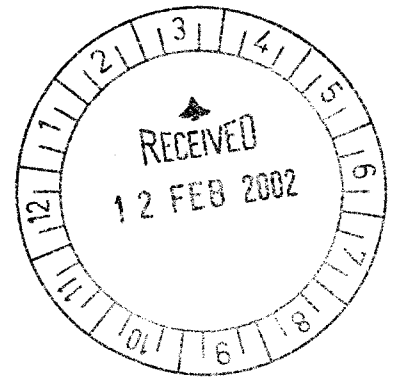


BIOINFORMATIC  
DATA

**MONSANTO**



**APPLICANT:** **Monsanto Australia Limited**

**A416**  
**CP4 EPSPS gene in Roundup Ready® Corn Line NK603**

---

**SUBMISSION:** Application to Australia New Zealand Food Authority  
for the inclusion of corn containing the CP4 EPSPS  
gene by Monsanto in Standard A18 - Food Derived  
From Gene Technology

**VOLUME:** 1 of 1

**SUPPORTING INFORMATION**

**DATE:** 11 February 2002

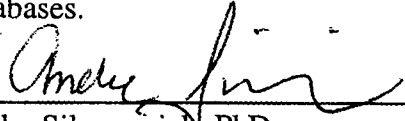
---

**PREPARED BY:** Megan Shaw  
Regulatory Product Manager

®/™ Registered/trademark of Monsanto Company USA, used under licence by Monsanto Australia Limited

## Summary of Bioinformatic Analyses for CP4 EPSPS L214P produced in corn event NK603

Bioinformatic analyses were performed for CP4 EPSPS L214P on Nov. 6, 2001 as part of the NK603 safety assessment described in MSL-17600. The bioinformatic data were reviewed by Roderick McCoy MD, PhD and Andre Silvanovich PhD. This communication contains a brief description of the methods and results and an exact copy of the original data set as can be found in the Monsanto Archives (700 Chesterfield Parkway North, St. Louis, MO, 63198). The data set is composed of four types of sequence searches: ALLERGENSEARCH analysis using CP4 EPSPS L214P as a query to search the ALLERGEN3 database and FASTA analyses using CP4 EPSPS L214P as a query to search the ALLERGEN3, TOXIN5 and ALLPEPTIDES databases.

  
Andre Silvanovich PhD  
Product Safety Center  
Protein Team Lead  
Monsanto Company Product Safety Center  
800 North Lindbergh BLVD.  
St Louis, MO  
63197

*Feb 10, 2002*  
Date

### Bioinformatic Analysis Methods

All analyses were performed using the UNIX based Genetics Computer Group (GCG) software package (version 10.0, Madison, WI) on a standard personal computer using the Windows 95 operating system and supported with Reflection X Client Manager network software (version 7.20, WRQ, Inc. Seattle, WA). Database construction was performed in the SeqLab environment of GCG. Searches were performed using command line entry through an X-terminal window.

#### *Public domain database preparation*

The database ALLPEPTIDES was comprised of publicly available protein sequences in SwissProt version 39+, TrEMBL (updated weekly) and GenPept version 124.

#### *Allergen database preparation*

A preliminary list of sequences for the ALLERGEN3 database was extracted from the ALLPEPTIDES database by executing the STRINGSEARCH function using the keyword "allergen". Non-allergen entries were identified by retrieving each flatfile and excluding the irrelevant entries from the list. The resulting list was compared to previously compiled allergen and gliadin databases (Hileman and Astwood, 1999; Metcalfe et al., 1996). Sequences present in the previous databases and missing in the STRINGSEARCH list were added to the list. Duplicate sequences were identified by performing a FASTA search on each individual sequence

against the list and removed. A sequence was considered unique if a single amino acid difference existed when compared to the remainder of the list. Thus, the database may contain multiple isoforms of the same allergen (e.g., there are 36 entries for Bet v 1). The list was finalized by adding additional allergen sequences identified by: i) comparison of the list to allergens in the publicly available list located on the Internet site, <ftp://biobase.dk/resources/pub/who-iuis/allergen.list>; and ii) performing an extensive search of the current literature using the publicly available Internet PubMed and Entrez information retrieval systems (Schuler et al., 1996). For publications retrieved in which only an N-terminal sequence was reported, the sequence was added if eight or more residues were provided. Newly identified allergens were assembled from public domain databases including Genbank and EMBL release 119 (Benson et al., 1997; Stoesser et al., 1997), PIR release (George et al., 1997), the NRL3D release 56 of RCSB PDB (Berman et al., 2000; Bernstein et al., 1977) and SwissProt version 38 (Bairoch and Apweiler, 1997). DATASET was used to compile the allergen and gliadin database that contains 658 separate protein sequences.

#### *Toxin database preparation*

The toxin (TOXIN5) sequence database was assembled from public domain databases SwissProt version 38+, TrEMBL (updated biweekly) and GenPept version 116 as previously described (Hileman and Astwood, 1999a).

#### *Sequence database searches*

The sequence similarity of CP4 EPSPS L214P was assessed by comparison to sequences in the ALLPEPTIDES, TOXIN5 and ALLERGEN3 databases using the FASTA algorithm (Pearson and Lipman, 1988). The FASTA sequence alignment tool was used to assess structural similarity. Specific parameters used for these analyses included an expectation score threshold of 10, a wordsize (k-tuple) of 2, a gap-creation penalty of 12 and gap-extension penalty of 2. The BLOSUM50 scoring matrix was used. The BLOSUM matrix series (Henikoff and Henikoff, 1992) was derived from a set of aligned, ungapped regions from protein families called the BLOCKS database. Sequences from each block were clustered. Two sequences were put into the same cluster whenever the percent of identical residues exceeded the specified level (Henikoff and Henikoff, 1996). Thus, the BLOSUM50 matrix will identify blocks of conserved residues that are at least 50% identical. BLOSUM50 works well for sequence alignments with gaps and it has become a standard matrix for recognizing distant evolutionary relationships (Pearson, 1996, Pearson, 2000).

The extent of similarity was evaluated by visual inspection of the alignment, the calculated percent identity and the E score value. The E score (expectation score) is a statistical measure of the likelihood that the observed similarity score could have occurred by chance in a search of a database this size (Pearson, 2000). The value of the E score reflects the degree of similarity and depends on the overall length of joined (gapped) local sequence alignments, the quality (percent identity, similarity) of the overlap and the size of the database. A larger E score value indicates a lower degree of similarity between the query sequence and the sequence from the database. For polypeptides, a sequence similarity may indicate sequence homology (i.e., representing a

sequence derived from a common ancestor gene with potentially homologous function). Sequences that share extensive amino acid sequence identity and/or similarity throughout the entire alignment or identified domains were considered to be biologically relevant homologues. A similar amino acid was defined as a non-identical, but physicochemically related, amino acid. Similar amino acids are structurally related and share polar, hydrophobic, or charged states. Such substitutions are referred to as "conservative" since they are unlikely to change the structures and, by inference, the functions of homologous proteins.

In addition to the FASTA comparisons of protein sequence to allergens (to assess overall structural similarity), an 8-mer search was performed. An algorithm called IDENTITYSEARCH (also referred to as ALLERGENSEARCH) was used to identify matches of eight linearly contiguous amino acid identities between the query sequence and sequences within the allergen and gliadin database (ALLERGEN3). The algorithm was run from a UNIX terminal window in GCG. This program compares the query sequence to each protein sequence in the allergen database using a sliding window of eight amino acids, with a seven amino acid overlap relative to the preceding window. Matches (if found) are printed in the file output. IDENTITYSEARCH may be more accurate than FASTA for identification of immunologically relevant epitopes because it only requires that the specified window size (eight amino acid residues) match. The FASTA algorithm is a powerful tool for identification of sequence similarity and is more appropriately used to assess structural similarity, while a short local overlap may be ignored for a higher scoring alignment containing a longer overlap but relatively few linearly contiguous amino acid identities.

#### Results:

A FASTA search of the ALLPEPTIDES database using CP4 EPSPS L214P as the query yielded 265 alignments whose E score was lower than the default cutoff value of 10. Inspection of these alignments positively identified the query sequence as being CP4 EPSPS L214P.

Bioinformatic analysis using CP4 EPSPS L214P as the query sequence for FASTA searches of the ALLERGEN3 yielded six alignments whose E score was lower than the default cutoff value of 10. Visual inspection of these alignments indicated that they were of low quality and are unlikely to reflect conserved structure or be predictive of potential allergenicity. Consistent with the FASTA search of the ALLERGEN3 database, the allergensearch algorithm did not identify any eight contiguous amino acid window(s) in CP4 EPSPS L214P that display 100% identity with any eight amino acid window of any protein sequence in the ALLERGEN3 database.

The results obtained for searches of the ALLPEPTIDES and ALLERGEN3 databases using CP4 EPSPS L214P parallel the bioinformatic assessment that was obtained for CP4 EPSPS (Rice *et al.*, 2001).

A FASTA search of the TOXIN5 database using CP4 EPSPS L214P as the query yielded 13 alignments whose E score was lower than the default cutoff value of 10. Visual inspection of these alignments indicated that they were of low quality and are unlikely to reflect conserved structure or be predictive of potential toxicity.

References:

- Bairoch, A., and Apweiler, R. (1997). The SWISS-PROT protein sequence data bank and its supplement TrEMBL. *Nucl Acids Res* 25:31-36.
- Benson, D. A., Boguski, M. S., Lipman, D. J., and Ostell, J. (1997). GenBank. *Nucl Acids Res* 25:1-6.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000). The Protein Data Bank. *Nucl Acids Res* 28:235-242.
- Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F., Jr., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977). The Protein Data Bank. A computer-based archival file for macromolecular structures. *Eur J Biochem* 80:319-324.
- George, D. G., Dodson, R. J., Garavelli, J. S., Haft, D. H., Hunt, L. T., Marzec, C. R., Orcutt, B. C., Sidman, K. E., Srinivasarao, G. Y., Yeh, L. S. L., Arminski, L. M., Ledley, R. S., Tsugita, A., and Barker, W. C. (1997). The Protein Information Resource (PIR) and the PIR-International Protein Sequence Database. *Nucl Acids Res* 25:24-28.
- Henikoff, S., and Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* 89:10915-10919.
- Henikoff, J. G., and Henikoff, S. (1996). Blocks database and its applications. *Methods Enzymol* 266:88-105.
- Hileman, R. E., and Astwood, J. D. (1999). Bioinformatics analysis of CP4 EPSPS protein sequence utilizing an allergen database. Technical Report MSL-16267, Monsanto Company, St. Louis, MO
- Hileman, R. E., and Astwood, J. D. (1999a). Bioinformatics analysis of CP4 EPSPS protein sequence utilizing toxin and public domain genetic databases Monsanto. Technical Report MSL-16268, Monsanto Company, St. Louis MO.
- Metcalf, D. D., Astwood, J. D., Townsend, R., Sampson, H. A., Taylor, S. L., and Fuchs, R. L. (1996). Assessment of the allergenic potential of foods derived from genetically engineered crop plants. *Crit Rev Food Sci Nutr* 36:S165-186.
- Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85:2440-2448.
- Pearson, W. R. (1996). Effective protein sequence comparison. *Methods Enzymol* 266:227-258.

Pearson, W. R. (2000). Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol* 132:185-219.

Rice, E. A., Goodman, R. E., Silvanovich, A., Hileman, R. E. and Astwood, J. D. (2001). Bioinformatic analysis of the CP4 EPSPS protein Utilizing ALLERGEN3 and current Public domain sequence databases. Technical Report MSL-17172, Monsanto Company, St. Louis MO.

Schuler, G. D., Epstein, J. A., Ohkawa, H., and Kans, J. A. (1996). Entrez: Molecular biology database and retrieval system. *Methods Enzymol* 266:141-162.

Stoesser, G., Sterk, P., Tuli, M. A., Stoehr, P. J., and Cameron, G. N. (1997). The EMBL nucleotide sequence database. *Nucl Acids Res* 25:7-13.

allergensearch -INfile1=nk603cp4.pep -INFILE2=allergen3:\*\* -  
OUTfile=nk603cp4.allergensearch -WINDOWsize=8

Query sequence(s): nk603cp4.pep  
Search sequence(s): allergen3:\*\*  
Window size: 8

nk603cp4.allergensearch Tue Nov 6 09:29:42 2001 ..

*Amelia*

12/20/2001

Page 1 of 1

# SEQUENCE\_LIST 1.0

(Peptide) FASTA of: nk603cp4.pep from: 1 to: 455 November 6, 2001 09:28

REFORMAT of: nk603cp4.txt check: 7462 from: 1 to: 455 November 6, 2001 09:27  
(No documentation)

TO: allergen3:\* \* Sequences: 659 Symbols: 141,391 Word Size: 2

Databases searched:  
, Release 1.0, Released on 13Oct2000, Formatted on 13Oct2000

Scoring matrix: GenRunData:blosun50.cmp  
Variable pamfactor used  
Gap creation penalty: 12 Gap extension penalty: 2

## Histogram Key:

Each histogram symbol represents 2 search set sequences  
z-scores computed from opt scores

z-score	obs	exp
(=)	(*)	
< 20	1	0: =
22	0	0:
24	0	0:
26	0	0:
28	0	0:
30	0	1: *
32	7	4: = * =
34	12	10: = * = * =
36	15	20: = * = * = *
38	37	32: = * = * = * = * = *
40	33	45: = * = * = * = * = *
42	37	55: = * = * = * = * = *
44	80	61: = * = * = * = * = *
46	65	62: = * = * = * = * = *
48	63	59: = * = * = * = * = *
50	54	54: = * = * = * = * = *
52	47	47: = * = * = * = * = *
54	49	41: = * = * = * = * = *
56	26	34: = * = * = * = * = *
58	23	28: = * = * = * = * = *
60	26	23: = * = * = * = * = *
62	18	18: = * = * = * = * = *
64	9	14: = * = * = * = * = *
66	6	11: = * = * = * = * = *
68	14	9: = * = * = * = * = *
70	13	7: = * = * = * = * = *
72	6	5: = * = * = * = * = *
74	2	4: = * = * = * = * = *
76	8	3: = * = * = * = * = *
78	2	3: = * = * = * = * = *
80	0	2: *
82	0	2: *
84	2	1: *
86	0	1: *
88	1	1: *
90	0	1: *
92	2	0: =
94	0	0:
96	0	0:
98	1	0: =
100	0	0:
102	0	0:
104	0	0:
106	0	0:

108 0 0:  
110 0 0:  
112 0 0:  
114 0 0:  
116 0 0:  
118 0 0:  
>120 0 0:

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc	E(658) ..
AL:S70378					
! S70378 Dermatophagoides farinae Der...	36	36	79	98.2	0.77
AL:D10448					
! D10448 Dermatophagoides farinae Der...	36	36	75	93.6	1.4
AL:S63984					
! major allergen Api g 1 - Apium grav...	38	61	75	93.2	1.4
AL:D10449					
! D10449 Dermatophagoides farinae Der...	36	36	69	87.3	3.1
AL:P04389					
! P04389 aspergillus restrictus, and ...	52	52	68	84.7	4.3
AL:A46497					
! major allergen I - Aspergillus fumi...	52	52	68	84.7	4.3
\\End of List					

nk603cp4.pep  
AL:S70378

LOCUS S70378\_1  
DEFINITION Der f II=group II major allergen [Dermatophagoides farinae=mites, mRNA Partial, 517 nt]; group II major allergen; This sequence comes from Fig. 4.  
DATE 22-SEP-1994 . . .

SCORES Init1: 36 Initn: 36 Opt: 79 z-score: 98.2 E(): 0.77  
Smith-Waterman score: 79; 30.6% identity in 72 aa overlap

nk603cp4.pep	190	200	210	220	230	240																																																					
	G	I	T	T	V	I	E	P	I	M	T	R	D	H	T	E	K	M	L	Q	G	F	G	A	N	P	T	V	E	T	D	A	D	-	G	V	R	T	I	R	L	E	G	R	G	L	T	G	Q	V	I	D	V	P	G	D	P	S	S
S70378	E	I	K	V	M	V	D	G	C	H	S	D	P	C	I	I	H	R	G	K	P	F	T	L	E	A	L	F	D	A	N	Q	N	T	K	T	A	K	I	E	I	K	A	S	L	D	G	L	E	I	D	V	P	G	-	I	D	T	
	30	40	50	60	70	80																																																					

nk603cp4.pep	250	260	270	280	290	300																																																					
	T	A	F	P	L	V	A	A	L	L	V	P	G	S	D	V	T	I	-	-	-	L	N	V	-	L	M	N	P	T	R	T	G	L	I	L	T	Q	E	M	G	A	D	I	E	V	I	N	P	R	L	A	G	G	E	D	V	A	D
S70378	N	A	C	H	F	V	K	C	P	L	V	K	G	Q	Q	Y	D	I	K	Y	T	W	N	V	P	K	I	A	P	K	S	E	N	V	V	T	V	K	L	I	G	D	N	G	V	L	A	C	A	I	A	T	H	G	K	I	R	D	
	90	100	110	120	130	140																																																					

nk603cp4.pep	310	320	330	340	350	360																																																			
	L	R	V	R	S	S	T	L	K	G	V	T	P	E	D	R	A	P	S	M	I	D	E	Y	P	I	L	A	V	A	A	F	A	E	G	A	T	V	M	N	G	L	E	L	R	V	K	E	S	D	R	L	S	A	V	A	N

nk603cp4.pep  
AL:D10448

LOCUS DEPPER2\_1  
DEFINITION Dermatophagoides farinae mRNA for mite allergen Der f II precursor, complete cds, clone:pFL2.  
DATE 02-FEB-1999  
ACCESSION D10448  
NID g217305 . . .

SCORES Init1: 36 Initn: 36 Opt: 75 z-score: 93.6 E(): 1.4  
Smith-Waterman score: 75; 29.2% identity in 72 aa overlap





!!SEQUENCE\_LIST 1.0

(Peptide) FASTA of: nk603cp4.pep from: 1 to: 455 November 6, 2001 09:29

REFORMAT of: nk603cp4.txt check: 7462 from: 1 to: 455 November 6, 2001 09:27  
(No documentation)

TO: toxin5: \* Sequences: 12,771 Symbols: 3,906,811 Word Size: 2

Databases searched:

, Release 1.0, Released on 28Aug2001, Formatted on 28Aug2001

Scoring matrix: GenRunData:blosun50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

#### Histogram Key:

Each histogram symbol represents 21 search set sequences

Each inset symbol represents 1 search set sequences

z-scores computed from opt scores

z-score obs exp  
(=) (\*)

```
< 20 23 0:==
22 0 0:
24 0 0:
26 0 0:
28 4 3:*
30 13 18:*
32 54 68:==*
34 157 184:==*
36 312 378:==*
38 621 624:==*
40 1135 871:==*
42 1119 1065:==*
44 1238 1174:==*
46 1214 1196:==*
48 975 1145:==*
50 950 1045:==*
52 884 919:==*
54 711 785:==*
56 619 656:==*
58 519 538:==*
60 495 436:==*
62 385 350:==*
64 245 278:==*
66 249 220:==*
68 188 173:==*
70 152 135:==*
72 107 106:==*
74 94 83:==*
76 67 64:==*
78 69 50:==*
80 41 39:==*
82 20 30:==*
84 24 23:==*
86 19 18:*
88 23 14:*
90 10 11:*
92 4 8:*
94 5 7:*
96 5 5:*
98 5 4:*
100 5 3:*
102 2 2:*
104 0 2:*
```

```
106 1 1:*
108 5 1:*
110 2 1:*
112 0 1:*
114 0 1:*
116 0 0:
118 0 0:
>120 1 0:==
```

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

	init1	initn	opt	z-sc	E(12735)...
TXN5:PHL2_BACCE					
! P11889 bacillus cereus. sphingomyel...	43	67	108	129.6	0.26
TXN5:AF297061_16					
! AF297061 Escherichia coli Escherich...	44	44	83	111.2	2.8
TXN5:Q9EZD8					
! Q9ezd8 escherichia coli. hypothetical...	44	44	83	111.2	2.8
TXN5:Q9S5W5					
! Q9s5w5 escherichia coli. udp-n-acet...	45	68	91	108.1	4.1
TXN5:BAB37491					
! Bab37491 escherichia coli o157:h7. ...	45	68	91	108.1	4.1
TXN5:MURA_ECOLI					
! P28909 escherichia coli. udp-n-acet...	45	68	91	108.1	4.1
TXN5:BAB34019					
! Bab34019 escherichia coli o157:h7. ...	31	31	85	108.0	4.2
TXN5:BAB38133					
! Bab38133 escherichia coli o157:h7. ...	75	75	84	107.9	4.2
TXN5:BAB36642					
! Bab36642 escherichia coli o157:h7. ...	38	38	84	107.1	4.7
TXN5:BAB37530					
! Bab37530 escherichia coli o157:h7. ...	41	41	87	102.9	8.1
TXN5:BAB35429					
! Bab35429 escherichia coli o157:h7. ...	51	75	92	102.6	8.4
TXN5:P74994					
! P74994 yersinia pestis, and yersini...	44	44	82	101.4	9.8
TXN5:Q9R3L5					
! Q9r3l5 yersinia pestis, and escheri...	44	44	82	101.4	9.9
\\End of List					

nk603cp4.pep  
TXN5:PHL2\_BACCE

Description: P11889 bacillus cereus. sphingomyelinase c precursor (ec 3.1.4.12) (sphingomyeli

Accession/ID: P11889

====General comments=====

ID PHL2\_BACCE STANDARD; PRT; 333 AA.  
AC P11889; . . .

SCORES Init1: 43 Initn: 67 Opt: 108 z-score: 129.6 E(): 0.26  
Smith-Waterman score: 108; 28.2% identity in 131 aa overlap

```
230 240 250 260 270 280
nk603cp4.pep RTIRLEGRGKLTGQVIDVPGDPSTAFPLVAALLVPGSDVTILN-VLMNPTRTGLILTLQ
PHL2_BACCE TLKVMTHNVYMLSTNLYPNWQTERADLIGAADYIKNQDVVILNEVPDNDASDRLGLNLK
40 50 60 70 80 90

290 300 310 320 330 340
nk603cp4.pep EMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDIEYPILA-VAAAFAGA
PHL2_BACCE KEYPNQTVAILGRSSGSEWDKTLGNYSSS-----TPEDGGVAIVSKWPIAEKIYVFAKGC
100 110 120 130 140 150
```

```
350 360 370 380 390
nk603cp4.pep TVMN-----GLEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAA
```

Monsanto Product Safety Center

Date 11/06/01

Study # - - - -

Page 1 of -

PHL2\_BACCE | : : : : : | : : : | : : |  
GPDNLSNKGfVYTKIKKNDRFVHVI-GTHLQAEDSMCGKTS PASVRTNQLKEIQDFIKNK  
160 170 180 190 200

nk603cp4.pep 400 410 420 430 440 450  
VATHLDHRIAMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA

PHL2\_BACCE NIPNNEYVLIGGDMVNKINAENNNDSEYASMFKTLNASVPSYTGHTATWDATTNSIAKY  
210 220 230 240 250 260

nk603cp4.pep  
TXN5:AF297061\_16

Description: AF297061 Escherichia coli Escherichia coli enterotoxin EspC  
(espC) gene, complet  
Accession/ID: AF297061

====General comments=====

LOCUS AF297061\_16 [AF297061]

DEFINITION Escherichia coli enterotoxin EspC (espC) gene, complete cds; and . . .

SCORES Initl: 44 Initn: 44 Opt: 83 z-score: 111.2 E(): 2.8  
Smith-Waterman score: 83; 33.8% identity in 74 aa overlap

50 60 70 80 90 100  
nk603cp4.pep TGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGN-AATGCRLT  
AF297061\_16 ||| : : : : : ||| : : : : :  
MTPRLLIIDEIGYLPFSQEEAKLYFQVIKRYEKSA  
10 20 30

110 120 130 140 150 160  
nk603cp4.pep MGLVGVDYF---DSTFIGDASLTKRPMGRVLNPLREMGVQKSEGDRLPVTLRGPKTPT  
AF297061\_16 ||| : : : : : ||| : : : : :  
MILTSNLSFGQWDQTFAGDAALTSAMLDRIH--YSHVIQIKGESYRLR  
40 50 60 70 80

170 180 190 200 210 220  
nk603cp4.pep PITVRVPMASAVKSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGV

nk603cp4.pep  
TXN5:Q9EZD8

Description: Q9ezd8 escherichia coli. hypothetical 9.6 kda protein. 3/2001  
Accession/ID: Q9EZD8

====General comments=====

ID Q9EZD8 PRELIMINARY; PRT; 83 AA.

AC Q9EZD8;

DT 01-MAR-2001 (TrEMBLrel. 16, Created) . . .

SCORES Initl: 44 Initn: 44 Opt: 83 z-score: 111.2 E(): 2.8  
Smith-Waterman score: 83; 33.8% identity in 74 aa overlap

50 60 70 80 90 100  
nk603cp4.pep TGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGN-AATGCRLT  
Q9EZD8 ||| : : : : : ||| : : : : :  
MTPRLLIIDEIGYLPFSQEEAKLYFQVIKRYEKSA  
10 20 30

110 120 130 140 150 160  
nk603cp4.pep MGLVGVDYF---DSTFIGDASLTKRPMGRVLNPLREMGVQKSEGDRLPVTLRGPKTPT  
Q9EZD8 ||| : : : : : ||| : : : : :  
MILTSNLSFGQWDQTFAGDAALTSAMLDRIH--YSHVIQIKGESYRLR  
40 50 60 70 80

170 180 190 200 210 220  
nk603cp4.pep PITVRVPMASAVKSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGV

nk603cp4.pep  
TXN5:Q9S5W5

Description: Q9s5w5 escherichia coli. udp-n-acetylglucosamine enolpyruvoyl  
transferase. 6/200  
Accession/ID: Q9S5W5

====General comments=====

ID Q9S5W5 PRELIMINARY; PRT; 419 AA.  
AC Q9S5W5; . . .

SCORES Initl: 45 Initn: 68 Opt: 91 z-score: 108.1 E(): 4.1  
Smith-Waterman score: 234; 25.1% identity in 442 aa overlap

10 20 30 40 50 60  
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
Q9S5W5 ||| : : : : : ||| : : : : :  
MDKFRVQGPTELQGEVTISGAKNAALPILFAALLAEFPVEIQNVPLKDVDTSM  
10 20 30 40 50

70 80 90 100 110  
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMG-LVGVDYFDFDST  
Q9S5W5 ||| : : : : : ||| : : : : :  
KLLSQLGAKVERNGSVHI-DARDVNVPFCAP---YDLVKTRASIWALGPLVARFQGQGV  
60 70 80 90 100 110

120 130 140 150 160 170  
nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQKSEGDRLPVTLRGP-KTPTPTITVRVPM-ASAQV  
Q9S5W5 ||| : : : : : ||| : : : : :  
LPGGCTIGARFPVDLHISGLGATIKLEEG-YVKASVDGRKGAHIVMDKVSVGATVTI  
120 130 140 150 160

180 190 200 210 220 230  
nk603cp4.pep KSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTG  
Q9S5W5 ||| : : : : : ||| : : : : :  
MCAATLAEGTTIENAAAREPEIV--DTANFLITLGA---KISGQGTDRIVIEGVERLGG  
170 180 190 200 210 220

240 250 260 270 280 290  
nk603cp4.pep QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT-LQEMGADIEVINPRL  
Q9S5W5 ||| : : : : : ||| : : : : :  
GVYRVLDPDRIETGTFVAAAISRGK---IICRNAQPDTLDAVLAKLRDAGADIEV-----  
230 240 250 260 270

300 310 320 330 340 349  
nk603cp4.pep AGGEDVADLRVRSSTLKGVTV---PEDRAP-SMIDEYPIAVAAAPAGATVMNGLEE--  
Q9S5W5 ||| : : : : : ||| : : : : :  
--GEDWISLDMHGKREKAVNVRTAPHPAFPTDMQAQFTLLNLVAE-GTGFITETVFENRF  
280 290 300 310 320 330

350 360 370 380 390 400 409  
nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGPRPDGKGLGNASGAAVATHLDHRIAMS  
Q9S5W5 ||| : : : : : ||| : : : : :  
MHVPELSRMGAHAE-IESNTVIC-----HGVEKLSGAQVMA-TDLRASAS  
340 350 360 370

410 420 430 440 450  
nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Q9S5W5 ||| : : : : : ||| : : : : :  
LVLACGIAEGTPTVDRIYHIDRGYERIEDKLRLGANIERVKGE  
380 390 400 410

nk603cp4.pep  
TXN5:BAB37491

Description: Bab37491 escherichia coli o157:h7. udp-n-glucosamine  
1-carboxyvinyltransferase.  
Accession/ID: 000000

====General comments=====

ID BAB37491 PRELIMINARY; PRT; 419 AA.

AC BAB37491; . . .

SCORES Init1: 45 Initn: 68 Opt: 91 z-score: 108.1 E(): 4.1  
Smith-Waterman score: 234; 25.1% identity in 442 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
BAB37491 MDKFRVQGPTKLQGEVTISGAKNAALPILFAALLAEFVEIQNVPKLKDVTSM

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMG-LVGVY-DFDST  
BAB37491 KLLSQLGAKVERNGSVHI-DARDVNVEFCAP---YDLVKTMRASIWALGPLVARFGQQQVS

nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGP-KTPTPIYRVPM-ASAQV  
BAB37491 LPGGCTIGARPVDLHISGLEQLGATIKLEEG-YVKASVDRLKGAHIVMDKVSVGATVTI

nk603cp4.pep KSAVLLAGLNPFGITTVIEPIMTRDHEKMLQGFANPTVETDAGVVRTIRLEGRGKLTG  
BAB37491 MCAATLAEGTTIENAREPEIV--DTANFLITLGA---KISGQGTDRIVEGVERLGG

nk603cp4.pep QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMPNTRTGLILT-LQEMGADIEVINPRL  
BAB37491 GVYRVLPDRIETGTFLVAAISRKG---ICRNTQPDTLDAVLAKLRDAGADIEV----

nk603cp4.pep AGGEDVADLRVRSSTLKGVTV---PEDRAPs-MIDEYPILAVAAFAEGATVMNGLEE--  
BAB37491 --GEDWISLDMHGKRPKAVNVRTAPHPAFPTDMQAQFTLLNLVAB-GTGFITETVFENRF

nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHRIAMS  
BAB37491 MHVPELSRMGAHAE-IESNTVIC-----HGVEKLSGAQVMA-TDLRASAS

nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
BAB37491 LVLACGIAEGTTVVDRYIHIDRGYERIEDKLRLGANIERVKGE

nk603cp4.pep  
TXN5:MURA\_ECOLI

Description: P28909 escherichia coli. udp-n-acetylglucosamine  
1-carboxyvinyltransferase (ec 2  
Accession/ID: P28909

=====  
ID MURA\_ECOLI STANDARD; PRT; 419 AA.  
AC P28909; . . .

SCORES Init1: 45 Initn: 68 Opt: 91 z-score: 108.1 E(): 4.1  
Smith-Waterman score: 234; 25.1% identity in 442 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG

MURA\_ECOLI MDKFRVQGPTKLQGEVTISGAKNAALPILFAALLAEFVEIQNVPKLKDVTSM

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMG-LVGVY-DFDST  
MURA\_ECOLI KLLSQLGAKVERNGSVHI-DARDVNVEFCAP---YDLVKTMRASIWALGPLVARFGQQQVS

nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGP-KTPTPIYRVPM-ASAQV  
MURA\_ECOLI LPGGCTIGARPVDLHISGLEQLGATIKLEEG-YVKASVDRLKGAHIVMDKVSVGATVTI

nk603cp4.pep KSAVLLAGLNPFGITTVIEPIMTRDHEKMLQGFANPTVETDAGVVRTIRLEGRGKLTG  
MURA\_ECOLI MCAATLAEGTTIENAREPEIV--DTANFLITLGA---KISGQGTDRIVEGVERLGG

nk603cp4.pep QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMPNTRTGLILT-LQEMGADIEVINPRL  
MURA\_ECOLI GVYRVLPDRIETGTFLVAAISRKG---ICRNTQPDTLDAVLAKLRDAGADIEV----

nk603cp4.pep AGGEDVADLRVRSSTLKGVTV---PEDRAPs-MIDEYPILAVAAFAEGATVMNGLEE--  
MURA\_ECOLI --GEDWISLDMHGKRPKAVNVRTAPHPAFPTDMQAQFTLLNLVAB-GTGFITETVFENRF

nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHRIAMS  
MURA\_ECOLI MHVPELSRMGAHAE-IESNTVIC-----HGVEKLSGAQVMA-TDLRASAS

nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
MURA\_ECOLI LVLACGIAEGTTVVDRYIHIDRGYERIEDKLRLGANIERVKGE

nk603cp4.pep  
TXN5: BAB34019

Description: Bab34019 escherichia coli o157:h7. putative fimbrial-like  
protein. 6/2001  
Accession/ID: 000000

=====  
ID BAB34019 PRELIMINARY; PRT; 171 AA.  
AC BAB34019; . . .

SCORES Init1: 31 Initn: 31 Opt: 85 z-score: 108.0 E(): 4.2  
Smith-Waterman score: 85; 28.0% identity in 125 aa overlap

nk603cp4.pep DTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVG-VYDFDSTFIGDASLTKR-PMGR  
BAB34019 MRRVFFSCFCGLLWSSSGWAADPLGTININLHGNVVDVFSCT-VNTADIDKTVDLGR

nk603cp4.pep VLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTIYRVPMASAVLLAGLNPFGIT  
BAB34019 W--PTTQL---LNAGDTTALLFFSLRLEGCP-PGSVAILFTGTPASDANLLA-LDDPAMA

60 70 80 90 100  
 nk603cp4.pep TVIEPI MTRDHT ETKMLQGF GANPTVET DADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFF  
 BAB34019 QTVA-IELRNSDRSRLALGEASPT EVDANGNVTLNFFANYRALASGVVRPGVAKADAIFM  
 110 120 130 140 150 160

260 270 280 290 300 310  
 nk603cp4.pep LVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLAGGEDVADLRVRSSTL  
 BAB34019 INYN  
 170

nk603cp4.pep  
 TXN5: BAB38133

Description: Bab38133 escherichia coli o157:h7. hypothetical 17.0 kda protein.  
 6/2001

Accession/ID: 000000

====General comments=====

ID BAB38133 PRELIMINARY; PRT; 149 AA.  
 AC BAB38133; . . .

SCORES Init1: 75 Initn: 75 Opt: 84 z-score: 107.9 E(): 4.2  
 Smith-Waterman score: 84; 31.0% identity in 58 aa overlap

130 140 150 160 170 180  
 nk603cp4.pep ASLTKRPMGRVLNPLREMGVQVKS EGDRLPVTLRGPKTPTPTYRVFMSAQVKS AVLL  
 BAB38133 MAINFSPKVGEILECNFGNYPVFSQNGPFSTTYDGRIPPEMIKNRLVVVL  
 10 20 30 40 50

190 200 210 220 230 240  
 nk603cp4.pep AGLNTPGTTTIEPI-MTRDHT ETKMLQGF GANPTVET DADGVRTIRLEGRGKLTGQVIDV  
 BAB38133 NG-KINGNAFIVVPLSTTRDH-DKLRGMHVEIASNVINDLQFFDQQIRWAKTDLVQVQS  
 60 70 80 90 100

nk603cp4.pep  
 TXN5: BAB36642

Description: Bab36642 escherichia coli o157:h7. putative fimbrial protein.  
 6/2001

Accession/ID: 000000

====General comments=====

ID BAB36642 PRELIMINARY; PRT; 166 AA.  
 AC BAB36642; . . .

SCORES Init1: 38 Initn: 38 Opt: 84 z-score: 107.1 E(): 4.7  
 Smith-Waterman score: 88; 30.8% identity in 104 aa overlap

10 20 30  
 nk603cp4.pep MLHGASSR PATARKSSGLSGTVRIPGDKSISHR-----  
 BAB36642 LLLCSIVSTFVMAGSKSVAMVLRVLVDAPPCTVTGASVEFGNVFISKIDGVSYKRPIDY  
 20 30 40 50 60 70

40 50 60 70 80  
 nk603cp4.pep SFMFGLGLASGETRI-----TGLLEGEDVINTEGKAMQAMGARIRKEGDTWIIDGVNGGGLL  
 BAB36642 SLVCNNLAMDDLRLNMQATTVINGETVIDTGIA--GFGIRIQVSDHSILD-----L  
 80 90 100 110 120

90 100 110 120 130 140  
 nk603cp4.pep APEA--PLDFGNAATGCRLTMGLVGVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKS  
 :| | :| :| :|

BAB36642 TPGAWLPFNFFSSGALALEAVPVVQSGVSLTAAEFSASATIVVDYQ  
 130 140 150 160

nk603cp4.pep  
 TXN5: BAB37530

Description: Bab37530 escherichia coli o157:h7. serine endoprotease. 6/2001  
 Accession/ID: 000000

====General comments=====

ID BAB37530 PRELIMINARY; PRT; 455 AA.  
 AC BAB37530;  
 DT 14-JUN-2001 (EMBLrel. 63, Created) . . .

SCORES Init1: 41 Initn: 41 Opt: 87 z-score: 102.9 E(): 8.1  
 Smith-Waterman score: 87; 20.9% identity in 172 aa overlap

50 60 70 80 90 100  
 nk603cp4.pep LLEGEDVINTEGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAAT---GCRLT  
 BAB37530 EGTASQGGKPIPEEFKFFGDDLDPQAPQPFEGLSGVIIINASKGYVLTNHVINQAQKIS  
 60 70 80 90 100 110

110 120 130 140 150 160  
 nk603cp4.pep MGLVGVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKS EGDRLPVTLRGPKTPTPTIT  
 BAB37530 IQLENDGREFDAKLIGSDDQSDIALLIQNP SKL--TQIAIADSKLRVGFVAVGNNPFG  
 120 130 140 150 160 170

170 180 190 200 210 220  
 nk603cp4.pep YRVFMSAQVKS AVLLAGLNTPGITTVIEPI MTRDHT ETKMLQGF GANPTVET DADGVRTI  
 BAB37530 LG-QTATSGIVSALGRSGLNLEGLNF IQ---TDASINRGNSG-GALLNLNGLIGLIGINTA  
 180 190 200 210 220 230

230 240 250 260 270 280  
 nk603cp4.pep RLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGA  
 BAB37530 ILAPGGGSGVIGFAIPSNMARTLAQQLIDFGEIKRGLLGKGTMSADI AKAFNLDVQRG  
 240 250 260 270 280 290

nk603cp4.pep  
 TXN5: BAB35429

Description: Bab35429 escherichia coli o157:h7. putative biga-like protein.  
 6/2001

Accession/ID: 000000

====General comments=====

ID BAB35429 PRELIMINARY; PRT; 1011 AA.  
 AC BAB35429; . . .

SCORES Init1: 51 Initn: 75 Opt: 92 z-score: 102.6 E(): 8.4  
 Smith-Waterman score: 92; 24.8% identity in 129 aa overlap

20 30 40 50 60  
 nk603cp4.pep TARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDV-----INTGK  
 BAB35429 DQGGNTALNGGTVMTVDGDNISLNNDGKTTAIGEGSVVGILTGDNITINNNGETEVDGGT  
 280 290 300 310 320 330

70 80 90 100 110 119  
 nk603cp4.pep AMQAMG--ARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 BAB35429 AVIINGDNTKLN TAGDSTITNG-GTGS LINGD-----NARVDNQGTMSVDGENSTGSKI  
 340 350 360 370 380 390

120 130 140 150 160 170 179  
 nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKS EGDRLPVTLRGPKTPTPTIT YRVFMSAQVKS A

BAB35429 :||: |: |: ||: |: :|  
VG DGATIKQE-GDLYVSGGAHGIDV---DGNDTFVSNKGNITVIEDNSIGMLLDGPGVSV  
400 410 420 430 440

180 190 200 210 220 230 239  
nk603cp4.pep VLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI

BAB35429 INMGDLNVGQAAAGENAIGIQIDGDNATFVNVDISATNAGTGVSVAGDKANISLAGGLD  
450 460 470 480 490 500

nk603cp4.pep  
TXN5:P74994

Description: P74994 yersinia pestis, and yersinia pseudotuberculosis.  
hypothetical 29.3 kda p

Accession/ID: P74994

====General comments=====

ID P74994 PRELIMINARY; PRT; 259 AA.  
AC P74994; . . .

SCORES Init1: 44 Initn: 44 Opt: 82 z-score: 101.4 E(): 9.8  
Smith-Waterman score: 82; 35.1% identity in 74 aa overlap

50 60 70 80 90 100  
nk603cp4.pep TGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGN-AATGCRLT  
P74994 ADLLLQLSTAQRQGRYKTTLQRGVMAPRLLIIDEIGYLPFSQEEAKLFFQVIKRYEKSA  
140 150 160 170 180 190

110 120 130 140 150 160  
nk603cp4.pep MGLVGVYDF---DSTFIGDASLTKRPMGRVNLPLREMGVQVKSEGDRLPVTLRGPKTPT  
P74994 MILTSNLPFGQWDQTFAGDAALTSAMLDRIHLH--HSHVVQIKGESYRLRQKRKAGVIAEA  
200 210 220 230 240 250

170 180 190 200 210 220  
nk603cp4.pep PITYRVPMASAVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGV  
P74994 NPE

nk603cp4.pep  
TXN5:Q9R3L5

Description: Q9r315 yersinia pestis, and escherichia coli. putative is100  
transposase. 6/2001

Accession/ID: Q9R3L5

====General comments=====

ID Q9R3L5 PRELIMINARY; PRT; 260 AA.  
AC Q9R3L5; . . .

SCORES Init1: 44 Initn: 44 Opt: 82 z-score: 101.4 E(): 9.9  
Smith-Waterman score: 82; 35.1% identity in 74 aa overlap

50 60 70 80 90 100  
nk603cp4.pep TGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGN-AATGCRLT  
Q9R3L5 ADLLLQLSTAQRQGRYKTTLQRGVMAPRLLIIDEIGYLPFSQEEAKLFFQVIKRYEKSA  
140 150 160 170 180 190

110 120 130 140 150 160  
nk603cp4.pep MGLVGVYDF---DSTFIGDASLTKRPMGRVNLPLREMGVQVKSEGDRLPVTLRGPKTPT  
Q9R3L5 MILTSNLPFGQWDQTFAGDAALTSAMLDRIHLH--HSHVVQIKGESYRLRQKRKAGVIAEA  
200 210 220 230 240 250

170 180 190 200 210 220

nk603cp4.pep PITYRVPMASAVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGV

Q9R3L5 NPE  
260

! Distributed over 1 thread.

! Start time: Tue Nov 6 09:29:03 2001

! Completion time: Tue Nov 6 09:29:12 2001

! CPU time used:

! Database scan: 0:00:07.7

! Post-scan processing: 0:00:00.9

! Total CPU time: 0:00:08.7

! Output File: nk603cp4\_toxin5

# !!SEQUENCE\_LIST 1.0

(Peptide) FASTA of: nk603cp4.pep from: 1 to: 455 November 6, 2001 17:10

REFORMAT of: nk603cp4.txt check: 7462 from: 1 to: 455 November 6, 2001

09:27

(No documentation)

TO: allpeptides: \* Sequences: 1,475,823 Symbols: 461,001,819 Word Size: 2

Sequences skipped due to type mismatch with query: 11501

Databases searched:

SWISS-PROT, Release 1.0, Released on 3Nov2001, Formatted on 3Nov2001

GenBank, Release 124.0, Released on 15Jun2001, Formatted on 10Jul2001

Scoring matrix: GenRunData:blosum50.cmp

Variable pamfactor used

Gap creation penalty: 12 Gap extension penalty: 2

## Histogram Key:

Each histogram symbol represents 2242 search set sequences

Each inset symbol represents 19 search set sequences

z-scores computed from opt scores

z-score	obs	exp	
(=)	(*)		
< 20	4580	0:===	
22	12	0:=	
24	42	1:*	
26	99	31:*	
28	428	335:*	
30	2009	2033:*	
32	6967	7859:***	
34	19778	21313:*****	
36	43159	43773:*****	
38	76220	72340:*****	
40	110780	100908:*****	
42	130213	123347:*****	
44	134479	136064:*****	
46	133700	138584:*****	
48	132889	132678:*****	
50	121275	121070:*****	
52	105403	106440:*****	
54	92658	90919:*****	
56	72907	75945:*****	
58	61546	62349:*****	
60	49172	50507:*****	
62	39576	40491:*****	
64	30780	32203:*****	
66	24366	25452:*****	
68	18938	20020:*****	
70	14361	15689:*****	
72	10820	12259:*****	
74	8921	9558:*****	
76	6443	7439:*****	
78	5242	5782:*****	
80	3716	4490:*****	
82	3096	3435:*****	
84	2351	2721:*****	
86	1703	2105:*****	
88	1495	1629:*****	
90	1072	1260:*****	
92	912	975:*****	
94	674	755:*****	
96	589	584:*****	
98	474	452:*****	
100	410	350:*****	

102	345	271:*	*****
104	271	209:*	*****
106	137	162:*	*****
108	114	125:*	*****
110	101	97:*	*****
112	86	75:*	*****
114	63	58:*	*****
116	45	45:*	*****
118	38	35:*	*****
>120	368	27:*	*****

Joining threshold: 37, opt. threshold: 25, opt. width: 16, reg.-scaled

The best scores are:

	initl	initn	opt	z-sc	E(1475339)...
SWISSPROT:AROA_AGRSP					
! Q9r4e4 agrobacterium sp. (strain cp...	2896	2896	2900	3136.1	1.1e-166
TREMBL_NEW:CAC41690					
! CAC41690 rhizobium meliloti (sinorh...	2672	2672	2676	2894.2	3.1e-153
SWISSPROT:AROA_PSES2					
! P56952 pseudomonas sp. (strain pg29...	2127	2413	2433	2631.9	1.3e-138
TREMBL_MAIN:Q9AGV2					
! Q9agv2 brucella abortus. 5-enolpyru...	2077	2353	2380	2574.2	2.1e-135
GP_BCT2:AF326475_1					
! AF326475 Brucella melitensis biovar...	2077	2353	2380	2574.2	2.1e-135
GP_BCT2:AP003006_13					
! AP003006 Mesorhizobium loti Mesorhi...	1300	2247	2289	2476.4	5.9e-130
TREMBL_MAIN:Q98CC1					
! Q98cc1 rhizobium loti (mesorhizobiu...	1300	2247	2289	2476.4	5.9e-130
GP_BCT1:AE006017_6					
! AE006017 Caulobacter crescentus Cau...	611	1423	1473	1595.4	6.9e-81
TREMBL_MAIN:Q9A2H2					
! Q9a2h2 caulobacter crescentus. 3-ph...	611	1423	1473	1595.4	6.9e-81
TREMBL_MAIN:Q9HZ69					
! Q9hz69 pseudomonas aeruginosa. stil...	472	1221	1279	1382.5	5e-69
GP_BCT1:AE004740_5					
! AE004740 Pseudomonas aeruginosa Pse...	472	1221	1279	1382.5	5e-69
GP_BCT1:AE004043_7					
! AE004043 Xylella fastidiosa 9a5c Xy...	474	1124	1193	1293.0	4.8e-64
TREMBL_MAIN:Q9PB21					
! Q9pb21 xylella fastidiosa. 3-phosph...	474	1124	1193	1293.0	4.8e-64
GP_BCT3:D90914_87					
! D90914 Synechocystis sp. Synechocys...	480	967	1140	1235.8	7.4e-61
GP_BCT3:X75325_1					
! X75325 Synechocystis sp. Synechocys...	480	967	1140	1235.8	7.4e-61
SWISSPROT:AROA_SYNY3					
! Q59975 synechocystis sp. (strain pc...	480	967	1140	1235.8	7.4e-61
GP_BCT3:Z29339_1					
! Z29339 Dichelobacter nodosus D.nodo...	442	961	1056	1145.2	8.2e-56
SWISSPROT:AROA_BACNO					
! Q46550 bacteroides nodosus (dichelo...	442	961	1056	1145.2	8.2e-56
TREMBL_MAIN:Q9KCA6					
! Q9kca6 bacillus halodurans. 3-phosp...	560	1049	1030	1117.3	2.9e-54
GP_BCT2:AP001512_254					
! AP001512 Bacillus halodurans Bacill...	560	1049	1030	1117.3	2.9e-54
SWISSPROT:AROA_AQUAE					
! O67494 aquifex aeolicus. 3-phosphos...	449	561	1011	1096.8	4.1e-53
GP_BCT1:AE000744_15					
! AE000744 Aquifex aeolicus Aquifex a...	449	561	1011	1096.8	4.1e-53
TREMBL_NEW:CAC97267					
! CAC97267 listeria innocua. aroe pro...	436	962	1008	1093.6	6.2e-53
TREMBL_MAIN:Q99Z83					
! Q99z83 streptococcus pyogenes. puta...	435	951	1002	1087.1	1.4e-52
GP_BCT1:AE006573_2					
! AE006573 Streptococcus pyogenes M1 ...	435	951	1002	1087.1	1.4e-52
TREMBL_NEW:CAD00001					
! CAD00001 listeria monocytogenes. ar...	424	940	993	1077.4	4.9e-52
SWISSPROT:AROA_STREN					
! Q9s400 streptococcus pneumoniae. 3-...	386	889	963	1045.0	3.1e-50
GP_BCT2:AF169483_1					

AF169483 Streptococcus pneumoniae S... 385 891 962 1043.9 3.6e-50  
 GP\_BCT1:AE006404\_9  
 AE006404 Lactococcus lactis subsp. ... 412 936 962 1043.9 3.6e-50  
 SWISSPROT:AROA\_LACLA  
 Q9ceu0 lactococcus lactis (subsp. l... 412 936 962 1043.9 3.6e-50  
 TREMBL\_NEW:AAL00033  
 Aa100033 streptococcus pneumoniae r... 386 897 959 1040.6 5.5e-50  
 GP\_BCT3:X78413\_2  
 X78413 Lactococcus lactis L.lactis ... 400 964 952 1033.1 1.4e-49  
 SWISSPROT:AROA\_LACLC  
 P43905 lactococcus lactis (subsp. c... 400 964 952 1033.1 1.4e-49  
 TREMBL\_MAIN:Q9ANY6  
 Q9any6 enterococcus faecalis (strep... 376 885 928 1007.2 4e-48  
 GP\_BCT2:AF318277\_6  
 AF318277 Enterococcus faecalis Ente... 376 885 928 1007.2 4e-48  
 GP\_BCT2:Z99115\_200  
 Z99115 Bacillus subtilis Bacillus s... 525 943 926 1005.1 5.3e-48  
 GP\_BCT2:M80245\_18  
 M80245 Bacillus subtilis B.subtilis... 525 943 926 1005.1 5.3e-48  
 SWISSPROT:AROA\_BACSV  
 P20691 bacillus subtilis. 3-phospho... 525 943 926 1005.1 5.3e-48  
 GP\_BCT1:AE001715\_8  
 AE001715 Thermotoga maritima Thermo... 326 788 834 906.0 1.7e-42  
 SWISSPROT:AROA\_THEMA  
 Q9wyi0 thermotoga maritima. 3-phosp... 326 788 834 906.0 1.7e-42  
 TREMBL\_NEW:BAB57626  
 Bab57626 staphylococcus aureus subs... 271 666 806 875.4 8.7e-41  
 GP\_BCT2:AP003362\_152  
 AP003362 Staphylococcus aureus subs... 271 666 806 875.4 8.7e-41  
 TREMBL\_MAIN:Q99U25  
 Q99u25 staphylococcus aureus subsp... 271 666 806 875.4 8.7e-41  
 GP\_BCT2:AP003134\_5  
 AP003134 Staphylococcus aureus subs... 271 666 806 875.4 8.7e-41  
 GP\_BCT3:L05004\_2  
 L05004 Staphylococcus aureus Staphy... 271 526 765 831.2 2.5e-38  
 SWISSPROT:AROA\_STAAU  
 Q05615 staphylococcus aureus. 3-pho... 271 526 765 831.2 2.5e-38  
 SWISSPROT:AROA\_HELFI  
 Q9zkf7 helicobacter pylori j99 (cam... 220 571 708 769.7 6.8e-35  
 GP\_BCT1:AE001527\_6  
 AE001527 Helicobacter pylori J99 He... 220 571 708 769.7 6.8e-35  
 GP\_BCT1:AE000556\_2  
 AE000556 Helicobacter pylori 26695 ... 218 568 701 762.1 1.8e-34  
 SWISSPROT:AROA\_HELFI  
 P56197 helicobacter pylori (campylo... 218 568 701 762.1 1.8e-34  
 GP\_BCT2:AL139076\_205  
 AL139076 Campylobacter jejuni Campy... 199 675 654 711.4 1.2e-31  
 SWISSPROT:AROA\_CAMJE  
 P52312 campylobacter jejuni. 3-phos... 199 675 654 711.4 1.2e-31  
 GP\_BCT2:X89371\_2  
 X89371 Campylobacter jejuni C.jejun... 198 671 644 700.6 4.8e-31  
 TREMBL\_MAIN:Q9RHZ8  
 Q9rhz8 pseudomonas stutzeri (pseudo... 419 579 622 681.6 5.5e-30  
 GP\_BCT1:AF038578\_6  
 AF038578 Pseudomonas stutzeri Pseud... 419 579 622 681.6 5.5e-30  
 SWISSPROT:AROA\_HAESO  
 P52310 haemophilus somnus. 3-phosph... 133 304 461 502.9 4.9e-20  
 GP\_BCT3:L47538\_1  
 L47538 Haemophilus somnus Haemophil... 133 304 461 502.9 4.9e-20  
 GP\_BCT3:U89948\_1  
 U89948 Mannheimia haemolytica Paste... 141 299 443 483.5 5.9e-19  
 SWISSPROT:AROA\_PASHA  
 P54220 pasteurella haemolytica. 3-p... 141 299 443 483.5 5.9e-19  
 GP\_BCT3:U03068\_1  
 U03068 Mannheimia haemolytica Paste... 141 299 438 478.0 1.2e-18  
 GP\_BCT3:L04686\_2  
 L04686 Haemophilus influenzae Haemo... 170 312 437 477.0 1.4e-18  
 SWISSPROT:AROA\_YERPE  
 Q60112 yersinia pestis. 3-phosphosh... 176 359 430 469.6 3.5e-18  
 TREMBL\_NEW:CAC90219  
 Cac90219 yersinia pestis. 3-phospho... 176 359 430 469.5 3.6e-18

GP\_BCT1:AE004251\_5  
 AE004251 Vibrio cholerae Vibrio cho... 162 444 425 464.1 7.1e-18  
 TREMBL\_MAIN:Q9KRBO  
 Q9krb0 vibrio cholerae. 3-phosphosh... 162 444 425 464.1 7.1e-18  
 SWISSPROT:AROA\_YEREN  
 P19688 yersinia enterocolitica. 3-p... 185 373 422 460.9 1.1e-17  
 GP\_BCT3:X82415\_1  
 X82415 Klebsiella pneumoniae K.pneu... 151 403 420 458.7 1.4e-17  
 SWISSPROT:AROA\_KLEPN  
 P24497 klebsiella pneumoniae. 3-pho... 151 403 420 458.7 1.4e-17  
 SWISSPROT:AROA\_HAEIN  
 Q03421 haemophilus influenzae. 3-ph... 170 315 420 458.6 1.4e-17  
 GP\_BCT3:U32833\_2  
 U32833 Haemophilus influenzae Rd Ha... 170 315 420 458.6 1.4e-17  
 SWISSPROT:AROA\_BURPS  
 P39915 burkholderia pseudomallei (p... 178 367 414 452.5 3.1e-17  
 GP\_BCT3:X77019\_1  
 X77019 Burkholderia pseudomallei P... 178 367 414 452.5 3.1e-17  
 TREMBL\_NEW:AAK97382  
 Aak97382 yersinia ruckeri. 5-enolpy... 167 361 408 445.8 7.5e-17  
 SWISSPROT:AROA\_METTH  
 O26860 methanobacterium thermoautot... 136 465 396 432.9 3.9e-16  
 GP\_BCT1:AE000855\_4  
 AE000855 Methanothermobacter therma... 136 465 396 432.9 3.9e-16  
 GP\_BCT1:AE005280\_4  
 AE005280 Escherichia coli O157:H7 E... 163 413 392 428.5 6.9e-16  
 GP\_BCT3:D90728\_13  
 D90728 Escherichia coli Escherichia... 163 413 392 428.5 6.9e-16  
 GP\_BCT1:AE000193\_3  
 AE000193 Escherichia coli K12 Esche... 163 413 392 428.5 6.9e-16  
 GP\_BCT2:AP002553\_239  
 AP002553 Escherichia coli O157:H7 E... 163 413 392 428.5 6.9e-16  
 GP\_BCT3:D90729\_6  
 D90729 Escherichia coli Escherichia... 163 413 392 428.5 6.9e-16  
 SWISSPROT:AROA\_ECOLI  
 P07638 escherichia coli, and escher... 163 413 392 428.5 6.9e-16  
 GP\_BCT1:AF101225\_1  
 AF101225 Shigella sonnei Shigella s... 163 413 391 427.4 7.9e-16  
 SWISSPROT:AROA\_SHISO  
 Q9zff7 shigella sonnei. 3-phosphosh... 163 413 391 427.4 7.9e-16  
 GP\_BCT2:M20023\_1  
 M20023 Bordetella pertussis B.pertu... 123 224 391 427.2 8.1e-16  
 SWISSPROT:AROA\_BORPE  
 P12421 bordetella pertussis. 3-phos... 123 224 391 427.2 8.1e-16  
 GP\_BCT3:X00557\_1  
 X00557 Escherichia coli E.coli gene... 160 407 386 422.0 1.6e-15  
 SWISSPROT:AROA\_EDWIC  
 Q9x4h2 edwardsiella ictaluri. 3-pho... 136 357 386 422.0 1.6e-15  
 GP\_BCT2:AF110153\_2  
 AF110153 Edwardsiella ictaluri Edwa... 136 357 386 422.0 1.6e-15  
 SWISSPROT:AROA\_BORBR  
 Q9rnd7 bordetella bronchiseptica (a... 123 212 385 420.7 1.9e-15  
 GP\_BCT2:AF182427\_3  
 AF182427 Bordetella bronchiseptica ... 123 212 385 420.7 1.9e-15  
 GP\_BCT3:U82268\_1  
 U82268 Shigella dysenteriae Shigell... 163 413 382 417.7 2.7e-15  
 SWISSPROT:AROA\_SHIDY  
 O87006 shigella dysenteriae. 3-phos... 163 413 382 417.7 2.7e-15  
 TREMBL\_NEW:AAL19912  
 Aal19912 salmonella typhimurium lt2... 148 382 378 413.4 4.8e-15  
 GP\_BCT1:AE006122\_8  
 AE006122 Pasteurella multocida Past... 124 283 373 407.8 9.8e-15  
 SWISSPROT:AROA\_PASMU  
 Q04570 pasteurella multocida. 3-pho... 124 283 373 407.8 9.8e-15  
 SWISSPROT:AROA\_SALTY  
 P07637 salmonella typhimurium. 3-ph... 142 376 372 406.9 1.1e-14  
 GP\_BCT3:M10947\_1  
 M10947 Salmonella typhimurium S.typ... 142 376 372 406.9 1.1e-14  
 GP\_BCT3:Y10355\_2  
 Y10355 Salmonella typhimurium S.typ... 142 376 372 406.9 1.1e-14  
 SWISSPROT:AROA\_SALGL



! P22299 salmonella gallinarum. 3-pho... 147 348 370 404.7 1.4e-14  
GP\_BCT3:M62801\_2  
! M62801 Salmonella enterica subsp. e... 147 348 370 404.7 1.4e-14  
GP\_BCT2:AP001118\_291  
! AP001118 Buchnera sp. APS Buchnera ... 133 293 369 403.7 1.7e-14  
SWISSPROT:AROA\_BUCAI  
! P57396 buchnera aphidicola (subsp. ... 133 293 369 403.7 1.7e-14  
TREMBL\_NEW:CAD05378  
! Cad05378 salmonella enterica subsp.... 147 348 365 399.3 2.9e-14  
SWISSPROT:AROA\_SALTI  
! P19786 salmonella typhi. 3-phosphos... 147 348 361 395.0 5e-14  
GP\_BCT3:X54545\_1  
! X54545 Salmonella typhi S.typhi gen... 147 348 361 395.0 5e-14  
GP\_BCT3:Z14100\_1  
! Z14100 Pasteurella multocida P.mult... 124 283 358 391.6 7.8e-14  
TREMBL\_MAIN:Q9K4A7  
! Q9K4A7 streptomyces coelicolor. 3-p... 104 259 357 390.5 8.9e-14  
GP\_BCT3:AL359214\_9  
! AL359214 Streptomyces coelicolor A3... 104 259 357 390.5 8.9e-14  
TREMBL\_MAIN:Q9L213  
! Q9L213 streptomyces coelicolor. 3-p... 144 327 349 381.9 2.7e-13  
GP\_BCT3:AL138598\_28  
! AL138598 Streptomyces coelicolor A3... 144 327 349 381.9 2.7e-13  
TREMBL\_NEW:AAK86449  
! Aak86449 agrobacterium tumefaciens.... 156 380 329 360.5 4.2e-12  
SWISSPROT:AROA\_ARCFU  
! O28775 archaeglobus fulgidus. prob... 103 283 323 354.2 9.5e-12  
GP\_BCT1:AE000999\_8  
! AE000999 Archaeglobus fulgidus Arc... 103 283 323 354.2 9.5e-12  
TREMBL\_MAIN:Q9JTT3  
! Q9JTT3 neisseria meningitidis (sero... 159 238 315 345.3 3e-11  
GP\_BCT3:AL162756\_195  
! AL162756 Neisseria meningitidis Z24... 159 238 315 345.3 3e-11  
GP\_BCT1:AE001310\_1  
! AE001310 Chlamydia trachomatis Chla... 67 202 306 335.4 1e-10  
SWISSPROT:AROA\_CHLTR  
! O84371 chlamydia trachomatis. 3-pho... 67 202 306 335.4 1e-10  
TREMBL\_MAIN:Q9JYU1  
! Q9JYU1 neisseria meningitidis (sero... 156 235 305 334.5 1.2e-10  
GP\_BCT1:AE002492\_7  
! AE002492 Neisseria meningitidis MC5... 156 235 305 334.5 1.2e-10  
GP\_PLN2:AP002542\_28  
! AP002542 Oryza sativa Oryza sativa ... 102 178 289 317.9 1e-09  
GP\_PLN1:AB016765\_1  
! AB016765 Oryza sativa Oryza sativa ... 102 178 289 317.9 1e-09  
TREMBL\_MAIN:O80428  
! O80428 oryza sativa (rice). 3-phosp... 102 178 289 317.9 1e-09  
GP\_PLN4:M61905\_1  
! M61905 Nicotiana tabacum N.tabacum ... 105 161 286 315.6 1.3e-09  
SWISSPROT:AROA2\_TOBAC  
! P23281 nicotiana tabacum (common to... 105 161 286 315.6 1.3e-09  
TREMBL\_MAIN:Q9ZEQ0  
! Q9ZEQ0 actinobacillus pleuropneumon... 137 164 285 314.6 1.5e-09  
GP\_BCT2:AJ012748\_1  
! AJ012748 Actinobacillus pleuropneum... 137 164 285 314.6 1.5e-09  
GP\_BCT3:AL583919\_127  
! AL583919 Mycobacterium leprae Mycob... 70 211 286 314.0 1.6e-09  
TREMBL\_MAIN:Q9CCI3  
! Q9CCI3 mycobacterium leprae. putati... 70 211 286 314.0 1.6e-09  
TREMBL\_MAIN:Q9HQC1  
! Q9HQC1 halobacterium sp. (strain nr... 151 398 282 309.5 2.9e-09  
GP\_BCT1:AE005049\_6  
! AE005049 Halobacterium sp. NRC-1 Ha... 151 398 282 309.5 2.9e-09  
GP\_BCT3:U67500\_3  
! U67500 Methanococcus jannaschii Met... 156 294 280 307.5 3.7e-09  
SWISSPROT:AROA\_METJA  
! Q57925 methanococcus jannaschii. pr... 156 294 280 307.5 3.7e-09  
GP\_BCT2:AP000060\_34  
! AP000060 Aeropyrum pernix Aeropyrum... 130 249 275 302.2 7.5e-09  
SWISSPROT:AROA\_AERPE  
! Q9yek9 aeropyrum pernix. probable 3... 130 249 275 302.2 7.5e-09

GP\_BCT1:AE002332\_14  
! AE002332 Chlamydia muridarum Chlamy... 186 337 274 300.9 8.8e-09  
SWISSPROT:AROA\_CHLMU  
! Q9pk28 chlamydia muridarum. 3-phosp... 186 337 274 300.9 8.8e-09  
TREMBL\_MAIN:Q97KM2  
! Q97km2 clostridium acetobutylicum. ... 154 383 266 292.4 2.6e-08  
GP\_BCT3:M62708\_1  
! M62708 Mycobacterium tuberculosis M... 76 222 264 289.9 3.6e-08  
SWISSPROT:AROA\_MYCTU  
! P22487 mycobacterium tuberculosis. ... 76 222 264 289.9 3.6e-08  
GP\_BCT3:X52269\_1  
! X52269 Mycobacterium tuberculosis M... 76 222 264 289.9 3.6e-08  
GP\_BCT3:Z95121\_2  
! Z95121 Mycobacterium tuberculosis M... 76 222 264 289.9 3.6e-08  
GP\_BCT1:AE007144\_4  
! AE007144 Mycobacterium tuberculosis... 76 222 264 289.9 3.6e-08  
GP\_BCT1:AE001959\_10  
! AE001959 Deinococcus radiodurans De... 96 309 261 286.5 5.5e-08  
SWISSPROT:AROA\_DEIRA  
! Q9rvd3 deinococcus radiodurans. 3-p... 106 401 260 285.8 6.1e-08  
SWISSPROT:AROA1\_TOBAC  
! P23981 nicotiana tabacum (common to... 134 272 261 285.8 6.1e-08  
GP\_PLN4:M61904\_1  
! M61904 Nicotiana tabacum N.tabacum ... 134 272 261 285.8 6.1e-08  
SWISSPROT:AROA\_LYCES  
! P10748 lycopersicon esculentum (tom... 133 243 252 276.0 2.1e-07  
GP\_PLN4:M21071\_1  
! M21071 Lycopersicon esculentum Toma... 133 243 252 276.0 2.1e-07  
GP\_PLN3:M21084\_1  
! M21084 Petunia x hybrida P.hybrida ... 136 279 249 272.8 3.2e-07  
SWISSPROT:AROA\_PETHY  
! P11043 petunia hybrida (petunia). 3... 136 279 249 272.8 3.2e-07  
GP\_PAT1:A82498\_1  
! A82498 Petunia x hybrida Sequence 1... 136 279 249 272.8 3.2e-07  
GP\_BCT3:X72784\_1  
! X72784 Synechocystis sp. PCC 6803 S... 173 214 238 272.4 3.4e-07  
GP\_PLN3:L18918\_1  
! L18918 Pneumocystis carinii Pneumoc... 56 122 251 267.6 6.3e-07  
SWISSPROT:AROA\_PNECA  
! Q12659 p pentafunctional arom polyp... 56 122 251 267.6 6.3e-07  
GP\_BCT1:L05002\_1  
! L05002 Aeromonas salmonicida Aeromo... 134 162 239 263.3 1.1e-06  
SWISSPROT:AROA\_AERSA  
! Q03321 aeromonas salmonicida. 3-pho... 134 162 239 263.3 1.1e-06  
TREMBL\_MAIN:Q9AT37  
! Q9at37 lolium rigidum. 3-phosphoshi... 102 151 237 262.5 1.2e-06  
GP\_PLN2:AF349754\_1  
! AF349754 Lolium rigidum Lolium rigi... 102 151 237 262.5 1.2e-06  
GP\_PLN2:AF360224\_1  
! AF360224 Arabidopsis thaliana Arabi... 123 239 238 260.9 1.5e-06  
TREMBL\_NEW:AAK64123  
! Aak64123 arabidopsis thaliana (mous... 123 239 238 260.9 1.5e-06  
TREMBL\_MAIN:Q9FVP6  
! Q9fvp6 arabidopsis thaliana (mouse-... 123 239 238 260.9 1.5e-06  
GP\_PLN1:AC084414\_12  
! AC084414 Arabidopsis thaliana Arabi... 123 239 238 260.9 1.5e-06  
GP\_PLN1:AC084242\_2  
! AC084242 Arabidopsis thaliana Arabi... 123 239 238 260.9 1.5e-06  
GP\_PAT1:A59345\_1  
! A59345 Zea mays Sequence 2 from Pat... 136 300 237 260.9 1.5e-06  
GP\_PLN4:X63374\_1  
! X63374 Zea mays Z.mays mRNA for EPS... 136 300 237 260.9 1.5e-06  
TREMBL\_MAIN:O24566  
! O24566 zea mays (maize). 3-phosphos... 136 300 237 260.9 1.5e-06  
GP\_PAT1:A59404\_1  
! A59404 Zea mays Sequence 2 from Pat... 136 300 237 260.9 1.5e-06  
GP\_PAT1:A69535\_1  
! A69535 unidentified Sequence 3 from... 136 300 237 260.9 1.5e-06  
SWISSPROT:AROA\_BRANA  
! P17688 brassica napus (rape). 3-pho... 124 268 235 257.7 2.2e-06  
GP\_PLN3:X51475\_1

! X51475 Brassica napus Brassica napu... 124 268 235 257.7 2.2e-06  
 SWISSPROT:AROA\_CORGL  
 ! Q9z470 corynebacterium glutamicum (... 97 226 233 256.8 2.5e-06  
 GP\_PAT2:AX064129\_1  
 ! AX064129 Corynebacterium glutamicum... 97 226 233 256.8 2.5e-06  
 GP\_BCT2:AF114233\_1  
 ! AF114233 Corynebacterium glutamicum... 97 226 233 255.8 2.9e-06  
 GP\_PAT1:A59347\_1  
 ! A59347 Zea mays Sequence 4 from Pat... 136 276 232 255.5 3e-06  
 GP\_PAT1:A59406\_1  
 ! A59406 Zea mays Sequence 4 from Pat... 136 276 232 255.5 3e-06  
 GP\_PAT1:A69537\_1  
 ! A69537 unidentified Sequence 5 from... 136 276 232 255.5 3e-06  
 TREMBL\_NEW:AAL07437  
 ! Aal07437 oryza sativa (rice). epsp ... 136 273 232 254.5 3.4e-06  
 TREMBL\_NEW:AAL06593  
 ! Aal06593 oryza sativa (rice). 5-eno... 136 273 232 254.5 3.4e-06  
 TREMBL\_NEW:BAB61062  
 ! Bab61062 oryza sativa (rice). 3-pho... 136 273 232 254.5 3.4e-06  
 GP\_PAT1:A18838\_1  
 ! A18838 Aeromonas salmonicida aroA g... 124 152 229 252.5 4.4e-06  
 SWISSPROT:AROA\_ARATH  
 ! P05466 arabidopsis thaliana (mouse-... 125 241 230 252.3 4.5e-06  
 GP\_PLN2:X06613\_1  
 ! X06613 Arabidopsis thaliana Arabido... 125 241 230 252.3 4.5e-06  
 GP\_PLN1:AC002387\_14  
 ! AC002387 Arabidopsis thaliana Arabi... 125 241 230 252.3 4.5e-06  
 TREMBL\_NEW:CAD01096  
 ! Cad01096 eleusine indica (goosegras... 136 300 229 252.2 4.5e-06  
 SWISSPROT:AROA\_BUCAP  
 ! Q59178 buchnera aphidicola (subsp. ... 103 298 227 250.3 5.8e-06  
 GP\_BCT2:L43549\_5  
 ! L43549 Buchnera aphidicola Buchnera... 103 298 227 250.3 5.8e-06  
 TREMBL\_NEW:CAD01095  
 ! Cad01095 eleusine indica (goosegras... 136 276 221 243.6 1.4e-05  
 TREMBL\_MAIN:Q9K9D5  
 ! Q9K9D5 bacillus halodurans. 3-phosp... 115 242 216 238.2 2.7e-05  
 GP\_BCT2:AP001516\_142  
 ! AP001516 Bacillus halodurans Bacill... 115 242 216 238.2 2.7e-05  
 TREMBL\_MAIN:Q9YC47  
 ! Q9yc47 aeropyrum pernix. 427aa long... 82 105 210 232.0 6.1e-05  
 GP\_BCT2:AP000061\_266  
 ! AP000061 Aeropyrum pernix Aeropyrum... 82 105 210 232.0 6.1e-05  
 SWISSPROT:ARO1\_SCHPO  
 ! Q9p7r0 s pentafunctional arom polyp... 109 179 216 229.8 8e-05  
 GP\_PLN4:AL157734\_2  
 ! AL157734 Schizosaccharomyces pombe ... 109 179 216 229.8 8e-05  
 GP\_PLN4:X06077\_1  
 ! X06077 Saccharomyces cerevisiae Yea... 82 122 212 225.5 0.00014  
 GP\_PLN4:Z48179\_2  
 ! Z48179 Saccharomyces cerevisiae S.c... 82 122 212 225.5 0.00014  
 SWISSPROT:ARO1\_YEAST  
 ! P08566 s pentafunctional arom polyp... 82 122 212 225.5 0.00014  
 SWISSPROT:ARO1\_EMENI  
 ! P07547 e pentafunctional arom polyp... 110 181 210 223.3 0.00018  
 GP\_PLN2:X05204\_1  
 ! X05204 Emericella nidulans Aspergil... 110 181 210 223.3 0.00018  
 GP\_BCT3:AL445063\_278  
 ! AL445063 Thermoplasma acidophilum T... 76 235 188 208.5 0.0012  
 SWISSPROT:AROA\_THEAC  
 ! Q9h1e6 thermoplasma acidophilum. pr... 76 235 188 208.5 0.0012  
 TREMBL\_NEW:BAB67386  
 ! Bab67386 sulfolobus tokodaii. 408aa... 89 117 186 206.4 0.0016  
 GP\_BCT1:AE006665\_12  
 ! AE006665 Sulfolobus solfataricus Su... 111 183 182 201.9 0.0028  
 TREMBL\_MAIN:Q980I5  
 ! Q980I5 sulfolobus solfataricus. 3-p... 111 183 182 201.9 0.0028  
 TREMBL\_NEW:CAD00604  
 ! Cad00604 listeria monocytogenes. ud... 49 75 180 199.5 0.0039  
 TREMBL\_NEW:CAC97896  
 ! Cac97896 listeria innocua. udp-n-ac... 49 75 180 199.5 0.0039

GP\_BCT1:AE001684\_6  
 ! AE001684 Chlamydomophila pneumoniae C... 115 297 180 199.3 0.004  
 GP\_BCT2:AP002548\_241  
 ! AP002548 Chlamydomophila pneumoniae J... 115 297 180 199.3 0.004  
 GP\_BCT1:AE002240\_9  
 ! AE002240 Chlamydomophila pneumoniae A... 115 297 180 199.3 0.004  
 SWISSPROT:AROA\_CHLPN  
 ! Q9z6m0 chlamydia pneumoniae (chlamy... 115 297 180 199.3 0.004  
 TREMBL\_NEW:AAK64441  
 ! Aak64441 myxococcus xanthus. udp-gl... 85 108 175 194.3 0.0076  
 GP\_BCT3:AL138851\_8  
 ! AL138851 Streptomyces coelicolor A3... 59 141 175 193.9 0.008  
 TREMBL\_MAIN:Q9LIU5  
 ! Q9LIU5 streptomyces coelicolor. udp... 59 141 175 193.9 0.008  
 TREMBL\_MAIN:Q99Z78  
 ! Q99Z78 streptococcus pyogenes. puta... 69 97 173 192.1 0.01  
 GP\_BCT1:AE006573\_8  
 ! AE006573 Streptococcus pyogenes M1 ... 69 97 173 192.1 0.01  
 TREMBL\_MAIN:Q9SON1  
 ! Q9son1 streptomyces lividans. udp-n... 59 141 170 188.5 0.016  
 GP\_BCT1:AB033486\_2  
 ! AB033486 Streptomyces lividans Stre... 59 141 170 188.5 0.016  
 TREMBL\_NEW:CAD00630  
 ! Cad00630 listeria monocytogenes. mu... 89 112 165 183.4 0.031  
 TREMBL\_MAIN:Q9ALU1  
 ! Q9alul streptococcus sanguis. udp-n... 48 75 163 182.4 0.035  
 GP\_BCT2:AF343003\_1  
 ! AF343003 Streptococcus sanguinis St... 48 75 163 182.4 0.035  
 GP\_PLN4:S76670\_1  
 ! S76670 Pneumocystis carinii arom=en... 91 91 154 180.6 0.044  
 TREMBL\_MAIN:Q12430  
 ! Q12430 pneumocystis carinii. epsp s... 91 91 154 180.6 0.044  
 GP\_PLN3:U31057\_1  
 ! U31057 Pneumocystis carinii f. sp. ... 91 91 154 180.6 0.044  
 TREMBL\_NEW:AAF86297  
 ! Aaf86297 listeria monocytogenes. ud... 76 76 156 173.7 0.11  
 TREMBL\_NEW:CAC97923  
 ! Cac97923 listeria innocua. murz pro... 78 78 156 173.7 0.11  
 GP\_BCT1:AF072894\_2  
 ! AF072894 Listeria monocytogenes Lis... 76 76 156 173.5 0.11  
 TREMBL\_MAIN:Q9KKF2  
 ! Q9kkf2 listeria monocytogenes. udp-... 76 76 156 173.5 0.11  
 SWISSPROT:AROA\_PYRAB  
 ! Q9v1hl pyrococcus abyssi. probable ... 79 268 153 170.7 0.16  
 GP\_BCT2:AJ248284\_158  
 ! AJ248284 Pyrococcus abyssi Pyrococc... 79 268 153 170.7 0.16  
 TREMBL\_MAIN:Q12733  
 ! Q12733 pneumocystis carinii. 3-phos... 69 69 144 169.8 0.18  
 GP\_PLN4:S76668\_1  
 ! S76668 Pneumocystis carinii arom=en... 69 69 144 169.8 0.18  
 GP\_PLN3:U31056\_1  
 ! U31056 Pneumocystis carinii f. sp. ... 69 69 144 169.8 0.18  
 GP\_PLN3:U31053\_1  
 ! U31053 Pneumocystis carinii f. sp. ... 55 55 140 165.5 0.31  
 TREMBL\_MAIN:Q12654  
 ! Q12654 pneumocystis carinii. 3-phos... 55 55 140 165.5 0.31  
 TREMBL\_MAIN:Q9A0I4  
 ! Q9a0i4 streptococcus pyogenes. puta... 52 52 148 165.1 0.32  
 GP\_BCT1:AE006528\_2  
 ! AE006528 Streptococcus pyogenes M1 ... 52 52 148 165.1 0.32  
 GP\_BCT2:Z99123\_7  
 ! Z99123 Bacillus subtilis Bacillus s... 45 73 147 163.9 0.37  
 GP\_BCT2:Z99122\_207  
 ! Z99122 Bacillus subtilis Bacillus s... 45 73 147 163.9 0.37  
 SWISSPROT:MUA2\_BACSU  
 ! P19670 bacillus subtilis. udp-n-ace... 45 73 147 163.9 0.37  
 GP\_BCT2:Z49782\_15  
 ! Z49782 Bacillus subtilis B.subtilis... 45 73 147 163.9 0.37  
 TREMBL\_MAIN:Q12655  
 ! Q12655 pneumocystis carinii. 3-phos... 55 86 138 163.3 0.4  
 GP\_PLN3:U31054\_1

! U31054 Pneumocystis carinii f. sp. ... 55 86 138 163.3 0.4  
TREMBL\_MAIN:Q978S3  
! Q978s3 thermoplasma volcanium. 5-en... 85 197 145 162.1 0.47  
GP\_BCT2:AP000996\_78  
! AP000996 Thermoplasma volcanium The... 85 197 145 162.1 0.47  
GP\_BCT3:AJ295297\_1  
! AJ295297 Mycobacterium chelonae Myc... 62 88 140 156.5 0.96  
TREMBL\_MAIN:Q9EXE3  
! Q9exe3 mycobacterium chelonae. udp-... 62 88 140 156.5 0.96  
TREMBL\_MAIN:Q9KP62  
! Q9kp62 vibrio cholerae. udp-n-acety... 50 115 138 154.4 1.3  
GP\_BCT1:AE004321\_8  
! AE004321 Vibrio cholerae Vibrio cho... 50 115 138 154.4 1.3  
SWISSPROT:MURA\_BUCAI  
! P57466 buchnera aphidicola (subsp. ... 39 39 137 153.3 1.5  
GP\_BCT2:AP001119\_68  
! AP001119 Buchnera sp. APS Buchnera ... 39 39 137 153.3 1.5  
TREMBL\_MAIN:Q9RVA6  
! Q9rva6 deinococcus radiodurans. udp... 49 83 134 149.9 2.2  
GP\_BCT1:AE001962\_5  
! AE001962 Deinococcus radiodurans De... 49 83 134 149.9 2.2  
GP\_PLN3:U31055\_1  
! U31055 Pneumocystis carinii f. sp. ... 55 55 125 149.3 2.4  
TREMBL\_MAIN:Q12656  
! Q12656 pneumocystis carinii. 3-phos... 55 55 125 149.3 2.4  
GP\_BCT3:D64000\_90  
! D64000 Synechocystis sp. Synechocys... 79 105 133 148.7 2.6  
SWISSPROT:MURA\_SYNY3  
! Q55673 synechocystis sp. (strain pc... 79 105 133 148.7 2.6  
SWISSPROT:RPOA\_SHESP  
! P74963 shewanella sp. (strain db670... 56 56 131 148.4 2.7  
GP\_BCT3:D83194\_1  
! D83194 Shewanella sp. Barophilic ba... 56 56 131 148.4 2.7  
SWISSPROT:MUA1\_LACLA  
! Q9cip4 lactococcus lactis (subsp. l... 55 55 132 147.8 2.9  
GP\_BCT1:AE006268\_6  
! AE006268 Lactococcus lactis subsp. ... 55 55 132 147.8 2.9  
TREMBL\_MAIN:Q99SD4  
! Q99sd4 staphylococcus aureus subsp.... 41 65 131 146.8 3.4  
GP\_BCT2:AP003364\_190  
! AP003364 Staphylococcus aureus subs... 41 65 131 146.8 3.4  
GP\_BCT2:AP003136\_76  
! AP003136 Staphylococcus aureus subs... 41 65 131 146.8 3.4  
TREMBL\_NEW:BA58286  
! Bab58286 staphylococcus aureus subs... 41 65 131 146.8 3.4  
TREMBL\_NEW:CAC49012  
! Cac49012 rhizobium meliloti (sinorh... 37 37 131 144.9 4.3  
GP\_BCT3:U31523\_1  
! U31523 Escherichia coli Escherichia... 47 80 114 142.8 5.6  
\\End of List

nk603cp4.pep  
SWISSPROT:AROA\_AGRSP

ID AROA\_AGRSP STANDARD; PRT; 455 AA.  
AC Q9R4E4;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 2896 Initn: 2896 Opt: 2900 z-score: 3136.1 E(): 1.1e-166  
Smith-Waterman score: 2900; 99.6% identity in 455 aa overlap

	10	20	30	40	50	60
nk603cp4.pep	MLHGASSRPATARKSSGLSGTVRI	PGDKSISHR	SFMFGGLASGETRITGL	LEGEDVINTG		
AROA_AGRSP	MSHGASSRPATARKSSGLSGTVRI	PGDKSISHR	SFMFGGLASGETRITGL	LEGEDVINTG		
	10	20	30	40	50	60

	70	80	90	100	110	120
nk603cp4.pep	KAMQAMGARIRKEGDTWIIDGVNGG	LLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI				
AROA_AGRSP	KAMQAMGARIRKEGDTWIIDGVNGG	LLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI				
	70	80	90	100	110	120
	130	140	150	160	170	180
nk603cp4.pep	GDASLTKRPMGRVLNPLREMGVQVKS	EDGDRLPVTLRGPKTPTPTIYRVPMASQVKS	SAV			
AROA_AGRSP	GDASLTKRPMGRVLNPLREMGVQVKS	EDGDRLPVTLRGPKTPTPTIYRVPMASQVKS	SAV			
	130	140	150	160	170	180
	190	200	210	220	230	240
nk603cp4.pep	LLAGLNTPGITTVIEPIIMTRDHT	TEKMLQGFGANPTVETDADGVRTIRLEGR	KGKLTGQVID			
AROA_AGRSP	LLAGLNTPGITTVIEPIIMTRDHT	TEKMLQGFGANPTVETDADGVRTIRLEGR	KGKLTGQVID			
	190	200	210	220	230	240
	250	260	270	280	290	300
nk603cp4.pep	VPGDPSSTAFPLVAALLVPGSDVT	ILNVLNPNTRTGLILTLQEMGADIEVIN	PRLAGGED			
AROA_AGRSP	VPGDPSSTAFPLVAALLVPGSDVT	ILNVLNPNTRTGLILTLQEMGADIEVIN	PRLAGGED			
	250	260	270	280	290	300
	310	320	330	340	350	360
nk603cp4.pep	VADLRVRSSTLKGVTVPEDRAPSM	IDEPILAVAAFAEGATVMNGLEBLRVKES	DRLSA			
AROA_AGRSP	VADLRVRSSTLKGVTVPEDRAPSM	IDEPILAVAAFAEGATVMNGLEBLRVKES	DRLSA			
	310	320	330	340	350	360
	370	380	390	400	410	420
nk603cp4.pep	VANGLKLVGDCDEGETSLVVRGR	PDGKGLGNASGAAVATHLDRHRIAMS	FLVGLVSENP			
AROA_AGRSP	VANGLKLVGDCDEGETSLVVRGR	PDGKGLGNASGAAVATHLDRHRIAMS	FLVGLVSENP			
	370	380	390	400	410	420
	430	440	450			
nk603cp4.pep	VTVDDATMIATSFPEFMDLMAGL	GAKIELSDTKAA				
AROA_AGRSP	VTVDDATMIATSFPEFMDLMAGL	GAKIELSDTKAA				
	430	440	450			

nk603cp4.pep  
TREMBL\_NEW:CAC41690

ID CAC41690 PRELIMINARY; PRT; 455 AA.  
AC CAC41690;  
DT 06-SEP-2001 (EMBLrel. 63, Created)  
DT 06-SEP-2001 (EMBLrel. 63, Last sequence update)  
DT 06-SEP-2001 (EMBLrel. 63, Last annotation update)  
DE PUTATIVE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE PROTEIN (EC . . .

SCORES Init1: 2672 Initn: 2672 Opt: 2676 z-score: 2894.2 E(): 3.1e-153  
Smith-Waterman score: 2676; 90.3% identity in 455 aa overlap

	10	20	30	40	50	60
nk603cp4.pep	MLHGASSRPATARKSSGLSGTVRI	PGDKSISHR	SFMFGGLASGETRITGL	LEGEDVINTG		
CAC41690	MSHGASNPRPATARKSSDLKGT	LRIPGDKSISHR	SFMFGGLAAGETRITGL	LEGEDVINTG		
	10	20	30	40	50	60
	70	80	90	100	110	120
nk603cp4.pep	KAMQAMGARIRKEGDTWIIDGVNGG	LLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI				
CAC41690	KAMQAMGARIRKEGDTWIIDGVNGG	LLAPEAPLDFGNAGTGCRLTMGLVGVYDFDSTFI				
	70	80	90	100	110	120
	130	140	150	160	170	180

nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMSAQVKSAAV  
CAC41690 GDASLTKRPMGRVLDPLREMGVQVKSABGDRLPVTLRGPKTPNPITYRVPMSAQVKSAAV  
130 140 150 160 170 180

190 200 210 220 230 240  
nk603cp4.pep LLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVID  
CAC41690 LLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVID  
190 200 210 220 230 240

250 260 270 280 290 300  
nk603cp4.pep VPGDPSSTAFPLVAALLVPGSDVTILNVLNMPNTRTGLILTLQEMGADIEVINPRLAGGED  
CAC41690 VPGDPSSTAFPLVAGLIVPGSDITILNVLNMPNTRTGLILTLQEMGANIEVMNKRLAGGED  
250 260 270 280 290 300

310 320 330 340 350 360  
nk603cp4.pep VADLRVRSSTLKGVTVPEDRAPSMIDEYFILAFAAFAEGATVMNGLEELRVKESDRLSA  
CAC41690 VADLRVRHSELKGVTVPEDRAPSMIDEYFILAFAAFAEGATVMNGLEELRVKESDRLSA  
310 320 330 340 350 360

370 380 390 400 410 420  
nk603cp4.pep VANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAATHLDHRIAMSFVLMGLVSENP  
CAC41690 VADGLKNGVDCDEGEASLVVRGRPGGKGLGKISGGQVKTLDHRIAMSFVLMGLASEHP  
370 380 390 400 410 420

430 440 450  
nk603cp4.pep VTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
CAC41690 VTVDATMIATSFPEFMGLMTGLGAKIEEAENKAA  
430 440 450

nk603cp4.pep  
SWISSPROT:AROA\_PSES2

ID AROA\_PSES2 STANDARD; PRT; 449 AA.  
AC P56952;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 2127 Initn: 2413 Opt: 2433 z-score: 2631.9 E(): 1.3e-138  
Smith-Waterman score: 2433; 82.7% identity in 450 aa overlap

10 20 30 40 50 60  
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AROA\_PSES2 MSHSASPKPATARRSEALTGEIRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
10 20 30 40 50 60

70 80 90 100 110 120  
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLGVGYDFDSTFI  
AROA\_PSES2 RAMQAMGAKIRKEGDVWIINGVNGGCLLOPEAALDFGNAGTGARLTMGLGVGYDMKTSFI  
70 80 90 100 110 120

130 140 150 160 170 180  
nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMSAQVKSAAV  
AROA\_PSES2 GDASLSKRPMPGRVLNPLREMGVQVEAADGDRMPLTLIGPKTANPITYRVPMSAQVKSAAV  
130 140 150 160 170 180

190 200 210 220 230 240  
nk603cp4.pep LLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVID  
AROA\_PSES2 LLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVID  
190 200 210 220 230 240

AROA\_PSES2 LLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVID  
190 200 210 220 230 240

250 260 270 280 290 300  
nk603cp4.pep VPGDPSSTAFPLVAALLVPGSDVTILNVLNMPNTRTGLILTLQEMGADIEVINPRLAGGED  
AROA\_PSES2 VPGDPSSTAFPLVAALLVPGSDVTIRNVLNMPNTRTGLILTLQEMGADIEVINPRLAGGED  
250 260 270 280 290 300

310 320 330 340 350 360  
nk603cp4.pep VADLRVRSSTLKGVTVPEDRAPSMIDEYFILAFAAFAEGATVMNGLEELRVKESDRLSA  
AROA\_PSES2 VADLRVRASLKGVTVPEDRAPSMIDEYFILAFAAFAEGATVMNGLEELRVKESDRLSA  
310 320 330 340 350 360

370 380 390 400 410 420  
nk603cp4.pep VANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAATHLDHRIAMSFVLMGLVSENP  
AROA\_PSES2 VARGLEANGVDCDEGETSLVVRGRPDGKGLG---GGTVATHLDHRIAMSFVLMGLAEKP  
370 380 390 400 410

430 440 450  
nk603cp4.pep VTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AROA\_PSES2 VTVDSSNMIATSFPEFMDMMFGLGAKIELSIL  
420 430 440

nk603cp4.pep  
TREMBL\_MAIN:Q9AGV2

ID Q9AGV2 PRELIMINARY; PRT; 480 AA.  
AC Q9AGV2;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE 5-ENOLPYRUVYL SHIKIMATE 3-PHOSPHATE SYNTHASE (EC 2.5.1.19) . . .

SCORES Init1: 2077 Initn: 2353 Opt: 2380 z-score: 2574.2 E(): 2.1e-135  
Smith-Waterman score: 2380; 80.9% identity in 450 aa overlap

10 20 30  
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSI  
Q9AGV2 MTTQYYYARETALISQSRGVSAPKCDCEKSMSSHACPKPATARRHSQALTGEIRIPGDKSI  
10 20 30 40 50 60

40 50 60 70 80 90  
nk603cp4.pep SHRSFMFGGLASGETRITGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAP  
Q9AGV2 STRSFMFGGLASGETRITGLLEGEDVINTGRAMQAMGARIRKEGDTWIINGVNGGCLLP  
70 80 90 100 110 120

100 110 120 130 140 150  
nk603cp4.pep EAPLDFGNAATGCRLTMGLGVGYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSSEDG  
Q9AGV2 EAPLDFGNAGTGARLTMGLGVGYDMKTSFIGDASLSKRPMPGRVLNPLREMGVQVEAAEGD  
130 140 150 160 170 180

160 170 180 190 200 210  
nk603cp4.pep RLPVTLRGPKTPTPTITYRVPMSAQVKSALLAGLNTPGITTVIEPIMTRDHTKMLQGF  
Q9AGV2 RMPLTLIGPRTANPIAYRVPMSAQVKSALLAGLNTPGITTVIEPIMTRDHTKMLQGF  
190 200 210 220 230 240

220 230 240 250 260 270  
nk603cp4.pep GANPTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLN  
Q9AGV2 GADLTVETDKDGVVRHRIIVGQKLTGQVIDVPGDPSSTAFPLVPALLVEGSEVTIRNVLN  
250 260 270 280 290 300

nk603cp4.pep 280 290 300 310 320 330  
 NPTRTGLILTLQEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI  
 Q9AGV2 310 320 330 340 350 360  
 NPTRTGLILTLQEMGADIEIIDPRLAGGEDVADLRVKASKLKGVVPPERAPSMIDEYPV

nk603cp4.pep 340 350 360 370 380 390  
 LAVAAAFAGATVMNGLEELRVKESDRLSAVANGLKNGVDCDEGETSLVVRGRPDGKGL  
 Q9AGV2 370 380 390 400 410 420  
 LAIXASFAEGETVMDGLDELVRKESDRLLAAVARGLEANGVDCTEGEMSLTVRGRPGGKGL

nk603cp4.pep 400 410 420 430 440 450  
 GNASGAAVATHLDHRIAMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELS  
 Q9AGV2 430 440 450 460 470  
 G---GGTVATHLDHRIAMSFLVMGLASEKPVTVDDSTMIATSFPEFMMAGLGAKIAES

nk603cp4.pep DTKAA  
 Q9AGV2 GAE  
 480

nk603cp4.pep  
 GP\_BCT2:AF326475\_1  
 LOCUS AF326475\_1 [AF326475]  
 DEFINITION Brucella melitensis biovar Abortus 5-enolpyruvyl shikimate  
 3-phosphate synthase (aroA) and CMP kinase (cmk) genes, complete  
 cds.  
 DATE 29-MAR-2001  
 ACCESSION AF326475 . . .

SCORES Init1: 2077 Initn: 2353 Opt: 2380 z-score: 2574.2 E(): 2.1e-135  
 Smith-Waterman score: 2380; 80.9% identity in 450 aa overlap

nk603cp4.pep 10 20 30  
 MLHGASSRPATARKSSGLSGTVRIPGDKSI  
 AF326475\_1 10 20 30 40 50 60  
 MTTQYYIARETALISQSRGVSAPKCDCEKSMHSACPKPATARHSQALGTEIRIPGDKSI

nk603cp4.pep 40 50 60 70 80 90  
 SHRSFMFGGLASGETRITGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGLLAP  
 AF326475\_1 70 80 90 100 110 120  
 STRSFMFGGLASGKTRITGLLEGEDVINTGRAMQAMGARIRKEGDTWIIINGVNGCLLP

nk603cp4.pep 100 110 120 130 140 150  
 EAPLDFGNAATGCRLTMGLVGVDYDFDSTFIGDASLTKEPMGRVNLPLREMGVQVKSSEDG  
 AF326475\_1 130 140 150 160 170 180  
 EAPLDFGNAGTGARLTMLGVGTIDMKTSFIGDASLSKRPGRVNLPLREMGVQVEAAEGD

nk603cp4.pep 160 170 180 190 200 210  
 RLPVTLRGPKTPITIRVPMASAVQVSAVLLAGLNTPGITTVIEPIIMTRDHEKMLQGF  
 AF326475\_1 190 200 210 220 230 240  
 RMPLTLIGPRTANPIAYRVPMASAVQVSAVLLAGLNTPGITTVIEPIIMTRDHEKMLQGF

nk603cp4.pep 220 230 240 250 260 270  
 GANPTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLM  
 AF326475\_1 250 260 270 280 290 300  
 GADLTVEYDKDGVRIHIVGQGLTGQITIDVPGDPSSTAFPLVPALLVEGSEVTIRNVLM

280 290 300 310 320 330

nk603cp4.pep NPTRTGLILTLQEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI  
 AF326475\_1 310 320 330 340 350 360  
 NPTRTGLILTLQEMGADIEIIDPRLAGGEDVADLRVKASKLKGVVPPERAPSMIDEYPV

nk603cp4.pep 340 350 360 370 380 390  
 LAVAAAFAGATVMNGLEELRVKESDRLSAVANGLKNGVDCDEGETSLVVRGRPDGKGL  
 AF326475\_1 370 380 390 400 410 420  
 LAIXASFAEGETVMDGLDELVRKESDRLLAAVARGLEANGVDCTEGEMSLTVRGRPGGKGL

nk603cp4.pep 400 410 420 430 440 450  
 GNASGAAVATHLDHRIAMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELS  
 AF326475\_1 430 440 450 460 470  
 G---GGTVATHLDHRIAMSFLVMGLASEKPVTVDDSTMIATSFPEFMMAGLGAKIAES

nk603cp4.pep DTKAA  
 AF326475\_1 GAE  
 480

nk603cp4.pep  
 GP\_BCT2:AP003006\_13  
 LOCUS AP003006\_13 [AP003006]  
 DEFINITION Mesorhizobium loti DNA, complete genome, section 13/21;  
 DATE 15-MAY-2001  
 ACCESSION AP003006  
 NID  
 ORGANISM Mesorhizobium loti . . .

SCORES Init1: 1300 Initn: 2247 Opt: 2289 z-score: 2476.4 E(): 5.9e-130  
 Smith-Waterman score: 2289; 77.6% identity in 451 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AP003006\_13 10 20 30 40 50 60  
 MSHAAAKPATARKSQUALSGTARVPGDKSISHRSFMFGGLASGETRITGLLEGEDVMTG

nk603cp4.pep 70 80 90 100 110 120  
 KAMQAMGARIRKEGDTWIIDGVNGGLLAPLDFGNAATGCRLTMGLVGVDYDFDSTFI  
 AP003006\_13 70 80 90 100 110 120  
 AAMKAMGAHIEKRGAEWVIRGTGNCALLQPEGPLDFGNAGTGSRLTMGLVGTIDMETTFI

nk603cp4.pep 130 140 150 160 170 179  
 GDASLTKRPMGRVNLPLREMGVQV-KSEGDRLPVTLRGPKTPITIRVPMASAVQVKS  
 AP003006\_13 130 140 150 160 170 180  
 GDASLSGRFPMGRVLEPLRQMGVQVLKATPGDRMPITLHGPKHAAPITYRVPMASAVQVKS

nk603cp4.pep 180 190 200 210 220 230 239  
 VLLAGLNTPGITTVIEPIIMTRDHEKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI  
 AP003006\_13 190 200 210 220 230 240  
 VLLAGLNTPGITTVIEPIIMTRDHEKMLQGFGANLVSVDTERGVRIHIFIEGQRLTGQTI

nk603cp4.pep 240 250 260 270 280 290 299  
 DVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLILTLQEMGADIEVINPRLAGGE  
 AP003006\_13 250 260 270 280 290 300  
 AVPGDPSSAGFPLVAALLVPGSDITIEVLMNPTRTGLLLTLQEMGGQIDILNPNRAGGE

nk603cp4.pep 300 310 320 330 340 350 359  
 DVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFAGATVMNGLEELRVKESDRLS

AP003006\_13 DVADLRVRYSELKGVAVPPERAPSMID EY PVLAVAASFAEGETLMQGLEELRVKESDRLS  
310 320 330 340 350 360

nk603cp4.pep 360 370 380 390 400 410  
AVANGLKLVGDCDEGETSLVVRGRPDGKGLG---NAGSAAVATHLDHRIAMSFVLMGLV  
AP003006\_13 360 370 380 390 400 410  
AVANGLKLVGDCDEGETSLVVRGRPDGKGLG---NAGSAAVATHLDHRIAMSFVLMGLV  
370 380 390 400 410 420

nk603cp4.pep 420 430 440 450  
SENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AP003006\_13 420 430 440 450  
TEKPTVDDQAMIAATSFPEFMDLMAGLGAKIELSDTKAA  
430 440 450

nk603cp4.pep  
TREMBL\_MAIN:Q98CC1

ID Q98CC1 PRELIMINARY; PRT; 452 AA.  
AC Q98CC1;  
DT 01-OCT-2001 (TREMBLrel. 18, Created)  
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . . .

SCORES Init1: 1300 Initn: 2247 Opt: 2289 z-score: 2476.4 E(): 5.9e-130  
Smith-Waterman score: 2289; 77.6% identity in 451 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFPMFGLASGETRITGLLEGEDVINTG  
Q98CC1 10 20 30 40 50 60  
MSHAAAKPATARKSQALSGTARVPGDKSISHRSFPMFGLASGETRITGLLEGEDVMTG

nk603cp4.pep 70 80 90 100 110 120  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVDFDSTFI  
Q98CC1 70 80 90 100 110 120  
AAMKAMGAHIEKGAEWVIRGTGNGALLQPEGPLDFGNAATGSRLTMGLVGVDFDSTFI

nk603cp4.pep 130 140 150 160 170 179  
GDASLTKRPMGRVNLPLREMGVQV-KSEGDRLPVTLRGFKTPTPTTYRPMASQVKS  
Q98CC1 130 140 150 160 170 180  
GDASLSGRPMGRVLEPLRQMGVQVLKATPGDRMPI TLHGPKHAAPTITYRPMASQVKS

nk603cp4.pep 180 190 200 210 220 230 239  
VLLAGLNTPGITTVIEPIMTRDHTKMLQGFANPTVETDADGVRTIRLEGRGKLTGQVI  
Q98CC1 180 190 200 210 220 230 240  
VLLAGLNTPGITTVIEPIMTRDHTKMLKGFANLSVETDERGVRRHIFIEGQRLTGQTI

nk603cp4.pep 240 250 260 270 280 290 299  
DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIEVINPLAGGE  
Q98CC1 240 250 260 270 280 290 300  
AVPGDPSSAGFPLVAALLVPGSDITIEVNLNPTRTGLLLTLQEMGGQIDILNPNRAGGE

nk603cp4.pep 300 310 320 330 340 350 359  
DVADLRVRSSTLKGVTVPEDRAPSMID EY PVLAVAASFAEGETLMQGLEELRVKESDRLS  
Q98CC1 300 310 320 330 340 350 360  
DVADLRVRYSELKGVAVPPERAPSMID EY PVLAVAASFAEGETLMQGLEELRVKESDRLS

nk603cp4.pep 360 370 380 390 400 410  
AVANGLKLVGDCDEGETSLVVRGRPDGKGLG---NAGSAAVATHLDHRIAMSFVLMGLV  
Q98CC1 360 370 380 390 400 410  
AVANGLKLVGDCDEGETSLVVRGRPDGKGLG---NAGSAAVATHLDHRIAMSFVLMGLV  
370 380 390 400 410 420

nk603cp4.pep 420 430 440 450  
SENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Q98CC1 420 430 440 450  
TEKPTVDDQAMIAATSFPEFMDLMAGLGAKIELSDTKAA

nk603cp4.pep  
GP\_BCT1:AE006017\_6

LOCUS AE006017\_6 [AE006017]  
DEFINITION Caulobacter crescentus section 343 of 359 of the complete genome;  
identified by match to protein family HMM.  
DATE 28-MAR-2001  
ACCESSION AE006017  
NID . . .

SCORES Init1: 611 Initn: 1423 Opt: 1473 z-score: 1595.4 E(): 6.9e-81  
Smith-Waterman score: 1473; 56.1% identity in 451 aa overlap

nk603cp4.pep 10 20 30 40 50 59  
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFPMFGLASGETRITGLLEGEDVINT  
AE006017\_6 10 20 30 40 50  
MSLAGLKSAPGGA-----LRGIVRAPGDKSISHRSMLGALATGTTTVEGLLEGDDVLAT

nk603cp4.pep 60 70 80 90 100 110  
GKAMQAMGARIRKEG-DTWI IDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVDFDST  
AE006017\_6 60 70 80 90 100 110  
ARAMQAFGARIEREGVGRWRIG--KGGFEFVDVIDCGNAGTGVRLIMGAAAGFAMCAT

nk603cp4.pep 120 130 140 150 160 170  
FIGDASLTKRPMGRVNLPLREMGVQVKSSEDGRLPVTLRGPKTPTPTTYRPMASQVKS  
AE006017\_6 120 130 140 150 160 170  
FTGDSLGRPMGRVLDPLARMGATWLGRDKGRPLTLKGGNL-RGLNYTLPMASQVKS

nk603cp4.pep 180 190 200 210 220 230  
AVLLAGLNTPGITTVIEPIMTRDHTKMLQGFANPTVETDADG---VRTIRL-EGRGKL  
AE006017\_6 180 190 200 210 220 230  
AVLLAGLHAEGGVIEFEATRDHTERMLRAFGAEVIVEDRKAGDKTFRHVRLEPGQ-KL

nk603cp4.pep 240 250 260 270 280 290  
TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIEVINPR  
AE006017\_6 240 250 260 270 280 290  
TGTHVAVPGDPSSTAFPLVAALLVPGSEVTVEGVMLNELRTGLFTTLQEMGADLVISNR

nk603cp4.pep 300 310 320 330 340 350  
LAGGEDVADLRVRSSTLKGVTVPEDRAPSMID EY PVLAVAASFAEGETLMQGLEELRVKE  
AE006017\_6 300 310 320 330 340 350  
VASGEEVGDIATARYSQLKGVVPPERAPSMID EY PVLAVAASFAEGETLMQGLEELRVKE

nk603cp4.pep 360 370 380 390 400 410  
SDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFVLMG  
AE006017\_6 360 370 380 390 400  
SDRLSLTANGKACGVQVVEEPEGFIVTGT---GQPPKGGATVVTGHDHRIAMSHLILG

nk603cp4.pep 420 430 440 450  
LVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AE006017\_6 420 430 440 450  
MAAQAEVAVDEPGMIATSFPGFADLMRGLGATLAEA

nk603cp4.pep  
TREMBL\_MAIN:Q9A2H2

ID Q9A2H2 PRELIMINARY; PRT; 443 AA.  
AC Q9A2H2;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . . .

SCORES Init1: 611 Initn: 1423 Opt: 1473 z-score: 1595.4 E(): 6.9e-81  
Smith-Waterman score: 1473; 56.1% identity in 451 aa overlap

```
nk603cp4.pep 10 20 30 40 50 59
               MLHGASSRPATARKSSGLSGTVRIPGDKSISHSFMSGGLASGETRITGLLEGEDVINT
Q9A2H2         MSLAGLSAPGGA-----LRGIVRAPGDKSISHSRMLGALATGTTTVEGLEGGDDVLAT
               10 20 30 40 50
```

```
nk603cp4.pep 60 70 80 90 100 110
               GKAMQAMGARIRKEG-DTWIIDGVNGGLLAPEAPLDFGNAATGTCRLTMGLVGVYDFDST
Q9A2H2         ARAMQAFGARIEREGVGRWRIEG--KGGFEEPVDVDCGNAGTGVRLLIMGAAAGFAMCAT
               60 70 80 90 100 110
```

```
nk603cp4.pep 120 130 140 150 160 170
               FIGDASLTKRPMGRVNLPLREMGVQVKSSEDGDRPLVTLRGPKTPTPIYRVPMASQVKS
Q9A2H2         FTGDQSLRGRPMGRVLDPLARMGATWLRDCKGRLPLTLKGNNL-RGLNYTLPMASQVKS
               120 130 140 150 160 170
```

```
nk603cp4.pep 180 190 200 210 220 230
               AVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFANPTVETDADG---VRTIRL-RGRGKL
Q9A2H2         AVLLAGLHAEGGVEVIEPEATRDHTERMLRAFGAEVIVEDRKAGDKTFRHVRLPEQG-KL
               180 190 200 210 220 230
```

```
nk603cp4.pep 240 250 260 270 280 290
               TGQVIDVPGDPSSTAPFLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPR
Q9A2H2         TGTHVAVPGDPSAAPPFLVAALLVPGSEVTVTEGVMLNELRTGLFTTLQEMGADLVISNVR
               240 250 260 270 280 290
```

```
nk603cp4.pep 300 310 320 330 340 350
               LAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKE
Q9A2H2         VASGEVGDITARYSQLKGVVPPERAPSMIDEYPILAVAAAFASGETVMRGVGMVRVKE
               300 310 320 330 340 350
```

```
nk603cp4.pep 360 370 380 390 400 410
               SDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFVLMG
Q9A2H2         SDRISLTANGLKACGVQVVERPEGFIVTGT---GQPPKGGATVVTHGDHRIAMSHLILG
               360 370 380 390 400
```

```
nk603cp4.pep 420 430 440 450
               LVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Q9A2H2         MAAQAEVAVDEPGMIATSFPGFADLMRGLGATLAEA
               410 420 430 440
```

nk603cp4.pep  
TREMBL\_MAIN:Q9HZ69

ID Q9HZ69 PRELIMINARY; PRT; 746 AA.  
AC Q9HZ69;  
DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE STILL FRAMESHIFT 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE . . .

SCORES Init1: 472 Initn: 1221 Opt: 1279 z-score: 1382.5 E(): 5e-69  
Smith-Waterman score: 1279; 48.5% identity in 441 aa overlap

```
nk603cp4.pep 10 20 30 40
               MLHGASSRPATARKSSGLSGTVRIPGDKSISHSFMSGGLA
Q9HZ69         RARVAREHFSKILARRAYVDAMHNNDLIYLAQPGGSLSGTIRVPGDKSISHSRIMLGSLA
               290 300 310 320 330 340
```

```
nk603cp4.pep 50 60 70 80 90 100
               SGETRITGLLEGEDVINTGKAMQAMGARIRKEGDTWI-IDGVNGGLLAPEAPLDFGNA
Q9HZ69         EGTTEVEGFLEGEDALATIQAFRDMGVVIEGPONGRVTVHGVGLHGLKAPPGPPIYLGNSG
               350 360 370 380 390 400
```

```
nk603cp4.pep 110 120 130 140 150 160
               TGCRLTMGLVGVYDFDSTFIGDASLTKRPMGRVNLPLREMGVQVKSSEDGDRPLVTLRGPK
Q9HZ69         TSMRLLSGLLAAQPFDSLTLTGDASLSKRPMMNRVAKPLREMGAVIETGPEGREPMTIRGGQ
               410 420 430 440 450 460
```

```
nk603cp4.pep 170 180 190 200 210 220
               TPTPIYRVPMASQVKS AVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFANPTVETDA
Q9HZ69         RLTMGHYDMPMASQVKSCLLLAGLYAAGETSVTEPAPTRDHTERMLRGFGYFVVVEGS-
               470 480 490 500 510 520
```

```
nk603cp4.pep 230 240 250 260 270 280
               DGVRTIRLEGKGLTGQVIDVPGDPSSTAPFLVAALLVPGSDVTILNVLNPNTRTGLILT
Q9HZ69         ----TAKVESGHKLSATHIEVPADISSAAFFLVAASIAEGSELVQLQHVGINPTRVGVIEI
               530 540 550 560 570 580
```

```
nk603cp4.pep 290 300 310 320 330 340
               LQEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEG
Q9HZ69         LRLMGGDLSENQREVGGEPVADIRVRSARLKGIDIPEDLVPLAIDEPFVLVFAAACAEAG
               590 600 610 620 630 640
```

```
nk603cp4.pep 350 360 370 380 390 400
               ATVMNGLEELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGA
Q9HZ69         RTVLRGAELRVKESDRIQVMADGLKALGVKAEPPTDGIIVIEG-----GAFGGGEVWA
               650 660 670 680 690
```

```
nk603cp4.pep 410 420 430 440 450
               HLDHRIAMSFVLMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Q9HZ69         HGDHRIAMSFVSLASLRSGPIRIHDCANVATSPFNFLALCAQTGIRVAVENN
               700 710 720 730 740
```

nk603cp4.pep  
GP\_BCT1:AE004740\_5

LOCUS AE004740\_5 [AE004740]  
DEFINITION Pseudomonas aeruginosa PA01, section 301 of 529 of the complete genome.  
DATE 30-AUG-2000  
ACCESSION AE004740  
NID . . .

SCORES Init1: 472 Initn: 1221 Opt: 1279 z-score: 1382.5 E(): 5e-69  
Smith-Waterman score: 1279; 48.5% identity in 441 aa overlap

```
nk603cp4.pep 10 20 30 40
               MLHGASSRPATARKSSGLSGTVRIPGDKSISHSFMSGGLA
```

AE004740\_5 RARVAREHFSKILARRAYVDAMHNDLIYLAQPGGSLSGTIRVPGDKSISHRSIMLGSLA  
290 300 310 320 330 340

nk603cp4.pep SGETRITGLLEGEDVINTGKAMQAMGARIRKEGDTWI-IDGVGNGGLLAPEAPLDFGNAA  
50 60 70 80 90 100  
AE004740\_5 EGTTEVEGFLEGEDALATIQAQPRDMGVVIEGPONGRVTVHGVGLHGLKAPPGPIYLGNSG  
350 360 370 380 390 400

nk603cp4.pep TGCRLTMGLVGVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSEDGRLPVTLRGPK  
110 120 130 140 150 160  
AE004740\_5 TSMRLLSGLLAAQPFDSLTLTGASLSKRPMNRVAKPLREMGAVIETGPEGRPPMTIRGGQ  
410 420 430 440 450 460

nk603cp4.pep TPTPITYRVPMASAVKSAVLLAGLNTPTGITTVEIPIMTRDHEKMLQGFANPTVETDADGVRTIRL  
170 180 190 200 210 220  
AE004740\_5 RLTMGHYDMFPMASAVKSAVLLAGLNTPTGITTVEIPIMTRDHEKMLQGFANPTVETDADGVRTIRL  
470 480 490 500 510 520

nk603cp4.pep DGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLIILT  
230 240 250 260 270 280  
AE004740\_5 ----TAKVESGHKLSATHIEVPADISSAAFFLVAASIAEGSELVLQHVGINPRTVGVIEI  
530 540 550 560 570 580

nk603cp4.pep LQEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFAG  
290 300 310 320 330 340  
AE004740\_5 LRLMGGDLSENQREVGGEVPADIRVRSARLKGIDIPEDLVPLAIDEPVLFVAAACAG  
590 600 610 620 630 640

nk603cp4.pep ATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPRDGGKLGNASGAAT  
350 360 370 380 390 400  
AE004740\_5 RTVLRGABELRVKESDRLQVMDAGLKGVAEPTPDGIVIEG-----GAFGGGEVWA  
650 660 670 680 690

nk603cp4.pep HLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLAKIELSDTKAA  
410 420 430 440 450  
AE004740\_5 HGDHRIAMSFVSLRASGPRIHDCANVATSPNPLALCAQTGIRVAENN  
700 710 720 730 740

nk603cp4.pep  
GP\_BCT1:AE004043\_7

LOCUS AE004043\_7 [AE004043]  
DEFINITION Xylella fastidiosa 9a5c, section 189 of 229 of the complete genome;  
similar to SP|P20691 (percent identity: 46 %/quer.  
DATE 15-JUN-2001  
ACCESSION AE004043  
NID . . .

SCORES Init1: 474 Initn: 1124 Opt: 1193 z-score: 1293.0 E(): 4.8e-64  
Smith-Waterman score: 1193; 46.7% identity in 439 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITG  
10 20 30 40 49  
AE004043\_7 MYCRRSHLKKPMSHRTHDYWIAHQGTPLHGVLSPGDKSISHRAVMAALADGTSRIDG  
10 20 30 40 50 60

nk603cp4.pep LLEGEDVINTGKAMQAMGARIRKEGDTW-IIDGVGNGGLLAPEAPLDFGNAATGCRLTMG  
50 60 70 80 90 100  
AE004043\_7 FLEAEDTCSTAEILARLGVIETPLSTQRIVHGVGDGLQASHIPLDCGNAGTGMRLLAG  
110 120 130 140 150 160

70 80 90 100 110 120

nk603cp4.pep LVGVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSEDGRLPVTLRGPKTPTPITYR  
110 120 130 140 150 160  
AE004043\_7 LLVAQPFDSVLVGDAASLSKRPMRRVTDPLSQMGARIDTSDDGTPPLRIYGGQLLHGIDFI  
130 140 150 160 170 180

nk603cp4.pep VPMASAVKSAVLLAGLNTPTGITTVEIPIMTRDHEKMLQGFANPTVETDADGVRTIRL  
170 180 190 200 210 220  
AE004043\_7 SPVASAQKSAVLLAGLNTPTGITTVEIPIMTRDHEKMLQGFANPTVETDADGVRTIRL  
190 200 210 220 230

nk603cp4.pep EGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGLIILTQEMGADI  
230 240 250 260 270 280  
AE004043\_7 RGGQRLCATDITIPADFSAAFYLVAAASVPIGSDITLRAVGLNPRRIGLLTVLRIMGANI  
240 250 260 270 280 290

nk603cp4.pep EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFAGATVMNGLE  
290 300 310 320 330 340  
AE004043\_7 VESNRHEGGGEPVVDLRVRYAPLQGTVPEDLVADMIDEFPALFVAAAAGGQTVVSGAA  
300 310 320 330 340 350

nk603cp4.pep ELRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPRDGGKLGNASGAATVLDHRIAM  
350 360 370 380 390 400  
AE004043\_7 ELRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPRDGGKLGNASGAATVLDHRIAM  
360 370 380 390 400

nk603cp4.pep SFLVMGLVSENPTVDDATMIATSFPEFMDLM--AGLGAIELSDTKAA  
410 420 430 440 450  
AE004043\_7 AFSIAGQLSVSTVRIEDVANVATSFDPYETLARSAGFGLVYCDPA  
410 420 430 440 450

nk603cp4.pep  
TREMBL\_MAIN:Q9PB21

ID Q9PB21 PRELIMINARY; PRT; 454 AA.  
AC Q9PB21;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . . .

SCORES Init1: 474 Initn: 1124 Opt: 1193 z-score: 1293.0 E(): 4.8e-64  
Smith-Waterman score: 1193; 46.7% identity in 439 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITG  
10 20 30 40 49  
Q9PB21 MYCRRSHLKKPMSHRTHDYWIAHQGTPLHGVLSPGDKSISHRAVMAALADGTSRIDG  
10 20 30 40 50 60

nk603cp4.pep LLEGEDVINTGKAMQAMGARIRKEGDTW-IIDGVGNGGLLAPEAPLDFGNAATGCRLTMG  
50 60 70 80 90 100  
Q9PB21 FLEAEDTCSTAEILARLGVIETPLSTQRIVHGVGDGLQASHIPLDCGNAGTGMRLLAG  
70 80 90 100 110 120

nk603cp4.pep LVGVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSEDGRLPVTLRGPKTPTPITYR  
110 120 130 140 150 160  
Q9PB21 LLVAQPFDSVLVGDAASLSKRPMRRVTDPLSQMGARIDTSDDGTPPLRIYGGQLLHGIDFI  
130 140 150 160 170 180



nk603cp4.pep 170 180 190 200 210 220  
VPMASAVKSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRL  
Q9PB21 190 200 210 220 230  
SPVASAQIKSAVLLAGLYARNETVVRPHPTRDYTERMLTAFGVDIDVSTGC-----ARL

nk603cp4.pep 230 240 250 260 270 280  
EGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADI  
Q9PB21 240 250 260 270 280 290  
RGGQRLCATDITIPADFSAAFYLAASVPGSDITLRAVLNPRRIGLLTVRLMGANI

nk603cp4.pep 290 300 310 320 330 340  
EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLE  
Q9PB21 300 310 320 330 340 350  
VESNRHEQGSGPVVDLRVRYAPLQGTVPEDLVADMIDEFPALFVAAAAEGQTVVSGAA

nk603cp4.pep 350 360 370 380 390 400  
ELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAM  
Q9PB21 360 370 380 390 400  
ELRVKESDRLAAMVTGLRVLVGVQVDETADGATIHGGPIGHG-----TINSHGDHRIAM

nk603cp4.pep 410 420 430 440 450  
SFLVMGLVSENPVTVDATMIATSFPEFMDLM--AGLGAKIELSDTKAA  
Q9PB21 410 420 430 440 450  
AFSIAGQLSVSTVRIEDVANVATSFPEFMDLM--AGLGAKIELSDTKAA

nk603cp4.pep  
GP\_BCT3:D90914\_87

LOCUS D90914\_87 [D90914]  
DEFINITION Synechocystis sp. PCC6803 complete genome, 16/27, 1991550-2137258;  
ORF\_ID:slr0444.  
DATE 07-FEB-1999  
ACCESSION D90914  
NID . . .

SCORES Initl: 480 Initn: 967 Opt: 1140 z-score: 1235.8 E(): 7.4e-61  
Smith-Waterman score: 1140; 46.3% identity in 443 aa overlap

nk603cp4.pep 10 20 30 40 50  
MLHGASSRPATARKSSG--LSGTVRIPGDKSISHRSFMFGGLASGETRITGLLE  
D90914\_87 10 20 30 40 50  
MALLSLNNHQSHQRLTVNPPAQGVALTGRLRVPGDKSISHRALMLGAIATGETIIEGLLL

nk603cp4.pep 60 70 80 90 100 110  
GEDVINTGKAMQAMGARIRK-EGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLV-  
D90914\_87 60 70 80 90 100 110  
GEDPRSTAHCFRAMEISELNSEKIIVQGRGLQLQEPSTVLDAGNSGTTMRLMLGLLA

nk603cp4.pep 120 130 140 150 160 169  
GVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRV  
D90914\_87 120 130 140 150 160 169  
GQKDCFLTFTVGGDDSLRHRPMSRVIQPLQMGAKIWARSNKGKFAFLAVQGSQK-PIHYHS

nk603cp4.pep 170 180 190 200 210 220 229  
PMSAQVKSALLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLE  
D90914\_87 170 180 190 200 210 220 229  
PIASAQVKSCLLLAGLTTEGDTTTEPALSRDHSERMLQAFGAKLTIDFVT---HSVTVH

nk603cp4.pep 230 240 250 260 270 280 289  
GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADI  
D90914\_87 230 240 250 260 270 280 289  
GPAHLTGQVRVVGPDISSAAFVLAASILPGSELLVENVGINPTRTGVLEVLQMGADIT

D90914\_87 240 250 260 270 280 290  
GPAHLTGQVRVVGPDISSAAFVLAASILPGSELLVENVGINPTRTGVLEVLQMGADIT

nk603cp4.pep 290 300 310 320 330 340 349  
VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEE  
D90914\_87 290 300 310 320 330 340 350  
PENERLVTGEFVADLRVRASHLQCGTFFGGETIIPRLIDEIPILAVAAFAEGTTRIEDAAE

nk603cp4.pep 350 360 370 380 390 400 409  
LRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
D90914\_87 350 360 370 380 390 400 410  
LRVKESDRLAIASELGKMGAKVTEFDDGLEIQG---GSPL---QGAEDVSLTDHRIAMA

nk603cp4.pep 410 420 430 440 450  
FLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
D90914\_87 410 420 430 440  
LAIAALGSGGQTIINRAEAAAIISYPEFFGTGQVAQG

nk603cp4.pep  
GP\_BCT3:X75325\_1

LOCUS X75325\_1 [SSAROAGEN]  
DEFINITION Synechocystis sp. (PCC6803) aroA gene;  
3-phosphoshikimate 1-carboxyvinyltransferase.  
DATE 10-OCT-1994  
ACCESSION X75325  
NID . . .

SCORES Initl: 480 Initn: 967 Opt: 1140 z-score: 1235.8 E(): 7.4e-61  
Smith-Waterman score: 1140; 46.3% identity in 443 aa overlap

nk603cp4.pep 10 20 30 40 50  
MLHGASSRPATARKSSG--LSGTVRIPGDKSISHRSFMFGGLASGETRITGLLE  
X75325\_1 10 20 30 40 50  
MALLSLNNHQSHQRLTVNPPAQGVALTGRLRVPGDKSISHRALMLGAIATGETIIEGLLL

nk603cp4.pep 60 70 80 90 100 110  
GEDVINTGKAMQAMGARIRK-EGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLV-  
X75325\_1 60 70 80 90 100 110  
GEDPRSTAHCFRAMEISELNSEKIIVQGRGLQLQEPSTVLDAGNSGTTMRLMLGLLA

nk603cp4.pep 120 130 140 150 160 169  
GVYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRV  
X75325\_1 120 130 140 150 160 170  
GQKDCFLTFTVGGDDSLRHRPMSRVIQPLQMGAKIWARSNKGKFAFLAVQGSQK-PIHYHS

nk603cp4.pep 170 180 190 200 210 220 229  
PMSAQVKSALLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLE  
X75325\_1 170 180 190 200 210 220 230  
PIASAQVKSCLLLAGLTTEGDTTTEPALSRDHSERMLQAFGAKLTIDFVT---HSVTVH

nk603cp4.pep 230 240 250 260 270 280 289  
GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADI  
X75325\_1 230 240 250 260 270 280 290  
GPAHLTGQVRVVGPDISSAAFVLAASILPGSELLVENVGINPTRTGVLEVLQMGADIT

nk603cp4.pep 290 300 310 320 330 340 349  
VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEE  
X75325\_1 290 300 310 320 330 340 350  
PENERLVTGEFVADLRVRASHLQCGTFFGGETIIPRLIDEIPILAVAAFAEGTTRIEDAAE

300 310 320 330 340 350  
 nk603cp4.pep 350 360 370 380 390 400 409  
 LRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDRHRIAMS  
 X75325\_1 LRVKESDRLAALIASLGKMGAKVTEFDDGLEIQG---GSPL---QGAEVDSLTDHRIAMA  
 360 370 380 390 400 410

410 420 430 440 450  
 nk603cp4.pep FLVMGLVSENPVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 X75325\_1 LAIAALGSGGQTIIINRAEAAAIISYEFFGTLGQVAQG  
 420 430 440

nk603cp4.pep  
 SWISSPROT:AROASYN3

ID AROASYN3 STANDARD; PRT; 447 AA.  
 AC Q59975; Q59974;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 480 Initn: 967 Opt: 1140 z-score: 1235.8 E(): 7.4e-61  
 Smith-Waterman score: 1140; 46.3% identity in 443 aa overlap

nk603cp4.pep 10 20 30 40 50  
 MLHGASSRPATARKSSG--LSGTVRIPGDKSISHRSFPMFGLASGETRITGLLE  
 AROASYN3 MALLSLNNHQSRLQRLTVNPPAQGVALTGRLRVPGDKSISHRALMLGAIATGETIEGLLL  
 10 20 30 40 50 60

nk603cp4.pep 60 70 80 90 100 110  
 GEDVINTGKAMQAMGARIRK-EGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLV-  
 AROASYN3 GEDPRSTAHCFRAMGAIEISELNSEKIIIVQGRGLGQLQEPSTVLDAAGNSGTTMLMLGLLA  
 70 80 90 100 110 120

nk603cp4.pep 120 130 140 150 160 169  
 GYVDFDSTFIGDASLTFRPMGRVNLNPREMGVQVKSE-DGDRLPVTLRGPKTPTPTITYRV  
 AROASYN3 GQKDCFLTFTVGDDSLRHRPMSRVIQPLQMGAKIWARSNKGFAPLAVQGSQK-PIHYHS  
 130 140 150 160 170

nk603cp4.pep 170 180 190 200 210 220 229  
 PMASQVKSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLE  
 AROASYN3 PIASQVKSCLLLAGLTTEGDTTVTEPALSRDHSEMLQAFGAKLTIDPVT---HSVTYH  
 180 190 200 210 220 230

nk603cp4.pep 230 240 250 260 270 280 289  
 GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIE  
 AROASYN3 GPAHLTGQVRVVGDISSAAFVLAASILPGSELLVENVGINPTRTGVLEVLAAQMGADIT  
 240 250 260 270 280 290

nk603cp4.pep 290 300 310 320 330 340 349  
 VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDIEYPIILAVAAAFAGATVMNGLEE  
 AROASYN3 PENERLVTGEVADLRVRASHLQCTFGGEIIPRLIDEIPILAVAAAFAGETTRIEDAAE  
 300 310 320 330 340 350

nk603cp4.pep 350 360 370 380 390 400 409  
 LRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDRHRIAMS  
 AROASYN3 LRVKESDRLAALIASLGKMGAKVTEFDDGLEIQG---GSPL---QGAEVDSLTDHRIAMA  
 360 370 380 390 400 410

410 420 430 440 450  
 nk603cp4.pep FLVMGLVSENPVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROASYN3 LAIAALGSGGQTIIINRAEAAAIISYEFFGTLGQVAQG  
 420 430 440

nk603cp4.pep  
 GP\_BCT3:Z29339\_1

LOCUS Z29339\_1 (DNEPS3PS)  
 DEFINITION D.nodosus (VCS1001) aroA gene for enolpyruvylshikimate 3-phosphate  
 synthase.  
 DATE 25-OCT-1994  
 ACCESSION Z29339  
 NID . . .

SCORES Init1: 442 Initn: 961 Opt: 1056 z-score: 1145.2 E(): 8.2e-56  
 Smith-Waterman score: 1056; 42.0% identity in 438 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGSLGTVRIPGDKSISHRSFPMFGLASGETRITGLLEGEDEVINTG  
 Z29339\_1 MMTNIWHTAPVSALSGEITICGDKSMHRALLAALAEQTEIRGFLACADCLATR  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110 120  
 KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
 Z29339\_1 QALRALGVDIQREKEIVTIRGVGFLGLQPPKAPLNMQNSGTSMLLAGILAAQRFESVLC  
 60 70 80 90 100 110

nk603cp4.pep 130 140 150 160 170 180  
 GDASLTFRPMGRVNLNPREMGVQVKSEGDRLPVTLRGPKTPTPTITYRVPMASQVKSASV  
 Z29339\_1 GDESLEKRPQMORIITPLVQMGAKIVSHSNFTAPLHISG-RPLTGIDYALPLPSAQLKSCL  
 120 130 140 150 160 170

nk603cp4.pep 190 200 210 220 230 240  
 LLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRKLTGQVID  
 Z29339\_1 ILAGLLADGTTTRLHTCGISRDRHTERMLPLFGGALEIKKEQ-----IIVTGGQKLHGCULD  
 180 190 200 210 220 230

nk603cp4.pep 250 260 270 280 290 300  
 VPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLAGGED  
 Z29339\_1 IVGDLASAAFFMVAALAPRAEVVIRNVGINPTAAIITLQKMGGRIELHHQRFWGAEP  
 240 250 260 270 280 290

nk603cp4.pep 310 320 330 340 350 360  
 VADLRVRSSTLKGVTVPEDRAPSMIDIEYPIILAVAAAFAGATVMNGLEELRVKESDRLSA  
 Z29339\_1 VADIVVYHSLRGITVAPEWIANAIIDELPIFFIAACAEGTTFVGNLSSELRVKESDRLLAA  
 300 310 320 330 340 350

nk603cp4.pep 370 380 390 400 410 420  
 VANGKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDRHRIAMSFLVMGLVSENP  
 Z29339\_1 MAQNLQTLGVACDVGADFIHIYGRSDRQFLP-----ARVNSFGDHRHRIAMSLAVAGVRAAGE  
 360 370 380 390 400

nk603cp4.pep 430 440 450  
 VTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 Z29339\_1 LLIDDGAVAAVSMFQPRDFAAAIAGMNVGEKDAKNCHD  
 410 420 430 440

nk603cp4.pep

```

ID  AROA_BACNO          STANDARD;          PRT;    443 AA.
AC  Q46550;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

```

```

              10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
              |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
AROABACNO    MMTNIWHTAPVSALSGETITCGDKSMHSRALLLALAEQGTEIRGFLACADCLATR
              10      20      30      40      50

```

```

              70      80      90     100     110     120
nk603cp4.pep KAMQAMGARIRKEGDWTIIDGVNGGLLAPEAPLDFGNAATGCRLTMGLVGVDYDFTFI
               :|::||:|::| || | | ::||::|::| | |::|::|::|::|::|::|::|::
AROABACNO    QALRALGVDIQREKEIVTIRGVFLGLQPPKAPLNMMNSGTSMRLLAGILAAQRFSVLC
               60      70      80      90     100     110

```

```
nk603cp4.pep      GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPIITYRVPMASAQVKS AV
                  ||| | | | | | : | : | : | : | : | : | : | : | : | : | : | : 
AROABACNO         GDESLEKRPQMRIITPLVMGAIVSHSNFTAPLHISGRPLTGIDYALPLPSAQLKSC L
                  120       130       140       150       160       170
```

```

          190      200      210      220      230      240
nk603cp4.pep LLAGLNTPGITTVEIPIINTRDHTEKMLQGFGANPIVETDADGVRTIRLEGRGKLTGQVID
               :|||:|||:|:|:::|||:|||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
ARO_A_BACNO ILAGLLADGTTRLHTCGISRDMTERMLPLFGGALEIKKEQ-----IIVTGGQKLHGCVLD
              180      190      200      210      220      230

```

```

                250      260      270      280      290      300
nk603cp4.pep VPGDPSSTAPFLVAALLVPGSDVTILNVLNNPRTGLILTQEMGADIEVINPRLAGGED
               : || |::| ::||::| ::||::| ::||::| ::||::| ::||::| ::||::|
AROA_BACNO   IVGDLASAAAFMVAAALIAPRAEVIIRNVGINPTRAAIITLQKMGRIELHHQRFWGAEP
              240      250      260      270      280      290

```

[illegible]

```

          370          380          390          400          410          420
nk603cp4.pep VANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFVLMGLVSENP
               : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AROABACNO    MAQNQLTGLVACDVGADFIHIYGRSDRQFLP---ARVNSFGDHRIRAMSLAVAGVRAAGE
               360          370          380          390          400

```

nk603cp4.pep                    430                    440                    450  
                                  : : | : : : | : : | : | : : : : : : : : : : : :  
ARO\_A\_BACNO            LLIDDGAVAAVSMQPFQDFAAAGMNVGKDAKNCHD  
                                  410                    420                    430                    440

nk603cp4.pep  
TREMBL MAIN:09KCA6

```

ID   Q9KCA6             PRELIMINARY;          PRT;   431 AA.
AC   Q9KCA6;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5-

```

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPIGDKSISHRSPFMGGSLASGETRITGLLEGEDVINTG  
Q9KCA6 MENKTVIPHAKGLGTIKVPDGKSIHRAVMFGALAGKTTTVEGFPLPGADCLSTI

```

              70      80      90      100      110      120
nk603cp4.pep KAMQANGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRLTMGLVGVDYDFDSTFI
              :::::  :::::  :::::  :::::  :::::  :::::
Q9KCA6        SCFQKLGVSIEQAEEERVTVKGGKWDGLREPSDILDVNGSGTTTRLILGLSTLPFHSVII
              60      70      80      90      100      110

```

```

              130      140      150      160      170      179
nk603cp4.pep GDASLTTRPMGRVLNPLREMGVQVKSED-GDRLPVTLRGPKTPTPTITYRVPMASAOVKSA
               |||::|||::|:|::|::|::|::|:|::|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Q9KCA6        GDESIGKRPMKRVTLEPLKSMGAQIDGRDHGNLTPLSIRGGQLKG-IDFHSPVASAQMKSAA
              120      130      140      150      160      170

```

```

      180       190       200       210       220       230       239
nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHTKMLQGFANPTVETDADGVRTIRLEGRGLTGQVI
              :|||:::||:|||||::|::|::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Q9KCA6        ILLAGLRAEGKTSVTPEAKTRDHTERMLEAFGVN----IEKDGL-TVSIEGGOMLTGHV
              ||| |   ||| |   ||| |   ||| |   ||| |   ||| |   ||| |   ||| |

```

```

      240       250       260       270       280       290       299
nk603cp4.pep DVPGDPSSTAFPLVAALLVPGSDVTILNVLMPRTGLTLTQEMGADIEVINPRLAGGE
              ||||| |::| |||::| :|| ::|||::|::|::|::|::|::|::|::|::|
Q9KCA6        VVPGDISAAFFLVAGAMVPHSRITLTNVGINPTRAGILEVLVKMGATLMENERVQGGE
            230         240         250         260         270         280

```

```

          300      310      320      330      340      350      359
nk603cp4.pep DVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAF AEGATVMNGLEELRVKESDRLS
          |||| ::||::||: | :||| |||:| :| :|:: |||:||::||
Q9KCA6      FVADLTETSVLQGV EIGGDIIPRLIDEIPII AVLATQASGRVTIKDAEELVKKETNRID
          290      300      310      320      330      340

```

```

          360      370      380      390      400      410      419
nk603cp4.pep AVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLVLMGLVSEN
          :|::| :|:: :|:::| || :|:::| ||| :|:: :|:::|
Q9KCA6       TVVSELTKLGASIHATDDGMIIEGPTPLKG-----GVTVSSHGDHRIGMATAALAEK
          350      360      370      380      390      400

```

```

          420      430      440      450
nk603cp4.pep PVTVD DATMIATSFPEFMDLMAGLGAKIELSDTKAA
              |||::: ||::: ||:::
Q9KCA6        PVTVEGTEIAVSYSPSFFDHLDRLKSE
              410      420      430

```

nk603cp4.pep  
GP BCT2:AP001512 254

LOCUS AP001512\_254 [AP001512]  
DEFINITION Bacillus halodurans genomic DNA, section 6/14;  
BH1667.  
DATE 10-JAN-2001  
ACCESSION AP001512  
NID

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

nk603cp4.pep MLHGASSRPATARKSSGLSVTRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_AQUAE MKKIEKIKRVKGLRVPSSDKSITHRAFILGALASGETLVRKPLISGDTLATL

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRITMGLGVYDFDSTFI  
 AROA\_AQUAE EILKAIRTKVREGKEEVLEIGR-NYTFLEPHDVLDAKNSGTTARIMSGVLSTQPFFSVLT

```

nk603cp4.pep      130      140      150      160      170      179
                  GDASLTKRPMGRVLNPLRREMGVQVKS-EDGDRLPVTLRGPKTPTPTITYRVFPMASAQVKSA
                  |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
AROA_AQUAE        GDES LKNRMLRVVEPLRREMGAKIDGREEGNKLPPIAIRGGNLKG-ISYFNKKSSAQVKSA
                  120      130      140      150      160      170

nk603cp4.pep      180      190      200      210      220      230      239
                  VLLAGLNTPGITVTIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEBGRGKLTGQVI
                  :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
AROA_AQUAE        LLLAGLRARGMTTEVVPEYPLSRDHTERMLKLGAEVITIEPERG-HIVKIKGGQELQGTEV
                  180      190      200      210      220

nk603cp4.pep      240      250      260      270      280      290      299
                  DVFGDPSSSTAFFPLVAALLVPGSDVTILNVLNMPNTRTGLILTQEMGADIEVINPRLAGGE
                  :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
AROA_AQUAE        YCPADPSSAAAYFAALATLAPEGEIRLKEVLLNPNTRDGFYRKLIEMGGDISFENYRELSNE
                  230      240      250      260      270      280

nk603cp4.pep      300      310      320      330      340      350
                  DVADLVRVS-STLKGVTVPEDRAPSMIDEPYILAVAAAFAEGATVMNGLEELRVKESDRL
                  :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
AROA_AQUAE        PMADLVVRPVDNLNPKVKVSPPEVPTLIDEIPTLAVLMAFADGVSEVKAGELRYKESDRI
                  290      300      310      320      330      340

nk603cp4.pep      360      370      380      390      400      410
                  SAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLVMGLVSE
                  :|:::|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
AROA_AQUAE        KAIVTNLRKLGQVREFEDGFAIHGTKEIKG-----GVIETPKDRIAMAFAVLGLVVE
                  350      360      370      380      390      400

nk603cp4.pep      420      430      440      450
                  NPVTVD DATMIATSFPEFMDLMAGLGAKIELSDTKAA
                  :|:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
AROA_AQUAE        BEVIIDHPCEVTVSYPEFNEIDILKVVEF
                  410      420      430

```

nk603cp4.pep  
GP\_BCT1:AE000744\_15

```

LOCUS      AE000744_15 [AE000744]
DEFINITION Aquifex aeolicus section 76 of 109 of the complete genome.
DATE       25-MAR-1998
ACCESSION  AE000744
NID
ORGANISM   Aquifex aeolicus . . .

```

SCORES Init1: 449 Initn: 561 Opt: 1011 z-score: 1096.8 E(): 4.1e-53  
Smith-Waterman score: 1011; 40.9% identity in 423 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVTRIPGDKSISHRSFMFGGLASGETRITGLLEGGDVINTG
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AE000744_15 MKKIEKIKRVKGELRVPSDKSITHRAFILGALASGETLVRKPLISGDTLATL
      10      20      30      40      50

      70      80      90      100      110      120
nk603cp4.pep KAMQAMGARIRKEGDTWIDGVNGGGLLAPEAPLDFGNAAATGCRLTMGVLGVYDFDSTFI
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AE000744_15 EILKAIKTVREGKEEVLIEGR-NYTFLEPHDVLDAKNSGTTARINMSGVLSTQPPFSVL
      60      70      80      90      100      110

      130      140      150      160      170      179
nk603cp4.pep GDASLTKEPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTITYRVPMAAQVKS
      | | | | : | | | : | | | : | | | : | | | : | | | : | | | :
AE000744_15 GDES LKNR FMLRVVEPLREMGAKIDGREGNKLP I A I R G N L K G - I S Y F N K S S A Q V K S
      120      130      140      150      160      170

      180      190      200      210      220      230      239

```

NK603 analysis  
Page 15 of 120

420                      430                      440                      450

nk603cp4.pep  
SWISSPROT:AROA\_STREN

ID AROA\_STRPN STANDARD; PRT; 427 AA.  
AC Q9S400;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 386 Initn: 889 Opt: 963 z-score: 1045.0 E(): 3.1e-50  
Smith-Waterman score: 1045; 42.6% identity in 430 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AROA\_STRPN MKLKTNIIRHLHGSIIRVPGDKSISHRSIIFGSLAEGETKVIYDILRGEDVLSTM

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
AROA\_STRPN QVFRDLGVEIEDKDGVIITVQGVGMAGLKAPQNALNMGNSGTSIRLISGVLGADFEVEMF

nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRVPMASAOVKSA  
AROA\_STRPN GDDSLSKRPMDRVTLPLKMGVSIISGQTERDLPPLRLKGTKNLRPIHYELPIASAOVKSA

nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI  
AROA\_STRPN LMFALQAKGESVIIIEKEYTRNHTEDMLKQFGGHLV---DG-KKITVQGPQKLTGQKV

nk603cp4.pep DVPGDPSSAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEV--INPLAG  
AROA\_STRPN VVPGDISSAAFVWLAGLIAFNSRLVLQNVGINETRTGIIDVIRAMGGKLEITEIDPVAKS

nk603cp4.pep GEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDR  
AROA\_STRPN ---ATLIVESSDLKGTIEIGGALIPRLIDELPIIALLATQAQGVTVIKDAEELVKKETDR

nk603cp4.pep LSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDRHRIAMSFVLMGL-V  
AROA\_STRPN IQVVADALNSMGADITPTADGMIKKGK-----SALHGARVNTFGDHRIGMMTAIAALLV

nk603cp4.pep SENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AROA\_STRPN ADGEVELDRAEAINTSYPSFDDLESIHG

nk603cp4.pep  
GP\_BCT2:AF169483\_1

LOCUS AF169483\_1 [AF169483]  
DEFINITION Streptococcus pneumoniae 5-enolpyruvylshikimate-3-phosphate  
synthase (aroA) gene, complete cds;  
sixth enzyme in the shikimate pathway.  
DATE 27-JAN-2000  
ACCESSION AF169483 . . .

SCORES Init1: 385 Initn: 891 Opt: 962 z-score: 1043.9 E(): 3.6e-50  
Smith-Waterman score: 1044; 42.6% identity in 430 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AF169483\_1 MKLKTNIIRHLHGSIIRVPGDKSISHRSIIFGSLAEGETKVIYDILRGEDVLSTM

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
AF169483\_1 QVFRDLGVEIEDKDGVIITVQGVGMAGLKAPQNALNMGNSGTSIRLISGVLGADFEVEMF

nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRVPMASAOVKSA  
AF169483\_1 GDDSLSKRPMDRVTLPLKMGVSIISGQTERDLPPLRLKGTKNLRPIHYELPIASAOVKSA

nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI  
AF169483\_1 LMFALQAKGESVIIIEKEYTRNHTEDMLKQFGGHLV---DG-KKITVQGPQKLTGQKV

nk603cp4.pep DVPGDPSSAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEV--INPLAG  
AF169483\_1 VVPGDISSAAFVWLAGLIAFNSRLVLQNVGINETRTGIIDVIRAMGGKLEITEIDPVAKS

nk603cp4.pep GEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDR  
AF169483\_1 ---ATLIVESSDLKGTIEIGGALIPRLIDELPIIALLATQAQGVTVIKDAEELVKKETDR

nk603cp4.pep LSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDRHRIAMSFVLMGL-V  
AF169483\_1 IQVVADALNSMGADITPTADGMIKKGK-----SALHGARVNTFGDHRIGMMTAIAALLV

nk603cp4.pep SENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AF169483\_1 ADGEVELDRAEAINTSYPSFDDLESIHG

nk603cp4.pep  
GP\_BCT1:AE006404\_9

LOCUS AE006404\_9 [AE006404]  
DEFINITION Lactococcus lactis subsp. lactis IL1403 section 166 of 218 of the  
complete genome;  
EVIDENCE EXPERIMENTAL PMID:7823907 BIO01.02  
Aromatic amino-acid family.  
DATE 14-MAY-2001 . . .

SCORES Init1: 412 Initn: 936 Opt: 962 z-score: 1043.9 E(): 3.6e-50  
Smith-Waterman score: 1034; 42.1% identity in 435 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AE006404\_9 MKLKINSQGLKGLKVPDGSISHRSIMFGSIAGKGTVIYDILRGEDVLSTI

```

      10      20      30      40      50
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVDYDFDSTFI
      70      80      90     100     110     120
AE006404_9 EAFRAMGVEIEDKGEVITVHGKGISELKAPEKALDMGNSGTSTRLLSGILAGLPFETTLF
      60      70      80      90     100     110

      130     140     150     160     170     179
nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSE-DGDRLPVTLRGPKTPTPTITYRVPMASAQVKS
      110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179
AE006404_9 GDDSLSKRPMDRVATPLQLMGAEITGQTDKVKLPMTIKGSTHLKAIDYVLPVASAQVKS
      120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179

      180     190     200     210     220     230     239
nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHEKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI
      170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239
AE006404_9 VIFPAALQAEGLTKVVEKEKTRSHTEMLVQFGGELKV-SD----KTILVPGGQKLVGQKV
      180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239

      240     250     260     270     280     290     299
nk603cp4.pep DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLAGGE
      230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299
AE006404_9 VVPGDISAFAFWLVAALVVENSELILENVGVNETRTGIIIEVIQAMGGQLEILEQDNVA--
      230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299

      300     310     320     330     340     350     359
nk603cp4.pep DVADLRVRSSTLKGVTVPEDRAPSMIDEYPIILAVAAFAEGATVMNGLEELRVKESDRLS
      290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359
AE006404_9 KAATLKVKASQLKGTEISGDLIPRLIDELPIIALLATQAQGTIIRDAELKVKETDRIA
      290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359

      360     370     380     390     400     410     419
nk603cp4.pep AVANGLKLVGDCDEGETSLVVRGPRDGKGLGNASGAATHLDHRIAMSFVLMGLVSEN
      350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419
AE006404_9 VVANALNSMGAKIQPTDDGMIIQG---GTLK-HAPENSINTLGDHRIAMMAIAALLVKN
      350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419

      420     430     440     450
nk603cp4.pep -PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
AE006404_9 GEIELERAEAIQTSYPSFFDLEQLSENI
      410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500

```

nk603cp4.pep  
SWISSPROT:AROA\_LACLA

```

ID   AROA_LACLA      STANDARD;          PRT;   430 AA.
AC   Q9CEU0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

```

SCORES Initl: 412 Initn: 936 Opt: 962 z-score: 1043.9 E(): 3.6e-50  
Smith-Waterman score: 1034; 42.1% identity in 435 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVIRPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
      50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
AROA_LACLA      MKLKINSQGLKGRLLKVPFGDKSISHRSIMFGSIAGKTVIYDILRGEDVLSTI
      10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500

```

```

      70      80      90     100     110     120
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVDYDFDSTFI
      60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
AROA_LACLA      EAFRAMGVEIEDKGEVITVHGKGISELKAPEKALDMGNSGTSTRLLSGILAGLPFETTLF
      60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500

```

```

      130     140     150     160     170     179
nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSE-DGDRLPVTLRGPKTPTPTITYRVPMASAQVKS
      120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 
```



```

* AAL00033 180 190 200 210 220 230
LMFAALQAKGESVIEKECTRNHTEDMLKQFGGHLV---DG-KKITVQGPQKLTGQKV

nk603cp4.pep 240 250 260 270 280 290
DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTQEMGADIEV--INPRLAG
VVPGDISSAAFVWVAGLINPNSHLVQNVGINETRTGIDVIRAMGGKLEVFIDEVAKS
AAL00033 240 250 260 270 280 290
VVPGDISSAAFVWVAGLINPNSHLVQNVGINETRTGIDVIRAMGGKLEVFIDEVAKS

nk603cp4.pep 300 310 320 330 340 350
GEDVADLRVRSSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDR
AAL00033 300 310 320 330 340 350
GEDVADLRVRSSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDR

nk603cp4.pep 360 370 380 390 400 410
LSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLVGLV-
AAL00033 360 370 380 390 400 410
LSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLVGLV-

nk603cp4.pep 420 430 440 450
SENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AAL00033 420 430 440 450
SENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA

nk603cp4.pep
GP_BCT3:X78413_2

LOCUS X78413_2 [LLTYRAPH]
DEFINITION L.lactis tyrA, aroA, aroK and pheA genes.
DATE 23-FEB-1995
ACCESSION X78413
NID
ORGANISM Lactococcus lactis . . .

SCORES Init1: 400 Initn: 964 Opt: 952 z-score: 1033.1 E(): 1.4e-49
Smith-Waterman score: 1033; 41.6% identity in 435 aa overlap

10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
X78413_2 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG

10 20 30 40 50 60
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI
X78413_2 KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI

10 20 30 40 50 60
nk603cp4.pep GDASLTKRPMGRVNLPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRVPMASAVKSKA
X78413_2 GDASLTKRPMGRVNLPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRVPMASAVKSKA

10 20 30 40 50 60
nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI
X78413_2 VLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI

10 20 30 40 50 60
nk603cp4.pep DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTQEMGADIEVINPRLAGGE
X78413_2 DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTQEMGADIEVINPRLAGGE

```

```

230 240 250 260 270 280
nk603cp4.pep DVADLRVRSSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLS
X78413_2 DVADLRVRSSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLS

300 310 320 330 340 350
nk603cp4.pep AVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLVGLVSEN
X78413_2 AVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLVGLVSEN

360 370 380 390 400 410 419
nk603cp4.pep -PVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
X78413_2 -PVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA

420 430 440 450
nk603cp4.pep GEIELERAETIQTSYPSFFDDLEKLSGNL
X78413_2 GEIELERAETIQTSYPSFFDDLEKLSGNL

nk603cp4.pep
SWISSPROT:AROA_LACLC

ID AROA_LACLC STANDARD; PRT; 430 AA.
AC P43905;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 400 Initn: 964 Opt: 952 z-score: 1033.1 E(): 1.4e-49
Smith-Waterman score: 1033; 41.6% identity in 435 aa overlap

10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AROA_LACLC MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG

70 80 90 100 110 120
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI
AROA_LACLC KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI

130 140 150 160 170 179
nk603cp4.pep GDASLTKRPMGRVNLPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRVPMASAVKSKA
AROA_LACLC GDASLTKRPMGRVNLPLREMGVQVKSE-DGDRLPVTLRGPKTPTPIYRVPMASAVKSKA

180 190 200 210 220 230 239
nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI
AROA_LACLC VLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI

240 250 260 270 280 290 299
nk603cp4.pep DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTQEMGADIEVINPRLAGGE
AROA_LACLC DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTQEMGADIEVINPRLAGGE

```



LOCUS Z99115\_200 [BSUB0012]  
DEFINITION Bacillus subtilis complete genome (section 12 of 21): from 2195541 to 2409220;  
DATE 26-NOV-1997  
ACCESSION Z99115  
NID . . .

SCORES Init1: 525 Initn: 943 Opt: 926 z-score: 1005.1 E(): 5.3e-48  
Smith-Waterman score: 1037; 41.2% identity in 434 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
Z99115\_200 MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAAGTTTVKNFLPGADCLSTI

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGVYDFDSTFI  
Z99115\_200 DCFRKMVGVIHQSSSDVVIHGKIDALKEPESLLDVGNSTGTTIRLMGLAGRPFYSAVA

nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSSEDGRL-FVTLRGPKTPTPIYRVPMAAQVKSA  
Z99115\_200 GDESIARKPMKRVTPELKKMGAKIDGRAGGEFTPLSVSGASLKG-IDYVSPVASAQIKSA

nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHEKMLQGFGANPTVETDADGVRTIRLEGRKLTGQVI  
Z99115\_200 VLLAGLQAEGETTVTEPHKSRDHTERMLSAFG---VKLSEDQT-SVSIAGGQKLTAAADI

nk603cp4.pep DVPGDPSSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLAGGE  
Z99115\_200 FVPGDISAFAFLAAGAMVNSRIVLKNVGLNPTRTGIIDVLQNMGALEIKPSADSGAE

nk603cp4.pep DVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLS  
Z99115\_200 PYGDLIIETSSSLKAVEIGGDIIPRLIDEIPIIALLATQAEGETTVIKDAAELKVKETNRID

nk603cp4.pep AVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAATHLDHRIAMSFLVMGLVSEN  
Z99115\_200 TVVSELRLKLGAEIETADGMKVYKQTLKG-----GAAVSSHGDHRIAMGLGIASCITEE

nk603cp4.pep PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Z99115\_200 PIEIEHTDAIHVSYPTEFFHLNKLKSKKS

nk603cp4.pep  
GP\_BCT2:M80245\_18

LOCUS M80245\_18 [BACVARGNS]  
DEFINITION B.subtilis dbpA, mtr(A,B), gerC(1-3), ndk, cheR, aro(B,E,F,H),  
trp(A-F), hisH, and tyrA genes, complete cds.  
DATE 17-AUG-1994  
ACCESSION M80245  
NID . . .

SCORES Init1: 525 Initn: 943 Opt: 926 z-score: 1005.1 E(): 5.3e-48  
Smith-Waterman score: 1037; 41.2% identity in 434 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
M80245\_18 MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAAGTTTVKNFLPGADCLSTI

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGVYDFDSTFI  
M80245\_18 DCFRKMVGVIHQSSSDVVIHGKIDALKEPESLLDVGNSTGTTIRLMGLAGRPFYSAVA

nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSSEDGRL-FVTLRGPKTPTPIYRVPMAAQVKSA  
M80245\_18 GDESIARKPMKRVTPELKKMGAKIDGRAGGEFTPLSVSGASLKG-IDYVSPVASAQIKSA

nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHEKMLQGFGANPTVETDADGVRTIRLEGRKLTGQVI  
M80245\_18 VLLAGLQAEGETTVTEPHKSRDHTERMLSAFG---VKLSEDQT-SVSIAGGQKLTAAADI

nk603cp4.pep DVPGDPSSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLAGGE  
M80245\_18 FVPGDISAFAFLAAGAMVNSRIVLKNVGLNPTRTGIIDVLQNMGALEIKPSADSGAE

nk603cp4.pep DVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLS  
M80245\_18 PYGDLIIETSSSLKAVEIGGDIIPRLIDEIPIIALLATQAEGETTVIKDAAELKVKETNRID

nk603cp4.pep AVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAATHLDHRIAMSFLVMGLVSEN  
M80245\_18 TVVSELRLKLGAEIETADGMKVYKQTLKG-----GAAVSSHGDHRIAMGLGIASCITEE

nk603cp4.pep PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
M80245\_18 PIEIEHTDAIHVSYPTEFFHLNKLKSKKS

nk603cp4.pep  
SWISSPROT:AROA\_BACSU

ID AROA\_BACSU STANDARD; PRT; 428 AA.  
AC P20691;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 525 Initn: 943 Opt: 926 z-score: 1005.1 E(): 5.3e-48  
Smith-Waterman score: 1037; 41.2% identity in 434 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AROA\_BACSU MKRDKVQTLHGEIHIPGDKSISHRSVMFGALAAAGTTTVKNFLPGADCLSTI

70 80 90 100 110 120  
 nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
 AROA\_BACSU DCFRKMGVHIEQSSSDVVIHGKGIDALKEPESLLDVNSGTTIRLMGLILAGRPFYSAVA  
 60 70 80 90 100 110

130 140 150 160 170 179  
 nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSEGDRL-PVTLRGPKTPTPTITRVPMASAVQVKS  
 AROA\_BACSU GDESIAKRPMKRVTEPLKMGAKIDGRAGGFTPLSVSGASLKG-IDVVSFVASAIIKSA  
 120 130 140 150 160 170

180 190 200 210 220 230 239  
 nk603cp4.pep VLLAGLNTPGITTVIEPIIMTRDHTKMLQGGFANPTVETDADGVRTIRLEGRGKLTGQV  
 AROA\_BACSU VLLAGLQAEGTITVTEPHKSRDHTERMLSAFG---VKLSEDQT-SVSIAGGOKLTAADI  
 180 190 200 210 220

240 250 260 270 280 290 299  
 nk603cp4.pep DVPGPSSTAFFLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIEVINPRLAGG  
 AROA\_BACSU FVPGDISAAFFLAAGAMVPSRIVLKNVGLNPTRTGTIIDVLQNMGALEIKPSADSGAE  
 230 240 250 260 270 280

300 310 320 330 340 350 359  
 nk603cp4.pep DVADLRVRSSSTLKGVTVPEDRAPSMIDEYPIILAVAAFAEGATVMNGLEELRVKESDR  
 AROA\_BACSU PYGDLIIETSSLKAVEIGGDIIPRLIDEPIIALLATQAEGTTVIKDAEELKVKETNRID  
 290 300 310 320 330 340

360 370 380 390 400 410 419  
 nk603cp4.pep AVANGLKLVGDCDEGETSLVVRGPRDGKGLGNASGAAVATHLDHRIAMSLVLMGLVSEN  
 AROA\_BACSU TVVSELRLKLGABIEPTADGMKVYKQTLKG---GAAVSSHGDHRIAMSLGIASCTEE  
 350 360 370 380 390 400

420 430 440 450  
 nk603cp4.pep PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROA\_BACSU PIEIEHTDAIHVSYPTEFFHLNKLKSKS  
 410 420

nk603cp4.pep  
 GP\_BCT1:AE001715\_8

LOCUS AE001715\_8 [AE001715]  
 DEFINITION Thermotoga maritima section 27 of 136 of the complete genome;  
 similar to GB:M80245 SP:P20691 PID:143816  
 GB:AL009126 percent identity: 66.27; identified by  
 sequence similarity; putative.  
 DATE 02-JUN-1999 . . .

SCORES Init1: 326 Initn: 788 Opt: 834 z-score: 906.0 E(): 1.7e-42  
 Smith-Waterman score: 921; 40.2% identity in 420 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AE001715\_8 MLVPPDKSITHRALILSALAETESTLYNLRCLDTERTH  
 10 20 30 40

70 80 90 100 110 120  
 nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
 AE001715\_8 DILEKLGTRF--EGD-WEKMKVFPKPFABIEPLFCGNSGTTTIRLMGSLVASYEMFTVLY  
 50 60 70 80 90  
 130 140 150 160 170 180

nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITRVPMASAVQVKS  
 AE001715\_8 GDPSLSRPMPRRVIEPLEMMGARFMAQQNNYLPMAIKGNHL-SGISYKTFVASAVQVKS  
 100 110 120 130 140 150

190 200 210 220 230 240  
 nk603cp4.pep LLAGLNTPGITTVIEPIIMTRDHTKMLQGGFANPTVETDADGVRTIRLEGRGKLTGQV  
 AE001715\_8 LLAGLRASGRTIVIEPAKSRDHTERMLKNLG---VFVEVEGTRVV-LEP-ATFRGFTMK  
 160 170 180 190 200 210

250 260 270 280 290  
 nk603cp4.pep VPGDPSSTAFFLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIE--VINPRLAGG  
 AE001715\_8 VPGDISAAFFVVLGAHPNARITVTDVGLNPTRTGLLEVMKLMGANLEWITEENL---  
 220 230 240 250 260

300 310 320 330 340 350  
 nk603cp4.pep EDVADLRVRSS-TLKGVTVPEDRAPSMIDEYPIILAVAAFAEGATVMNGLEELRVKESDR  
 AE001715\_8 EPIGTVRVETSPNLKGVVVPEHLVPLMIDELPLVALLGVFAEGETVVRNAEELRKESDR  
 270 280 290 300 310 320

360 370 380 390 400 410  
 nk603cp4.pep LSAVANGLKLVGDCDEGETSLVVRGPRDGKGLGNASGAAVATHLDHRIAMSLVLMGLVSEN  
 AE001715\_8 IRLVENFKRLGVEIEEFKDGKIVGKQSIK-----GSVDPEGDHRLMAMLSIAGLV  
 330 340 350 360 370 380

420 430 440 450  
 nk603cp4.pep ENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AE001715\_8 EEGVDVKDHECAVVSFPNFYELLERVVIS  
 390 400 410

nk603cp4.pep  
 SWISSPROT:AROA\_THEMEA

ID AROA\_THEMEA STANDARD; PRT; 410 AA.  
 AC Q9WYI0;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 326 Initn: 788 Opt: 834 z-score: 906.0 E(): 1.7e-42  
 Smith-Waterman score: 921; 40.2% identity in 420 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_THEMEA MLVPPDKSITHRALILSALAETESTLYNLRCLDTERTH  
 10 20 30 40

70 80 90 100 110 120  
 nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
 AROA\_THEMEA DILEKLGTRF--EGD-WEKMKVFPKPFABIEPLFCGNSGTTTIRLMGSLVASYEMFTVLY  
 50 60 70 80 90

130 140 150 160 170 180  
 nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITRVPMASAVQVKS  
 AROA\_THEMEA GDPSLSRPMPRRVIEPLEMMGARFMAQQNNYLPMAIKGNHL-SGISYKTFVASAVQVKS  
 100 110 120 130 140 150

190 200 210 220 230 240  
 nk603cp4.pep LLAGLNTPGITTVIEPIIMTRDHTKMLQGGFANPTVETDADGVRTIRLEGRGKLTGQV  
 190 200 210 220 230 240



NK603 analysis  
Page 24 of 120

DEFINITION Staphylococcus aureus dehydroquinase synthase (aroB) gene, 3' end  
cds; 3-phosphoshikimate-1-carboxyvinyltransferase (aroA) gene,  
complete cds; ORF3, complete cds;  
DATE 21-OCT-1993  
ACCESSION L05004 . . .

SCORES Init1: 271 Initn: 526 Opt: 765 z-score: 831.2 E(): 2.5e-38  
Smith-Waterman score: 765; 34.2% identity in 424 aa overlap

```
10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
L05004_2 MVNEQIIDISGPLKGEIEVPGDKSMTHRAIMLASLAEGVSTIYKPLLGEDCRRTM
```

```
70 80 90 100 110 120
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI
L05004_2 DIFRHLGVEIKEDDEKLVTSPGYQ-VNTPHQVLYTGNSGTTTLLAGLLSGLGNSVLS
```

```
130 140 150 160 170 180
nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPIITYRVEMASAOVKSVA
L05004_2 GDVSIKGRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPIITYRVEMASAOVKSVA
```

```
190 200 210 220 230 240
nk603cp4.pep LLAGLNTPGITTVIEPIIMTRDHEKMLQGFANPTVETDADGVRTIRLEGRGKLTGQVID
L05004_2 LFAFLSKPTIIEKELDVSRNHTETMPKHF--NIPIEAEGLSINTTPEAIR-YIKPADFH
```

```
250 260 270 280 290 300
nk603cp4.pep VPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIEVINPRLAGGED
L05004_2 VPGDISSAAFFIVAALITPGSDVTIHNVGINQTRSGIIDIIVKMGNIQLFN-QTTGAEP
```

```
310 320 330 340 350 359
nk603cp4.pep VADLRVR-SSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLS
L05004_2 TASIRIQYTPMLQPTIEGELVPAIDELPVIALCTQAVGTSTIKDAEELVKETNRID
```

```
360 370 380 390 400 410 419
nk603cp4.pep AVANGLKLNVDCEGETSLVVRGPRDGKGLGNASGAAVATHLDHRIAMSFLVMGLVSEN
L05004_2 TTADMLNLGLFELQPTNDGLIHH--PS-EFKTNATDILT---DHRIGMMLAVACVLSSE
```

```
420 430 440 450
nk603cp4.pep PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
L05004_2 PVKIKQFDVAVNSFPGLPKLKLQNEG
```

nk603cp4.pep  
SWISSPROT:ARO\_A\_STA

ID ARO\_A\_STA STANDARD; PRT; 430 AA.  
AC Q05615;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 271 Initn: 526 Opt: 765 z-score: 831.2 E(): 2.5e-38

Smith-Waterman score: 765; 34.2% identity in 424 aa overlap

```
10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
ARO_A_STA MVNEQIIDISGPLKGEIEVPGDKSMTHRAIMLASLAEGVSTIYKPLLGEDCRRTM
```

```
70 80 90 100 110 120
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI
ARO_A_STA DIFRHLGVEIKEDDEKLVTSPGYQ-VNTPHQVLYTGNSGTTTLLAGLLSGLGNSVLS
```

```
130 140 150 160 170 180
nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPIITYRVEMASAOVKSVA
ARO_A_STA GDVSIKGRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPIITYRVEMASAOVKSVA
```

```
190 200 210 220 230 240
nk603cp4.pep LLAGLNTPGITTVIEPIIMTRDHEKMLQGFANPTVETDADGVRTIRLEGRGKLTGQVID
ARO_A_STA LFAFLSKPTIIEKELDVSRNHTETMPKHF--NIPIEAEGLSINTTPEAIR-YIKPADFH
```

```
250 260 270 280 290 300
nk603cp4.pep VPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIEVINPRLAGGED
ARO_A_STA VPGDISSAAFFIVAALITPGSDVTIHNVGINQTRSGIIDIIVKMGNIQLFN-QTTGAEP
```

```
310 320 330 340 350 359
nk603cp4.pep VADLRVR-SSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLS
ARO_A_STA TASIRIQYTPMLQPTIEGELVPAIDELPVIALCTQAVGTSTIKDAEELVKETNRID
```

```
360 370 380 390 400 410 419
nk603cp4.pep AVANGLKLNVDCEGETSLVVRGPRDGKGLGNASGAAVATHLDHRIAMSFLVMGLVSEN
ARO_A_STA TTADMLNLGLFELQPTNDGLIHH--PS-EFKTNATDILT---DHRIGMMLAVACVLSSE
```

```
420 430 440 450
nk603cp4.pep PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
ARO_A_STA PVKIKQFDVAVNSFPGLPKLKLQNEG
```

nk603cp4.pep  
SWISSPROT:ARO\_HELPJ

ID ARO\_HELPJ STANDARD; PRT; 429 AA.  
AC Q9ZKF7;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 220 Initn: 571 Opt: 708 z-score: 769.7 E(): 6.8e-35  
Smith-Waterman score: 708; 34.0% identity in 418 aa overlap

```
10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
ARO_HELPJ MIELDINADKSLSHRAVIFSLLAQKPCFVRNFMGDCVLSL
```

nk603cp4.pep KAMQAMGARIRKEG-DTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMTGLVGVYDFDSTF  
 AROA\_HELJPJ EIAQNILGAKVENTAKNSPKITPPTT--IKEPNKILNCNNSGTMRLYSGLLSAQKGLFVL  
 120 130 140 150 160 170  
 nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIT-YRVPMAAQVKS  
 AROA\_HELJPJ SGDNSLNARPMKRIIEPLKAFGAKILGREDNHFAPLAIVGGPLKA-CDYESPIASAAQVKS  
 180 190 200 210 220 230  
 nk603cp4.pep AVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIR-LEGRGKLTGQ  
 AROA\_HELJPJ AFILSALQAQGISAYKESELSRNHTEIMLKSLGANIQ---NQDGVLKISPLEK--PLESF  
 240 250 260 270 280 290  
 nk603cp4.pep VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIE-VINPRLA  
 AROA\_HELJPJ DFTIANDPSSAFFLALACAITPKSRLLLNPNTRIEAFEVLKMGAGHIEYVIQSK--  
 300 310 320 330 340 350  
 nk603cp4.pep GGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIAVAAAFAGATVMNGLEELRVKESD  
 AROA\_HELJPJ DLEVIGDIYIEHAPLKAISIDQNTA-SLIDEIPALSIAMLFAGKGSVMVRNAKDLRAKESD  
 360 370 380 390 400 410  
 nk603cp4.pep RLSAVANGLKLVGDCDEGETSLVVRGPRDQKGL---GNASGAAVATHLDHRIAMSFLV  
 AROA\_HELJPJ RIKAVVSNFKALGIECEEFEEDGFYIEGLGDASQLKHFSKIKPFIKSFNDHRIAMSFAV  
 420 430 440 450  
 nk603cp4.pep MGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROA\_HELJPJ LTLAL--PLEIDNLECANISFPTFQLWLNLFKKRSLNGN  
 460 470 480 490  
 nk603cp4.pep  
 GP\_BCT1:AE001527\_6  
 LOCUS AE001527\_6 [AE001527]  
 DEFINITION Helicobacter pylori, strain J99 section 88 of 132 of the complete  
 genome;  
 similar to H. pylori 26695 gene HP0401.  
 DATE 20-JAN-1999  
 ACCESSION AE001527 . . .  
 SCORES Init1: 220 Initn: 571 Opt: 708 z-score: 769.7 E(): 6.8e-35  
 Smith-Waterman score: 708; 34.0% identity in 418 aa overlap  
 10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AE001527\_6 MIELDINADSKSLSHRAVIFSLLAQKPCFVRNFMGEDCLSSL  
 70 80 90 100 110 119  
 nk603cp4.pep KAMQAMGARIRKEG-DTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMTGLVGVYDFDSTF  
 AE001527\_6 EIAQNILGAKVENTAKNSPKITPPTT--IKEPNKILNCNNSGTMRLYSGLLSAQKGLFVL  
 120 130 140 150 160 170  
 nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIT-YRVPMAAQVKS

AE001527\_6 SGDNSLNARPMKRIIEPLKAFGAKILGREDNHFAPLAIVGGPLKA-CDYESPIASAAQVKS  
 180 190 200 210 220 230  
 nk603cp4.pep AVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIR-LEGRGKLTGQ  
 AE001527\_6 AFILSALQAQGISAYKESELSRNHTEIMLKSLGANIQ---NQDGVLKISPLEK--PLESF  
 240 250 260 270 280 290  
 nk603cp4.pep VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIE-VINPRLA  
 AE001527\_6 DFTIANDPSSAFFLALACAITPKSRLLLNPNTRIEAFEVLKMGAGHIEYVIQSK--  
 300 310 320 330 340 350  
 nk603cp4.pep GGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIAVAAAFAGATVMNGLEELRVKESD  
 AE001527\_6 DLEVIGDIYIEHAPLKAISIDQNTA-SLIDEIPALSIAMLFAGKGSVMVRNAKDLRAKESD  
 360 370 380 390 400 410  
 nk603cp4.pep RLSAVANGLKLVGDCDEGETSLVVRGPRDQKGL---GNASGAAVATHLDHRIAMSFLV  
 AE001527\_6 RIKAVVSNFKALGIECEEFEEDGFYIEGLGDASQLKHFSKIKPFIKSFNDHRIAMSFAV  
 420 430 440 450  
 nk603cp4.pep MGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AE001527\_6 LTLAL--PLEIDNLECANISFPTFQLWLNLFKKRSLNGN  
 460 470 480 490  
 nk603cp4.pep  
 GP\_BCT1:AE000556\_2  
 LOCUS AE000556\_2 [AE000556]  
 DEFINITION Helicobacter pylori 26695 section 34 of 134 of the complete genome;  
 similar to PID:895778 SP:P52312 percent identity:  
 53.56; identified by sequence similarity; putative.  
 DATE 06-APR-1999  
 ACCESSION AE000556 . . .  
 SCORES Init1: 218 Initn: 568 Opt: 701 z-score: 762.1 E(): 1.8e-34  
 Smith-Waterman score: 701; 34.0% identity in 423 aa overlap  
 10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AE000556\_2 MIELDINADSKSLSHRAVIFSLLAQKPCFVRNFMGEDCLSSL  
 70 80 90 100 110 119  
 nk603cp4.pep KAMQAMGARIRKEG-DTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMTGLVGVYDFDSTF  
 AE000556\_2 EIAQNILGAKVENTAKNSPKITPPTT--IKEPNKILNCNNSGTMRLYSGLLSAQKGLFVL  
 120 130 140 150 160 170  
 nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIT-YRVPMAAQVKS  
 AE000556\_2 SGDNSLNARPMKRIIEPLKAFGAKILGREDNHFAPLVILG--SPLKACHYESPIASAAQVKS  
 180 190 200 210 220 230  
 nk603cp4.pep SAVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIR-LEGRGKLTG  
 AE000556\_2 SAFILSALQAQGASTYKESELSRNHTEIMLKSLGADIH---NQDGVLKISPLEK--PLEA



```

      240      250      260      270      280      290
nk603cp4.pep QVIDVPGDPSSTAFFLVAALLVPGSDVTILNVLNMPNTRTGLILTQEMGADIE-VINPRL
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
AROAH_HELPY FDFTIANDPSSAFFALACAITPKSRLLLNKVVLLNPTRIEAFVLLKKMGASIEYAIOQSK-
      220      230      240      250      260      270

```

nk603cp4.pep      360            370            380            390            400            410  
LSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSPLVMGLVS

400                      410                      420

```

      420          430          440          450
nk603cp4.pep  ENPVTVD DATMIATSFPEFMDLMAGLGA KIELSDTKAA
               :  :  :  :  :  :  :  :  :  :  :  :  :  :
AL139076_205 --GIEIDSDCIKTSFNFIEILSNLGARIDY
               400          410          420

```

nk603cp4.pep  
SWISSPROT:AROA CAMJE

```
ID      AROA_CAMJE      STANDARD;      PRT;      428 AA.
AC      P52312; Q9FP36;
DT      01-OCT-1996 (Rel. 34, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5-
```

SCORES Init1: 199 Initn: 675 Opt: 654 z-score: 711.4 E(): 1.2e-31  
Smith-Waterman score: 788; 34.3% identity in 428 aa overlap

```

              10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRFSFMFGLSGASGETRITGLLEGEDVINTG
              | : | | | | | : : | : : | : : | : : |
AROA_CAMJE   MKIYKLTQTPVNAILENIAADKSISHRFAISLLTQEENKAQNYLLAQDTLNTL
              10      20      30      40      50

```

```

              70          80          90          100          110          119
nk603cp4.pep  KAMQAMGARIRKEGDTWIIDVGNGGLLAPEAPLDFGNAATGCRLTMG-LVGVDYDFDSTF
               : : : | : | : : : : : : : : | : | : | : | : | : | : :
AROA_CAMJE    EIIKNLGAKIEQKDCSVKI--IPPKEILSPNCILDCGNSGTAMRLMIGFLAGISGF-FVL
               60          70          80          90          100          110

```

```

      120       130       140       150       160       170
nk603cp4.pep IGDSLTKRPMGRVLNPLREMGVQVKSEGDRL-PVTLRGPKTPTPITYRVPMASAQVKS
              || :||| | :|| :||::: || :| : : : : ||| |
AROA_CAMJE   SGDKYLNRRPMRISKPLTGIGARIYGRNEANLPCIEGQKLKA-FNFKSEISSAQVKT
               120       130       140       150       160

```

```

      180          190          200          210          220          230
nk603cp4.pep AVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGK-LTGO
               |:|::: : | | ::||:|:||: | |::: ||: :::: | | :|
AROA_CAMJE    AMILSAFRNDVCTFSEISLSRHNSENMLKAMKA-P-IRVSNDGL-SLEINPLKKPLKAQ
              170         180         190         200         210         220

```

nk603cp4.pep VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMPRTTGLILTLOEMGADIEVINPRLAG  
| : | : | | : : : | : : | : | : | | | : | : | :  
AROA\_CAMJE NIIIPNDPSSAFYFLVAAILPKSQIILKNILLNPTRIEAYKILQKMGAKELEMTITQ-ND

240      250      260      270      280      290

```

      300       310       320       330       340       350
nk603cp4.pep GEDVADLRVRSSLT LKGVTVPEDRAPSMID EYPILAVAAFAEAGATVMNGLEELRVKESDR
               |::|||:| |::| :| | | | | | | | | | | | | | | | | | | | | | |
AROAcAMJE    FETIGEIRVESSKLNGIEV-KDNIAWLIDEAPALAI AFALAKGSSSLINAKELRVKESDR
              290         300         310         320         330         340

```

```

          360      370      380      390      400      410
nk603cp4.pep LSAVANGKLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFLVMGLVS
          ::::: ||| ||: | : : : : | : : : : ||||| : : ||:
AROQ_CAMJE   IAVMVENLKLCCVEARELDDGFETGGCELKS-----SKKSYGDHRIAMSFAILGLLC
          350      360      370      380      390

```

```

          420          430          440          450
nk603cp4.pep ENPVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
              : : | : | | | : : : : | | : |
AROA_CAMJE   --GLEIDSDCIKTSEPFNFIELSLNIGRIDY

```

nk603cp4.pep  
GP\_BCT2:X89371\_2

```

LOCUS      X89371_2 [CJDNAPALG]
DEFINITION C.jejuni DNA for pheS, aroA, lytB and Sl genes.
DATE       28-DEC-1997
ACCESSION  X89371
NID
ORGANISM   Campylobacter jejuni . . .

```

SCORES Init1: 198 Initn: 671 Opt: 644 z-score: 700.6 E(): 4.8e-31  
Smith-Waterman score: 778; 34.0% identity in 427 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSPFMFGGLASGETRITGLLEGEDVINTG  
X89371\_2 MKIYKLTQTPVNAILENIAADKSISHRFAIFSLTQEEKNAQNYLLAQDTLNTL

```

              70      80      90      100      110      119
nk603cp4.pep KAMQAMGARIRKEGDTWIIDVGNGGGLLAPEAPLDFGNAATGCRLTMG-LVGVYDFDSTF
              ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
X89371_2      EIIKNLGAKIEQKDSCVKI--IPPKELISPNCILDCNGSGTAMRLMIGFLAGISGF-FVL
              60      70      80      90      100      110

```

[illegible]

```

      180       190       200       210       220       230
nk603cp4.pep AVLLAGLNTPTGITTVEIPIMTRDHTKMLQGFGANPVTETDADGVRTIRLEGRKGLTGQV
              |:|::: : :: | :||:|:|::| :| :|::|:|
X89371_2      AMILSAFRANNVCASFSEISLRNHSENMLKAMKA-P-IRVSNDGLSLEISPLKKPLKAQN
              170       180       190       200       210       220

```

```

      240      250      260      270      280      290
nk603cp4.pep IDVPGDPSSTAFPLVAALLVPGSDVTILNLVLMNPTRTGLILTIQEMGADIEVINPRLAGG
| : : : | | : : : | : : : | : : : | : : : | : : : | : : :
X89371_2      I I I P N D P S S A F Y F A L A A I I L P K S Q I I L K N I L L N P T R I E A Y K I I L K M G A K L E M T I T Q - N D F
      230      240      250      260      270      280

```

```

          300      310      320      330      340      350
nk603cp4.pep EDVADLRVRSSLTGKGVTPEDRAPMSIDEYPILAVAAFAEGATVMNGLEELRVKESDRI
              |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
X89371_2      ETIGIEIRVSSKLNIGIEV-KDNIAWLIDEAPALAIAPALAKGKSSLINAKELRVKESDRI
              290      300      310      320      330      340

```

```

          360      370      380      390      400      410
nk603cp4.pep SAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLVLMGLVSE
          :::::| | | | : : : | : | : : : | | | | | | : : | |
X89371_2      AVMVENLKLQGEARELDDGFIEGGCELKS-----SKIKSYGDHRIAMSFALLGLLC-
          350      360      370      380      390

```

```

          420          430          440          450
nk603cp4.pep NPVTVDDATMIATSPFPEFMDLMAGLGAKIELSDTKAA
          : : ||: | |||: |: ::::: |||: :
X89371_2      -GIEIDSDCIKTSFPNFTIELSNLGARIDY
          400          410          420

```

nk603cp4.pep  
TREMBL MAIN:O9RHZ8

ID Q9RHZ8 PRELIMINARY; PRT; 207 AA.  
AC O9RHZ8;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 5-ENOLPYRUVYLSHIKIMATE 3-P SYNTHASE (FRAGMENT) . . . .

SCORES Initl: 419 Initn: 579 Opt: 622 z-score: 681.6 E(): 5.5e-30  
 Smith-Waterman score: 622; 51.0% identity in 198 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 Q9RHZ8 MHSNDLIFLANPGGSLTGQLRVPGDKSISHRSIMLGSIAEGTTEEXEGFLEGEDALATX

nk603cp4.pep KAMQAMGARIRKEGDTWI-IDGVNGGGLLAPEAPLDFGNAATGCRLTMGVLGVYDFDSTF  
 Q9RHZ8 QAFRXMGVVIIEGPHQGRVTVHGVGLHGLQAPPPIYLGNSGTSMRLLAGLLAAQPFDDTL

nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITYRVPMASQVKS  
 Q9RHZ8 SGDASLTKRPMNRVAKPLREMGAVIETAAEGRPLTIRGGKLSGMHYDMPMASQVKS

nk603cp4.pep VLLAGLNTPGITTVIEPIMTRDTEKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI  
 Q9RHZ8 LLLAGLYAAGKTSVTEPAPTRDHTERMQQ

nk603cp4.pep  
 GP\_BCT1:AF038578\_6

LOCUS AF038578\_6 [AF038578]  
 DEFINITION Pseudomonas stutzeri gyrase A subunit (gyrA) gene, partial cds;  
 3-phosphoserine aminotransferase (serC), chorismate  
 mutase/prephenate dehydratase (aroQp/pheA), imidazole acetol  
 phosphate aminotransferase (hisHb), and cyclohexadienyl  
 dehydrogenase (tyrAc) genes, complete cds; and . . .

SCORES Initl: 419 Initn: 579 Opt: 622 z-score: 681.6 E(): 5.5e-30  
 Smith-Waterman score: 622; 51.0% identity in 198 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AF038578\_6 MHSNDLIFLANPGGSLTGQLRVPGDKSISHRSIMLGSIAEGTTEEXEGFLEGEDALATX

nk603cp4.pep KAMQAMGARIRKEGDTWI-IDGVNGGGLLAPEAPLDFGNAATGCRLTMGVLGVYDFDSTF  
 AF038578\_6 QAFRXMGVVIIEGPHQGRVTVHGVGLHGLQAPPPIYLGNSGTSMRLLAGLLAAQPFDDTL

nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITYRVPMASQVKS  
 AF038578\_6 SGDASLTKRPMNRVAKPLREMGAVIETAAEGRPLTIRGGKLSGMHYDMPMASQVKS

nk603cp4.pep VLLAGLNTPGITTVIEPIMTRDTEKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVI  
 AF038578\_6 LLLAGLYAAGKTSVTEPAPTRDHTERMQQ

nk603cp4.pep  
 SWISSPROT:AROA\_HAESO

ID AROA\_HAESO STANDARD; PRT: 432 AA.  
 AC P52310;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Initl: 133 Initn: 304 Opt: 461 z-score: 502.9 E(): 4.9e-20  
 Smith-Waterman score: 505; 28.8% identity in 437 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_HAESO MEKLTLSPISRIDGEINLPKSKLSNRALLLAALAKGTTQVTNLLDSDDIRYML

nk603cp4.pep KAMQAMGARIRKEGDTWI-IDGVNGGGLLAPEAPLDFGNAATGCR-LTMGLGVYDFDSDS  
 AROA\_HAESO NALKALGVNYQLSDDKTVCVVEGIGGAFQWQNGLSLFLGNAGTAMRPLAALAKLGDTES

nk603cp4.pep TFI--GDASLTKRPMGRVLNPLREMGVQVKSEGDRLP-VTLRGPKTPTPTITYRVPMASA  
 AROA\_HAESO EVILTGEPRMKERPIKHLVDALRQTGANIQYLENDGYPLAIRNQIGFGKXQVQIDGSISS

nk603cp4.pep QVKSALLAGLNTPGITTV--IEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLE  
 AROA\_HAESO QFLTALLMAAPLGEEDMEIEILGELVSKPYIDITPAMMKDFGIN-----VDDYNYQRFL

nk603cp4.pep GRGK---LTGQVIDVPDPSSTAFPLVAALLVPGSDVTLNVLNMPRTTGLILT--LQEM  
 AROA\_HAESO IKGKQYYISPTQYLVVEGDASSASYFLAAAA-IK GK-VKVTGIGRNSIQGDRLFDVLAQM

nk603cp4.pep GADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYFILAFAAFAEGATVM  
 AROA\_HAESO GA-----KVTWGEDF--IQVEKSELKIDMDMNHIP---DAAMTIAITLFAQGETVI

nk603cp4.pep NGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVVRGRPDGKGLGNASGAATHLDH  
 AROA\_HAESO RNINYNRVKEDRLTAIATELRKLGAEVEEGED--FIRIQP--LALDKFKHAEIATYNDH

nk603cp4.pep RIAMSFVLMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROA\_HAESO RIAMCFSLIAL--SDTSVTILDPACTAKTFPTFYFSEFEKISKNQ

nk603cp4.pep  
 GP\_BCT3:L47538\_1

LOCUS L47538\_1 [HEA3P1C]  
 DEFINITION Haemophilus somnus 3-phosphoshikimate 1-carboxyvinyltransferase  
 gene, complete cds.

DATE 04-JAN-1996  
ACCESSION L47538  
NID . . .

SCORES Init1: 133 Initn: 304 Opt: 461 z-score: 502.9 E(): 4.9e-20  
Smith-Waterman score: 505; 28.8% identity in 437 aa overlap

```
10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
L47538_1 MEKLTLPISRIDGEINLPGSKSLNRAALLAALAKGTQVTNLLDSDDIRYML
10 20 30 40 50
```

```
70 80 90 100 110
nk603cp4.pep KAMQAMGARIRKEGDTWI--IDGVGNGLLAPEAPLDFGNAATGCR--LTMGLVGVYDFD
L47538_1 NALKALGVNYQLSDDKTVCVVEGIGGAFQWQNGLSLFLGNAGTAMRPLAALCLKGDTES
60 70 80 90 100 110
```

```
120 130 140 150 160 170
nk603cp4.pep TFL--GDASLTKRPMGRVLNPLREMGVQVKSDEGDRLP-VTLRGPKTPTPTITYRVMASA
L47538_1 EVILTGEPRMKERPIKHLVDALRQTGANIQYLENDGYPLAIRNQGIFGKQVQIDGSISS
120 130 140 150 160 170
```

```
180 190 200 210 220 229
nk603cp4.pep QVKSALLAGLNTPGITTV--IEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLE
L47538_1 QFLTALLMAAPLGEEDMEIEILGELVSKPYIDITPAMMKDFGIN-----VDDYNQRF
180 190 200 210 220
```

```
230 240 250 260 270 280
nk603cp4.pep GRGK---LTQGVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPRTGLILT--LQEM
L47538_1 IKGKQYISPTYLVEGDASSASYFLAAA--IKGK-VKVTGIRNSIQGDRFLFADVLAQM
230 240 250 260 270 280
```

```
290 300 310 320 330 340
nk603cp4.pep GADIEVINPRLAGEEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIAVALAAFAEGATVM
L47538_1 GA-----KVTWGEDF--IQVEKSELKIDMMNHIP---DAAMTIAITALFAQGETVI
290 300 310 320 330
```

```
350 360 370 380 390 400
nk603cp4.pep NGLLEELRVKESDRLSAVANGLKNGVDCDEGETSLVVRGPRDGKGLGNASGAAVATHLDH
L47538_1 RNIYNWRVKETDRLTATLRLKLGAEVEEGED--FIRIQP--LALDKFKHAEIATYNDH
340 350 360 370 380 390
```

```
410 420 430 440 450
nk603cp4.pep RIAMSFVLVGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEISDTKAA
L47538_1 RIAMCFSIAL--SDTSVTILDPACTAKTFTPTTFSEFEKISKNQ
400 410 420 430
```

nk603cp4.pep  
GP\_BCT3:U89948\_1

LOCUS U89948\_1 [PHU89948]  
DEFINITION Pasteurella haemolytica serotype 2 aroA gene, complete cds.  
DATE 17-NOV-1997  
ACCESSION U89948  
NID  
ORGANISM Mannheimia haemolytica . . .

SCORES Init1: 141 Initn: 299 Opt: 443 z-score: 483.5 E(): 5.9e-19  
Smith-Waterman score: 477; 27.9% identity in 448 aa overlap

```
10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
U89948_1 MEKLTLPISRVEGEINLPGSKSLNRAALLAALATGTQVTNLLDSDDIRHML
10 20 30 40 50
```

```
70 80 90 100 110
nk603cp4.pep KAMQAMGARIRKEGDTWI--IDGVGNGLLAPEAPLDFGNAATGCR--LTMGLVGVYDF
U89948_1 NALKALGVNYQLSDDKTVCVVEGIGGAFQWQNGLSLFLGNAGTAMRPLAALCLKGAEATA
60 70 80 90 100 110
```

```
120 130 140 150 160 169
nk603cp4.pep DSTFIGDASLTKRPMGRVLNPLREMGVQVKSDEGDRLPV-----LRGPKTPTPTITYRV
U89948_1 QIILTGEPRMKERPIKHLVDALRQVGAEVQYLENEGYPLAISNSGLQGGKQVQIDGSISS
120 130 140 150 160 170
```

```
170 180 190 200 210 220
nk603cp4.pep PMASAVQKSAVLL--AGLNTPGITTVI--EPIMTRDHTEKMLQGFGANPTVETDADGVRTIR
U89948_1 QFLTALLMSAPLAESDMEIEIIGDLVSKPYI--DITLSMMNDFGI--TVE-NRD-YKTFL
180 190 200 210 220
```

```
230 240 250 260 270 280
nk603cp4.pep LEGR--GKLTGQ--VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPRTGLILT--TLQE
U89948_1 VKGKQGYVAPQGNLYVEGDASSASY--FLASGAIKGK-VKVTGIGKKSIGQDRFLFADVLEK
230 240 250 260 270 280
```

```
290 300 310 320 330 340
nk603cp4.pep MGADIEVINPRLAGEEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIAVALAAFAEGATV
U89948_1 MGA-----KITWGEDF--IQAEQSLKGVDMMDNHIP---DAAMTIAITALFAEGETV
290 300 310 320 330
```

```
350 360 370 380 390 400
nk603cp4.pep MNGLEELRVKESDRLSAVANGLKNGVDCDEGETSLVVRGPRDGKGLGNASGAAVATHLD
U89948_1 IRNIYNWRVKETDRLTATLRLKLGAEVEEGED--FIRIQP--LALENQHAIEIETYN
340 350 360 370 380 390
```

```
410 420 430 440 450
nk603cp4.pep HRIAMSFVLVGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEISDTKAA
U89948_1 HRMAMCFSIAL--SNTSEVTILDPACTAKTFTPTTFSEFEKLSVR
400 410 420 430
```

nk603cp4.pep  
SWISSPROT:AROA\_PASHA

ID AROA\_PASHA STANDARD; PRT; 432 AA.  
AC P54220; P96968;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 141 Initn: 299 Opt: 443 z-score: 483.5 E(): 5.9e-19  
Smith-Waterman score: 477; 27.9% identity in 448 aa overlap

```
10 20 30 40 50 60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AROA_PASHA MEKLTLPISRVEGEINLPGSKSLNRAALLAALATGTQVTNLLDSDDIRHML
10 20 30 40 50
```

```
70 80 90 100 110
nk603cp4.pep KAMQAMGARIRKEGDTWI--IDGVGNGLLAPEAPLDFGNAATGCR--LTMGLVGVYDF
```



```

      230      240      250      260      270      280
nk603cp4.pep G-RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGA
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
L04686_2      GNQSYISPNKYLVEGDASSASY-FLAAGAIK GK-VKVTGIGKNSIQGDRFLFADVLEKMG
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      290      300      310      320      330      340
nk603cp4.pep DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNG
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
L04686_2      -----KITWGEDF--IQAEHAELNGIDMDMNHIP---DAAMTIATTALEFSGNETVIRN
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      350      360      370      380      390      400
nk603cp4.pep LEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRI
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
L04686_2      IYNWRVKETDRLTAMATELRKVGAEVEEGED--FIRIQP--LALNQFKHANIETYNDRHM
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      340      350      360      370      380      390
      410      420      430      440      450
nk603cp4.pep AMSFLVMGLVSENFTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
L04686_2      AMCFSLIAL-SNTPVTILDPKCTAKTPTPTFNEFEKICKLN
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      400      410      420      430

```

nk603cp4.pep  
SWISSPROT:AROA\_YERPE

```

ID   AROA_YERPE   STANDARD;       PRT;   424 AA.
AC   Q60112;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

```

SCORES Init1: 176 Initn: 359 Opt: 430 z-score: 469.6 E(): 3.5e-18  
Smith-Waterman score: 519; 27.4% identity in 435 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
AROA_YERPE      MLESLETLQPIALVNGTVNLPGSKSVSNRALLAALAEQTTQLNNVLDSDDIRHML
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      70      80      90      100      110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--GNAATGCRLTMGLVGVDYDFDST
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
AROA_YERPE      NALQALGVDFRLSADRTCCCEVDGLGGLVAEQPLSLFLGNAGTAMRPLAAVLCGLNSDIV
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      120      130      140      150      160      170
nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTITYRVPMAAQVK
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
AROA_YERPE      LTGEPRMKERPIGHLVDALRQGGQIDYLEQENYPLRLRGGFRRGGLTVDGRVSSQFLT
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      180      190      200      210      220      230
nk603cp4.pep SAVLLAGLNTPGIIT-VIEPIMTR---DHTEKMLQGGFANPTVETDADGVRTIRLEG--R
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
AROA_YERPE      ALLMTAPLAEQDTTIRIMGDLVSKPYIDITLHLMKAFG----IDVGHENYQIFHIKGGQT
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      240      250      260      270      280      289
nk603cp4.pep GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTR--TGLILTQEMGADI
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
AROA_YERPE      YRSPGTYL-VEGDASSASYFLAAAA-IKGGTVRVVTGIGKKSQGDITKFAVLEKMGAA---
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      290      300      310      320      330      340      349
nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGLEE

```

```

      290      300      310      320      330
AROA_YERPE      ---KVTWGGDY--IECSRGELOGIDMDMNHIP---DAAMTIATTALEFSGNETVIRN
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      350      360      370      380      390      400      409
nk603cp4.pep LRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
AROA_YERPE      WRVKETDRLTAMATELRKVGAEVEEGED--YIRVVPPLQ---LTAADIGTYDDHRMAMC
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      410      420      430      440      450
nk603cp4.pep FLVMGLVSENFTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
AROA_YERPE      FSLVAL-SDTPVTILDPKCTAKTPTPTDYFEQFARK
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      400      410      420

```

nk603cp4.pep  
TREMBL\_NEW:CAC90219

```

ID   CAC90219   PRELIMINARY;       PRT;   428 AA.
AC   CAC90219;
DT   01-NOV-2001 (EMBLrel. 63, Created)
DT   01-NOV-2001 (EMBLrel. 63, Last sequence update)
DT   01-NOV-2001 (EMBLrel. 63, Last annotation update)
DE   3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) . . .

```

SCORES Init1: 176 Initn: 359 Opt: 430 z-score: 469.5 E(): 3.6e-18  
Smith-Waterman score: 521; 28.1% identity in 438 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
CAC90219      MLESLETLQPIALVNGTVNLPGSKSVSNRALLAALAEQTTQLNNVLDSDDIRHML
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      70      80      90      100      110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--FGNAATGCRLTMGLVGVDYDFDST
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
CAC90219      NALQALGVDFRLSADRTCCCEVDGLGGLVAEQPLSLFLGNAGTAMRPLAAVLCGLNSDIV
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      120      130      140      150      160      170
nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTITYRVPMAAQVK
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
CAC90219      LTGEPRMKERPIGHLVDALRQGGQIDYLEQENYPLRLRGGFRRGGLTVDGRVSSQFLT
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      180      190      200      210      220      230
nk603cp4.pep SAVLLAGLNTPGIIT-VIEPIMTR---DHTEKMLQGGFANPTVETDADGVRTIRLEG--
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
CAC90219      TALLMTAPLAEQDTTIRIMGDLVSKPYIDITLHLMKAFG----IDVGHENYQIFHIKGGQ
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      240      250      260      270      280
nk603cp4.pep RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTR--TGLILTQEMGADI
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
CAC90219      TYRSPGTYL-VEGDASSASYFLAAAA-IKGGTVRVVTGIGKKSQGDITKFAVLEKMGAA--
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      290      300      310      320      330      340
nk603cp4.pep EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGLE
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
CAC90219      -----KVTWGGDY--IECSRGELOGIDMDMNHIP---DAAMTIATTALEFSGNETVIRN
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
      350      360      370      380      390      400
nk603cp4.pep ELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAM
      1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
CAC90219      NWRVKETDRLTAMATELRKVGAEVEEGED--YIRVVPPLQ---LTAADIGTYDDHRMAM

```

340 350 360 370 380 390

nk603cp4.pep 410 420 430 440 450  
SFLVMGLVSENFVTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
CAC90219 CFSLVAL-SDTPVTILDPKCTAKTFPDYFQFARLSQLA  
400 410 420

nk603cp4.pep  
GP\_BCT1:AE004251\_5

LOCUS AE004251\_5 [AE004251]  
DEFINITION Vibrio cholerae chromosome I, section 159 of 251 of the complete  
chromosome;  
similar to SP:P07638 GB:X00557 PID:1066024  
PID:40966 GB:U00096; identified by sequence similarity;  
putative. . . .

SCORES Init1: 162 Initn: 444 Opt: 425 z-score: 464.1 E(): 7.1e-18  
Smith-Waterman score: 529; 28.4% identity in 437 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AE004251\_5 MESLTLQPIELISGEVNLPGSKSVSNRALLAALASGTTTLNLLSDDIRHML  
10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLD--FGNAATGCRLTMGLVGVYDFDST  
AE004251\_5 NALTKLGVNRYLSADKTTCEVEGLGQAFHTTQPLELFLGNAGTAMRPLAAALCLGQGDYV  
60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
FIGDASLTKRPMGRVNLPLREMGVQVKSSEDGDRLP-VTLRGPKTPTPTIYRVPMASQVK  
AE004251\_5 LTGEPRMKERPIGHLVLDALRQAGAQIEYLEQENFPPLRIQGTGLQAGTVTIDGSISSQFL  
120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
SAVLLAGLNTPGITT--VIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRG  
AE004251\_5 TAFLEMSAPLAQGVKTIKIVGELVSKPYIDITLHIMEQFGVQVI---NHQYQEFVIPAGQS  
180 190 200 210 220 230

nk603cp4.pep 240 250 260 270 280 289  
KLT-GQVIDVPGDPSSSTAFPLVAALLVPGSDVTILNVLNPTRTGLIL--TLQEMGADIE  
AE004251\_5 YVSPGQFL-VEGDASSASYFLAAAA-IKGGEVKVTGIGKNSIQGDIQFADALEKMGQIE  
240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340 349  
VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAFAGATVMNGLEE  
AE004251\_5 W-----GDDY--VIARRGELNAVDLDFNHIP---DAAMTIATTALFAKGTTAIRNVYN  
290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400 409  
LRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
AE004251\_5 WRVKETDRLAAMATELRKVGVATVEGEDFIVIT--PPTKLIH---AAIDTYDDHRMAMC  
340 350 360 370 380 390

nk603cp4.pep 410 420 430 440 450  
FLVMGLVSENFVTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AE004251\_5 FSLVAL-SDTPVTINDPKCTSKTFPDYFDFKFAQLSR  
400 410 420

nk603cp4.pep  
TREMBL\_MAIN:Q9KRB0

ID Q9KRB0 PRELIMINARY; PRT; 426 AA.  
AC Q9KRB0;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 162 Initn: 444 Opt: 425 z-score: 464.1 E(): 7.1e-18  
Smith-Waterman score: 529; 28.4% identity in 437 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
Q9KRB0 MESLTLQPIELISGEVNLPGSKSVSNRALLAALASGTTTLNLLSDDIRHML  
10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLD--FGNAATGCRLTMGLVGVYDFDST  
Q9KRB0 NALTKLGVNRYLSADKTTCEVEGLGQAFHTTQPLELFLGNAGTAMRPLAAALCLGQGDYV  
60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
FIGDASLTKRPMGRVNLPLREMGVQVKSSEDGDRLP-VTLRGPKTPTPTIYRVPMASQVK  
Q9KRB0 LTGEPRMKERPIGHLVLDALRQAGAQIEYLEQENFPPLRIQGTGLQAGTVTIDGSISSQFL  
120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
SAVLLAGLNTPGITT--VIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRG  
Q9KRB0 TAFLEMSAPLAQGVKTIKIVGELVSKPYIDITLHIMEQFGVQVI---NHQYQEFVIPAGQS  
180 190 200 210 220 230

nk603cp4.pep 240 250 260 270 280 289  
KLT-GQVIDVPGDPSSSTAFPLVAALLVPGSDVTILNVLNPTRTGLIL--TLQEMGADIE  
Q9KRB0 YVSPGQFL-VEGDASSASYFLAAAA-IKGGEVKVTGIGKNSIQGDIQFADALEKMGQIE  
240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340 349  
VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAFAGATVMNGLEE  
Q9KRB0 W-----GDDY--VIARRGELNAVDLDFNHIP---DAAMTIATTALFAKGTTAIRNVYN  
290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400 409  
LRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
Q9KRB0 WRVKETDRLAAMATELRKVGVATVEGEDFIVIT--PPTKLIH---AAIDTYDDHRMAMC  
340 350 360 370 380 390

nk603cp4.pep 410 420 430 440 450  
FLVMGLVSENFVTDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Q9KRB0 FSLVAL-SDTPVTINDPKCTSKTFPDYFDFKFAQLSR  
400 410 420

nk603cp4.pep  
SWISSPROT:AROA\_YEREN

ID AROA\_YEREN STANDARD; PRT; 427 AA.  
AC P19688;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Initl: 185 Initn: 373 Opt: 422 z-score: 460.9 E(): 1.1e-17  
Smith-Waterman score: 519; 28.6% identity in 440 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
X82415\_1 MESLTLQPIARVDGTVNLPGSKSVSNRALLAALARGTTVLTNLLSDDDVRHML  
10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--GNAATGCRLTMGLVGVYDFDST  
X82415\_1 NALSALGVHYVLSSDTRCEVTGTGGPLQAGSALEFLGNAGTAMRPLAAALCLGKNDIV  
60 70 80 90 100 110

nk603cp4.pep FIGDASLTAKRPMGRVNLPLREMGVQV--KSEGDGR-LPVTLRGPKTPTPTITYRVFMSA  
X82415\_1 LTGEPRMKERPIGHLVDALRQGAQIDYLEQENYRRCIAGGFRGKLTVDGVSQFLTA  
120 130 140 150 160 170

nk603cp4.pep QVKSALLAGLNTPGIT--TVIEPIMTRDHTKMLQGFANPTVET---DADGVRTIRL  
X82415\_1 LLMTAPLABQDTEIQGELVSKPYI--DITLHLMKAFGVDDVHHYQIFHIKGGQTYRS  
180 190 200 210 220 230

nk603cp4.pep EGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTR--TGLILTLOEMGA  
X82415\_1 PG-----IYLVGEDASSASYFLAAAA--IKGGTVRVGTGIGKQSVQGDTKFADVLEKMG  
240 250 260 270 280

nk603cp4.pep DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEXPIILAVAAFAEGATVMNG  
X82415\_1 -----KISWGDDY--IECSRGELOQIDMDMNHIP---DAAMTIATLAFADGPTVIRN  
290 300 310 320 330

nk603cp4.pep LEELRVKESDRLSAVANGLKLVGDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHRI  
X82415\_1 IYNWRVKETDRLSAMATELRKVGAEEVEGQD--YIRVVPPAQLIA----AEIGTYNDHRM  
340 350 360 370 380

nk603cp4.pep AMSFLVGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
X82415\_1 AMCFSLVAL-SDTPVTILDPKCTAKTFPDYFEQLARLSQIA  
390 400 410 420

nk603cp4.pep  
GP\_BCT3:X82415\_1

LOCUS X82415\_1 [KPAROA]  
DEFINITION K.pneumoniae aroA gene.  
DATE 17-FEB-1997  
ACCESSION X82415  
NID  
ORGANISM Klebsiella pneumoniae . . .

SCORES Initl: 151 Initn: 403 Opt: 420 z-score: 458.7 E(): 1.4e-17  
Smith-Waterman score: 505; 29.2% identity in 435 aa overlap

10 20 30 40 50 60

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
X82415\_1 MESLTLQPIARVDGTVNLPGSKSVSNRALLAALARGTTVLTNLLSDDDVRHML  
10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--GNAATGCRLTMGLVGVYDFDST  
X82415\_1 NALSALGVHYVLSSDTRCEVTGTGGPLQAGSALEFLGNAGTAMRPLAAALCLGKNDIV  
60 70 80 90 100 110

nk603cp4.pep FIGDASLTAKRPMGRVNLPLREMGVQVKS--EDGRLPVTLRGPKTPTPTITYRVFMSA  
X82415\_1 LTGEPRMKERPIGHLVDALRQGAQIDYLEQENYRRCIAGGFRGKLTVDGVSQFLTA  
120 130 140 150 160 170

nk603cp4.pep SAVLLAGLNTPGITTVI--EPIMTR---DHTKMLQGFANPTVETDADGVRTIRLEGRG  
X82415\_1 TALLMASFLAPQDVTIAIKGELVSRPYIDITLHLMKTFGVE--VENQAYQRFIVRGNQOY  
180 190 200 210 220 230

nk603cp4.pep KLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGADIEV  
X82415\_1 QSPGDYL-VEGDASSASY-FLAAGAIGKGTVKVTGIGRNSVQGDIRFADVLEKMGATV--  
240 250 260 270 280

nk603cp4.pep INPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEXPIILAVAAFAEGATVMNGLEE  
X82415\_1 -----TWGEDYIACR---GELNAIDMDMNHIP---DAAMTIATLAFARGTPTTLRNIYN  
290 300 310 320 330

nk603cp4.pep LRVKESDRLSAVANGLKLVGDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHRIAMS  
X82415\_1 WVRKETDRLFAMATELRKVGAEEVEGQD--YIRITPPL-----TLQFAEIGTYNDHRMAMC  
340 350 360 370 380 390

nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
X82415\_1 FSLVAL-SDTPVTILDPKCTAKTFPDYFGQLARISTLA  
400 410 420

nk603cp4.pep  
SWISSPROT:AROAKLEPN

ID AROAKLEPN STANDARD; PRT; 427 AA.  
AC P24497;  
DT 01-MAR-1992 (Rel. 21, Created)  
DT 01-MAR-1992 (Rel. 21, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Initl: 151 Initn: 403 Opt: 420 z-score: 458.7 E(): 1.4e-17  
Smith-Waterman score: 505; 29.2% identity in 435 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
X82415\_1 MESLTLQPIARVDGTVNLPGSKSVSNRALLAALARGTTVLTNLLSDDDVRHML  
10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--GNAATGCRLTMGLVGVYDFDST  
X82415\_1 NALSALGVHYVLSSDTRCEVTGTGGPLQAGSALEFLGNAGTAMRPLAAALCLGKNDIV  
60 70 80 90 100 110



AROA\_KLEPN NALSALGVHYVLLSSDRTRCEVTGTGGPLQAGSALEPLGNAGTAMRPLAAALCLGNSDIV  
 60 70 80 90 100 110  
 nk603cp4.pep 120 130 140 150 160 170  
 FIGDASLTAKRPMGRVNLPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIYRVPMASAOVK  
 AROA\_KLEPN 120 130 140 150 160 170  
 LTGEPRMKRPIGHLVDALRQGAQIDYLEQENYPPFLRLRGFTGGDVEVDGSVSS-QFL  
 nk603cp4.pep 180 190 200 210 220 230  
 SAVLLAGLNTPGITTVI--EPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRG  
 AROA\_KLEPN 180 190 200 210 220 230  
 TALLMASPLAQDVTIAIKGELVSRPYIDITLHLMKTFGVE--VENQAYQRFIVRGNQY  
 nk603cp4.pep 240 250 260 270 280 290  
 KLTGGQVIDVPGDPSSSTAFPLVAALLVPGSDVTILNVLNMPRTTGLILT--LOEMGADIEV  
 AROA\_KLEPN 240 250 260 270 280  
 QSPGQYL-VEGDASSASY-FLAAGAIKGGTVKVTGIGRNSVQGDIFADVLEKMGATV--  
 nk603cp4.pep 300 310 320 330 340 349  
 INPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGLEE  
 AROA\_KLEPN 290 300 310 320 330  
 -----TWGEDYIACTR--GELNAIDMDMNHIP---DAAMTIATAALFARGTTTLRNIN  
 nk603cp4.pep 350 360 370 380 390 400 409  
 LRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGPRDGKGLGNASGAAVATHLDHRIAMS  
 AROA\_KLEPN 340 350 360 370 380 390  
 WRVKETDRLFAMATELRKVGAEVEEGED--YIRITPPL---TLQFAEIGTYNDRHMA  
 nk603cp4.pep 410 420 430 440 450  
 FLVVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEIELSDTKAA  
 AROA\_KLEPN 400 410 420  
 FSLVAL-SDTPVTIILDPKCTAKTFPDYFQGLARISTLA

nk603cp4.pep  
 SWISSPROT:AROA\_HAEIN

ID AROA\_HAEIN STANDARD; PRT; 432 AA.  
 AC Q03421;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 170 Initn: 315 Opt: 420 z-score: 458.6 E(): 1.4e-17  
 Smith-Waterman score: 476; 27.0% identity in 437 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 U32833\_2 10 20 30 40 50  
 MEKITLAPISAVEGTINLPGSKSLSNRALLAALAKGTTKVTNLLDSDDIRHML

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRKEGDTWIIDGVGNGGLLAPEAPLDF--GNAATGCR-LTMGLV--GVYDF  
 U32833\_2 60 70 80 90 100 110  
 NALKALGVRYQLSDDKTICEVEGLGTFNIQDNLSLFLGNAGTAMRPLTAALCLKGKTES

nk603cp4.pep 120 130 140 150 160 170  
 DSTFFIGDASLTAKRPMGRVNLPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIYRVPMASA  
 AROA\_HAEIN 120 130 140 150 160 170  
 EIILTGEPRMKRPIHLVDALRQAGADIRYLENEGYPFLAIRNKGKIGKVKIDGSISS

nk603cp4.pep 180 190 200 210 220 229  
 QVKSALLAGLNTPGITTV--IEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLE  
 AROA\_HAEIN 180 190 200 210 220 230  
 QFLTALLMSAPLAENDTEIEIIGELVSKPYIDITLAMRDFG----VQVENHHYQKFOVK

nk603cp4.pep 230 240 250 260 270 280  
 G-RGKLTGGQVIDVPGDPSSSTAFPLVAALLVPGSDVTILNVLNMPRTTGLILT--LOEMGA  
 AROA\_HAEIN 240 250 260 270 280  
 GNQSYISPKNYLVEGDASSASY-FLAAGAIKGGTVKVTGIGKNSIQGDRFLFADVLEKMG

nk603cp4.pep 290 300 310 320 330 340  
 DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNG  
 AROA\_HAEIN 290 300 310 320 330  
 -----KITWGEDF--IQAEHAELNGIDMDMNHIP---DAAMTIATTALFANSETVIRN

nk603cp4.pep 350 360 370 380 390 400  
 LEEVRKESDRLSAVANGKLKNGVDCDEGETSLVVRGPRDGKGLGNASGAAVATHLDHRI  
 AROA\_HAEIN 340 350 360 370 380 390  
 IYNWRVKETDRLTAMATELRKVGAEVEGEDFIRIQPLP----LNQFKHANITETYNDRHM

nk603cp4.pep 410 420 430 440 450  
 AMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEIELSDTKAA  
 AROA\_HAEIN 400 410 420 430  
 AMCFSLIAL-SNTPVTIILDPKCTAKTFPTFFNEFEKICLKN

nk603cp4.pep  
 GP\_BCT3:U32833\_2

LOCUS U32833\_2 [U32833]  
 DEFINITION Haemophilus influenzae Rd section 148 of 163 of the complete  
 genome;  
 similar to GB:L04686 SP:Q03421 PID:148865 percent  
 identity: 97.69; identified by sequence similarity;  
 putative. . . .

SCORES Init1: 170 Initn: 315 Opt: 420 z-score: 458.6 E(): 1.4e-17  
 Smith-Waterman score: 476; 27.0% identity in 437 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 U32833\_2 10 20 30 40 50  
 MEKITLAPISAVEGTINLPGSKSLSNRALLAALAKGTTKVTNLLDSDDIRHML

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRKEGDTWIIDGVGNGGLLAPEAPLDF--GNAATGCR-LTMGLV--GVYDF  
 U32833\_2 60 70 80 90 100 110  
 NALKALGVRYQLSDDKTICEVEGLGTFNIQDNLSLFLGNAGTAMRPLTAALCLKGKTES

nk603cp4.pep 120 130 140 150 160 170  
 DSTFFIGDASLTAKRPMGRVNLPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIYRVPMASA  
 U32833\_2 120 130 140 150 160 170  
 EIILTGEPRMKRPIHLVDALRQAGADIRYLENEGYPFLAIRNKGKIGKVKIDGSISS

nk603cp4.pep 180 190 200 210 220 229  
 QVKSALLAGLNTPGITTV--IEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLE  
 U32833\_2 180 190 200 210 220 230  
 QFLTALLMSAPLAENDTEIEIIGELVSKPYIDITLAMRDFG----VQVENHHYQKFOVK

AROA\_BURPS GSSMIVCPR--GELRAAVRSDSNSIPDAAMTLATRSAGARWAATANHIRVAGEGDGS  
280 290 300 310 320 330

```

      290      300      310      320      330      340
nk603cp4.pep DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDVEYFILAVAAAFAGATVMNG
               ::|||  :::::  :::::  :::::  :::::  :::::  :::::  :::::
U32833_2      -----KITWGDEF--IQAEHAELNGIDMDMNHIP--DAAMTIATTALFANSETVIRN
               290      300      310      320      330

```

nk603cp4.pep  
GP\_BCT3:x77019\_1

LOCUS X77019\_1 [PPAROA]  
DEFINITION P.pseudomallei aro A gene.  
DATE 18-FEB-1994  
ACCESSION X77019  
NID  
ORGANISM Burkholderia pseudomallei . . .

```

      350          360          370          380          390          400
nk603cp4.pep LEELRVKESDRLSAVANGKLNGVDCDEGETSLVVGRGPDGKGLGNASGAAVATHLDHRI
               : : ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
U32833_2     IYNWRVKETDLRTAMATELRKVGAEEVGEDFIRIQPLP----LNQFKHANIEITYNDHRM
               340          350          360          370          380          390

```

SCORES Init1: 178 Initn: 367 Opt: 414 z-score: 452.5 E(): 3.1e-17  
Smith-Waterman score: 422; 32.7% identity in 309 aa overlap

```

          410         420        430       440      450
nk603cp4.pep AMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
              |||::|||::|||::|   |::|||::|
U32833_2     AMCFSLIAL-SNTPVTILDPKCTARTFTFFNEFEKICLKN
             400           410          420          430

```

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVIRPGDKSISHRSMFPGGLASGETRITGLLEGEDVINTG
              :::||:|||||:||:|:::||:|:||:|||||:::||:|
X77019_1      MAFQWPRFPLQPWRHVTHLRLPGDKSISNRSLLLGALAEQVTEVTGLLSDSDARAML
              10      20      30      40      50

```

nk603cp4.pep  
SWISSPROT:AROA\_BURPS

```

              70          80          90         100        110
nk603cp4.pep KAMQAMGARIR--KEGDTWIIDGVNGGLLAPEAPLDFGNAAATGCRLTMGLVGVDYFDST
               ::::|::|:      :||   || | |::|::|::|:    :::|:  |:|
X77019_1     NALRDLGGVVIIEGFHQGRCTVHGVLHGLKAPPGPFLF L GNA GTAM RPLSAALAQPFDDT
             60       70       80       90      100      110
```

```
ID      AROA_BURPS          STANDARD;          PRT;          409 AA.
AC      P39915;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5-
```

```

120      130      140      150      160      170
nk603cp4.pep FIGDASLTAKRPMGRVLNPLREMG--VQVKSEGDRLPVTLRG----PKTPTPITYRVP
: || :||:|:|:|:| ||| : :|| :|:| : : :|
X77019_1      LRGDPMSERPINRLVDALREMGAVIEYLAQEG-YPPLTLIRGGGSVSSQFLTALLMTAPM
120      130      140      150      160      170

```

SCORES Init1: 178 Initn: 367 Opt: 414 z-score: 452.5 E(): 3.1e-17  
Smith-Waterman score: 422; 32.7% identity in 309 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSMFGGSLASGETRITGLLEGEDVINTG  
:::||:|||||::|:::|:| |||:::|:  
AROA\_BURPS MAFQWPRFPQPWRHVTGHLRLPGDKSISNRSLLLGALAEGVTEVTGLLSDDDARAML  
10 20 30 40 50 60

```

      180      190      200      210      220      229
nk603cp4.pep ASAQVKSALLAGLNTPGITTVIEP--INTRDHTKMLQGFGANPTVETDADGVRTIRLE
      |||||:|::||: || : | : |||||::: :| :|
X77019_1      ASAQIKSGLLLSKPYID-ITLNVMPFGVPVTRDHTERIF-----AVSAIRYP
      180      190      200      210      220

```

```

              70          80          90          100          110
nk603cp4.pep KAMQAMGARIR--KEGDTWIIDGVNGGLLAPEAPLDGNAATGCRLTMGLGVGYDFDST
              :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
ARO_A_BURPS  NALRDLGVVIEGPHQGRC-TVHGVGLHGLKAPPGLPLFLNAGTAMRPLSAALALQPFDDT
              60          70          80          90          100          110

```

```

      230      240      250      260      270      280
nk603cp4.pep GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTQEMGA---
              :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
X77019_1      -----SPAVLRLEGDATSASYFLAAG-IGKVPVVTGIGRHSMQGDSWFFPRALRRMGARSC
              230      240      250      260      270

```

```

120      130      140      150      160      170
nk603cp4.pep FIGDASLTQKPMGRVLNPLREMG--VQVKSEGDRLPVTLRG----PKTPTPTITYRVP
: || ::||::||:: |||| :: ::| ::| ::|
ARO_A_BURPS  LTGDPMSERPIINRLVDALREMGAVIEYLAQEG-YPPLTIRGGGSVSQFLTALMTAPM
120      130      140      150      160      170

```

```

          290      300      310      320      330      340
nk603cp4.pep -DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAF AEGATVMN
              :  :: ||   ||   | : |  |::
X77019_1      GSSMIVCPR--GELRAAVRSDSNSIPDAAMTLATRSAGARWAATANHIRVAGEGDGSAV
              280      290      300      310      320      330

```

```

      180      190      200      210      220      229
nk603cp4.pep  ASAQVKSALLAGLNTPGITTVIEP--IMTRDHTKMLQGFGANPTVETDADGVRTIRLE
               ||||:|::||:  ||  :  : |||||:::  :|:|
AROABURPS     ASAQIKSGLLLSKPYID-ITLNVMPFGVPTRDHTERIF-----AVSAIRYP
               180      190      200      210      220

```

nk603cp4.pep  
TREMBL NEW:AAK97382

```

ID  AAK97382      PRELIMINARY;      PRT;    427 AA.
AC  AAK97382;
DT  01-NOV-2001  (EMBLrel. 63, Created)
DT  01-NOV-2001  (EMBLrel. 63, Last sequence update)
DT  01-NOV-2001  (EMBLrel. 63, Last annotation update)
DE  5-ENOLPYRUVYL-SHIKIMATE-3-PHOSPHATE SYNTHASE

```

```

      230      240      250      260      270      280
nk603cp4.pep GRGKLTGQVIDVPGDPSSTAFLVAALLVPGSDVTILNVLNPNTRTGLIILTQEMCA---
               :  | :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
AROABURPS    -----SPAVLRLEGDATSASYFLAAG-IGKVPVTGIGRHSMOGDSWFFPRALRRMGARSC
               230      240      250      260      270

```

SCORES Init1: 167 Initn: 361 Opt: 408 z-score: 445.8 E(): 7.5e-17  
Smith-Waterman score: 503; 29.5% identity in 440 aa overlap

nk603cp4.pep -DIEVINRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAF AEGATVMN  
: : : | | | : | : : :

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG

```

AAK97382      MLES LTLQPIALINGTVNLP GSKSVSNRALLAALAE GTRLHNLDSDDIRHML
               10      20      30      40      50
nk603cp4.pep  KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDF--GNAATGCR-LTMGLVGVYDFDS
               70      80      90      100     110
AAK97382      NALKSLGVNYRLSADRTQCDVDGLGGPLVADKTLELFLGNA GTAMRPLAALACLGYSS-DI
               60      70      80      90      100     110
nk603cp4.pep  TFIGDASLTKRPMGRVLNPLREMGVQVKS--EDGDRLPVTLRGPKTPTPTITRVPMASAOV
               120     130     140     150     160     170
AAK97382      VLTGEERMKERPIGHLVDAL--EGGAQIDYLEQENYPPLRIRRGFGRGELTVDGSVSS--QF
               120     130     140     150     160     170
nk603cp4.pep  KSAVLLAG-LNTPGITTVE--PIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGR
               180     190     200     210     220     230
AAK97382      LTALLMAAPLATQDTHIRIQGDLVSRPYIDITLHLMRSFG---VEVTHQNYQVFHIQG-
               180     190     200     210     220
nk603cp4.pep  GKL---TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTR--TGLILTQEMGA
               240     250     260     270     280
AAK97382      GQTYHSPGEYL--VEGDASSASYFLAAA--IKGGTVRVGTIGKSVQGDTKFADVLEKMG
               230     240     250     260     270     280
nk603cp4.pep  DIEVINPRLAGGEDVDALVRSSSTLKGVTVPEDRAPSMIDEYPI LAVA AAFAGATVMNG
               290     300     310     320     330     340
AAK97382      IHHW-----GDDY--IECSRGE LRDIDMDMNHIP---DAAMTIATAALFAKGP TTI RN
               290     300     310     320     330
nk603cp4.pep  LEE LR VKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGA AVATHL DHRI
               350     360     370     380     390     400
AAK97382      IYNWRVKETDRLSAMATELRKVGAEEVEEQD--YIRVPPAHLIA---AEIGTYNDHRM
               340     350     360     370     380
nk603cp4.pep  AMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
               410     420     430     440     450
AAK97382      AMCFSLVAL--SDTFVTILD PKCTAKTFPDYFEQLARLSVLA
               390     400     410     420

nk603cp4.pep
SWISSPROT: AROA_METTH

ID   AROA_METTH      STANDARD;      PRT;   419 AA.
AC   O26860;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PROBABLE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) . . .

SCORES   Init1: 136   Initn: 465   Opt: 396   z-score: 432.9 E(): 3.9e-16
Smith-Waterman score: 664;   33.8% identity in 450 aa overlap

nk603cp4.pep  MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
               10      20      30      40      50      60
AROA_METTH    MDLTVEKSGNLEGTVKAPPSKSYTHRAVIAALAEGVSEIRDPLIAEDTLSSL
               10      20      30      40      50

nk603cp4.pep  KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDFGNAATGCRLTMGLVGVYDFDSTFI
               70      80      90      100     110     120
AROA_METTH    NACRAFGIRV-DEGDAWTVHGSG--GELETPDDVIYLGNSGTTLRMTSVAGLAENYTVLT
               60      70      80      90      100     110

```

```

               60      70      80      90      100     110
nk603cp4.pep  GDASLTKRPMGRVLNPLREMGVQVKS EDGDRLP-VTLRGPKTPTPTITRVPMASAOVKSA
               130     140     150     160     170
AROA_METTH    GDES LRTRPMQPLLDALRPLGV EALSSRMNGLPPIIVRGGLRGGSTSIRGDVSS--QFISS
               120     130     140     150     160     170

               180     190     200     210     220     230
nk603cp4.pep  VLLAGLNTPGITTVE--PIMTR---DHTEKMLQGFGANPTVETD-ADGVRTIRLEGRGKL
               180     190     200     210     220
AROA_METTH    ILIAAPLTGVEVMVEGDFISRPYVDMTVDMVERF---SVPVDYSEG--TFRVEP-AVY
               180     190     200     210     220

               240     250     260     270     280     290
nk603cp4.pep  TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTG--LILTL-QEMGADIEVI
               240     250     260     270     280
AROA_METTH    RGLDYTVEGDYSSASY-LAGAVAAAGDVL IEN-LFRDSRQGRDIILDIISDMGAEVR-
               230     240     250     260     270     280

               300     310     320     330     340     350
nk603cp4.pep  NPRLAGGEDVDALVRSS--TLKGVTVPEDRAPSMIDEYPI LAVA AAFAGATVMNGLEEL
               300     310     320     330
AROA_METTH    -----GED--HVR IASTGELSGVSVNLHDA PDL L---PTVAVLGALATGRTEIGGVEHA
               290     300     310     320

               360     370     380     390     400     410
nk603cp4.pep  RVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGA AVATHL DHRIAMSP
               360     370     380     390     400
AROA_METTH    RYKETDRISTCAAE LRRLGVDVTBLPDGM IEG-----GASG GTVWSHGDRHLAMAF
               330     340     350     360     370     380

               420     430     440     450
nk603cp4.pep  LVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
               420     430     440
AROA_METTH    TLIGL--REGITIRDAEVFSVSFPDFPERMMQIGCRMNLS
               390     400     410

nk603cp4.pep
GP_BCT1: AE000855_4

LOCUS      AE000855_4 [AE000855]
DEFINITION Methanobacterium thermoautotrophicum from bases 691033 to 702432
              (section 61 of 148) of the complete genome;
              Function Code: 5.14 - L-Amino Acid Metabolism,
              Phenylalanine--tyrosine and tryptophan biosynthesis ;
              similar to, pir:LN:F64362 AC:F64362, p()=1.8E-98, pid=45%. . . .

SCORES   Init1: 136   Initn: 465   Opt: 396   z-score: 432.9 E(): 3.9e-16
Smith-Waterman score: 664;   33.8% identity in 450 aa overlap

nk603cp4.pep  MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
               10      20      30      40      50      60
AE000855_4    MDLTVEKSGNLEGTVKAPPSKSYTHRAVIAALAEGVSEIRDPLIAEDTLSSL
               10      20      30      40      50

               70      80      90      100     110     120
nk603cp4.pep  KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDFGNAATGCRLTMGLVGVYDFDSTFI
               70      80      90      100     110
AE000855_4    NACRAFGIRV-DEGDAWTVHGSG--GELETPDDVIYLGNSGTTLRMTSVAGLAENYTVLT
               60      70      80      90      100     110

               130     140     150     160     170
nk603cp4.pep  GDASLTKRPMGRVLNPLREMGVQVKS EDGDRLP-VTLRGPKTPTPTITRVPMASAOVKSA
               130     140     150     160     170
AE000855_4    GDES LRTRPMQPLLDALRPLGV EALSSRMNGLPPIIVRGGLRGGSTSIRGDVSS--QFISS
               120     130     140     150     160     170

```

nk603cp4.pep 180 190 200 210 220 230  
VLLAGLNTPTGITTVE-PIMTR---DHTEKMLQGGFANPTVETD-ADGVRTIRLEGKGL  
AE000855\_4 180 190 200 210 220  
ILIAAPLTEGVEMVEGDFISRPYVDMTVDVMERF---SVPVDYSEG--TFRVEP-AVY

nk603cp4.pep 240 250 260 270 280 290  
TGQVIDVPGDPSSTAFFLVAALLVPGSDVTILNVLNPTRTG--LILTL-QEMGADIEVI  
AE000855\_4 230 240 250 260 270 280  
RGLDYTVEGDYSSASY-LAGAVAAAGGDVLLEN-LFRDSRQGGRIILDIISDMGAEVRR-

nk603cp4.pep 300 310 320 330 340 350  
NPRLAGGEDVADLRVRSS-TLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGLEE  
AE000855\_4 290 300 310 320 330 340  
-----GED--HVRIASGTGELSGVSVNLHDAPDLL---FTVAVLGALATGRTEIGGVEHA

nk603cp4.pep 360 370 380 390 400 410  
RVKESDRLSAVANGLKNGVDCDEGETSLVVRGRPDGKGLGNASGAATHLDRHRIAMSF  
AE000855\_4 330 340 350 360 370 380  
RYKETDRISTCAAEALRRLLGVDTLPDGMIIIEG-----GASGGTVWSHGDRHRLAMAF

nk603cp4.pep 420 430 440 450  
LVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AE000855\_4 390 400 410  
TLIGL--REGITIRDAEVFSVSPDFPERMMQIGCRMNLS

nk603cp4.pep  
GP\_BCT1:AE005280\_4

LOCUS AE005280\_4 [AE005280]  
DEFINITION Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 104 of 155;  
Residues 1 to 427 of 427 are 100.00 pct identical  
to residues 1 to 427 of 427 from Escherichia coli K-12  
Strain MG1655: B0908. . . .

SCORES Init1: 163 Initn: 413 Opt: 392 z-score: 428.5 E(): 6.9e-16  
Smith-Waterman score: 493; 27.8% identity in 435 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHSRSMFGGLASGETRITGLLEGEDVINTG  
AE005280\_4 10 20 30 40 50  
MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTNLLSDDDVRHML

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDF--GNAATGCRLTMGLVGVDYDFDST  
AE005280\_4 60 70 80 90 100 110  
NALTALGVSYTSLADTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAAALCLGSNDIV

nk603cp4.pep 120 130 140 150 160 170  
FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIRVPMASQVK  
AE005280\_4 120 130 140 150 160 170  
LTGEPKMKERPIGHLVDALRLGGAKITYLEQENYPPRLRQGGFTGGNVVDVGSVSS-QFL

nk603cp4.pep 180 190 200 210 220 230  
SAVLLAGLNTPT-GITTVIEPIMTR---DHTEKMLQGGFANPTVETDADGVRTIRLEG-R  
AE005280\_4 180 190 200 210 220 230  
TALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFG----VEIENQHYQQFVVKGGQ

nk603cp4.pep 240 250 260 270 280 289  
GKLTGQVIDVPGDPSSTAFFLVAALLVPGSDVTILNVLNPTRTGLILT--LQEMGADIE  
D90728\_13 230 240 250 260 270 280  
SYQSPGTYLVEGDASSASYFLAAAA-IGGGTVKVTGIGRNSMQGDIRFADVLEKMGATI-

AE005280\_4 230 240 250 260 270 280  
SYQSPGTYLVEGDASSASYFLAAAA-IGGGTVKVTGIGRNSMQGDIRFADVLEKMGATI-

nk603cp4.pep 290 300 310 320 330 340 349  
VINPRLAGGEDVADLRVRSS-TLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGLEE  
AE005280\_4 290 300 310 320 330  
-----CWGDDY--ISCTRGELNAIDMDMNHIP---DAAMTIATAALFAKGTTLRLNIYN

nk603cp4.pep 350 360 370 380 390 400 409  
LRVKESDRLSAVANGLKNGVDCDEGETSLVVRGRPDGKGLGNASGAATHLDRHRIAMS  
AE005280\_4 340 350 360 370 380 390  
WRVKETDRLFAMATELRKVGAEEVEGHD--YIRITPPEK----LNFABEATYNDHRMAMC

nk603cp4.pep 410 420 430 440 450  
FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AE005280\_4 400 410 420  
FSLVAL-SDTFTILDPKCTAKTFPDYFEQLARISQAA

nk603cp4.pep  
GP\_BCT3:D90728\_13

LOCUS D90728\_13 [D90728]  
DEFINITION Escherichia coli genomic DNA. (20.4 - 20.8 min);  
ORF\_ID: o217#5  
similar to PIR Accession Number A30370.  
DATE 07-FEB-1999  
ACCESSION D90728 . . .

SCORES Init1: 163 Initn: 413 Opt: 392 z-score: 428.5 E(): 6.9e-16  
Smith-Waterman score: 493; 27.8% identity in 435 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHSRSMFGGLASGETRITGLLEGEDVINTG  
D90728\_13 10 20 30 40 50  
MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTNLLSDDDVRHML

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDF--GNAATGCRLTMGLVGVDYDFDST  
D90728\_13 60 70 80 90 100 110  
NALTALGVSYTSLADTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAAALCLGSNDIV

nk603cp4.pep 120 130 140 150 160 170  
FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIRVPMASQVK  
D90728\_13 120 130 140 150 160 170  
LTGEPKMKERPIGHLVDALRLGGAKITYLEQENYPPRLRQGGFTGGNVVDVGSVSS-QFL

nk603cp4.pep 180 190 200 210 220 230  
SAVLLAGLNTPT-GITTVIEPIMTR---DHTEKMLQGGFANPTVETDADGVRTIRLEG-R  
D90728\_13 180 190 200 210 220  
TALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFG----VEIENQHYQQFVVKGGQ

nk603cp4.pep 240 250 260 270 280 289  
GKLTGQVIDVPGDPSSTAFFLVAALLVPGSDVTILNVLNPTRTGLILT--LQEMGADIE  
D90728\_13 230 240 250 260 270 280  
SYQSPGTYLVEGDASSASYFLAAAA-IGGGTVKVTGIGRNSMQGDIRFADVLEKMGATI-

```

          410          420          430          440          450
nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
               |:::|:::|:::|:::|:::|:::|:::|:::|
D90728_13     FSLVAL-SDTPVTILDPKCTAKTFPDYFEQLARISQAA
               400          410          420

```

```

LOCUS      AE000193_3 [AE000193]
DEFINITION Escherichia coli K12 MG1655 section 83 of 400 of the complete
            genome;
            o427; 99 pct identical to AROA_ECOLI SW: P07638.
DATE       01-DEC-2000
ACCESSION  AE000193

```

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRFSFMGGLASGETRITGLLEGEDVINTG  
AE000193\_3 MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTNLLDSDDVHRHML

```

nk603cp4.pep      120       130       140       150       160       170
                  FIGDASLTGRPMGRVLNPLREMGVQVKSE-EDGDRLPVTLRGPKTTPPTITYRVPMSAAQVK
AE000193_3        :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
                  LTGEPRMKERPIGLHVDALRLGGAKITYLEQENYPPLRLQGFGTTGGNVDDGSVSS-QFL
                  120       130       140       150       160       170

```

nk603cp4.pep      240          250          260          270          280          289  
                  GKLTQGVIDVPGDPSSTAFPLVAALLVPGSVDVTILNVLMPNPTRTGLILT--LQEMGADIE  
                  : : : | ||| :::: :|| : :| : : : | : : : : | : : |||  
AE000193\_3       SYSGPGTYLVEGDASSASYFLAAA-IGGGTVKVTIGRNSMQGDIRFADVLEKMGATI-  
                  230          240          250          260          270          280

nk603cp4.pep      350            360            370            380            390            400            409  
                  LRVKESDRLSAVANGKLNGVDCDEGTSLVVGRPGKGLGNASGAAVATHLDHRIAMS  
                  |||||:||| :||: | : ||::||: :||| :||| :||| :||| :||| :||| :||| :|||  
 AE000193\_3      WRVKETDRLFAMATELRKVGAEEVEGHD--YIRITPPEK-----LNFAEIATYNDRHMAMC  
                  340            350            360            370            380            390            400

nk603cp4.pep

GP\_BCT3:D90729\_6

LOCUS D90729\_6 [D90729]  
 DEFINITION Escherichia coli genomic DNA. (20.5 - 20.9 min);  
 ORF\_ID:o217#5  
 similar to PIR Accession Number A30370.  
 DATE 07-FEB-1999  
 ACCESSION D90729 . . .

SCORES Init1: 163 Initn: 413 Opt: 392 z-score: 428.5 E(): 6.9e-16  
 Smith-Waterman score: 493; 27.8% identity in 435 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 D90729\_6 MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTLNLLSDDDVRHIL

70 80 90 100 110  
 nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--GNAATGCRITMGLVGVYDFDST  
 D90729\_6 NALTALGVSYTLSADTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAALCLGSSNDIV

120 130 140 150 160 170  
 nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS--EDGDRLPVTLRGPKTPTPTITYRVPMASQVK  
 D90729\_6 LTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPRLQGFTGGNVDDVSGVSS--QFL

180 190 200 210 220 230  
 nk603cp4.pep SAVLLAGLNTPT--GITTIVIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEG-R  
 D90729\_6 TALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTKTFG---VEIENQHYQQFVVKGGQ

240 250 260 270 280 289  
 nk603cp4.pep GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGADIE  
 D90729\_6 SYQSPGTLYVEGDASSASYFLAAAA--IKGGTVKVTGIGRNSMQGDIRFADVLEKMGATI-

290 300 310 320 330 340 349  
 nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEE  
 D90729\_6 -----CWGDDY--ISCTRGELNAIDMDMNHIP---DAAMTIATAALFAKGTTLRLNIYN

350 360 370 380 390 400 409  
 nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
 D90729\_6 WRVKETDRLFAMATELRKVGAEEVGHGHD--YIRITPPEK---LNFABEATYNDHRMAMC

410 420 430 440 450  
 nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 D90729\_6 FSLVAL--SDTPVTILDPKCTAKTFFPDYFQLARISQAA

nk603cp4.pep  
 SWISSPROT:ARO\_A\_ECOLI

ID ARO\_A\_ECOLI STANDARD; PRT; 427 AA.  
 AC P07638; P78222;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 163 Initn: 413 Opt: 392 z-score: 428.5 E(): 6.9e-16  
 Smith-Waterman score: 493; 27.8% identity in 435 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 ARO\_A\_ECOLI MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTLNLLSDDDVRHML

70 80 90 100 110  
 nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--GNAATGCRITMGLVGVYDFDST  
 ARO\_A\_ECOLI NALTALGVSYTLSADTRCEIIGNGGPLHAEGALELFLGNAGTAMRPLAALCLGSSNDIV

120 130 140 150 160 170  
 nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS--EDGDRLPVTLRGPKTPTPTITYRVPMASQVK  
 ARO\_A\_ECOLI LTGEPRMKERPIGHLVDALRLGGAKITYLEQENYPPRLQGFTGGNVDDVSGVSS--QFL

180 190 200 210 220 230  
 nk603cp4.pep SAVLLAGLNTPT--GITTIVIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEG-R  
 ARO\_A\_ECOLI TALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTKTFG---VEIENQHYQQFVVKGGQ

240 250 260 270 280 289  
 nk603cp4.pep GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGADIE  
 ARO\_A\_ECOLI SYQSPGTLYVEGDASSASYFLAAAA--IKGGTVKVTGIGRNSMQGDIRFADVLEKMGATI-

290 300 310 320 330 340 349  
 nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEE  
 ARO\_A\_ECOLI -----CWGDDY--ISCTRGELNAIDMDMNHIP---DAAMTIATAALFAKGTTLRLNIYN

350 360 370 380 390 400 409  
 nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
 ARO\_A\_ECOLI WRVKETDRLFAMATELRKVGAEEVGHGHD--YIRITPPEK---LNFABEATYNDHRMAMC

410 420 430 440 450  
 nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 ARO\_A\_ECOLI FSLVAL--SDTPVTILDPKCTAKTFFPDYFQLARISQAA

nk603cp4.pep  
 GP\_BCT1:AF101225\_1

LOCUS AF101225\_1 [AF101225]  
 DEFINITION Shigella sonnei 3-enolpyruvylshikimate-5-phosphate synthetase  
 (aroA) gene, complete cds.  
 DATE 11-NOV-1998  
 ACCESSION AF101225  
 NID . . .

SCORES Init1: 163 Initn: 413 Opt: 391 z-score: 427.4 E(): 7.9e-16  
 Smith-Waterman score: 492; 27.8% identity in 435 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG



```

          240          250          260          270          280
nk603cp4.pep  RGKLTGQ-VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTR--TGLILTLQEMGAD
              :  :  |  :  :  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
M20023_1      DAVYVPGGRMAIEGDASTASY-FLALGAIGGGPVRTVGVEDSIQGDDVAFAATLAAMGAD
              240          250          260          270          280          290

```

```

      290      300      310      320      330      340
nk603cp4.pep IEVINPRLAGGEDVADLRV-RSSTLKGVTVPEDRAPMSIDEYPIILAVAAAPAEAGATVMNG
      ::  :|  |  ::  ::||  ::||  ::  :  :  |  |  :|  :|  :|  :|  :|  :|
M20023_1      VRY-GP---GWIEETGRVRAEGGRLLKAFDAFNLIP---DAAMTATLALYADGPCRLRN
      300      310      320      330      340

```

```

      350          360          370          380          390          400
nk603cp4.pep LEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRI
              : |||:|||: |: |: |: |: | |: |: |: |: |: |: |: |: |: |: |: 
M20023_1     IGSWRVKETDRIHAMHTELEKLGAGVQSGADWLEF-APPEPGGWRAH---IGTWDDHRM
              350          360          370          380          390          400

```

```

      410      420      430      440      450
nk603cp4.pep  AMSFLVMLGVSENPVTVDATMIATSPFPEFMDLMAGLGAKIELSDTKAA
               |||:::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
M20023_1      AMCFLLAAF-GPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
               410      420      430      440

```

```

ID  AROA_BORPE          STANDARD;          PRT;    442 AA.
AC  P12421;
DT  01-OCT-1989 (Rel. 12, Created)
DT  01-OCT-1989 (Rel. 12, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5-

```

```

      10      20      30      40      50      60
nk603cp4.ppep MLHGASSRPATARKSSGLSGTVIRPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
                | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
AROA_BORPE      MSGLAYLDLPAARLARGEVALPGSKSISNRVLLLAALAEGSTTEITGLLSDSDTRVML
                10      20      30      40      50

```

```

              70          80          90         100        110       119
nk603cp4.pep KAMQAMGARIRKEGDTWI-IDGVNGGLLAPEAPLDFGNAAATGCRLTMGLVGVDYFDSTF
               |: :| : : | : |: || : : :|| : ||| :| : : :   | : : :    | : 
AROA_BORPE   AALRQLGVSVGEADCVCTIEGVAR--FPTEQAELFLNAGTAFRPPLTAALMALMGDDYRL
             60      70     80      90      100     110

```

```

120      130      140      150      160      170
nk603cp4.pep IGDSLTKRPMGRVLNPLREMGVQVK--SEDGRLPVTLRGPKTPTPTITYRVP-MASAQV
          |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
AROA_BORPE  SGVPRMHERPIGDLVDLRQFGAGIEYLQGAG-YPPLRIGGGSIRVDGPFVRVEGSVSSQF
          |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
          120      130      140      150      160      170

```

```

      180      190      200      210      220      230
nk603cp4.pep KSAVLLAG-----LNTPGIIT-VIEPIIMTRDTEKMLQGGFANPTVETDADGVRTIRLEG
              :::::      |||  |  :::::  |  :::  |  :::  |  :::  |  :::  |
AROA_BORPE   LTALLMAAPLARRSGDITIEVVGELISKPYIEITLN-LMARFGVSVRRDGWRAFTIAR
              180      190      200      210      220      230

```

nk603cp4.pep      240      250      260      270      280  
 RGKLTGQ-VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMPNTR--TGLILTLQEMGAD  
 : | : : | | : : : : | : { : | : : | : : : : | | | |

```

      290      300      310      320      330      340
nk603cp4.pep IEVINPRLAGGEDVADLRV--RSSTLKGVTVPEDRAPSMIDEYPI LAVAAAPAEAGATVMNG
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
AROA_BORPE VRY-GP---GWIETRGVRVAEGRRLKAFDADFNLIPI---DAAMTAATLALYADGPGCLRLN
      300      310      320      330      340

```

```

      350          360          370          380          390          400
nk603cp4.pep LEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRRPDPKGKLGNASGAAVATHLDHRI
              |||:|||::||::||::||::||::||::||::||::||::||::||::||
AROABORPE    IGSWRVKETDRIHAMTELEKLGAGVQSGADWLEV-APPEPGGWRDAH---IGTWDDHRM
              350          360          370          380          390          400

```

```

      410      420      430      440      450
nk603cp4.pep AMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
AROA_BORPE  AMCFLLAAF-GPAAVRILDPGCVSKTFPDYFDVYAGLLAARD
      410      420      430      440

```

```

LOCUS      X00557_1 [ECAROA]
DEFINITION E.coli gene araA for 5-enolpyruvylshikimate 3-phosphate synthase
            (EPSP synthase, EC 2.5.1.19, alternative name 3-phosphoshikimate
            1-carboxyvinyltransferase);
            EPSP synthase.
DATE      12-SEP-1993

```

[illegible]

```

              70          80          90          100         110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDF--GNAATGCRLTMGLVGVDYFDST
               :|::||:      :|   :|::||| |::|:  ::|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:~
X00557_1     NALTALGVSYTLSADRTCEIINGCGPLHAEGALELFLGNAGTAMRPLAALCLGSNDIV
               60          70          80          90          100         110

```

[illegible]

```

180      190      200      210      220      230
nk603cp4.pep SAVLLAGLNTPT--GITTVEIEPTMR---DHTEKMLQGFGANPTVETDADGCVTRIRLEG-R
:|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
X00557_1      TALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTFTG----VEIENQHYQQFVVKGGQ
180      190      200      210      220

```

```

      240      250      260      270      280      289
nk603cp4.pep  GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILT--LQEMGADIE
               : : : | | : : | | : : | : : : : : : : : : | : | |
X00557_1      SYQSPGTYLVEGDASSASYFLAAAA-IKGGTVKVTGIGRNSMQGDIRFADVLEKMGATI-
      230      240      250      260      270      280

```

```

      290      300      310      320      330      340      349
mk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMID EYPILAVAAFAEGATVMNGLEE
              |:| : : : : : : : : : : : : : : : : : : :
X00557_1      -----CWGGDY--ISCTRGELNAIDMDMNHIP--DAAMTIATAALFAKGTTRLRNIYN
              290      300      310      320      330      340

```



nk603cp4.pep 350 360 370 380 390 400 409  
 LRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
 X00557\_1 340 350 360 370 380 390  
 WRVKETDRLFAMATELRKVGAEEVGHGHD--YIRITPPEK---LNFAEITATYNDHRMAMC

nk603cp4.pep 410 420 430 440 450  
 FLVMGLVSENVPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 X00557\_1 400 410 420  
 FSLVAL-SDTPVTILDPKCTAKTFPDYFEQLARISQAA

nk603cp4.pep  
 SWISSPROT:AROA\_EDWIC

ID AROA\_EDWIC STANDARD; PRT; 428 AA.  
 AC Q9X4H2;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 136 Initn: 357 Opt: 386 z-score: 422.0 E(): 1.6e-15  
 Smith-Waterman score: 475; 28.3% identity in 442 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_EDWIC 10 20 30 40 50  
 MSSALTLPQVRRFSGEINLPGSKSVSNRALLLAAQARGVTRLNHLNLLSDDVRYML

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLD--FGNAATGCRLTMGLVGVYDFDST  
 AROA\_EDWIC 60 70 80 90 100 110  
 DALKALGVRYQLSDCRTRCEVQGLGGTSLAHGALTFLGNAGTAMRPLAALSLGLRDI

nk603cp4.pep 120 130 140 150 160 170  
 FIGDASLTKRPMGRVNLPLREMGVQVKSSEDGDRLP-VTLRGPKTPPTITYRVPMSAQVK  
 AROA\_EDWIC 120 130 140 150 160 170  
 LTGEPRMKERPIAHLVTALRQGGAGVDYLEDGYPPVRLHGGFNGGEISVDGVS--QFL

nk603cp4.pep 180 190 200 210 220 230  
 SAVLLAGLNTPGIT--TVIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEGRG  
 AROA\_EDWIC 180 190 200 210 220  
 TALMAAPMAAEETRIITILGELVSKPYIAITLAMMRAFG---VEVENHAYRHFVVRG--

nk603cp4.pep 240 250 260 270 280  
 KLTGQVIDVP-----GDPSSSTAPFLVAALLVPGSDVTILNVLNPTRTGLILT--LQEM  
 AROA\_EDWIC 230 240 250 260 270 280  
 ---GQVYQAPSDYLVGEGDASSASYFLAGAAIAGGT-VRVTGIGRHSMQGDIHFADVLEKM

nk603cp4.pep 290 300 310 320 330 340  
 GADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVM  
 AROA\_EDWIC 290 300 310 320 330  
 GAQVEWGNNDYIACTRD-----SLHGIDMDMNAIP---DAAMTIATTALFAKGPTTL

nk603cp4.pep 350 360 370 380 390 400  
 NGLLELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDH  
 AROA\_EDWIC 340 350 360 370 380  
 RNIYNWRVKETDRLAAMASELRKVGAUVEEGTDFL--RIEPPAQ---LQAAQIATYNDH

410 420 430 440 450

nk603cp4.pep RIAMSFLVMGLVSENVPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROA\_EDWIC 390 400 410 420  
 RMAMCFSLVAL-SGTPVTICDPGCTAKTFPDYFRQFSALCHPR

nk603cp4.pep  
 GP\_BCT2:AF110153\_2

LOCUS AF110153\_2 [AF110153]  
 DEFINITION Edwardsiella ictaluri phosphoserine transaminase (serC) gene,  
 partial cds; and 5-enolpyruvylshikimate-3-phosphate synthase  
 (aroA) gene, complete cds.  
 DATE 02-MAY-1999  
 ACCESSION AF110153 . . .

SCORES Init1: 136 Initn: 357 Opt: 386 z-score: 422.0 E(): 1.6e-15  
 Smith-Waterman score: 475; 28.3% identity in 442 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AF110153\_2 10 20 30 40 50  
 MSSALTLPQVRRFSGEINLPGSKSVSNRALLLAAQARGVTRLNHLNLLSDDVRYML

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLD--FGNAATGCRLTMGLVGVYDFDST  
 AF110153\_2 60 70 80 90 100 110  
 DALKALGVRYQLSDCRTRCEVQGLGGTSLAHGALTFLGNAGTAMRPLAALSLGLRDI

nk603cp4.pep 120 130 140 150 160 170  
 FIGDASLTKRPMGRVNLPLREMGVQVKSSEDGDRLP-VTLRGPKTPPTITYRVPMSAQVK  
 AF110153\_2 120 130 140 150 160 170  
 LTGEPRMKERPIAHLVTALRQGGAGVDYLEDGYPPVRLHGGFNGGEISVDGVS--QFL

nk603cp4.pep 180 190 200 210 220 230  
 SAVLLAGLNTPGIT--TVIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEGRG  
 AF110153\_2 180 190 200 210 220  
 TALMAAPMAAEETRIITILGELVSKPYIAITLAMMRAFG---VEVENHAYRHFVVRG--

nk603cp4.pep 240 250 260 270 280  
 KLTGQVIDVP-----GDPSSSTAPFLVAALLVPGSDVTILNVLNPTRTGLILT--LQEM  
 AF110153\_2 230 240 250 260 270 280  
 ---GQVYQAPSDYLVGEGDASSASYFLAGAAIAGGT-VRVTGIGRHSMQGDIHFADVLEKM

nk603cp4.pep 290 300 310 320 330 340  
 GADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVM  
 AF110153\_2 290 300 310 320 330  
 GAQVEWGNNDYIACTRD-----SLHGIDMDMNAIP---DAAMTIATTALFAKGPTTL

nk603cp4.pep 350 360 370 380 390 400  
 NGLLELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDH  
 AF110153\_2 340 350 360 370 380  
 RNIYNWRVKETDRLAAMASELRKVGAUVEEGTDFL--RIEPPAQ---LQAAQIATYNDH

nk603cp4.pep 410 420 430 440 450  
 RIAMSFLVMGLVSENVPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AF110153\_2 390 400 410 420  
 RMAMCFSLVAL-SGTPVTICDPGCTAKTFPDYFRQFSALCHPR

nk603cp4.pep  
 SWISSPROT:AROA\_BORBR

ID AROA\_BORBR STANDARD; PRT; 442 AA.  
 AC Q9RND7;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 123 Initn: 212 Opt: 385 z-score: 420.7 E(): 1.9e-15  
 Smith-Waterman score: 385; 26.9% identity in 442 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_BORBR MSGLAYLDLPAARLARGEVALPGSKSISNRVLLAALAEGSTTEITGLDSDTRVML

70 80 90 100 110 119  
 nk603cp4.pep KAMQAMGARIRKEGDTWI-IDGVGNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 AROA\_BORBR AALRQLGVSVGEVADGRVTIEGVAR--FPTEQAELFLGNAGTAFRPLTAALALMGGDYRL

120 130 140 150 160 170  
 nk603cp4.pep IGDASLTKRPMGRVNLNPLREMGVQVK--SEDGDRLPVTLRGPKTPTPIYRVP-MASAOV  
 AROA\_BORBR SGVPRMHERPIGDLVDALRQFGAGIEYLGQAG-YPLRIGGSGIRVDGPVRVEGSSVSQF

180 190 200 210 220 230  
 nk603cp4.pep KSAVLLAG-----LNTPGITT-VIEPIMTRDHTKMLQGFANPTVETDADGVRTIRLEG  
 AROA\_BORBR LTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLN-LMARFGVSVRRDGWRAFTIAR

240 250 260 270 280  
 nk603cp4.pep RGKLTGQ-VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTR--TGLILTQEMGAD  
 AROA\_BORBR DAAVRGPGRMALIEGDASTASY-FLALGAIGGGPVRVTGVEDSIQGDVAFATLAAMGAD

290 300 310 320 330 340  
 nk603cp4.pep IEVINPRLAGGEDVADLRV-RSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNG  
 AROA\_BORBR VRY-GP---GWIEIRGVRVAEGGRLKAFDADFNLP---DAAMTAATLALYADGPCRLRN

350 360 370 380 390 400  
 nk603cp4.pep LEELRVKESDRLSAVANGLKLVDCDEGETSLVVRGPDGKGLGNASGAAVATHLDHRI  
 AROA\_BORBR IGSWRVKETDRIHAMHTELEKLGAGVQSGADWLEVA--PPAPG--GWRDAHIGTWDDHRM

410 420 430 440 450  
 nk603cp4.pep AMSFLVMGLVSENP--VTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROA\_BORBR AMCF---SLAAGFPAARVILDPGCVSKTFPDYFDVYAGLLAARD

nk603cp4.pep  
 GP\_BCT2:AF182427\_3

LOCUS AF182427\_3 [AF182427]  
 DEFINITION Bordetella bronchiseptica prephenate dehydratase (pheA) gene,  
 partial cds; prephenate dehydrogenase (tyrA),  
 5-enolpyruvylshikimate-3-phosphate synthase (aroA), and cytidine  
 monophosphate kinase (cmk) genes, complete cds; and ribosomal  
 protein S1 (rpsA) gene, partial cds; . . .

SCORES Init1: 123 Initn: 212 Opt: 385 z-score: 420.7 E(): 1.9e-15  
 Smith-Waterman score: 385; 26.9% identity in 442 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AF182427\_3 MSGLAYLDLPAARLARGEVALPGSKSISNRVLLAALAEGSTTEITGLDSDTRVML

70 80 90 100 110 119  
 nk603cp4.pep KAMQAMGARIRKEGDTWI-IDGVGNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 AF182427\_3 AALRQLGVSVGEVADGRVTIEGVAR--FPTEQAELFLGNAGTAFRPLTAALALMGGDYRL

120 130 140 150 160 170  
 nk603cp4.pep IGDASLTKRPMGRVNLNPLREMGVQVK--SEDGDRLPVTLRGPKTPTPIYRVP-MASAOV  
 AF182427\_3 SGVPRMHERPIGDLVDALRQFGAGIEYLGQAG-YPLRIGGSGIRVDGPVRVEGSSVSQF

180 190 200 210 220 230  
 nk603cp4.pep KSAVLLAG-----LNTPGITT-VIEPIMTRDHTKMLQGFANPTVETDADGVRTIRLEG  
 AF182427\_3 LTALLMAAPVLARRSGQDITIEVVGELISKPYIEITLN-LMARFGVSVRRDGWRAFTIAR

240 250 260 270 280  
 nk603cp4.pep RGKLTGQ-VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTR--TGLILTQEMGAD  
 AF182427\_3 DAAVRGPGRMALIEGDASTASY-FLALGAIGGGPVRVTGVEDSIQGDVAFATLAAMGAD

290 300 310 320 330 340  
 nk603cp4.pep IEVINPRLAGGEDVADLRV-RSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNG  
 AF182427\_3 VRY-GP---GWIEIRGVRVAEGGRLKAFDADFNLP---DAAMTAATLALYADGPCRLRN

350 360 370 380 390 400  
 nk603cp4.pep LEELRVKESDRLSAVANGLKLVDCDEGETSLVVRGPDGKGLGNASGAAVATHLDHRI  
 AF182427\_3 IGSWRVKETDRIHAMHTELEKLGAGVQSGADWLEVA--PPAPG--GWRDAHIGTWDDHRM

410 420 430 440 450  
 nk603cp4.pep AMSFLVMGLVSENP--VTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AF182427\_3 AMCF---SLAAGFPAARVILDPGCVSKTFPDYFDVYAGLLAARD

nk603cp4.pep  
 GP\_BCT3:U82268\_1

LOCUS U82268\_1 [SDU82268]  
 DEFINITION Shigella dysenteriae EPSP synthase AroA (aroA) gene, complete cds.  
 DATE 26-AUG-1998  
 ACCESSION U82268  
 NID  
 ORGANISM Shigella dysenteriae . . .

SCORES Init1: 163 Initn: 413 Opt: 382 z-score: 417.7 E(): 2.7e-15  
 Smith-Waterman score: 483; 27.6% identity in 435 aa overlap

10 20 30 40 50 60  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 U82268\_1 MESLTLQPIARVDGTINLPGSKSVSNRALLAALAHGKTVLTNLLDSDVVRHML

```

      10      20      30      40      50
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDF--GNAATGCRLTMGLVGVDYFDST
      70      80      90     100     110
U82268_1  NALAALGVSYTSLADTRCEIIGNGGSLHAEGALELFLGNAGTAMRPLAALCLGNSDIV
      60      70      80      90     100     110

      120     130     140     150     160     170
nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTITYRVPMSAQVK
      70      80      90     100     110
U82268_1  LTGEPRMKERPIGHLVDALRLGRAKITYLEQENYPPLRLQGQFTGGNVVDGVS--QFL
      120     130     140     150     160     170

      180     190     200     210     220     230
nk603cp4.pep SAVLLAGLNTPT--GITTVEIPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEG-R
      70      80      90     100     110
U82268_1  TALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTKTFG----VEIENQHYQQFVVKGGQ
      180     190     200     210     220

      240     250     260     270     280     289
nk603cp4.pep GKLTGQVIDVPDPSSTAFPLVAALLVFGSDVTILNVLNPNTRTGLILT--LQEMGADIE
      70      80      90     100     110
U82268_1  SYQSPGTYLVEGDASSASYFLAAAA-IKGGTVKVTGIGRNSMQGDIRFADVLEKMGATT-
      230     240     250     260     270     280

      290     300     310     320     330     340     349
nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAFAGATVMNGLEE
      70      80      90     100     110
U82268_1  -----CWGDDY--ISCTRGELNAIDMDMNHIP---DAAMTIATAALFARKGTTTLRNIYN
      290     300     310     320     330

      350     360     370     380     390     400     409
nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS
      70      80      90     100     110
U82268_1  WRVKETDRLFAMATELRKVGAEVEEGHD--YIRITPPEK----LNFAEITATYNDHRMAMC
      340     350     360     370     380     390

      410     420     430     440     450
nk603cp4.pep FLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      70      80      90     100     110
U82268_1  FSLVAL-SDTPVTILDPKCTAKTFPDYFEQLARISQAA
      400     410     420

```

nk603cp4.pep  
SWISSPROT: AROA\_SHIDY

```

ID   AROA_SHIDY   STANDARD;       PRT;   427 AA.
AC   O87006;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

```

SCORES Init1: 163 Initn: 413 Opt: 382 z-score: 417.7 E(): 2.7e-15  
Smith-Waterman score: 483; 27.6% identity in 435 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
      70      80      90     100     110
AROAS_SHIDY  MESLTLQPIARVDGTINLPGSKSVSNRALLLAALAHGKTVLTNLLDSDDVHRHML
      10      20      30      40      50

```

```

      70      80      90     100     110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDF--GNAATGCRLTMGLVGVDYFDST
      70      80      90     100     110
AROAS_SHIDY  NALAALGVSYTSLADTRCEIIGNGGSLHAEGALELFLGNAGTAMRPLAALCLGNSDIV
      60      70      80      90     100     110

```

```

      120     130     140     150     160     170
nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTITYRVPMSAQVK
      120     130     140     150     160     170
AROAS_SHIDY  LTGEPRMKERPIGHLVDALRLGRAKITYLEQENYPPLRLQGQFTGGNVVDGVS--QFL

      180     190     200     210     220     230
nk603cp4.pep SAVLLAGLNTPT--GITTVEIPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEG-R
      180     190     200     210     220
AROAS_SHIDY  TALLMTAPLAPEDTVIRIKGDLVSKPYIDITLNLMTKTFG----VEIENQHYQQFVVKGGQ

      240     250     260     270     280     289
nk603cp4.pep GKLTGQVIDVPDPSSTAFPLVAALLVFGSDVTILNVLNPNTRTGLILT--LQEMGADIE
      230     240     250     260     270     280
AROAS_SHIDY  SYQSPGTYLVEGDASSASYFLAAAA-IKGGTVKVTGIGRNSMQGDIRFADVLEKMGATT-

      290     300     310     320     330     340     349
nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAFAGATVMNGLEE
      290     300     310     320     330
AROAS_SHIDY  -----CWGDDY--ISCTRGELNAIDMDMNHIP---DAAMTIATAALFARKGTTTLRNIYN

      350     360     370     380     390     400     409
nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS
      340     350     360     370     380     390
AROAS_SHIDY  WRVKETDRLFAMATELRKVGAEVEEGHD--YIRITPPEK----LNFAEITATYNDHRMAMC

      410     420     430     440     450
nk603cp4.pep FLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
      400     410     420
AROAS_SHIDY  FSLVAL-SDTPVTILDPKCTAKTFPDYFEQLARISQAA

```

nk603cp4.pep  
TREMBL\_NEW: AAL19912

```

ID   AAL19912     PRELIMINARY;       PRT;   427 AA.
AC   AAL19912;
DT   01-NOV-2001 (EMBLrel. 63, Created)
DT   01-NOV-2001 (EMBLrel. 63, Last sequence update)
DT   01-NOV-2001 (EMBLrel. 63, Last annotation update)
DE   3-ENOLPYRUVYLSHIKIMATE-5-PHOSPHATE SYNTHETASE (EC 2.5.1.19). . . .

```

SCORES Init1: 148 Initn: 382 Opt: 378 z-score: 413.4 E(): 4.8e-15  
Smith-Waterman score: 477; 28.8% identity in 437 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
      70      80      90     100     110
AAL19912     MESLTLQPIARVDGAINLPGSKSVSNRALLLAALACGKTALTNLLDSDDVHRHML
      10      20      30      40      50

```

```

      70      80      90     100     110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDF--GNAATGCRLTMGLVGVDYFDST
      70      80      90     100     110
AAL19912     NALSALGINVYTLADTRCEDITGNGGALRAPGALELFLGNAGTAMRPLAALCLGQNEIV
      60      70      80      90     100     110

```

```

      120     130     140     150     160     170
nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTITYRVPMSAQVK
      120     130     140     150     160     170
AAL19912     LTGEPRMKERPIGHLVDSLQGGANIDYLEQENYPPLRLQGQFTGGDIEVDGVS--QFL
      120     130     140     150     160     170

```

```

      180     190     200     210     220     230
nk603cp4.pep SAVLLAGLNTPT--GITTVEIPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEG-R

```

240            250            260            270            280            290

nk603cp4.pep      MGADIEVINPRLAGGEDVADLRVRSSLTGKGVTVPEDRAPFSMIDEYFIPILAVAAAFEGATV  
                     |||||            : : |            : : : : :            : : : : :            : : : : :  
AE006122\_8        MGAHI-----TWGDDF--IQVEKGNLKGIDMMNHIP---DAAMTIATTALFAEGSTV  
                                300            310            320            330            340

```

          350      360      370      380      390      400
nk603cp4.pep MNGLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRRGPRDGKGLGNASGAAVATHLD
               :  ::  : ||| : ||| : : : : : : : : : : : : : : : : : : :
AB006122_8    IRNIYNWRVKETDRLTAMATELRKVGAEVEEGED--FIRIQP--LNLAQFQHAIEITYND
               350      360      370      380      390

```

nk603cp4.pep      410      420      430      440      450  
HRIAMSFVLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
|:|:| |::| |:|:| |:|:| |:|:| |:|:|  
AE006122\_8    HRMAMCFALIAL-SQTSVTILDPSTAKTFTPTFFDTFLRLTHAES  
400      410      420      430      440

nk603cp4.pep  
SWISSPROT:AROA PASMU

```

ID      AROA_PASMU      STANDARD;      PRT;      440 AA.
AC      Q04570;
DT      01-OCT-1993 (Rel. 27, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE

```

SCORES Init1: 124 Initn: 283 Opt: 373 z-score: 407.8 E(): 9.8e-15  
Smith-Waterman score: 480; 28.9% identity in 440 aa overlap

```

              10      20      30      40      50      60
nk603cp4.pep  MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSMFGLGASGETRITGLLEGEDVINTG
              | : | | : | | : | : | : | : | : | : | : | : | : | : | : | : |
ARO_A_PASMU   MIKDATAITLNPISYIEGEVRLPGSKSLSNRALLSALAKGKTLTNLLSDSDVHRHML
              10      20      30      40      50      60

```

```

              70          80          90         100         110
nk603cp4.pep KAMQAMGARIRKEGDTWI--IDGVNGGLLAPEAPLDFGNAATGCR-LTMGL-VGVYDFD
               ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
ARO_A_PASMU  NALKELGVTVQLSEDKSVCEIEGLGRAFEWQSGLALFLGNAGTAMRPLTAALCLSTPNKE
              60       70       80       90      100      110

```

```

              120      130      140      150      160      170
nk603cp4.pep S-----TFIGDASLTKRPMGRVLNPLREMGVQVKSE-EDGRLPVTLRGPKTPTPTIYRVPM
               :   ::|::|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
AROAA_PASMU   GKNEIVLTGEPRMKERPIQHLVDALCQAAGIEQYLEBQEGYPFTAIRNTGLKGGRIQIDGS
              120      130      140      150      160      170

```

```

               180         190         200         210         220
nk603cp4.pep ASAQVKSALLAGLNTPGIIT--VIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTI
              :|::||:||::||::||::||::||::||::||::||::||::||::||::||::||::||
AROA_PASMU    VSSQLFTALLMAAPMAEADTIEIIIGELVSKPYIDITLKMMQTFFG----VEVENAQYQRF
             180        190        200        210        220        230

```

```

      230      240      250      260      270      280
nk603cp4.pep  RL EGRGKL-TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILT--LQE
               ::||::: : : : || ||::: ||: : : : || : : : : || : : : :
AROA_PASMU    LVKGHQQYQSPHRFLVEGDASSYFLAAAA-IGKK-VKVTGVGKNSIQGDRLFADVLEK
               240      250      260      270      280      290

```

```

          290      300      310      320      330      340
nk603cp4.pep MGADIEVINPRLAGGEDVADLRVRSSSTLKGVTVPEDRAPSMIDEYPI LAVA AAFABEGATV
              ||| | : |: : ::::: ||: : : | : : | : : |||||
AROA_PASMU   MGAHI-----TWGDDF--IQVEKGNLKGIDMDMHIP---DAAMTIATTALFAEGETV
                300      310      320      330      340

```

NK  
D

nk603cp4.pep MNGLEELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLD  
 AROA\_PASMU IRNIYNWRVKETDRLTAMATELRKVGAEEVEEGD--PIRIQP--LNLAQFQHAEIETYN  
 350 360 370 380 390 400

nk603cp4.pep HRIAMSLVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROA\_PASMU HRMAMCFALIAL-SQTSVTILDPSCTAKTPTFTFDTFLRLTHAES  
 410 420 430 440 450

nk603cp4.pep  
 SWISSPROT:AROA\_SALTY

ID AROA\_SALTY STANDARD; PRT; 427 AA.  
 AC P07637;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 01-APR-1988 (Rel. 07, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 142 Initn: 376 Opt: 372 z-score: 406.9 E(): 1.1e-14  
 Smith-Waterman score: 471; 28.6% identity in 437 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_SALTY MESLTLQPIARVDGAINLPGSKSVSNRALLAALPCGKTALTNLDSDDVRHML  
 10 20 30 40 50 60

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLL-APEA-PLDFGNAATGCRLTMGLVGVYDFDST  
 AROA\_SALTY NALSALGINITYLSADTRCDITGNGGALRAPGALFLGNAGTAMRPLAALCLGQNEIV  
 70 80 90 100 110

nk603cp4.pep FIGDASLTAKRMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTITYRVPMASQVK  
 AROA\_SALTY LTGEPRMKERPIGHVDSLRQGGANIDYLEQENYPLRLRGFTGGDIEVDGSVSS-QFL  
 120 130 140 150 160 170

nk603cp4.pep SAVLLAGLNTF--GITTVEIPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRG  
 AROA\_SALTY TALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLMTFTG---VEIANHHYQQFVVKGGQ  
 180 190 200 210 220 230

nk603cp4.pep KL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILT--LQEMGADI  
 AROA\_SALTY QYHSPGRYL-VEGDASSASY-FLAAGAIKGGTVKVTGIGRKSMDGDIRFADVLEKMGATI  
 240 250 260 270 280

nk603cp4.pep EVINPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGL  
 AROA\_SALTY -----TWGDDFIACR---GELHAIDMDMNHIP---DAAMTIATTALFAKGTTLRLNI  
 290 300 310 320 330

nk603cp4.pep EELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIA  
 AROA\_SALTY YNWRVKETDRLFAMATELRKVGAEEVEEGHD--YIRITPPAK-LQHAD---IGTYNDHRMA  
 350 360 370 380 390 400

nk603cp4.pep MSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA

AROA\_SALTY MCFSLVAL-SDTPVTILDPKCTAKTTPDYFEQLARMSTPA  
 390 400 410 420

nk603cp4.pep  
 GP\_BCT3:M10947\_1

LOCUS M10947\_1 [STYAROAEM]  
 DEFINITION S.typhimurium aroA locus 5-enolpyruvylshikimate-3-phosphate  
 synthase (aroA protein) gene, complete cds;  
 5-enolpyruvylshikimate-3-phosphate synthase, EC  
 2.5.1.19.  
 DATE 26-APR-1993 . . .

SCORES Init1: 142 Initn: 376 Opt: 372 z-score: 406.9 E(): 1.1e-14  
 Smith-Waterman score: 471; 28.6% identity in 437 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 M10947\_1 MESLTLQPIARVDGAINLPGSKSVSNRALLAALPCGKTALTNLDSDDVRHML  
 10 20 30 40 50 60

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLL-APEA-PLDFGNAATGCRLTMGLVGVYDFDST  
 M10947\_1 NALSALGINITYLSADTRCDITGNGGALRAPGALFLGNAGTAMRPLAALCLGQNEIV  
 70 80 90 100 110

nk603cp4.pep FIGDASLTAKRMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTITYRVPMASQVK  
 M10947\_1 LTGEPRMKERPIGHVDSLRQGGANIDYLEQENYPLRLRGFTGGDIEVDGSVSS-QFL  
 120 130 140 150 160 170

nk603cp4.pep SAVLLAGLNTF--GITTVEIPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRG  
 M10947\_1 TALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLMTFTG---VEIANHHYQQFVVKGGQ  
 180 190 200 210 220 230

nk603cp4.pep KL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILT--LQEMGADI  
 M10947\_1 QYHSPGRYL-VEGDASSASY-FLAAGAIKGGTVKVTGIGRKSMDGDIRFADVLEKMGATI  
 240 250 260 270 280

nk603cp4.pep EVINPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGL  
 M10947\_1 -----TWGDDFIACR---GELHAIDMDMNHIP---DAAMTIATTALFAKGTTLRLNI  
 290 300 310 320 330

nk603cp4.pep EELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIA  
 M10947\_1 YNWRVKETDRLFAMATELRKVGAEEVEEGHD--YIRITPPAK-LQHAD---IGTYNDHRMA  
 350 360 370 380 390 400

nk603cp4.pep MSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 M10947\_1 MCFSLVAL-SDTPVTILDPKCTAKTTPDYFEQLARMSTPA  
 390 400 410 420

nk603cp4.pep  
 GP\_BCT3:Y10355\_2

LOCUS Y10355\_2 [STSERCARO]  
 DEFINITION S.typhimurium serC and aroA genes.  
 DATE 13-JAN-1997  
 ACCESSION Y10355  
 NID  
 ORGANISM Salmonella typhimurium . . .

SCORES Init1: 142 Initn: 376 Opt: 372 z-score: 406.9 E(): 1.1e-14  
 Smith-Waterman score: 471; 28.6% identity in 437 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 Y10355\_2 MESLTLQPIARVDGAINLPGSKSVSNRALLAALPCGKTALTNNLSDSDVRHML

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLL-APEA-PLDFGNAATGCRLTMGLVGVYDFDST  
 Y10355\_2 NALSALGINITYLSADTRCDITGNGGALRAPGALEFLGNAGTAMRPLAALCLGQNEIV

nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPIYRVPMASQVK  
 Y10355\_2 LTGEPRMKERPIGHLVDSLQGGANIDYLEQENYPLRLRGFTGGDIEVDGVSVS-QFL

nk603cp4.pep SAVLLAGLNTPT--GITTVEIPIMTR---DHTEKMLQGFANPTVETDADGVRTIRLEGRG  
 Y10355\_2 TALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLMTFTG---VEIANHHYQQFVVKGGQ

nk603cp4.pep KL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGADI  
 Y10355\_2 QYHSPGRYL-VEGDASSASY-FLAAGAIGGTVKVTGIGRKSMDQDIRFADVLEKMGATI

nk603cp4.pep EVINPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGL  
 Y10355\_2 -----TWGDDFIACR---GELHAIDMDMNHIP---DAAMTIATTALFAKGTTLRLNI

nk603cp4.pep EELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIA  
 Y10355\_2 YNWRVKETDRLFAMATELRKVGAEEVEEGHD--YIRITPPAK-LQHAD---IGTYNDHRMA

nk603cp4.pep MSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLAKIELSDTKAA  
 Y10355\_2 MCFSLVAL-SDTPVTILDPKCTAKTFFPDYFEQLARMSTPA

nk603cp4.pep  
 SWISSPROT:AROASALGL

ID AROASALGL STANDARD; PRT; 427 AA.  
 AC P22299;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 147 Initn: 348 Opt: 370 z-score: 404.7 E(): 1.4e-14  
 Smith-Waterman score: 469; 28.6% identity in 437 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROASALGL MESLTLQPIARVDGAINLPGSKSVSNRALLAALACGKTVLTNNLSDSDVRHML

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLL-APEA-PLDFGNAATGCRLTMGLVGVYDFDST  
 AROASALGL NALSALGINITYLSADTRCDITGNGGALRAPGALEFLGNAGTAMRPLAALCLGQNEIV

nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPIYRVPMASQVK  
 AROASALGL LTGEPRMKERPIGHLVDSLQGGANIDYLEQENYPLRLRGFTGGDIEVDGVSVS-QFL

nk603cp4.pep SAVLLAGLNTPT--GITTVEIPIMTR---DHTEKMLQGFANPTVETDADGVRTIRLEGRG  
 AROASALGL TALLMTAPLAPKDTIIRVKGELVSKPYIDITLNLMTFTG---VEIANHHYQQFVVKGGQ

nk603cp4.pep KL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGADI  
 AROASALGL QYHSPGRYL-VEGDASSASY-FLAAGAIGGTVKVTGIGRKSMDQDIRFADVLEKMGATI

nk603cp4.pep EVINPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNGL  
 AROASALGL -----TWGDDFIACR---GELHAIDMDMNHIP---DAAMTIATTALFAKGTTLRLNI

nk603cp4.pep EELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIA  
 AROASALGL YNWRVKETDRLFAMATELRKVGAEEVEEGHD--YIRITPPAK-LQHAD---IGTYNDHRMA

nk603cp4.pep MSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLAKIELSDTKAA  
 AROASALGL MCFSLVAL-SDTPVTILDPKCTAKTFFPDYFEQLARMSTPA

nk603cp4.pep  
 GP\_BCT3:M62801\_2

LOCUS M62801\_2 [STYSERARO]  
 DEFINITION Salmonella gallinarum 3-phosphoserine aminotransferase (serC) and  
 5-enolpyruvylshikimate 3-phosphate synthase (aroA) genes, complete  
 cds.  
 DATE 26-APR-1993  
 ACCESSION M62801 . . .

SCORES Init1: 147 Initn: 348 Opt: 370 z-score: 404.7 E(): 1.4e-14  
 Smith-Waterman score: 469; 28.6% identity in 437 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 M62801\_2 MESLTLQPIARVDGAINLPGSKSVSNRALLAALACGKTVLTNNLSDSDVRHML



```

                240      250      260      270      280
nk603cp4.pep  KL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVITILNVLNPNTRTGLILT--LQEMGADI
                :  |  :  |  |  |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
CAD05378      QYHSPGRYL-VEGDASSASY-FLAAGAIGKGTVKVTGIGRKSMQGDIRFADVLEKMGATI
                230      240      250      260      270      280

```

```

          290      300      310      320      330      340
nk603cp4.pep EVINPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGL
              :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
AROA_SALT1    -----TWGDDFIACTR--GELHAIDMDMNHIP---DAAMTIATALTFAKGTITLRLNI
              290      300      310      320      330      340

```

350                      360                      370                      380                      390                      400



nk603cp4.pep EELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIA  
 AROA\_SALTI YNWRVKETDRLFAMATELRKVGAEEVEGHD--YIRITPPAK-LQHAD---IGTYNDRHMA  
 340 350 360 370 380

nk603cp4.pep MSFLVMGLVSENPTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA  
 AROA\_SALTI MCFSLVAL-SDTFVTILDPKCTAKTFPDYFEQLARMSTPA  
 390 400 410 420

nk603cp4.pep  
 GP\_BCT3:X54545\_1

LOCUS X54545\_1 [ST5E3PS]  
 DEFINITION S.typhi gene for 5-enolpyruvylshikimate 3-phosphate synthase (EC  
 2.5.1.19);  
 5-enolpyruvylshikimate 3-phosphate synthase.  
 DATE 12-SEP-1993  
 ACCESSION X54545 . . .

SCORES Init1: 147 Initn: 348 Opt: 361 z-score: 395.0 E(): 5e-14  
 Smith-Waterman score: 460; 27.9% identity in 437 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 X54545\_1 MESLTLQPIARVDGAINLPKSKSVSNRALLLAALACGKTVLTNLLSDDDVRHML  
 10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDF--GNAATGCRLTGMGLGVGYDFDST  
 X54545\_1 NALSALGINYTLASDRTRCDITGNGGFLRASGTLEFLGNAGTAMRPLAALCLGQNEIV  
 60 70 80 90 100 110

nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTTYRVPMASAOVK  
 X54545\_1 LTGEPRMKERPIGHLVDSLROGGANIDYLEQENYPLRLRGFGIGGDIIEVDGSSVS-QFL  
 120 130 140 150 160 170

nk603cp4.pep SAVLLAGLNTPE--GITTVEIPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRG  
 X54545\_1 TALLMTAPLAPEDTIIIRVKGELVSKPYIDITLNLMTFTG---VEIANHHYQQFVVKGGQ  
 180 190 200 210 220

nk603cp4.pep KL--TGQVIDVPGDPSSTAFLVAALLVPGSDVTILNVLNPTRTGLILT--LQEMGADI  
 X54545\_1 QYHSPGRYL-VEGDASSASY-FLAAGGKGGTVKVTGIGGKSMQGDIFADVLHKMGATI  
 230 240 250 260 270 280

nk603cp4.pep EVINPRLAGGED-VADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGL  
 X54545\_1 -----TWGDDFIACTR--GELHAIDMDMNHIP---DAAMTIATPALFAKGTITLRNI  
 290 300 310 320 330

nk603cp4.pep EELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIA  
 X54545\_1 YNWRVKETDRLFAMATELRKVGAEEVEGHD--YIRITPPAK-LQHAD---IGTYNDRHMA  
 340 350 360 370 380

nk603cp4.pep MSFLVMGLVSENPTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA  
 410 420 430 440 450

X54545\_1 MCFSLVAL-SDTFVTILDPKCTAKTFPDYFEQLARMSTPA  
 390 400 410 420

nk603cp4.pep  
 GP\_BCT3:Z14100\_1

LOCUS Z14100\_1 [PMAROA]  
 DEFINITION P.multocida aroA gene.  
 DATE 30-JUN-1993  
 ACCESSION Z14100  
 NID  
 ORGANISM Pasteurella multocida . . .

SCORES Init1: 124 Initn: 283 Opt: 358 z-score: 391.6 E(): 7.8e-14  
 Smith-Waterman score: 467; 28.8% identity in 441 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 Z14100\_1 MIKDATAITLNPISYIEGEVRLPGSKSLNRRALLSALAKGKTTLTNLLSDDDVRHML  
 10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEGDTWI--IDGVNGGGLLAPEAPLDFGNAATGCR-LTMGL-VGVYDFD  
 Z14100\_1 NALKEGVTYQLSEDKSVCEIEGLGRAFEWQSGLALFLGNAGTAMRPLTAALCLSTPNRE  
 60 70 80 90 100 110

nk603cp4.pep S----TFIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTTYRVPM  
 Z14100\_1 GKNEIVLTGEPKMRPIQHLVDALCQAGAEIQYLEQEGYPPFIATRNGLKGGRIQIDGS  
 120 130 140 150 160 170

nk603cp4.pep ASAQVKSAVLLAGLNTPGITT--VIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTI  
 Z14100\_1 VSSQFLTALLMAAPMAEADTEIEIIGELVSKPYIDITLKMMTFTG---VEVENQAYQRF  
 180 190 200 210 220 230

nk603cp4.pep RLEGRGKL-TGQVIDVPGDPSSTAFLVAALLVPGSDVTILNVLNPTRTGLILT--LQE  
 Z14100\_1 LVKGHQQQSPHRFLVEGDASSASYFLAAAA-IKKG-VKVTGVGKNSIQGDRLPADVLEK  
 240 250 260 270 280 290

nk603cp4.pep MGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATV  
 Z14100\_1 MGAHI-----TWGDDF--IQVEKGNLKGIDMDMNHIP---DAAMTIATPALFAEGETV  
 300 310 320 330 340

nk603cp4.pep MNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLD  
 Z14100\_1 IRNIYNWRVKETDRLTAMATELRKVGAEEVEGHD--FIRIQP--LNLAQFQHAELNIH-D  
 350 360 370 380 390

nk603cp4.pep HRIAMSFLVMGLVSENPTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA  
 Z14100\_1 HRMAMCFALIAL-SKTSVTILDPSCTAKTFPTFLILFTLNTREVAYR  
 400 410 420 430 440

nk603cp4.pep  
 TREMBL\_MAIN:Q9K4A7

ID Q9K4A7 PRELIMINARY; PRT; 438 AA.

AC Q9K4A7;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . .

SCORES Init1: 104 Initn: 259 Opt: 357 z-score: 390.5 E(): 8.9e-14  
 Smith-Waterman score: 436; 28.3% identity in 456 aa overlap

```

nk603cp4.pep  MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVI
Q9K4A7          MTNVNPTHALWPAP-HASGAVDATVHVPGSKSVTNRALVLAALASEPGWLRRPLRSRDTL
               10      20      30      40      50
nk603cp4.pep  NTGKAMQAMGARIRK-----EGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVG
Q9K4A7          LMAEALRTLGVIEEGVGPEGTGEFW---RVIPAGLRGP-ATVDVGNAGTVMRFLPPVAT
               60      70      80      90     100     110
nk603cp4.pep  VYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGPKT--PTPITYRV
Q9K4A7          LADGAVRFDGDPRSYERPLHGVIDALRVLGARIDDDGRGALPLTVHGGGALEGGPVEIDA
               120     130     140     150     160     170
nk603cp4.pep  PMASAQVKSALLAGLN-TPGI---TTVIEPIMTR-DHTEKMLQGFGANPTVET-DADG
Q9K4A7          S-SSSQFVSALLSGPRFNQGVVHRHTGSALPSMPHIRMVDMRLRAVGAQ--VDTPESSG
               180     190     200     210     220     230
nk603cp4.pep  VRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQ
Q9K4A7          EPNVWRVTPGALLGRDLTVPEPLSN-AQPFLLAALVTGGKVVIPDW---PSRT-----T
               240     250     260     270     280
nk603cp4.pep  EMGADIEVINPRLAGGEDVADLRVR---SSTLKGVTVPEDRAPSMIDEYPIAVAAFAE
Q9K4A7          QPGDRLREIFTDMGGSCELTDFGLVFTGSGAIHGIDVDLSEVGELT---PGIAAVALAD
               290     300     310     320     330
nk603cp4.pep  GATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVA
Q9K4A7          SPSTLRGVAHLRLHETDRLAALTKEINELGGDVTETADGLHIRP-----RLHGGVVFH
               340     350     360     370     380     390
nk603cp4.pep  THLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAG-LGAKIELSDTKAA
Q9K4A7          TYDDHRMATAGAVLGLAVEG-VQIENVATTAKTLPDFPDLTGMLGA
               400     410     420     430

```

nk603cp4.pep  
 GP\_BCT3:AL359214\_9

LOCUS AL359214\_9 [SC7E4]  
 DEFINITION Streptomyces coelicolor cosmid 7E4;  
 SC7E4.09c, aroA2, 3-phosphoshikimate  
 1-carboxyvinyltransferase, len: 438 aa; highly similar to  
 SW:ARO\_A\_MYCTU (EMBL:X52269) Mycobacterium tuberculosis  
 3-phosphoshikimate 1-carboxyvinyltransferase (EC 2.5.1.19) . . .

SCORES Init1: 104 Initn: 259 Opt: 357 z-score: 390.5 E(): 8.9e-14

Smith-Waterman score: 436; 28.3% identity in 456 aa overlap

```

nk603cp4.pep  MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVI
AL359214_9     MTNVNPTHALWPAP-HASGAVDATVHVPGSKSVTNRALVLAALASEPGWLRRPLRSRDTL
               10      20      30      40      50
nk603cp4.pep  NTGKAMQAMGARIRK-----EGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVG
AL359214_9     LMAEALRTLGVIEEGVGPEGTGEFW---RVIPAGLRGP-ATVDVGNAGTVMRFLPPVAT
               60      70      80      90     100     110
nk603cp4.pep  VYDFDSTFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGPKT--PTPITYRV
AL359214_9     LADGAVRFDGDPRSYERPLHGVIDALRVLGARIDDDGRGALPLTVHGGGALEGGPVEIDA
               120     130     140     150     160     170
nk603cp4.pep  PMASAQVKSALLAGLN-TPGI---TTVIEPIMTR-DHTEKMLQGFGANPTVET-DADG
AL359214_9     S-SSSQFVSALLSGPRFNQGVVHRHTGSALPSMPHIRMVDMRLRAVGAQ--VDTPESSG
               180     190     200     210     220     230
nk603cp4.pep  VRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQ
AL359214_9     EPNVWRVTPGALLGRDLTVPEPLSN-AQPFLLAALVTGGKVVIPDW---PSRT-----T
               240     250     260     270     280
nk603cp4.pep  EMGADIEVINPRLAGGEDVADLRVR---SSTLKGVTVPEDRAPSMIDEYPIAVAAFAE
AL359214_9     QPGDRLREIFTDMGGSCELTDFGLVFTGSGAIHGIDVDLSEVGELT---PGIAAVALAD
               290     300     310     320     330
nk603cp4.pep  GATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVA
AL359214_9     SPSTLRGVAHLRLHETDRLAALTKEINELGGDVTETADGLHIRP-----RLHGGVVFH
               340     350     360     370     380     390
nk603cp4.pep  THLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAG-LGAKIELSDTKAA
AL359214_9     TYDDHRMATAGAVLGLAVEG-VQIENVATTAKTLPDFPDLTGMLGA
               400     410     420     430

```

nk603cp4.pep  
 TREMBL\_MAIN:Q9L213

ID Q9L213 PRELIMINARY; PRT; 440 AA.  
 AC Q9L213;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . .

SCORES Init1: 144 Initn: 327 Opt: 349 z-score: 381.9 E(): 2.7e-13  
 Smith-Waterman score: 540; 28.7% identity in 446 aa overlap

```

nk603cp4.pep  MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDV
Q9L213          MSATDPRLKAETLTIRTLTGFDVTVRVLGSKSYTNRYLAIASLGGQETVIDNALLSDDT
               10      20      30      40      50      60

```

```

      60      70      80      90     100     110
nk603cp4.pep INTGKAMQAMGARIRKEGD--TWIIDGVGNG-GLLAPEAPLDFGNAATGCRLTMGLVG VY
Q9L213      VYFSRAIETFG-HVTCDDIHATARIRVTPTRGFRAPSEDI FVGAGTPLRPLISMAGHA
      70      80      90     100     110

      120     130     140     150     160     170
nk603cp4.pep DFDSTFIGDASLTKRPMGRVLNPLREMGVQVKS EGD-RLPVTLRGPKTPTPTIYRVPMA
Q9L213      DGTITITGNARMQERPMGDL LKALPALGVDATAVRNGSGPPVRVVGSGFKGGATSISGAV
      120     130     140     150     160     170

      180     190     200     210     220     230
nk603cp4.pep SAQVKS AVLLAGLNTPGIT--TVIEPIMTRDHT EKMLQGFGANPTVETDADGVRTIRLEG
Q9L213      SSQFTSSLIINALRAQTDTEITISDDL VSKPYVEMTLAGL-AEMGVSVD RDGYRRFTVPS
      180     190     200     210     220     230

      240     250     260     270     280
nk603cp4.pep RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVL MNPTR--TGLILTLQEMGADI
Q9L213      GQQARGGQVTVPEPDASGMSYFLAAAILQ-SRVVIPGIGAGSHQGDVHLVQALERMGCRT
      240     250     260     270     280     290

      290     300     310     320     330     340
nk603cp4.pep EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMID EYPI LAVAAAFAGATVMNGLE
Q9L213      EV-----GDD--SITVTGGPLRGIDIDMEAMPDVV---PSLAIVAAYAE GTTRITNIA
      300     310     320     330     340

      350     360     370     380     390     400
nk603cp4.pep ELRVKESDRLSAVANGKLNGVDCDEGETSLVVRG RPDGKGLGNASGAAVATHLDHRIAM
Q9L213      SLRVKECDRIA AVTTELKMGIDVEEHS DAMYITG-----GTPHGAVIDT YDDHRIAM
      350     360     370     380     390

      410     420     430     440     450
nk603cp4.pep SFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGA KIELSDTKAA
Q9L213      TFAIGGLRTEG-VVIKDPGCVAKSFPAFWQTLDTLHPDLESTK
      400     410     420     430     440

```

nk603cp4.pep  
GP\_BCT3:AL138598\_28

LOCUS AL138598\_28 [SC1A2]  
DEFINITION Streptomyces coelicolor cosmid 1A2;  
SC1A2.28c, aroA, 3-phosphoshikimate  
1-carboxyvinyltransferase (EC 2.5.1.19), len: 440 aa.  
Highly similar to many e.g. Escherichia coli  
SW:ARO A\_ECOLI(EMBL:X00557) 3-phosphoshikimate . . .

SCORES Init1: 144 Initn: 327 Opt: 349 z-score: 381.9 E(): 2.7e-13  
Smith-Waterman score: 540; 28.7% identity in 446 aa overlap

```

      10      20      30      40      50
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDV
AL138598_28 MSATPDPR LKAETLTIRTLTGFDET VRVLGSKSYTNRYLAIASLSGQETVIDNALLSDDT
      10      20      30      40      50      60

      60      70      80      90     100     110
nk603cp4.pep INTGKAMQAMGARIRKEGD--TWIIDGVGNG-GLLAPEAPLDFGNAATGCRLTMGLVG VY
AL138598_28 VYFSRAIETFG-HVTCDDIHATARIRVTPTRGFRAPSEDI FVGAGTPLRPLISMAGHA
      70      80      90     100     110

      120     130     140     150     160     170
nk603cp4.pep DFDSTFIGDASLTKRPMGRVLNPLREMGVQVKS EGD-RLPVTLRGPKTPTPTIYRVPMA

```

```

      120     130     140     150     160     170
AL138598_28 DGTITITGNARMQERPMGDL LKALPALGVDATAVRNGSGPPVRVVGSGFKGGATSISGAV
      120     130     140     150     160     170

      180     190     200     210     220     230
nk603cp4.pep SAQVKS AVLLAGLNTPGIT--TVIEPIMTRDHT EKMLQGFGANPTVETDADGVRTIRLEG
AL138598_28 SSQFTSSLIINALRAQTDTEITISDDL VSKPYVEMTLAGL-AEMGVSVD RDGYRRFTVPS
      180     190     200     210     220     230

      240     250     260     270     280
nk603cp4.pep RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVL MNPTR--TGLILTLQEMGADI
AL138598_28 GQQARGGQVTVPEPDASGMSYFLAAAILQ-SRVVIPGIGAGSHQGDVHLVQALERMGCRT
      240     250     260     270     280     290

      290     300     310     320     330     340
nk603cp4.pep EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMID EYPI LAVAAAFAGATVMNGLE
AL138598_28 EV-----GDD--SITVTGGPLRGIDIDMEAMPDVV---PSLAIVAAYAE GTTRITNIA
      300     310     320     330     340

      350     360     370     380     390     400
nk603cp4.pep ELRVKESDRLSAVANGKLNGVDCDEGETSLVVRG RPDGKGLGNASGAAVATHLDHRIAM
AL138598_28 SLRVKECDRIA AVTTELKMGIDVEEHS DAMYITG-----GTPHGAVIDT YDDHRIAM
      350     360     370     380     390

      410     420     430     440     450
nk603cp4.pep SFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGA KIELSDTKAA
AL138598_28 TFAIGGLRTEG-VVIKDPGCVAKSFPAFWQTLDTLHPDLESTK
      400     410     420     430     440

```

nk603cp4.pep  
TREMBL\_NEW:AAK86449

ID AAK86449 PRELIMINARY; PRT; 425 AA.  
AC AAK86449;  
DT 06-SEP-2001 (EMBLrel. 63, Created)  
DT 06-SEP-2001 (EMBLrel. 63, Last sequence update)  
DT 06-SEP-2001 (EMBLrel. 63, Last annotation update)  
DE AGR\_C\_1140P. . . .

SCORES Init1: 156 Initn: 380 Opt: 329 z-score: 360.5 E(): 4.2e-12  
Smith-Waterman score: 483; 27.7% identity in 448 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AAK86449      MIELTITPPGHPLSGKVEPPGSKSTINRALLLAGLAKGKSHLSGALKSDDTLYMA
      10      20      30      40      50

      60      70      80      90     100     110     119
nk603cp4.pep KAMQAMGARIRK-EGD TWIIDGVGNGGLLAPEAPLDFGNAATGCRLTMGLVG VYDFDSTF
AAK86449      EALREMGKVKVPEDATT FVVEGTGV--LQOPEKPLFLGNAGTATRFLTAAGALVDGAVII
      60      70      80      90     100     110

      120     130     140     150     160     170
nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKS EGDRLPVTLRGPKTPTP---ITYRVPMA SAQV
AAK86449      DGDEHMRKRPIPLVQALRALGVEADAPT GCP-PVTVRGKGMGFPKGSVTIDANLSSQYV
      120     130     140     150     160     170

      180     190     200     210     220     230
nk603cp4.pep KSAVLLAGLNTPGITTVI--EPIMTR---DHTEKMLQGF GANPTVETDADGVRTIRLEGR
AAK86449      SALLMAAACGDKPVDI ILKGEHIGAKGYIDLTTSAMEAFGAK--VERVSNAIWRVHTGY

```

180 190 200 210 220 230

nk603cp4.pep GKLTGQVIDVPDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVI  
 AAK86449 ---TATDFHIEPDASAATY-LWGAELLTGGAIDI---GTP-----AD-KFT

240 250 260 270 280 290

nk603cp4.pep NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFAGATVMNGLEELR  
 AAK86449 QPDAKAYEVMAQFPHLPAEIDG-----SQMQDAIPTIAVIAAFNPTVRFVGIANLR

300 310 320 330 340 350

nk603cp4.pep VKESDRLSAVANGKLK--NGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
 AAK86449 VKECDRIKAVSLGLNEIREGLAHEEGD-DLIVHADPSLAG--QTVASIDTFADHRIAMS

360 370 380 390 400 409

nk603cp4.pep FLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AAK86449 FALAAL-KIGGIAIQNPACVAKTYPGWKALASLGVDYTEKESAAEPQH

410 420 430 440 450

nk603cp4.pep  
 SWISSPROT:ARO\_ARCFU

ID ARO\_ARCFU STANDARD; PRT; 416 AA.  
 AC O28775;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE PROBABLE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) . . .

SCORES Init1: 103 Initn: 283 Opt: 323 z-score: 354.2 E(): 9.5e-12  
 Smith-Waterman score: 548; 28.9% identity in 432 aa overlap

10 20 30 40 50 60

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 ARO\_ARCFU MDVIVRKGEIRGKAKPPASKSYTHRAFIASL-SPSARVVNPLISEDITISTL

70 80 90 100 110 120

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
 ARO\_ARCFU NACKKRIGAAVLKKGNEWLFSGVD--GVEA-EGYFNFANSQTTLRIPTGLLSLSPFRSVVD

130 140 150 160 170 180

nk603cp4.pep GDASLTKRPMGRVNLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTIYRVPMASQVKS AV  
 ARO\_ARCFU GDES LRKRPNGELVLALSKLGARFKGREPYTPFPFSVQGVKGEVEIEAP--SSQFVSSL

190 200 210 220 230 239

nk603cp4.pep LLAGLNTPGITTV-IEPIMTRDHTEKMLQGFGANPTVETDADGVRTIRLEGRKLTGQVI  
 ARO\_ARCFU LFALSLAEGDSSLRVEKVKSQPYIDVTLVDLRESG-VKVEREG-NFYHIPGSQSFKLRRY

240 250 260 270 280 290

nk603cp4.pep DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTL-QEMGADIEVINPRLAGG  
 ARO\_ARCFU DVPADFSSASY-LIAAGLIAG-EVVLEGMFESAQGDRIKVIDICREMGGSVEWDKRGV--

300 310 320 330 340 350

nk603cp4.pep EDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFAGATVMNGLEELRVKESDRL  
 ARO\_ARCFU -----IRAERSELEGVEVDASDIPDLV---PTIAVLA AVAKGKTRIYNAEHLRIKIDRI

360 370 380 390 400 410

nk603cp4.pep SAVANGLKLNVDCEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFLVMGLVSE  
 ARO\_ARCFU EGIHQNLKALGVESKPLKDLIIGK-----GKGEFRGV-VDSFGDHRMALAFSLGLLGE

420 430 440 450

nk603cp4.pep NPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 ARO\_ARCFU --VKCRNAEVSVSFPGYFRVLESLGASVIRL

nk603cp4.pep  
 GP\_BCT1:AE000999\_8

LOCUS AE000999\_8 [AE000999]  
 DEFINITION Archaeoglobus fulgidus section 108 of 172 of the complete genome;  
 similar to GB:L77117 SP:Q57925 PID:1591205 percent  
 identity: 41.47; identified by sequence similarity;  
 putative.  
 DATE 15-DEC-1997 . . .

SCORES Init1: 103 Initn: 283 Opt: 323 z-score: 354.2 E(): 9.5e-12  
 Smith-Waterman score: 548; 28.9% identity in 432 aa overlap

10 20 30 40 50 60

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AE000999\_8 MDVIVRKGEIRGKAKPPASKSYTHRAFIASL-SPSARVVNPLISEDITISTL

70 80 90 100 110 120

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
 AE000999\_8 NACKKRIGAAVLKKGNEWLFSGVD--GVEA-EGYFNFANSQTTLRIPTGLLSLSPFRSVVD

130 140 150 160 170 180

nk603cp4.pep GDASLTKRPMGRVNLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTIYRVPMASQVKS AV  
 AE000999\_8 GDES LRKRPNGELVLALSKLGARFKGREPYTPFPFSVQGVKGEVEIEAP--SSQFVSSL

190 200 210 220 230 239

nk603cp4.pep LLAGLNTPGITTV-IEPIMTRDHTEKMLQGFGANPTVETDADGVRTIRLEGRKLTGQVI  
 AE000999\_8 LFALSLAEGDSSLRVEKVKSQPYIDVTLVDLRESG-VKVEREG-NFYHIPGSQSFKLRRY

240 250 260 270 280 290

nk603cp4.pep DVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTL-QEMGADIEVINPRLAGG  
 AE000999\_8 DVPADFSSASY-LIAAGLIAG-EVVLEGMFESAQGDRIKVIDICREMGGSVEWDKRGV--

300 310 320 330 340 350

nk603cp4.pep EDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFAGATVMNGLEELRVKESDRL  
 AE000999\_8 -----IRAERSELEGVEVDASDIPDLV---PTIAVLA AVAKGKTRIYNAEHLRIKIDRI

360 370 380 390 400 410

nk603cp4.pep SAVANGLKLNVDCEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFLVMGLVSE

AE000999\_8 EGIHQNLKALGVESKPLKGLIIGK-----GKGEFRGV-VDSFGDHRMALAFSLGLLGE  
340 350 360 370 380

nk603cp4.pep NPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
420 430 440 450

AE000999\_8 --VKCRNAEVVSVSFPGYFRVLESGLGASVIRL  
390 400 410

nk603cp4.pep  
TREMBL\_MAIN:Q9JTT3

ID Q9JTT3 PRELIMINARY; PRT; 433 AA.  
AC Q9JTT3;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 159 Initn: 238 Opt: 315 z-score: 345.3 E(): 3e-11  
Smith-Waterman score: 395; 25.5% identity in 455 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
10 20 30 40 50 60  
Q9JTT3 MTESLRLPAASLKPS----TVALPGSKSISNRTLLAALSDNACEIHSLLKSDDTDRML  
10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEG-DTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMLGVGVYDFDSTF  
70 80 90 100 110 119  
Q9JTT3 EALDKLGVIEYLAEDRLKVHGTG-GRFPNRTADLFLGNAGTAFRPLTAALAVLGDDYRL  
60 70 80 90 100 110

nk603cp4.pep IGDASLTKRPMGRVNLPLREMGVQVKSEGDRLPVTLRGPKTPTPTITYRVP---ASAQV  
120 130 140 150 160 170  
Q9JTT3 HGVARMHPIGDLADALRIAGADVEYLKGEHYPLHIGERQDNG-ERVIPIKGNVSSQF  
120 130 140 150 160 170

nk603cp4.pep KSAVLLAGLNTPTGIT---TVIEPIMTR---DHTEKMLQGFGANPTVETDAGVTRIRLEG  
180 190 200 210 220 230  
Q9JTT3 LTALLMA-LPLTGQAFEIRMVGEISKPYIDITLKLMAQFG---VQVINEGYRVFKIPA  
180 190 200 210 220

nk603cp4.pep RGKL-TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILT--LQEMGAD  
240 250 260 270 280  
Q9JTT3 DAHYHAPHLHVEGDASSASY-FLAAGLIAATPVVRVTGIGANSIQGDVAFACELEKIGAD  
230 240 250 260 270 280

nk603cp4.pep IEVINPRLAGGEDVADL-RVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGAT-VMN  
290 300 310 320 330 340  
Q9JTT3 V-----VWGENFVEVSRPKERAVRAFDLDANHIP---DAAMTLAIVA-LATGQTCTLR  
290 300 310 320 330

nk603cp4.pep GLEELRVKESDRLSAVANGLKLNVDCEGETSLVVRGRPDGKGLGNASGAAVATHLDHR  
350 360 370 380 390 400  
Q9JTT3 NIGSWRVKETDRIAAMANELRKLGAKEAEAEIHTP-PTTP---TPDAVIDTYDDHR  
340 350 360 370 380 390

nk603cp4.pep IAMSFLVMGLVSENFPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
410 420 430 440 450  
Q9JTT3 MAMCFSLVSL-LDVPVVINDPKCTHKTFTPTFYDFVSSLTETAE

400 410 420 430

nk603cp4.pep  
GP\_BCT3:AL162756\_195

LOCUS AL162756\_195 [NMA522491]  
DEFINITION Neisseria meningitidis serogroup A strain Z2491 complete genome;  
segment 5/7;  
NMA1644, aroA, 5-enolpyruvoylshikimate-3-phosphate  
synthase, len: 433aa; similar to many eg. SW:P07638  
(AROA\_ECOLI) aroA, 5-enolpyruvoylshikimate-3-phosphate . . .

SCORES Init1: 159 Initn: 238 Opt: 315 z-score: 345.3 E(): 3e-11  
Smith-Waterman score: 395; 25.5% identity in 455 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
10 20 30 40 50 60  
AL162756\_195 MTESLRLPAASLKPS----TVALPGSKSISNRTLLAALSDNACEIHSLLKSDDTDRML  
10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEG-DTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMLGVGVYDFDSTF  
70 80 90 100 110 119  
AL162756\_195 EALDKLGVIEYLAEDRLKVHGTG-GRFPNRTADLFLGNAGTAFRPLTAALAVLGDDYRL  
60 70 80 90 100 110

nk603cp4.pep IGDASLTKRPMGRVNLPLREMGVQVKSEGDRLPVTLRGPKTPTPTITYRVP---ASAQV  
120 130 140 150 160 170  
AL162756\_195 HGVARMHPIGDLADALRIAGADVEYLKGEHYPLHIGERQDNG-ERVIPIKGNVSSQF  
120 130 140 150 160 170

nk603cp4.pep KSAVLLAGLNTPTGIT---TVIEPIMTR---DHTEKMLQGFGANPTVETDAGVTRIRLEG  
180 190 200 210 220 230  
AL162756\_195 LTALLMA-LPLTGQAFEIRMVGEISKPYIDITLKLMAQFG---VQVINEGYRVFKIPA  
180 190 200 210 220

nk603cp4.pep RGKL-TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILT--LQEMGAD  
240 250 260 270 280  
AL162756\_195 DAHYHAPHLHVEGDASSASY-FLAAGLIAATPVVRVTGIGANSIQGDVAFACELEKIGAD  
230 240 250 260 270 280

nk603cp4.pep IEVINPRLAGGEDVADL-RVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGAT-VMN  
290 300 310 320 330 340  
AL162756\_195 V-----VWGENFVEVSRPKERAVRAFDLDANHIP---DAAMTLAIVA-LATGQTCTLR  
290 300 310 320 330

nk603cp4.pep GLEELRVKESDRLSAVANGLKLNVDCEGETSLVVRGRPDGKGLGNASGAAVATHLDHR  
350 360 370 380 390 400  
AL162756\_195 NIGSWRVKETDRIAAMANELRKLGAKEAEAEIHTP-PTTP---TPDAVIDTYDDHR  
340 350 360 370 380 390

nk603cp4.pep IAMSFLVMGLVSENFPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
410 420 430 440 450  
AL162756\_195 MAMCFSLVSL-LDVPVVINDPKCTHKTFTPTFYDFVSSLTETAE  
400 410 420 430

nk603cp4.pep  
GP\_BCT1:AE001310\_1

LOCUS AE001310\_1 [AE001310]  
DEFINITION Chlamydia trachomatis section 37 of 87 of the complete genome.

DATE 30-OCT-2000  
ACCESSION AE001310  
NID  
ORGANISM Chlamydia trachomatis . . .

SCORES Initl: 67 Initn: 202 Opt: 306 z-score: 335.4 E(): 1e-10  
Smith-Waterman score: 442; 27.4% identity in 434 aa overlap

```
nk603cp4.pep 10 20 30 40 50 60
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AE001310_1 MVSSNQDLLISPSIPYGEIAVPPSKSHSLRAILFASLSKGTIIENCFLSPDSQAML
10 20 30 40 50
```

```
nk603cp4.pep 70 80 90 100 110 120
KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRLTMLGVGVYDFDSTFI
AE001310_1 TACEKMGAVHVRIGDSLHIQGNPDPHHCHPRY-FHMGNSGIALRFLTALSTLSPTPLIT
60 70 80 90 100 110
```

```
nk603cp4.pep 130 140 150 160 170 180
GDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMSAQVKS AV
AE001310_1 GSHTLKRRIAPLLSSLKQLGAHIRQKTSSSIPFTIHGPLSPGHVT--ISGQDSQYASAL
120 130 140 150 160 170
```

```
nk603cp4.pep 190 200 210 220 230
LLAGLNTFGITTVIEPIIMTRDHEKMLQGFGA-NPTVETDADGVRTIRLEGRGKLTGQV
AE001310_1 AITAALAPYPLSFSIENLKERPFWDLTLDWLHSLNISFLRDQD---SLTFPGQSLESFS
180 190 200 210 220 230
```

```
nk603cp4.pep 240 250 260 270 280 290
IDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTG---LILTQEMGADIEVINPRL
AE001310_1 YSVPGDYSSAFLASFGLLSSSKPTILRNLSQDSQGDKLLFSLKQLGAHI-----L
240 250 260 270 280
```

```
nk603cp4.pep 300 310 320 330 340 350
AGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKES
AE001310_1 IGKHHI---EMHPSSFSGGEIDMD--P-FIDALPILAVLCCFAKNPSRLYNALGAKDKES
290 300 310 320 330
```

```
nk603cp4.pep 360 370 380 390 400 410
DRLSAVANGLKLVGDCDEGETSLVVRGPDGKGLGNASGAAVATHLDHRIAMSFLVMGL
AE001310_1 NRIEAIAHELQKMGGSVHPTRDGLYIE--PS-----RLHGAVVDSDHNDHRIAMALAVAGV
340 350 360 370 380 390
```

```
nk603cp4.pep 420 430 440 450
-VSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
```

```
AE001310_1 HASSGQTLNCNTQCINKSFYFVIAAQTLANVRHYQADFPLRSSFCR
400 410 420 430 440
```

nk603cp4.pep  
SWISSPROT:AROA\_CHLTR

ID AROA\_CHLTR STANDARD; PRT; 440 AA.

AC O84371;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Initl: 67 Initn: 202 Opt: 306 z-score: 335.4 E(): 1e-10  
Smith-Waterman score: 442; 27.4% identity in 434 aa overlap

```
nk603cp4.pep 10 20 30 40 50 60
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AROA_CHLTR MVSSNQDLLISPSIPYGEIAVPPSKSHSLRAILFASLSKGTIIENCFLSPDSQAML
10 20 30 40 50
```

```
nk603cp4.pep 70 80 90 100 110 120
KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRLTMLGVGVYDFDSTFI
AROA_CHLTR TACEKMGAVHVRIGDSLHIQGNPDPHHCHPRY-FHMGNSGIALRFLTALSTLSPTPLIT
60 70 80 90 100 110
```

```
nk603cp4.pep 130 140 150 160 170 180
GDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMSAQVKS AV
AROA_CHLTR GSHTLKRRIAPLLSSLKQLGAHIRQKTSSSIPFTIHGPLSPGHVT--ISGQDSQYASAL
120 130 140 150 160 170
```

```
nk603cp4.pep 190 200 210 220 230
LLAGLNTFGITTVIEPIIMTRDHEKMLQGFGA-NPTVETDADGVRTIRLEGRGKLTGQV
AROA_CHLTR AITAALAPYPLSFSIENLKERPFWDLTLDWLHSLNISFLRDQD---SLTFPGQSLESFS
180 190 200 210 220 230
```

```
nk603cp4.pep 240 250 260 270 280 290
IDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTG---LILTQEMGADIEVINPRL
AROA_CHLTR YSVPGDYSSAFLASFGLLSSSKPTILRNLSQDSQGDKLLFSLKQLGAHI-----L
240 250 260 270 280
```

```
nk603cp4.pep 300 310 320 330 340 350
AGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKES
AROA_CHLTR IGKHHI---EMHPSSFSGGEIDMD--P-FIDALPILAVLCCFAKNPSRLYNALGAKDKES
290 300 310 320 330
```

```
nk603cp4.pep 360 370 380 390 400 410
DRLSAVANGLKLVGDCDEGETSLVVRGPDGKGLGNASGAAVATHLDHRIAMSFLVMGL
AROA_CHLTR NRIEAIAHELQKMGGSVHPTRDGLYIE--PS-----RLHGAVVDSDHNDHRIAMALAVAGV
340 350 360 370 380 390
```

```
nk603cp4.pep 420 430 440 450
-VSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
```

```
AROA_CHLTR HASSGQTLNCNTQCINKSFYFVIAAQTLANVRHYQADFPLRSSFCR
400 410 420 430 440
```

nk603cp4.pep  
TREMBL\_MAIN:Q9JYU1

ID Q9JYU1 PRELIMINARY; PRT; 433 AA.

AC Q9JYU1;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)

DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Initl: 156 Initn: 235 Opt: 305 z-score: 334.5 E(): 1.2e-10  
Smith-Waterman score: 380; 25.1% identity in 443 aa overlap

```
nk603cp4.pep 10 20 30 40 50 60
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
Q9JYU1 MTESVRLPVARLKPSTVALPGSKSISNRTLLAALSDNACEIHSLLKSDTDRML
10 20 30 40 50
```

```
70 80 90 100 110 119
```

Monsanto Product Safety Center

Date 11/06/01

NK603 analysis

Page 56 of 120

nk603cp4.pep KAMQAMGARIRKEG-DTWIIDGVNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 Q9JYU1 EALDKLGVQIEYLAEDRLKVHGTG-GRFPNRTADLFLGNAGTAFRPLTAALAVLGGDYHL  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
 IGDASLTAKRMGRVNLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTITYRVPM---ASAQV  
 Q9JYU1 HGVPFRHERPIGDLVDALRIAGADVEYLKGEHYPLHIGERQDNG-ERVIPIKGNVSSQF  
 120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
 KSAVLLAGLNTPGIT---TVIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEG  
 Q9JYU1 LTALLMA-LPLTGQAPEIRMVGEISKPYIDITLKLMAQFG---VQVINEGYRVFKIPA  
 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280  
 RGKL-TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGAD  
 Q9JYU1 DAHYHAPEHLHVEGDASSASY-FLAAGLIAATFVRVTGIGANSIQGDVAFARELEKIGAD  
 230 240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340  
 IEVINPRLAGGEDVADL-RVRSSTLKGVTVPEDRAPSMIDEYPIILAVAAFAEGAT-VMN  
 Q9JYU1 V-----VWGENFVEVSRPKERAVQSFDLDANHIP---DAAMTLAIVA-LATGQTCTLR  
 290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400  
 GLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHR  
 Q9JYU1 NIGSWRVKETDRIAAMANELRKLGAKEVEAEAIHITP-PETL---TPDAVIDTYDDHR  
 340 350 360 370 380 390

nk603cp4.pep 410 420 430 440 450  
 IAMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 Q9JYU1 MAMCFSLVSLGV-PVVINDPKCTHKTFFPYFDFVSSLTETAE  
 400 410 420 430

nk603cp4.pep  
 GP\_BCT1:AE002492\_7

LOCUS AE002492\_7 [AE002492]  
 DEFINITION Neisseria meningitidis serogroup B strain MC58 section 134 of 206  
 of the complete genome;  
 similar to GB:M20023 SP:P12421 PID:144037 percent  
 identity: 69.88; identified by sequence similarity;  
 putative. . . .

SCORES Init1: 156 Initn: 235 Opt: 305 z-score: 334.5 E(): 1.2e-10  
 Smith-Waterman score: 380; 25.1% identity in 443 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIKPKSISHRSMFGGLASGETRITGLLEGEDVINTG  
 AE002492\_7 MTESVRLPVARLKPSTVALPGSKSISNRTLLLAALSDNACEIHSLLKSDDTDRML  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110 119  
 KAMQAMGARIRKEG-DTWIIDGVNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 AE002492\_7 EALDKLGVQIEYLAEDRLKVHGTG-GRFPNRTADLFLGNAGTAFRPLTAALAVLGGDYHL  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
 IGDASLTAKRMGRVNLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTITYRVPM---ASAQV  
 120 130 140 150 160 170

AE002492\_7 HGVPFRHERPIGDLVDALRIAGADVEYLKGEHYPLHIGERQDNG-ERVIPIKGNVSSQF  
 120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
 KSAVLLAGLNTPGIT---TVIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEG  
 AE002492\_7 LTALLMA-LPLTGQAPEIRMVGEISKPYIDITLKLMAQFG---VQVINEGYRVFKIPA  
 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280  
 RGKL-TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILT--LQEMGAD  
 AE002492\_7 DAHYHAPEHLHVEGDASSASY-FLAAGLIAATFVRVTGIGANSIQGDVAFARELEKIGAD  
 230 240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340  
 IEVINPRLAGGEDVADL-RVRSSTLKGVTVPEDRAPSMIDEYPIILAVAAFAEGAT-VMN  
 AE002492\_7 V-----VWGENFVEVSRPKERAVQSFDLDANHIP---DAAMTLAIVA-LATGQTCTLR  
 290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400  
 GLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHR  
 AE002492\_7 NIGSWRVKETDRIAAMANELRKLGAKEVEAEAIHITP-PETL---TPDAVIDTYDDHR  
 340 350 360 370 380 390

nk603cp4.pep 410 420 430 440 450  
 IAMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AE002492\_7 MAMCFSLVSLGV-PVVINDPKCTHKTFFPYFDFVSSLTETAE  
 400 410 420 430

nk603cp4.pep  
 GP\_PLN2:AP002542\_28

LOCUS AP002542\_28 [AP002542]  
 DEFINITION Oryza sativa genomic DNA, chromosome 6, PAC clone:P0679C08;  
 contains ESTs  
 D40305(S2184),D24295(R1678),C73562(E11038),  
 AU101334(E11038).  
 DATE 26-JAN-2001 . . .

SCORES Init1: 102 Initn: 178 Opt: 289 z-score: 317.9 E(): 1e-09  
 Smith-Waterman score: 297; 24.5% identity in 404 aa overlap

nk603cp4.pep 40 50 60 70 80 89  
 SHRSFMFGGLASGETRITGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGLLA-  
 AP002542\_28 MLEALKALGLSVEADKAVKRAVVVCGGKFFPV  
 10 20 30

nk603cp4.pep 90 100 110 120 130 140  
 -----PEAPLDFGNAATGCRLTMGLVGVYDFDSTFIGDA---SLTKRPMGRVNLNPLREMGV  
 AP002542\_28 EKDAKEEVLFLGNAGTAMRPLTAAVTAAGGNATYVLDGVPRMRERPIGDLVVLGKLGA  
 40 50 60 70 80 90

nk603cp4.pep 150 160 170 180 190  
 QVKSSEDGRLP-VTLRG-----PKTPTPTITYRVPMASQVKSASVLLAG-LNTPGITT-VIE  
 AP002542\_28 DVDCEFLGTECFVRVKIGIGLPGGKVKLSGSI---SSQYLSALLMAAPALGDEVEIBIID  
 100 110 120 130 140

nk603cp4.pep 200 210 220 230 240 250  
 PIMTRDHTE---KMLQGFGANPTVETDADGVRTIRLEGKLTGQVIDVPGDPSSTAFPL  
 AP002542\_28 KLISIPVYEMTLRLMERFGVKAH-HSDSWDRFYIKGGQKYKSPGNAY-VEGDASSASYPL  
 150 160 170 180 190 200

310            320            330            340            350            360

nk603cp4.pep      370            380            390            400            410            420  
NGVDCDEGETSLVVRRGRPDGKGLGNASGAAVATHLDHRIAMSLVMGLVSENFVTVDAT  
|:::||       ::|: .     :|: | |:|:|:    ::|||:





nk603cp4.pep KAMQAMGARIRKEGDTWI--IDGVNGGGLLAPEAPLDFGNAATGCR-LTMGLV--GVYDF  
 Q9ZEQ0 NALKALGVNYLSSEDKTVCTVEGVGGAFNWKNGLALFLGNAGTAMRPLTAALCLKGSSEA  
 nk603cp4.pep DSTFIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIYRVPMASA  
 Q9ZEQ0 EVVLTGEPRMKERPIKHLVDALRQAGASVQYLENEGYPFAIRNSGLKGGKVQIDGSISS  
 nk603cp4.pep QVKSALLAGLNTPGITT--VIEPIMTR---DHTEKMLQGFANPTVETDADGVRTIRLE  
 Q9ZEQ0 QFLTALLMAAPLAEGDMEIEIIGELVSKPYIDITLAMMKDFG----VKVENRNYQTFVVK  
 nk603cp4.pep G-RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLIL--TLQEMGA  
 Q9ZEQ0 GNQSYLSPEKYLVEGDASSASY-FLAAGAIKKG-KKVTGIGKNSIQGDRLFANVLEAMGA  
 nk603cp4.pep DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNG  
 Q9ZEQ0 -----KITWGDDF--IQAEQGLKGGMDMMNHIP---DAAMTIATAALFAEGKR  
 nk603cp4.pep LEELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRI  
 nk603cp4.pep GP\_BCT2:AJ012748\_1  
 LOCUS AJ012748\_1 [APL012748]  
 DEFINITION Actinobacillus pleuropneumoniae aroA gene.  
 DATE 18-NOV-1998  
 ACCESSION AJ012748  
 NID  
 ORGANISM Actinobacillus pleuropneumoniae . . .  
 SCORES Init1: 137 Initn: 164 Opt: 285 z-score: 314.6 E(): 1.5e-09  
 Smith-Waterman score: 300; 26.0% identity in 339 aa overlap  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AJ012748\_1 MEKITLAPISRVEGEINLPGSKSLNRAALLAAAKGTTKVTNLLDSDDIRHML  
 nk603cp4.pep KAMQAMGARIRKEGDTWI--IDGVNGGGLLAPEAPLDFGNAATGCR-LTMGLV--GVYDF  
 AJ012748\_1 NALKALGVNYLSSEDKTVCTVEGVGGAFNWKNGLALFLGNAGTAMRPLTAALCLKGSSEA  
 nk603cp4.pep DSTFIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIYRVPMASA  
 AJ012748\_1 EVVLTGEPRMKERPIKHLVDALRQAGASVQYLENEGYPFAIRNSGLKGGKVQIDGSISS  
 nk603cp4.pep QVKSALLAGLNTPGITT--VIEPIMTR---DHTEKMLQGFANPTVETDADGVRTIRLE  
 nk603cp4.pep QVKSALLAGLNTPGITT--VIEPIMTR---DHTEKMLQGFANPTVETDADGVRTIRLE

AJ012748\_1 QFLTALLMAAPLAEGDMEIEIIGELVSKPYIDITLAMMKDFG----VKVENRNYQTFVVK  
 nk603cp4.pep G-RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLIL--TLQEMGA  
 AJ012748\_1 GNQSYLSPEKYLVEGDASSASY-FLAAGAIKKG-KKVTGIGKNSIQGDRLFANVLEAMGA  
 nk603cp4.pep DIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFAGATVMNG  
 AJ012748\_1 -----KITWGDDF--IQAEQGLKGGMDMMNHIP---DAAMTIATAALFAEGKR  
 nk603cp4.pep LEELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRI  
 nk603cp4.pep GP\_BCT3:AL583919\_127  
 LOCUS AL583919\_127 [MLEPRTN3]  
 DEFINITION Mycobacterium leprae strain TN complete genome; segment 3/10;  
 Similar to Mycobacterium tuberculosis  
 3-phosphoshikimate 1-carboxyvinyltransferase aroA OR  
 RV3227 OR MTCY20B11.02 SW:AROAMYTU (P22487) fasta  
 scores: E(): 0, 79.0% id in 428 aa, and to Yersinia . . .  
 SCORES Init1: 70 Initn: 211 Opt: 286 z-score: 314.0 E(): 1.6e-09  
 Smith-Waterman score: 364; 28.8% identity in 445 aa overlap  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLAS----GETRITGLLEGEDV  
 AL583919\_127 MSIESWTAIPVASTFPVQANVTIPGSKSQTNRALMLAALAAAGQGTSTIGGALRSRDT  
 nk603cp4.pep INTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCR-LTMGLVGVYDFD  
 AL583919\_127 ELMISALRTLGLHVDGASVLTVNGRITPG---EDAQVDCGLAGTVLRFVPLAALSANP  
 nk603cp4.pep STFIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKTPTPTIYRVPM-ASAQ  
 AL583919\_127 VTFDGDGEQARARPIITPLDALRGLGVPPV---DGVGLPFQVQSGSVAGGTVAIDASASSQ  
 nk603cp4.pep VKSAVLL-AGLNTPGITT--VIEPIMTRDH---TEKMLQGFANPTVETDADGVRTIRLE  
 AL583919\_127 FVSGLLCAASFQGTGVTQHTGSPVPSAPHIAMTVMMLRQAGVQ--VD-DSVGNRWQVRP  
 nk603cp4.pep GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILTLQEMGADIE  
 AL583919\_127 GTVAARHWVVE-PDLTNAVAF--LAAAASVSGGTVRI-----TGWPKTSVQPADNIL  
 nk603cp4.pep VINPRLAGGEDVAD--LRVRSSTL-KGVTVPEDRAPSMIDEYPILAVAAAFAGATV--M  
 AL583919\_127 NILFRLNVVVNQTSDFLEVQGSTVYDGFVDLDRDVGELT---PSVAALALATPGVSQSL  
 nk603cp4.pep NGLEELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDH

AL583919\_127 HGIAHLRGHETDRLAALSTEINRLGGDCQETS DGLIITATPLRPGVWRA-----YADH  
 340 350 360 370 380

nk603cp4.pep RIAMSLV MGLVSEN PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 410 420 430 440 450  
 AL583919\_127 RMAMAGAI VGL-RVSGVEVDDIGATHKTL P QFFPQLWANMLKRSTG  
 390 400 410 420 430

nk603cp4.pep  
 TREMBL\_MAIN:Q9CCI3

ID Q9CCI3 PRELIMINARY; PRT; 430 AA.  
 AC Q9CCI3;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE PUTATIVE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYL TRANSFERASE . . .

SCORES Init1: 70 Initn: 211 Opt: 286 z-score: 314.0 E(): 1.6e-09  
 Smith-Waterman score: 364; 28.8% identity in 445 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVIRPGDKSISHRSFMFGGLAS----GETRITGLLEGEDV  
 10 20 30 40 50  
 Q9CCI3 MSIESWTPAVSTFPVQANVTIPGSKSQTNRALMLAALAAQGGTSTIGGALRSRDT  
 10 20 30 40 50

nk603cp4.pep INTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCR LTMGLVGVYDFD  
 60 70 80 90 100 110  
 Q9CCI3 ELMISALRTLGLHVEAGSVLT VNGRITPG---EDAQVDCGLAGTVLRFPPLAALSANP  
 60 70 80 90 100 110

nk603cp4.pep STFIGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPIYRVPM-ASAQ  
 120 130 140 150 160 170  
 Q9CCI3 VTFDGEQARARPIITPLLDALRGLGVFV---DGVGLPFQVQSGSVAGGTVAIDASASSQ  
 120 130 140 150 160 170

nk603cp4.pep VKSAVLL-AGLNTPGITT--VIEPIIMTRDH---TEKMLQGFGANPTVETDAGVRTIRLE  
 180 190 200 210 220 229  
 Q9CCI3 FVSGLLLCASFTQGVTVQHTGSPVPSAPHIAMTVMLRQAGVQ--VD-DSVGNRQWVRP  
 180 190 200 210 220

nk603cp4.pep GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPNTRTGLILTQEMGADIE  
 230 240 250 260 270 280 289  
 Q9CCI3 GTVAARHWVVE-PDLTNAVAF--LAAAASVGGTVRI-----TGWEKTSVQPADNII  
 230 240 250 260 270

nk603cp4.pep VINPRLAGGEDVAD--LRVRSSTL-KGVTVPEDRAPSMIDEYPI LAVAAAFABGATV--M  
 290 300 310 320 330 340  
 Q9CCI3 NILFRNLNVVNQTD SFLEVQGSTVYDGFVDLRLDVGELT---PSVAALAAALTPGVSQSL  
 280 290 300 310 320 330

nk603cp4.pep NGLLEELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGA AVATHLDH  
 350 360 370 380 390 400  
 Q9CCI3 HGIAHLRGHETDRLAALSTEINRLGGDCQETS DGLIITATPLRPGVWRA-----YADH  
 340 350 360 370 380

nk603cp4.pep RIAMSLV MGLVSEN PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 410 420 430 440 450  
 Q9CCI3 RMAMAGAI VGL-RVSGVEVDDIGATHKTL P QFFPQLWANMLKRSTG

390 400 410 420 430

nk603cp4.pep  
 TREMBL\_MAIN:Q9HQC1

ID Q9HQC1 PRELIMINARY; PRT; 439 AA.  
 AC Q9HQC1;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . . .

SCORES Init1: 151 Initn: 398 Opt: 282 z-score: 309.5 E(): 2.9e-09  
 Smith-Waterman score: 625; 32.2% identity in 441 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVIRPGDKSISHRSFMFGGLASGETRITGLLEGEDVINT  
 10 20 30 40 50 59  
 Q9HQC1 MPWAALLAGMHATVSPSRVVRGRARAPPSKSYTHRALAAGYADGETTVVRDPLVSADTRAT  
 10 20 30 40 50 60

nk603cp4.pep GKAMQAMG-ARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCR LTMGLVGVYDFDST  
 60 70 80 90 100 110  
 Q9HQC1 ARAVELLGGAAARENGD-WVVTGFGSRPAI-PDAVIDCANSGMTMLRVTA AALADGTTV  
 70 80 90 100 110

nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVKSEGD-RLPVTLRGPKTPTPIYRVPM-ASAQV  
 120 130 140 150 160 170  
 Q9HQC1 LTGDESLRARPHGPLLDALSGLGGTARSTRNGQAPLVVDGPVSGGSVA--LPGDVSSQF  
 120 130 140 150 160 170

nk603cp4.pep KSAVLLAG-LNTPGI---TTVIEPIIMTRDHTKMLQGFGANPTVETDAGVRTIRLEGR  
 180 190 200 210 220 230  
 Q9HQC1 VTALLMAGAVTETGIETDLTTELKSAPYVDITLDVLDAGVGAS-ETAA-GYR-VRGGQA  
 180 190 200 210 220 230

nk603cp4.pep GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPNTRTG---LILTLQEMGADI  
 240 250 260 270 280  
 Q9HQC1 YAPSGAEYAVPGDFSSASYLLAAGALAAADGA AVVVEGMHPSAQGDAAIVDVLERMGADI  
 240 250 260 270 280 290

nk603cp4.pep EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFABGATVMMNGLE  
 290 300 310 320 330 340  
 Q9HQC1 DW-----DTESGVITVQRELSGVEVGADTPDLL---PTIAVLGAAADGTRITDAE  
 300 310 320 330 340

nk603cp4.pep ELRVKESDRLSAVANGLKLVGDCDEGETSLVVRGRPDGKGLGNASGA AVATHLDHRIAM  
 350 360 370 380 390 400  
 Q9HQC1 HVRVYKETDRVAAMAE SSKLGASVEERPEDELVVRG---GDTELSGASVDGRGDRHLVM  
 350 360 370 380 390

nk603cp4.pep SFLVMGLVSEN PVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 410 420 430 440 450  
 Q9HQC1 ALAVAGLVADGETTIAGSEHVDVSFPDFFEVLGAGLADTDG  
 400 410 420 430

nk603cp4.pep  
 GP\_BCT1:AE005049\_6

LOCUS AE005049\_6 [AE005049]  
 DEFINITION Halobacterium sp. NRC-1 section 80 of 170 of the complete genome;

Psc.  
DATE 12-FEB-2001  
ACCESSION AE005049  
NID . . .

SCORES Init1: 151 Initn: 398 Opt: 282 z-score: 309.5 E(): 2.9e-09  
Smith-Waterman score: 625; 32.2% identity in 441 aa overlap

```
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINT
AE005049_6 MPWAALLAGMHATVSPSRVGRARAPPSKSYTHRALAAGYADGETVVRDPLVSADTRAT
```

```
nk603cp4.pep GKAMQAMG-ARIRKEGDTWIIDGVGNGLLAPEAPLDFGNAATGCRLTMGLVGVDYFDST
AE005049_6 ARAVELLGGAAARENGD-WVVTGFGSRPAI-PDAVIDCANSGTTMRLVTAAALADGTTV
```

```
nk603cp4.pep FIGDASLTKRPMGRVNLNPLREMGVQKSEGDG-RLPVTLRGPKTPTPITYRVPM-ASAQV
AE005049_6 LTGDESLRAPHGPLLDALSLGGTARTRGNGQAPLVVDGFPVSGSVA--LPGDVSSQF
```

```
nk603cp4.pep KSAVLLAG-LNTPGI---TTVIEPIMTRDHEKMLQGFGANPTVETDADGVRTIRLEGR
AE005049_6 VTALLMAGAVTETGIETDLTTELKSAFYVDITLDVLDFAFGVAS-ETAA-GYR-VRGGQA
```

```
nk603cp4.pep GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTG---LILTQEMGADI
AE005049_6 YAPSGAEYAVFGDFSSASYLLAAGALAADGAADVVEGMHPSAQGDAAIVDVLERMGADI
```

```
nk603cp4.pep EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFABGATVMNGLE
AE005049_6 DW-----DTESGVITVQRSELSGVEVGADTPDLL---PTIAVLGAAADGTPTRITDAE
```

```
nk603cp4.pep ELRVKESDRLSAVANGLKINGVDCDEGETSLVVVRGPDGKGLGNASGAAVATHLDHRIAM
AE005049_6 HVRVKETDRVAAMASLSKLGASVEERPDELVRG----GDTELSGASVDGRGDHRLVM
```

```
nk603cp4.pep SFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AE005049_6 ALAVAGLVADGETTIAGSEHVDVSFPDFFEVLGAGLADTDG
```

nk603cp4.pep  
GP\_BCT3:U67500\_3

LOCUS U67500\_3 [U67500]  
DEFINITION Methanococcus jannaschii section 42 of 150 of the complete genome;  
similar to GB:L04686 SP:Q03421 PID:148865 percent  
identity: 38.11; identified by sequence similarity;  
putative.  
DATE 28-JAN-1998 . . .

SCORES Init1: 156 Initn: 294 Opt: 280 z-score: 307.5 E(): 3.7e-09  
Smith-Waterman score: 610; 31.0% identity in 451 aa overlap

```
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
U67500_3 MYLLIVKKTDRLEGIVKAPPSKSYTHRAVIGASLADGVSRINPLWGADCLSSV
```

```
nk603cp4.pep KAMQAMGARIR--KEGDTWIIDGVGNGLLAPEAPLDFGNAATGCRLTMGLVG-VYDFDS
U67500_3 HGCRLMGANIELDKEKDEWIVKG---GELKTPDNIIDIGNSGTTLRILTSTIASQIPKGYA
```

```
nk603cp4.pep TFIGDASLTKRPMGRVNLNPLREMGVQKSEGDG-RLPVTLRGPKT-PTPITYRVPMASQA
U67500_3 ILTGDDSIKRPMPQLLDALKQLNIEAFSSKLDGTAPIIVKSGKIYGNVVKIRGDISSQF
```

```
nk603cp4.pep VKSAVLLAGLNTPGITTVIE-PIMTR---DHTEKMLQGGFANPTVETDADGVRTIRLEGR
U67500_3 ITSLMMLLPFNKEDTEIILTSPLKSKFYIDITLDILNKFGIK--IDKTDNG---FLVYGN
```

```
nk603cp4.pep GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTG---LILTQEMGADI
U67500_3 QKYKPIDYIVEGYSSASY-LIAAGVLINSNITIEIENLFAN-SKQGDKAIINIVKEMGADI
```

```
nk603cp4.pep EVINPR-LAGGEVDADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAAFABGAT-VMNG
U67500_3 KVKKDKVIEEGE-----YSLKGIDVDVKDIPDLV---PTIAVLGCGFAEGKTEIYNG
```

```
nk603cp4.pep LEEELRVKESDRLSAVANGLKINGVDCDEGETSLVVVRGPRPDGKGLGNASGAAVATHLDHRI
U67500_3 -EHVRLKECDRLRACAVELKKMGADIEEKPDGLIIRGVKKLKG-----AKINTYDHRL
```

```
nk603cp4.pep AMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
U67500_3 VMAFTIAGLKAEGETIEEGEAVKISFPNFVDVMKSLGANIEVK
```

nk603cp4.pep  
SWISSPROT:AROA\_METJA

ID AROA\_METJA STANDARD; PRT; 429 AA.  
AC Q57925;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE PROBABLE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) . . .

SCORES Init1: 156 Initn: 294 Opt: 280 z-score: 307.5 E(): 3.7e-09  
Smith-Waterman score: 610; 31.0% identity in 451 aa overlap

```
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AROA_METJA MYLLIVKKTDRLEGIVKAPPSKSYTHRAVIGASLADGVSRINPLWGADCLSSV
```

nk603cp4.pep KAMQAMGARIR--KEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 AROA\_METJA HGCRMLGANIELDKKEDWIVKG---GELKTPDNIIDIGNSGTTLRILTSIASQIPKGYA  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
 TFIFGASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMASQVKS  
 AROA\_METJA ILTGDDSIKRPMPQPLLDALKQLNIEAFSSKLDGTAPIIVKSGKIYGNVVKIRGDISSQF  
 120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
 VKSAVLLAGLNTPGITTVIE-PIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGR  
 AROA\_METJA ITSLLMLLPFNKEDTEIILTSPKSKPYIDITLDILNKPGIK---IDKTONG---FLVYGN  
 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280  
 GKLTVGQVIDVPGDPSSTAFFPLVAALLVPGSDVTILNLMNPTRTG---LILTQEMGADI  
 AROA\_METJA QKYKPIDYIVVEGYSSASY-LIAAGVLINSNITENLFAN-SKQGDKAIINIVKEMGADI  
 230 240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340  
 EVINPR-LAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIAVAAAFAGAT-VMNG  
 AROA\_METJA KVKKDKVIIIEG-----YSLKGIIDVDVKDIPDLV---PTIAVLGCFAGKTEIYNG  
 290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400  
 LELRLVKESDRLSAVANGLKINGVDCDEGETSLVVVRGPDGKGLGNASGAAVATHLDHRI  
 AROA\_METJA -EHVRLKECDRLRACAVELKKMGADIEEKPDGLIIRGVKKLKG-----AKLNTYHDHRL  
 340 350 360 370 380

nk603cp4.pep 410 420 430 440 450  
 AMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AROA\_METJA VMAFTIAGLKAEGETIIEGEEAVKISFPNPVDMVKSLGANIEVK  
 390 400 410 420

nk603cp4.pep  
 GP\_BCT2:AP000060\_34

LOCUS AP000060\_34 [AP000060]  
 DEFINITION Aeropyrum pernix genomic DNA, section 3/7;  
 similar to PIR:D69202 percent identity:38.995 in  
 418aa. motif=EPSP synthase signatures.  
 DATE 06-APR-2000  
 ACCESSION AP000060 . . .

SCORES Init1: 130 Initn: 249 Opt: 275 z-score: 302.2 E(): 7.5e-09  
 Smith-Waterman score: 527; 31.4% identity in 439 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_AERPE MVWLRAPDRVVVHPSTVEGRVEAPPSSKSYTHRMFLALLARGSVVRRPLVSNDTLATL  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110 119  
 KAMQAMGARIR-KEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 AROA\_AERPE NAVALLGKPKRLGRGVAEVEGGEVRRGAVVYAA---GSGTT-IRIANG-VAHSAEATL  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
 I-GDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMASQVKS  
 AROA\_AERPE LYGDSESLNRRPVHPLSEALRSMGARVCDTGGNP-PVKVSGPLRRASVEVDAAISS-QFAT  
 120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
 AVLLAGLNTPGITTVIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRGKLT  
 AROA\_AERPE SLLIAGSRLGEFELSAARLSSRGYVDITLESLSMFG---VRVEREGYRLFLRGTPKPV  
 180 190 200 210 220

AP000060\_34 LYGDSESLNRRPVHPLSEALRSMGARVCDTGGNP-PVKVSGPLRRASVEVDAAISS-QFAT  
 120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
 AVLLAGLNTPGITTVIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRGKLT  
 AP000060\_34 SLLIAGSRLGEFELSAARLSSRGYVDITLESLSMFG---VRVEREGYRLFLRGTPKPV  
 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 290  
 GQVIDVPGDPSSTAFFPLVAALLVPGSDVTILN-VLMNPTRTGLILTQEMGADIEVINPR  
 AP000060\_34 DAA--VPGDYSSASFMLAAGAIAGRVEVEGLRPVDPQPDPRR-IVELLRSMGARV-----R  
 230 240 250 260 270

nk603cp4.pep 300 310 320 330 340 350  
 LAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIAVAAAFAGATVMNGLEELRVKE  
 AP000060\_34 VEGGVVAVE---STGPLEFVDVLDGSP---DLAPVAAVLAAYARGVSRRLRGLERLKYKE  
 280 290 300 310 320 330

nk603cp4.pep 360 370 380 390 400 410  
 SDRLSAVANGLKINGVDCDEGETSLVVVRGPDGKGLGNASGAAVATHLDHRIAMSFVLMG  
 AP000060\_34 SDRLSAIAWNLARLGVEARVRRGILEIRG-----GGVEGGVARSWGDHRIAMAMAVAG  
 340 350 360 370 380

nk603cp4.pep 420 430 440 450  
 LVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AP000060\_34 LGARRPVAVEGFSRVDPDSYPGFLDLARLGARVEAVKGGGV  
 390 400 410 420

nk603cp4.pep  
 SWISSPROT:AROA\_AERPE

ID AROA\_AERPE STANDARD; PRT; 427 AA.  
 AC Q9YEK9;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE PROBABLE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) . . .

SCORES Init1: 130 Initn: 249 Opt: 275 z-score: 302.2 E(): 7.5e-09  
 Smith-Waterman score: 527; 31.4% identity in 439 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AROA\_AERPE MVWLRAPDRVVVHPSTVEGRVEAPPSSKSYTHRMFLALLARGSVVRRPLVSNDTLATL  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110 119  
 KAMQAMGARIR-KEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTF  
 AROA\_AERPE NAVALLGKPKRLGRGVAEVEGGEVRRGAVVYAA---GSGTT-IRIANG-VAHSAEATL  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
 I-GDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMASQVKS  
 AROA\_AERPE LYGDSESLNRRPVHPLSEALRSMGARVCDTGGNP-PVKVSGPLRRASVEVDAAISS-QFAT  
 120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
 AVLLAGLNTPGITTVIEPIMTR---DHTEKMLQGFGANPTVETDADGVRTIRLEGRGKLT  
 AROA\_AERPE SLLIAGSRLGEFELSAARLSSRGYVDITLESLSMFG---VRVEREGYRLFLRGTPKPV  
 180 190 200 210 220

300            310            320            330            340            350

nk603cp4.pep      360      370      380      390      400      410  
 ESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGA AVATHLDHRIAMSFVLM

nk603cp4.pep  
GP\_BCT3:M62708\_1

```

          420          430          440          450
nk603cp4.pep  GL-VSENPTVDDATMTATSFPEFMDLMAGLGAKIELSDTKAA
               | : - : : : : : : : | : | : | : | :
AROA_CHLMU    GVHASSGQTLLCNTQCYNKSFPHFVIAAQTLHANIRRHQADFSLRSSLCR
               400          410          420          430          440

```

```

LOCUS       M62708_1 [MSGARO]
DEFINITION  M.tuberculosis aroA gene, complete cds;
             key enzyme from the shikimate biosynthesis pathway.
DATE        26-APR-1993
ACCESSION   M62708
NID

```

nk603cp4.pep  
TREMBL\_MAIN:Q97KM2

```
ID Q97KM2 PRELIMINARY; PRT; 428 AA.
AC Q97KM2;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE. . . .
```

SCORES Init1: 76 Initn: 222 Opt: 264 z-score: 289.9 E(): 3.6e-08  
Smith-Waterman score: 386; 29.4% identity in 439 aa overlap

SCORES Init1: 154 Initn: 383 Opt: 266 z-score: 292.4 E(): 2.6e-08  
Smith-Waterman score: 482; 26.0% identity in 435 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRFSFMGGLASGETRITGLLEGEDVINTG  
Q97KM2 MNCVKINPCCLKGDIKIPPSKSLGHRRAICAAALSEESTIENISYSKDIKATC

```

          10      20      30      40      50
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHSRSMFVGGGLAS----GETRITGLLEGEIV
              || :||:| :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
M62708_1      MKTPWAPTAPTVPVRAITVTPGSKSQTNRAVLVLAALAAAOGRGASTISGALSRDT

```

nk603cp4.pep      KAMQANGARIRKEGDTWIIDGVGNGLLAPAE-APLDFGNAATGCRLTMGLVGVDSTF  
                     |::|||:::        : : |:| : : |:::| : : : : : : : :  
Q97KM2             IGMSKLGALIIEDAKDNSTLKIKKQKLVSKEKYIDCSSESGTVRFLIPISLIEERNVVF  
                     |         70                     80                     90                     100                     110

```

          60           70           80           90          100          110
nk603cp4.pep INTGKAMQAMGARIRKEGDWTWIIDGVNGGLLAPEAPLDFGNAAATGCRLTMGLVGVDYFD
              ::::| : | : : ||||| : ||| : ||| : ||| : ||| : 
M62708_1      ELMIDLALQTILGRVDSVLTVSGRIEPG---PGARVDCGLAGTIVLRFPPLAALSVPF
               60       70       80       90        100        110

```

```

      120       130       140       150       160       170
nk603cp4.pep IGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTP----TPITRYRVPMASAQ
               :::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Q97KM2        DGQGKLSYRPLDSYFNIFDEKEIAYSHPEGKVLPLQGIKRLLKACMFNLPGNISSQPIISGL
               ::::~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:~::~:
      120       130       140       150       160       170

```

nk603cp4.pep      120          130          140          150          160          170  
STFIGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPIYRVPM-ASAQ  
||| : | : ||::||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:  
M62708\_1    VTFDGDDQARGRPIAPLLDALRELGVAV---DGTGLPFVRVRCNGSLAGGTVAIDASASSQ  
             120          130          140          150          160

[illegible]

```

      180          190          200          210          220          229
nk603cp4.pep VKSAVLLAGLN--TFGITT--VIEPIIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLE
              |:|::||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
M62708_1     FVSGLLSAASFDTGLTVQHTGSSLP SAPHIAMTAAMLRLQAGV-----DIDDSIPNRWQ
            170          180          190          200          210          220

```

```

          240          250          260          270          280          290
nk603cp4.pep  GQVIDVPGDPSSTAFFLVAALLVPGSDVITLNVLMNPTRTGLILTLQEMGADIEVINPRL
               |||::||::||::||::||::||::||::||::||::||::||::||::||::||::||
Q97KM2        GTKYKVEGDFSQAQFWSLGSILNGNINCKDLN-----ISSLQGDKIVILDILK-KM
               230          240          250          260          270

```

```

      230      240      250      260      270      280
nk603cp4.pep GR-GKLTGQVIDVPGDPSSTAFLVLAALLVPGSDVTILN---VLMNPTRTGLILTLQEMG
              |  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
M62708_1      VRPGFVAARRWDIEPDLTN-AVAFLSAAVVSGGTVRITGWPRVSVQPADHILAI-LRQLN
              230      240      250      260      270      280

```

```

          300      310      320      330      340      350
nk603cp4.pep AGGEDVADLRVRSSTLKGVTVPEDRAPSMID EYPILAVAAFAEGATVMNGLEELRVKES
               :|:|  :::|  :|::  ::|::  |||:|:|:|:|:|:|  ::  :|:|:|
Q97KM2        GGAIDEKSFSSKKSHTHGIVIDASQCPLDV---PTLSVVAALSEGTTKIVNAARLRIKES
               280      290      300      310      320      330

```

```

      290      300      310      320      330      340
nk603cp4.pep ADIEVINPRLAGGEDVADLRVRSSST-LKGVTVPEDRAPSMID EYPILAVAAFAEGATV-
               |||:      ::|||:      | | | | :      | : | | :      ::|
M62708_1      A--VVIHAD-----SSLEVRGPTGYDGFVDV-DLRVAGELT--PSVAALAAALSPGSVS
               290      300      310      320

```

nk603cp4.pep      360            370            380            390            400            410  
                  DRRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLHDHRIAMSFVLVMGL  
                  |||::||: | : |:: | | : ::|: ||         : : | ||| : : : |  
 Q97KM2          DRLLKAMATELNKLGAEVVELEDGLLEGKEKLGK-----GEVESWNDHRIAMALGIAAL  
                  340            350            360            370            380

```

          350      360      370      380      390      400
nk603cp4.pep -MNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHL
               ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
M62708_1     RLSGIAHLRGHETDRLAALSTEINRLGGTCRETPDLGVITATPLRPGIWR-----YA
           330       340       350       360       370       380

```

nk603cp4.pep VSENPTVDDATMIATSFPEFMDLMAGLAKIELSDTKAA  
| : || :: : : | : | : : || ::  
Q97KM2 RCEESVTINGSECVSKSPQFWSDLKQLGGDVHEWSLGE  
390 400 410 420

nk603cp4.pep DHRIAMSFVLVGLVSENPVTVDATMIATSPFPEFMDLMAGLGAKEIELSDTKAA  
|||::|: ::|| : | ||| : : ::|| | |  
M62708\_1 DHRMAMAGATIGLRVAG-VEVDDIAATTKTLPEFPRLNAEMVPGQGWGYPQPRSGQRAR  
300 400 500 600 700 800 900 1000

M62708\_1      RATGQSGG  
450

nk603cp4.pep  
SWISSPROT:AROA\_MYCTU

ID AROA\_MYCTU STANDARD; PRT; 450 AA.  
AC P22487;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 01-AUG-1991 (Rel. 19, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 76 Initn: 222 Opt: 264 z-score: 289.9 E(): 3.6e-08  
Smith-Waterman score: 386; 29.4% identity in 439 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLAS----GETRITGLLEGEDV  
ARO\_AMYCTU MKTWPAPTAPTTPVRATVTVPGSKSQTNRALVLAALAAQAQGRGASTISGALRSRDT  
10 20 30 40 50  
10 20 30 40 50

nk603cp4.pep INTGKAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRITMGLVGVDYDFD  
ARO\_AMYCTU ELMLDALQTLGLRVGVSELTVSGRIEPG---PGARVDCGLAGTVLRFVPLAALGSVP  
60 70 80 90 100 110  
60 70 80 90 100 110

nk603cp4.pep STFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPIYRVPM-ASAQ  
ARO\_AMYCTU VTFDGDQQAARGRIAPLLDALRELGVAV---DGTGLPFRVRNGSLAGGTVAIDASASSQ  
120 130 140 150 160 170  
120 130 140 150 160

nk603cp4.pep VKSAVLLAGLN-TPGIT--VIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLE  
ARO\_AMYCTU FVSGLLLSAASFTDGLTVQHTGSSLPASPHIAMTAAMLROAGV-----DIDDSTPNRWQ  
180 190 200 210 220 229  
170 180 190 200 210 220

nk603cp4.pep GR-GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILN---VLMNPTRTGLILTLQEMG  
ARO\_AMYCTU VRPGPVAARRWDIEPDLTN-AVAFLSAAVVSGGTVRITGWPRVSVQPADHILAI-LRQLN  
230 240 250 260 270 280  
230 240 250 260 270 280

nk603cp4.pep ADIEVINPRLAGGEDVADLRVRSST-LKGVTVPEDRAPSMIDEPILAVAAAFAGATV-  
ARO\_AMYCTU A--VVIHAD-----SSLEVRGPTGYDGFVDV-DLRAVGELT--PSVAALASPGSVS  
290 300 310 320 330 340  
290 300 310 320

nk603cp4.pep -MNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGPDGKGLGNASGAAVATHL  
ARO\_AMYCTU RLSGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATPLRPGIWRA-----YA  
350 360 370 380 390 400  
330 340 350 360 370 380

nk603cp4.pep DHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
ARO\_AMYCTU DHRMAMAGAIIGLRVAG-VEVDIAATTKTLPEFPRLWAEMVGPQGQGWYPQPRSGQRAR  
410 420 430 440 450  
390 400 410 420 430 440

ARO\_AMYCTU RATGQSGGG  
450

nk603cp4.pep  
GP\_BCT3:X52269\_1

LOCUS X52269\_1 [MTAROA]  
DEFINITION M. tuberculosis aroA gene for 5-enolpyruvylshikimate-3-phosphate  
synthase (EC 2.5.1.19);  
aroA protein (AA 1-450).  
DATE 10-FEB-1999

ACCESSION X52269 . . .

SCORES Init1: 76 Initn: 222 Opt: 264 z-score: 289.9 E(): 3.6e-08  
Smith-Waterman score: 386; 29.4% identity in 439 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLAS----GETRITGLLEGEDV  
X52269\_1 MKTWPAPTAPTTPVRATVTVPGSKSQTNRALVLAALAAQAQGRGASTISGALRSRDT  
10 20 30 40 50  
10 20 30 40 50

nk603cp4.pep INTGKAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRITMGLVGVDYDFD  
X52269\_1 ELMLDALQTLGLRVGVSELTVSGRIEPG---PGARVDCGLAGTVLRFVPLAALGSVP  
60 70 80 90 100 110  
60 70 80 90 100 110

nk603cp4.pep STFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPIYRVPM-ASAQ  
X52269\_1 VTFDGDQQAARGRIAPLLDALRELGVAV---DGTGLPFRVRNGSLAGGTVAIDASASSQ  
120 130 140 150 160 170  
120 130 140 150 160

nk603cp4.pep VKSAVLLAGLN-TPGIT--VIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLE  
X52269\_1 FVSGLLLSAASFTDGLTVQHTGSSLPASPHIAMTAAMLROAGV-----DIDDSTPNRWQ  
180 190 200 210 220 229  
170 180 190 200 210 220

nk603cp4.pep GR-GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILN---VLMNPTRTGLILTLQEMG  
X52269\_1 VRPGPVAARRWDIEPDLTN-AVAFLSAAVVSGGTVRITGWPRVSVQPADHILAI-LRQLN  
230 240 250 260 270 280  
230 240 250 260 270 280

nk603cp4.pep ADIEVINPRLAGGEDVADLRVRSST-LKGVTVPEDRAPSMIDEPILAVAAAFAGATV-  
X52269\_1 A--VVIHAD-----SSLEVRGPTGYDGFVDV-DLRAVGELT--PSVAALASPGSVS  
290 300 310 320 330 340  
290 300 310 320

nk603cp4.pep -MNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGPDGKGLGNASGAAVATHL  
X52269\_1 RLSGIAHLRGHETDRLAALSTEINRLGGTCRETPDGLVITATPLRPGIWRA-----YA  
350 360 370 380 390 400  
330 340 350 360 370 380

nk603cp4.pep DHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
X52269\_1 DHRMAMAGAIIGLRVAG-VEVDIAATTKTLPEFPRLWAEMVGPQGQGWYPQPRSGQRAR  
410 420 430 440 450  
390 400 410 420 430 440

X52269\_1 RATGQSGGG  
450

nk603cp4.pep  
GP\_BCT3:Z95121\_2

LOCUS Z95121\_2 [MTY20B11]  
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 139/162;  
Rv3227, (MTCY20B11.02), len: 450, AroA,  
3-phosphoshikimate 1-carboxyvinyl transferase (EC  
2.5.1.19; 5-enolpyruvylshikimate-3-phosphate synthase).  
FASTA best: M62708|MSGARO\_1, AROA\_MYCTU; contains PS00885 . . .

SCORES Init1: 76 Initn: 222 Opt: 264 z-score: 289.9 E(): 3.6e-08  
Smith-Waterman score: 386; 29.4% identity in 439 aa overlap



```

      10      20      30      40      50
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLAS----GETRITGLLEGEDV
Z95121_2      MKTWPAPTAPTFRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDT
      10      20      30      40      50
      60      70      80      90     100     110
nk603cp4.pep INTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFD
Z95121_2      ELMLDALQTLGLRVGVSELTVSGRIEPG---PGARVDCGLAGTVLRFVPLAALGSVP
      60      70      80      90     100     110
      120     130     140     150     160     170
nk603cp4.pep STFIGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPITYRVPM-ASAO
Z95121_2      VTFDGDQQAARGRIAPLLDALRELGVAV---DGTGLPFRVRNGSLAGGTVAIDASASSQ
      120     130     140     150     160
      180     190     200     210     220     229
nk603cp4.pep VKSAVLLAGLN-TPGITT--VIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLE
Z95121_2      FVSGLLLSAASFTDGLTVQHTGSSLPASPHIAMTAAMLROAGV-----DIDDSTPNRWQ
      120     130     140     150     160
      170     180     190     200     210     220
      230     240     250     260     270     280
nk603cp4.pep GR-GKLTGQVIDVPGDPSSSTAFPLVAALLVPGSDVTILN---VLMNPTRTGLILTLQEMG
Z95121_2      VRPGFVAARRWDIEPDLTN-AVAFLSAAVVSGGTVRITGWPRVSVQPADHILAI-LRQLN
      230     240     250     260     270     280
      290     300     310     320     330     340
nk603cp4.pep ADIEVINPRLAGGEDVADLRVRSST-LKGVTVPEDRAPSMIDEPILAVAAFAEGATV-
Z95121_2      A--VVIHAD-----SSLEVRGPTGYDGFVDV-DLRAVGELT--PSVAALALASPGSVS
      290     300     310     320
      350     360     370     380     390     400
nk603cp4.pep -MNGLEELRVKESDRLSAVANGKLNQVDCDEGETSLVVRGPRDGKGLGNASGAAVATHL
Z95121_2      RLSGIAHLRGHETDRLAALSTEINRLGGTCRETDPGLVITATPLRPGIWR-----YA
      330     340     350     360     370     380
      410     420     430     440     450
nk603cp4.pep DHRIAMSFVLMGLVSENFVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Z95121_2      DHRMAMAGAIIGLRVAG-VEVDDIAATTKTLPEFPRLWAEMVGPQGQGWGYPQPRSGQRAR
      390     400     410     420     430     440
Z95121_2      RATGQSGGG
      450

```

nk603cp4.pep  
GP\_BCT1:AE007144\_4

LOCUS AE007144\_4 [AE007144]  
DEFINITION Mycobacterium tuberculosis CDC1551, section 230 of 280 of the complete genome; similar to SP:P24497 GB:X82415 PID:562760; identified by sequence similarity; putative.  
DATE 27-APR-2001 . . .

SCORES Init1: 76 Initn: 222 Opt: 264 z-score: 289.9 E(): 3.6e-08  
Smith-Waterman score: 386; 29.4% identity in 439 aa overlap

```

      10      20      30      40      50
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLAS----GETRITGLLEGEDV
AE007144_4      MKTWPAPTAPTFRATVTVPGSKSQTNRALVLAALAAQGRGASTISGALRSRDT

```

```

      10      20      30      40      50
nk603cp4.pep INTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFD
AE007144_4      ELMLDALQTLGLRVGVSELTVSGRIEPG---PGARVDCGLAGTVLRFVPLAALGSVP
      60      70      80      90     100     110
      120     130     140     150     160     170
nk603cp4.pep STFIGDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPITYRVPM-ASAO
AE007144_4      VTFDGDQQAARGRIAPLLDALRELGVAV---DGTGLPFRVRNGSLAGGTVAIDASASSQ
      120     130     140     150     160
      180     190     200     210     220     229
nk603cp4.pep VKSAVLLAGLN-TPGITT--VIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLE
AE007144_4      FVSGLLLSAASFTDGLTVQHTGSSLPASPHIAMTAAMLROAGV-----DIDDSTPNRWQ
      170     180     190     200     210     220
      230     240     250     260     270     280
nk603cp4.pep GR-GKLTGQVIDVPGDPSSSTAFPLVAALLVPGSDVTILN---VLMNPTRTGLILTLQEMG
AE007144_4      VRPGFVAARRWDIEPDLTN-AVAFLSAAVVSGGTVRITGWPRVSVQPADHILAI-LRQLN
      230     240     250     260     270     280
      290     300     310     320     330     340
nk603cp4.pep ADIEVINPRLAGGEDVADLRVRSST-LKGVTVPEDRAPSMIDEPILAVAAFAEGATV-
AE007144_4      A--VVIHAD-----SSLEVRGPTGYDGFVDV-DLRAVGELT--PSVAALALASPGSVS
      290     300     310     320
      350     360     370     380     390     400
nk603cp4.pep -MNGLEELRVKESDRLSAVANGKLNQVDCDEGETSLVVRGPRDGKGLGNASGAAVATHL
AE007144_4      RLSGIAHLRGHETDRLAALSTEINRLGGTCRETDPGLVITATPLRPGIWR-----YA
      330     340     350     360     370     380
      410     420     430     440     450
nk603cp4.pep DHRIAMSFVLMGLVSENFVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AE007144_4      DHRMAMAGAIIGLRVAG-VEVDDIAATTKTLPEFPRLWAEMVGPQGQGWGYPQPRSGQRAR
      390     400     410     420     430     440
AE007144_4      RATGQSGGG
      450

```

nk603cp4.pep  
GP\_BCT1:AE001959\_10

LOCUS AE001959\_10 [AE001959]  
DEFINITION Deinococcus radiodurans R1 section 96 of 229 of the complete chromosome 1; similar to GB:L77117 SP:Q57925 PID:1591205 percent identity: 53.76; identified by sequence similarity; putative. . . .

SCORES Init1: 96 Initn: 309 Opt: 261 z-score: 286.5 E(): 5.5e-08  
Smith-Waterman score: 554; 31.8% identity in 449 aa overlap

```

      10      20      30      40
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRI
AE001959_10 AALPGGALYPVRMSDALPATFDVIVHPARELRGELRAQPSKNYTRYLLAALAEGETRV
      20      30      40      50      60      70
      50      60      70      80      90     100
nk603cp4.pep TGLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEA--PLDFGNAATGCRL
      50      60      70      80      90     100

```

```

      120          130          140          150          160          170
nk603cp4.pep  TFIFGDA--SLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITYRVPMSAQ
              ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
AROA_DEIRA    TFVTDYPDSLGKRPGQDILLEALERLGAWVSSNDG-RLPISVSGPVRGGTVEVSAERS-C
      120          130          140          150          160          170

```

210            220            230            240            250            260            269

nk603cp4.pep FGANPTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNLV  
 ARO1\_TOBAC FGVSVHEHTSSWDKF-LVRGGQKYKSPGKAY-VEGDASSASYFLAGAA-VTGGTVTVVEGCG  
 300 310 320 330 340 350

nk603cp4.pep MNPTRTGLILT--LQEMGADIEVI-NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMID  
 ARO1\_TOBAC TSSLQGDVVKFAEVLEKMGAEVTVTENSVTVKGPFRNSGSMKH--LRAVDVNMNMKMP---D  
 360 370 380 390 400

nk603cp4.pep EYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVVRGRPD  
 ARO1\_TOBAC VAMTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGLATVVEGSDYCIIT--PP  
 410 420 430 440 450 460

nk603cp4.pep GKGLGNASGAAVATHLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAK  
 ARO1\_TOBAC EKL--NVT--EIDTYDDHRMAMAF-SLAACADVPTIKDPGCTRKTFPNYFDVLQYQSKH  
 470 480 490 500 510

nk603cp4.pep IELSDTKAA

nk603cp4.pep  
 GP\_PLN4:M61904\_1

LOCUS M61904\_1 [TOBEPSPS1]  
 DEFINITION N.tabacum 5-enolpyruvylshikimate-3-phosphate synthase mRNA,  
 complete cds.  
 DATE 14-JUL-1993  
 ACCESSION M61904  
 NID . . .

SCORES Initl: 134 Initn: 272 Opt: 261 z-score: 285.8 E(): 6.1e-08  
 Smith-Waterman score: 417; 26.1% identity in 444 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSMFMGGLASGETRI  
 M61904\_1 LRVAKSPFRISASVVTQAQKNEIVLQPIKDISGTVKLPKSKLSNRILLALLSEGRITV  
 60 70 80 90 100 110

nk603cp4.pep TGLLEGEDVINTGKAMQAMGARIR--KEGDTWIIDGVGN---GGLLAPEAPLDFGNAAT  
 M61904\_1 DNLLSSDDIHMYLGALKTLGLHVEDDNENQRAIVEGCGGQFPVGGKSEEEIQFLGNAGT  
 120 130 140 150 160 170

nk603cp4.pep GCRLTMGLGVYDFDSTFIGDA--SLTKRPMGRVLNPLREMGVQKSEDDGRLP---VTL  
 M61904\_1 AMRPLTAAVTVAGGHSRYVLDGVPRMRERPIGDLVDGLKQLGAEDVDCFLGTNCPVPRIVS  
 180 190 200 210 220 230

nk603cp4.pep RG--PKTPTPIYRVPMASAVKSAVLLAG-LNTPGITT-VIEPIMTRDHT---KMLQG  
 M61904\_1 KGGLPGGKVKLSGSI---SSQYLTALLMAAPLALGDVEIEIDKLISVPYVEMTLKLMER  
 240 250 260 270 280 290

nk603cp4.pep FGANPTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNLV  
 M61904\_1 FGVSVHEHTSSWDKF-LVRGGQKYKSPGKAY-VEGDASSASYFLAGAA-VTGGTVTVVEGCG  
 300 310 320 330 340 350

nk603cp4.pep MNPTRTGLILT--LQEMGADIEVI-NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMID  
 M61904\_1 TSSLQGDVVKFAEVLEKMGAEVTVTENSVTVKGPFRNSGSMKH--LRAVDVNMNMKMP---D  
 360 370 380 390 400

nk603cp4.pep EYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVVRGRPD  
 M61904\_1 VAMTLAVVALFADGPTAIRDVASWRVKETERMIAICTELRKLGLATVVEGSDYCIIT--PP  
 410 420 430 440 450 460

nk603cp4.pep GKGLGNASGAAVATHLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAK  
 M61904\_1 EKL--NVT--EIDTYDDHRMAMAF-SLAACADVPTIKDPGCTRKTFPNYFDVLQYQSKH  
 470 480 490 500 510

nk603cp4.pep IELSDTKAA

nk603cp4.pep  
 SWISSPROT:AROALYCES

ID AROALYCES STANDARD; PRT; 520 AA.  
 AC P10748;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE, CHLOROPLAST PRECURSOR . . .

SCORES Initl: 133 Initn: 243 Opt: 252 z-score: 276.0 E(): 2.1e-07  
 Smith-Waterman score: 403; 25.8% identity in 445 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSMFMGGLASGETRI  
 AROALYCES RVVRKSSFRISASVATAEKPHIEVLXPIKDISGTVKLPKSKLSNRILLALLSEGRITV  
 60 70 80 90 100 110

nk603cp4.pep TGLLEGEDVINTGKAMQAMGARIR--KEGDTWIIDGVGN---GGLLAPEAPLDFGNAAT  
 AROALYCES DNLLSSDDIHMYLGALKTLGLHVEDDNENQRAIVEGCGGQFPVGGKSEEEIQFLGNAGT  
 120 130 140 150 160 170

nk603cp4.pep GCRLTMGLGVYDFDSTFIGDA--SLTKRPMGRVLNPLREMGVQKSEDDGRLP---VTL  
 AROALYCES AMRPLTAAVTVAGGHSRYVLDGVPRMRERPIGDLVDGLKQLGAEDVDCFLGTNCPVPRIVS  
 180 190 200 210 220 230

nk603cp4.pep RG--PKTPTPIYRVPMASAVKSAVLLAG-LNTPGITT-VIEPIMTRDHT---KMLQG  
 AROALYCES KGGLPGGKVKLSGSI---SSQYLTALLMAAPLALGDVEIEIDKLISVPYVEMTLKLMER  
 240 250 260 270 280 290

nk603cp4.pep FGANPTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNLV  
 AROALYCES FGV--FVEHSSGWDRLFLVKGQKYKSPGKAF-VEGDASSASYFLAGAA-VTGGTVTVVEGCG  
 300 310 320 330 340 350

nk603cp4.pep LMNPTRTGLILT--LQEMGADIEVI-NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMI  
 AROALYCES GTSSLQGDVVKFAEVLEKMGAEVTVTENSVTVKGPFRNSGSMKH--LRAVDVNMNMKMP---D  
 360 370 380 390 400





nk603cp4.pep RGRPDGKGLGNASGAAVATHLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMA  
 A82498\_1 TP-PEKLNVTD-----IDTYDDHRMAMAF-SLAACADVPTINDPGCTRKTFPNYFDVLQ  
 460 470 480 490 500 510

450  
 nk603cp4.pep GLGAKIELSDTKAA

A82498\_1 QYSKH

nk603cp4.pep  
 GP\_BCT3:X72784\_1

LOCUS X72784\_1 [SSAROA]  
 DEFINITION Synechocystid sp. (PCC 6803) putative aroA gene.  
 DATE 19-JUL-1993  
 ACCESSION X72784  
 NID  
 ORGANISM Synechocystis sp. PCC 6803 . . .

SCORES Init1: 173 Initn: 214 Opt: 238 z-score: 272.4 E(): 3.4e-07  
 Smith-Waterman score: 238; 46.2% identity in 93 aa overlap

70 80 90 100 110 120  
 nk603cp4.pep QAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLV-GVYDFDSTFIGD  
 X72784\_1 LDAGNSGTMLRLMLGLLAGQKDCFLFTVTGD  
 10 20 30

130 140 150 160 170 180  
 nk603cp4.pep ASLTKRPMGRVLNPLREMGVQVKE-DGDRLPVTLRGPKTPTPTITYRVPMASAVQKSAVL  
 X72784\_1 DSLRHRPMSRVIQPLQMGAKIWARSNKGKAPLAVQGSQK-PIHYHSPIASAVQKSCLL  
 40 50 60 70 80

190 200 210 220 230 240  
 nk603cp4.pep LAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQVIDV  
 X72784\_1 LAG  
 90

nk603cp4.pep  
 GP\_PLN3:L18918\_1

LOCUS L18918\_1 [PMCROMX]  
 DEFINITION Pneumocystis carinii pentafunctional enzyme (arom) gene, complete  
 cds.  
 DATE 29-APR-1994  
 ACCESSION L18918  
 NID . . .

SCORES Init1: 56 Initn: 122 Opt: 251 z-score: 267.6 E(): 6.3e-07  
 Smith-Waterman score: 376; 24.3% identity in 449 aa overlap

10 20 30 40 50  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFPMFGGLASGETRITGL  
 L18918\_1 SSVSDDDIRTILSQNILLYGIPLNAFQKHTTITLPGSKSISNRALILASLSNGICYLKNF  
 380 390 400 410 420 430

60 70 80 90 100  
 nk603cp4.pep LEGEDVINTGKAMQAMGA---RIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTM  
 L18918\_1 LHSDDTYMYLSALEKLNAAEFKWEQDGVLVVKG-KSGYLENPQMELYLGNSGTTARFLT  
 440 450 460 470 480 490

110 120 130 140 150 160  
 nk603cp4.pep GLVGVYDFDST-----FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKT  
 L18918\_1 SICTLVQPNRSRENHLILTGSNRMKQRIPIGLVDALKNNGGCCIEYLELENCPLLIK-PKE  
 500 510 520 530 540 550

170 180 190 200 210  
 nk603cp4.pep PTPITYRVPM---SAQVKSALLAG-----LNTPGITTVIEPIMTRDHTKMLQGF  
 L18918\_1 IGLYGNINLSATVSSQYVSSILMCSFYAKTQVTLSLIGGKPISQPIY--DMTISMSSSF  
 560 570 580 590 600 610

220 230 240 250 260  
 nk603cp4.pep GANPTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNV--  
 L18918\_1 GIKVTRSHSKENTYYIP-KGCTYCPSEYI-IEGDATSATYPLAIAAITGGS-CTISNVGS  
 620 630 640 650 660 670

270 280 290 300 310 320  
 nk603cp4.pep --LMNPTRTGLILTLQEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAFSMID  
 L18918\_1 ASLQGDSEKSEYI-LKPMGCEV-VQSPTTTYIKGPPKGLKLS--LGSINME-----SMTD  
 680 690 700 710 720

330 340 350 360 370 380  
 nk603cp4.pep EYFILAFAA--APAEAG--ATVMNGLEELRVKESDRLSAVANGKLVDCDEGETSLVV  
 L18918\_1 TFLTAAVLASVAYEESKPYVTKITGISNRIKCNINAMVCELKKGIEAGELPDGIYV  
 730 740 750 760 770 780

390 400 410 420 430 440  
 nk603cp4.pep RGRPDGKGLGNASGAAVATHLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMA  
 L18918\_1 KALNTSNLPYSVEG--INCYNDRHRIAMSFVLAACISSKPTTILDKACVNKTWFIYWDILN  
 790 800 810 820 830 840

450  
 nk603cp4.pep GLGAKIELSDTKAA

L18918\_1 STFKVQMKGIEFDLNPNTINSSILHHPSECTIPLIGMRGAGKTTLGQLAANFLGREFIDLD  
 850 860 870 880 890 900

nk603cp4.pep  
 SWISSPROT:ARO1\_PNECA

ID ARO1\_PNECA STANDARD; PRT; 1581 AA.  
 AC Q12659;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE PENTAFUNCTIONAL AROM POLYPEPTIDE [INCLUDES: 3-DEHYDROQUINATE SYNTHASE . . .

SCORES Init1: 56 Initn: 122 Opt: 251 z-score: 267.6 E(): 6.3e-07  
 Smith-Waterman score: 376; 24.3% identity in 449 aa overlap

10 20 30 40 50  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFPMFGGLASGETRITGL  
 ARO1\_PNECA SSVSDDDIRTILSQNILLYGIPLNAFQKHTTITLPGSKSISNRALILASLSNGICYLKNF  
 380 390 400 410 420 430

60 70 80 90 100  
 nk603cp4.pep LEGEDVINTGKAMQAMGA---RIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTM  
 ARO1\_PNECA LHSDDTYMYLSALEKLNAAEFKWEQDGVLVVKG-KSGYLENPQMELYLGNSGTTARFLT  
 440 450 460 470 480 490

110 120 130 140 150 160  
 nk603cp4.pep GLVGVYDFDST-----FIGDASLTKRPMGRVLNPLREMGVQVKS-EDGDRLPVTLRGPKT



nk603cp4.pep 240 250 260 270 280  
 GKLT--GQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPT-RTGLILTLQEMGADI  
 AROA\_AERSA QSVISPGDFL-VEGDASSASY-FLAAGAIKKG-VRVTGIGKHSIGDIHFADVLERMGA--  
 240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340  
 EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAFAGATVMNGLE  
 AROA\_AERSA -----RITWGDDF--IEAQGPHGVDMNMNHPDVGHDS--GQSHCLPRVPHSQHLQ  
 290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400  
 ELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDH-RIA  
 AROA\_AERSA -LAVRD-DRCTPCTHGHRRRAQAGVSEEGTTFITRDAAD-----PAQARRDRLHQRISIA  
 340 350 360 370 380

nk603cp4.pep 410 420 430 440 450  
 MSFLVMGLVSENFPVTVDATMIATSFPEFMDLMAGLGAKEIELSDTKAA  
 AROA\_AERSA MCPSLVAL-SDIAVTINDPGCTSKTFPDYDKLASVSQAV  
 390 400 410 420

nk603cp4.pep  
 TREMBL\_MAIN:Q9AT37

ID Q9AT37 PRELIMINARY; PRT; 347 AA.  
 AC Q9AT37;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 102 Initn: 151 Opt: 237 z-score: 262.5 E(): 1.2e-06  
 Smith-Waterman score: 238; 26.6% identity in 354 aa overlap

nk603cp4.pep 60 70 80 90 100  
 EGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGLLA-----PEAPLDFGNAATGCR-  
 Q9AT37 PLSVEADKVAKRAVVVCGGRFPXKDAKEVKLFLGNAGTAMRP  
 10 20 30 40

nk603cp4.pep 110 120 130 140 150  
 LTMGLVGVYDFDSTFIGDA--SLTKRPMGRVLNPLREMGVQKSEGD-DRLPVTLRG---  
 Q9AT37 LTAAVVAAGG-NATYVLDGVPRMRERPIGDLVVGKQLGANVDCFLGTDCPPVRINGIGG  
 50 60 70 80 90 100

nk603cp4.pep 160 170 180 190 200 210  
 PKTPTPTITYRVPMASAVKSAVLLAG-LNTPGIIT-VIEPIMTRDHT---KMLQGFGA  
 Q9AT37 LPGAQVKSLSGSI---SSQYLSLLMAAPLALGDVEIEIDKLISVPYVEMTLRLMERFGV  
 110 120 130 140 150 160

nk603cp4.pep 220 230 240 250 260  
 NPTVETDADGVRTIRLEG--RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTI----L  
 Q9AT37 --TAE-HSDSWDRFYIKGGQKXKSPGNAY-VEGDASSASYFLAGAAITGGT-VTVQGCCT  
 170 180 190 200 210

nk603cp4.pep 270 280 290 300 310 320  
 NVLMNPTRTGLILTLQEMGADIEVINPRLA-GGEDVADLRVRSSTLKGVTVPEDRAPSMI  
 Q9AT37 TSLQGDVKFAEVLEM--MGAKVTWTDTSVTVTGPPRQPFGRKH--LKAVDNMMNKMP---  
 220 230 240 250 260

nk603cp4.pep 330 340 350 360 370 380  
 DEYPI LAVAAAFAGATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRP

Q9AT37 270 280 290 300 310 320  
 DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAICTELTKLGATVEEGPDYCIIT--P

nk603cp4.pep 390 400 410 420 430 440  
 DGKGLGNASGAAVATHLDHRIAMSLVMGLVSENFPVTVDATMIATSFPEFMDLMAGLGA  
 Q9AT37 PEKL--NVT--AIDTYDDHRMAMA  
 330 340

nk603cp4.pep  
 GP\_PLN2:AF349754\_1

LOCUS AF349754\_1 [AF349754]  
 DEFINITION Lolium rigidum 5-enolpyruvylshikimate 3-phosphate synthase (epsp-s)  
 mRNA, epsp-s-S allele, partial cds;  
 shikimate pathway enzyme.  
 DATE 18-MAR-2001  
 ACCESSION AF349754 . . .

SCORES Init1: 102 Initn: 151 Opt: 237 z-score: 262.5 E(): 1.2e-06  
 Smith-Waterman score: 238; 26.6% identity in 354 aa overlap

nk603cp4.pep 60 70 80 90 100  
 EGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGLLA-----PEAPLDFGNAATGCR-  
 AF349754\_1 PLSVEADKVAKRAVVVCGGRFPXKDAKEVKLFLGNAGTAMRP  
 10 20 30 40

nk603cp4.pep 110 120 130 140 150  
 LTMGLVGVYDFDSTFIGDA--SLTKRPMGRVLNPLREMGVQKSEGD-DRLPVTLRG---  
 AF349754\_1 LTAAVVAAGG-NATYVLDGVPRMRERPIGDLVVGKQLGANVDCFLGTDCPPVRINGIGG  
 50 60 70 80 90 100

nk603cp4.pep 160 170 180 190 200 210  
 PKTPTPTITYRVPMASAVKSAVLLAG-LNTPGIIT-VIEPIMTRDHT---KMLQGFGA  
 AF349754\_1 LPGAQVKSLSGSI---SSQYLSLLMAAPLALGDVEIEIDKLISVPYVEMTLRLMERFGV  
 110 120 130 140 150 160

nk603cp4.pep 220 230 240 250 260  
 NPTVETDADGVRTIRLEG--RGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTI----L  
 AF349754\_1 --TAE-HSDSWDRFYIKGGQKXKSPGNAY-VEGDASSASYFLAGAAITGGT-VTVQGCCT  
 170 180 190 200 210

nk603cp4.pep 270 280 290 300 310 320  
 NVLMNPTRTGLILTLQEMGADIEVINPRLA-GGEDVADLRVRSSTLKGVTVPEDRAPSMI  
 AF349754\_1 TSLQGDVKFAEVLEM--MGAKVTWTDTSVTVTGPPRQPFGRKH--LKAVDNMMNKMP---  
 220 230 240 250 260

nk603cp4.pep 330 340 350 360 370 380  
 DEYPI LAVAAAFAGATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRP  
 AF349754\_1 DVAMTLAVVALFADGPTAIRDVASWRVKETERMVAICTELTKLGATVEEGPDYCIIT--P  
 270 280 290 300 310 320

nk603cp4.pep 390 400 410 420 430 440  
 DGKGLGNASGAAVATHLDHRIAMSLVMGLVSENFPVTVDATMIATSFPEFMDLMAGLGA  
 AF349754\_1 PEKL--NVT--AIDTYDDHRMAMA  
 330 340

nk603cp4.pep  
 GP\_PLN2:AF360224\_1



LOCUS AF360224.1 [AF360224]  
DEFINITION Arabidopsis thaliana putative 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase (F27K7.11) mRNA, complete cds.  
DATE 22-MAR-2001  
ACCESSION AF360224  
NID . . .

SCORES Initl: 123 Initn: 239 Opt: 238 z-score: 260.9 E(): 1.5e-06  
Smith-Waterman score: 394; 24.2% identity in 475 aa overlap

```
nk603cp4.pep          9      10
                        MLHGASSRP-----ATARKSSGL-----
                        ||::|||
AF360224.1  RLSSPAVQISLHSQTRKNFRQSWGLKKSDDLMLNGSEIRPVKVRASVSTA EKASEIVLQPI
              30      40      50      60      70      80

nk603cp4.pep  ---SGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDVINTGKAMQAMGARIR--KE
              20      30      40      50      60      70
AF360224.1  REISGLIKLP GSKSLSNRILLALSEGTTVVDNLLNSDDIN YMLDALKILGLNVETHSE
              90     100     110     120     130     140

nk603cp4.pep  GDTWIIDGVNGGGLLAP-----EAPLDFGNAATGCRLTMGLVGVYDFSTFIGDA--SLT
              80      90      100     110     120
AF360224.1  NNRAVVEGCG-GVFPASIDSKSDIELYLGNA GTAMRPLTA AVTAAGGNASVYLDGVPRMR
              150     160     170     180     190     200

nk603cp4.pep  KRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTIRVPMA---SAQVKS AVLLA
              130     140     150     160     170     180
AF360224.1  ERPIGDLVVLGKQLGADVECTLTGNCPPVRVNANGGLP-GGKV KLSGSISSQYL TALLMA
              210     220     230     240     250     260

nk603cp4.pep  G-LNTPGITT-VIEPIMTRDHT E---KMLQGFGANPTVETDADGVRTIRLEGRGKLTGQV
              190     200     210     220     230
AF360224.1  APLALGDVEIEIVDKLISVPYVEMTLKLMERFGVSAEHS ESWDRF-FVKGGQKYKSPGNA
              270     280     290     300     310     320

nk603cp4.pep  IDVPGDPSSTAFPLVAALLVPGSDVTILNVL MNPTRTGLILT--LQEMGADIE-VINPRL
              240     250     260     270     280     290
AF360224.1  Y-VEGDASSASYFLAGAA-ITGETVTVEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVT
              330     340     350     360     370     380

nk603cp4.pep  AGGEDVADLRVRSSSTLKGVTVPEDRAPSMIDEYPI LAVAAAF AEGATVMNGLEELRVKES
              300     310     320     330     340     350
AF360224.1  VTGPSRDAFGMRH--LRAIDVNMNKMP---DVAMT LAVVALFADGPTTIRDVASWRVKET
              390     400     410     420     430

nk603cp4.pep  DRLSAVANGLKLVDCDEGETSLVVRGRPDGKGLGNASGA AVATHLDHRIAMSFVLMGL
              360     370     380     390     400     410
AF360224.1  ERMAICTELRKLGATVEEGSDYCVIT--PPKK-----VKPAE IDTYDDHRMAMAF-SLAA
              440     450     460     470     480     490

nk603cp4.pep  VSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
              420     430     440     450
AF360224.1  CADVPITINDPGCTRKTFPDYFQVLERITKH
              500     510     520
```

nk603cp4.pep  
TREMBL\_NEW:AAK64123

ID AAK64123 PRELIMINARY; PRT; 521 AA.  
AC AAK64123;

DT 06-SEP-2001 (EMBLrel. 63, Created)  
DT 06-SEP-2001 (EMBLrel. 63, Last sequence update)  
DT 06-SEP-2001 (EMBLrel. 63, Last annotation update)  
DE PUTATIVE 5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE (EPSP) SYNTHASE. . . .

SCORES Initl: 123 Initn: 239 Opt: 238 z-score: 260.9 E(): 1.5e-06  
Smith-Waterman score: 394; 24.2% identity in 475 aa overlap

```
nk603cp4.pep          9      10
                        MLHGASSRP-----ATARKSSGL-----
                        ||::|||
AAK64123  RLSSPAVQISLHSQTRKNFRQSWGLKKSDDLMLNGSEIRPVKVRASVSTA EKASEIVLQPI
              30      40      50      60      70      80

nk603cp4.pep  ---SGTVRI PGDKSISHRSFMFGGLASGETRITGLLEGEDVINTGKAMQAMGARIR--KE
              20      30      40      50      60      70
AAK64123  REISGLIKLP GSKSLSNRILLALSEGTTVVDNLLNSDDIN YMLDALKILGLNVETHSE
              90     100     110     120     130     140

nk603cp4.pep  GDTWIIDGVNGGGLLAP-----EAPLDFGNAATGCRLTMGLVGVYDFSTFIGDA--SLT
              80      90      100     110     120
AAK64123  NNRAVVEGCG-GVFPASIDSKSDIELYLGNA GTAMRPLTA AVTAAGGNASVYLDGVPRMR
              150     160     170     180     190     200

nk603cp4.pep  KRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTIRVPMA---SAQVKS AVLLA
              130     140     150     160     170     180
AAK64123  ERPIGDLVVLGKQLGADVECTLTGNCPPVRVNANGGLP-GGKV KLSGSISSQYL TALLMA
              210     220     230     240     250     260

nk603cp4.pep  G-LNTPGITT-VIEPIMTRDHT E---KMLQGFGANPTVETDADGVRTIRLEGRGKLTGQV
              190     200     210     220     230
AAK64123  APLALGDVEIEIVDKLISVPYVEMTLKLMERFGVSAEHS ESWDRF-FVKGGQKYKSPGNA
              270     280     290     300     310     320

nk603cp4.pep  IDVPGDPSSTAFPLVAALLVPGSDVTILNVL MNPTRTGLILT--LQEMGADIE-VINPRL
              240     250     260     270     280     290
AAK64123  Y-VEGDASSASYFLAGAA-ITGETVTVEGCGTTS LQGDVKFAEVLEKMGCKVSWTENSVT
              330     340     350     360     370     380

nk603cp4.pep  AGGEDVADLRVRSSSTLKGVTVPEDRAPSMIDEYPI LAVAAAF AEGATVMNGLEELRVKES
              300     310     320     330     340     350
AAK64123  VTGPSRDAFGMRH--LRAIDVNMNKMP---DVAMT LAVVALFADGPTTIRDVASWRVKET
              390     400     410     420     430

nk603cp4.pep  DRLSAVANGLKLVDCDEGETSLVVRGRPDGKGLGNASGA AVATHLDHRIAMSFVLMGL
              360     370     380     390     400     410
AAK64123  ERMAICTELRKLGATVEEGSDYCVIT--PPKK-----VKPAE IDTYDDHRMAMAF-SLAA
              440     450     460     470     480     490

nk603cp4.pep  VSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
              420     430     440     450
AAK64123  CADVPITINDPGCTRKTFPDYFQVLERITKH
              500     510     520
```

nk603cp4.pep  
TREMBL\_MAIN:Q9FVP6

ID Q9FVP6 PRELIMINARY; PRT; 521 AA.  
AC Q9FVP6;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

SCORES Init1: 123 Initn: 239 Opt: 238 z-score: 260.9 E(): 1.5e-06  
Smith-Waterman score: 394; 24.2% identity in 475 aa overlap

nk603cp4.pep VSENPVTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA  
::|:|::| :||:::  
Q9FVP6 CADVPITINDPGCTRKTFPDYFQVLERITKH  
500 510 520

LOCUS AC084414\_12 [AC084414]  
DEFINITION Arabidopsis thaliana chromosome 1 BAC F27K7 genomic sequence,  
complete sequence.  
DATE 04-NOV-2000  
ACCESSION AC084414  
NID

```

          420          430          440          450
nk603cp4.pep VSENFVTVDATMIATSFPEPFMDLMAGLGAKEISDTKAA
          ::||::||::||::||::||::||::||::||::||::||
AC084414_12 CADVPITINDPGCTRKTFPDYQVQLERITKH
          500          510          520

```

```

LOCUS      AC084242.2 [AC084242]
DEFINITION Arabidopsis thaliana chromosome 1 BAC T24P22 genomic sequence,
            complete sequence.
DATE       19-JAN-2001
ACCESSION  AC084242
NTD

```

SCORES Initl: 123 Initn: 239 Opt: 238 z-score: 260.9 E(): 1.5e-06  
Smith-Waterman score: 394; 24.2% identity in 475 aa overlap

```
nk603cp4.pep      9      10
                  MLHGASSRP-----ATARKSSGL-----
AC084242_2      30      40      50      60      70      80
RLSSPAVQISLHSQTRKNFRQSWGLKKSDMLNGSEIRPVKVRASVSTA EKASEIVLQPI

nk603cp4.pep      20      30      40      50      60      70
---SGTVRIIPGDKSISHRSFMPGGLASGETRITGLLEGEDVINTGKAMQAMGARIR--KE
AC084242_2      90      100      110      120      130      140
REISGLIKLPGSKSLSNRILLALSEGTTVVNDLLNSDDINMLDALIKILGLNVETHSE

nk603cp4.pep      80      90      100      110      120
GDTWIIDGVGNGLLAP-----EAPLDFGNAATGCRLTMGLVGVYDFDSTFIGDA--SLT
AC084242_2      150      160      170      180      190      200
NNRAVVEGCG--GVFPASIDSKSDIELYLGNAAGTAMRPLTA AVTAAGGNASYVLDGVPRMR

nk603cp4.pep      130      140      150      160      170      180
KRPMGRVLNPLREMGVQVKSSEDGDRPLVTLRGPKTPTPTITRYVPM---SAQVKSALLA
AC084242_2      210      220      230      240      250      260
ERPIGDLVVGKQLGADVECTLTNCPPVRVANGGLF--GGKVLKSGSISSQYLTALLMA

nk603cp4.pep      190      200      210      220      230
G-LNTPGITT-VIEPIIMTRDHT---KMLQGFGANPTVETDADGVRITRLEGRGKLTQGV
AC084242_2      270      280      290      300      310      320
APLALGDVEIEIVDKLISVPYVEMTLKMERFGVSAEHSSEWDRF--FVKGGQKYKSPGNA

nk603cp4.pep      240      250      260      270      280      290
IDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILT--LQEMGADIE-VINPRL
AC084242_2      330      340      350      360      370      380
Y-VEGDASSASYFLAGAA--ITGETVTVEGCGTTSLOGDVKFAEVLEKMGCKVSWTENSVT

nk603cp4.pep      300      310      320      330      340      350
AGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAFAEGATVMNGLEELRVKES
AC084242_2      390      400      410      420      430
VTGPSRDAPGMRH--LRAIDVNMNKMFP---DVAMTLAVVALFADGPTTIRDVASWRVKET

nk603cp4.pep      360      370      380      390      400      410
DRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGA AVATHLDHRIAMSFLVMGL
AC084242_2      440      450      460      470      480      490
ERMIAICTELRLKLGATVEEGSDYCVIT--PPKK---VKFAEIDTYDDHRMAMAF-SLAA

nk603cp4.pep      420      430      440      450
VSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AC084242_2      500      510      520
CADVPITINDPGCTRKTFPDYFQVLERITKH
```

nk603cp4.pep  
GP\_PAT1:A59345\_1

LOCUS A59345\_1 [A59345]  
DEFINITION Sequence 2 from Patent WO9704114;  
unnamed protein product.  
DATE 06-MAR-1998  
ACCESSION A59345  
NID . . .

SCORES Initl: 136 Initn: 300 Opt: 237 z-score: 260.9 E(): 1.5e-06  
Smith-Waterman score: 405; 26.4% identity in 447 aa overlap

```
nk603cp4.pep      10      20      30      40      50      60
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG
A59345_1      10      20      30      40      50
AGAEIEIVLQPIKEISGTVKLPKSGKSLSNRILLALSEGTTVVNDLLNSSEIVHYML

nk603cp4.pep      70      80      90      100      110
KAMQAMGARIRKEGDTWIIDGVGNGLLA-----PEAPLDFGNAATGCRLTMGLVGVYDF
A59345_1      60      70      80      90      100      110
GALRTLGLSVBADKAAKRAVVVCGGKFPVEDAKEEVQLFLGNAGTAMRPLTA AVTAAGG

nk603cp4.pep      120      130      140      150      160
DSTFIGDA--SLTKRPMGRVLNPLREMGVQVKSSEDG--DRLPVTLRG---PKTPTITRY
A59345_1      120      130      140      150      160      170
NATYVLDGVPRMRERPIGDLVVGKQLGADVDCFLGTDCCPPVRVNGIGGLPGGKVLKSGS

nk603cp4.pep      170      180      190      200      210      220
VPMASQVKSALLAG--LNTPGITT-VIEPIIMTRDHT---KMLQGFGANPTVETDADGV
A59345_1      180      190      200      210      220      230
I---SSQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAE---HSDSW

nk603cp4.pep      230      240      250      260      270
RTIRLEGRGKL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTI---LNVLMNPTRTGLI
A59345_1      240      250      260      270      280
DRFYIKGGQKYKSPKNAYVEGDASSASYFLAGAAITGGT--VTVEGCGTTSLOGDVKFAEV

nk603cp4.pep      280      290      300      310      320      330
LTLQEMGADIEVINPRLA--GGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAF
A59345_1      290      300      310      320      330      340
LEM--MGAKVITWETSIVTVGPPREPFGGRKH--LKAIDVNMNKMFP---DVAMTLAVVALF

nk603cp4.pep      340      350      360      370      380      390
AEGATVMNGLEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAA
A59345_1      350      360      370      380      390
ADGFTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEKL--NVT--A

nk603cp4.pep      400      410      420      430      440      450
VATHLDHRIAMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
A59345_1      400      410      420      430      440
IDTYDDHRMAMAF--SLAACAEVPTVIRDPGCTRKTFPDYFQVLERITKH
```

nk603cp4.pep  
GP\_PLN4:X63374\_1

LOCUS X63374\_1 [ZMEPSPS]  
DEFINITION 3-mays mRNA for EPSP-synthase;  
3-phosphoshikimate 1-carboxyvinyltransferase.  
DATE 04-SEP-1996  
ACCESSION X63374  
NID . . .

SCORES Initl: 136 Initn: 300 Opt: 237 z-score: 260.9 E(): 1.5e-06  
Smith-Waterman score: 405; 26.4% identity in 447 aa overlap

```
nk603cp4.pep      10      20      30      40      50      60
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG
X63374_1      10      20      30      40      50
AGAEIEIVLQPIKEISGTVKLPKSGKSLSNRILLALSEGTTVVNDLLNSSEIVHYML

nk603cp4.pep      70      80      90      100      110
```

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLA-----PEAPLDFGNAATGCRLTMGLVGVYDF  
 X63374\_1 GALRTLGLSVEADKAARAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGG  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160  
 DSTFIGDA--SLTKRPMGRVLNPLREMGVQVKSSEDG--DRLPVTLRG---PKTPTPTITYR  
 X63374\_1 NATYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPFVRVNGIGGLPGGKVKLSGS  
 120 130 140 150 160 170

nk603cp4.pep 170 180 190 200 210 220  
 VPMASAVKSAVLLAG--LNTPGITT--VIEPIMTRDHTD---KMLQGFGANPTVETDADGV  
 X63374\_1 I---SSQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAE---HSDSW  
 180 190 200 210 220 230

nk603cp4.pep 230 240 250 260 270  
 RTIRLEGRGKL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTI---LNVLMNPTRTGLI  
 X63374\_1 DRFYIKGGQYKSPKNAYVEGDASSASYFLAGAAITGGT--VTVEGCGTTSIQGDVVKFAEV  
 240 250 260 270 280

nk603cp4.pep 280 290 300 310 320 330  
 LTLQEMGADIEVINPRLA--GGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAF  
 X63374\_1 LEM--MGAKVTWTETSVITVTPGPPREPFGGRKH--LKAIDVNMNMKMP---DVAMTLAVVALF  
 290 300 310 320 330 340

nk603cp4.pep 340 350 360 370 380 390  
 AEGATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGPDGKGLGNASGAA  
 X63374\_1 ADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEKL--NVT--A  
 350 360 370 380 390

nk603cp4.pep 400 410 420 430 440 450  
 VATHLDHRIAMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKEIELSDTKAA  
 X63374\_1 IDTYDDHRMAMAF--SLAACAEVFTVIRDPGCTRTKTFPDYFDVLSTFVKV  
 400 410 420 430 440

nk603cp4.pep  
 TREMBL\_MAIN:024566

ID 024566 PRELIMINARY; PRT; 444 AA.  
 AC 024566;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5- . . .

SCORES Init1: 136 Initn: 300 Opt: 237 z-score: 260.9 E(): 1.5e-06  
 Smith-Waterman score: 405; 26.4% identity in 447 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPIGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG  
 024566 AGAEEIVLQPIKEISGTIVKLPKSGKSLNRILLALLAALSEGTTVDNLLNSDEVHYML  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRKEGDTWIIDGVNGGGLA-----PEAPLDFGNAATGCRLTMGLVGVYDF  
 024566 GALRTLGLSVEADKAARAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGG  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160  
 DSTFIGDA--SLTKRPMGRVLNPLREMGVQVKSSEDG--DRLPVTLRG---PKTPTPTITYR  
 024566 NATYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPFVRVNGIGGLPGGKVKLSGS  
 120 130 140 150 160 170

nk603cp4.pep 170 180 190 200 210 220  
 VPMASAVKSAVLLAG--LNTPGITT--VIEPIMTRDHTD---KMLQGFGANPTVETDADGV  
 024566 I---SSQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAE---HSDSW  
 180 190 200 210 220 230

024566 NATYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPFVRVNGIGGLPGGKVKLSGS  
 120 130 140 150 160 170

nk603cp4.pep 170 180 190 200 210 220  
 VPMASAVKSAVLLAG--LNTPGITT--VIEPIMTRDHTD---KMLQGFGANPTVETDADGV  
 024566 I---SSQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAE---HSDSW  
 180 190 200 210 220 230

nk603cp4.pep 230 240 250 260 270  
 RTIRLEGRGKL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTI---LNVLMNPTRTGLI  
 024566 DRFYIKGGQYKSPKNAYVEGDASSASYFLAGAAITGGT--VTVEGCGTTSIQGDVVKFAEV  
 240 250 260 270 280

nk603cp4.pep 280 290 300 310 320 330  
 LTLQEMGADIEVINPRLA--GGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAF  
 024566 LEM--MGAKVTWTETSVITVTPGPPREPFGGRKH--LKAIDVNMNMKMP---DVAMTLAVVALF  
 290 300 310 320 330 340

nk603cp4.pep 340 350 360 370 380 390  
 AEGATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGPDGKGLGNASGAA  
 024566 ADGPTAIRDVASWRVKETERMVAIRTELTKLGASVEEGPDYCIIT--PPEKL--NVT--A  
 350 360 370 380 390

nk603cp4.pep 400 410 420 430 440 450  
 VATHLDHRIAMSFLVMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKEIELSDTKAA  
 024566 IDTYDDHRMAMAF--SLAACAEVFTVIRDPGCTRTKTFPDYFDVLSTFVKV  
 400 410 420 430 440

nk603cp4.pep  
 GP\_PAT1:A59404\_1

LOCUS A59404\_1 [A59404]  
 DEFINITION Sequence 2 from Patent WO9704103;  
 unnamed protein product.  
 DATE 06-MAR-1998  
 ACCESSION A59404  
 NID . . .

SCORES Init1: 136 Initn: 300 Opt: 237 z-score: 260.9 E(): 1.5e-06  
 Smith-Waterman score: 405; 26.4% identity in 447 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPIGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG  
 A59404\_1 AGAEEIVLQPIKEISGTIVKLPKSGKSLNRILLALLAALSEGTTVDNLLNSDEVHYML  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRKEGDTWIIDGVNGGGLA-----PEAPLDFGNAATGCRLTMGLVGVYDF  
 A59404\_1 GALRTLGLSVEADKAARAVVVGCGGKFPVEDAKEEVQLFLGNAGTAMRPLTAAVTAAGG  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160  
 DSTFIGDA--SLTKRPMGRVLNPLREMGVQVKSSEDG--DRLPVTLRG---PKTPTPTITYR  
 A59404\_1 NATYVLDGVPRMRERPIGDLVVGLKQLGADVDCFLGTDCPFVRVNGIGGLPGGKVKLSGS  
 120 130 140 150 160 170

nk603cp4.pep 170 180 190 200 210 220  
 VPMASAVKSAVLLAG--LNTPGITT--VIEPIMTRDHTD---KMLQGFGANPTVETDADGV  
 A59404\_1 I---SSQYLSALLMAAPLALGDVEIEIIDKLISIPYVEMTLRLMERFGVKAE---HSDSW  
 180 190 200 210 220 230

280            290            300            310            320            330

nk603cp4.pep EYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPD



nk603cp4.pep  
GP\_PAT2:AX064129\_1

LOCUS AX064129\_1 [AX064129]  
DEFINITION Sequence 411 from Patent WO0100843;  
RXA02458.  
DATE 24-JAN-2001  
ACCESSION AX064129  
NID . . .

SCORES Init1: 97 Initn: 226 Opt: 233 z-score: 256.8 E(): 2.5e-06  
Smith-Waterman score: 379; 25.7% identity in 443 aa overlap

nk603cp4.pep 10 20 30 40 50  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMPGGLASGETRITGLLEGED  
AX064129\_1 MVFVSDSSISLPIWDAPRARGPIVSDLAIPGSKSITNRLILAAALASTPSTIIDVLRSD  
10 20 30 40 50 60

nk603cp4.pep 60 70 80 90 100 110  
VINTGKAMQAMGARIRKEG-DTWIIDGVNGGGLLAPEALDFGNAATGCRLTMGLVGVYD  
AX064129\_1 TDLMTDGLRSLGITITEEAVDRYRVEP---GQLSA--GSVECGLAGTVMRFLPPVAAFAD  
70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
FDSTFFIGDASLTTRKPMGRVLNPLREMGVQVKSSEDGRLFTLRGPKTPTPTITVRVMA-S  
AX064129\_1 GPVHFDGDPQARVRPMTSILDALRSLGVEV---DNNNLPFTVNAGEVPEGGVVEIDASGS  
120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 230  
AQVKSALLAGLNTPGITTVIEPIIMTRDHTKMLQG-FGANPTVETDADGVRTIRLEGR-  
AX064129\_1 SQFVSGLLLSA-----PRFKNGVTVKHVGGRLPSPHIEMTVDMLRSAGIEIEE  
180 190 200 210 220

nk603cp4.pep 240 250 260 270 280  
-----GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILT--  
AX064129\_1 SENQWVHPGEILGRWRIEPLDLSN-ATPFLAAAVTGGTIKINHWPIKTTQPGDAIRSI  
230 240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 339  
LQEMGADIEVINPRLAGGEDVADLRVRSST-LKGVTVPEDRAPSMIDEYPILAVAAFAE  
AX064129\_1 LERMGCVELV---AQGEGY-DLSVTGPVALKGIEIDMSDIGELT---PTVAALALAS  
290 300 310 320 330

nk603cp4.pep 340 350 360 370 380 390 399  
GATVMNGLEELRVKESDRLSAVANGLKLVGDCDEGETSLVVGRPDGKGLGNASGAAVA  
AX064129\_1 TESRLTGIAHLRGHETDRLAALTAENKLGKCTELKDGILLIE--P-----ASLHGGVWH  
340 350 360 370 380

nk603cp4.pep 400 410 420 430 440 450  
THLDHRIAMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AX064129\_1 SYADHRMATAGAIIGLAVDG-VQVEDIKTTSKTFPGFENVWEEMVG  
390 400 410 420 430

nk603cp4.pep  
GP\_BCT2:AF114233\_1

LOCUS AF114233\_1 [AF114233]  
DEFINITION Corynebacterium glutamicum 5-enolpyruvylshikimate 3-phosphate  
synthase (aroA) gene, complete cds;  
EPSP synthase.  
DATE 07-FEB-1999

ACCESSION AF114233 . . .

SCORES Init1: 97 Initn: 226 Opt: 233 z-score: 255.8 E(): 2.9e-06  
Smith-Waterman score: 379; 25.7% identity in 444 aa overlap

nk603cp4.pep 10 20 30 40  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMPGGLA  
AF114233\_1 SPVVNKTNLPHITSIMVVFVSDSSISLPIWDAPRARPVSDLAIPGSKSITNRLILAAAL  
60 70 80 90 100 110

nk603cp4.pep 50 60 70 80 90 100  
SGETRITGLLEGEDVINTGKAMQAMGARIRKEG-DTWIIDGVNGGGLLAPEALDFGNA  
AF114233\_1 STPSTIIDVLRSDTDLMTDGLRSLGITITEEAVDRYRVEP---GQLSA--GSVECGLAG  
120 130 140 150 160

nk603cp4.pep 110 120 130 140 150 160  
TGCRLTMGLVGVYDFDSTFFIGDASLTTRKPMGRVLNPLREMGVQVKSSEDGRLFTLRGPK  
AF114233\_1 TVMRFLPPVAAFADGPFVHFDGDPQARVRPMTSILDALRSLGVEV---DNNNLPFTVNAGE  
170 180 190 200 210 220

nk603cp4.pep 170 180 190 200 210  
TPTPTITVRVMA-SAQVKSALLAGLNTPGITTVIEPIIMTRDHTKMLQG-FGANPTVET  
AF114233\_1 VPEGGVVEIDASGSSQFVSGLLLSA-----PRFKNGVTVKHVGGRLPSPHIE  
230 240 250 260 270

nk603cp4.pep 220 230 240 250 260  
DADGVRTIRLEGR-----GKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNV  
AF114233\_1 TVDMLRSAGIEIEESENQWVHPGEILGRWRIEPLDLSN-ATPFLAAAVTGGTIKINHW  
280 290 300 310 320 330

nk603cp4.pep 270 280 290 300 310 320  
LMNPTRTGLILT--LQEMGADIEVINPRLAGGEDVADLRVRSST-LKGVTVPEDRAPSMI  
AF114233\_1 PIKTTQPGDAIRSI LERMGCVELV---AQGEGY-DLSVTGPVALKGIEIDMSDIGELT  
340 350 360 370 380

nk603cp4.pep 330 340 350 360 370 380  
DEYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGLKLVGDCDEGETSLVVGRPD  
AF114233\_1 ---PTVAALALASTESRLTGIAHLRGHETDRLAALTAENKLGKCTELKDGILLIE--P  
390 400 410 420 430 440

nk603cp4.pep 390 400 410 420 430 440  
DGKGLGNASGAAVATHLDHRIAMSFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGA  
AF114233\_1 ----ASLHGGVWHVSYADHRMATAGAIIGLAVDG-VQVEDIKTTSKTFPGFENVWEEMVG  
450 460 470 480 490

nk603cp4.pep 450  
KIELSDTKAA

nk603cp4.pep  
GP\_PAT1:A59347\_1

LOCUS A59347\_1 [A59347]  
DEFINITION Sequence 4 from Patent WO9704114;  
unnamed protein product.  
DATE 06-MAR-1998  
ACCESSION A59347  
NID . . .

SCORES Init1: 136 Initn: 276 Opt: 232 z-score: 255.5 E(): 3e-06  
Smith-Waterman score: 400; 26.2% identity in 447 aa overlap







[illegible]

```

                                10           20           30           40           50          60
nk603cp4.pep MLHGASSRPAATARKSSGLSGTVTRIPGDKSISHRFSFMFGLASGETRITGLEGEDVINTQ
                | : || |:| |:| ::::|| | ||:||:::|
A18838_1       MNSLRLEPISRVAAGEVNLPKGSKSVSNGALLLAALARGTTRLTNLLSDDIRHMLD
                        10         20        30        40        50


                                70           80           90           100          110
nk603cp4.pep KAMQAMGARIRKEGDWTWIIDGVGNGLLAPAEPLDF--GNAATGCRLTMGLVGVDYFDSTI
                | : : : : : | | | | | : | | : | : | : : : : 
A18838_1       AALTQLGVKYLSADKTCTEVTGHGRSFSAVASPVNLFNLGAGTAGMRPLCAALCLGSGSEYM
                       60      70      80      90     100     110


                   120          130          140          150          160          170
nk603cp4.pep FIGDASLTKRPMGRVLNPLERMGVQVKSEDGDRFPVTLRGPKTPPTITYRVPMASAOVRM
                 : | : : : | : | : | : : : : : : : | : : | : : | : : | : : |
A18838_1       LGGEPRMEERIPIHLVDCALKAKIAHIQYLKKDGYPPLVVDAKLWGGDVHDVDSVSSQFLI
                      120    130    140    150    160    170


                   180          190          200          210          220          230
nk603cp4.pep SAVLLAGLNTPPGITIVIEPI-----MTRDHTEKMLQGCGANPTVETDADGVRTIRLEGFI
                 : | : : | : | : | : | : | : | : | : : : : | : : : : | : : | : : |
A18838_1       TAFDMAA---PAMAPVIPRIHKIGELSVSKPYIDITLHMNSSGGVV-IEHDNYKLYFYIKGN
                     180    190    200    210    220    230


                           240          250          260          270          280
nk603cp4.pep GKLT--GQVIDVPFGDPSSATAFFLVAAILLVPSDVDTIILNVLMNPTRTGILILTQEIMGADI
                  :: : | : v | | | : : : : | : : : : : | : : : : | : : : : | : : | : : |
A18838_1       QSIVSPGDFLI-VEGDASSASY-FLAAGAIGKG-VRVGTGIKHSIGDIHFADVLERMGAA--
                         240    250    260    270    280

```

330                    340                    350                    360                    370                    380

nk603cp4.pep DEYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRF

NK603 analysis  
Page 86 of 120



SCORES Init1: 136 Initn: 276 Opt: 221 z-score: 243.6 E(): 1.4e-05  
Smith-Waterman score: 403; 26.1% identity in 448 aa overlap

```
nk603cp4.pep 10 20 30 40 50 60
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
CAD01095 AGAEVVLQPIKEISGVVVKLPKSKSLNRIILLSSALAEGTTVDNLLNSEDVHYML
```

```
nk603cp4.pep 70 80 90 100 110
KAMQAMGARIRKEGDTWIIDGVNGGGLLA-----PEAPLDFGNAATGCRLTMGLVGVYD
CAD01095 GALKTLGLSVEADKAARAVVVGCGGKFPVEKDAKEEVQLFLGNAGTAMRSLTAAVTAAG
```

```
nk603cp4.pep 120 130 140 150 160
FDSTFIGDA--SLTKRPMGRVLNPLREMGVQKSEDD--RLPVTLRG--PKTPTPTITY
CAD01095 GNATYVLDGVPFRMRERPIGDLVVLKQLGADVDCFLGTDCPPVRVKKGIGGLPGGKVKLSG
```

```
nk603cp4.pep 170 180 190 200 210 220
RVPMSAQVKSALLAG--LNTPGITT--VIEPIIMTRDHTD---KMLQGFANPTVETDADG
CAD01095 SI---SSQYLSALLMAAPLALGDVIEIIDKLISIPVVMERFVGVKAE---HSDS
```

```
nk603cp4.pep 230 240 250 260 270
VRTIRLEGRGKL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTI---LNVLMNPTRTGL
CAD01095 WDRFYIKGGQYKSPKNAYVEGDASSASYFLAGAAITGGT--VTVEGCGTTSLQGDVKFAE
```

```
nk603cp4.pep 280 290 300 310 320 330
ILTLQEMGADIEVINPRLA--GGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAA
CAD01095 VLEM--MGAKVTWTETSVTVTPQREPFGKRH--LKAIDVNMNMKP--DVAMTLAVVAL
```

```
nk603cp4.pep 340 350 360 370 380 390
FAEGATVMNGLEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVGRPDGKGLGNASGA
CAD01095 FADGPTARDVASWRVKETERMVAIRTELTKLGASVEEGLDYCIIT--PPEKL--NVT--
```

```
nk603cp4.pep 400 410 420 430 440 450
AVATHLDHRIAMSLVMGLVSENFVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
CAD01095 AIDTYDDHRMAMAF--SLAACADVFTVIRDPGCTRKTPFDYFDVLSFVKV
```

nk603cp4.pep  
TREMBL\_MAIN:Q9K9D5

ID Q9K9D5 PRELIMINARY; PRT; 447 AA.  
AC Q9K9D5;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . . .

SCORES Init1: 115 Initn: 242 Opt: 216 z-score: 238.2 E(): 2.7e-05  
Smith-Waterman score: 444; 26.4% identity in 440 aa overlap

```
nk603cp4.pep 10 20 30 40
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRIT
Q9K9D5 MTRFDENARSPWTPHLDVKTVELFPLNQRLDGSITLPGSKSLTNRALIISALANSDSMLT
```

```
10 20 30 40 50 60
nk603cp4.pep 50 60 70 80 90 100
GLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMG
Q9K9D5 GMLKSDDTYWCIQALKRLGVQINVQGETTSIRGIGGQW---KSSSLYIGAAGTLARFLLG
```

```
nk603cp4.pep 110 120 130 140 150 160
LVGVYDFDSTFI--GDASLTGRPMGRVLNPLREMGVQVQK--SEDGDRLPVTLRGPKTPTPI
Q9K9D5 ALAISRSNGWEIASQSMKRPVIEPLVGVRLGATIHLYLRREG--FYPLSIHNGLAGGT
```

```
nk603cp4.pep 170 180 190 200 210 220
TYRVPMSAQVKSALLAG--LNTPGITT--VIEPIIMTRDHTD---TEKMLQGFANPTVETDA
Q9K9D5 VRLSGQMSSQYISGLLIAAPYADTPVTITVQGSIVQHAYVFTLHLMKSFGAQ--VEYDQ
```

```
nk603cp4.pep 230 240 250 260 270 279
DGVRTIRLEGRGKL--TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLMNPTRTGL--IL
Q9K9D5 Q-LQTVVHPT--PYTCQDIDLEADASTACYFLAALTKGR--IRLNNLTASTTQPDHLML
```

```
nk603cp4.pep 280 290 300 310 320 330
TL--QEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAFA
Q9K9D5 TVFEKMGCTV-----TRGSSFIELEGVSQKGGFQISMNE---MSDQALTLAAIAPFA
```

```
nk603cp4.pep 340 350 360 370 380 390
EGATVMNGLEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVGRPDGKGLGNASGA
Q9K9D5 DGPITITDVEHIRYHESDRIAVICEALTRLGIQVDEFEDGLTVY--P---GTPKPTLHPL
```

```
nk603cp4.pep 400 410 420 430 440 450
ATHLDHRIAMSLVMGLVSENFVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Q9K9D5 STYDDHRVAMSLSLIGTKVKG--LRINDPGCVAKTCPSYFQLLEQLGIQVHYQ
```

nk603cp4.pep  
GP\_BCT2:AP001516\_142

LOCUS AP001516\_142 [AP001516]  
DEFINITION Bacillus halodurans genomic DNA, section 10/14;  
DATE 10-JAN-2001  
ACCESSION AP001516  
NID  
ORGANISM Bacillus halodurans . . .

SCORES Init1: 115 Initn: 242 Opt: 216 z-score: 238.2 E(): 2.7e-05  
Smith-Waterman score: 444; 26.4% identity in 440 aa overlap

```
nk603cp4.pep 10 20 30 40
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRIT
AP001516_142 MTRFDENARSPWTPHLDVKTVELFPLNQRLDGSITLPGSKSLTNRALIISALANSDSMLT
```

```
nk603cp4.pep 50 60 70 80 90 100
GLLEGEDVINTGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMG
AP001516_142 GMLKSDDTYWCIQALKRLGVQINVQGETTSIRGIGGQW---KSSSLYIGAAGTLARFLLG
```

nk603cp4.pep LVGVYDFDSTFI--GDASLTKRPMGRVNLPLREMGVQVK--SEGDRLPVTLRGPKTPTPI  
 AP001516\_142 ALAISRSNGNWEIEASQSMKRPIELVGVLELREGATIHVLRREG-FYPLSIHNGLAGGT  
 nk603cp4.pep TYRVFMASAVKSAVLLAG--LNTPGITTVIEPIIMTRDH---TEKMLQGGFANPTVETDA  
 AP001516\_142 VRLSGQMSSQYISGLLIAAPYADTFVTTITVQGSIVQHAYVFLTLHLKMSFGAQ--VEYDQ  
 nk603cp4.pep DGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGL-IL  
 AP001516\_142 Q-LQTIIVVHT-PYTCQDIDLEADASTACYFLALAAALTKGR-IRLNNLTASTTQPDHLML  
 nk603cp4.pep TL-QEMGADIEVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFA  
 AP001516\_142 TVFEKMGCTV-----TRGSSFIELEGVSQKGGFQISMNE---MSDQALTLAAIAPFA  
 nk603cp4.pep EGATVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVVRGPRDGKGLGNASGAAV  
 AP001516\_142 DGPITITDVEHIRVHESDRIAVICEALTRLGIQVDEFEDGLTVY--P---GTPKPTLHPL  
 nk603cp4.pep ATHLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEISDTKAA  
 AP001516\_142 STYDDHVRVAMSLSLIGTKVKG-LRLNDPGCAKTCFSYFQLEQLGIQVHYQ

nk603cp4.pep  
 TREMBL\_MAIN:Q9YC47

ID Q9YC47 PRELIMINARY; PRT; 427 AA.  
 AC Q9YC47;  
 DT 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE 427AA LONG HYPOTHETICAL 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE. . .

SCORES Init1: 82 Initn: 105 Opt: 210 z-score: 232.0 E(): 6.1e-05  
 Smith-Waterman score: 330; 26.2% identity in 446 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 Q9YC47 MDEGGVAVLNLVVKPSQIGGEIIAPSSKGHTIAALFASLLAGGRSKISEPLYSRDVDTAI  
 nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAP--LDFGNAATGCRLTMGLVGVYDFDST  
 Q9YC47 YAVRRFGGLVQKGERDLIVDSPPR-----PWWPRVNVNCRSSSTVLRHSIVTAALAPGISL  
 nk603cp4.pep FIGDASLTKRPMGRVNLPLREMGVQKSEGDRLPVTLRGPKTPTPIYRVFMASAVKVS  
 Q9YC47 VYGDSTNSTPVSSELAKKIGAEVVTING-RPPVAVKGPLARGCCEGETVGASDGESL  
 nk603cp4.pep AVLLAGLNTPGITTVIEPIIMTRD--HTE---KMLQGGFANPTVETDADGVRTIRLEGRGK

Q9YC47 AALL--LASPLLGFALKRAGTHSWHHVAVALHVLRGFGAK--ISFDGGIYYPDKPYGPGA  
 nk603cp4.pep LTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVL-MNPTRTGLILT-LQEMGADIEVI  
 Q9YC47 YA-----VPGDLYVNA--PLLLAGAIAGR-ITVRRLLDPEDPQGEKVFNLNLAQAGA-----  
 nk603cp4.pep NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELR  
 Q9YC47 --KLKAGESSVEV-VGTGSLEGFEADVSETPSLA---PVLAVLAAYAKGRSVIKGISHLR  
 nk603cp4.pep VKESDRLSAVANGKLNGVDCDE--GETSLVVRGPRDGKGLGNASGAAVATHLDHRIAMS  
 Q9YC47 LKEGGNKLKPLMSNLRRLKVKAKPRCGDCLEVYGEVYVEG-GTAKGYG-----DPRMTMA  
 nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEISDTKAA  
 Q9YC47 FAVAGLASRKGVRVTGASRYRDYYPGFVEDLRSVGAVIEAD  
 nk603cp4.pep GP\_BCT2:AP000061\_266  
 LOCUS AP000061\_266 [AP000061]  
 DEFINITION Aeropyrum pernix genomic DNA, section 4/7;  
 similar to PIR:D69202 percent identity:30.935 in 417aa.  
 DATE 06-APR-2000  
 ACCESSION AP000061 . . .  
 SCORES Init1: 82 Initn: 105 Opt: 210 z-score: 232.0 E(): 6.1e-05  
 Smith-Waterman score: 330; 26.2% identity in 446 aa overlap  
 nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AP000061\_266 MDEGGVAVLNLVVKPSQIGGEIIAPSSKGHTIAALFASLLAGGRSKISEPLYSRDVDTAI  
 nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAP--LDFGNAATGCRLTMGLVGVYDFDST  
 AP000061\_266 YAVRRFGGLVQKGERDLIVDSPPR-----PWWPRVNVNCRSSSTVLRHSIVTAALAPGISL  
 nk603cp4.pep FIGDASLTKRPMGRVNLPLREMGVQKSEGDRLPVTLRGPKTPTPIYRVFMASAVKVS  
 AP000061\_266 VYGDSTNSTPVSSELAKKIGAEVVTING-RPPVAVKGPLARGCCEGETVGASDGESL  
 nk603cp4.pep AVLLAGLNTPGITTVIEPIIMTRD--HTE---KMLQGGFANPTVETDADGVRTIRLEGRGK  
 AP000061\_266 AALL--LASPLLGFALKRAGTHSWHHVAVALHVLRGFGAK--ISFDGGIYYPDKPYGPGA  
 nk603cp4.pep LTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVL-MNPTRTGLILT-LQEMGADIEVI  
 AP000061\_266 YA-----VPGDLYVNA--PLLLAGAIAGR-ITVRRLLDPEDPQGEKVFNLNLAQAGA-----

nk603cp4.pep NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIILAVAAFAEGATVMNGLEELR  
 AP000061\_266 --KLKAGESSVEV-VGTGSLGFEADVSETPSLA---PVLAVLAAYAKGRSVIKGISHLR  
 nk603cp4.pep VKESDRLSAVANGKLKLVGDCDE--GETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
 AP000061\_266 LKEGGLNKLPLMSNLRRLKVKAKPRCGGDCLEVYGEVYVGG-GTAKGYG-----DPRMTMA  
 nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLAKIELSDTKAA  
 AP000061\_266 FAVAGLASRKGVVVTGASRYRDYYPGFVEDLRSVGAVIEAD  
 nk603cp4.pep  
 SWISSPROT:ARO1\_SCHPO  
 ID ARO1\_SCHPO STANDARD; PRT; 1573 AA.  
 AC Q9P7R0;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE PENTAFUNCTIONAL AROM POLYPEPTIDE [INCLUDES: 3-DEHYDROQUINATE SYNTHASE . . .

SCORES Initl: 109 Initn: 179 Opt: 216 z-score: 229.8 E(): 8e-05  
 Smith-Waterman score: 415; 27.3% identity in 451 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLL  
 ARO1\_SCHPO ATVVSDDDIRFILSRDVKVDEFTKSSWDVVVTPPGSKSISNRLVLAAMNGTCRLTNML  
 nk603cp4.pep EGEDVINTGKAMQAMGARI---RKEGDTWIDGVNGGGLLA-PEAPLDFGNAATGCRLTM  
 ARO1\_SCHPO HSDDTQFMMSALESLSGAATFSWEDGGETLVVKG--NGGKLAVKEELYLGNAAGTAARFLT  
 nk603cp4.pep G---LVGVYDFDSTFI-GDASLTKRPMGRVNLPLREMGVQVKS-EDGDRLPVTL---RGP  
 ARO1\_SCHPO GIAALVSSKDGAKVVLVTGNHRMKVRPIGVLVDALRANGCEINYLEKQGSPLPLDSSKNGL  
 nk603cp4.pep KTPTPTIYRVPMASQVKSAVLLAGLNTPGITTVI---EPI---MTRDHTKMLQGFGANP  
 ARO1\_SCHPO KGGI-IELAATVSSQYVSSILMCAPYASQPVTLKLVGGKPIISQLYIDMTIAMMASFGVNV  
 nk603cp4.pep TVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNV---LM  
 ARO1\_SCHPO TKSTTEENTYNIPC-GKYQNPPIHY-EIESDASSATYPLAIAA-ITGKTCTVPNIGSASLQ  
 nk603cp4.pep NPTRTGLILTLQEMGADIE-VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEY  
 ARO1\_SCHPO GDARFACDV-LRPMGCTVEQTATSTTVQGPFGTKL---PLESIDM-ETMTDAFLTASV

nk603cp4.pep ILAVAAFAEG--ATVMNGLEELRVKESDRLSAVANGKLKLVGDCDEGETSLVVRGRPDG  
 ARO1\_SCHPO VAAVACNVSEGDVPVTRITGIANQRVKECNRIAMVHELAKFGVRTGELEDGIYIFGK-NY  
 nk603cp4.pep KGLGNASGAAVATHLDHRIAMSFLVMGLVSENPTVDDATMIATSFPEFMD-LMAGLGA  
 ARO1\_SCHPO KELKKPE-EGITYYDDHRIAMSFSVLSLICPSRTLIIDKACVEKTFWYVWVDLHQSFGVK  
 nk603cp4.pep IELSDTKAA  
 ARO1\_SCHPO LTGATSVASDPLKGSISKNASIILIGMRGAKTTTIGKIIAKQLNFKFLDLDELLEDYLEM  
 nk603cp4.pep  
 GP\_PLN4:AL157734\_2

LOCUS AL157734\_2 [SPAC1834]  
 DEFINITION S.pombe chromosome I cosmid c1834;  
 SPAC1834.02, len:1594, SIMILARITY:Saccharomyces  
 cerevisiae, ARO1\_YEAST, pentafunctional arom polypeptide  
 [includes: 3-dehydroquininate synthase(ec 4.6.1.3);  
 3-dehydroquininate dehydratase, (1588 aa), fasta scores: . . .

SCORES Initl: 109 Initn: 179 Opt: 216 z-score: 229.8 E(): 8e-05  
 Smith-Waterman score: 415; 27.3% identity in 451 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLL  
 AL157734\_2 ATVVSDDDIRFILSRDVKVDEFTKSSWDVVVTPPGSKSISNRLVLAAMNGTCRLTNML  
 nk603cp4.pep EGEDVINTGKAMQAMGARI---RKEGDTWIDGVNGGGLLA-PEAPLDFGNAATGCRLTM  
 AL157734\_2 HSDDTQFMMSALESLSGAATFSWEDGGETLVVKG--NGGKLAVKEELYLGNAAGTAARFLT  
 nk603cp4.pep G---LVGVYDFDSTFI-GDASLTKRPMGRVNLPLREMGVQVKS-EDGDRLPVTL---RGP  
 AL157734\_2 GIAALVSSKDGAKVVLVTGNHRMKVRPIGVLVDALRANGCEINYLEKQGSPLPLDSSKNGL  
 nk603cp4.pep KTPTPTIYRVPMASQVKSAVLLAGLNTPGITTVI---EPI---MTRDHTKMLQGFGANP  
 AL157734\_2 KGGI-IELAATVSSQYVSSILMCAPYASQPVTLKLVGGKPIISQLYIDMTIAMMASFGVNV  
 nk603cp4.pep TVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNV---LM  
 AL157734\_2 TKSTTEENTYNIPC-GKYQNPPIHY-EIESDASSATYPLAIAA-ITGKTCTVPNIGSASLQ  
 nk603cp4.pep NPTRTGLILTLQEMGADIE-VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEY  
 AL157734\_2 GDARFACDV-LRPMGCTVEQTATSTTVQGPFGTKL---PLESIDM-ETMTDAFLTASV  
 nk603cp4.pep ILAVAAFAEG--ATVMNGLEELRVKESDRLSAVANGKLKLVGDCDEGETSLVVRGRPDG



NK603 analysis  
Page 91 of 120

```

      390      400      410      420      430
nk603cp4.pep GKGL---GNASG-AAVATHLDHRIAMSF-LVMGLVSE-----NPVTVDATMIATSF
      800      810      820      830      840      850
Z48179_2      IKDLKVPSSDSSGPGVGVCTYDDHRVAMSFSLLAGMVNSQNERDEVANPVRILERHCTGKTW

      440      450
nk603cp4.pep PEFMD-LMAGLGAKIELSDTKAA
      860      870      880      890      900      910
Z48179_2      PGWWDVLHSELGAKLDGAEPLECTSKKNKSKSVVIIGMRAAGKTTISKWCASALGYKLVD

nk603cp4.pep
SWISSPROT:ARO1_YEAST

ID   ARO1_YEAST      STANDARD;          PRT; 1588 AA.
AC   P08566;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PENTAFUNCTIONAL AROM POLYPEPTIDE [INCLUDES: 3-DEHYDROQUINATE SYNTHASE . . .

SCORES   Init1: 82   Initn: 122   Opt: 212   z-score: 225.5 E(): 0.00014
Smith-Waterman score: 424;   28.1% identity in 466 aa overlap

      10      20      30      40      50
nk603cp4.pep      MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGE
      370      380      390      400      410      420
ARO1_YEAST      DEDLRFILTDLTVLPFKDIPADQQKVVIPGSKSISNRALILAAALGEGQCKIKNLLHSD

      60      70      80      90      100      110
nk603cp4.pep      DVINTGKAMQAM-GARIRKE--GDTWIIDGVNGGLLAPEPLDFGNAATGCRLLTMGLVVG
      460      470      480      490      500      510
ARO1_YEAST      DTKHMLTAVHELKATISWEDNGETVVGEGHGGSTLSACADPLYLGNAGTASRFLTSLAA

      120      130      140      150      160
nk603cp4.pep      VYDFDST-----FIGDASLTKRPMGRVNLPLREMGVQVKSEDDG-RLPVTLRGPKT--PT
      520      530      540      550      560      570
ARO1_YEAST      LVNSTSSQKIYIVLTGNARMQORPIAPLVDLSLRANGTKIEYLNNEGSLPIKVYTDVSFVKGK

      170      180      190      200      210
nk603cp4.pep      PITRYVPMAAQVKSAVLLAGLNTPGITTVI---EPI--MTRDHTKMLQGFANPTVET
      580      590      600      610      620
ARO1_YEAST      RIELAATVSSQVYVSSILMCAPYAEEFVTLALVGGKPISKLYVDMTIKMEKFGIN--VET

      220      230      240      250      260      270
nk603cp4.pep      DADGVRTIRLEGRGKLTGQVIDVPGDPSSTAPFLVAALLVPGSDVTILNV---LMNPTR
      630      640      650      660      670      680
ARO1_YEAST      STTEPYTYIIPKGHYINPSEYVIESDASSATYPLAFAAMT-GTTVTVPNIGFESLQGDAR

      280      290      300      310      320      330
nk603cp4.pep      TGLILTLQEMGADI-EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAV
      690      700      710      720      730
ARO1_YEAST      FARDV-LKPMGCKITQTATSTTVSGPPVGTLK---PLKHV---DMEP-MTDAFLTACV

      340      350      360      370      380
nk603cp4.pep      AAAPAEGA-----TVMNGLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPD
      740      750      760      770      780      790
ARO1_YEAST      VAAISHSDSPNSANTTTIEGIANQVRVKECNRIILAMATELAKFGVKTTTELPGDIQVHGLNS

      390      400      410      420      430
nk603cp4.pep      GKGL---GNASG-AAVATHLDHRIAMSF-LVMGLVSE-----NPVTVDATMIATSF

```

```

      800      810      820      830      840      850
ARO1_YEAST      IKDLKVPSSDSSGPGVGVCTYDDHRVAMSFSLLAGMVNSQNERDEVANPVRILERHCTGKTW

      440      450
nk603cp4.pep PEFMD-LMAGLGAKIELSDTKAA
      860      870      880      890      900      910
ARO1_YEAST      PGWWDVLHSELGAKLDGAEPLECTSKKNKSKSVVIIGMRAAGKTTISKWCASALGYKLVD

nk603cp4.pep
SWISSPROT:ARO1_EMENI

ID   ARO1_EMENI      STANDARD;          PRT; 1586 AA.
AC   P07547;
DT   01-APR-1988 (Rel. 07, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   PENTAFUNCTIONAL AROM POLYPEPTIDE [INCLUDES: 3-DEHYDROQUINATE SYNTHASE . . .

SCORES   Init1: 110   Initn: 181   Opt: 210   z-score: 223.3 E(): 0.00018
Smith-Waterman score: 393;   25.6% identity in 465 aa overlap

      10      20      30
nk603cp4.pep      MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFG
      370      380      390      400      410      420
ARO1_EMENI      SAIGTPYETRASVVANEDIRVVLAPSIEVHPGVAHSSNVICAP---PGSKSISNRALVLA

      40      50      60      70      80      90
nk603cp4.pep      GLASGETRITGLLEGEDVINTGKAMQAMGARI---RKEGDTWIIDGVNGGLLAPEAPLD
      430      440      450      460      470      480
ARO1_EMENI      ALGSGTCRIKNLHSDDETEVMLNALERLGAATFSWEEEGEVLVNKGK-GNLQASSSPLY

      100      110      120      130      140      150
nk603cp4.pep      FGNAATGCRLL---TMGLVGVYDFDSTFI-GDASLTKRPMGRVNLPLREMGVQVKSEDDG-
      490      500      510      520      530      540
ARO1_EMENI      LGNAGTASRFLTTVATLANSSSTVDSVLTGNRMKQRPIGDLVDALTANVLPLNTSKGRA

      160      170      180      190      200
nk603cp4.pep      RLP--VTLRGFKTPTPTITYRVPMASQVKSAVLLAGLNTPGITTVI---EPI-MTR--DHT
      550      560      570      580      590      600
ARO1_EMENI      SLPLKIAASGGFAGGNINLAQVSSQVSSLLMCAPYAKEPVTLRVLVGGKPISQPIDMT

      210      220      230      240      250      260
nk603cp4.pep      EKMLQGFANPTVETDADGVRTIRLEGRGKLTGQVIDVPGDPSSTAPFLVAALLVPGSDV
      610      620      630      640      650      660
ARO1_EMENI      TAMMRSFGIDVQKSTTEHTYHIP-QGRVYNPAEYV-IESDASCATYPLAVAA-VTGTTC

      270      280      290      300      310
nk603cp4.pep      TILNV---LMNPTRTGLILTQEMGADIEVI---NPRLAGGEDVADLRVRSSTLKGVTVP
      670      680      690      700      710
ARO1_EMENI      TVPNIGSASLQGDARFA-VEVLRPMGCTVEQTETSTVTVTGFSDDGI---LRATSKRGYGTN

      320      330      340      350      360      370
nk603cp4.pep      EDRAPSMI---DEYFILAFAAFAEGATVMNGLEELRVKESDRLSAVANGKLNGVDCDE
      720      730      740      750      760      770
ARO1_EMENI      DRCVPRCFRTGSHRPMKESQTT---PPVSSGIANQVRVKECNRIKAMKDELAKFGVICRE

      380      390      400      410      420      430
nk603cp4.pep      GETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSF-LVMGLVSENPVTVDATMIATSF
      780      790
ARO1_EMENI      HDDGLEIDG-IDRSNLRQPVGG-VFCYDDHRVAFSFSVLSLVTPOPTLILEKECVGKTWP

```



SWISSPROT:AROA\_THEAC

ID AROA\_THEAC STANDARD; PRT; 410 AA.  
AC Q9HLE6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE PROBABLE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) . . .

SCORES Init1: 76 Initn: 235 Opt: 188 z-score: 208.5 E(): 0.0012  
Smith-Waterman score: 316; 26.6% identity in 428 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGG--LASGETRITGLLEGEDVIN  
AROA\_THEAC MTKVIYSGSGGTVALPSSKSFTQR-YILGVSFLNKSVTNLNVYVITITGDDISIA

nk603cp4.pep TGKAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDFGNAATGCRLTMGLVGVYDFDSTFI  
AROA\_THEAC LDIAQRA-GANITMNDSSIKI---RTSFTCP-SDIYVGESATSYRIALGLLGSAGCITH

nk603cp4.pep FIGDASLTKRPMGRVLNPLREMGVQVK-SEGDRLPVTLRGPKTPTPTIYRVPMASQVKS  
AROA\_THEAC VKGDPLAKRPMDFLVKALEANDVHLKLNEDG---FYDVGSESKKRYI-EVDGVSSQFV

nk603cp4.pep SAVLLAGLNTFGITTVIEPIIMTRDH---TEKMLQGFANPTVETDADGVRTIRLEGRGKL  
AROA\_THEAC SSMIFYARKGGGEFLVKNMRS PGYVYITKRVLYDLGYFVNIE-----KTITINPSG-V

nk603cp4.pep TGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIEVINPR  
AROA\_THEAC WKTAIDVEPDYSSLAFFLVGLLSENVDVRF-----NIKRMRSRI---QPDSVLLDMFKDN

nk603cp4.pep LAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAAFAGATVMNGLEELRVKE  
AROA\_THEAC IAIIDRT--LRVLPGRIDRTITVDADHNP---DLCPPISVGIFSEHGVEIDNYARLKTKE

nk603cp4.pep SDRLSAVAN-GLKLNVDCEGETSLVVRGPRPDGKGLGNASGAAVATHLDHRIAMSFLVM  
AROA\_THEAC SNRYEGIIDMASRFGAIVEDNGKDLFIKRG--DLKDPGTLS-----YTDHRMIMSAAVA

nk603cp4.pep GLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AROA\_THEAC AAASGFEVEIENETKVSFPGFFKELSKFANVSES

nk603cp4.pep  
TREMBL\_NEW: BAB67386

ID BAB67386 PRELIMINARY; PRT; 408 AA.  
AC BAB67386;  
DT 01-NOV-2001 (EMBLrel. 63, Created)  
DT 01-NOV-2001 (EMBLrel. 63, Last sequence update)  
DT 01-NOV-2001 (EMBLrel. 63, Last annotation update)  
DE 408AA LONG HYPOTHETICAL 3-PHOSPHOSHIKIMATE . . .

SCORES Init1: 89 Initn: 117 Opt: 186 z-score: 206.4 E(): 0.0016  
Smith-Waterman score: 351; 25.2% identity in 440 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
BAB67386 MLVEINPSKIYKVKAPQSKSFGIRLVLYSLK--ESKLDNLIPSDDDVNAI

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDFGNAATGCRLTMGLVGVYDFDSTFI  
BAB67386 NVVKQLGVSV--EGTYFKREKE---LVTPKF-LYFGGSATTLRMSIPILSVLGVDTIID

nk603cp4.pep GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTIYRVPMASQVKS  
BAB67386 GDETLLRKRPLNAIICAL-EGSVSFSS---SLLPTKISGKLKENFVRIEGGESSQYISGFI

nk603cp4.pep LLAGLNTFGITTVIEPIIMTRDH---TEKMLQGFANPTVETDADGVRTIRLEGRGKLTKGQ  
BAB67386 YAFSLVGGGEIEIIPISSKSYIYLTIELLNSLGNIMKGNK-----IYVB-KGDFKPY

nk603cp4.pep VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTLQEMGADIEVIN-PRLA  
BAB67386 IGKVPGDYALASF-YASSIVSGGEIVIEDVYELPNFDG-----DHSIVNFYKMM

nk603cp4.pep GGED-VADLR--VRSST-LKGVTVPEDRAPSMIDEPILAVAAAFAGATVMNGLEELRV  
BAB67386 GABSIVKDNKWKVKSSEKLNIEVNVDDYP---DLAPSIASLAPFSSPTTIKGIKRLKT

nk603cp4.pep KESDRLSAVANGLKLNVDCEGETSLVVRGPRPDGKGLGNASGAAVATHLDHRIAMSFLV  
BAB67386 KESNRVVTISETLSKFGKVEYDEDKIVY--PSEVKAGH-----VICPNDRHRIAMLASV

nk603cp4.pep MGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
BAB67386 LSPFKSGG--TIEKAEVKNKSNPNFWKDLISLNGRIIR

nk603cp4.pep  
GP\_BCT1: AE006665\_12

LOCUS AE006665\_12 [AE006665]  
DEFINITION Sulfolobus solfataricus section 24 of 272 of the complete genome;  
Amino Acid Biosynthesis, Aromatic amino acids.  
DATE 26-APR-2001  
ACCESSION AE006665  
NID . . .

SCORES Init1: 111 Initn: 183 Opt: 182 z-score: 201.9 E(): 0.0028  
Smith-Waterman score: 365; 25.5% identity in 443 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG



NK603 analysis  
Page 96 of 120

nk603cp4.pep 360 370 380 390 400 410  
 VKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFL  
 AE001684\_6 DKESDRILAITEELQKMGACIQPTHGGLLVNPSF-----LYGAVLDSHDDHRIAMALT  
 340 350 360 370 380

nk603cp4.pep 420 430 440 450  
 VMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AE001684\_6 IAALYASGDSRIHNTACVRKTFPNFVQTLNIMEARIEECHDNYSMWSTHKRKVFARESF  
 390 400 410 420 430 440

nk603cp4.pep  
 GP\_BCT2:AP002548\_241

LOCUS AP002548\_241 [AP002548]  
 DEFINITION Chlamydophila pneumoniae J138 genomic DNA, complete sequence,  
 section 4/4;  
 DATE 08-DEC-2000  
 ACCESSION AP002548  
 NID . . .

SCORES Init1: 115 Initn: 297 Opt: 180 z-score: 199.3 E(): 0.004  
 Smith-Waterman score: 415; 27.6% identity in 442 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AE002548\_241 MLTYKVSPSSVYGNAPFIPSSKSHTLRAILWASVAEGKSIINYLDSPDTEAMI  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110 119  
 KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPL-DFGNAATGCRITMGLVGVYDFDSTF  
 AE002548\_241 CACKQMGASIKKFPQ--ILEIVGNPLAIFPKYTLIDAGNSGIVLRFMTALACVFSKEITV  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
 IGDASLTKRPMGRVNLPLREMGVQVK-SEGDRLPVTLRGPKTPTPTIYRVPMASAOVKS  
 AE002548\_241 TGSSQLQRRPMAPLLQALRNFGASFHFSSDKSVLPFTMSGPLRSA--YSDVEGSDSQFAS  
 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
 AVLLA-GLNT-PGITTVEIPIMTR--DHTEKMLQGFANPTVETDADGVRTIRLEGRGKL  
 AE002548\_241 ALAVACSLAEGPCSFITIEPKERPWFDSLWLEKLHL-PYSCSDT---TYSFPGSSHP  
 170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 290  
 TGQVIDVPGDPSSTAFPLVAALLVPG-SDVTILNV-LMNPTRTGLILTL-QEMGADIEVI  
 AE002548\_241 QGFSYHVTDGDFSSAAFIAAAALLSKSLQPIRLRLDILDIOGDKIFFSLMQLGASIQY-  
 230 240 250 260 270 280

nk603cp4.pep 300 310 320 330 340 350  
 NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELR  
 AE002548\_241 -----DNEEILVFPSSFSGGSIDMD---GCIDALPILTVLCCFADSPSHLYNARS  
 290 300 310 320 330

nk603cp4.pep 360 370 380 390 400 410  
 VKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFL  
 AE002548\_241 DKESDRILAITEELQKMGACIQPTHGGLLVNPSF-----LYGAVLDSHDDHRIAMALT  
 340 350 360 370 380

420 430 440 450

nk603cp4.pep VMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AP002548\_241 IAALYASGDSRIHNTACVRKTFPNFVQTLNIMEARIEECHDNYSMWSTHKRKVFARESF  
 390 400 410 420 430 440

nk603cp4.pep  
 GP\_BCT1:AE002240\_9

LOCUS AE002240\_9 [AE002240]  
 DEFINITION Chlamydophila pneumoniae AR39, section 68 of 94 of the complete  
 genome;  
 3-phosphoshikimate 1-carboxyvinyltransferase;  
 identified by match to PFAM protein family HMM PF00275.  
 DATE 30-MAY-2000 . . .

SCORES Init1: 115 Initn: 297 Opt: 180 z-score: 199.3 E(): 0.004  
 Smith-Waterman score: 415; 27.6% identity in 442 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AE002240\_9 MLTYKVSPSSVYGNAPFIPSSKSHTLRAILWASVAEGKSIINYLDSPDTEAMI  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110 119  
 KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPL-DFGNAATGCRITMGLVGVYDFDSTF  
 AE002240\_9 CACKQMGASIKKFPQ--ILEIVGNPLAIFPKYTLIDAGNSGIVLRFMTALACVFSKEITV  
 60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
 IGDASLTKRPMGRVNLPLREMGVQVK-SEGDRLPVTLRGPKTPTPTIYRVPMASAOVKS  
 AE002240\_9 TGSSQLQRRPMAPLLQALRNFGASFHFSSDKSVLPFTMSGPLRSA--YSDVEGSDSQFAS  
 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
 AVLLA-GLNT-PGITTVEIPIMTR--DHTEKMLQGFANPTVETDADGVRTIRLEGRGKL  
 AE002240\_9 ALAVACSLAEGPCSFITIEPKERPWFDSLWLEKLHL-PYSCSDT---TYSFPGSSHP  
 170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 290  
 TGQVIDVPGDPSSTAFPLVAALLVPG-SDVTILNV-LMNPTRTGLILTL-QEMGADIEVI  
 AE002240\_9 QGFSYHVTDGDFSSAAFIAAAALLSKSLQPIRLRLDILDIOGDKIFFSLMQLGASIQY-  
 230 240 250 260 270 280

nk603cp4.pep 300 310 320 330 340 350  
 NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELR  
 AE002240\_9 -----DNEEILVFPSSFSGGSIDMD---GCIDALPILTVLCCFADSPSHLYNARS  
 290 300 310 320 330

nk603cp4.pep 360 370 380 390 400 410  
 VKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFL  
 AE002240\_9 DKESDRILAITEELQKMGACIQPTHGGLLVNPSF-----LYGAVLDSHDDHRIAMALT  
 340 350 360 370 380

nk603cp4.pep 420 430 440 450  
 VMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 AE002240\_9 IAALYASGDSRIHNTACVRKTFPNFVQTLNIMEARIEECHDNYSMWSTHKRKVFARESF  
 390 400 410 420 430 440

nk603cp4.pep  
 SWISSPROT:AROA\_CHLPN





```

      10      20      30      40      50      60
nk603cp4.pep KAMQAMGARIR--KEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVG--VYDFD
AL138851_8  GLLQLHGVTVRPGEEPGLVLDPTH---VESANVADIDAHAGSSRIPLFCGPLLHRLG

      70      80      90     100     110
nk603cp4.pep STFI---GDASLTKRPMGRVLNPLREMGVQV--KSEGDRL--PVTLRGPKTPTPTITYRVP
AL138851_8  HAFIPGLGGCDIGGRPIDFHFQVLRQFGAKIEKRADGQYLEAPQRLRGTKINLP--YPSV

      120     130     140     150     160     170
nk603cp4.pep MASAOVK--SAVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLE
AL138851_8  GATEQVLLTAVLAEGV--TELSNAAVEPEI--EDLICVLQKMGAIAMTD---RTIRVT

      180     190     200     210     220     229
nk603cp4.pep GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIE
AL138851_8  GVDELGGYTHRALSDRLEAASWASAALATEG--NVYVRGAQQRSMMT--FLNTRFKVGGAFE

      230     240     250     260     270     280     289
nk603cp4.pep VINPRLAGGEDVADLRV--RSSTLKGVTVPEDRAPSM--IDEYPILAVAAFAEGATVMN-
AL138851_8  I-----DDEGIRFWHPGGRKLSIALETDVHFGFQTDWQQPLVVALTQATGLSIVHE

      290     300     310     320     330     340
nk603cp4.pep GLEELRVKESDRLSAVANGLKL-----NGVDCDEGETSLVVRGPDGKGLGNASGAAVAT
AL138851_8  TVYESRLGFTSALNQMGAGHIQLYRECLGGSDCRFGQRNPLHSAVVSQPT--KLEGADLVI

      350     360     370     380     390     400
nk603cp4.pep HLDHRIAMSFVLMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AL138851_8  P-DLRGGFSYLAALAAQGTSRVHGIDLINRGYENFMDKLVELGAKVELPGKALG

      410     420     430     440     450
nk603cp4.pep HLDHRIAMSFVLMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AL138851_8  P-DLRGGFSYLAALAAQGTSRVHGIDLINRGYENFMDKLVELGAKVELPGKALG

nk603cp4.pep
TREMBL_MAIN:Q9L1U5

ID   Q9L1U5      PRELIMINARY;      PRT;      448 AA.
AC   Q9L1U5;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   UDP-N-ACETYLGLUCOSAMINE TRANSFERASE. . . .

SCORES   Init1: 59   Initn: 141   Opt: 175   z-score: 193.9 E(): 0.008
Smith-Waterman score: 271;   24.6% identity in 452 aa overlap

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG
Q9L1U5      MTVNGADDVLLVHGGTPLEGEIRVRGAKNLPKAMVAALLGSAPSRLRNVDIRDVVRV
      10      20      30      40      50      60
nk603cp4.pep KAMQAMGARIR--KEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVG--VYDFD
Q9L1U5      GLLQLHGVTVRPGEEPGLVLDPTH---VESANVADIDAHAGSSRIPLFCGPLLHRLG
      70      80      90     100     110

```

```

      120     130     140     150     160     170
nk603cp4.pep STFI---GDASLTKRPMGRVLNPLREMGVQV--KSEGDRL--PVTLRGPKTPTPTITYRVP
Q9L1U5      HAFIPGLGGCDIGGRPIDFHFQVLRQFGAKIEKRADGQYLEAPQRLRGTKINLP--YPSV
      120     130     140     150     160     170
nk603cp4.pep MASAOVK--SAVLLAGLNTPGITTVIEPIMTRDHTKMLQGFGANPTVETDADGVRTIRLE
Q9L1U5      GATEQVLLTAVLAEGV--TELSNAAVEPEI--EDLICVLQKMGAIAMTD---RTIRVT
      180     190     200     210     220     229
nk603cp4.pep GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIE
Q9L1U5      GVDELGGYTHRALSDRLEAASWASAALATEG--NVYVRGAQQRSMMT--FLNTRFKVGGAFE
      230     240     250     260     270     280     289
nk603cp4.pep VINPRLAGGEDVADLRV--RSSTLKGVTVPEDRAPSM--IDEYPILAVAAFAEGATVMN-
Q9L1U5      I-----DDEGIRFWHPGGRKLSIALETDVHFGFQTDWQQPLVVALTQATGLSIVHE
      290     300     310     320     330     340
nk603cp4.pep GLEELRVKESDRLSAVANGLKL-----NGVDCDEGETSLVVRGPDGKGLGNASGAAVAT
Q9L1U5      TVYESRLGFTSALNQMGAGHIQLYRECLGGSDCRFGQRNPLHSAVVSQPT--KLEGADLVI
      350     360     370     380     390     400
nk603cp4.pep HLDHRIAMSFVLMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Q9L1U5      P-DLRGGFSYLAALAAQGTSRVHGIDLINRGYENFMDKLVELGAKVELPGKALG
      410     420     430     440     450
nk603cp4.pep HLDHRIAMSFVLMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Q9L1U5      P-DLRGGFSYLAALAAQGTSRVHGIDLINRGYENFMDKLVELGAKVELPGKALG

nk603cp4.pep
TREMBL_MAIN:Q99Z78

ID   Q99Z78      PRELIMINARY;      PRT;      419 AA.
AC   Q99Z78;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE   PUTATIVE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE. . . .

SCORES   Init1: 69   Initn: 97   Opt: 173   z-score: 192.1 E(): 0.01
Smith-Waterman score: 237;   24.1% identity in 444 aa overlap

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG
Q99Z78      MRKIIINGKALSGEVAVSGAKNSVVALIPAILADDIVLDGVPAISDVDSLI
      10      20      30      40      50
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFG--NAATGCRLTMG-LVGVYDFDS
Q99Z78      EIMELMGATVNVHGDITLEIDPRG-----VDIPMPYGKINSRLRASYFYGSLGRFGQAV
      60      70      80      90     100
nk603cp4.pep TFI---GDASLTKRPMGRVLNPLREMGVQV--KSEGDRLPVTLRGPKTPTPTITYRVPMASAOV
Q99Z78      VGLPGGCDLGPRIIDLHLKAFAMGVEV--SYEGENMNLSTNGQKIHGAHIY--MDTVSVA
      110     120     130     140     150     160
nk603cp4.pep KSAVLLAGLNTPGITTVIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEGRGK

```

```

Q99Z78      TINTMVAATKAQG-KTVIENAAAREPEIIDVATLLNNMGAH---IRGAGTDIITIIGVQK
170         180         190         200         210         220

nk603cp4.pep 240         250         260         270         280         290
LTGQVIDVPGD--PSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILTLQEMGADIEVI
Q99Z78      LHGTRHQVIPDRIEAGTYIALAAI---GKGVKITNVLYEHLES-FIAKLEEMGVRMTV-
230         240         250         260         270

nk603cp4.pep 300         310         320         330         340
NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEY--PILAVAAAFABG-ATVMNGLE
Q99Z78      -----EEDAI FVEKQESLKAITIKTSPYPGFATDLQQPLTPLLK-ADGRGTIIDTIY
280         290         300         310         320

nk603cp4.pep 350         360         370         380         390         400
ELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAM
Q99Z78      EKRNHVPEL-----MRM-GADIS-----VIGGQIVYQGPSRLTGAQVKA-TDLRAGA
330         340         350         360         370

nk603cp4.pep 410         420         430         440         450
SFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Q99Z78      ALVTAGLIAEGKTEITNIEFILRGYASIIAKLTALGADIQLIED
380         390         400         410

nk603cp4.pep
GP_BCT1:AE006573_8

LOCUS      AE006573_8 [AE006573]
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 102 of 167 of
the complete genome;
Best Blastp hit = sp|P19670|MUA2_BACSU
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 2
(ENOYL PYRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE . . .

SCORES      Init1: 69      Initn: 97      Opt: 173      z-score: 192.1 E(): 0.01
Smith-Waterman score: 237;      24.1% identity in 444 aa overlap

nk603cp4.pep 10         20         30         40         50         60
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG
AE006573_8      MRKIIINGKALSGEVAVSGAKNSVVALIPAILADDIVILDGVPAISDVDSLI
10         20         30         40         50

nk603cp4.pep 70         80         90         100        110
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFG--NAATGCRLTMG-LVGVDYFDS
AE006573_8      EIMELMGATVNYHGDLEIDPRG----VQDIPMPYKINSLRASYFYGSLLGRFGQAV
60         70         80         90         100

nk603cp4.pep 120        130        140        150        160        170
TFI-GDASLTKRPMGRVLNPLREMGVQVKS EDGRLFVTLRGPKTPTPTITYRVPMSAQV
AE006573_8      VGLPGGCDLGPFPIDLHLKAFAMGVEV-SYEGENMNLSTNGQKINGAHYI-MDTVSVGA
110        120        130        140        150        160

nk603cp4.pep 180        190        200        210        220        230
KSAVLLAGLNTPGIITVIEPIMTRDH---TEKMLQGFANPTVETDADGVRTIRLEGRGK
AE006573_8      TINTMVAATKAQG-KTVIENAAAREPEIIDVATLLNNMGAH---IRGAGTDIITIIGVQK
170        180        190        200        210        220

nk603cp4.pep 240        250        260        270        280        290
LTGQVIDVPGD--PSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILTLQEMGADIEVI
AE006573_8      LHGTRHQVIPDRIEAGTYIALAAI---GKGVKITNVLYEHLES-FIAKLEEMGVRMTV-

```

```

230         240         250         260         270

nk603cp4.pep 300         310         320         330         340
NPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEY--PILAVAAAFABG-ATVMNGLE
AE006573_8      -----EEDAI FVEKQESLKAITIKTSPYPGFATDLQQPLTPLLK-ADGRGTIIDTIY
280         290         300         310         320

nk603cp4.pep 350         360         370         380         390         400
ELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAM
AE006573_8      EKRNHVPEL-----MRM-GADIS-----VIGGQIVYQGPSRLTGAQVKA-TDLRAGA
330         340         350         360         370

nk603cp4.pep 410         420         430         440         450
SFLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AE006573_8      ALVTAGLIAEGKTEITNIEFILRGYASIIAKLTALGADIQLIED
380         390         400         410

nk603cp4.pep
TREMBL_MAIN:Q9S0N1

ID      Q9S0N1      PRELIMINARY;      PRT;      446 AA.
AC      Q9S0N1;
DT      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE      UDP-N-ACETYLGLUCOSAMINE TRANSFERASE (FRAGMENT). . . .

SCORES      Init1: 59      Initn: 141      Opt: 170      z-score: 188.5 E(): 0.016
Smith-Waterman score: 266;      24.6% identity in 452 aa overlap

nk603cp4.pep 10         20         30         40         50         60
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMPGGLASGETRITGLLEGEDVINTG
Q9S0N1      VNGADDVLLVHGGTPLEGEIRVRGAKNLVFNELVAALLGSAPSRLRNVFDIRDVRVVR
10         20         30         40         50

nk603cp4.pep 70         80         90         100        110
KAMQAMGARIR--KEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGVLG--VYDFD
Q9S0N1      GLLQLHGVTVRPGEEPGLVLDPTH---VESANVADIDAHAGSSRIPIILFCGPELLHRLG
60         70         80         90         100        110

nk603cp4.pep 120        130        140        150        160        170
STFI---GDASLTKRPMGRVLNPLREMGVQV-KSEGDRL--PVTLRGPKTPTPTITYRVP
Q9S0N1      HAFIPGLGGCDIGGRPIDFHFVLRQFGAKIEKRAQGQYLEAPQRLRGTKINLP--YPSV
120        130        140        150        160        170

nk603cp4.pep 180        190        200        210        220        229
MASAQVK-SAVLLAGLNTPGIITVIEPIMTRDHTEKMLQGFANPTVETDADGVRTIRLE
Q9S0N1      GATEQVLLTAVLAEGV-TELSNAAVEPEI--EDLICVLQKMGAIAMDTD---RTIRVT
180        190        200        210        220

nk603cp4.pep 230        240        250        260        270        280        289
GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILTLQEMGADIE
Q9S0N1      GVDELGGYTHRALSDRLAASWASAALATEG-NVYVRGAQQRSMNT-FLNTFRKVGGAPE
230        240        250        260        270        280

nk603cp4.pep 290        300        310        320        330        340
VINPRLAGGEDVADLRV--RSSTLKGVTVPEDRAPSM-IDEYPILAVAAAFABGATVMN-
Q9S0N1      I-----DDEGIRFWHPGGRLKSIALETDVHPGFTDQWQPLVVALTQATGLSIVHE
290        300        310        320        330

```

nk603cp4.pep 350 360 370 380 390 400  
GLEELRVKESDRLSAVANGKLK-----NGVDCDEGETSLVVRGRPDGKGLGNASGAAVAT  
Q9S0N1 TVYESRLGFTSALNQMGAGHIQLYRECLGGSDCRFGQRNLFSAVVSQPT--KLEGADLVI  
340 350 360 370 380 390

nk603cp4.pep 410 420 430 440 450  
HLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Q9S0N1 P-DLRGGFSYLIAALAAQGTSSRVHGIDLINRGYENFMDKLVELGAKVELPGKALG  
400 410 420 430 440

nk603cp4.pep  
GP\_BCT1:AB033486\_2

LOCUS AB033486\_2 [AB033486]  
DEFINITION Streptomyces lividans hup, murA genes for histone-like protein H51,  
UDP-N-acetylglucosamine transferase, complete and partial sequence.  
DATE 19-OCT-1999  
ACCESSION AB033486  
NID . . .

SCORES Init1: 59 Initn: 141 Opt: 170 z-score: 188.5 E(): 0.016  
Smith-Waterman score: 266; 24.6% identity in 452 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AB033486\_2 VNGADDVLLVHGGTPLEGEIRVRGAKNLVFNELVAALLGSAPSRLRNVPDIRDVRVVR  
10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIR--KEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGV--VYDFD  
AB033486\_2 GLLQLHGVTVTRPGEPEPGLVLDPTH---VESANVADIDAHAGSSRIPLFCGPLLHRLG  
60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
STFI---GDASLTKRPMGRVLNPLREMGVQV--KSEDGDR--PVTLRGPKTPTPTITYRVP  
AB033486\_2 HAFIPGLGGCIDIIGRPIDFHPDVRQFGAKIEKRADGQYLEAPQRLRGTKINLP--YPSV  
120 130 140 150 160 170

nk603cp4.pep 180 190 200 210 220 229  
MASAQVK--SAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFANPTVETDADGVRTIRLE  
AB033486\_2 GATEQVLLTAVLAEGV--TELSNAAVEPEI--EDLICVLQKMGAIAMDTD---RTIRVT  
180 190 200 210 220

nk603cp4.pep 230 240 250 260 270 280 289  
GRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIE  
AB033486\_2 GVDELGGYTHRALSDRLAASWASAALATEG--NVYVRGAQQRSMT--FLNTRFKVGGAFE  
230 240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340  
VINPRLAGGEDVADLRV--RSSTLKGVTVPEDRAPSM--IDEYPILAVAAAFAGATVMN--  
AB033486\_2 I-----DDEGIRFWHPGGRLKSIALETDVHPGFQTDWQQPLVVALTQATGLSIVHE  
290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400  
GLEELRVKESDRLSAVANGKLK-----NGVDCDEGETSLVVRGRPDGKGLGNASGAAVAT  
AB033486\_2 TVYESRLGFTSALNQMGAGHIQLYRECLGGSDCRFGQRNLFSAVVSQPT--KLEGADLVI  
340 350 360 370 380 390

nk603cp4.pep 410 420 430 440 450  
HLDHRIAMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA

AB033486\_2 P-DLRGGFSYLIAALAAQGTSSRVHGIDLINRGYENFMDKLVELGAKVELPGKALG  
400 410 420 430 440

nk603cp4.pep  
TREMBL\_NEW:CAD00630

ID CAD00630 PRELIMINARY; PRT; 423 AA.  
AC CAD00630;  
DT 01-NOV-2001 (EMBLrel. 63, Created)  
DT 01-NOV-2001 (EMBLrel. 63, Last sequence update)  
DT 01-NOV-2001 (EMBLrel. 63, Last annotation update)  
DE MURZ PROTEIN. . . .

SCORES Init1: 89 Initn: 112 Opt: 165 z-score: 183.4 E(): 0.031  
Smith-Waterman score: 262; 24.3% identity in 449 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
CAD00630 MTDRLIIQGGKLSGTLQVDGAKNSAVALIPAAILAESEVVLEGLPDISDVYTLTY  
10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGV--DFDST  
CAD00630 DILEELGGSVRYDNKTAIIDPTD---MLS--MPLPSGNVKK--LRASYLYMGAMLRFFKA  
60 70 80 90 100

nk603cp4.pep 120 130 140 150 160 170  
FIG---DASLTKRPMGRVLNPLREMGVQVQKSEDDGDRLPVTLRGPKTPTPTITYRVMASAQ  
CAD00630 VIGLPGGCYLSPRPIDQHIKGFALGAKVTNEQG---AIYLRADLKGARIY--LDVVSVG  
110 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
VKSALLAGLNTPGITTVIEPIIMTRDHTKMLQGFANPTVETDADGVRTIRLEGRG  
CAD00630 ATINIMLAARVAKG--KTVIENAAKEPEIIDVATLLTNMGA---IKGA--GTDITIRITGV  
170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 289  
KLTVG---QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIE  
CAD00630 HLHGCHHTI--IPDRIEAGTFMVLAA--ASGKGIRIENVI--PTHLEGTIAKLTEMGVPM  
230 240 250 260 270

nk603cp4.pep 290 300 310 320 330 340  
VINPRLAGG--EDVADLRVRSSTLKGVTVPEDRAPSMIDEPYILAVAAAFAGATVMN--  
CAD00630 IEEDSIFIGEVENIKKVDIKTYAYPGF--PTDL-----QQPLTALLTR--AEGSSVITDT  
280 290 300 310 320

nk603cp4.pep 350 360 370 380 390 400  
LEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRI  
CAD00630 IYPSRFKHIAELERMGGKFKL-----EGRSVAVI-----SGPAKLGQSKV--TATDLRA  
330 340 350 360 370

nk603cp4.pep 410 420 430 440 450  
AMSFVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
CAD00630 GAALVIAGLAEGRTEIHGVEHIERGYSKIIIEKLSAIGADITRSSTAETNI  
380 390 400 410 420

nk603cp4.pep  
TREMBL\_MAIN:Q9ALU1

ID Q9ALU1 PRELIMINARY; PRT; 357 AA.  
AC Q9ALU1;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE UDP-N-ACETYLGALUCOSAMINE-LIKE PROTEIN (FRAGMENT) . . . .

SCORES Init1: 48 Initn: 75 Opt: 163 z-score: 182.4 E(): 0.035  
Smith-Waterman score: 167; 22.4% identity in 335 aa overlap

nk603cp4.pep 10 20 30 40 50 59  
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLAS-GETRITGLLEGEDVINT  
Q9ALU1 MEKIIIIQGGDNRLVGKVKIEGAKNAVLPLLAATVLASEGQSVLKNVPVLSVFTM  
10 20 30 40 50

nk603cp4.pep 60 70 80 90 100 110  
GKAMQAMGARI--RKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGVYDFDS  
Q9ALU1 NNVVRGLNTQVDFNQEENTVVVDATQP---LSEEPYKYVSKMRASIVVLGPVLARNGHA  
60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
--TFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMASAO  
Q9ALU1 KVSMPGGCTIGSRPIDLHLKGLAMGAQITQTAGY---IEAKAERLKGAIHY-MDFPSVG  
120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
VKSAVLLAGLNTPGIITTVIEPIMTRDHTKMLQGFANPTVETDADGVRTIRLEGRGKLT  
Q9ALU1 ATQNIMMAATLAQG-TTVIENA-AREPEIVDLALFLNEMGAKVRGAGTETLAIVGVDQLR  
170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 290  
GQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRL  
Q9ALU1 GAKHNVVQDRIRAGTFMVAAMTSG-DLLIEDAIWEHNRP-LLSKMQEMGVEVT-----  
230 240 250 260 270

nk603cp4.pep 300 310 320 330 340 350  
AGGEDVADLRVRS--STLKGVTVPEDRAPSM-IDEYPILAVAAFAEGATVMNGLEELRV  
Q9ALU1 ---EEDEGIRIRSDVSKLRPVSVKTLPPYGFPTDMQAQFTALMAMAKGESTMIETVFENR  
280 290 300 310 320 330

nk603cp4.pep 360 370 380 390 400 410  
KESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFLV  
Q9ALU1 FQHLEEMRRMGLHSDIMRDTAR  
340 350

nk603cp4.pep  
GP\_BCT2:AF343003\_1

LOCUS AF343003\_1 [AF343003]  
DEFINITION Streptococcus sanguinis UDP-N-acetylglucosamine-like protein gene,  
partial cds.  
DATE 14-FEB-2001  
ACCESSION AF343003  
NID . . .

SCORES Init1: 48 Initn: 75 Opt: 163 z-score: 182.4 E(): 0.035  
Smith-Waterman score: 167; 22.4% identity in 335 aa overlap

nk603cp4.pep 10 20 30 40 50 59  
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLAS-GETRITGLLEGEDVINT  
10 20 30 40 50 59

AF343003\_1 MEKIIIIQGGDNRLVGKVKIEGAKNAVLPLLAATVLASEGQSVLKNVPVLSVFTM  
10 20 30 40 50

nk603cp4.pep 60 70 80 90 100 110  
GKAMQAMGARI--RKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGVYDFDS  
AF343003\_1 NNVVRGLNTQVDFNQEENTVVVDATQP---LSEEPYKYVSKMRASIVVLGPVLARNGHA  
60 70 80 90 100 110

nk603cp4.pep 120 130 140 150 160 170  
--TFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMASAO  
AF343003\_1 KVSMPGGCTIGSRPIDLHLKGLAMGAQITQTAGY---IEAKAERLKGAIHY-MDFPSVG  
120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
VKSAVLLAGLNTPGIITTVIEPIMTRDHTKMLQGFANPTVETDADGVRTIRLEGRGKLT  
AF343003\_1 ATQNIMMAATLAQG-TTVIENA-AREPEIVDLALFLNEMGAKVRGAGTETLAIVGVDQLR  
170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 290  
GQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRL  
AF343003\_1 GAKHNVVQDRIRAGTFMVAAMTSG-DLLIEDAIWEHNRP-LLSKMQEMGVEVT-----  
230 240 250 260 270

nk603cp4.pep 300 310 320 330 340 350  
AGGEDVADLRVRS--STLKGVTVPEDRAPSM-IDEYPILAVAAFAEGATVMNGLEELRV  
AF343003\_1 ---EEDEGIRIRSDVSKLRPVSVKTLPPYGFPTDMQAQFTALMAMAKGESTMIETVFENR  
280 290 300 310 320 330

nk603cp4.pep 360 370 380 390 400 410  
KESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFLV  
AF343003\_1 FQHLEEMRRMGLHSDIMRDTAR  
340 350

nk603cp4.pep  
GP\_PLN4:S76670\_1

LOCUS S76670\_1 [S76670]  
DEFINITION arom=enolpyruvylshikimate-3-phosphate synthase type B [Pneumocystis  
carinii, isolate from infected ferret lung, Genomic, 324 nt];  
This sequence comes from Fig. 2; EPSP synthase B.  
DATE 27-SEP-1995  
ACCESSION S76670 . . .

SCORES Init1: 91 Initn: 91 Opt: 154 z-score: 180.6 E(): 0.044  
Smith-Waterman score: 154; 34.3% identity in 102 aa overlap

nk603cp4.pep 300 310 320 330 340 350  
DVADLRVRSSTLKGVTVPEDRAPSMIDEYPILA-VAAFAEGA-TVMNGLEELRVKESDR  
S76670\_1 FLTTAILASVAYKESESCITEITGISNRKECNR  
10 20 30

nk603cp4.pep 360 370 380 390 400 410  
LSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFLVGLVS  
S76670\_1 INAMVCEPKFGIEAGELPDGIYVRALNISDLLSPING--IDCHNDHRIAMSFSVLACVS  
40 50 60 70 80 90

nk603cp4.pep 420 430 440 450  
ENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
S76670\_1 PNPTIILNKTCVNKT  
100

nk603cp4.pep  
TREMBL\_MAIN:Q12430

ID Q12430 PRELIMINARY; PRT; 108 AA.  
AC Q12430;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE EPSP SYNTHASE (FRAGMENT) . . . .

SCORES Init1: 91 Initn: 91 Opt: 154 z-score: 180.6 E(): 0.044  
Smith-Waterman score: 154; 34.3% identity in 102 aa overlap

```

      300      310      320      330      340      350
nk603cp4.pep DVADLRVRSSTLKGVTVPEDRAPSMIDEYPILA-VAAAFAGA-TVMNGLEELRVKESDR
Q12430          ||| ||| :|: | :|: | :|: | :|: |
                  FLTTAILASVAYKESESCITEITGISNQRIKECNR
                        10      20      30
```

```

      360      370      380      390      400      410
nk603cp4.pep LSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAATHLDHRIAMSFVMGLVS
Q12430          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  INAMVCEFKKFGIEAGELPDGIYVRALNISDLLSPTNG--IDCHNDHRIAMSFVLAQVS
                        40      50      60      70      80      90
```

```

      420      430      440      450
nk603cp4.pep ENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
Q12430          ||: :|: | :|: | :|: | :|: | :|: |
                  PNPTIILNKTCVNKT
                        100
```

nk603cp4.pep  
GP\_PLN3:U31057\_1

LOCUS U31057\_1 [PCU31057]  
DEFINITION Pneumocystis carinii f. sp. mustelae B (arom) gene, EPSP synthase  
domain, partial cds;  
encodes EPSP synthase domain.  
DATE 17-APR-1996  
ACCESSION U31057 . . .

SCORES Init1: 91 Initn: 91 Opt: 154 z-score: 180.6 E(): 0.044  
Smith-Waterman score: 154; 34.3% identity in 102 aa overlap

```

      300      310      320      330      340      350
nk603cp4.pep DVADLRVRSSTLKGVTVPEDRAPSMIDEYPILA-VAAAFAGA-TVMNGLEELRVKESDR
U31057_1          ||| ||| :|: | :|: | :|: | :|: |
                  FLTTAILASVAYKESESCITEITGISNQRIKECNR
                        10      20      30
```

```

      360      370      380      390      400      410
nk603cp4.pep LSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAATHLDHRIAMSFVMGLVS
U31057_1          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  INAMVCEFKKFGIEAGELPDGIYVRALNISDLLSPTNG--IDCHNDHRIAMSFVLAQVS
                        40      50      60      70      80      90
```

```

      420      430      440      450
nk603cp4.pep ENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
U31057_1          ||: :|: | :|: | :|: | :|: | :|: |
                  PNPTIILNKTCVNKT
                        100
```

nk603cp4.pep  
TREMBL\_NEW:AAF86297

ID AAF86297 PRELIMINARY; PRT; 423 AA.  
AC AAF86297;  
DT 01-NOV-2001 (EMBLrel. 63, Created)  
DT 01-NOV-2001 (EMBLrel. 63, Last sequence update)  
DT 01-NOV-2001 (EMBLrel. 63, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE-1-CARBOXYVINYL TRANSFERASE. . . .

SCORES Init1: 76 Initn: 76 Opt: 156 z-score: 173.7 E(): 0.11  
Smith-Waterman score: 235; 23.4% identity in 449 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLGEGEDVINTG
AAF86297          ||: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  MTDKLIIRGGKKLAGTLQVDGAKNSAVALIPAILAESEVLEGLPDISDVHTLY
                        10      20      30      40      50
```

```

      70      80      90      100      110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTGMLVGVI--DFDST
AAF86297          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  NILEELGGTVRYDNKTAVIDPTDMISM-----PLPSGNVKK-LRASYILMGAMLRFKKA
                        60      70      80      90      100
```

```

      120      130      140      150      160      170
nk603cp4.pep FIG---DASLTKRPMGRVLNPLREMGVQVSEGDRLPVTLRGPKTPTPIYRVFMASAQ
AAF86297          ||: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  VIGLPGGCYLGPRPIQHIKGFALGAKVTNEQG---ATYLRADLEKGARIY-LDVVSVG
                        110      120      130      140      150      160
```

```

      180      190      200      210      220      230
nk603cp4.pep VKSAVLLAGLNTPGIITVIEPIIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEGRG
AAF86297          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  ATINIMLAARAKG-KTVIENAAKEPEIIDVATLLTNMGA---IIKGA-GTDTIRITGVE
                        170      180      190      200      210      220
```

```

      240      250      260      270      280      289
nk603cp4.pep KLTG--QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRT-GLITLQEMGADIE
AAF86297          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  HLHGCHHTI-IPDRIEAGTFMVLAA--ASGKGVIENVI--PTHLEGIIAKLTEMGVPM
                        230      240      250      260      270
```

```

      290      300      310      320      330      340      349
nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEE
AAF86297          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  IEEDAIFVGE-VE--KIKKVDIKTYAYPG--FPTDLQQ-PLTALLTR-AEGSSVITDT--
                        280      290      300      310      320
```

```

      350      360      370      380      390      400      409
nk603cp4.pep LRVKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAATHLDHRIAMS
AAF86297          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  --IYPS-RFKHIAEIERMGGKFKLEGRSAVI-----NGPVQLQGSKV-TATDLRAGAA
                        330      340      350      360      370
```

```

      410      420      430      440      450
nk603cp4.pep FLVMGLVSENPTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA
AAF86297          :|: :|: | :|: | :|: | :|: | :|: | :|: | :|: |
                  LVIAALLADGETIEHGVHEIERGYSKIIEKLSAIGANITRSSAAETKL
                        380      390      400      410      420
```

nk603cp4.pep  
TREMBL\_NEW:CAC97923

ID CAC97923 PRELIMINARY; PRT; 423 AA.  
AC CAC97923;  
DT 01-NOV-2001 (EMBLrel. 63, Created)  
DT 01-NOV-2001 (EMBLrel. 63, Last sequence update)  
DT 01-NOV-2001 (EMBLrel. 63, Last annotation update)  
DE MURZ PROTEIN. . . .





nk603cp4.pep CDEGETSLVVRGRPDGKGLGNASGA--AVATHLDHRIAMSFLVMGLVSENFPVTDDATMI  
| : : : : | : : : : | : : : : | : : : : | : :  
380 390 400 410 420 429



U31053\_1 AGELPDGIYVKA----QNISNLSYPVEGINCYNDRHRIAMSFVSLACISQKPTTILDKACV  
50 60 70 80 90 100

430 440 450  
nk603cp4.pep ATSFPEFMDLMAGLGAKIELSDTKAA

U31053\_1 NKT

nk603cp4.pep  
TREMBL\_MAIN:Q12654

ID Q12654 PRELIMINARY; PRT; 108 AA.  
AC Q12654;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (3- . . .

SCORES Init1: 55 Initn: 55 Opt: 140 z-score: 165.5 E(): 0.31  
Smith-Waterman score: 140; 30.0% identity in 90 aa overlap

320 330 340 350 360 370  
nk603cp4.pep KGVTVPEDRAPSMIDIEYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGKLNGVD  
Q12654 FLTAAVLASVAYBESKPYITKITGISNQRICEKNRIDAMICELKKFGIE  
10 20 30 40

380 390 400 410 420 429  
nk603cp4.pep CDEGETSLVVRGRPDGKGLGNASGA--AVATHLDHRIAMSFVLMGLVSENPTVDDATMI  
Q12654 AGELPDGIYVKA----QNISNLSYPVEGINCYNDRHRIAMSFVSLACISQKPTTILDKACV  
50 60 70 80 90 100

430 440 450  
nk603cp4.pep ATSFPEFMDLMAGLGAKIELSDTKAA

Q12654 NKT

nk603cp4.pep  
TREMBL\_MAIN:Q9A0I4

ID Q9A0I4 PRELIMINARY; PRT; 423 AA.  
AC Q9A0I4;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
DE PUTATIVE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE . . .

SCORES Init1: 52 Initn: 52 Opt: 148 z-score: 165.1 E(): 0.32  
Smith-Waterman score: 210; 22.2% identity in 442 aa overlap

10 20 30 40 50 59  
nk603cp4.pep MLHGASSRPATARKSSGSLSGTVRIIPGDKSISHRSFMFGGLAS-GETRITGLLEGEDVINT  
Q9A0I4 MDKIIIEGGQTRLEGEVVEIEGAKNAVLPLLAASILPSKSKTILRNVPIILSDVFTM  
10 20 30 40 50

60 70 80 90 100 110  
nk603cp4.pep GKAMQAMGARI--RKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGVYDFDS  
Q9A0I4 NNVVRGLDIRVDFNEAANEITVDASGH---ILDEAPYEYVSQMRASIVVLGPILARNGHA  
60 70 80 90 100 110

120 130 140 150 160 170  
nk603cp4.pep --TFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTTITYRVPMASAQ

Q9A0I4 KVSMPGGCTIGSRPINLHLKGLEAMGATITQKGGD---ITAQADRLQGAMIIY-MDFPSPVG  
120 130 140 150 160

180 190 200 210 220 230  
nk603cp4.pep VKSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLT  
Q9A0I4 ATQNLMAATLADGVTT--IENAAREPEIVDLAQ--FLNKMGARIRGAGTETLTITGVTHLR  
170 180 190 200 210 220

240 250 260 270 280 290  
nk603cp4.pep GQVIDVPGDFSTAFPLVAALLVPGSDVTILNVLNMPTRTGLILTLQEMGADIEVINPRL  
Q9A0I4 GVEHDVVQDRIEAGTFMVAAMTSG--NVLIRDAVWEHNR--LISKLMEMGVSVT-----  
230 240 250 260 270

300 310 320 330 340 350  
nk603cp4.pep AGGEDVADLRVRSST--LKGVTVPEDRAPSMIDIEYPILAVAAFAEGATVMNGLEELRVK  
Q9A0I4 ---EEEGYIRVQANTPKLKPVTVKTLPHPG---FPT-DMQAQFTALMAVNG-ESTMVE  
280 290 300 310 320

360 370 380 390 400 410  
nk603cp4.pep E--SDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSF  
Q9A0I4 TVFENRFQHLEE--MRRMGLQSEILRETAMIHG---GRQL---QGAPVMS-TDLRASAALI  
330 340 350 360 370 380

420 430 440 450  
nk603cp4.pep VMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Q9A0I4 LTGIVAQGVTVNNVLVHLDRGYYQFHEKLAKLGATISRSSEV  
390 400 410 420

nk603cp4.pep  
GP\_BCT1:AE006528\_2

LOCUS AE006528\_2 [AE006528]  
DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 57 of 167 of  
the complete genome;  
Best Blastp hit = sp|P70965|MUA1\_BACSU  
UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 1  
(ENOYLPHYRUVATE TRANSFERASE) (UDP-N-ACETYLGLUCOSAMINE . . .

SCORES Init1: 52 Initn: 52 Opt: 148 z-score: 165.1 E(): 0.32  
Smith-Waterman score: 210; 22.2% identity in 442 aa overlap

10 20 30 40 50 59  
nk603cp4.pep MLHGASSRPATARKSSGSLSGTVRIIPGDKSISHRSFMFGGLAS-GETRITGLLEGEDVINT  
AE006528\_2 MDKIIIEGGQTRLEGEVVEIEGAKNAVLPLLAASILPSKSKTILRNVPIILSDVFTM  
10 20 30 40 50

60 70 80 90 100 110  
nk603cp4.pep GKAMQAMGARI--RKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRITMGLVGVYDFDS  
AE006528\_2 NNVVRGLDIRVDFNEAANEITVDASGH---ILDEAPYEYVSQMRASIVVLGPILARNGHA  
60 70 80 90 100 110

120 130 140 150 160 170  
nk603cp4.pep --TFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTTITYRVPMASAQ  
AE006528\_2 KVSMPGGCTIGSRPINLHLKGLEAMGATITQKGGD---ITAQADRLQGAMIIY-MDFPSPVG  
120 130 140 150 160

180 190 200 210 220 230  
nk603cp4.pep VKSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLT  
AE006528\_2 ATQNLMAATLADGVTT--IENAAREPEIVDLAQ--FLNKMGARIRGAGTETLTITGVTHLR

170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 290  
GQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTQEMGADIEVINPRL  
AE006528\_2 GVEHDVVQDRIEAGTFPMVAAAMTSGLNLRDAVWEHNRPLISKLMEMGVSVT-----  
230 240 250 260 270

nk603cp4.pep 300 310 320 330 340 350  
AGGEDVADLRVRSST--LKGVTVPEDRAPSMIDEYPIILAVAAAFAGATVMNGLEELRVK  
AE006528\_2 ---EEYGIQVQANTPKLKPVTIKTLPHPG---FPT-DMQAQFTALMAVNG-ESTMVE  
280 290 300 310 320

nk603cp4.pep 360 370 380 390 400 410  
E--SDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFL  
AE006528\_2 TVFENRFQHLEE-MRRMGLQSEILRETAMIHG---GRQL---QGAPVMS-TDLRASAALI  
330 340 350 360 370 380

nk603cp4.pep 420 430 440 450  
VMGLVSENFPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AE006528\_2 LTGIVAQQVTVNNLVHLDRGGYQFHEKLAKLGATISRSSEV  
390 400 410 420

nk603cp4.pep  
GP\_BCT2:Z99123\_7

LOCUS Z99123\_7 [BSUB0020]  
DEFINITION Bacillus subtilis complete genome (section 20 of 21): from 3798401  
to 4010550;  
alternate gene name: murZ, lssF, rev-4.  
DATE 26-NOV-1997  
ACCESSION Z99123 . . .

SCORES Init1: 45 Initn: 73 Opt: 147 z-score: 163.9 E(): 0.37  
Smith-Waterman score: 229; 23.4% identity in 444 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPARATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
Z99123\_7 MEKLNIAAGGDSLNGTVHISGAKNSAVALIPATILANSEVTIEGLPEISDIETLR  
10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVY--DFDST  
Z99123\_7 DLLKEIGNVHFNENGEMVVDPTSMISM-----PLPNGKVKK-LRASYYLMGAMLRGFKQA  
60 70 80 90 100

nk603cp4.pep 120 130 140 150 160 170  
FIG---DASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGFKTPTPIYRVPMASAQ  
Z99123\_7 VIGLPGGCHLGRPRIDQHIKGFALGAETVNEQG---AIYLAERLRGARIY-LDVVSVG  
110 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
VKSALLAGLNTPGITTVIEPIIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEGRG  
Z99123\_7 ATINIMLAALAEK-KTIIENAAKEPEIIDVATLLTSMGA---KIKGAGTNVIRIDGVK  
170 180 190 200 210

nk603cp4.pep 240 250 260 270 280 290  
KLTG-QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRT-GLILTQEMGADIEV  
Z99123\_7 ELHGCKHTIIPDRIEAGTFMIAGAAM--GKEVIIDNVI--PTHLESITAKLREMGYHIET  
220 230 240 250 260 270

300 310 320 330 340

nk603cp4.pep INPRL--AGGE-DVADLRVRSSTLKGVTVPEDRAPSMIDEYPIILAVAAAFAGATVM-NG  
Z99123\_7 SDDQLLIVGGQKNLKPVDVKTLLVYPGF--PTDL-----QQPMTALLTR-AKGTSSVVTDT  
280 290 300 310 320

nk603cp4.pep 350 360 370 380 390 400  
LEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRI  
Z99123\_7 IYSARFKHIDELRRMGANMKV-----EGRSAIIT-----GPVELOQAKVKAS-DLRA  
330 340 350 360 370

nk603cp4.pep 410 420 430 440 450  
AMSFLVMGLVSENFPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Z99123\_7 GSCLVVAGLMADGVTEITGLEHIDRGYSSELEKKLEGLGATIWRRERMTDEEIEQLQNS  
380 390 400 410 420

nk603cp4.pep  
GP\_BCT2:Z99122\_207

LOCUS Z99122\_207 [BSUB0019]  
DEFINITION Bacillus subtilis complete genome (section 19 of 21): from 3597091  
to 3809700;  
alternate gene name: murZ, lssF, rev-4.  
DATE 24-JUN-1999  
ACCESSION Z99122 . . .

SCORES Init1: 45 Initn: 73 Opt: 147 z-score: 163.9 E(): 0.37  
Smith-Waterman score: 229; 23.4% identity in 444 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPARATARKSSGLSGTVRIIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
Z99122\_207 MEKLNIAAGGDSLNGTVHISGAKNSAVALIPATILANSEVTIEGLPEISDIETLR  
10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEAPLDFGNAATGCRLTMGLVGVY--DFDST  
Z99122\_207 DLLKEIGNVHFNENGEMVVDPTSMISM-----PLPNGKVKK-LRASYYLMGAMLRGFKQA  
60 70 80 90 100

nk603cp4.pep 120 130 140 150 160 170  
FIG---DASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGFKTPTPIYRVPMASAQ  
Z99122\_207 VIGLPGGCHLGRPRIDQHIKGFALGAETVNEQG---AIYLAERLRGARIY-LDVVSVG  
110 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
VKSALLAGLNTPGITTVIEPIIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEGRG  
Z99122\_207 ATINIMLAALAEK-KTIIENAAKEPEIIDVATLLTSMGA---KIKGAGTNVIRIDGVK  
170 180 190 200 210

nk603cp4.pep 240 250 260 270 280 290  
KLTG-QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRT-GLILTQEMGADIEV  
Z99122\_207 SDDQLLIVGGQKNLKPVDVKTLLVYPGF--PTDL-----QQPMTALLTR-AKGTSSVVTDT  
280 290 300 310 320

nk603cp4.pep 350 360 370 380 390 400  
LEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRI



DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (3- . . .

SCORES Init1: 55 Initn: 86 Opt: 138 z-score: 163.3 E(): 0.4  
 Smith-Waterman score: 138; 31.8% identity in 88 aa overlap

nk603cp4.pep KGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGKLNGVD  
 Q12655 FLTTTILSSIAHKESKPCVTKITGISNQRVKECNRIEAMVCELKKFGIQ  
 10 20 30 40

nk603cp4.pep CDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFVLMGLVSENPTVDDATMIAT  
 Q12655 VKELPDGIYVKSINISNLLPPKNG--VNCYNDHRIAMSFVLAACISPKPIVILNKTVCNK  
 50 60 70 80 90 100

nk603cp4.pep SFPEFMDLMAGLGAKIELSDTKAA  
 Q12655 T

nk603cp4.pep  
 GP\_PLN3:U31054\_1

LOCUS U31054\_1 [PCU31054]  
 DEFINITION Pneumocystis carinii f. sp. oryctolagi (arom) gene, EPSP synthase  
 domain, partial cds;  
 encodes EPSP synthase domain.  
 DATE 17-APR-1996  
 ACCESSION U31054 . . .

SCORES Init1: 55 Initn: 86 Opt: 138 z-score: 163.3 E(): 0.4  
 Smith-Waterman score: 138; 31.8% identity in 88 aa overlap

nk603cp4.pep KGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGKLNGVD  
 U31054\_1 FLTTTILSSIAHKESKPCVTKITGISNQRVKECNRIEAMVCELKKFGIQ  
 10 20 30 40

nk603cp4.pep CDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFVLMGLVSENPTVDDATMIAT  
 U31054\_1 VKELPDGIYVKSINISNLLPPKNG--VNCYNDHRIAMSFVLAACISPKPIVILNKTVCNK  
 50 60 70 80 90 100

nk603cp4.pep SFPEFMDLMAGLGAKIELSDTKAA  
 U31054\_1 T

nk603cp4.pep  
 TREMBL\_MAIN:Q978S3

ID Q978S3 PRELIMINARY; PRT; 410 AA.  
 AC Q978S3;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)  
 DE 5-ENOLPYRUVYLSHIKIMATE 3-PHOSPHATE SYNTHASE. . . .

SCORES Init1: 85 Initn: 197 Opt: 145 z-score: 162.1 E(): 0.47  
 Smith-Waterman score: 309; 23.7% identity in 438 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 Q978S3 MIVKVSNLGGSGIAEMPSSKSTQRYVLASAFLNKSVVLNGITITNDDVAM  
 10 20 30 40 50

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRLTMGLVGVYDFDSTFI  
 Q978S3 RIAESVSGSTITNNRSIKI---SSNFKCFEE-IYVGESGTSYRLSIGLLAASGCVTRIK  
 60 70 80 90 100

nk603cp4.pep GDASLTKRPMGRVLNPLREMVGQVQKSEGDRLPVTLRGPKTPTPITYRVPMASAQVKSAV  
 Q978S3 GEDSLAKRPIEPLLMALGENGVKFERNEAGFYNVDRNSQKKH---VEIEGSSSQFVSSL  
 110 120 130 140 150 160

nk603cp4.pep LLAGLNTPGITTVIEPIIMTRDH---TEKMLQGFGANPTVETDADGVTRIRLEGRKLTGQ  
 Q978S3 MLYYAKKGGGEFTARNIKSIGVYVYITKRVLYDLGYFANIE---RTITINPTG-VWKT  
 170 180 190 200 210

nk603cp4.pep VIDVPGDPSSTAPFLVAALLVPGSDVTILNVLNMPTRTGLILTLQEMGADIEVI-NPRLA  
 Q978S3 AIDVEPDYSSMAFFMVLGLLSDSDVDFR-----RIKRISRIQPDVSILDFKNNILI  
 220 230 240 250 260

nk603cp4.pep GGEDVDALRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESD  
 Q978S3 NGEIE---RVISGINEPVSVDADMNP---DLCPELSVIGIFSQYQIRNYERLKTKEKN  
 270 280 290 300 310 320

nk603cp4.pep RLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFVLMGLV  
 Q978S3 RYEGIIDLAERFGANVEDNGQDLFIK--P-----GSVRFDPVISYKDHRIAMASIASLI  
 330 340 350 360 370

nk603cp4.pep SENPVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
 Q978S3 GGFPTVIENAEKTAKSFPFGFAELSKFANVEELA  
 380 390 400 410

nk603cp4.pep  
 GP\_BCT2:AP000996\_78

LOCUS AP000996\_78 [AP000996]  
 DEFINITION Thermoplasma volcanium genomic DNA, complete sequence, section 6/6;  
 DATE 07-JUN-2001  
 ACCESSION AP000996  
 NID  
 ORGANISM Thermoplasma volcanium . . .

SCORES Init1: 85 Initn: 197 Opt: 145 z-score: 162.1 E(): 0.47  
 Smith-Waterman score: 309; 23.7% identity in 438 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
 AP000996\_78 MIVKVSNLGGSGIAEMPSSKSTQRYVLASAFLNKSVVLNGITITNDDVAM  
 10 20 30 40 50 60

```

10      20      30      40      50
nk603cp4.pep 70      80      90      100     110     120
KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRITMGLVGVYDFDSTFI
AP000996_78 RIAESVSGSTITINNRSIKI-----SSNFKCPEE-IYVGESGTSYRLSIGLLAASGCVTRIK
60      70      80      90      100

nk603cp4.pep 130     140     150     160     170     180
GDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTITRVPMASAVQKSAV
AP000996_78 GEDSLAKRPIEPLLMALGENGKVKFERNEAGFYNVNDRNSQKKH---VEIEGSSSQFVSSL
110     120     130     140     150     160

nk603cp4.pep 190     200     210     220     230
LLAGLNTPGITTVIEPIIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEGRGKLTGQ
AP000996_78 MLYYAKKGGGEFTARNIKSIGVYITKRVLYDLGYFANIE-----RTITINPTG-VWKT
170     180     190     200     210

nk603cp4.pep 240     250     260     270     280     290
VIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVI-NPRLA
AP000996_78 AIDVPEPYSSMAFFMVLGLLSDSDVDRF-----RIKRISRIQPDSDVILDLFKNNILI
220     230     240     250     260

nk603cp4.pep 300     310     320     330     340     350
GGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGATVMNGLEELRVKESD
AP000996_78 NGEEL---RVISGINEPVSVDADMNP---DLCPPLSVIGIFSQYGVQRNRYERLKTESN
270     280     290     300     310     320

nk603cp4.pep 360     370     380     390     400     410
RLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLFVGMGLV
AP000996_78 RYEGIIDLAERFGANVEDNGQDLFIK--P-----GSVRFPPDVISYKDHRRMMAASIASLI
330     340     350     360     370

nk603cp4.pep 420     430     440     450
SENPVTVDATMIATSFPEFMDLMAGLGAKEIELSDTKAA
AP000996_78 GGFPTVIENAEKTAQSFPGFFAELSKFANVEELA
380     390     400     410

nk603cp4.pep
GP_BCT3:AJ295297_1

LOCUS      AJ295297_1 [MCH295297]
DEFINITION Mycobacterium chelonae murA gene for UDP-N-acetylglucosamine
              enolpyruvyl transferase.
DATE       06-JAN-2001
ACCESSION  AJ295297
NID . . .

SCORES      Init1: 62      Initn: 88      Opt: 140      z-score: 156.5 E(): 0.96
Smith-Waterman score: 241;      24.3% identity in 440 aa overlap

nk603cp4.pep 10      20      30      40      50      60
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AJ295297_1    MSEHFLVTGGARLTGNVAVTGAKNSVLKLMAAALLAEGTSTITNCPDILDVPLMA
10      20      30      40      50

nk603cp4.pep 70      80      90      100     110
KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNA---ATGCRITMGLVGV-VYDF
AJ295297_1    DVLRGLGATVELEGDVVRITSPDE-----PKYEADFAAVRQFRASVCLVG-PLVGRCRRA
60      70      80      90      100

nk603cp4.pep 120     130     140     150     160     170
DSTFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTITRVPMASAV
AJ295297_1    KVALPGGDAIRSRPLDMHQAGLRQLGAECTIEHG---CVVAEAKALHGADIQLEFPSPVG
110     120     130     140     150     160

nk603cp4.pep 180     190     200     210     220     230
VKSALLAGLNTPGITTVIEPIIMTRDHTE---KMLQGFGANPTVETDADGVRTIRLEGRGK

```

```

120     130     140     150     160     170
nk603cp4.pep DSTFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTITRVPMASAV
AJ295297_1    KVALPGGDAIRSRPLDMHQAGLRQLGAECTIEHG---CVVAEAKALHGADIQLEFPSPVG
110     120     130     140     150     160

nk603cp4.pep 180     190     200     210     220     230
VKSALLAGLNTPGITTVIEPIIMTRDHTE---KMLQGFGANPTVETDADGVRTIRLEGRGK
AJ295297_1    ATENILMAAVLADGVTIHNAAAREPDVVDLCTMLVQMGA---QIEGAGTSTLTIVIGVPK
170     180     190     200     210     220

nk603cp4.pep 240     250     260     270     280     290
LTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLIL-TLQEMGADIEVIN
AJ295297_1    LHPTEHRVIGDRIVAATWGLAAAMTRG-DVTVTGV--DPEHLQLVLHLHDAGATVTQND
230     240     250     260     270

nk603cp4.pep 300     310     320     330     340     350
PRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEGAT-VMNGLEELR
AJ295297_1    ---NGFRVVQYE-RPKAINVATLFFPGFPT--DLQPMAGLAAIADGTSMITENVFEAR
280     290     300     310     320     330

nk603cp4.pep 360     370     380     390     400     410
VKESDRLSAVANGKLKNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSLF
AJ295297_1    FRFVEKM---MRL-GADARTDGHAVVRG-----IRQLSSAPVWSS-DIRAGAGLV
340     350     360     370

nk603cp4.pep 420     430     440     450
VMGLVSENPTVDATMIATSFPEFMDLMAGLGAKEIELSDTKAA
AJ295297_1    LPGLVADGVTEVYDVPHIDRGYPLFVENLQSLGAEVERVS
380     390     400     410

nk603cp4.pep
TREMBL_MAIN:Q9EXE3

ID      Q9EXE3      PRELIMINARY;      PRT;      417 AA.
AC      Q9EXE3;
DT      01-MAR-2001 (TREMBLrel. 16, Created)
DT      01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE      UDP-N-ACETYLGLUCOSAMINE ENOLPYRUVYL TRANSFERASE. . . .

```

```

SCORES      Init1: 62      Initn: 88      Opt: 140      z-score: 156.5 E(): 0.96
Smith-Waterman score: 241;      24.3% identity in 440 aa overlap

nk603cp4.pep 10      20      30      40      50      60
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
Q9EXE3      MSEHFLVTGGARLTGNVAVTGAKNSVLKLMAAALLAEGTSTITNCPDILDVPLMA
10      20      30      40      50

nk603cp4.pep 70      80      90      100     110
KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNA---ATGCRITMGLVGV-VYDF
Q9EXE3      DVLRGLGATVELEGDVVRITSPDE-----PKYEADFAAVRQFRASVCLVG-PLVGRCRRA
60      70      80      90      100

nk603cp4.pep 120     130     140     150     160     170
DSTFIGDASLTKRPMGRVLNPLREMGVQVKSSEDGRLPVTLRGPKTPTPTITRVPMASAV
Q9EXE3      KVALPGGDAIRSRPLDMHQAGLRQLGAECTIEHG---CVVAEAKALHGADIQLEFPSPVG
110     120     130     140     150     160

nk603cp4.pep 180     190     200     210     220     230
VKSALLAGLNTPGITTVIEPIIMTRDHTE---KMLQGFGANPTVETDADGVRTIRLEGRGK

```

```

10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRI PGDKSISHSRFMGGLASGETRI TGLLEGEDVINTG
Q9KP62      MEKFRVIGSTQPLQGEVTISGAKNAALFILFASILAE EEPVEVANVPHLRIDIDITM
10      20      30      40      50

70      80      90      100     110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGLLAPEAPLDFGNAATGCRLTMG- LVGVY-DFPDST
Q9KP62      ELLERLIGAKVERNGSV----HVDAGPINQYCAPYDLVKTMRASI WALGELVARFGQGQVS
60      70      80      90      100     110

120     130     140     150     160     170
nk603cp4.pep FIGDASLTKRPMGRVLNPLIREMGVQVKSEDDRLPVTLRGPKTPTPTIY-RVFM-ASAQV
Q9KP62      LPGGCAIGARPVDLHITGLEQLGATITLEDG-YVKAHV DGR LQGAHIVMDKVS VGATITII
120     130     140     150     160     170

180     190     200     210     220     230
nk603cp4.pep KSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFGANPTVETDADGVRTIRLEGRGKLTG
Q9KP62      MCAATLAEGTTVLNDNAAREPEIV--DTAMFLNKLGA----KISGAGTDSITIEGVERLGG
180     190     200     210     220

240     250     260     270     280     290
nk603cp4.pep QVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGADIEVINPRLA
Q9KP62      GKHAVVPDRIETGTFLVAAAVSRGK-IVCRNTHAHLLA-LVAKLEEA GAETEC-----

```

```

                230          240          250          260          270
nk603cp4.pep   GGEDVADLRVRSSSTLKGVTV---PEDRAPS-MIDEYPILAVAAAFAGATVMNGLEELR-
                ||| : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : |
Q9KP62         -GEDWISLDMTGRELKAVTVRTAPHGFPPTDMAQAQTLLNMMAK--GGGVITETIFENRF
                280          290          300          310          320          330

                300          310          320          330          340          350
nk603cp4.pep   GGEDVADLRVRSSSTLKGVTV---PEDRAPS-MIDEYPILAVAAAFAGATVMNGLEELR-
                ||| : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : |
Q9KP62         -GEDWISLDMTGRELKAVTVRTAPHGFPPTDMAQAQTLLNMMAK--GGGVITETIFENRF
                280          290          300          310          320          330

                360          370          380          390          400          409
nk603cp4.pep   --VKESDRLSAVANGLKLNGVDCDEGETSLVVVRGRPDGKLGNAASGAAVATHLDHRIAMS
                ::| : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : |
Q9KP62         MHVPFLKRMGAKA-----EIEGNTVICGDVD----RLSQAQYMA-TDLRASAS
                340          350          360          370

                410          420          430          440          450
nk603cp4.pep   FLVMGLVSENPVTVDATTMIATSFPEFMDLMAGLGAKIELSDTKAA
                ::: : | : | : ||| : | : | : | : | : | : | : | : | : | : | : | : |
Q9KP62         LVIAGCIAKGETIVDRIYIHDRGYERIEDKLSALGANIERFRD
                380          390          400          410

nk603cp4.pep   GP_BCT1:AE004321_8

LOCUS      AE004321_8 [AE004321]
DEFINITION Vibrio cholerae chromosome I, section 229 of 251 of the complete
            chromosome;
            similar to GB:M92358 SP:P28909 PID:146902
            PID:606127 GB:U00096; identified by sequence similarity;
            putative. . . .
```

SCORES Init1: 50 Initn: 115 Opt: 138 z-score: 154.4 E(): 1.3  
Smith-Waterman score: 219; 24.4% identity in 442 aa overlap

[illegible]

nk603cp4.pep --VKESDRLSAVANGKLNGLVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
AE004321\_8 MHVPELKRMGAKA-----EIEGNTVICGDVD---RLSGAQVMA-TDLRASAS  
340 350 360 370

nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AE004321\_8 LVIAGCIAKETIVDRIYHIDRGYERIEDKLSALGANIERFRD  
380 390 400 410

nk603cp4.pep  
SWISSPROT:MURA\_BUCAI

ID MURA\_BUCAI STANDARD; PRT; 416 AA.  
AC P57466;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7) . . .

SCORES Init1: 39 Initn: 39 Opt: 137 z-score: 153.3 E(): 1.5  
Smith-Waterman score: 150; 21.3% identity in 441 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
MURA\_BUCAI MEKLYVEGNKILNGHVIIISGSKNAALPILFMTILTEGKIKIGNIPNLTIDINIAL  
10 20 30 40 50 60

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDFGNAATGCRLTMG-LVGVDYDFDSTF  
MURA\_BUCAI KLLVYLGVKITGN-ETLCIDA-SSINIFCP--PYNLINKIRASIWMGLPGLARFGKAKIF  
70 80 90 100 110 119

nk603cp4.pep I-GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITY-RVPM-ASAQV  
MURA\_BUCAI LPGAQCKIGSRPIDLHLNGLTQLGATINLKN-NCIDAYVKGRQLQGKYILMEKISVGATITI  
120 130 140 150 160 170

nk603cp4.pep KSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFANPTVETDADGVRTIRLEGRGKLTG  
MURA\_BUCAI MSAATLAKGSTIIDNAACEPEIV--DIAKFLNTLGADII----GAGSNKICIKGVKLKLTG  
180 190 200 210 220 230

nk603cp4.pep QVIDVPGDPSSSTAFPLVAALLVPGSDVTILNVLNPNTR-TGLILTLQEMGADIEVINPRL  
MURA\_BUCAI GTHQVIPDRIETGTPLVAAAASQGH-ITCHKT--EPKHLTNVLMKLTGAGAKIKT-----  
240 250 260 270 280 290

nk603cp4.pep AGGEDVA--DLR-VRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEG-ATVMNGLEELR  
MURA\_BUCAI --GKDWKILDMRGKRPKSLNICTAPYPGFPT--DMQAQFALLNSISKIGITITETIFENR  
300 310 320 330 340 350

nk603cp4.pep VKESDRLSAVANGKLNGLVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFL  
MURA\_BUCAI FIYTSSELIRMGAKIKIKN-----NTIICYGIP--KLISNVFSS-----DLRASATLI  
360 370 380 390 400 410

nk603cp4.pep VMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
MURA\_BUCAI LAGCIAAGITIVNHTYHLVRGYESFPKLNKIGANIKII  
380 390 400 410

MURA\_BUCAI : | : : : | : : : : | : : | : : :  
LAGCIAAGITIVNHTYHLVRGYESFPKLNKIGANIKII  
380 390 400 410

nk603cp4.pep  
GP\_BCT2:AP001119\_68

LOCUS AP001119\_68 [AP001119]  
DEFINITION Buchnera sp. APS genomic DNA, complete sequence, segment 2/2;  
BU386.  
DATE 07-SEP-2000  
ACCESSION AP001119  
NID . . .

SCORES Init1: 39 Initn: 39 Opt: 137 z-score: 153.3 E(): 1.5  
Smith-Waterman score: 150; 21.3% identity in 441 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AP001119\_68 MEKLYVEGNKILNGHVIIISGSKNAALPILFMTILTEGKIKIGNIPNLTIDINIAL  
10 20 30 40 50 60

nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLLAPEALDFGNAATGCRLTMG-LVGVDYDFDSTF  
AP001119\_68 KLLVYLGVKITGN-ETLCIDA-SSINIFCP--PYNLINKIRASIWMGLPGLARFGKAKIF  
70 80 90 100 110 119

nk603cp4.pep I-GDASLTKRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITY-RVPM-ASAQV  
AP001119\_68 LPGAQCKIGSRPIDLHLNGLTQLGATINLKN-NCIDAYVKGRQLQGKYILMEKISVGATITI  
120 130 140 150 160 170

nk603cp4.pep KSAVLLAGLNTPGITTVIEPIIMTRDHTKMLQGFANPTVETDADGVRTIRLEGRGKLTG  
AP001119\_68 MSAATLAKGSTIIDNAACEPEIV--DIAKFLNTLGADII----GAGSNKICIKGVKLKLTG  
180 190 200 210 220 230

nk603cp4.pep QVIDVPGDPSSSTAFPLVAALLVPGSDVTILNVLNPNTR-TGLILTLQEMGADIEVINPRL  
AP001119\_68 GTHQVIPDRIETGTPLVAAAASQGH-ITCHKT--EPKHLTNVLMKLTGAGAKIKT-----  
240 250 260 270 280 290

nk603cp4.pep AGGEDVA--DLR-VRSSTLKGVTVPEDRAPSMIDEYPILAVAAFAEG-ATVMNGLEELR  
AP001119\_68 --GKDWKILDMRGKRPKSLNICTAPYPGFPT--DMQAQFALLNSISKIGITITETIFENR  
300 310 320 330 340 350

nk603cp4.pep VKESDRLSAVANGKLNGLVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMSFL  
AP001119\_68 FIYTSSELIRMGAKIKIKN-----NTIICYGIP--KLISNVFSS-----DLRASATLI  
360 370 380 390 400 410

nk603cp4.pep VMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
AP001119\_68 LAGCIAAGITIVNHTYHLVRGYESFPKLNKIGANIKII  
380 390 400 410

nk603cp4.pep  
TREMBL\_MAIN:Q9RVA6

ID Q9RVA6 PRELIMINARY; PRT; 426 AA.  
AC Q9RVA6;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE. . . .

SCORES Init1: 49 Initn: 83 Opt: 134 z-score: 149.9 E(): 2.2  
Smith-Waterman score: 203; 22.0% identity in 441 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
Q9RVA6      MQLTPLHVTGGRQLGGEIAYQHSKNAALPIIVATLLSSEPITLHGIPRLSDVYTIL
      10      20      30      40      50

      70      80      90      100     110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLAPE-----APLDFGNAATGCRLTMGLVGVYDF
Q9RVA6      ELMHHLGTR-----HAWV--GPNSLTLTHTPEIENTDAPYALVSKMRASFIVLGAILARAG
      60      70      80      90      100

      120     130     140     150     160     170
nk603cp4.pep DST--FIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMAS
Q9RVA6      TATVSMPPGGCAWGPFPVDQHVKALRALGAEEVI--EDGNGFFAHREGSLNGSFVPELLTVGG
      110     120     130     140     150     160

      180     190     200     210     220     230
nk603cp4.pep AQVKSAVLLAGLNTPGITTVIEPIIMTRDHT--KMLQGFGANPTVETDADGVRTIRLEGR
Q9RVA6      TH--NAILAAALGE-GTVTLENASIDTDVVDVIEFLNALGA----RIEGAGTNTTITHGV
      170     180     190     200     210     220

      240     250     260     270     280     290
nk603cp4.pep GKLTG-QVIDVPGDPSSTAPPLVAALLVPGSDVTILNVLMNPTRTGLILTLQEMGADIEV
Q9RVA6      KELHGGGEYTVIPDRIEAGTFMMLAA--ATRSRLTLTNVRPDHLRA-VTSKLAEMGVDIQE
      230     240     250     260     270

      300     310     320     330     340     349
nk603cp4.pep INPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSM-IDEYPILAVAAAFAGATVMNGLEE
Q9RVA6      DGNRM-----VVDARGRD--LKPVNVTTSYPGFPDLPQMSALLATVPGSTSVIQ--DP
      280     290     300     310     320

      350     360     370     380     390     400     409
nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS
Q9RVA6      VY---PDLRTHVAELHRMGATITVSGYTQ-VIQG-----GTLHAAPVKAA---DLRAGAA
      330     340     350     360     370

      410     420     430     440     450
nk603cp4.pep FLVMGLVSENPTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA
Q9RVA6      LFIAGLTCEGETVIDGVQYLNRYERLAERLRGIGGEVVMQPEPMLAADD
      380     390     400     410     420

```

nk603cp4.pep  
GP\_BCT1:AE001962\_5

LOCUS AE001962\_5 [AE001962]  
DEFINITION Deinococcus radiodurans R1 section 99 of the complete  
chromosome 1;  
similar to SP:P19670 PID:143432 PID:853767  
GB:AL009126 percent identity: 66.11; identified by  
sequence similarity; putative. . . .

SCORES Init1: 49 Initn: 83 Opt: 134 z-score: 149.9 E(): 2.2  
Smith-Waterman score: 203; 22.0% identity in 441 aa overlap

```

      10      20      30      40      50      60
nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG
AE001962_5    MQLTPLHVTGGRQLGGEIAYQHSKNAALPIIVATLLSSEPITLHGIPRLSDVYTIL
      10      20      30      40      50

      70      80      90      100     110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDGVNGGGLAPE-----APLDFGNAATGCRLTMGLVGVYDF
AE001962_5    ELMHHLGTR-----HAWV--GPNSLTLTHTPEIENTDAPYALVSKMRASFIVLGAILARAG
      60      70      80      90      100

      120     130     140     150     160     170
nk603cp4.pep DST--FIGDASLTKRPMGRVLNPLREMGVQVKSSEDGDRLPVTLRGPKTPTPTITYRVPMAS
AE001962_5    TATVSMPPGGCAWGPFPVDQHVKALRALGAEEVI--EDGNGFFAHREGSLNGSFVPELLTVGG
      110     120     130     140     150     160

      180     190     200     210     220     230
nk603cp4.pep AQVKSAVLLAGLNTPGITTVIEPIIMTRDHT--KMLQGFGANPTVETDADGVRTIRLEGR
AE001962_5    TH--NAILAAALGE-GTVTLENASIDTDVVDVIEFLNALGA----RIEGAGTNTTITHGV
      170     180     190     200     210     220

      240     250     260     270     280     290
nk603cp4.pep GKLTG-QVIDVPGDPSSTAPPLVAALLVPGSDVTILNVLMNPTRTGLILTLQEMGADIEV
AE001962_5    KELHGGGEYTVIPDRIEAGTFMMLAA--ATRSRLTLTNVRPDHLRA-VTSKLAEMGVDIQE
      230     240     250     260     270

      300     310     320     330     340     349
nk603cp4.pep INPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSM-IDEYPILAVAAAFAGATVMNGLEE
AE001962_5    DGNRM-----VVDARGRD--LKPVNVTTSYPGFPDLPQMSALLATVPGSTSVIQ--DP
      280     290     300     310     320

      350     360     370     380     390     400     409
nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS
AE001962_5    VY---PDLRTHVAELHRMGATITVSGYTQ-VIQG-----GTLHAAPVKAA---DLRAGAA
      330     340     350     360     370

      410     420     430     440     450
nk603cp4.pep FLVMGLVSENPTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA
AE001962_5    LFIAGLTCEGETVIDGVQYLNRYERLAERLRGIGGEVVMQPEPMLAADD
      380     390     400     410     420

nk603cp4.pep
GP_PLN3:U31055_1

LOCUS      U31055_1 [PCU31055]
DEFINITION Pneumocystis carinii f. sp. hominis (arom) gene, EPSP synthase
domain, partial cds;
encodes EPSP synthase domain.

DATE      17-APR-1996
ACCESSION U31055 . . .

SCORES Init1: 55 Initn: 55 Opt: 125 z-score: 149.3 E(): 2.4
Smith-Waterman score: 125; 29.1% identity in 103 aa overlap

      300     310     320     330     340     350
nk603cp4.pep DVADLRVRSSTLKGVTVPEDRAPSMID EYPILAVAAAFAGA---TVMNGLEE LRVKESD
U31055_1      FLTTTILA-SVAYKESKPCITKITGISNQRKECN
      10      20      30

```



nk603cp4.pep 360 370 380 390 400 410  
RLSAVANGKLNGVDCDEGETSLVVRGPRDGLGNASGAAVATHLDHRIAMSFVLMGLV  
U31055\_1 RIKAMVIELAKFGIEAGELPDGIYVKSLLDISNLLSPKNG--INCYNDRHRIAMSFVSLACI  
40 50 60 70 80 90

nk603cp4.pep 420 430 440 450  
SENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
U31055\_1 SPKPTIILGKTCVNKT  
100

nk603cp4.pep  
TREMBL\_MAIN: Q12656

ID Q12656 PRELIMINARY; PRT; 108 AA.  
AC Q12656;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE 3-PHOSPHOSHIKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (3- . . .

SCORES Init1: 55 Initn: 55 Opt: 125 z-score: 149.3 E(): 2.4  
Smith-Waterman score: 125; 29.1% identity in 103 aa overlap

nk603cp4.pep 300 310 320 330 340 350  
DVADLRVRSSTLKGVTVPEDRAPSMIDEYPIILAVAAFAEGA---TVMNGLEELRVKESD  
Q12656 FLTTTILA-SVAYKESKPCITKITGISNQRKCEK  
10 20 30

nk603cp4.pep 360 370 380 390 400 410  
RLSAVANGKLNGVDCDEGETSLVVRGPRDGLGNASGAAVATHLDHRIAMSFVLMGLV  
Q12656 RIKAMVIELAKFGIEAGELPDGIYVKSLLDISNLLSPKNG--INCYNDRHRIAMSFVSLACI  
40 50 60 70 80 90

nk603cp4.pep 420 430 440 450  
SENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
Q12656 SPKPTIILGKTCVNKT  
100

nk603cp4.pep  
GP\_BCT3:D64000\_90

LOCUS D64000\_90 [SYCSLRB]  
DEFINITION Synechocystis sp. PCC6803 complete genome, 19/27, 2392729-2538999;  
ORF\_ID:slr0017.  
DATE 13-FEB-1999  
ACCESSION D64000  
NID . . .

SCORES Init1: 79 Initn: 105 Opt: 133 z-score: 148.7 E(): 2.6  
Smith-Waterman score: 227; 23.0% identity in 443 aa overlap

nk603cp4.pep 10 20 30 40 50  
MLHGASSRPATARKSSGLSGTVRIPGDKSISHSRSMFGG-LASGETRITGLLE  
D64000\_90 MLQVATPTATEETQTVEIEGRASLKGVEVRISGAKN-SALAIMAGTILCSDDCRLSNLPA  
10 20 30 40 50

nk603cp4.pep 60 70 80 90 100 109  
GEDVINTGKAMQAMGARIRKEGDTWIIDG--VNGGGLLAPEAPLDFGNAATGCRLTMG-L  
D64000\_90 LADIDKMCQILAAIGVNLERDGLRVVDSSNVGHG----P-APYELVSQLRASFFVIGPL

60 70 80 90 100 110  
nk603cp4.pep 110 120 130 140 150 160  
VGVDYDFDST-FIGDASLTKRPMGRVLNPLREMGVQVKSEDDRLPVTLRG-PKTPTPTITY  
D64000\_90 LSRLGMTKVPLPGGCAIGTRPVDLHVRLQAMGADVHIEHG-VVNASIKGHSRRLTGAKI  
120 130 140 150 160 170

nk603cp4.pep 170 180 190 200 210 220  
RVFMASAVQKSAVLLAGLNTPTGITTVEIPIIMTRDHTKMLQGFANPTVETDADGVRTIR  
D64000\_90 YLDYPSVGATETILMAATLAEG-ETVIDNA-AREPEIVDLANFCRSMGAQIRGDGSSRII  
180 190 200 210 220 230

nk603cp4.pep 230 240 250 260 270 280  
LEGRGKLTGQVIDVPGDPSST-AFPLVAALLVPGSDVTILNVLMNPTRTGLILTLQEMGA  
D64000\_90 INGVEKL--HSTDFPIIPRIEAAATFLCAGAITHSISLFPVVDHLASA-IAKLREIGP  
240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340  
DIEVINPRLAGGEDVADLRVRSSTLKGV---TVPEDRAPSMIDEYPIILAVAAFAEGATV  
D64000\_90 EVAIDAP-----DRVRLIPRD--LRGTDIETLPYPGFPT--DMQAQFMALLAVSEGNVS  
290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400  
MN-GLEELRVKESDRLSAVANGKLNGVDCDEGETSLVVRGPRDGLGNASGAAVATHL  
D64000\_90 VTTETVENRLQHVAEQLQRMGANIKLKG-----NAAFIQGVF-----FLSGAPVMS-T  
340 350 360 370 380

nk603cp4.pep 410 420 430 440 450  
DHRIAMSFVLMGLVSENPVTVDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
D64000\_90 DLRAAALVIAGLAAEGKTIQVGLRHLDGRYEDIEGKLRKLGRITRIEQPVTL  
390 400 410 420 430

nk603cp4.pep  
SWISSPROT: MURA\_SYNY3

ID MURA\_SYNY3 STANDARD; PRT; 438 AA.  
AC Q55673;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.7) . . .

SCORES Init1: 79 Initn: 105 Opt: 133 z-score: 148.7 E(): 2.6  
Smith-Waterman score: 227; 23.0% identity in 443 aa overlap

nk603cp4.pep 10 20 30 40 50  
MLHGASSRPATARKSSGLSGTVRIPGDKSISHSRSMFGG-LASGETRITGLLE  
MURA\_SYNY3 MLQVATPTATEETQTVEIEGRASLKGVEVRISGAKN-SALAIMAGTILCSDDCRLSNLPA  
10 20 30 40 50

nk603cp4.pep 60 70 80 90 100 109  
GEDVINTGKAMQAMGARIRKEGDTWIIDG--VNGGGLLAPEAPLDFGNAATGCRLTMG-L  
MURA\_SYNY3 LADIDKMCQILAAIGVNLERDGLRVVDSSNVGHG----P-APYELVSQLRASFFVIGPL  
60 70 80 90 100 110

nk603cp4.pep 110 120 130 140 150 160  
VGVDYDFDST-FIGDASLTKRPMGRVLNPLREMGVQVKSEDDRLPVTLRG-PKTPTPTITY  
MURA\_SYNY3 LSRLGMTKVPLPGGCAIGTRPVDLHVRLQAMGADVHIEHG-VVNASIKGHSRRLTGAKI  
120 130 140 150 160 170

nk603cp4.pep 170 180 190 200 210 220  
RVPMSAQVKSALLAGLNTPGITTVIEPIIMTRDHEKMLQGFANPTVETDADGVRTIR  
MURA\_SYNY3 YLDYPSVGATETILMAATLAEG-ETVIDNA-AREPEIVDLANFCRSMGAQIRGDSSRII  
180 190 200 210 220 230

nk603cp4.pep 230 240 250 260 270 280  
LEGRGKLTGQVIDVPGDPSST-AFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEMGA  
MURA\_SYNY3 INGVEKL--HSTDFFIIPDRIEAATFLCAGAITHSEISLFPVVPDLASA-IAKLREIGP  
240 250 260 270 280

nk603cp4.pep 290 300 310 320 330 340  
DIEVINPRLAGGEDVADLRVRSSTLKGVT--TPEDRAPSMIDEYPI-LAVAAAF-AEAGATV  
MURA\_SYNY3 EVAIDAP-----DRVRLIPRD--LRGTDIETLPYPGFPT--DMQAQFMALLAVSEGNV  
290 300 310 320 330

nk603cp4.pep 350 360 370 380 390 400  
MN-GLEELRVKESDRLSAVANGKLKNGVDCDEGETSLVVGRPDGKGLGNASGA-AVATHL  
MURA\_SYNY3 VTETVFENRLQHVAELQRMGANIKLKG-----NAAFIQGVF-----FLSGAPVMS-T  
340 350 360 370 380

nk603cp4.pep 410 420 430 440 450  
DHRIAMSLVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEI-ELSDTKAA  
MURA\_SYNY3 DLRSAAALVIAAGLAEGKTIVQGLRHLDRGYEDIEGKLRKLGAKLT-RIEQPVTL  
390 400 410 420 430

nk603cp4.pep  
SWISSPROT:RPOA\_SHESP

ID RPOA\_SHESP STANDARD; PRT; 328 AA.  
AC P74963;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE DNA-DIRECTED RNA POLYMERASE ALPHA CHAIN (EC 2.7.7.6) (TRANSCRIPTASE . . .

SCORES Init1: 56 Initn: 56 Opt: 131 z-score: 148.4 E(): 2.7  
Smith-Waterman score: 131; 28.1% identity in 128 aa overlap

nk603cp4.pep 230 240 250 260 270 280  
TIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEM  
RPOA\_SHESP VEIDGVLHEYSSIEGVQEDVLEILLNLKGLAIKLEGKDEALLTLTKSGTGFVLAGDITHD  
60 70 80 90 100 110

nk603cp4.pep 290 300 310 320 330  
GADIEVINPR-----LAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI--LAVAAAF  
RPOA\_SHESP G-DVEIINPEHVIRNLTGQADIS-MRIRVAKGRGYVPASTRIQSDDEERPIGRLLVDAAF  
120 130 140 150 160 170

nk603cp4.pep 340 350 360 370 380 390  
AEGATVMNGLEELRVKESDRLSAVANGKLKNG-VDCDEGETSLVVGRPDGKGLGNASGA  
RPOA\_SHESP SPVSRIAYNVESARVEQRTNLDKLVDMETNGTLDPEEAIRRAATILAEQLDAFVLDLDRV  
180 190 200 210 220 230

nk603cp4.pep 400 410 420 430 440 450  
AVATHLDHRIAMSLVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEI-ELSDTKAA  
RPOA\_SHESP TEPEQKEEKPEFDPILLRPVDDLELTVRSANCLKAEAIHYIGDLVQRTEVELLKTPTNLGK  
240 250 260 270 280 290

nk603cp4.pep

GP\_BCT3:D83194\_1

LOCUS D83194\_1 [D83194]  
DEFINITION Barophilic bacterium gene encoding RNA polymerase alpha subunit,  
complete cds.  
DATE 06-FEB-1999  
ACCESSION D83194  
NID . . .

SCORES Init1: 56 Initn: 56 Opt: 131 z-score: 148.4 E(): 2.7  
Smith-Waterman score: 131; 28.1% identity in 128 aa overlap

nk603cp4.pep 230 240 250 260 270 280  
TIRLEGRGKLTGQVIDVPGDPSSTAFPLVAALLVPGSDVTILNVLNPNTRTGLILTLQEM  
D83194\_1 VEIDGVLHEYSSIEGVQEDVLEILLNLKGLAIKLEGKDEALLTLTKSGTGFVLAGDITHD  
60 70 80 90 100 110

nk603cp4.pep 290 300 310 320 330  
GADIEVINPR-----LAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI--LAVAAAF  
D83194\_1 G-DVEIINPEHVIRNLTGQADIS-MRIRVAKGRGYVPASTRIQSDDEERPIGRLLVDAAF  
120 130 140 150 160 170

nk603cp4.pep 340 350 360 370 380 390  
AEGATVMNGLEELRVKESDRLSAVANGKLKNG-VDCDEGETSLVVGRPDGKGLGNASGA  
D83194\_1 SPVSRIAYNVESARVEQRTNLDKLVDMETNGTLDPEEAIRRAATILAEQLDAFVLDLDRV  
180 190 200 210 220 230

nk603cp4.pep 400 410 420 430 440 450  
AVATHLDHRIAMSLVLMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKEI-ELSDTKAA  
D83194\_1 TEPEQKEEKPEFDPILLRPVDDLELTVRSANCLKAEAIHYIGDLVQRTEVELLKTPTNLGK  
240 250 260 270 280 290

nk603cp4.pep  
SWISSPROT:MUA1\_LACLA

ID MUA1\_LACLA STANDARD; PRT; 421 AA.  
AC Q9CIP4;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYVINYLTRANSFERASE 1 (EC 2.5.1.7) . . .

SCORES Init1: 55 Initn: 55 Opt: 132 z-score: 147.8 E(): 2.9  
Smith-Waterman score: 165; 21.3% identity in 447 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
MUA1\_LACLA MKKIVINGGKRISGTIPISGAKNSVVALIPATILANDVVTLEGVPDISDVASLV  
10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
KAMQAMGARIRK--EGDTWIIDGVNGGGLLAPEAPLDFG--NAATGCRLTMG-LVGVYDF  
MUA1\_LACLA EIMEIMGAKIERNLEEGRLVIDTRS-----VVSRLPYGKINSLSRASYFNGALLGRFGQ  
60 70 80 90 100

nk603cp4.pep 120 130 140 150 160 170  
DST-FIGDASLTKRPMGRVLNPLREMGVQ-VKSEGDRLFVTLRGPKTPTPTITRVPMAS  
MUA1\_LACLA ATVGLPGGCDLGPRTDLHDKAFKAGKIQEAEHLEV-IGDSLVTGTTIYMDVVSVG  
110 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230

nk603cp4.pep AOVKSALLAGLNTPGITTVIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEG  
 MUA1\_LACLA ATINT--MLAASRAKGLT-I IENAAREPEIIDVATLINMGA---QVRGAGTDIIRITG  
 170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280  
 RGKLTG-QVIDVPGD-PSSTAPFLVAALLVPGSDVTILNVLNPNTRTGLIITLQEMGADI  
 MUA1\_LACLA VDEMGAQHTVIPDRIEAGTYLALAAAM---GDGVIIENVIYEHLES-FIAKLEEMGVGL  
 230 240 250 260 270

nk603cp4.pep 290 300 310 320 330 340  
 EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEY--PILAVAAFAEG-ATVMN  
 MUA1\_LACLA TI---R---EDSIEVH-KSENLSKNITSVPYPGFATDLQOPITPLLLK-AKGRGSIVD  
 280 290 300 310 320

nk603cp4.pep 350 360 370 380 390 400  
 GLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHR  
 MUA1\_LACLA TIYQKRVNHVPELARM--GANISVLD-DR---IIYDAPN---ELTGSCVQA-TDLR  
 330 340 350 360 370

nk603cp4.pep 410 420 430 440 450  
 IAMSFLVMGLVSENPTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA  
 MUA1\_LACLA AGAALVTAGIIASGTTKISNIEFILRGYDHIIEKLTAVGVDIQLIEE  
 380 390 400 410 420

nk603cp4.pep  
 GP\_BCT1:AE006268\_6

LOCUS AE006268\_6 [AE006268]  
 DEFINITION Lactococcus lactis subsp. lactis IL1403 section 30 of 218 of the  
 complete genome;  
 EVIDENCE BY HOMOLOGY BIO03.02 Murein sacculus and  
 peptidoglycan. CELL WALL FORMATION. ADDS ENOLPYRUVYL TO  
 UDP-N-ACETYLGLUCOSAMINE. FIRST COMMITTED STEP IN . . .

SCORES Init1: 55 Initn: 55 Opt: 132 z-score: 147.8 E(): 2.9  
 Smith-Waterman score: 165; 21.3% identity in 447 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSMFGGLASGETRITGLLEGEDVINTG  
 AE006268\_6 MKKIVINGGKRISGTIPISGAKNSVVALIPATILANDVVTLEGVPDISDVASLV  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRK--EGDTWIIDGVNGGLLAPEAPLDFG--NAATGCRLTMG-LVGVDYF  
 AE006268\_6 EIMEIMGAKIERNLEECRLVIDTRS---VVSRLPYGKINSRLASYYFNGALLGRFGQ  
 60 70 80 90 100

nk603cp4.pep 120 130 140 150 160 170  
 DST-FIGDASLTKRPMGRVLNPLREMGVQ-VKSEGDGRLPVTLRGPKTPTPTIYRVPMAS  
 AE006268\_6 ATVGLPGGCDLGPRTDLHDKAFKALGAKKIQEEEAHLEV-IGDSLVTGTTIYMDVVSVG  
 110 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
 AOVKSALLAGLNTPGITTVIEPIMTRDH---TEKMLQGFGANPTVETDADGVRTIRLEG  
 AE006268\_6 ATINT--MLAASRAKGLT-I IENAAREPEIIDVATLINMGA---QVRGAGTDIIRITG  
 170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280  
 RGKLTG-QVIDVPGD-PSSTAPFLVAALLVPGSDVTILNVLNPNTRTGLIITLQEMGADI  
 240 250 260 270 280

AE006268\_6 VDEMGAQHTVIPDRIEAGTYLALAAAM---GDGVIIENVIYEHLES-FIAKLEEMGVGL  
 230 240 250 260 270

nk603cp4.pep 290 300 310 320 330 340  
 EVINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEY--PILAVAAFAEG-ATVMN  
 AE006268\_6 TI---R---EDSIEVH-KSENLSKNITSVPYPGFATDLQOPITPLLLK-AKGRGSIVD  
 280 290 300 310 320

nk603cp4.pep 350 360 370 380 390 400  
 GLEELRVKESDRLSAVANGLKLNQVDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHR  
 AE006268\_6 TIYQKRVNHVPELARM--GANISVLD-DR---IIYDAPN---ELTGSCVQA-TDLR  
 330 340 350 360 370

nk603cp4.pep 410 420 430 440 450  
 IAMSFLVMGLVSENPTVDDATMIATSPPEFMDLMAGLGAKIELSDTKAA  
 AE006268\_6 AGAALVTAGIIASGTTKISNIEFILRGYDHIIEKLTAVGVDIQLIEE  
 380 390 400 410 420

nk603cp4.pep  
 TREMBL\_MAIN:Q99SD4

ID Q99SD4 PRELIMINARY; PRT; 419 AA.  
 AC Q99SD4;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)  
 DE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYLVINYL TRANSFERASE 2. . . .

SCORES Init1: 41 Initn: 65 Opt: 131 z-score: 146.8 E(): 3.4  
 Smith-Waterman score: 173; 20.4% identity in 446 aa overlap

nk603cp4.pep 10 20 30 40 50 60  
 MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSMFGGLASGETRITGLLEGEDVINTG  
 Q99SD4 MAQEVIKIRGRTLNQEVNISGAKNSAVAIIPATLLAQGHVKLEGLPQISDVKTIV  
 10 20 30 40 50

nk603cp4.pep 70 80 90 100 110  
 KAMQAMGARIRKEGDTWIIDG--VNGGGLLAPEAPLDFGNAATGCRLTMGLVGVDY--DFD  
 Q99SD4 SLLEDLNKASLNGTELEVDTTEIQNAAL--PN-----NKVESLRASYMMGAMLGFRFK  
 60 70 80 90 100

nk603cp4.pep 120 130 140 150 160 170  
 STFI---GDASLTKRPMGRVLNPLREMGVQVKSDEGDRLPVTLRGPKTPTPTIYRVPMAS  
 Q99SD4 KCVIGLPGGCPGPRPIDQHIKGFKALGAIEDSSTSMKIEAKELKGAHIF---LDMVS  
 110 120 130 140 150 160

nk603cp4.pep 180 190 200 210 220 230  
 AOVKSALLAGLNTPGITTVIEPIMTRDHT---EKMLQGFGANPTVETDADGVRTIRLEG  
 Q99SD4 VGATINIMLAAYATG-QTVIENAAKEPEVVDVANFLTSMGAN---IKGAGTSTIKING  
 170 180 190 200 210 220

nk603cp4.pep 240 250 260 270 280 289  
 RGKLTGQVIDVPGD-PSSTAPFLVAALLVPGSDVTILNVLNPNTRTGLIITLQEMGADI  
 Q99SD4 VKELHGSSEYQVIPDRIEAGTYMCIAAAC--GENVILNNIVPKHVET-LTAKFSELGVNVD  
 230 240 250 260 270

nk603cp4.pep 290 300 310 320 330 340 349  
 VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPIAVAAFAEGATVMNGLEE  
 Q99SD4 VRDERIRINNAPYQFVDIKTLVYPGFATDLQOPIT---PLLFMANG---PSFVTDITIYP  
 280 290 300 310 320 330

```

          350      360      370      380      390      400      409
nk603cp4.pep LRVKESDRLSAVANGKLNGVDCEGETSLVRGRPDGKGLNAGSAAVAATHLDHRIAMS
               | :   | :   | :   | :   | :   | :   | :   | :   | :   |
Q99SD4        ERFKHVEELKRMGAN-----IEVDEGTATI---KPS----TLHGAEVYAS-DLRAGAC
                   340             350                 360       370

```

```

          410          420          430          440          450
nk603cp4.pep FLVVMGLVSENPVTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA
               ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Q99SD4        LTIAGLIAEGVTTIYNVKHIYRGYTDIVEHLKAGADIWTEIV
               380        390        400        410

```

nk603cp4.pep  
GP BCT2:AP003364\_190

```

LOCUS      AP003364_190 [AP003364]
DEFINITION Staphylococcus aureus subsp. aureus Mu50 genomic DNA, complete
            sequence, section 7/9;
            SAV2124.
DATE       30-MAY-2001
ACCESSION  AP003364 . . .

```

SCORES Init1: 41 Initn: 65 Opt: 131 z-score: 146.8 E(): 3.4  
Smith-Waterman score: 173; 20.4% identity in 446 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AP003364\_190 MAQEVIKIRGGRTLNGEVNIISGAKNSAVAIIPATLLAQGHVKLEGLPQISDVKTLV

```

              70          80          90          100          110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDG--VGNGLLAPEAPLDFGNAATGCRLTMGLGVVY--DFD
              :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
AP003364_190 SLLEDLNKASLNGTELEVDTTEIQNAAL--PN-----NKVESLRASYMMGAMGLGRFK
              60          70          80          90          100

```

```

      120          130          140          150          160          170
nk603cp4.pep STFI---GDASLTWRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPTITRYVFMA
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AP003364_190 KCVIGLPGGCPLGPRPIDQHIGKFKALGAIEDSSSTSMKIEAKELKGAHIF---LDMVS
      110          120          130          140          150          160

```

```

      180      190      200      210      220      230
nk603cp4.pep  AQVKSAVLLAGLNTPGITTVIEPIIMTRDHT---EKM LQGFANPTVETDADGVRTIRLEG
      : : : : : | : | | | : : : : : | : : | | :
AP003364_190  VGATINIMLAAVIATG-QTVIENAAKEPEVVDVANFLTSMGAN----IKGAGTSTIKING
      170      180      190      200      210      220

```

```

          240           250           260           270           280           289
nk603cp4.pep RGKLTGQVIDV--PGDPSSTAFPLVAALLVPGSVDVTILNVLNNPRTGLILTLEMGADIE
              :| | : : | | : :: :|| : |: : || : : :| :| : :| ::::
AP003364_190 VKELHGSEYQVIPDRLEGTYMCIAAAC--GENVLNNIVPKHVET-LTAKFSSELGVNVL
               230             240             250             260             270

```

```

          290      300      310      320      330      340      345
nk603cp4.pep VINPRLAGGEDVADLVRVSSLKGVTVTPEDRAPSMIDYEYIPILAVAAAFAGATVMNGLEE
          | : | : : : | : | : : : | : | : : | : : | : : | : : |
AP003364_190 VRDERIRINNAPYQFVDIKTLVYPGFATDLQOFIT---PLLFMANG---PSFVTDITY
          280      290      300      310      320      330

```

nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVGRPDGKGLGNASGAAVATHLDHRIAMS  
| : : : | : : : : : : : || : : : : : : | : : | : : | : :  
AP003364\_190 ERFKHVEELKRMGAN----IEVDGTATI---KPS-----TLHGAEVYAS-DLRAGA  
340 350 360 370

410                      420                      430                      440                      450

nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
::: ||::: |:|: :: | :::: :||| |  
AP003364\_190 LIIAGLIAEGVTTIYNVKKHIYRGYTDIVEHLKALGADIWTETV  
380 390 400 410

nk603cp4.pep  
GP\_BCT2:AP003136\_76

```

LOCUS      AP003136_76 [AP003136]
DEFINITION Staphylococcus aureus subsp. aureus N315 genomic DNA, complete
            genome, section 8/10;
            ORFID:SA1926.
DATE      12-JUN-2001
ACCESSION AP003136 . . .

```

SCORES Init1: 41 Initn: 65 Opt: 131 z-score: 146.8 E(): 3.4  
Smith-Waterman score: 173; 20.4% identity in 446 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
AP003136\_76 MAQEVIKIRGRTLNGEVNISGAKN SAVAIIPATLLAQGHVKLEGLPQISDVKTLV

```

              70      80      90      100      110
nk603cp4.pep KAMQAMGARIRKEGDTWIIDG--VGNGLLAP EAPLDFGNAATGCRLTMGLGVGY--DFD
              : : : : : : : : : : : : : : : : : : : : : : : : : : : :
AP003136_76  SLLEDLNKASLNGTELEVDTTEIQNAAL--PN-----NKVESLRASYMMGAMLRGPK
              60      70      80      90      100

```

```

      120          130          140          150          160          170
nk603cp4.pep STFI---GDASLTGRPMGRVLNPLREMGVQVKSEGDRLPVTLRGPKTPTPITYRVFMAS
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
AP003136_76 KCVIGLPGGCPLGPRPIDQHIGKFKALGAIEDSSSTSMKIEAKELKGAHIP---LDMVS
      110          120          130          140          150          160

```

```

      180      190      200      210      220      230
nk603cp4.pep  AQVKSALLAGLNTPGITTVIEPIIMTRDHT---EKMLQGPGANPTVETDADGVRTIRLEG
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
AP003136_76   VGATINIMLAAYATG-QTVIENAAKEPEVVDVANFLTSMGAN---IKGAGTSTIKING
      170      180      190      200      210      220

```

nk603cp4.pep      240            250            260            270            280            289  
                  RGKLTGQVIDV-PGDPSSTAFPLVAALLVP GSDVTILNVLNNPTRTGLILTQEMGADIE  
                  :| |: :|| | : :: :|| | :| : : :: : :|| | : || :| ::::  
 AP003136\_76   VKELHGSEYQVIPDRIEAGTYMCIAAAC--GENVILNNIVPKHVET-LTAKFSELGVNVVD  
                  230            240            250            260            270

```

      290      300      310      320      330      340      349
nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEYPI LAVAAAF AEGATVMNGLEE
      | : | : : : | : | | : : | : | : : | : : |
AP003136_76 VRDERIRINNAPYQFVDIKTLVYPGFATDLQOFIT--PLLFMANG---PSFVTDITYF
      280      290      300      310      320      330

```

350 360 370 380 390 400 409  
 nk603cp4.pep LRVKESDRLSAVANGKLINGVDCDEGETSLVVRGRPDGKGLGNASGAAVATHLDHRIAMS  
 | : : : : : : : | : : : : : : : | : : : : : : :  
 AP003136\_76 ERFKHVEELKRMGAN-----IEVDEGTATI---KPS-----TLHGAEVYAS-DLRAGAC  
 340 350 360 370

```

          410          420          430          440          450
nk603cp4.pep  FLV MGLVSEN PVT VDDATMIATSPFPEFMDLMAGLGAKIELSDTKAA
               ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
AP003136_76   LIIAGLIAEGVTTIYNVKHIYRGYTDIVEHLKALGADIWTETV
               380      390      400      410

```

nk603cp4.pep  
TREMBL NEW: BAB58286

ID BAB58286 PRELIMINARY; PRT; 419 AA.  
AC BAB58286;  
DT 14-JUN-2001 (EMBLrel. 63, Created)  
DT 14-JUN-2001 (EMBLrel. 63, Last sequence update)  
DT 14-JUN-2001 (EMBLrel. 63, Last annotation update)  
DE UDP-N-ACETYLGLUCOSAMINE 1-CARBOXYLVINYL TRANSFERASE 2. . . .

SCORES Init1: 41 Initn: 65 Opt: 131 z-score: 146.8 E(): 3.4  
Smith-Waterman score: 173; 20.4% identity in 446 aa overlap

nk603cp4.pep MLHGASSRPATARKSSGLSGTVRIPGDKSISHRSFMFGGLASGETRITGLLEGEDVINTG  
BAB58286 MAQEVIKIRGGRTLNGEVNISGAKNSAVAIIPATLLAQGHVKLEGLPQISDVKTLV

nk603cp4.pep KAMQAMGARIRKEGDTWIIDG--VGNGLLAPEAPLDFGNAATGCRLTMGLGVVY--DFD  
BAB58286 SLLEDLNKASLNGTELEVDTTEIQNAAL--PN-----NKVESLRASYMMGAMLRGFK

nk603cp4.pep STFI---GDASLTKRPMGRVLNPLREMVGQVKSEGDRLPVTLRGPKTPTPTITYRVPMAS  
BAB58286 KCVIGLPGGCPGPRPIDQHIGKFKALGAEIDESSTSMKIEAKELGAHIF---LDMVS

nk603cp4.pep AQVKSAVLLAGLNTPGITVIEPIIMTRDHT---EKMLQGFGANPTVETDADGVRTIRLEG  
BAB58286 VGATINIMLAAYATG-QTVIENAAKEPEVVDVANFLTSMGAN---IKGAGTSTIKING

nk603cp4.pep RGKLTGQVIDV-PGDPSTAFPLVAALLVPGSDVTILNVLNPTRTGLILTQEMGADIE  
BAB58286 VKELHGSEYQVIPDRIEAGTYMCIAAAC--GENVILNNIVPKHVET-LTAKFSELGVNVD

nk603cp4.pep VINPRLAGGEDVADLRVRSSTLKGVTVPEDRAPSMIDEPILAVAAFAEGATVMNGLEE  
BAB58286 VRDERIRINNAPYQFVDIKTLVYPGFATDLQQPIT---PLLFMANG---PSFVTDITYP

nk603cp4.pep LRVKESDRLSAVANGKLNGVDCDEGETSLVVRGPDGKGLGNASGAAVATHLDHRIAMS  
BAB58286 ERFKHVELKRMGAN-----IEVDEGTATI---KPS-----TLHGAEVYAS-DLRAGAC

nk603cp4.pep FLVMGLVSENPTVDDATMIATSFPEFMDLMAGLGAKIELSDTKAA  
BAB58286 LIIAGLIAEGVTTIYNVKHIYRGYTDIVEHLKALGADIWTEYV

nk603cp4.pep  
TREMBL\_NEW: CAC49012

ID CAC49012 PRELIMINARY; PRT; 557 AA.  
AC CAC49012;  
DT 06-SEP-2001 (EMBLrel. 63, Created)  
DT 06-SEP-2001 (EMBLrel. 63, Last sequence update)  
DT 06-SEP-2001 (EMBLrel. 63, Last annotation update)  
DE PUTATIVE OLIGOPEPTIDEMUREIN PEPTIDE ABC TRANSPORTER ATP-BINDING . . .

SCORES Init1: 37 Initn: 37 Opt: 131 z-score: 144.9 E(): 4.3  
Smith-Waterman score: 132; 23.9% identity in 222 aa overlap

nk603cp4.pep DFGNAATGCRLTMGLGVYDFDSTFIGDASLTKRPMGRVLNPLREMVGQVKSE--DGDRIL  
CAC49012 EPTSALDATVQKQVLELLKILVDETVGSIIILVTHDMGVISEITNRVLVMRKGQVVEADRT

nk603cp4.pep FVTLRGPK-----TPTPTITYRVPMASQVKSALLAGLNTPTGITTIVIEPIIMTRDHT  
CAC49012 TAILDQPRHDYTKKLLAAVP-RLRIPTRTAKAEDNGRQA--DSPAIDSDQNPLLIABGLS

nk603cp4.pep KML--QGFGANPTVETDADGVRTIRLE-GRGKLTGQVIDVPGDPSTAFPLVAAL-LVPG  
CAC49012 KQFAPOGFAWIGRGKPRFGLRDVGIRLARGSITG-IVGESGSGKITTFGRILAGLDTAPT

nk603cp4.pep SDVTI----LNVLMNPTRTGLILTQEMGADIEV-INPRLAGGEDVADLRVR-SSTLKG  
CAC49012 GRITIGEHFPDVSQSGRRSGLLRVQMFQDPSVSLNPRMTIGETL-DESIRFGARIRSG

nk603cp4.pep TVPEDRAPSMIDEPILAVAAFAEGATVMNGLEELRVKESDRLSAVANGKLNGVDCDE  
CAC49012 EEPADLA-AMMDRLGLPRSLGRHPQLSGGQKQVRCIARALLARPEIIVADEPTSALDV

nk603cp4.pep  
GP\_BCT3:U31523\_1

LOCUS U31523\_1 [ECU31523]  
DEFINITION Escherichia coli 5-enolpyruvylshikimate-3-phosphate-synthase (AroA)  
gene, partial cds.  
DATE 18-NOV-1995  
ACCESSION U31523  
NID . . .

SCORES Init1: 47 Initn: 80 Opt: 114 z-score: 142.8 E(): 5.6  
Smith-Waterman score: 114; 41.9% identity in 43 aa overlap

nk603cp4.pep GVDCEGETSLVVRGPDGKGLGNASGAAVATHLDHRIAMSFVLMGLVSENPTVDDATM  
U31523\_1 ATYNDRHMAAMCFSLVAL-SDTFVTILDPKC

nk603cp4.pep IATSFPEFMDLMAGLGAKIELSDTKAA  
U31523\_1 TAKTFPDYFQLARISQAA

! Distributed over 1 thread.  
! Start time: Tue Nov 6 16:56:11 2001  
! Completion time: Tue Nov 6 17:12:17 2001  
  
! CPU time used:  
! Database scan: 0:15:06.4  
! Post-scan processing: 0:00:22.7  
! Total CPU time: 0:15:29.1

