The Use of Molecular Techniques to Differentiate *Escherichia coli* Strains and Identify Genomic Changes in Evolving Populations of *E. coli* cells

Catherine Ruggiero

ABSTRACT

The use of molecular techniques is critical to detecting evolutionary change in populations. Random Amplified Polymorphic DNA (RAPD) analysis was performed to see whether genomic differences could be identified in evolving populations of *Escherichia coli*. The methodology was first applied to two different strains of *E. coli* (JM109 and K12) to confirm that the RAPD analysis was capable of detecting differences in closely related genomes. Following confirmation that this technique possessed the strength to differentiate the strains, RAPD analysis was performed on DNA extracted from evolving populations of *E. coli* cells. Sequencing information identified the DNA amplified by RAPDs as homologous to *E. coli* and also illustrated where differences between the two strains were located in the genome. Although genomic differences among evolving populations of a single *E. coli* strain were not identified here, I conclude that RAPD analysis can be used to differentiate between very similar genomes and could be useful in detecting evolutionary changes among populations of the same species or strain.

INTRODUCTION

Evolution can be monitored by examining changes in the genomes of populations. Detecting these changes requires a molecular technique that allows us to highlight differences between samples of genomic DNA. Ideally, such methods would also enable us to identify areas of the genome where changes have occurred in order to elucidate the type of change that has

taken place (i.e. inversions, deletions, changes in transposable element copy number, etc.). Exploring polymorphisms within and between species not only provides insight into genetic and phenotypic differences, but can also be useful in tracing evolutionary histories (Liti et al., 2009; Schacherer et al., 2009).

Two methodologies that show promise for detecting genomic-level changes among evolving populations include Restriction Fragment Length Polymorphism (RFLP) analysis and Random Amplified Polymorphic DNA (RAPD) analysis. Both techniques allow investigators to create a barcode that is specific to a species or strain based on banding patterns from DNA fractionated on a gel. The ability to quickly and easily identify a species or subspecies in this way could be an indispensable tool for researchers working on any number of organisms.

RFLP utilizes restriction enzymes to digest DNA at specific restriction sites to create DNA fragments of different sizes. Distinct organisms will have unique fragment lengths as a result of enzyme digestion of their genomes due to mutations or differences in the location of the restriction sites. This method is often paired with a radioactive probe via Southern Blotting to detect potential fragments of interest among the many that are produced by restriction enzymes. It can also be combined with PCR and sequencing to yield more specific information about an organism's genotype and even its host species (Kundu et al., 2009; Zhou et al. 2009).

RFLP has successfully been used to differentiate species and detect genetic variability among eukaryotic (Lee et al., 2009; Lyra et al., 2009) and prokaryotic species (Okhravi et al., 2000). It has also been used to examine evolving populations of *E. coli* at various generations (Papadopoulos et al., 1999; Schneider et al., 2000). RFLP without the use of Southern Blotting was attempted prior to the RAPD analysis presented in this paper. Although the methodology seemed promising for detecting genomic changes, it proved difficult with the time restraints of

this project and was abandoned in exchange for the RAPD analysis. More work with RFLP should be conducted in the future, as it holds many possibilities for distinguishing between species as well as evolving populations of the same species.

RAPD is a PCR-based analysis that uses small arbitrary primers to produce random PCR products. RAPD analysis has been used to create profiles and identify the degree of genetic variation in plant species (Da Mata et al., 2009; Narzary et al., 2009). It has also been used in *E. coli* to differentiate isolates from different sources (Taghi et al., 2008; Tseng et al., 2001), identify different strains (Kilic et al., 2004), and even detect genetic rearrangements due to environmental stressors (Jolivet-Gougeon et al., 2000). Therefore, it may be a useful tool to investigate evolutionary change between populations.

This experiment utilized the RAPD methodology to investigate differences among evolving populations of *E. coli*. It has been shown that bacterial populations accumulate a variety of genomic rearrangements, frame shifts, deletions, and gene inactivations during their evolution (Moran et al., 2009). In order to investigate whether genomic changes could be detected in evolving populations of *E. coli*, I first showed that RAPDs can be used to differentiate between closely related strains of *E. coli* (JM109 and K12). Once this was established, I used RAPD analysis to do a preliminary investigation into the differences between populations of K12 *E. coli* that have been evolving in the lab. By examining populations that had evolved for at least 2,000 generations under different conditions, as well as the starting population they evolved from, I was able to test whether the RAPD methodology could distinguish between samples of the same strain at different evolutionary time points. Although I did not detect any differences between the evolving populations themselves, or between the evolved populations and their starting

population, I am confident that the RAPD analysis used here can be employed to detect such changes if they exist.

MATERIALS & METHODS

DNA Samples of E. coli - Cell cultures of the K12 strain of *E. coli* bacteria that had been exposed to different evolutionary conditions were obtained. The starting population, known as K12 generation 0 (K12 Gen.0), was the population from which the two evolved samples of K12 had originated from. The two evolved samples were each grown in different volumes of media (2 ml for the small populations and 10 ml for the large populations) and bottlenecked every 24 hours at 10 cells (small populations) or 250 million cells (large populations). At the time of this experiment, the evolved populations had all passed 2000 generations. A cell culture of JM109 *E. coli* was used as a positive control to investigate differences between the genomes of two more distantly related populations, as JM109 is a direct descendent of K12 (Maillet et al., 2007). Genomic DNA was extracted from cell cultures of JM109 and K12 strains of *E. coli* using the DNeasy Blood & Tissue Kit (QIAGEN, USA). DNA samples were evaluated by spectrophotometry to check concentration and quality.

RAPD Analysis - PCR reactions were run with 10-mer RAPD primers OPA-01 through OPA-10, obtained from Operon Technologies (Alameda, CA). Each 25 µl reaction was carried out with a single primer in the following mixture: 12.5µl 2X GoTaq from Promega (Madison, WI), 1.0 µl primer (10 pmol/µl), 1.0 µl genomic DNA (20 ng/µl), and 10.5 µl dH_2O. The samples were amplified in a thermocycler with the following program: 1 cycle at 94°C for 5 minutes, followed by 50 cycles at 94°C for 30 seconds, 35°C for 1 minute, and 72°C for 2 minutes,

followed by 1 cycle at 72°C for 5 minutes and a hold at 4°C. PCR products were fractionated on a 1% agarose gel.

Sequencing Procedure - Bands of interest were cut out of the gel and purified using the QIAquick Gel Extraction Kit (QIAGEN, USA). The DNA obtained was then ligated into the pGEM-T Easy vector from Promega (Madison, WI) using the following ligation reaction: 2.5 µl 2X Ligation Buffer, 0.5 µl T4 Ligase, 0.5 µl pGEM-T vector (25 ng/µl), and 1.5 µl extracted DNA. The ligation reaction ran overnight at 4°C. Transformation of the vector into JM109 *E. coli* cells followed the manufacturer's protocol (Promega, Madison, WI). The transformation products were plated onto Amp/X-Gal/IPTG LB agar plates and incubated at 37°C overnight. Two white colonies, indicating that they contained the plasmid with the DNA insert, were removed from each plate and added to 5 ml LB media containing 5 µl Ampicillin to maintain selection pressure. These cultures were incubated overnight at 37°C. The plasmid was purified from the cell culture using the QIAprep Miniprep spin column (QIAGEN, USA). The isolated DNA was then sequenced by GENEWIZ (South Plainfield, NJ).

RESULTS

DNA extracted from JM109 and K12 Gen.0 *E. coli cells* was used in RAPD reactions with RAPD primers OPA-01 through OPA-10 (Figure 1). Primers that produced bands at different positions in the two DNA samples were assumed to be recognizing genomic differences between the strains. Primers OPA-04 and OPA-07 showed different banding patterns between JM109 and K12 Gen.0 and were therefore chosen for further examination using DNA samples from the evolving populations of K12 *E. coli*.

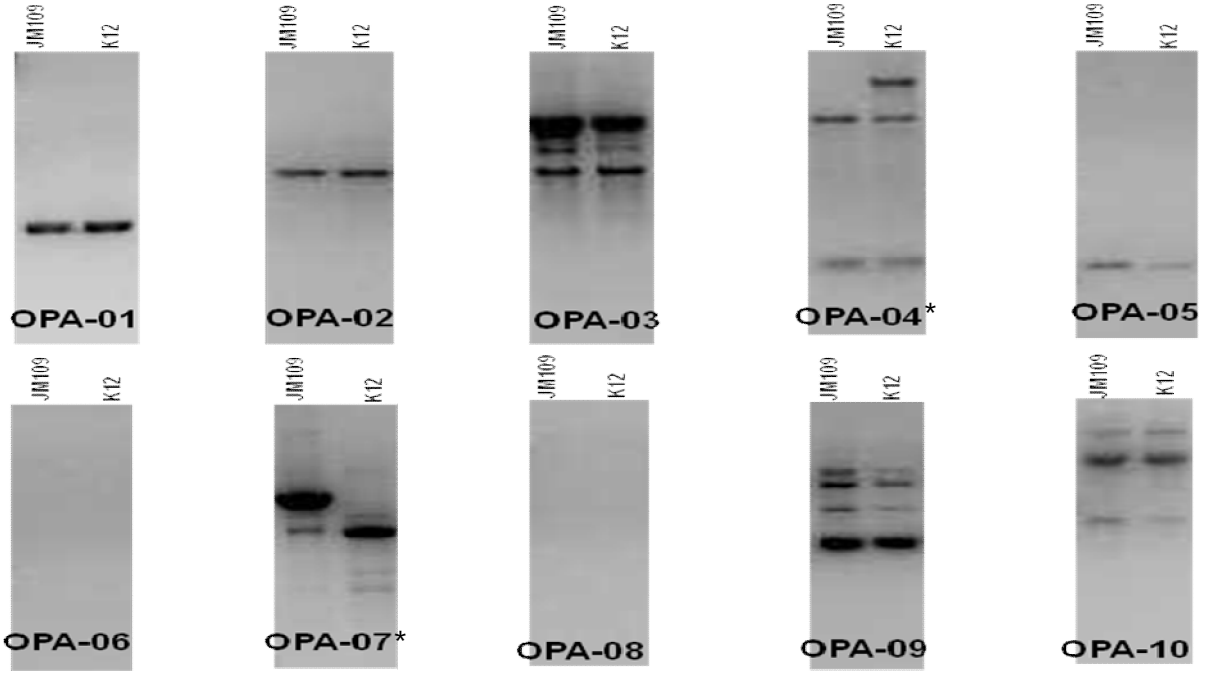


Figure 1. RAPD analysis of JM109 and K12, Gen.0 *E. coli* strains. Primers used in analysis (OPA-01 through OPA-10) are indicated below each gel picture. * Indicates primers chosen for further analysis.

The RAPD analysis was repeated using primers OPA-04 and OPA-07, this time including the small and large evolved population DNA samples of K12 *E. coli*. The goal was to investigate whether the evolved populations were different from their starting population. I also wanted to see if there were any differences between the small and large evolved populations since they were generated under such different conditions (see "Materials & Methods"). Different banding patterns were again observed between the two strains (JM109 and K12, Gen.0). However, no differences were seen between the small and large evolved populations (K12 Small Evolved and K12 Large Evolved). In addition, there were no notable differences between the evolved populations and the starting population they evolved from (K12 Gen.0) (Figure 2). Bands were cut out of this gel for sequencing to confirm the PCR reaction had amplified DNA from *E. coli* cells and also to identify regions where genomic differences existed.

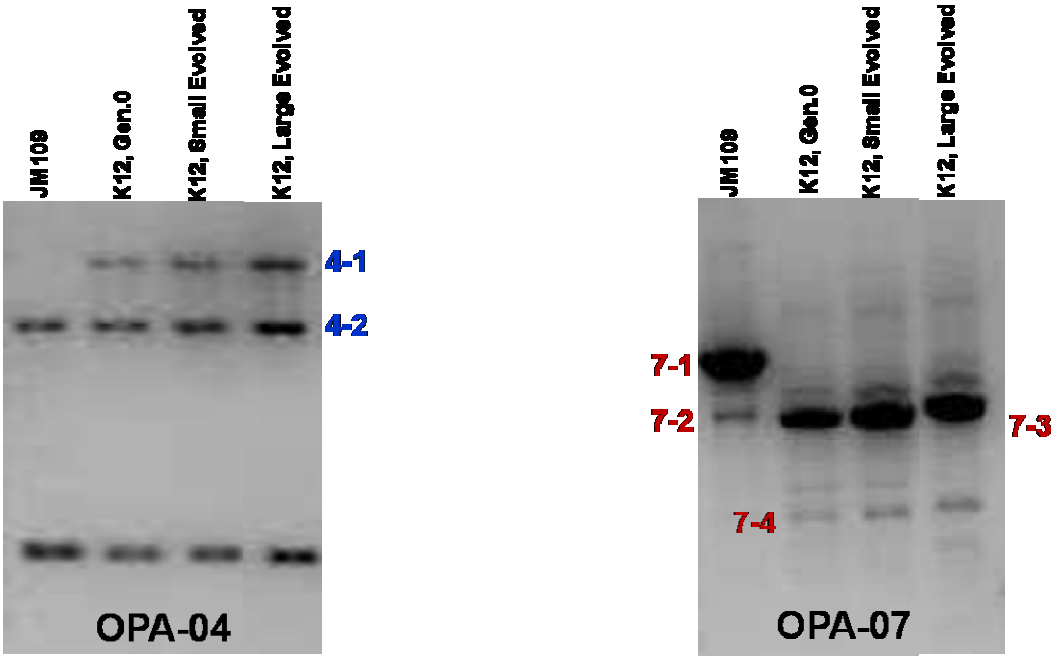
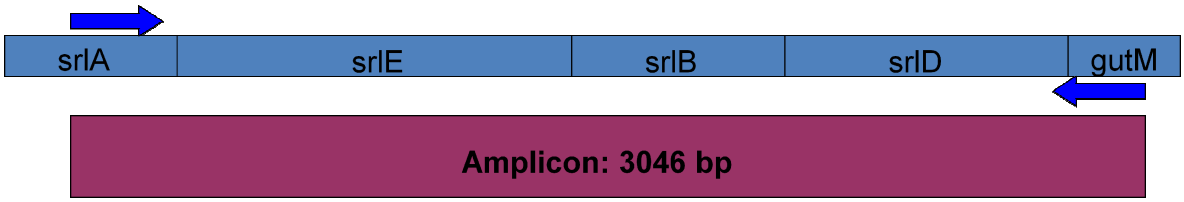


Figure 2. RAPD analysis of JM109 and K12 populations of *E. coli*. RAPD primers OPA-04 and OPA-07 were used to analyze samples of evolved small and large populations of K12 *E. coli* in addition to JM109 and K12 Gen.0 samples. Bands cut out for sequencing are numbered.

The DNA fragments were isolated from the agarose gel as described under "Materials & Methods". Sequencing of these fragments produced sequences that were homologous to *E. coli* with high sequence identity ($\geq$ 98%). Band 4-1 is a large amplicon (3046 base pairs) present only in the K12 samples that spans gene srlA (sorbitol-6-phosphate dehydrogenase) through gene gutM (DNA binding transcriptional activator of glucitol operon) (Figure 3A). Band 4-2 is a fragment present in all samples that is homologous to gene yfaW (predicted enolase). Band 7-1, present only in the JM109 sample, has a sequence that matches F plasmid DNA. Band 7-2 (from JM109) and 7-3 (from all K12 samples) match genes srlA and srlB (glucitol/sorbitol specific enzyme components II of the phosphotransferase system). Band 7-4, a small amplicon (582 base

pairs) present only in the K12 samples, lies within gene rhtA (homoserine and threonine efflux system) (Figure 3B).

A. Band 4-1



B. Band 7-4

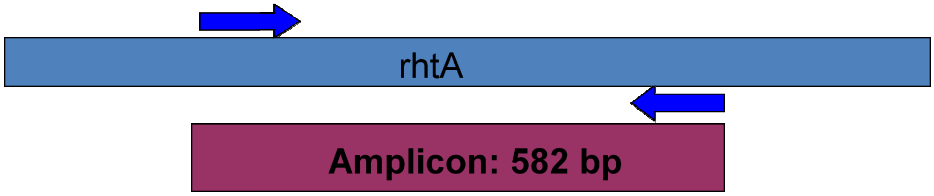


Figure 3. Sequence information for bands 4-1 and 7-4. **A.** Sequence information from *E. coli* based on BLAST matches to amplicon produced from RAPD analysis with primer OPA-04 and DNA from K12 *E. coli* samples. Arrows indicate position of primer OPA-04 annealing in *E. coli* sequence (slrA and gutM). **B.** Sequence information from *E. coli* based on BLAST matches to amplicon produced from RAPD analysis with primer OPA-07 and DNA from K12 *E. coli* samples. Arrows indicate position of primer OPA-07 annealing in *E. coli* sequence (rhtA).

DISCUSSION

Methods for investigating genomic differences can be utilized to identify different species or strains, and may be useful in detecting evolutionary changes in populations of the same strain. Here I showed that RAPDs can be used to differentiate between strains of *E. coli* bacteria. Sequencing of the fragments of DNA produced by PCR confirmed that *E. coli* DNA was amplified by the RAPD primers. Based on sequence information, it is clear that band 4-2 was the same fragment present in both strains JM109 and K12. The same is true for bands 7-2

and 7-3, which had sequence homology to the same genes based on BLAST results and are therefore the same fragment. Band 7-1, present in JM109 only, matched F plasmid DNA. The F plasmid has been removed from the K12 strain and therefore would not be expected to produce such a band with the OPA-07 primer (Blattner et al., 1997). Bands 4-1 and 7-4 were present in the K12 samples but not in JM109. Possible explanations for absence of these fragments in JM109, based on the sequence matches in K12 (Figure 3) include a lack of sequence homology in JM109 with the primer at these locations, rearrangement of the genome in this region that would result in one or both primer locations moving to another part of the genome, or an insertion between the two primers in this region. It is also possible that the genes identified by BLAST are mutated or missing in JM109. Although this is a less likely scenario, there is some evidence that the rhtA gene, responsible for the 7-4 band in K12, may not be present in all strains or substrains of *E. coli*, and no direct evidence was found that it is present in the JM109 strain.

Although it was not successful in this experiment, I remain confident that RAPDs could also detect any genomic changes among evolving populations of *E. coli*. Considering that JM109 is a direct descendent of K12, the differences observed between the two strains do show evidence of evolutionary change, although it is over a much greater time period and the two strains are known to have differences. The goal of this study was to investigate populations of the same strain that have been evolving for a relatively short time under different conditions, and see if genomic changes have occurred. Two possibilities exist to explain the result that no genomic differences between the evolved populations of K12 were detected in this experiment. One is that the evolved populations do have different genomes than the starting population, Gen.0, but these changes were simply not detectable with the two primers (OPA-04 and OPA-07) used in the analysis. More RAPD primers need to be used with DNA from the evolved populations to

determine if this is the case. The second possibility is that there may be no detectable genomic differences between the evolved populations and their starting populations at this time. Previous studies have examined evolved populations of *E. coli* at 10,000 generations (Papadopoulos et al., 1999; Schneider et al., 2000), and the samples used here had only passed approximately 2,000 generations. It may simply be that too few generations have passed to allow genomic changes to accumulate enough for RAPD analysis to be useful. If differences are not detected with the use of more primers, then perhaps the populations must evolve for a longer time before they can be analyzed with this methodology.

The RAPD methodology could be strengthened by using a greater number of primers or combinations of primers. Here, I used only 10 RAPD primers, and was able to successfully differentiate two *E. coli* strains. Using more primers or combinations of primers could allow investigators to produce a barcode for each strain that would reliably differentiate it from any other. The use of this molecular technique in conjunction with others like RFLP may prove to be a valuable and convenient way to differentiate between species or subspecies as well as to investigate evolutionary change among populations.

Acknowledgement – I would like to thank Dr. Berish Rubin for his guidance and advice. I would also like to thank Dr. Sylvia Anderson for her assistance in troubleshooting with Dr. Rubin and Bo Liu, and also for her help with choosing and obtaining the RAPD primers. Finally, I would like to thank Bo Liu and Leleesha Samaraweera for their constant assistance and support in the laboratory.

REFERENCES

Blattner, F.R., Plunkett, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Callado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B., and Shao, Y. (1997) *Science* **277**, 1432-44

Da Mata, T.L., Segeren, M.I., Fonseca, A.S., and Colombo, C.A. (2009) *Scientia Horticulturae* **121**. 92-96

Joliet-Gougeon, A., David-Jobert, S., Tamanai-Shacoori, Z., Ménard, C., and Cormier, M. (2000) *Appl. Environ. Microbiol.* **66**, 5484-5487

Kilic, A., Muz, A., Ertaş, H.B., and Özbey, G. (2009) *F.U. Sag.Bil. Vet. Derg.* **23**, 01-04

Kundu, J.K., Gadiou, S., and Cervena, G. (2009) *Virus Genes* **38**, 468-474

Lee, H.L., Cheon, D.S., and Choi, C. (2009) *Food Control* **20**, 623-626

Liti, G., Carter, D.M., Moses, A.M., Warringer, J., Parts, L., James, S.A., Davey, R.P., Roberts, I.N., Burt, A., Koufopanou, V., Tsai, I.J., Bergman, C.M., Bensasson, D., O'Kelly, M.J.T., and van Oudenaarden, A. (2009) *Nature* **458**, 337-342

Lyra, M.L., Klaczko, L.B., and Azeredo-Espin, A.M.L. (2009) *Med. Vet. Entemol.* **23**, 32-42

Maillet, I., Berndt, P., Malo, C., Rodriguez, S., Brunisholz, R.A., Pragai, Z., Arnold, S., Langen, H., and Wyss, M. (2007) *Proteomics* **7**, 1097-1106

Moran, N.A., McLaughlin, H.J., and Sorek, R. (2009) *Science* **323**, 379-382

Narzary, D., Mahar, K.S., Rana, T.S., and Ranade, S.A. (2009) *Scientia Horticulturae* **121**, 237-242.

Okhravi, N., Adamson, P., Matheson, M.M., Towler, H.M.A., and Lightman, S. (2000) *Invest. Opthalmol. Vis. Sci.* **41**, 1438-1447

Papadopoulos, D., Schneider, D., Meier-Eiss, J., Arber, W., Lenski, R.E., and Blot, M. (1999)

*Proc. Natl. Acad. Sci. USA***96**, 3807-3812

Salehi, T.Z., Madani, S.A., Karimi, V., and Khazaeli, F.A. (2008) *Braz. J. Microbiol.* **39**, 494-497

Schacherer, J., Shapiro, J.A., Ruderfer, D.M., and Kruglyak, L. (2009) *Nature* **458**, 342-347

Schneider, D., Duperchy, E., Coursange, E., Lenski, R.E., and Blot, M. (2000) *Genetics***156**, 477-488

Tsent, C., Ting, E., Johnson, D., Saluta, M., and Dunst, R. (2001) *Life Science News (Amersham Pharmacia Biotech)* **7**, 1-2

Zhou, P., Zhang, H., Lin, R.Q., Zhang, D.L., Song, H.Q., Su, C., and Zhu, X.Q. (2009) *Int. J. Parisitol.* **58**, 193-195