APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2000, p. 5104–5105 Vol. 66, No. 11
0099-2240/00/$04.00+0

Letter to the Editor

Genetic Diversity within *E. coli*

The results shown for the 13 strains from the *Escherichia coli* reference (ECOR) collection that Souza et al. included "as a reference" in their recent allozyme analysis of diverse natural *E. coli* isolates (8) fail to validate the newly derived phylogeny (Fig. 1 in reference 8), leave mysterious the correspondence between this phylogeny and that of the ECOR collection (1–5, 7), and raise serious concerns regarding the validity of the investigators' allozyme data and the conclusions derived therefrom (8).

In the dendrogram of Souza, et al. (Fig. 1 in reference 8), the eight ECOR group A control strains are variously placed in the new "ancestral cluster" (along with the four group B1 strains) and in clusters G and C (together with the sole ECOR group D control strain), i.e., across the breadth of the tree and in association with members of the two other ECOR groups studied. This conflicts with previous analyses of the ECOR strains, which with few exceptions have placed the group A strains close together, apart from representatives of other ECOR groups, irrespective of cluster analysis method (e.g., principal-component analysis, UPGMA, neighbor joining, or parsimony) or type of data set (e.g., starch or cellulose acetate gel allozyme analysis, comparative DNA sequencing, or PCR fingerprinting) (1–5, 7).

A direct comparison of allozyme data for the 13 ECOR controls as provided by Souza et al. (8), and as obtained from the ECOR database at the Thomas Whittam laboratory web site (http://www.bio.psu.edu/People/Faculty/Whittam/Lab/), for the six loci at which these two data sets coincide is informative (Table 1). Souza et al. report more polymorphisms at five of the six loci than are documented in the ECOR database. These discrepancies are most marked for the eight group A ECOR strains, among which Souza et al. list twice as many allelic variants (not counting nulls) over the six loci as does the ECOR database (Table 1). According to Souza et al., but not the ECOR data base, ECOR strains 5 and 8 exhibit distinct alleles for IDH, MDH, MPI, and/or PGM as compared with the other six group A strains (data not shown), evidence suggesting that different strains may have been tested as ECOR 5 and 8 in the different analyses. Still, Souza et al. report more diversity even among the remaining six group A ECOR strains than is found in the ECOR database (data not shown).

TABLE 1. Allozyme polymorphisms among 13 ECOR control strains[a] according to two different sources

Locus	No. of alleles per locus among ECOR control strains			
	All ECOR controls (n = 13)		Group A controls (n = 8)	
	Souza et al.[b]	ECOR database[c]	Souza et al.[b]	ECOR database[c]
ADH	3	2 (+ 1 null)	2	1 (+ 1 null)
G6PD (6G[b])	1	4	1	2
IDH	4	2	3	2
MDH	3	2	2	1
MPI	5 (+1 null)	3	4 (+ 1 null)	1
PGM	4	3	4	1
Total	20	16	16	8

[a] The ECOR controls included ECOR strains 1, 5, 7, 8, 10, 11, 12, and 23 (group A), 26, 27, 33, and 45 (group B1), and 41 (group D).
[b] Per reference 8.
[c] Per Thomas Whittam web site (http://www.bio.psu.edu/People/Faculty/Whittam/Lab/).

That these discrepancies are not simply a matter of cellulose acetate gels (8) versus starch gels (ECOR database) is shown by the data of Pupo et al. (5). Using cellulose acetate gel allozyme analysis, these investigators found the group A ECOR strains to be quite homogeneous across 10 enzyme loci, including the six shown in Table 1, and derived a neighbor joining tree quite similar to that obtained by Herzer et al. using starch gels and 38 loci (3, 5). Thus, the data of Souza et al. (8), at least for the group A ECOR strains, are of uncertain validity. Confidence in the remainder of these investigators' allozyme data set is weakened by the inclusion of an enzyme (ARK, arginine kinase) that is without precedent in the *E. coli* allozyme literature and the exceptional finding of two loci for ME (malic enzyme) (3, 6). Consequently, the tree (Fig. 1 in reference 8), the genetic diversity calculations (Table 2 in reference 8), and all inferences drawn from these analyses are suspect and provide no support for the revised understandings of phylogenetic relationships within *E. coli* proposed by the authors (8).

REFERENCES

1. **Desjardins, P., B. Picard, B. Kaltenbock, J. Elion, and E. Denamur.** 1995. Sex in *Escherichia coli* does not disrupt the clonal structure of the population: evidence from random amplified polymorphic DNA and restriction-fragment-length polymorphism. J. Mol. Evol. **41:**440–448.
2. **Guttman, D. S., and D. E. Dykhuizen.** 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science **266:**1380–1383.
3. **Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam.** 1990. Phylogenetic distribution of branched RNS-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. J. Bacteriol. **172:**6175–6181.
4. **Lecointre, G., L. Rachdi, P. Darlu, and E. Denamur.** 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. Mol. Biol. Evol. **15:**1685–1695.
5. **Pupo, G. M., D. K. R. Karaolis, R. Lan, and P. R. Reeves.** 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. Infect. Immun. **65:**2685–2692.
6. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:** 873–884.
7. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *In* F. C. Neidhardt, K. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
8. **Souza, V., M. Rocha, A. Valera, and L. E. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Appl. Environ. Microbiol. **65:**3373–3385.

James R. Johnson
Infectious Diseases (111F)
Minneapolis VA Medical Center
1 Veterans Drive
Minneapolis, MN 55417

Author's Reply

Although *Escherichia coli* is perhaps the best known organism in terms of genetics and physiology, we still have very limited knowledge of the ecology of its natural populations.

For this reason we decided to organize a collection of *E. coli* isolates from wild animals of different origins, orders, and diets and analyzed it in terms of sugar utilization, antibiotic resistance, toxin production, and plasmid profiles as well as a preliminary population genetics analysis based on 11 loci by using MLEE in cellulose acetate membranes (7).

Dr. Johnson is concerned with our MLEE analysis of the ECOR collection. We are completely confident of our results because we used a double blind method for the labeling of the strains, and we ran our samples as many times as necessary to be sure that all strains considered to have the same allele had exactly the same mobility. This rigorous reading explains the presence of new alleles when small differences in mobility were detected. The fact that we have a diversity for the ECOR different from that in reference 5 is due to differences in procedure. In their study, the ECOR was used as a standard, and these strains were "forced" to have the same alleles as the ones reported in the original work (4) (see Materials and Methods in reference 5). This had the artificial effect of obtaining all the readings of the gels relative to the described bands (6). We did not use the ECOR as a standard but just as additional strains. Besides, due to the needs of so many different strains, we grew the strains in glucose-enriched minimal media instead of LB. We also used a brand of cellulose acetate membranes and buffers and enzymes different from those used by Pupo et al. (5).

Dr. Johnson is also worried about our usage of the enzyme arginine kinase, EC 2.7.3.3. This is an allozyme (described in reference 2) that is repeatable and variable in most bacterial species that we have analyzed. This enzyme behaves as an "average" locus in *E. coli* (7), and its inclusion or deletion does not change the described patterns.

Another concern of Dr. Johnson is that the ECOR collection does not cluster in our dendrogram as previously described. We consider our dendrogram only a visual expression of the statistical relationships of a set of markers that are mainly (but not only) under strong genetic control. It is not surprising that the relationships among the strains may appear different when different sets of strains, different genetic markers, and different distance methods are used, in particular considering the small number of characters (11 loci) and the large number of strains (202). We avoided naming our dendrogram a "phylogenetic tree" as we were aware that recombination among and within loci is an important evolutionary force in this species (1, 3). For instance, in reference 3, the ECOR group A is separated into many different groups in the *finO*, *trpC*, and *gnd* trees.

We consider that neither our collection nor the ECOR collection represents the complete ecology of *E. coli*, and thus each collection helps us in a different way to understand what is *E. coli*. Because recombination is an important evolutionary force in this species, there is not a single phylogenetic tree for the strains, and in consequence we can analyze them only by using statistical descriptions of the allelic frequencies.

REFERENCES

1. **Dykhuizen, D. E., and L. Green.** 1991. Recombination in *Escherichia coli* and the definition of biological species. J. Bacteriol. **173:**7257–7268.
2. **Hebert, P. D., and M. J. Beaton.** 1993. Methodologies for allozyme analysis using cellulose acetate electrophoresis. Helena Laboratories, Beaumont, Tex.
3. **Lecointre, G., L. Rachdi, P. Darlu, and E. Denamur.** 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. Mol. Biol. Evol. **15:**1685–1695.
4. **Ochman, H., and R. K. Selander.** 1984. Standard reference strains of *Escherichia coli* from natural populations. J. Bacteriol. **157:**690–693.
5. **Pupo, G. M., D. K. R. Karaolis, L. Ruiting, and P. R. Reeves.** 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. Infect. Immun. **65:**2685–2692.
6. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Muser, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873–884.
7. **Souza, V., M. Rocha, A. Valera, and L. E. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. Appl. Environ. Microbiol. **65:**3373–3385.

Valeria Souza
Martha Rocha
Aldo Valera
Luis E. Eguiarte
Departamento de Ecología Evolutiva
Instituto de Ecología
Universidad Nacional Autónoma de México
Apartado Postal 70-275
Coyoacán 04510 México D.F., México