JOURNAL OF BACTERIOLOGY, June 1988, p. 2467–2471
0021-9193/88/062467-05$02.00/0
Copyright © 1988, American Society for Microbiology
Vol. 170, No. 6

Cloning and Nucleotide Sequence of the *aroA* Gene of *Bordetella pertussis*

DUNCAN J. MASKELL,* PIERRE MORRISSEY,† AND GORDON DOUGAN

Department of Molecular Biology, Wellcome Research Laboratories, Langley Park, South Eden Park Road, Beckenham, Kent BR3 3BS, United Kingdom

Received 1 December 1987/Accepted 29 February 1988

The *aroA* locus of *Bordetella pertussis*, encoding 5-enolpyruvylshikimate 3-phosphate synthase, has been cloned into *Escherichia coli* by using a cosmid vector. The gene is expressed in *E. coli* and complemented an *E. coli aroA* mutant. The nucleotide sequence of the *B. pertussis aroA* gene was determined and contains an open reading frame encoding 442 amino acids, with a calculated molecular weight for 5-enolpyruvylshikimate 3-phosphate synthase of 46,688. The amino acid sequence derived from the nucleotide sequence shows homology with the published amino acid sequences of *aroA* gene products of other microorganisms.

Bordetella pertussis, the agent of whooping cough in humans, is a fastidious organism, and its growth is extremely sensitive to inhibitory substances present in normal laboratory media (20). Nevertheless, *B. pertussis* can be grown on relatively simple defined media, a factor which would assist the genetic characterization of the metabolic activities of the organism. By using chemical mutagenesis and DNA transfer techniques, Kloos et al. (15) isolated and characterized a number of auxotrophic mutants of *B. pertussis*. More recently it has been shown that *B. pertussis* DNA cloned in *Escherichia coli* can complement some *E. coli* auxotrophic mutations (3). In the intensive search for a new acellular pertussis vaccine, much recent work on *B. pertussis* genes has concentrated on virulence determinants. Many of the virulence factors of *B. pertussis* are subject to regulation by the so-called *vir* locus, which appears to be a positive regulator controlling gene expression (27). Little work has been carried out on *B. pertussis* genes not under *vir* regulation, and the study of such genes may yield important information about the genetic organization of *B. pertussis*. A useful gene to study in this context is *aroA*.

The *aroA* gene of *E. coli* encodes the enzyme 5-enolpyruvylshikimate 3-phosphate synthase (EPSP synthase), which acts in the biosynthetic pathway leading to chorismate. This pathway is the only route in bacterial, fungal, and plant species for the biosynthesis of aromatic compounds including *p*-aminobenzoic acid, 2,3-dihydroxybenzoate, and aromatic amino acids. The *aroA* loci of *E. coli* (10), *Salmonella typhimurium* (25), *Aspergillus nidulans* (5), and *Saccharomyces cerevisiae* (9) have been sequenced and are highly conserved. In this manuscript we report the cloning and nucleotide sequence analysis of the *aroA* locus from *B. pertussis* and show that it, too, is homologous with the other *aroA* genes sequenced to date.

MATERIALS AND METHODS

Bacterial strains, plasmids, and bacteriophages. *B. pertussis* CN2992 was provided by P. Novotny (19). *E. coli* AB2829, which harbors a stable *aroA* mutation (12), was obtained from the *E. coli* Genetic Stock Center, Yale University School of Medicine, New Haven, Conn. Cosmid pHC79 (11) and plasmids pACYC184 (4) and pUC18 (Pharmacia, Milton Keynes, England) were used as cloning vectors. For sequencing, DNA fragments were cloned into M13mp18 or M13mp19 (Pharmacia). pHC79 carries ampicillin resistance (Ap^r) and tetracycline resistance (Tc^r). Tc^r can be insertionally inactivated by cloning into the *Bam*HI site. pACYC184 carries Tc^r and chloramphenicol resistance (Cm^r). Cm^r can be insertionally inactivated by cloning into the *Eco*RI site. pUC18 is Ap^r.

Media and reagents. *B. pertussis* was grown in Stainer-Scholte liquid medium at 37°C (24). *E. coli* strains were grown in Luria broth (LB) (18) or on LB – 1.6% (wt/vol) agar, the constituents for which were obtained from Difco Laboratories, Detroit, Mich. Minimal medium (MM) was as described previously (7). MM was solidified with 2% (wt/vol) Noble agar (Difco). Aromatic amino acids, *p*-aminobenzoic acid, and dihydroxybenzoate supplement (aro mix) were as described previously (7) and were obtained from Sigma Chemical Co., Poole, England. Antibiotics were included at appropriate concentrations.

Restriction endonucleases and T4 DNA ligase were obtained from Boehringer, Lewes, England, or GIBCO BRL, Paisley, Scotland, and were used as specified by the manufacturer. DNA polymerase I large fragment (Klenow enzyme) was FPLC pure from Pharmacia.

***B. pertussis* high-molecular-weight-DNA production.** Chromosomal DNA from *B. pertussis* was isolated by using a modification of the method of Hull et al. (13). A 200-ml culture of *B. pertussis* CN2992 was split into 25-ml aliquots and centrifuged at 10,000 rpm for 5 min in an SS-34 rotor (Ivan Sorvall, Inc., Norwalk, Conn.). The cell pellets were washed twice with 25% sucrose in 1 mM EDTA–10 mM Tris (pH 7.5). The washed pellets were each suspended in 2 ml of 25% sucrose in 1 mM EDTA–50 mM Tris (pH 8.0) at 4°C, 50 μl of 40-mg/ml lysozyme in 0.25 M Tris (pH 8.0) was added, and the preparation was left on ice for 5 min. Then 50 μl of 1-mg/ml proteinase K (Boehringer) was added and mixed well, and 0.4 ml of 0.5M EDTA (pH 8.0) was mixed in, followed by 250 μl of 10% (wt/vol) sodium *N*-lauroyl sarcosinate. This mixture was left on ice until lysis was complete (approximately 2 h) and then was kept at 50°C overnight. The lysed cells were extracted three times with equal volumes of 1:1 (vol/vol) phenol-chloroform and then once with chloroform alone. Then 1/10 volume of 3 M sodium acetate was added to the aqueous layer followed by 2 volumes of

* Corresponding author.
† Present address: Department of Medical Microbiology, Stanford University, Stanford, CA 94305.

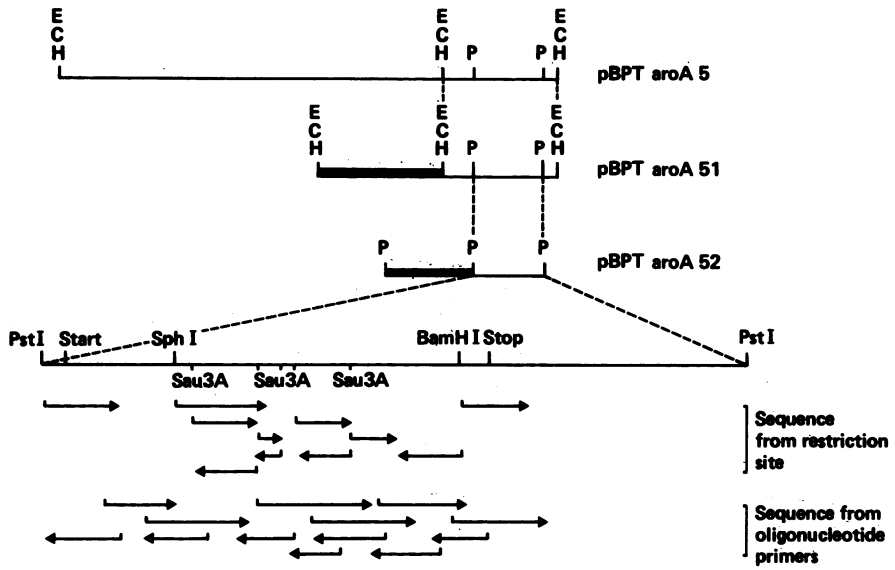

FIG. 1. Strategies used to clone and sequence the *B. pertussis aroA* gene. Abbreviations: E, *Eco*RI site; C, *Cla*I site; H, *Hin*dIII site; P, *Pst*I site. Dark blocks are known vector sequences. The vector/target join in pBPTaroA5 is unclear from restriction digests and has not been determined by DNA sequencing. Thus, no dark block is shown for pBPTaroA5. "Start" refers to the ATG at position 1 of Fig. 2, and "Stop" likewise refers to the TGA at position 1327.

ethanol. The precipitated DNA was wound off, dissolved in 1 mM EDTA–10 mM Tris hydrochloride (pH 8.0), and dialyzed against the same buffer.

Cloning of *B. pertussis* CN 2992 DNA in cosmid pHC79. Cosmid vector pHC79 (11) was digested to completion with *Bam*HI. High-molecular-weight *B. pertussis* DNA was partially digested with *Sau*3AI to give DNA in the size range 40 to 50 kilobases (kb). Vector and target DNA samples were combined at appropriate concentrations and ligated with T4 DNA ligase. Ligation mixtures were packaged in vitro by using an in vitro packaging system from Gigapack (Stratagene, Northumbria Biologicals Ltd., Cramlington, England). *E. coli* strains, grown in LB–0.4% maltose, were infected with the packaged DNA and plated out on LB agar–50 μg of ampicillin per ml. Ap^r colonies were replica plated onto LB agar containing ampicillin (50 μg/ml) or tetracycline (20 μg/ml). Ap^r Tc^s colonies were picked into microdilution plates containing (per well) LB–50% glycerol–50 μg of ampicillin per ml, grown at 37°C, and then stored frozen at −70°C. Large-scale plasmid isolation was carried out by the method described by Maniatis et al. (18), and rapid plasmid screening was done by the method of Birnboim and Doly (2).

Gel purification of restriction fragments. Restriction fragments were purified from low-melting-temperature agarose by hot phenol extraction as described previously (18).

DNA sequencing. DNA was sequenced by the dideoxynucleotide chain termination method with [^{35}S]dATP (Amersham International, Little Chalfont, England) and wedge gels (1, 22). For sequencing through severe compressions which arose in both directions at some points on the DNA, 20 to 40% (vol/vol) deionized formamide was added to the gel.

RESULTS AND DISCUSSION

Cloning of the *B. pertussis aroA* gene in *E. coli*. The initial cosmid bank ligations were packaged in vitro, and the resultant phage were used to infect *E. coli* AB2829 and *E. coli* HB101. When the packaged bank was used to infect *E. coli* HB101, approximately 10^5 Ap^r colonies were obtained, of which 600 were replica plated onto tetracycline. Of these, the Tc^s colonies made up 56%. This was not done for *E. coli* AB2829. Infected *E. coli* AB2829 cells were plated directly onto MM containing 50 μg of ampicillin per ml and incubated at 37°C for 3 days. Uninfected bacteria were also plated on MM and MM plus ampicillin as negative controls or on MM with aro mix as positive controls. No colonies were seen on the negative control plate, whereas bacteria grew well on the positive control plates. On the selection plate five colonies grew and were picked and streaked out onto MM plus ampicillin. Colony no. 5 grew well on solid and in liquid MM plus ampicillin. Recombinant cosmid DNA was isolated from this clone and was named pBPTaroA5. This plasmid was only 15.5 kb in size and may have resulted from spontaneous deletion of a larger cosmid. This putative larger cosmid may have had more than one copy of the pHC79 vector, since pBPT aroA5 has the *Eco*RI, *Hin*dIII, and *Cla*I cluster of sites, present in pHC79, represented twice. Since it is unclear what is vector and what is target sequence in pBPTaroA5, no vector sequence has been marked in Fig. 1. To subclone the *aroA* gene, we digested pBPTaroA5 with *Eco*RI, yielding two fragments of 12 and 3.5 kb. pACYC184 was digested with *Eco*RI and ligated with the fragments from pBPTaroA5. The ligation mix was transformed into *E. coli* AB2829, and transformants were selected on MM-tetracycline. A Tc^r Cm^s colony was selected. The isolate harbored a recombinant 7.5-kb plasmid which was named pBPTaroA51. A 2.2-kb *Pst*I fragment was further subcloned into pUC18 from pBPTaroA51 by complementation of *E. coli* AB2829, the resulting plasmid being named pBPTaroA52 (Fig. 1). A gel-purified *Pst*I fragment from pBPTaroA52 was also cloned into M13 for sequencing. The identity of this fragment with *B. pertussis* DNA was confirmed by hybridization (data not shown).

DNA sequencing. The DNA sequencing strategy is shown in Fig. 1. In addition to the *Pst*I fragment, *Sph*I, *Bam*HI-*Pst*I and *Sau*3AI fragments were cloned into M13mp18 and/or

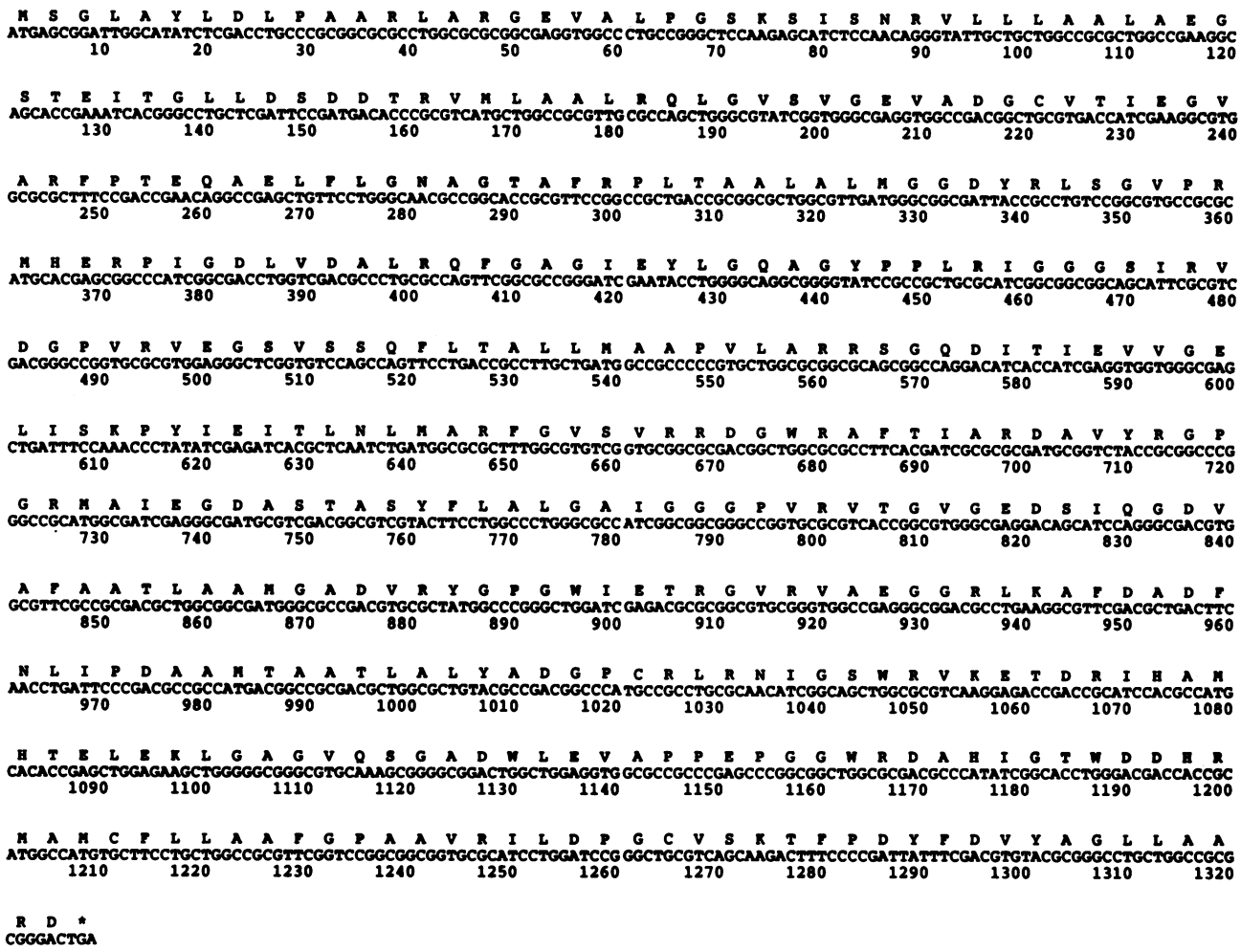

FIG. 2. Nucleotide and derived amino acid sequence of the *aroA* gene of *B. pertussis*.

M13mp19 and sequenced. Gaps in the sequence were filled in with synthetic oligonucleotides to prime sequencing reactions from CsCl-purified single-stranded recombinant M13 DNA (26). The nucleotide sequence is shown in Fig. 2. On translation in all six frames, one open reading frame was encountered, which showed similarity with the published sequence of the *E. coli aroA* gene. The *B. pertussis aroA* gene sequence consists of 1,329 nucleotides, including the TGA stop codon, encoding 442 amino acids. The deduced molecular weight of the *B. pertussis* enzyme is 46,688.

The nucleotides from the *PstI* site to the ATG start codon were sequenced and number 59 base pairs. The nucleotides 5′ to the ATG start codon have the sequence 5′-AATTGGCGCAAGGATTCCAA-3′. The underlined bases, extending from positions −10 to −7, are in good agreement with the consensus Shine-Dalgarno sequence (23). No promoterlike sequences have been identified 5′ to the coding region, although consensus promoters have not yet been compiled for *B. pertussis* housekeeping genes. We are currently investigating whether *B. pertussis aroA* is part of an operon with *serC*, as has been described for *E. coli* (8) and *Salmonella* strains (12). No sequence similar to the rho-independent terminator seen between *serC* and *aroA* in *E. coli* (8) was observed in this sequence.

Codon usage in *B. pertussis aroA*. The codon usage in *B. pertussis aroA* is shown in Table 1. There is a clear bias in codon usage, with strong preference for G or C in the third position. For Leu, Ser, Cys, and Arg, no codons with A or T in the third position are used. For the other amino acids, codons with A or T in the third position occur rarely.

In broad agreement with published estimates for the G+C content of *B. pertussis* DNA (14), the *aroA* gene coding sequence contains 71.3% G+C, with 13.3% A, 33.6% C, 37.7% G, and 15.3% T. This is higher than the content given for pertussis toxin, 62.2% (17), and that derived from the sequence of a *B. pertussis* serotype 2 fimbria, 61.55% G+C (16), and may reflect different evolutionary origins between normal housekeeping genes and virulence factors in *B. pertussis*.

Comparison of *B. pertussis aroA* DNA and protein sequences with other *aroA* sequences. It has been observed that the EPSP synthase amino acid sequences show marked similarity between genera (5, 6). There is considerable similarity between the *B. pertussis* EPSP synthase amino acid sequence and that in *E. coli* (approximately 57%) (Fig. 3). Considerable homology is also evident with the EPSP synthase amino acid sequences in the other microorganisms shown in Fig. 3. The *B. pertussis* sequence is most similar to the other two bacterial sequences shown, but there are positions where the *B. pertussis* sequence diverges from the

TABLE 1. *B. pertussis aroA* gene codon usage

Amino acid	Codon	No. of times used[a]	Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used	Amino acid	Codon	No. of times used
F	TTT	2	S	TCT	0	Y	TAT	5	C	TGT	0
F	TTC	13	S	TCC	6	Y	TAC	6	C	TGC	4
L	TTA	0	S	TCA	0	*	TAA	0	*	TGA	1
L	TTG	5	S	TCG	5	*	TAG	0	W	TGG	6
L	CTT	0	P	CCT	0	H	CAT	1	R	CGT	0
L	CTC	3	P	CCC	8	H	CAC	4	R	CGC	29
L	CTA	0	P	CCA	1	Q	CAA	1	R	CGA	0
L	CTG	39	P	CCG	13	Q	CAG	7	R	CGG	6
I	ATT	3	T	ACT	1	N	AAT	1	S	AGT	0
I	ATC	19	T	ACC	12	N	AAC	4	S	AGC	10
I	ATA	0	T	ACA	0	K	AAA	1	R	AGA	0
M	ATG	12	T	ACG	8	K	AAG	5	R	AGG	1
V	GTT	0	A	GCT	1	D	GAT	7	G	GGT	1
V	GTC	7	A	GCC	27	D	GAC	22	G	GGC	45
V	GTA	2	A	GCA	1	E	GAA	5	G	GGA	2
V	GTG	24	A	GCG	33	E	GAG	17	G	GGG	7

[a] Number of times the codon is used to code for the amino acid.

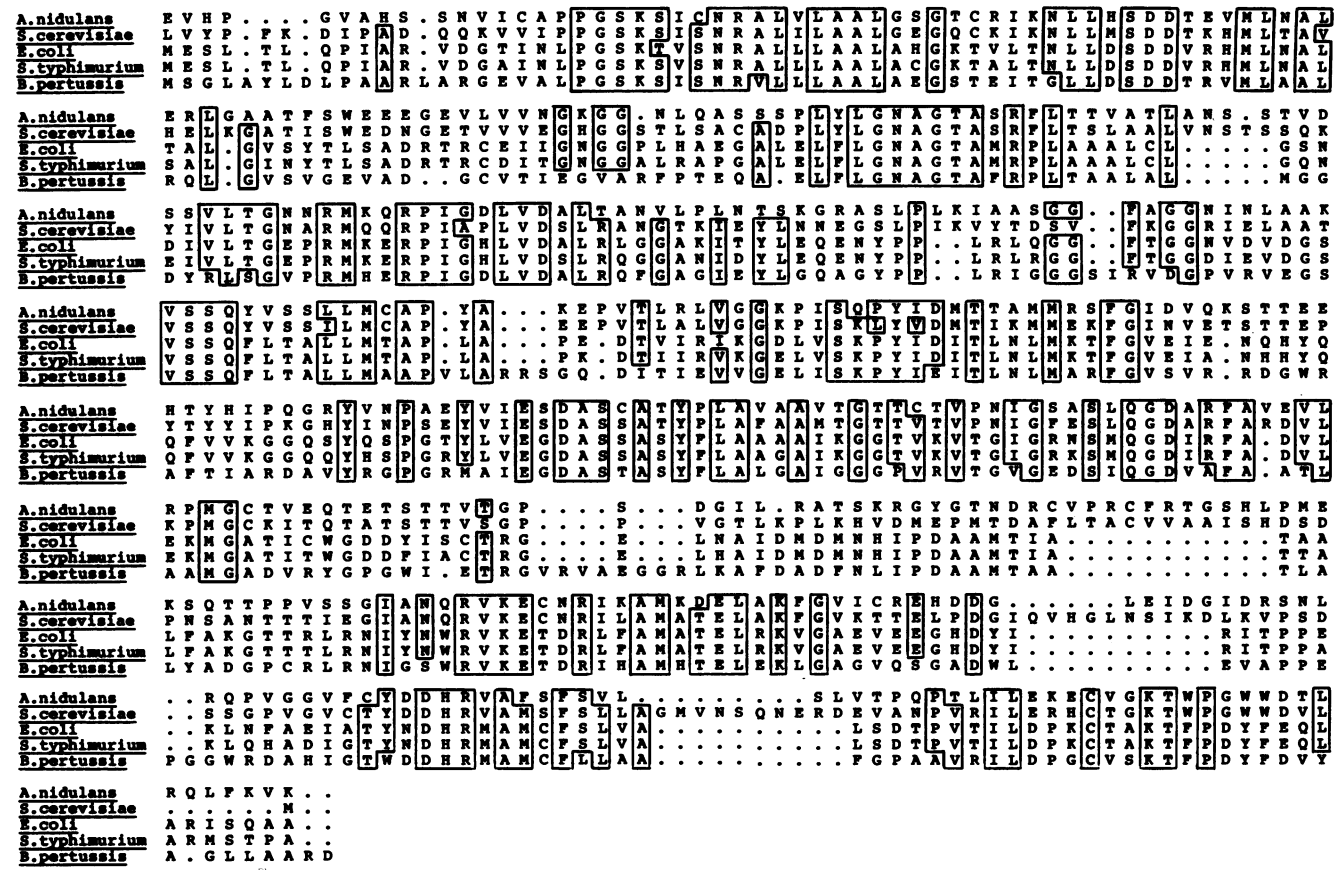

FIG. 3. Amino acid homologies between the *B. pertussis aroA* gene product and the other published *aroA* gene products. Boxes correspond to areas where the amino acids are identical in four or more of the species shown. Dots refer to regions of variable length between the sequences shown and are assigned according to the best fit for the region of homology.

bacterial sequences and is homologous with the fungal sequences.

Strikingly, there are some blocks of homology that are identical over a number of amino acids in all five sequences shown in Fig. 3, and there are also large areas where conservative amino acid substitutions have occurred. These observations suggest that there may have been evolutionary conservation of areas of the molecule essential to its function.

Conclusions. This report shows that the *aroA* gene of *B. pertussis* is highly homologous to those of other microorganisms and that its product can complement *aroA* mutations in *E. coli*. DNA sequencing demonstrates that the gene has a high G+C content, higher than that of two virulence factors recently sequenced, and a consensus Shine-Dalgarno sequence. No information is available on the promoter for this gene, and we are investigating whether this is due to its being in an operon with *serC*, as has been shown for other bacteria (8, 12).

ACKNOWLEDGMENTS

We thank Ian Charles and Neil Fairweather for helpful advice and review of the manuscript, Hugh Spence for making the oligos, Steve Nichol for running the sequence analysis programs, and Tina Silva for preparing the manuscript.

LITERATURE CITED

1. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
2. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
3. **Brownlie, R. M., J. G. Coote, and R. Parton.** 1986. Complementation of mutations in *Escherichia coli* and *Bordetella pertussis* by *B. pertussis* DNA cloned in a broad-host-range cosmid vector. J. Gen. Microbiol. **132**:3221–3229.
4. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from P15A cryptic miniplasmid. J. Bacteriol. **134**:1141–1156.
5. **Charles, I. G., J. W. Keyte, W. J. Brammar, M. M. Smith, and A. R. Hawkins.** 1986. The isolation and nucleotide sequence of the complex AROM locus of *Aspergillus nidulans*. Nucleic Acids Res. **14**:2201–2213.
6. **Coggins, J. R., K. Duncan, I. A. Anton, M. R. Boocock, S. Chaudhuri, J. M. Lambert, A. Lewendon, G. Millar, D. M. Mousdale, and D. D. S. Smith.** 1987. The anatomy of a multifunctional enzyme. Biochem. Soc. Trans. **15**:754–759.
7. **Dougan, G., D. Maskell, D. Pickard, and C. Hormaeche.** 1987. Isolation of stable *aroA* mutants of *Salmonella typhi* Ty2: properties and preliminary characterisation in mice. Mol. Gen. Genet. **207**:402–405.
8. **Duncan, K., and J. R. Coggins.** 1986. The *serC-aroA* operon of *Escherichia coli*. Biochem. J. **234**:49–57.
9. **Duncan, K., R. M. Edwards, and J. R. Coggins.** 1987. The pentafunctional *arom* enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. Biochem. J. **246**:375–386.
10. **Duncan, K., A. Lewendon, and J. R. Coggins.** 1984. The complete amino acid sequence of *Escherichia coli* 5-enolpyruvylshikimate 3-phosphate synthase. FEBS Lett. **170**:59–63.
11. **Hohn, B., and J. Collins.** 1980. A small cosmid for efficient cloning of large DNA fragments. Gene **11**:291–298.
12. **Hoiseth, S. K., and B. A. D. Stocker.** 1985. Genes *aroA* and *serC* of *Salmonella typhimurium* constitute an operon. J. Bacteriol. **163**:355–361.
13. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33**:933–938.
14. **Kersters, K., K.-H. Hinz, A. Hertle, P. Segers, A. Lievens, O. Siegmann, and J. DeLey.** 1984. *Bordetella avium* sp. nov. isolated from the respiratory tracts of turkeys and other birds. Int. J. Syst. Bacteriol. **34**:56–70.
15. **Kloos, W. E., W. J. Dobrogosz, J. W. Ezzell, B. R. Kimbro, and C. R. Manclark.** 1979. DNA-DNA hybridization, plasmids, and genetic exchange in the genus *Bordetella*, p. 70–80. *In* C. R. Manclark, and J. C. Hill (ed.), International Symposium on Pertussis. DHEW publication no. (NIH)79-1830. U.S. Department of Health, Education, and Welfare. National Institutes of Health, Bethesda, Md.
16. **Livey, I., C. J. Duggleby, and A. Robinson.** 1987. Cloning and nucleotide sequence analysis of the serotype 2 fimbrial subunit gene of *Bordetella pertussis*. Mol. Microbiol. **1**:203–209.
17. **Locht, C., and J. M. Keith.** 1986. Pertusis toxin gene: nucleotide sequence and genetic organization. Science **232**:1258–1264.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Novotny, P., and K. Cownley.** 1979. Effect of growth conditions on the composition and stability of the outer membrane of *Bordetella pertussis*, p. 99–123. *In* C. R. Manclark and J. C. Hill (ed.), International Symposium on Pertussis. DHEW publication no. (NIH)79-1830. U.S. Department of Health, Education, and Welfare. National Institutes of Health, Bethesda, Md.
20. **Parker, C.** 1976. Role of the genetics and physiology of *Bordetella pertussis* in the production of vaccine and the study of host-parasite relationships in pertussis. Adv. Appl. Microbiol. **20**:27–42.
21. **Pittard, J., and B. J. Wallace.** 1966. Distribution and function of genes concerned with aromatic biosynthesis in *Escherichia coli*. J. Bacteriol. **91**:1494–1508.
22. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
23. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
24. **Stainer, D. W., and M. J. Scholte.** 1970. A simple chemically defined medium for the production of phase I *Bordetella pertussis*. J. Gen. Microbiol. **63**:211–220.
25. **Stalker, D. M., W. F. Hiatt, and L. Comai.** 1985. A single amino acid substitution in the enzyme 5-enolpyruvylshikimate-3-phosphate synthase confers resistance to the herbicide glyphosate. J. Biol. Chem. **260**:4724–4728.
26. **Strauss, E. C., J. A. Kobori, G. Siu, and L. E. Hood.** 1986. Specific-primer-directed DNA sequencing. Anal. Biochem. **154**:353–360.
27. **Weiss, A. A., and S. Falkow.** 1984. Genetic analysis of phase change in *Bordetella pertussis*. Infect. Immun. **43**:263–269.