Cloning and DNA sequence analysis of the *serC*–*aroA* operon from *Salmonella gallinarum*; evolutionary relationships between the prokaryotic and eukaryotic *aroA*-encoded enzymes

HUGH G. GRIFFIN* and ANNETTE M. GRIFFIN

Division of Molecular Biology, AFRC Institute for Animal Health, Houghton Laboratory, Houghton, Huntingdon, Cambridgeshire PE17 2DA, UK

(Received 17 July 1990; revised 20 September 1990; accepted 28 September 1990)

The *serC*–*aroA* operon of *Salmonella gallinarum* was isolated from a gene library using a labelled oligonucleotide probe and by complementation of an *aroA* *Escherichia coli* strain. The nucleotide sequence of a 2.6 kbp fragment was determined. The predicted amino acid sequence of the *aroA* gene product was compared to the equivalent sequence from ten other organisms. Computer-generated evolutionary trees clearly divide the eleven sequences into four different groups: Gram-negative bacteria, Gram-positive bacteria, fungi and plants. These trees depict a close evolutionary relationship between the sequences from Gram-negative bacteria and higher plants.

Introduction

Fowl typhoid, the disease caused by *Salmonella gallinarum*, remains a significant worldwide problem, especially in those countries which are beginning to intensify their poultry industry (Pomeroy, 1984; Silva, 1985). Because of its economic importance and due to the emergence of antibiotic-resistant strains of *S. gallinarum* (Stuart *et al.*, 1962, 1967; Smith *et al.*, 1981) control of the disease by vaccination is desirable. Killed vaccines are of little practical value and an attenuated rough strain, although affording protection, has been shown to possess virulence for some breeds of chickens (Smith, 1956; Silva *et al.*, 1981).

Recent work with *S. typhi*, *S. typhimurium* and *S. dublin* indicates that strains of *Salmonella* harbouring stable mutations in the *serC*–*aroA* operon are attenuated and make excellent vaccines in mice and other animals (Hoiseth & Stocker, 1981; Stocker *et al.*, 1983; Smith *et al.*, 1984; Stocker, 1988). Attenuation probably results from a block in the aromatic biosynthetic pathway causing a requirement for *p*-aminobenzoic acid and dihydroxybenzoic acid, neither of which is readily

available in vertebrate tissue. An *aroA* strain of *S. gallinarum* may prove to be an ideal live vaccine against fowl typhoid and could also be used as a vehicle to deliver heterologous antigens to the chicken immune system.

As well as being of interest as a site of bacterial attenuation, the *serC*–*aroA* operon suggests the existence of an interesting regulatory link between the common aromatic biosynthetic pathway, the serine biosynthetic pathway, and the pyridoxine biosynthetic pathway (Hoiseth & Stocker, 1985). Much recent interest has been shown in the *aroA* gene product, 5-enolpyruvylshikimate 3-phosphate (EPSP) synthase, of plants, which is the primary target for the herbicide glyphosate (Stalker *et al.*, 1985; Steinrücken & Amrhein, 1980).

Sequence analysis of proteins and nucleic acids provides a new and powerful approach to measure evolutionary relationships between organisms. Genotypic information is superior to phenotypic information in two main ways: firstly, it is readily, reliably and precisely interpreted; and secondly, it is innately more informative. Data presented here brings the total of different *aroA* genes sequenced to date to eleven and provides an ideal opportunity to examine the evolutionary relationships between the EPSP synthase enzymes of bacteria, fungi and plants.

In this study we describe the cloning and characterization of the *serC*–*aroA* operon of *S. gallinarum*, compare the predicted protein sequence with that of other organisms and construct evolutionary trees showing the

Abbreviations: Ap, ampicillin; EPSP, 5-enolpyruvylshikimate 3-phosphate; MM, minimal medium; PSAT, 3-phosphoserine aminotransferase.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned the accession number M36178.

evolutionary relationships between prokaryotic and eukaryotic sequences. Implications for the evolutionary relationships between bacteria, fungi and plants are discussed.

Methods

Bacterial strains, plasmids and bacteriophages. *S. gallinarum* strain 9 was provided by P. Barrow (Barrow *et al.*, 1987). *Escherichia coli* AB2829, which harbours a stable *aroA* mutation, was obtained from G. Dougan (Maskell *et al.*, 1988). Cosmid pHC79, encoding ampicillin (Ap) and tetracycline resistance, plasmid pUC18, encoding Ap resistance, and bacteriophage M13mp19 were supplied by Gibco BRL. *E. coli* K12 strain TB1, F *ara* $\Delta(lac-proAB)$ *rpsL* ψ 80d *lacZ* Δ M15 *hsdR17* (*r₊ m₊*), was used as a host for plasmids; HB101 and JM101 (Sambrook *et al.*, 1989) were used as hosts for pHC79 and M13 respectively.

Media and reagents. Bacteria were grown in L broth or on L agar (Miller, 1972). Strains carrying M13 were grown in 2 $\times$ YT broth or on H agar containing X-Gal and IPTG (Miller, 1972). Minimal medium (MM) agar plates were as described by Miller (1972); vitamin B1 was added to a final concentration of 100 μ g ml⁻¹, and when required, tyrosine, tryptophan, phenylalanine, dihydroxybenzoic acid, and *p*-aminobenzoic acid (aromix) were added to a final concentration of 100 μ g ml⁻¹. Antibiotics and other chemicals were obtained from Sigma. Antibiotics were used at the recommended concentrations (Sambrook *et al.*, 1989). Radionucleotides [α -³²P]dATP, [γ -³²P]ATP and [α -³⁵S]dATP were obtained from Amersham International. Oligonucleotides were from Emzymatix.

DNA manipulations. Restriction endonuclease enzymes, T4 DNA ligase, exonuclease III, DNA polymerase I (Klenow fragment), S1 nuclease and T4 polynucleotide kinase were obtained from Gibco BRL and were used according to the manufacturer's instructions. Standard molecular cloning, hybridization, transformation and electrophoresis techniques were used (Sambrook *et al.*, 1989). DNA sequence analysis was performed using dideoxy chain terminators (Sanger *et al.*, 1977) using T7 polymerase (Tabor & Richardson, 1987) with a Sequenase version 2 kit (USB, Cleveland, Ohio). A 3.1 kbp fragment was subcloned into the *Pst*I site of M13mp19. Ordered sets of deletions were generated using exonuclease III (Henikoff, 1984) the primer site being protected by *Kpn*I digestion and deletion proceeding from a *Bam*HI site. End repair was effected with the Klenow fragment of DNA polymerase I and S1 nuclease, and the molecules recircularized using T4 DNA ligase. Overlapping sequencing runs were performed to determine the sequence of a 2.6 kbp fragment. Both strands were sequenced.

High molecular mass DNA was prepared from *S. gallinarum* by the method of Hull *et al.* (1981) as modified by Maskell *et al.* (1988). The cosmid library of *S. gallinarum* was prepared by first performing a partial digestion of high molecular mass chromosomal DNA with *Sau*3AI to yield fragments in the size range 30–40 kbp. The partially cleaved DNA was ligated into the *Bam*HI site of cosmid pHC79 and packaged into λ particles as described by Sambrook *et al.* (1989).

Computer analysis of sequence data. Sequences were read into a BBC microcomputer using a sonic digitizer (Graf/Bar Science Accessories Corporation) and data were analysed on Micro VAX 3600 using the computer programs of the University of Wisconsin Genetics Computer Group (UWGGC) (Devereux *et al.*, 1984), PHYLIP (Felsenstein, 1985) and CLUSTAL (Higgins & Sharp, 1989).

Evolutionary trees. The evolutionary relationship between the predicted protein sequence of the *S. gallinarum aroA* gene presented

here was compared to all available published EPSP synthase sequences. These include bacterial sequences from *E. coli* (Duncan *et al.*, 1984), *S. typhimurium* (Stalker, *et al.*, 1985), *Bordetella pertussis* (Maskell *et al.*, 1988), *Yersinia enterocolitica* (O'Gaora *et al.*, 1989) and *Bacillus subtilis* (Henner *et al.*, 1986); fungal sequences from *Saccharomyces cerevisiae* (Duncan *et al.*, 1987) and *Aspergillus nidulans* (Charles *et al.*, 1986); and plant sequences from *Petunia hybrida*, *Lycopersicon esculentum* (Gasser *et al.*, 1988) and *Arabidopsis thaliana* (Klee *et al.*, 1987).

The computer programs used to examine the potential evolutionary relationships between the eleven different *aroA* genes were FITCH and KITSCH, from the Phylogeny Inference Package (PHYLIP) of Felsenstein (1985), and CLUSTAL (Higgins & Sharp, 1989). These programs use distance matrix methods to generate evolutionary trees. FITCH and KITSCH use an algorithm based on the least squares method of Fitch & Margoliash (1967) while CLUSTAL utilizes the unweighted pair group maximum averages (UPGMA) algorithm. Details of how data were prepared for analysis by the FITCH and KITSCH programs are described by Griffin & Bournsnel (1990). Basically, trees were constructed from a table of distances generated by the UWGGC program GAP (Devereux *et al.*, 1984) for alignments between pairs of sequences. The three programs (FITCH, KITSCH and CLUSTAL) make various assumptions about the data; these assumptions are discussed in detail elsewhere (Bishop *et al.*, 1987; Felsenstein, 1985; Lewin, 1988; Taten, 1985). The programs FITCH and KITSCH were run on a Micro VAX 3600, and CLUSTAL was run on an IBM-compatible personal microcomputer.

Results and Discussion

Cloning of the serC-aroA operon of S. gallinarum 9

A cosmid gene library of *S. gallinarum* was constructed in pHC79 using both *E. coli* HB101 and the *aroA E. coli* AB2829 as hosts. Infected *E. coli* AB2829 cells were plated onto MM agar containing Ap and incubated at 37 °C for 3 d. One colony (cosmid 1) was picked and purified on MM plus Ap. This strain grew well on solid and liquid MM.

Infected *E. coli* HB101 cells were plated on L agar containing Ap. Colonies were replica-plated, transferred to nitrocellulose, and probed with a 25-mer oligonucleotide (5'-AATCCCTGACGTTACAACCCATCGC-3') labelled with [γ -³²P]ATP (Sambrook *et al.*, 1989). This sequence, from an area of high conservation near the 5' end of the *aroA* cistron of *E. coli* and *S. typhimurium*, should hybridize near the centre of the *S. gallinarum serC-aroA* operon because *serC* is promoter-proximal to *aroA* (Hoiseth & Stocker, 1985). Four positive colonies out of 1000 were detected. Caesium chloride purified DNA from one of these (cosmid 2) was used to transform *E. coli* AB2829. Transformants grew well in liquid and solid MM.

DNA from both cosmid 1 and cosmid 2 was digested with a range of restriction endonucleases and Southern hybridization performed using the 25-mer oligonucleotide probe. The probe hybridized to a 6.0 kbp *Eco*RI fragment, a 5.6 kbp *Sal*I fragment, and a 4.5 kbp *Eco*RI-*Sal*I fragment from cosmid 2, each of which was

Fig. 1. Nucleotide sequences of the *serC-aroA* operon from *S. gallinarum* (excluding the *serC* coding region, published by Griffin, 1990). The *serC* promoter elements (-35 and -10 regions) and the factor-independent RNA polymerase terminator site are indicated. Potential ribosome-binding sites (RBS) are also marked. The *aroA* structural gene and predicted amino acid sequence for EPSP synthase are shown (positions 1293–2576).

complemented the *aroA* mutation of *E. coli* AB2829. The DNA sequence of the ends of the inserted fragments was determined using forward and reverse primers (Sambrook *et al.*, 1989). The 3' end of the *aroA* gene was

1 100
AN evhpgvahssnvicapPGsKs
SC lvypfkdi padqkvvipPGsKs
AT maqvsricngvqmpslisnlsskssqrksplsvslktqghpraypisswglkksqmtligselrplkvmsvstaekaseivlqpireisqliklPGsKs
LE maqissmaqqigtlslnssnlsktqkgplvsnsllffgskkltqisakslgvfkksdvlrvvrkssfrisasvataekpheivlxpikdisgtvklPGsKs
PH maqinnmaqqigtln-pnsnfhkpg-vpksssfvfgskklkn-sansmlvlkksd-sif-mqkfcsfrisasvataekpseivlqpikeisgtvklPGsKs
EC mesltlqpiarvdtinlPGsKt
SG mesltlqpiarvdgainlPGsKs
ST mesltlqpiarvdgainlPGsKs
YE mlesltlhpialingtvnlPGsKs
BP msglayldlpaarlalargevalPGsKs
BS mkrdrvqtlhgeihiPGdKs

101 200
AN icnRaiLvlaALgsGtcriknlLhsdDtevmLnalerlgaatfsweeegevlvvnGkgg—n-lqassspLylGNaGTasRflttvatlansstvsds-s
SC isnRaiLilaALgeGqckiknlLhsdDtkhmlTavhelkgatiswedngetvvveGhgg—stlsacadpLylGNaGTasRfltslaalvnstssqkyi
AT lsnRi lllaALseGttvvdnlLnsdDihymLgalktl-glnvetdsennravveGcggi fpassidsksdieLylGNaGTamRpltaavtaaggns—y
LE lsnRi lllaALseGttvvdnlLnsdDihymLgalktl-glhveddnengraiveGcgggfvpvgkseeieqLflGNaGTamRpltaavtvagghsr—y
PH lsnRi lllaALseGttvvdnlLnsdDihymLgalktl-glhveedsanqravveGcgglfpvgkeskeeieqLflGNaGTamRpltaavtvagghsr—y
EC vsnRai lllaALahGktvltlnLdsdDvrhmLnaltal-gvstltsadrtceiiGnggpl—haegaleLflGNaGTamRplaaalcl—gsnd—i
SG vsnRai lllaALacGktvltlnLdsdDvrhmLnalsal-ginyttsadrtcditGnggpl—rappaleLflGNaGTamRplaaalcl—ggne—i
ST vsnRai lllaALpcGktaltlnLdsdDvrhmLnalsal-ginyttsadrtcditGnggal—rappaleLflGNaGTamRplaaalcl—ggne—i
YE vsnRai lllaALaeGttqlnlnLdsdDihymLnalg—gvkyrlsadrtcevdGlggkl—vaeqpleLflGNaGTamRplaaalcl—gknd—i
BP isnRv lllaALaeGsteitglLdsdDtrvmLaalrql-gvsvgevad—GcvtiegarvfpteqaeLflGNaGTafRpltaaalalmggd—y
BS ishRsvmfALaaGttvknflpgadC—LstidcfrkmgvhieqsssdvviHGk—idalkepesLdvGNSGTTiRlmlgil—agrpfs

201 300
AN vltGnnmkqRPigdlvdaLtanvplntskgraslpkiaasggfaggninlaakvsSqvssllmcApyakepv—tlrlvggkkipisqpyidmTtam
SC vltGnnmkqRPiaplvdsLrangtkieylnnegslpikvytdsvfkggrielaatvssSqvssllmcApyaeevp—tlalvggkkipisklyvdmTtkm
AT vldGvprmrERPIgdlvdgkqlgaedvdcslgtncppvrivskgglpggkvklsgsisSqyltal lmaApl—algdveieidklisvpyvemTlkl
LE vldGvprmrERPIgdlvdgkqlgaedvdcslgtncppvrivskgglpggkvklsgsisSqyltal lmaApl—algdveieidklisvpyvemTlkl
PH vldGvprmrERPIsdldvdgkqlgaedvdcflgtkcpvrivskgglpggkvklsgsisSqyltal lmaApl—algdveieidklisvpyvemTlkl
EC vltGepmrERPIghlvdaLrlggakitylegenypplrl—qggftggnvvdvsgvsSqfltal lmtApl—apedtirikgdlvskpyidiTlnl
SG vltGepmrERPIghlvdsLrqqganidylegenypplrl—rggftggdievdsgvsSqfltal lmtApl—apkdtiirvkgelvsckpyidiTlnl
ST vltGepmrERPIghlvdsLrqqganidylegenypplrl—rggftggdievdsgvsSqfltal lmtApl—apkdtiirvkgelvsckpyidiTlnl
YE vltGepmrERPIghlvdaLrqqgaqidylegenyr-ci—aggrfggkltdvsgvsSqfltal lmtApl—aeqdtiqiggelvskpyidiTlhl
BP rlsGvprmrERPIgdlvdaLrqqgaqieyldgagypplriggssrvdgpvrvegsSvSsqfltal lmaAplarrsgqditievgeliskpyieiTlnl
BS avaGdesiakRPmkrvteplLkmgakidraggeftplsvsgas-lkg—idyvspvaSaqiksavllAqlgaegtvtvtephksr—dhTerm

301 400
AN mrsFGidvqkstteehyhipggr-yvnpaeyviesDaSaAtypLavaA-vtgttctvpniGsaslggdarfavevLrpMGctveqtetststvtggsdg-
SC mekFGinvetsttepytyyikpgh-yinpseyviesDaSaAtypLafaa-mtgttvtvpniGfeslggdarfardvLkpMGckitqtatststvtgsgpvg-
AT merFGsvhsdswdr-ffvkgggkykspgnayvegDaSaAsyflAgaA-itgetvtvegcGttslggdvka-fvLekMGckvswtensvtvtgpprd-
LE merFGvfevhsdswdr-flvkgggkykspgkafvegDaSaAsyflAgaA-vtggvtvvegcGtsslggdvka-fvLekMGaevtwensvtvkgpprn-
PH merFGisvehssdswdr-ffvrqggkykspgkafvegDaSaAsyflAgaA-vtggitvvegcGtnslggdvka-fvLekMGaevtwensvtvkgpprs-
EC mktFGveienqh-yqq-fvvkgggysqsgtylvegDaSaAsyflAaaa-ikggtvkvgtgiGrnsnggdirfa-dvLekMGaticwgddyi-
SG mktFGveianhh-yqq-fvvkgggqyhsppgrylvegDaSaAsyflAaga-ikggtvkvgtgiGrksnggdirfa-dvLekMGatitwgddfi-
ST mktFGveianhh-yqq-fvvkgggqyhsppgrylvegDaSaAsyflAaga-ikggtvkvgtgiGrksnggdirfa-dvLekMGatitwgddfi-
YE mkaFGvdvhen-yqi-fhikgggtyrsggiylvegDaSaAsyflAaaa-ikggtvrvgtgiGkqsvggdtkfa-dvLekMGakiswgddyi-
BP marFGvsv-rrdgwra-ftiadvyrpggmaiegdDaSaAsyflALga-igggpvrvtgGedsiggdvafa-atLaamGadvrypggwietrgrvva-
BS lsaFGvkl sed-qtsvsiaggqkl-taadi fvpGDiSaAaffLaagAmvpsrivlknvGlnptrtgi—idvLqnMGakleikpsadsgaepygdI

401 500
AN ilratskrgygt-ndrcvprcfrtgshlpmeksqtpvssgianqrVKEcnRikamkdeLaKfGvicrehdDgleidgidr-snlrqpvq—gvfc
SC tlkplkhvdmepmtdafltacvvaishdsdpnsantttiegiangrVKEcnRilamatELaKfGvkttelPdgiqvhglnsikdlkvpsdssgpgvgt
AT afgmrhlraidvnmnkmpdvamtlavvalfadg—pttirdvaswrvKeteRmiaictELrKlGatveegsDycvitppk—vktaeidt
LE ssgmkhlraidvnmnkmpdvamtlavvalfadg—pttirdvaswrvKeteRmiaictELrKlGatveegsDyciitppek—lnvteidt
PH ssgmkhlraidvnmnkmpdvamtlavvalfadg—pttirdvaswrvKeteRmiaictELrKlGatveegsDyciitppek—lnvteidt
EC sctrgelnaidndmnhpdaamtiataalfakg—ttlrlniynwrvKeteDRlfamatELrKvGaeveeghDyiritppek—lnfaeiad
SG actrgelnaidndmnhpdaamtiataalfakg—ttlrlniynwrvKeteDRlfamatELrKvGaeveeghDyiritppek—lqhadiqt
ST actrgelnaidndmnhpdaamtiataalfakg—ttlrlniynwrvKeteDRlfamatELrKvGaeveeghDyiritppek—lqhadiqt
YE ecsrgelggidndmnhpdaamtiataalfakg—ptviriynwrvKeteDRlsamatELrKvGaeveeghDyirivppaq—liaaeigt
BP eggr-lkafadfnlidaamtiataalfakg—pcrlrniyswrvKeteDRihamtELrKlGagvqsgaDwleavapp—ggwrdahtigt
BS iietsslkaveiggdiprlideipiiallatgaegtvtvikdaaelkVKEtnRidttvsvELrKlGaeieptaDgmkyvgkqtlkg—gaavss

Fig. 2. Alignments of the amino acid sequences predicted for the EPSP synthase proteins from eleven different species. The sequences were aligned using the computer program CLUSTAL. In these comparisons, the gap penalty was set at 10. Minor adjustments to the alignments generated by the computer program were used to improve the clarity of the alignment. These involved the removal of gaps in the fungal and bacterial sequences introduced by the program between positions 1 and 100 of the alignment, and the use of another computer program, CLUSTALOUT (M. E. G. Boursnell, personal communication), to change the linelength of the alignment to 100, and introduce an uppercase letter to denote the positions of amino acids present in all eleven sequences. Bold type is used to indicate the positions of identity and conservative substitution. AN, *Aspergillus nidulans*; SC, *Saccharomyces cerevisiae*; AT, *Arabidopsis thaliana*; LE, *Lysopersicon esculentum*; PH, *Petunia hybrida*; EC, *Escherichia coli*; SG, *Salmonella gallinarum*; ST, *Salmonella typhimurium*; YE, *Yersinia enterocolitica*; BP, *Bordetella pertussis*; BS, *Bacillus subtilis*.

mapped very close to the *SalI* site of the 4.5 kbp *EcoRI-SalI* fragment. The *aroA* gene is 1.3 kbp in length (by analogy with *E. coli*); this leaves 3.2 kbp of DNA upstream of *aroA*, an ample amount to contain the *serC* gene (1.1 kbp in *E. coli*) and promoter region.

DNA sequence analysis

Sequence was determined for a 2.6 kbp region (Fig. 1), which was shown by homology with the *E. coli* sequence to contain the entire *serC-aroA* operon. The *serC* coding region has been published previously (Griffin, 1990). We have identified a putative promoter region and potential ribosome-binding sites (Fig. 1).

The *serC* and *aroA* open reading frames are separated by a 70 bp non-coding sequence. A sequence resembling a factor-independent RNA polymerase terminator site is located in this region (Fig. 1). The *E. coli* intergenic region also contains a terminator (Duncan & Coggins, 1986) but the longer 95 bp non-coding region of *Y. enterocolitica* apparently does not (O'Gaora *et al.*, 1989). The sequence of this region is unavailable for *S. typhimurium* or other bacterial species.

Comparison of the *S. gallinarum serC* and *aroA* sequences with those of other organisms

An alignment of the predicted protein sequence of the *aroA* gene of *S. gallinarum* with those of ten other organisms is presented in Fig. 2. This alignment should be very useful in the identification of residues important for polypeptide functioning since previous alignments have been based on a maximum of only five different

EPSP synthase sequences (Maskell *et al.*, 1988; O'Gaora *et al.*, 1989). Several areas of very high homology are evident. The putative glyphosate-binding site (Gasser *et al.*, 1988) is conserved in all eleven species sequenced to date (amino acids 172–186 as numbered in Fig. 2). The functions of the other conserved regions are uncertain. The protein sequences from the higher eukaryotes contain a 76 amino acid extension at the N-terminal region of the EPSP synthase sequence (Fig. 2). This is the transit peptide sequence which is required for transport of the enzyme into the chloroplast stroma (della-Cioppa *et al.*, 1986). There are nine base pair differences between the DNA sequences of *S. gallinarum* and *S. typhimurium*. These result in the production of four conservative amino acid substitutions: positions 112, 117, 159 and 247 in Fig. 2. Since the *aroA* gene is highly conserved in the majority of species studied so far, the small difference (less than 1%) between the *aroA* genes of *S. gallinarum* and *S. typhimurium* probably reflects genetic drift between the two *Salmonella* serotypes (Table 1). The *S. gallinarum aroA* protein is slightly more homologous (89.2%) than the *S. typhimurium* enzyme (88.8%) to the *E. coli aroA* gene product (Table 1).

The *serC* protein, 3-phosphoserine aminotransferase (PSAT), encoded by the same bacterial operon as EPSP synthase, also shows a high degree of conservation between different species (Table 2). We compared the *S. gallinarum* sequence to the three other available PSAT sequences (Duncan & Coggins, 1986; Misrahi *et al.*, 1987; O'Gaora *et al.*, 1989; van der Zel *et al.*, 1989). The resulting alignment is presented in Fig. 3. We noted the presence of several regions which were highly conserved even between the mammalian and bacterial sequences. The functions of these regions are unknown. This

```

1
EC Ma---QifNFssGPAmLPaeVLkqaqqELrdwnGLGtSVmEvSHRgkeFiqvaaeaEkdfRdLlnvPsNYKVlFchGGgrgQFaaVPLNllGdktt--AD
SG Ma---QvFNfssGPAmLPaeVLkllaqqELrdvhGLGtSVmEiSHRgkeFiqvaaeaEqdfRdLlniPsNYKVlFchGGgrgQFagVPLNllGdkt--AD
YE Mt---QvYNfSaGPAmLPveVLrraeqELrnwhGLGtSVmEiSHRskeFmqvaaeaEkdlRdLmqiPaNYKVlFchGGrraQFaaVPLNllGdsrs--AD
OC MdsprQivNFgpgPAkLPhsVLleiqkELldykGLGiSVlEmSHRssdFakivnntEnlvReLlavPdNYKVlFlqGGcgQFsaVPLNliGlkpgrcAD
* * * * *
101
EC YvdaGyWaasAikEAKKycpvnfdakvtvdglravkpmreWqLsdnAaYmhYCpNETidGaiidetPDfgadvvVaaDfSSStiLSrPiDVSryGVIyAG
SG YvdaGyWaasAikEAKKycapqiidaktvtdgkravkpmreWqLsdnAaYlhYCpNETidGaiidetPDfgpevvVtaDfSSStiLSaPlDVSryGVIyAG
YE YidGyWahsAvkEAqKycpvnidvtthdngvtgiapmkqWkLsdnAaYvhYCpNETidGlaineePDfgnkvvVa-DySSSiLSrPiDVSryGVIyAG
OC YvvtGaWsaAaeEAKKfgtvnivhpk1--gsytkipdpstWnLnpdAsYvyYcANETvhGvefdvPDvkgailVc-DmSSnflLSrPvDVSkfGVIyAG
* * * * *
201
EC AQKNiGpAGLTiVIVReDLLGkAniacPSiLdYsilndngSmfNTPPtFawYlsGLVfkWlKanGGvaeMdkinqqKaellYgvIDnsd-FYrndVakrN
SG AQKNiGpAGLTiVIVReDLLGkAhescPSiLdYtvlndndSmfNTPPtFawYlsGLVfkWlKaqGGvaaMhKinqqKaellYgvIDnsd-FYrndVaqaN
YE AQKNiGpAGLTiVIVReDLLGkArtelPSiLdYkvlaendSmfNTPPtFawYlsGLVfkWlKqGGlgeMgKrnqaKaellYgaIDrtd-FYrnqVataN
OC AQKNvGaAGvTvVIVRdDLLGfAlrecPSvLeYkvqatssSlyNTPPCFsiYvmGLVleWiKnnGGaaaMkKlstiKsqmiYeiIDnsqgFYvcpVeprN
* * * * *
301
EC RSrMNvPFq1a---DsaLdKlFLeesfaag1haLKGHRvvGGmRaSiYNAmplegVkaLtdFMveF-Errhg
SG RSrMNvPFq1a---DnaLdKvFLeesfaag1haLKGHRvvGGmRaSiYNAmplegVkaLtdFMidF-Errhg
YE RSrMNvPFqmi---DpsLdKlFLseaeag1qaLKGHRvvaGGmRaSiYNAmplegVkaLtdFMaeF-Errhg
OC RSrMNiPFrignakgDeaLeKrFLdkalelhmisLKGHRsvGGvRvSLYNAvtiedVqkLasFMknFLEmhql
* * * * *
373

```

Fig. 3. Alignments of the amino acid sequences predicted for the translated *serC* genes from four different species. The sequences were aligned using the computer program CLUSTAL. In these comparisons, the gap penalty was set at 10. CLUSTALOUT was used to change the linelength of the alignment to 100, and introduce an uppercase letter to denote the positions where the observed amino acid was present in all four sequences. An asterisk (*) indicates the position of a conservative substitution. EC, *Escherichia coli*; SG, *Salmonella gallinarum*; YE, *Yersinia enterocolitica*; OC, *Oryctolagus cuniculus* (rabbit).

Table 1. Percentage amino acid identity between the translated *aroA* genes from eleven different species

Identity scores were produced using the Wisconsin program GAP with a gap weight and gap length of 5.0 and 0.3 respectively. AN, *Aspergillus nidulans*; AT, *Arabidopsis thaliana*; BP, *Bordetella pertussis*; BS, *Bacillus subtilis*; EC, *Escherichia coli*; LE, *Lycopersicon esculentum*; PH, *Petunia hybrida*; SC, *Saccharomyces cerevisiae*; SG, *Salmonella gallinarum*; ST, *Salmonella typhimurium*; YE *Yersinia enterocolitica*.

	AN	AT	BP	BS	EC	LE	PH	SC	SG	ST	YE
AN	100.0										
AT	39.5	100.0									
BP	34.6	49.1	100.0								
BS	23.5	26.6	24.4	100.0							
EC	36.0	54.7	55.4	25.6	100.0						
LE	38.3	75.4	49.5	26.6	55.4	100.0					
PH	38.8	75.2	49.8	27.1	54.2	87.6	100.0				
SC	57.1	42.0	37.2	24.5	40.1	42.0	42.0	100.0			
SG	35.5	54.5	55.0	27.1	89.2	54.5	53.5	41.1	100.0		
ST	35.5	54.2	54.7	26.9	88.8	54.2	53.3	41.2	99.1	100.0	
YE	35.6	55.6	56.1	26.8	80.5	56.3	55.0	41.2	79.6	79.3	100.0

alignment also supports the theory that the sequence reported by Misrahi *et al.* (1987) is in fact the rabbit *serC* gene (van der Zel *et al.*, 1989).

Evolutionary relationships between the *aroA* gene products (EPSP synthase)

Trees depicting evolutionary relationships between the sequences of eleven different EPSP synthase enzyme

sequences from the prokaryotic and eukaryotic species are presented in Fig. 4. Three different methods were used to generate these trees, since it has previously been shown that different computer programs can generate different trees from the same data (Griffin & Boursnell, 1990). As mentioned in Methods, CLUSTAL utilizes a different algorithm from that used by KITSCH and FITCH, while KITSCH and FITCH differ from each other in that KITSCH assumes that a constant evolutionary clock is valid, i.e. the rate of evolution in each branch is the same.

Fig. 4. Evolutionary trees depicting the relative divergence of the translated *aroA* gene sequences. The trees were generated using three different computer programs: KITSCH (a), FITCH (b) and CLUSTAL (c). The amount of divergence is represented by the branch lengths, i.e. horizontal lines. Vertical separation is for clarity only. Internal branches emerge at nodes, each of which can be rotated if desired. The position of the root of the KITSCH is indicated by the horizontal line on the left of the tree. The direction of divergence is indicated by an arrow.

Table 2. *Percentage amino acid identity and similarity between the translated serC genes from four different species*

The computer program GAP was used to identify matches between identical and similar amino acids. Both sets of data are presented, with the percentage similarity scores given in parentheses. In these comparisons, the gap weight and gap length were set at 5.0 and 0.3 respectively. EC, *Escherichia coli*; OC, *Oryctolagus cuniculus* (rabbit); SG, *Salmonella gallinarum*; YE, *Yersinia enterocolitica*.

	EC	OC	SG	YE
EC	100.0			
OC	49.0 (68.8)	100.0		
SG	89.8 (95.6)	48.5 (67.1)	100.0	
YE	75.6 (86.7)	46.8 (65.7)	75.9 (85.6)	100.0

The existence of such an evolutionary clock is still controversial; however it has proven to be a very valuable 'dating' device in molecular evolution (reviewed by Lewin, 1988; Wilson *et al.*, 1987). In the KITSCH tree the branch lengths reflect the time elapsed

since divergence has occurred, while in the tree generated using FITCH, the branch lengths are proportional to the expected amount of change since divergence has occurred. Regardless of the computer program used, the evolutionary trees clearly divide the enzyme sequences into four different groups; Gram-negative bacteria, Gram-positive bacteria, plants and fungi. In fact the trees depict *B. subtilis* as being no more closely related to the Gram-negative bacterial species than it is to the fungal and plant species. Two pairs of species are consistently depicted as being closely related. These are the bacterial *Salmonella* species, *S. gallinarum*/*S. typhimurium*, and the members of the plant family *Solanaceae*, *L. esculentum*/*P. hybrida*. In addition to these similarities, the programs also consistently show that *B. pertussis* is the first Gram-negative bacterial species to emerge following the divergence of the ancestral gene into early fungal and bacterial/plant genes. The plant EPSP synthase enzyme differs from that of the fungal and bacterial enzymes in that it is synthesized as a precursor with an N-terminal transit peptide sequence (see alignment Fig. 2). We constructed trees using the mature plant EPSP synthase enzyme sequences, i.e. without the

transit sequences, and found that these were identical to the trees depicted in Fig. 4 (data not presented).

Our results show that the EPSP synthase enzymes are divided into four separate groups on the evolutionary trees: fungi, plants, Gram-negative bacteria and Gram-positive bacteria. Because plants and fungi are both eukaryotic organisms, we had expected that their EPSP synthase enzymes would be more related to each other than they would be to the bacterial enzymes. The trees presented here clearly show that the divergence between plant and Gram-negative bacterial EPSP synthases occurred some time later than the divergence of the fungal synthase from an ancestral enzyme. There are several possible reasons for this. For instance, the groupings depicted may simply reflect the fact that in fungi, the EPSP synthase is part of a pentafunctional enzyme, AROM, expressed as a polypeptide encoded by the *ARO1* gene (Charles *et al.*, 1986; Duncan *et al.*, 1987), while in bacteria and plants the EPSP synthase enzyme is encoded by an individual *aroA* structural gene (Klee *et al.*, 1987; Gasser *et al.*, 1988; Henner *et al.*, 1986; Maskell *et al.*, 1988; O'Gaora *et al.*, 1989; Duncan *et al.*, 1984; Stalker *et al.*, 1985). Our findings also support the hypothesis that plastids in plants are descendants of endosymbiotic prokaryotes that invaded ancestral plant cells (Margulis, 1970). Some genes present in these prokaryotes are proposed to have migrated to the nucleus of their eukaryotic host (Weeden, 1981; Gasser *et al.*, 1988). Such genes would retain homology with prokaryotic sequences. A similar hypothesis has been proposed to explain the evolutionary relationships between the glyceraldehyde-3-phosphate enzymes of plants, animals and fungi (Shin *et al.*, 1986).

We would like to thank the following for generously providing bacterial strains: P. Barrow, R. Curtiss, G. Dougan, T. Elliott, P. W. Postma, J. R. Roth, K. Sanderson and B. A. D. Stocker. Technical assistance was provided by K. Billington.

References

- BARROW, P. A., SIMPSON, J. M., LOVELL, M. A. & BINNS, M. M. (1987). Contribution of *Salmonella gallinarum* large plasmid toward virulence in fowl typhoid. *Infection and Immunity* **55**, 388–392.
- BISHOP, M. J., FRIDAY, A. E. & THOMPSON, A. E. (1987). Inference of evolutionary relationships. In *Nucleic Acids and Protein Sequence Analysis: a Practical Approach*, pp. 359–385. Edited by M. J. Bishop & C. J. Rawlings. Oxford: IRL Press.
- CHARLES, I. G., KEYTE, J. W., BRAMMAR, W. J., SMITH, M. M. & HAWKINS, A. R. (1986). The isolation and nucleotide sequence of the complex AROM locus of *Aspergillus nidulans*. *Nucleic Acids Research* **14**, 2201–2213.
- DELLA-CIOPPA, G., BAUER, C., KLEIN, B., SHAH, D., FRALEY, R. & KISHORE, G. (1986). Translocation of the precursor of 5-enolpyruvylshikimate-3-phosphate synthase into chloroplasts of higher plants in vitro. *Proceedings of the National Academy of Sciences of the United States of America* **83**, 6873–6877.
- DEVEREUX, J., HAEBERLI, P. & SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 216–223.
- DUNCAN, K. & COGGINS, J. R. (1986). The *serC-aroA* operon of *Escherichia coli*. *Biochemical Journal* **234**, 49–57.
- DUNCAN, K., LEWENDON, A. & COGGINS, J. R. (1984). The complete amino acid sequence of the *Escherichia coli* 5-enolpyruvylshikimate-3-phosphate synthase. *FEBS Letters* **170**, 59–63.
- DUNCAN, K., EDWARDS, R. M. & COGGINS, J. R. (1987). The pentafunctional *arom* enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. *Biochemical Journal* **246**, 375–386.
- FELSENSTEIN, J. (1985). Confidence limits on phylogenies; an approach using the bootstrap. *Evolution* **39**, 783–791.
- FITCH, W. M. & MARGOLISH, E. (1967). Construction of phylogenetic trees. *Science* **155**, 279–284.
- GASSER, C. S., WINTERS, J. A., HIRONAKA, C. M. & SHAH, D. M. (1988). Structure, expression, and evolution of the 5-enolpyruvylshikimate-3-phosphate synthase genes of petunia and tomato. *Journal of Biological Chemistry* **263**, 4280–4289.
- GRIFFIN, H. G. (1990). Nucleotide sequence of the *Salmonella serC* gene. *Nucleic Acids Research* **18**, 4260.
- GRIFFIN, A. M. & BOURSNELL, M. E. G. (1990). Analysis of the nucleotide sequence of DNA from the thymidine kinase gene from infectious laryngotracheitis virus; potential evolutionary relationships between the herpesvirus subfamilies. *Journal of General Virology* **71**, 841–850.
- HENIKOFF, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351–359.
- HENNER, D. J., BAND, L., FLAGGS, G. & CHEN, E. (1986). The organization and nucleotide sequence of the *Bacillus subtilis* *hisH*, *tyrA* and *aroE* genes. *Gene* **49**, 147–152.
- HIGGINS, D. G. & SHARP, P. M. (1989). Fast sensitive multiple sequence alignments on a micro computer. *CABIOS* **5**, 151–153.
- HOISETH, S. K. & STOCKER, B. A. D. (1981). Aromatic-dependent *Salmonella typhimurium* are non-virulent and are effective live vaccines. *Nature, London* **291**, 238–239.
- HOISETH, S. K. & STOCKER, B. A. D. (1985). Genes *aroA* and *serC* of *Salmonella typhimurium* constitute an operon. *Journal of Bacteriology* **163**, 355–361.
- HULL, R. A., GILL, R., HSU, P., MINSHEW, B. H. & FALKOW, S. (1981). Construction and expression of recombinant plasmids encoding type 1 on D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. *Infection and Immunity* **33**, 933–938.
- KLEE, H. J., MUSKOPF, Y. M. & GASSER, C. S. (1987). Cloning of the *Arabidopsis thaliana* gene encoding the 5-enolpyruvylshikimate-3-phosphate synthase: sequence analysis and manipulation to obtain glyphosate-tolerant plants. *Molecular and General Genetics* **210**, 437–442.
- LEWIN, R. (1988). Molecular clock turns a quarter century. *Science* **239**, 561–563.
- MARGULIS, L. (1970). Eucaryotic plants. In *Origin of Eucaryotic Cells*, pp. 276–293. Edited by L. Margulis. New Haven, CT: Yale University Press.
- MASKELL, D. J., MORRISSEY, P. & DOUGAN, G. (1988). Cloning and nucleotide sequence of the *aroA* gene of *Bordetella pertussis*. *Journal of Bacteriology* **170**, 2467–2471.
- MILLER, J. H. (1972). *Experiments in Molecular Genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- MISRAHI, M., ATGER, M. & MILGROM, E. (1987). A novel progesterone-induced messenger RNA in rabbit and human endometria. Cloning and sequence analysis of the complementary DNA. *Biochemistry* **26**, 3975–3982.
- O'GAORA, P., MASKELL, D., COLEMAN, D., CAFFERKEY, M. & DOUGAN, G. (1989). Cloning and characterisation of the *serC* and *aroA* genes of *Yersinia enterocolitica* and construction of an *aroA* mutant. *Gene* **84**, 23–30.
- POMEROY, B. S. (1984). Fowl typhoid. In *Diseases of Poultry*, 8th edn, pp. 79–91. Edited by M. S. Hofstad, H. J. Barries, B. W. Calnek, W. M. Reid & H. W. Yoder Jr. Ames: Iowa State University Press.
- SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

- SANGER, F., NICKLEN, S. & COULSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74**, 5463–5467.
- SHIN, M.-C., LAZAR, G. & HOWARD, H. M. (1986). Evidence in favour of the symbiotic origin of chloroplasts: primary structure and evolution of tobacco glyceraldehyde-3-phosphate dehydrogenases. *Cell* **47**, 73–80.
- SILVA, E. N. (1985). The *Salmonella gallinarum* problem in central and South America. In *Proceedings of the International Symposium on Salmonella, New Orleans, Louisiana 1984*, pp. 150–160. Edited by G. H. Snoeyenbos. Philadelphia: American Association of Avian Pathologists.
- SILVA, E. N., SNOEYENBOS, G. H., WEINACK, O. M. & SMYSER, C. F. (1981). Studies on the use of 9R strain of *Salmonella gallinarum* as a vaccine in chickens. *Avian Diseases* **25**, 38–52.
- SMITH, B. P., REINA-GUERRA, M., HOISETH, S. K., STOCKER, B. A. D., HABASHA, R., JOHNSON, E. & MERRITT, F. (1984). Aromatic-dependent *Salmonella typhimurium* as modified live vaccines for calves. *American Journal of Veterinary Research* **45**, 59–66.
- SMITH, H. W. (1956). The use of live vaccines in experimental *Salmonella gallinarum* infection in chickens with observations on their interference effect. *Journal of Hygiene* **54**, 419–432.
- SMITH, H. W., TUCKER, J. F. & LOVELL, M. A. (1981). Furazolidone resistance in *Salmonella gallinarum*: the relationship between *in vitro* and *in vivo* determination of resistance. *Journal of Hygiene* **87**, 71–81.
- STALKER, D. M., HIATT, W. F. & COMAI, L. (1985). A single amino acid substitution in the enzyme 5-enolpyruvylshikimate-3-phosphate synthase confers resistance to the herbicide glyphosate. *Journal of Biological Chemistry* **260**, 4724–4728.
- STEINRUCKEN, H. C. & AMRHEIN, N. (1980). The herbicide glyphosate is a potent inhibitor of 5-enolpyruvylshikimate acid-3-phosphate synthase. *Biochemical and Biophysical Research Communications* **94**, 1207–1212.
- STOCKER, B. A. D. (1988). Auxotrophic *Salmonella typhi* as live vaccine. *Vaccine* **6**, 141–145.
- STOCKER, B. A. D., HOISETH, S. K. & SMITH, B. P. (1983). Aromatic-dependent *Salmonella* sp. as live vaccine in mice and calves. *Developments in Biological Standardisation* **53**, 47–54.
- STUART, E. E., KENNUM, R. D. & BRUINS, H. W. (1962). Experimental studies on an isolate of *Salmonella gallinarum* apparently resistant to furazolidone. *Avian Diseases* **7**, 294–303.
- STUART, E. E., KENNUM, R. D. & BRUINS, H. W. (1967). The emergence of a furazolidone resistant strain of *Salmonella gallinarum*. *Avian Diseases* **11**, 139–145.
- TABOR, S. & RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 4767–4771.
- TATENO, Y. (1985). Theoretical aspects of molecular tree estimation. In *Population Genetics and Molecular Evolution*, pp. 239–312. Edited by T. Ohta & K. Aoki. Tokyo: Japan Science Society Press.
- WEEDEN, N. F. (1981). Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *Journal of Molecular Evolution* **17**, 133–139.
- WILSON, C. A., OCHMAN, H. & PRAGER, E. M. (1987). Molecular time scale for evolution. *Trends in Genetics* **3**, 241–247.
- VAN DER ZEL, A., LAM, H.-M. & WINKLER, M. E. (1989). Extensive homology between the *Escherichia coli* K-12 *SerC*(*PdxF*) amino-transferase and a protein encoded by a progesterone-induced mRNA in rabbit and human endometria. *Nucleic Acids Research* **20**, 8379.