The complete *Avr*II restriction map of the *Escherichia coli* genome and comparisons of several laboratory strains

Donna L.Daniels
University of Wisconsin, Laboratory of Genetics, Madison, WI 53706, USA

Received January 30, 1990; Revised and Accepted March 30, 1990

ABSTRACT

The complete 13 site *Avr*II restriction map of the genome of *E. coli* strain MG1655 is presented and compared with several other *E. coli* strains. The map was determined primarily by isolating individual *Avr*II fragments from pulsed-field gels, and hybridizing these large probes to a battery of mapped *E. coli* clones in lambda vectors. *Avr*II restriction patterns for eight other laboratory strains were determined and maps for seven of them deduced from the gel and comparisons between the strain genotypes, the MG1655 map, and *Avr*II sites in *E. coli* sequences taken from Genbank.

INTRODUCTION

Restriction enzymes have been used for the physical dissection of the genomes of organisms ranging from bacteriophages to mammels. Maps (an enumeration of the locations of restriction sites on a DNA molecule) provide a physical description of the molecule and provide a basis for the comparison of related genomes. Maps also facilitate a variety of experiments requiring the isolation of a particular region of DNA such as cloning, sequencing, or probe isolation.

The literature contains hundreds of physical maps of subregions of the *E. coli* genome in the form of restriction maps of clones. Several more extensive regions have been mapped by merging overlapping maps (1−8). Maps of enzymes cutting frequently (average fragment size <20 kb) are called high resolution maps. A map is complete if all sites for that enzyme have been mapped on the specified region of DNA. Partial maps are those where an occasonal site or fragment has been located on the DNA. Low resolution maps on the other hand contain infrequent sites and fragment length measurement in general requires the use of pulsed-field gels. Such a map is fundamentally different from the assembled maps because the fragment size data is obtained directly on whole, unfractionated *E. coli* DNA. Smith *et al.* published a complete, 23 site *Not*I map of *E. coli* strain EMG2 (9) and partial information is available for *Avr*II, *Spe*I, *Sfi*I, and *Not*I on this and other strains. (7,8,10) This paper presents complete *Avr*II restriction maps for *E. coli* strains MG1655, two versions of W3110, P678, EMG2, C600,GM48, and W3350.

MATERIALS AND METHODS

E. coli MG1655, EM62, C60, P678, W3350,and W3110 (BB) were obtained from Barbara Backman at the Genetic Stock Center. GM48 and W3110 (FB) were from Frederick R. Blattner's collection, GM48 having been obtained before that from J. Dahlberg and W3110 from W. Szybalski.

DNA of intact *E. coli* chromosomes was prepared as described (11). For enzyme digestion one plug (100 μl) was rinsed with 1 ml of digestion buffer (10 mM TRIS-HCl, pH 7.4, 10 mM $MgCl_2$, 60mM NaCl) for 10 min. then covered with 200 μl of digestion buffer, 2 μl of *Avr*II enzyme (1 unit) added, and incubated at 37° C for 4−8 hours. Pulsed gel electrophoresis was performed in horizontal submerged slab gels. Field pulsing was controlled by a DNASTAR, Inc. P200 system. Two electrode arrangements were used, FIGE (12) and CHEF (13,14). After electrophoresis DNA bands were visualized with EtBr and UV light. A gel slice was cut out with a scaple, DNA electroeluded and labeled by nick translation. Phage stocks were diluted to a titer of 5×10^3 to 1×10^5/ml and transferred to a 96 well plate. Seven microliters of each were spotted on a bacterial lawn with a 96 well replicator. Plates were incubated overnight

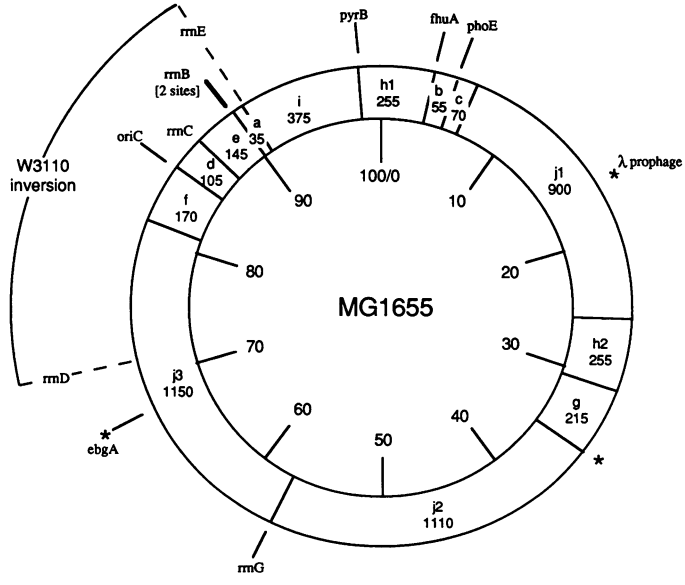

Figure 1: *Avr*II restriction map of *E. coli* MG1655. Numbers inside the circle indicate minutes on the *E. coli* genetic map. Lines between the two concentric circles represent AvrII sites. Fragments are labeled with their name and measured length in Kb. Lines to the outside of the circle indicate *Avr*II sies found by a scan of GenBank™. Asterisk (*) indicates sites which are variable among the strains studied here. Locations of several genes are indicated as is the section of the chromosome inverted in strain W3110.

at 37°C to allow spots to develop. Phage DNA was transferred to nitrocellulose and hybridized to labeled probes as described (15). 600 lambda clones were screened on a single filter.

GenBank™ searches were done using a program by DNASTAR, Inc., Madison, WI.

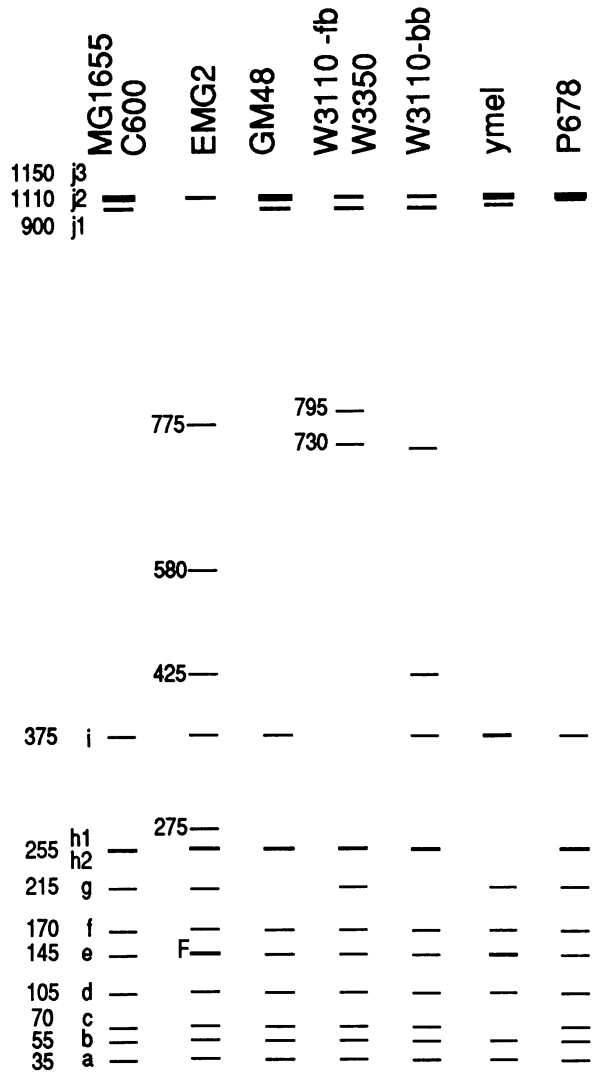

Figure 2: *Avr*II restriction patterns for several *E. coli* lab strains. The similarities and differences in sizes of *Avr*II digestion products for nine *E. coli* strains are compared graphically. Fragment names and lengths in kb are indicated.

RESULTS

E. coli MG1655 DNA was digested with a battery of restriction enzymes to identify enzymes which cut rarely. The recognitizion sequence of *Avr*II (CCTAGG) has as its center the rarest tetranucleotide in *E. coli* DNA, and *Avr*II was found to cut only 13 times.

The 13 site *Avr*II map of MG1655 DNA is shown in Figure 1. *Avr*II digested *E. coli* MG1655 DNA was electrophoresed under a variety of pulsing conditions and electrode arrangements. Bands on the gels were named A through J (smallest to largest) and fragment sizes were measured by comparison to 'λ-ladder' multimers as a size standard. Band names and fragment size are shown in Figures 1 and 2. Measurements were very reproducible from gel to gel and under various pulse regimes. Measurement of fragments A to I probably have less than 20 kb error and J_1, J_2, J_3, (at the extreme end of lambda ladder) are likely accurate within 150 kb. The sum of the 13 fragment lengths is 4835 kb within 3% of the genome size estimated by Kohara *et al.* (6). Fragments smaller than 20kb may have not been visible on gels.

Fragments were located relative to the genetic map by hybridization to a large (~600) set of *E. coli* clones in lambda vectors, nearly 200 of which had been previously located to the genetic map by comparison of the restriction map to restriction maps of published clones as described (7). This set of mapped clones covered 75% of the genetic map and the only region with more than a 5 min. space between mapped clones was around 40 min. When a clone whose genetic map location was known showed positive hybridization with an *Avr*II fragment that fragment could be assigned to the genetic map. Conversly, when clones whose map locations were not known hybridized to a located *Avr*II fragment they were then identified as mapping to that region of *E. coli* DNA. The ten *Avr*II fragments A through I (band H was a double) were located in this manner.

J_1, J_2, and J_3 could not be located in this manner because they were so large that the hybridizations showed too much background to be useful. This is because the spots of lysis of the lambda phage also contain *E. coli* host DNA which is transferred to the filter along with the lambda clone DNA. Since the λ clone DNA which hybridizes is at most 20kb long while the host DNA hybridizes over the length of the probe the signal to noise ratio gets worse as probes get larger. Fragments <400 kb were usable as probes.

After assigning locations to all the smaller fragments it was clear that one of the large fragments mapped to the 8 to 29′ region and two to the 35 to 82 minute region. J_1 was located to the 8

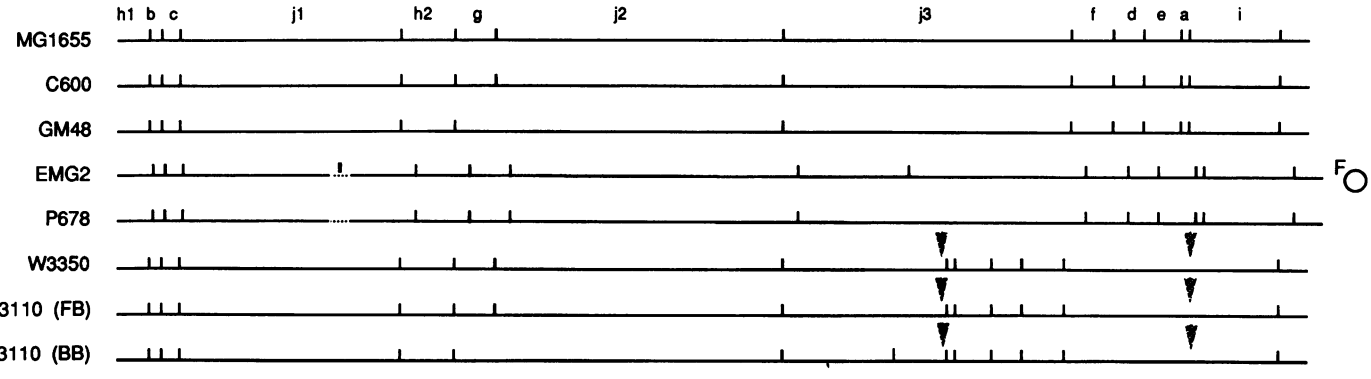

Figure 3: Derived maps of seven *E. coli* strains are aligned linearly for comparison. Vertical bars indicate *Avr*II sites. Hatched triangles show the location of the inversion point in W3110 and W3350.

to 29′ region because it was changed in size in a lambda lysogen (att λ is at 18′). J_3 was located to the 58 to 82′ region because it was not present in W3110 which has in inversion endpoint at 72′. By elimination J_2 was assigned to the 35 to 58′ region.

This MG1655 map was further verified by scanning GenBank™ for *E. coli* sequences with *Avr*II sites. Lines on the outside of the circle in Figure 1 show the results of the most recent search (GenBank™ release #61, Sept. 1989). Sites found in oriC, rrnB, pyrB, fhuA, phoE and rrnG correspond to mapped sites. Two other mapped *Avr*II sies are probably located in rrnC and rrnE for which sequence is not available. Of the other three rrn operons, rrnD and rrnA do not have *Avr*II sites and rrnH is located so close to fhuA that a fragment between them may have been too small to detect on the gels so we can not be certain that rrnH does not contain an *Avr*II site. The site in GenBank™ in ebgA is not present on the MG1655 map. Finally, the site between J_2 and J_3, which had not been precisely located on the MG1655 map by hybridization but which was presumed to be in rrnG, was accurately located by digesting our λ MG1655 clone containing rrnG with *Avr*II and verifying that the DNA was cut.

The digestion patterns for several other strains are shown in Figure 2. This information together with the known strain genotypes (16,17) and information in Figure 1 was used to deduce the maps of seven of the other strains. This is presented in Figure 3. There are several differences between the strains, including a surprizing difference between W3110 from two different sources. 1) Presence or absense of the F factor which is not cut by *Avr*II and runs in our gels at about 145 Kb. 2) Presence or absence of the lambda prophage (48.5 Kb). Wild-type lambda is cut by *Avr*II and the sequence predicts two cutsites 100 bp apart from each other (18). The prophage in P678 however appears to not be cut by *Avr*II while the prophage in EMG2 is cut. 3) Presense or absence of the cut-sites at 34 min and/or in ebgA at 68 min. 4) Inversion between rrnD at 72 min and rrnE at 91 min. One of the published IS5 sequences (19) has an *Avr*II cut site at the IS5/coli junction but neither the polymorphism at 34 min. nor at 68min. seems due to an IS5 difference (20).

DISCUSSION

The strategy used here to construct a complete *Avr*II restriction map is applicable to organisms where a significant number of mapped clones is available and to enzymes which cut the DNA into a few bands which are resolvable on gels. Conceptually it is similar to the strategy used by Smith *et al.*(9). for the determination of the *Not*I map, however, logistically it is much simplier to hybridize a few large fragment probes isolated from one or a few gels to a very large number of lambda clones than to prepare and label DNA from a large number of mapped clones each to be used as a probe against a blot of a gel. This simplification allows the screening of many more mapped clones and thus more accurate maps.

The *Avr*II maps of the eight *E. coli* strains presented here demonstrate that significant map differences do occur between *E. coli* lab strains. Cognizance of this fact is important when using restriction maps in the planning of experiments such as strategies for cloning, sequencing, or probe isolation.

ACKNOWLEDGEMENTS

This is paper No. 3095 of the Laboratory of Genetics, funded by NIH grant GM 35682 to DLD. I thank Guy Plunkett for help with the manuscript.

REFERENCES

1. von Meyenburg, K., and Hansen, F. G. 1980. in: *Mechanistic Studies of DNA Replication in Genetic Recombination*. Academic Press, Inc.
2. Hadley, R. D., Hu, M., Timmons, M., Yun, K. and Deonief, R. C. 1983. *Gene*, **22(2-3)**, 281-7.
3. Bouché, J.-P. 1982. *J. Mol. Biol.* **154**, 1-20.
4. Knott, V., Rees, D. J., Cheng, Z. and G. G. Brownlee. 1988. *Nucleic Acids Res.*, **25**, 2001-2612.
5. Knott, V., Blake, D. J., and G. G. Brownley. 1989. *Nucleic Acids Res.*, **17**, 5901-5912.
6. Kohara, Y., Akiyama, K., and Isono, K. 1987. *Cell*, **50**, 495-505.
7. Daniels, D. L., and Blattner, F. R. 1987. *Nature*, **325**, 831.
8. Daniels. D. L. 1990. in: *The Bacterial Chromosome*, K. Drlica and M. Riley, eds. American Society of Microbiology, Washington, D. C. pp 43-51.
9. Smith, C.L., Econome,J. G., Schutt, A., Klco, S., and Cantor, C. R. 1987. *Science*, **236**, 1448-1453.
10. Condemine, G. and Smith, C. L. 1990. in: *The Bacterial Chromosome*, K. Drlica and M. Riley, eds. American Society of Microbiology, Washington, D.C., pp 53-60.
11. Smith, C. L., Warburton, P., Gaal, A., and Cantor, C. R. 1986. in: *Genetic Engineering*, J. K. Setlow and A. Hollaender,eds. Plenum Press, NY, NY. **8**, 45-70.
12. Carle, G. F., Frank, M., and Olson, M. V. 1986. *Science*, **232**, 65-68.
13. Chu, G., Vollrath, D., and Davis, R. W. 1986. *Science*, **234**, 1582.
14. Birren, B. W., Lai, E., Clark, S. M., Hood. L., and Simon, M. I. 1988. *Nucleic Acids Research*, **16**, 7563-7582.
15. Benton, W. D., and Davis, R. W. 1977. *Science*, **196**, 180.
16. Bachmann, B. 1987. in: *Escherichia coli* and *salmonella typhimurium: cellular and molecular biology*. F. C. Neidhardt, editor-in-chief. American Society of Microbiology, pp 803-876.
17. Bachmann, B. 1987. in: *Escherichia coli* and *salmonella typhimurium: cellular and molecular biology*. F. C. Neidhardt, editor-in-chief. American Society of Microbiology, pp 1190-1219.
18. Daniels, D. L., Schroeder, J. L., Szybalski, W., Sanger, F., Coulson, A. R., Hong, G.-F., Hill, D. F., Petersen, G. B., and Blattner, F. R. 1983. Appendix II, in: *Lambda II*, Weisberg, Roberts, Stahl and Hendrix, eds. Cold Spring Harbor Lab. Press. pp 519-676.
19. Schoner, Brigitte and Kahan, Marc. 1981. The nucleotide sequence of IS5 from *Escherichia coli. Gene*, **14**, 165-174.
20. Maramatsu, Shuji, Kato, Masashi, Kohara, Yuji, and Mizuno, Takeshi. 1988. Insertion sequence IS5 contains a sharply curved DNA structure at its terminus. *Mol. Gen. Genet.* **214**, 433-438.