping, have a worldwide distribution and have been isolated over the last 40 years (Selander et al. 1986b; Johnson et al. 1991). The clonal structure of the population is amplified by spread in these cases. Our RFLP analysis explores a small part of the genome that corresponds to the 7 *rrn* operons. It can be hypothesized that some of these operons, as well as the carboxylesterase B locus, are in linkage disequilibrium with a block of virulence genes characteristic of these strains (Johnson 1991), or are directly involved in virulence.

Our data confirm that RAPD analysis is a very powerful technique for studying genetic structure and evolution of bacterial populations. Five primers yielded a tree similar to that obtained with 38 enzymes, and all the 72 ECOR were distinguished. When the statistical analyses were performed step by step, adding the data for one primer each time, the structure of the ECOR population was progressively refined. Wang et al. (1993) studied a total of 75 strains of *E. coli,* all causing diarrhea, and obtained from various parts of the world. They found that RAPD analysis with five primers was more sensitive for distinguishing between related bacterial strains than was MLEE with 39 enzymes. Statistical analysis of the RAPD data using FACs allows identification of the variables which characterize a group. These DNA bands can be easily excised from the agarose gel, cloned into a plasmid, sequenced, and used as a probe to convert them to RFLPs. This approach has led to the identification of

markers linked to a *Pseudomonas* resistance gene in tomato (Martin et al. 1991), and of ubiquitous somatic mutations in colorectal carcinoma cells (Ionov et al. 1993). In a strategy which can be compared to positional cloning (Ballabio 1993), it should be possible to characterize genes specific to a group of strains, and if those strains represent virulent clones, to isolate the virulence genes.

References

Achtman M (1994) Clonal spread of serogroup A meningococci: a paradigm for the analysis of microevolution in bacteria. Mol Microbiol 11:15–22

Ballabio A (1993) The rise and fall of positional cloning? Nature Genet 3:277–279

Bingen E, Denamur E, Lambert Zechovsky N, Aujard Y, Brahimi N, Geslin P, Elion J (1992) Analysis of DNA restriction fragment length polymorphism extends the evidence for breast milk transmission in *Streptococcus agalactiae* late-onset neonatal infection. J Infect Dis 165:569–573

Bisercic M, Feutruer JY, Reeves PR (1991) Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli:* evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. J Bacteriol 173:3894–3900

Brunham RC, Plummer FA, Stephens RS (1993) Bacterial antigenic variation, host immune response, and pathogen-host coevolution. Infect Immun 61:2273–2276

Caetano-Anollés (1993) Amplifying DNA with arbitrary oligonucleotide primers. PCR Methods Appl 3:85–94

Cavé H, Bingen E, Elion J, Denamur E (1994) Differentiation of *Escherichia coli* strains using randomly amplified polymorphic DNA analysis. Res Microbiol 145:141–150

Denamur E, Picard B, Decoux G, Denis JB, Elion J (1993) The absence of correlation between allozyme and *rrn* analysis indicates a high gene flow rate within human clinical *Pseudomonas aeruginosa* isolates. FEMS Microbiol Lett 110:275–280

Dykhuizen DE, Green L (1991) Recombination in *Escherichia coli* and the definition of biological species. J Bacteriol 173:7257–7268

Dykhuizen DE, Polin DS, Dunn JJ, Wilske B, Preac-Mursic V, Dattwyler RJ, Luft BJ (1993) *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. Proc Natl Sci USA 90:10163–10167

Goullet P, Picard B (1986) Highly pathogenic strains of *Escherichia coli* revealed by the electrophoretic patterns of carboxylesterase B. J Gen Microbiol 132:1853–1858

Goullet P, Picard B (1989) Comparative electrophoretic polymorphism of esterases and other enzymes in *Escherichia coli.* J Gen Microbiol 135:135–143

Greenacre MJ (1984) Theory and applications of correspondence analysis. Academic Press, Inc., London, p 364

Herzer PJ, Inouye S, Inouye M, Whittman TS (1990) Phylogenetic distribution of branched RNA-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli.* J Bacteriol 172: 6175–6181

Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363:558–561

Johnson JR (1991) Virulence factors in *Escherichia coli* urinary tract infection. Clin Microbiol Rev 4:80–128

Johnson JR, Goullet P, Picard B, Moseley SL, Roberts PL, Stamm WE (1991) Association of carboxylase B electrophoretic pattern with presence and expression of urovirulence factor determinants and antimicrobial resistance among trains of *Escherichia coli* that cause urosepsis. Infect Immun 59:2311–2315

Kan YW, Dozy A (1978) Polymorphism of DNA sequence adjacent to human β-globin structural gene: relationship to sickle mutation. Proc Natl Acad Sci USA 75:5631–5635

Karl SA, Avise JC (1992) Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science 256:100–102

Lenski RE (1993) Assessing the genetic structure of microbial populations. Proc Natl Acad Sci USA 90:4334–4336

Martin GB, Williams JGK, Tanksley SD (1991) Rapid identification of markers linked to a *Pseudomonas* resistance gene in tomato by using random primers and near-isogenic lines. Proc Natl Acad Sci USA 88:2336–2340

Maynard Smith J, Dowson CG, Spratt BG (1991) Localized sex in bacteria. Nature 349:29–31

Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? Proc Natl Acad Sci USA 90:4384–4388

Nelson K, Selander RK (1992) Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli.* J Bacteriol 174:6556–6895

Ochman H, Selander RK (1984) Standard reference strains of *Escherichia coli* from natural population. J Bacteriol 157:690–692

Parker JD, Rabinovitch PS, Burmer GC (1991) Targeted gene walking polymerase chain reaction. Nucleic Acids Res 19:3055–3060

Picard B, Picard-Pasquier N, Krishnamoorthy R, Goullet P (1991) Characterization of highly virulent *Escherichia coli* strains by ribosomal DNA restriction fragment length polymorphism. FEMS Microbiol Lett 82:183–188

Picard B, Journet-Mancy C, Picard-Pasquier N, Goullet P (1993) Genetic structures of the B_2 and B_1 *Escherichia coli* strains responsible for extra-intestinal infections. J Gen Microbiol 139:3079–3088

Selander RK, Levin BR (1980) Gene diversity and structure in *Escherichia coli* populations. Science 210:545–547

Selander RK, Caugant DA, Ochman H, Musser JM, Gilmour MN, Whittman TS (1986a) Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl Environ Microbiol 51:873–884

Selander RK, Korhonen TK, Väisänen-Rhen V, Williams PH, Pattison PE, Caugant DA (1986b) Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. Infect Immun 52:213–222

Selander RK, Caugant DA, Whittam TS (1987) Genetic structure and variation in natural population of *Escherichia coli.* In: Neidhardt FC, Ingragam JL, Low KB, Magasanik B, Schaechter M, Umbarger HE (eds): *Escherichia coli* and *Salmonella typhimurium:* cellular and molecular biology. American Society of Microbiology, Washington, DC, Vol 2, pp 1625–1648

Souza V, Nguyen TT, Hudson RR, Pinero D, Lenski RE (1992) Hierarchical analysis of linkage disequilibrium in *Rhizobium* populations: evidence for sex? Proc Natl Acad Sci USA 89:8389–8393

STAT.ITCF. (1988) Institut Technique des Céréales et des Fourages, Paris

Tibayrenc M, Neubauer K, Barnabé C, Guerrini F, Skarecky D, Ayala FJ (1993) Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc Natl Acad Sci USA 90:1335–1339

Wang G, Whittam TS, Berg CM, Berg DE (1993) RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. Nucleic Acids Res 21:5930–5933

Welsh J, McClelland M (1991) Genomic fingerprinting using arbitrarily primed PCR and a matrix of pairwise combinations of primers. Nucleic Acids Res 19:5275–5279

Welsh J, Petersen C, McClelland M (1991) Polymorphism generated by arbitrarily primed PCR in the mouse: application to strain identification and genetic mapping. Nucleic Acids Res 19:303–306

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18:6531–6535