

Sequence and organization of genes encoding enzymes involved in pyruvate metabolism in *Mycoplasma capricolum*

PENG-PENG ZHU AND ALAN PETERKOF SKY

Laboratory of Biochemical Genetics, National Heart, Lung and Blood Institute, Bethesda, Maryland 20892

(RECEIVED March 6, 1996; ACCEPTED May 29, 1996)

Abstract

The region of the genome of *Mycoplasma capricolum* upstream of the portion encompassing the genes for Enzymes I and II^{A_{glc}} of the phosphoenolpyruvate:sugar phosphotransferase system (PTS) was cloned and sequenced. Examination of the sequence revealed open reading frames corresponding to numerous genes involved with the oxidation of pyruvate. The deduced gene organization is *naox* (encoding NADH oxidase)-*lplA* (encoding lipoate-protein ligase)-*odpA* (encoding pyruvate dehydrogenase EI α)-*odpB* (encoding pyruvate dehydrogenase EI β)-*odp2* (encoding pyruvate dehydrogenase EII)-*dldH* (encoding dihydrolipoamide dehydrogenase)-*pta* (encoding phosphotransacetylase)-*ack* (encoding acetate kinase)-*orfA* (an unknown open reading frame)-*kdtB*-*ptsI*-*crr*. Analysis of the DNA sequence suggests that the *naox* and *lplA* genes are part of a single operon, *odpA* and *odpB* constitute an additional operon, *odp2* and *dldH* a third operon, and *pta* and *ack* an additional transcription unit. Phylogenetic analyses of the protein products of the *odpA* and *odpB* genes indicate that they are most similar to the corresponding proteins from *Mycoplasma genitalium*, *Acholeplasma laidlawii*, and Gram-positive organisms. The product of the *odp2* gene contains a single lipoyl domain, as is the case with the corresponding proteins from *M. genitalium* and numerous other organisms. An evolutionary tree places the *M. capricolum* *odp2* gene product in close relationship to the corresponding proteins from *A. laidlawii* and *M. genitalium*. The *dldH* gene encodes an unusual form of dihydrolipoamide dehydrogenase that contains an aminoterminal extension corresponding to a lipoyl domain, a property shared by the corresponding proteins from *Alcaligenes eutrophus* and *Clostridium magnum*. Aside from that feature, the protein is related phylogenetically to the corresponding proteins from *A. laidlawii* and *M. genitalium*. The phosphotransacetylase from *M. capricolum* is related most closely to the corresponding protein from *M. genitalium* and is distinguished easily from the enzymes from *Escherichia coli* and *Haemophilus influenzae* by the absence of the characteristic amino-terminal extension. The acetate kinase from *M. capricolum* is related evolutionarily to the homologous enzyme from *M. genitalium*. Map position comparisons of genes encoding proteins involved with pyruvate metabolism show that, whereas all the genes are clustered in *M. capricolum*, they are scattered in *M. genitalium*.

Keywords: acetate kinase; dihydrolipoamide dehydrogenase; lipoate-protein ligase; NADH oxidase; PTS; phosphotransacetylase; pyruvate dehydrogenase

Mycoplasmas, the smallest free-living organisms, contain the least complex genomes (Razin, 1985, 1992). They appear to have evolved from Gram-positive bacteria by selective elimination of nonessential genes (Maniloff, 1983). Growth of these organisms requires complex media, reflecting the loss of many genes encoding anabolic enzymes, but retention of genes encoding cat-

abolic pathways (Miles, 1992). *Mycoplasma capricolum* has a genome of 1,155 kb, sufficient to encode approximately 350 genes (Miyata et al., 1991).

It was shown previously (Cirillo, 1979) that *M. capricolum* can metabolize carbohydrates by the ubiquitous phosphoenolpyruvate:sugar phosphotransferase system (PTS) (for review, see Cirillo, 1979). This system promotes phosphotransfer from phosphoenolpyruvate to the heat-stable phosphocarrier protein HPr in a reaction proposed to be catalyzed by the homodimeric form of Enzyme I. Phosphorylated HPr then transfers a phosphoryl group to the sugar-specific acceptor proteins, referred to as Enzymes II. Each Enzyme II complex consists of one or two membrane embedded proteins or domains (IIC and IID), as well

Reprint requests to: Dr. Alan Peterkofsky, National Institutes of Health, Building 36, Room 4C-11, Bethesda, Maryland 20892; e-mail: alan@codon.nih.gov.

¹ Nucleotide sequence(s) reported in this paper has (have) been submitted to the GenBank™/EMBL Data Bank with accession number (U62057).

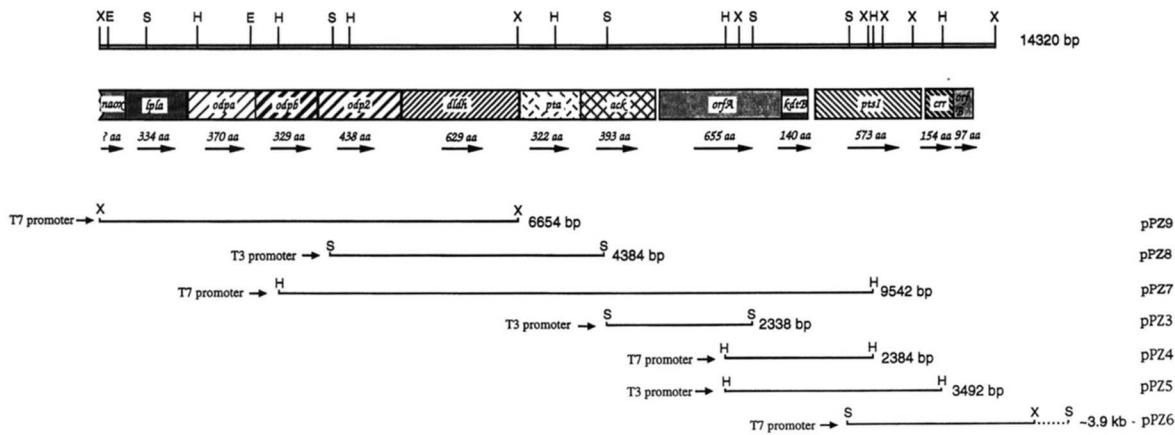


Fig. 1. Restriction and genetic maps of the region of the *M. capricolum* genome upstream of the *ptsI-crr* operon. The three clones that were isolated and sequenced are (A) pPZ7, the 9,542-bp clone derived from a partial *Hind* III digest, (B) pPZ8, the 4,384-bp *Spe* I clone, (C) pPZ9, the 6,654-bp *Xba* I clone. The internal *Xba* I site is located at base 10209 (see Fig. 2). The internal *Eco* I sites are located at bases 141 and 2396; the internal *Spe* I sites are located at bases 720, 3710, and 8094; the internal *Hind* III sites are at bases 1551, 2834, 3966, 7266, and 9992. Horizontal arrows indicate direction of transcription of the indicated genes. --- aa, the number of amino acid residues of the deduced open reading frame. *naox*, the gene encoding NADH oxidase; *lplA*, the gene encoding lipoate-protein ligase; *odpA*, the gene encoding the EI α subunit of pyruvate dehydrogenase; *odpB*, the gene encoding the EI β subunit of pyruvate dehydrogenase; *odp2*, the gene encoding the EII subunit of pyruvate dehydrogenase; *dldh*, the gene encoding dihydrolipoamide dehydrogenase; *pta*, the gene encoding phosphotransacetylase; *ack*, the gene encoding acetate kinase; *kdtB*, an open reading frame homologous to the *E. coli* KDTB protein product; *ptsI*, the gene encoding Enzyme I of the PTS; *crr*, the gene encoding Enzyme IIA glc of the PTS.

as two cytoplasmic proteins or domains (IIA and IIB) that pass the phosphoryl group from HPr to a specific sugar substrate during its translocation across the membrane (Postma et al., 1993).

In previous studies from this laboratory (Zhu et al., 1993, 1994), we reported the sequences of cloned *pts* genes from *M. capricolum*. It was shown that the *ptsH* gene, encoding HPr, constituted a unique monocistronic operon (the *ptsH* operon) in contrast to the typical occurrence in all *pts* operons sequenced to date of the *ptsH* gene immediately upstream of the *ptsI* gene, encoding Enzyme I. It was also established that *ptsI* is located in a dicistronic operon containing the gene encoding the glucose-specific Enzyme IIA (*crr*). In the present study, the region of the *M. capricolum* genome upstream of the *ptsI-crr* operon is characterized. It is shown that this region includes essentially all the genes necessary for the use of pyruvate.

Results

Cloning of the genes required for pyruvate metabolism

Previously we sequenced the region of the *M. capricolum* genome encoding the *ptsI-crr* operon (Zhu et al., 1994). Clones that were isolated in the course of that sequencing project were designated pPZ3, pPZ4, pPZ5, and pPZ6 (see Fig. 1). With the intention of investigating the organization of genes located upstream of the *ptsI-crr* operon, further clones were isolated. A partial digest with *Hind* III provided DNA for the isolation of a 9,542-bp clone, designated pPZ7, which was screened using an oligonucleotide based on the pPZ4 sequence (from bases 3466 to 3422 of the sequence published previously) (Zhu et al., 1994) (see Materials and methods). Other DNA clones were isolated by a similar approach. pPZ8 was isolated from an *Spe* I digest

and identified using an oligonucleotide sequence from pPZ7 corresponding to base numbers 8018–8062. The clone pPZ9 (from an *Xba* I digest) was also detected with a pPZ7 probe corresponding to bases 2882–2926 (see Fig. 1).

Sequence analysis of the genes required for pyruvate metabolism

Clones pPZ7, pPZ8, and pPZ9 were sequenced on both strands and then analyzed for open reading frames to reveal the gene organization illustrated in Figure 1. It was surprising to observe that a block of eight genes in this upstream region was in some way involved with the metabolism of pyruvate. The deduced gene organization, starting from the 5'-end of the sequence, is *naox* (encoding NADH oxidase), *lplA* (encoding lipoate-protein ligase), *odpA* (encoding the EI α subunit of pyruvate dehydrogenase), *odpB* (encoding the EI β subunit of pyruvate dehydrogenase), *odp2* (encoding the EII subunit of pyruvate dehydrogenase), *dldh* (encoding dihydrolipoamide dehydrogenase), *pta* (encoding phosphotransacetylase), *ack* (encoding acetate kinase). The region downstream of *ack* (*orfA*, *kdtB*, *ptsI*, *crr*, and *orfB*) has been described previously (Zhu et al., 1994).

The combination of the three sequenced clones (pPZ7, pPZ8, and pPZ9) allowed us to deduce a total of 10,214 bp of DNA sequence (see Fig. 2). Bases 1–280 coded for an open reading frame that is homologous to the carboxyl-terminal region of the amino acid sequence of NADH oxidase from *Enterobacter faecalis* (Genbank accession no. P37061) and *M. genitalium* (MG275 from the TIGR database) (Fig. 3). Comparison (by BESTFIT analysis) of the *M. capricolum* and *E. faecalis* sequences indicated 64% similarity and 39% identity of the compared sequences over a length of 88 residues. When the *M. capricolum* sequence was compared with the *M. genitalium* se-

A *ndox*

TCTAGATAGACCAGAATTATGCAACAGCAAATGAAGTTTATTCAAAGTTGGAGATAAAAAAACAGAAAATCATTGGTGTCAAGTAGCTAGTGAAACCCACTACTGAAGT
L D R P E F M S T A N E V L F K V V W D K K T R K I I G A Q V A S E K N H T E V
TATGTTATGTTAGCTTTAGGAATTCAAAGACCTTAACATTGATGAATTACCAATTGTTGAGCATTTCTCTTACACACTTAATTAACCCATTAAACTTTATTCGCTAGCAGGACT
M Y M L A L G I Q K D L T I D E L P L V D I F F L P H F N K P F N F I S L A G L
lpla
AGAAAGTTTATGGTTAAACTACTTTAAAGGAAATAGGAAATAGTTATGTTGATCTTAAGTATCATGATCCTGCTATGAATTAGCAATTGAAAGTATTTAACT
E V L G L N Y F K K E K * M I N N L I S K Y H D P A M N L A I E E Y L T
TATCATTATTAACGAAAAGAACCTATGTTTGTGACAAAATGCTAAACTATAGTTGTTGGAGAAATCAAAGCGTTTCTGCTAAATTAACCTAGAAGCTGCTCAAAGGATAAC
Y H Y K A K E P F W Q N A N T G D L G N V C Y S L I V D N S T D D V D Y Q K D N
GTTAAGATGTTAAAGAAATCAGGGGGAACTGTTTATCAGGATTITAGTTGATGTTATCTTGTAGTTGACAATTAACCTAGTGTGATTTGAAAGCAGTCAA
V K I V K R N T G G G T V Y Q D L G N V C Y S L I V D N S T D D V D Y Q K D N
CCTATTATCCTTAAATCAAAATATAATGCAATGTTCTGGAAGAAATGACATGGTGTGATGTTATAAGGTTTCAAGGAAATGCTCAATTAAACTAATGAAAGAAC
P I I T Y L N Q K N I N A M F S G R N D M V I D G Y K V S G N A Q L K T N E K T
CTAGTTCATGGTACATGCTATTGATGTTGTTGTTCAAAATGCTAAATTTTGTGATGCTAGGAAATTTAAACATCAAACTAGATCAAACCTGGCAGAGTTGAAAT
L V H G T L F D V D L S K M P K Y L V V W D P E K L K H Q Q I R S K P A R V R N
ATAAAAGAGTTCTTCAAAGACATTAATATGATATGTTAAGTACTTTTATTAAATGACCTAGTTGTTATGTTAAATGAAAGGAAATTTAAATGAAATTAACCTGATCAA
I K E F F K D I N I D I D L S T F I N D V V S S Y V K N E K I K W I E L T D Q E
AAGCAAACTTCAGTCAGAAAAGAACAACTTGTGATGAGACTGAACTTTGGGAAAATACCTGTTTTCTCTGTTAAAACAAATACCTGTTAAAGGTTTATTAC
K Q Y I Q S R K E T F D Q W D W T F G K N T S P L V K K Q Y L E S K G F I T
CTAAACTTGTGATGTCATAATGGAGTTTACTAACATTAAATTATGTTGATGTTTGTGAACTGAAAGACTGAAATTTAGAGCAGAACTAATTGTTGTTAAAGTTGATA
L N L D V D N G V I T N I K I Y G D F L G T Q G T E K L E A K L I G V K F D K K
GATGTTGAAAAGTTTAAATCAATTGACTTAAAGCAATTGCTAAAGAAATGCTAAAGGTTTCAAGTGTGATGTTACCAAACTTAAAGACTAATAAAATAAAATGAAAG
D V E K V L N Q F D L E A I F A K N F T S D D I T N L L F K D *
odpa
■ATAATCAAAATGACTTAAAGGAAATTGATCCTCTAAAGGAAATGTTGTTGTTGATAAAGGAAATTTAAATCTCAAGTGTGAACTGCAAATGCTAAAGGAAATTTCTGATCAA
M T Y L G K F D P L K E V C V L D K P V K I N P K L M P K I S D Q E
AATTCTTGAAGCATACAAAATATGAACTTATCTGCTAGACAAGATATTATCAAATACTATGCAACGTCAGGAAGATAATTATCATTTTATCTTCAACAGGACAAGCTGTA
I L E A Y K I M N L S R R Q D I Y Q N T M Q R Q G R L L S F L S S T G Q E A C E
GGTTGCATACATTAAATGCTAAACAAAACAGATCACTTTGTAAGTGGATATGAAATATTGCTGTGTTAGGCAATGGTCATTAGAAATAATTGCTATTGAAATTG
V A Y I N N K K T D H F V S R N N A A M Q L V R N I M L Y W I G
TAATGAACCGGGTAAAGCTCTGAGGAGTAATTGTTACCTCAAACATTGTTATTGGTTCAACATATTCTCAAGCTACAGGTTGCTGTGATAAATATGAAAACAGG
N E A G G K A P E G V N C L P N I V I G S Q Y S Q A T G I A F A D K Y R K T G
AGGAGTTGTTGTAACACTACTGGAGATGGTGGATCTAGTGAAAGTGAACATTATGAACTTGTGAAACCTTCAGCAAGTTCATGTTATTGTTATTGAAATAACAAATG
G V V V T T T G D G G S S E G E T Y E A M N F A K L H E V P C I F V I E N N K W
AGCTTACACAGCTAGAAGGAAACAACTTAACATTGTT
A I S T A R S E Q T K S I N F A V K G I A T G I P S I V D G N D Y L A C I G V
ATTTAAAGAAAGTTGTTGACTTGTGTTGAAAGGAAACGGCTCTGTTTGTGTTGACTTATGAGTGTGACTTATGAGTGTGACTTCTCTGTTGTTGTTGTTGTTGTT
F K E V V E Y V R K G N G P V L V E C D T Y R L G A H S S S D N P D A Y R P K G
TGAATTGAAAGAAATGGCTAAATTGATCCTTAAATTGATTAATTGATTAATTGATTAATTGATTAATTGATTAATTGATTAATTGATTAATTGATTAATTGATTAATTGAT
E F E E M A K F D P L I R L K Q Y L I S P N E S D N E Q D K F V
TGCTGATGAAATTGCTGTTGAAAGGAAATTTATGATCTAAATTGCAATTGTTAAATTATGATCTAAATTGCAATTGTTAAATTATGATCTAAATTGCAATTGTTAA
A D E F A W V E K N K N Y D L I D I F K Y Q Y D K M D I F L E E Q Y K E A K E F
CTTGAAAATACCCAGAACTTAAAGGAG
F E K Y P E S K E G G H H * M A I I N N I K A V T D D A L D C A M Q R D P N V I V
TTTGGTGAAGATGTTGGAACTGAAGGAGTTCTGGAGCTACTGGAGCTACTGGAGGTTAGCTTAAATTGAGGAAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
F G E D V G T E E G G V F R A T Q G L A V K F G N A D C F N P A I S E A M F A G V
GGTTTGGAAATGGCTATGAAATGGTGTGAAACCGAGTTTGGAGGTTAGCTTCTTACAAAATTTTACTAACATTCAAGAATGAGGAGGAGGAGGAGGAGGAGGAG
G L G M A M N G M K P V L E M Q F E G L G L A S L Q N I F T N I S R M R N R T R
GGTAACACTGCTCAAATGGTATTAGAGCTTCAAGGTTGGGGGGTATTCTGCTTGTGTTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
G K Y T A P M V I R M P M G G G I R L E H H A L E E A L E A V Y H I P G V Q I V
TGTCATCATGATGAACTAAAGGGTAATTGCTGCAATTGATGTT
C P S T P Y D T K G L I L A A I D S P D P V I V V E P T K L Y R A F K Q E V F D
GAACACTACATAGTACCAATTGGAGAGGTTATAAACTCAAGAGGAAATGATCTAATGTTGTTACTTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
E H Y I V P I G E G Y K I Q E G N D L T V V T Y G A Q T V D C Q K A I A L L K E
ACTCATCCAACTGCAACTTATGTT
T H P N A T I D L I D L R S P W D K K M V I E S V K K T G R L L V V H E A V
AAATCATTTTCAGTTGCTGAAATTATGCAACTGTTAATGAAAGATTTGTTGAAACATATAAGCACCTTATGAGTGTGACAGGTTATGTTGTTGTTGTTGTTGTT
K S F S V S A E I I A T V N E C F E Y I K A P L S R C T G Y D V I T P F D R G
odp2
GAAGGTTACTTCAGGAACTTAAACGGAG
E G Y F Q V N P K K V L K M Q E L L D F K P * M F K V K F A
TGACATGGTGAAGGCTAAACAGGAAACAGCTGCTGAAGTT
D I G E G L T E G T V A E V L V K V G D V V K E G Q S L Y F V E T D K V N S E I
ACCTGCTCAGTGGCTGGAAATATGCACTTAAACATTAAGCTGGACAAGAAATTAAAGTTGGAGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
P A P V A G K I A V I N I K A G Q E I K V G D V V M E I E D G S D T S A T S E P
AAAGGAGAAACAAATCAGAAAGCTTAAAGTGAAGTAGTTGAGGAAATGCTAGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG
K A E T K S E A K V E V V E E N A S V V G T A C T P V S N D V I V R K Q T T T V N K
ATCAAGTACTATCAACAGCTACCCATTAGCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCT
S S T I K A T P L A R K V A A D L N I D L S L V T P T G P N Q R I L V A D I K N
TCATCAAGCTTCACTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCTCAACCCATTAGCT
H Q A S S T Q L A S Q P I S Q P A P T P S P S A H Q T I A P T I K V V E P S A P
TTTATCTGAGATGAGCTTCAATTGAGTTGTTGAGAAAAGCTACAGTAAAAGCAATGCAATTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
L S W D E V P M N G V R K A T V K A M T K S H T E I A A F T G M K N T D I T E T
TCACAAAATGAGAATTAAGGAGATCATGCTGAGCTAGTGGAAATTAACTACTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT
H K M R T E L K D H A A A S G I K L T Y L A F I I K A V A K S L R D M P N I N V
AAGAGGTGATTTGCAAATACAAAATCCAATTATGCAACACATTAAATTGGAATTGCAAGTAGATAACCCAAACGGATTATGGTCCAGTTAAAGGTGCTGATCTTAA
R G D F A N N K I Q F M H N I N I G I A V D T P N G L M V P V I K G A D H L S V 4440

Fig. 2A. (Figure continues on next two pages.)

B	ATTTGAAATTGCAATTAAAATTAGTAGCAACTAGCAAATAAGGCCAAGATGGTAAATTACAAGAGCTGAATGACTGAAGCAACATTACTGTTCAACCTTGGTCAGTAGGTTAGA F E I A I K I S E L A N K A K D G K L T R A E M T E A T F T V S N F G S V G L D	4560
	TTATGCTACTCCTATTAACTCACAGAGTCGTCTGCTATTGGAGGTGGTACAATGCTCAACTCCCTTATATATTAAATGGTGAATTACAAAAAGATTATAATGCCATTATCAAT Y A T P I N S P E S A I L G V G T M S Q T P L Y I N G E L R F I M P L S M	4680
	ACTTGTCATAGAACATTGATGGTGCATGGAGATTITAATTAAGTACAAGATTACTTACAAACAGTTATTGTTATGTTATGTTAGTAAATTGTTAGGATAAAATATGGT T C D H R I I D G A R F L I K V Q D Y L S K P V L L F M *	4800
	TTAAGAAAAATTGCTGACATAGGTGAAGGTTAACAGAAGAACAGTCGCTGAAGTTTAGTTAAAGTGGTGTGTTAAAGAAGGCAACCTTAACTTTGTTGAAACTGATAA K V K F A D I G E G L T E G T V A E V L V K V G D V V K E G Q P L Y F V E T D K	4920
	AGTAAATAGTGAATTCCCTCCCGTTGCCGAAAATTGCAATAATCAATCTACTGGTCAAGAAATTAAAGTGGTGTGTTATTGAAATTGATGATGGAAAGTCACCTC V N S E I P S P V A G K I A I I N I S T G Q E I K V G D V V I E I D D G S S T S	5040
	TACAGCTTCACCTCAAAAGTGAAGTAGTGTGAAGAAAAGTCTAGTGTAGTTGGTGTCACTCCAGTTCAGTGTGTTACCAAGTAGGACCAAAAGCTGAAGCTAAAGT T A S T S K V E V V E E N A S V V G A T P V S N D V L P S R A P K P K A E A K V	5160
	TGAAGTAGTTGAAGAAAATGCTAGTGTAGTTGGTGTACTCCAGTTCAATGATGTTTACCAAGTAGGACCAAAAGCTGAAGCTCAAAGTGTGATGTTCAAATTGAAAGA E V V E E N A S V V G A T P V S N D V L P S R A P K P K V E A P K V D V Q I E D	5280
	TACATTTGATGTTGTAGTTGGTCAGGAATTGGTGTATGTTACTGCTATTAACTGCAATTAGGTTAAACCTTAAATTGAAAGAATTACTATGGTGGAGTTGTT T F D C V C V V G A G I G G Y V T A I K S A Q L G L K T L I I K E Y Y G G V C L	5400
	AAATGTTGGATGTTCTACAAAAACTTGTAACTCTCATGTTATCATGATATGATACTAAAGCAGAAATTAGGAAACTGTTACAAAATACTGAAAGTGTGATTAATTG A V G C I P T K T L L K T S H V Y H D I V H K A K E L G I V L Q N T E N V V I D	5520
	TTGAGCTCAAGCACTTGAAGAAAATGGTGTGTTAAGAAAATTACAGGGGGAGTTAAATTTAGATAAAAATAGTAACTCAAAATTAAAGGTAAGCTATTGCTTACAGATAA W A Q A L E R K N G V V K K L T G G V K Y L L D K N K V T Q I K G E A I A L D K	5640
	AAATAACAATTTCAGTAAATAAATAAAATTATCGTGTAAATAATTAGTTATTGCTCTGGATCAACCCCAATCATTTACACTTCAGGTTTACATGCAAGGAAGAAAAGATGAAATTAT N T I S V N N K N Y R V N N L V I A S G S T P N H L P L P G F D Q G R K D G I I	5760
	TATTGACTCACTGGAATTATCAGTCCAAAATTCTGAAACTTGTGAAATTGGTGGAGGAGTTGGTGTATTGAGTTAGCTGTTATTGCTGATGTTACAAAGTTAC I D S T G I L S V P K I P E T L V T V I G G G V I G T F S C L F A S L G T K V T	5880
	TGTTTACAGGATTCCAACATTAGGAGTAAATGCTAGATAAAAGATATTATTGATGCAATGACTAAAGAGTTAAAACAGATAACATATCCAAGTTTACAAATGCTCAGTTAAAGA V L Q G L P T I L E M L D K D I I D A M T K E L K N R Y N I Q V I T N A S V K E	6000
	ATTTAAAGATGGTCTGTAGTATATCAAATTGATGGTCAAGATCAAATGTTAAAGGAGAATATGTTAGATACTGGTGGAGCTAAACTTCAACTGGATTGAAACATTGGTT F K D G S V V Y Q I D G Q D Q M I K G E Y V L E S V G R K T S L T G F E N I G L	6120
	AGAAATTCACTCCAAGAAAAGGGTGTGTTAGGAGTCAAGAACTAACTTAGGTTATGCAATTGGTGTGTTAGGAGCTGGTGTAGCTAACATGCTAACAGTTAAAGG E L T P R K G V V V N E Y Q E T N L D G V G V Y A I G D V V W G K S M L A Q T A V K G	6240
	AGCTATTGTTCTGTCTAGATAAGATTGCTAAAGGCTATAAGGCTCATGCTGAAGATAATTGTTATGAAATTGACAAAGTTCATCATGTTATTACACACCCAGAAGTTCAAGT A I V A A N R I A K K A N K A H A E D I V M N Y D K V P S C I Y T H P E V S M I	6360
	TGTTAAGCAGACAACAAATTAAAGAAAATTGTAATACAACGCTTTAAATCCCATTGCAATTGGTAAAGCTTAGCTGATGATGATACTTCAGGATTGTTAAAATTAT G K T E Q Q L K Q E N I E Y K A F K F P F S A I G K A L A D D D T S G F V K I I	6480
	TGTTAGAACCTTAAACACAAACTTTAGGAGTGTGCAACATATTGGAATTAGGCTACTGAAATGATGTTGAAATTGATGTAAGGAAACATCACAGAAAATTGCTAA V E P K H Y K T T L G A T M I S E I T A V I E C E G T I T E I A N pta.	6600
	TACAATTCCACCTCACCAACAGCAATTGGAGAAGCAGCTAGAAACAGGAAAGCTTCAATTGATGTTGAAATTGATGAAATTGATGTTGAAATTGATGTTGAAATTGAT T I H P P T M S E A I G E A A E L E T G K A I H F *	6720
	AGAAATTAAACACTAGGTTAAACAGGCTATTGCAACCAAGTATTGCAATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCA E I K N Q L G L K S E K K S I V F P E A E S E I I I Q S V A K T L V D E K L G L P	6840
	ATTCTATTATTTAAATCTTCAAAAGAGTCCAGTGAATTAAAATTCTCATCAATTAAACTATTGAGTGAACACAAAAGAATTGTAAGAAGAATTGTAACAAACT I L L F K S S K P S E I G T C I L D E F D T K E F E E E F V K L	6960
	TAGAAAAGTAAAGCAACTATTGAGTTGCACACCAAGTATTGCAATTGCAATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCA R K G K A T I E V A H Q V M Q L P N Y I G A M L V K L N Q A D C M L S G L N N T	7080
	ACACAGCTGATACAATTGACCAAGCTTACAATTGAGGATATAATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCAAGTATTGCA T A D T I R P A L Q I I G T K P G Y N I A S S I F V U M S K G N E N Y I F T D C A	7200
	TTTAAATTTAAACACAAAGTGAACAAATTGCAATTGCAACAAAGCTTAAATGCAATTGCAATTGCAATTGCAATTGCAATTGCAATTGCAATTGCA L N I K P T E Q L V E I T Q M A V D F A K T V M N Q V A E A L S Y S T N G	7320
	TAGTGGTAAAGGTGAAGATGTTGATAGAGTACATCAAGCAGTTGAATTAAAAGTCAAGGTTGAGGTTGAAATTCAATTGATGCGCTTGTGATAAAAAC S G K G E D V D R V H Q A V E I L K S E K D Y V C E G E I Q F D A A F D K K T	7440
	TAGAGATAAGAAAATTAAAACACACAGATAATTGTTGTTTCCAGATATAATTGCTGAATTGCTGAATTGCTGAATTGCTGAATTGCTGAATTGCTGA R D K K F K N C S L L K Q T P D I F V F P D I N A G N I G Y K I A Q R M G G F E	7560
	AGCAATTGGACCTTTGTTAGGTTAAATCAACAGTTAATGACCTAAGTAGGGTCAACATTGCAATTGCAATTGCAATTGCAATTGCAATTGCA A I G P F V L G L N Q P V N D L S R G A T F V D V L N T A I M T L Y L S Y *	7680
	AGTAAATTGTTGAGTAAATTGAGGAGTGTGTTAGGTTAAATCAACAGTTAATGCAATTGCAATTGCAATTGCAATTGCAATTGCAATTGCA M I L V I N S G S S S I K F K L F D T S K A I E P I L D G L A E R I G I D G	7800
	GATTTTAAAGTTGAACATAATAAATATAATTGAGATCCACTCCAGATCATGAACTGTTCAATTAAATTGCAATTGCAATTGCAATTGCAATTGCA F L K F E H N N Q K Y K F E D P L P D H E H A I Q L I L N K L L E L K I I S N I	7920
	TTGATGAAATAAAGGTGTAGGTTTAGGTTCATGGTGGAAATTTCACATTCACTAAATTAAATGAGAAGTATTCAAGAAAGTGTAAATTGCTCCTTAC D E I K G V G F R V V H G G E I S H S S I N E E V L Q K I Q E S V K L A P L H	8040
	ATAATCCTGTCGAATTATTGCAATAAAAGCAGTAAACACTAACTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCA T N P A A I I I A I K A V K K L M P N T S M I A C F D T A F H Q T M P Q V N Y L Y S	8160
	CTGTTCTTATAATGATGAGAAGATTGGCGTAAGAAAATGGTTTCAAGGAAATTGTTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCA V P Y K W Y E E F G V R K Y G F H G I S Y E Y I V N K C E E I L N K K K E H L N	8280
	ATTAATAGTTGTCATTTAGGAAATGGCGCAAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCA L I V C H L G N G A S I S C I K D G K S Y D T S M G L T P L A G L M M G T R S G	8400
	GAGATATTGTTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCAATTGAGTATTGCA D I D V S I C E Y V A K Q T N S D I F A I T Q I L N K Q S G L G L S Q T S A D	8520
	ATATGAGAGATGTTAGAACATATGAGAACATTGAGTATTGAGTATTGAGTATTGAGTATTGAGTATTGAGTATTGAGTATTGAGTATTGAGTATTGAGTATTGAG M R D V L E Q Y D R N D K K A I I A V E K Y V Q V V A D F I V K Y A N Y L D S I	8640
	TTGATGCACTAGTCATTACAGCAGGAATTGGTGAAGAACATGAGATGTCATTAGAGATTATGAGTAAAGAGTTAACCTTGTGAGGTTAGGTTACAAATTGATGCA D A V V F T A G I G E N A D V I R D L I C K R V K L L G L Q I D Q E K N E S K Y	8760
	ATTGAGACTATAATTCTAGTAGGAAATCAAAATTCCAGTTATGCTATTAGAACAAATGAGAACAAATGAGAACAAATGAGAACAAATGAGAACAAATGAGAACAA S D Y K L I S S E K S K I P V Y A I R T N E E K M I C L D T L N L I K *	8880

Fig. 2B.

C	orfA	9000
AGTTAAAAAACTTAGCTTAAATGCTAAGTTTAATGTTAATAAAGAAACATAAAAGTATTAAAGTTTGTAGTTAATTCATTTATCAATAACAACT		
M K R T I K Y L S F L G L I P F L S I T T		
ATAAGTTGTGTTAACAGCTAAAGAAAATAATAAAATCAATTAAATTAGTCATTAAACAACCTTATTTATCTTAAACAGCTTGTAGTTAGATAATAAAACTAGAATCAA		9120
I S C V K Q A K E N N N K N Q L I S Q F K Q L I F I L N S F D L D N K K L E S K		
ATCATAAAAAGCTATTGAAAAAAAGTATTGTTAATAAAATTAGTAACTATAAATTAGAATTAACTATAAAGTTTAACAGATAATAAAAGTGGAAACTAAACAAATTAGTCAGTA		9240
I T K A I E K S D F N K I S N I N L E L T I K F L T R I K N E L E T K T I S Q L		
AATAAAAATGATAAGCTAGATATTAACTAAATAAAAGTCATTAGTCATTAATTAGAATTAACTATAAAGTTTAACAGATAATAAAAGTGGAAACTAAACAAATTAGTCAGTA		9360
N K N D K L D I L T K I K V H L G S L N L I E L V N I V D E L V N K L N Q K E E		
ATTAAAAATACTCATAAAGATAAAAATGAAAAAAATAAAAGATAATATAGAAGATATTGATGATTCAAAACTTGAGATTAGATAATAACCTACAGCACAACTACCCAGAT		9480
I K N T H K D K I E K N K D N I E D I D D S K L E I L E S K Y I P N Q H N Y P D		
TATGTTAAAACCTTTAACAGCTTACAGAGAAATTATAAGCTTATGATAGAATTTCAGTTAAATTAAAGATGGTTATTAGTAATTCAGCAATTGATGAGTTAGTAATAAATTAACTGAAACT		9600
V K K F K T V S A E E I Y K E L D Y R T F S I K K F L V K L K D G G L L S N G T		
GGAACCTGGTGTGATTAGATTATGATCATAAATATAGAATACATAAAATAAAATGTTATTGCAACAACTTACATGTTGGCAGATTAGCAATTCAACTGATGAACAAAT		9720
G T G W L L D Y H K Y S N T N K Y K M P I A T N L H V L A D F S N S L T D E Q N		
AAAGAATTAACTACTATGATCCATCAGGAATAAAGTGTAGGTTAGGAAAGCTGATAATGTTACTGATTTAGAGATAATAAAATTCAGAAATAATAATCAGAAATAATAAT		9840
K E F N Y Y D P S G N K V I G L G L G K A D N V T D F S R K N N N S K S E N N I		
GCTAATTATTATTTAAATAATCAGATTGAAAGATTATCTTAAAGTATTGAAAGTGTAAATTATCTAAAGGTTATTGCAACAACTTACATGTTGGCAGATTAGCAATTCAACTGATGAACAAAT		9960
A N Y Y L N N Q D F E N Y L K W F S V N K F S G K I S E P K I V F G A V D F		
ATGAAAGATCGTGTATTAAATCATTAGAAGCTTACAAAAAGAGCAATTAAATTATATACTATAAAATAATAAGTAAATTATGATGACAATAAGATAGCTGAATAAT		10080
M K D R A I K N H Y E A L Q K E A I N Y Y N Y K K N N N E I N D D N K I A W N N		
TTTTAAATAATAAAAGATATTCTATAATGATAGATTGAGCTTGCAGATTGATGTTGACCTAGATTGGTGTATAATTAAATCATGAATTCTAATGCTATTAGTGGTTA		10200
F L L N K D I P I M I D F A V F E F D V D L D L V D Y N L K S W I S N A I S G L		
GATAATTATCTAGA 10214		
D N Y L		

Fig. 2. Nucleotide sequence of the region of the *Mycoplasma* genome encoding genes involved with pyruvate metabolism.¹ A partial sequence of *naox* and the entire nucleotide sequences of *lplA*, *odpA*, *odpB*, *odp2*, *dldH*, *pta*, *ack*, and an ORF designated *orfA* (Zhu et al., 1994), as well as the deduced amino acid sequences, are shown. The putative promoter regions [i.e., 35 regions (underlined) and 10 regions (boxed)] are shown for the *lplA*, *odpA*, *odpB*, *odp2*, *pta*, and *orfA* genes. A sequence corresponding to a likely transcription termination site downstream of *ack* is marked with diverging arrows. Shine-Dalgarno ribosome binding sites for *lplA*, *odpA*, *odpB*, *odp2*, *dldH*, *pta*, *ack*, and *orfA* are highlighted with shaded boxes. Asterisks denote translation stop codons.

quence, it showed 77% similarity and 52% identity over a length of 92 residues.

After a gap of 11 bases, an open reading frame (from bases 292 to 1312) was identified. The end of this sequence was characterized by the presence of four in-frame stop codons (all TAA). Use of FASTA analysis of the deduced protein product led to its identification as the gene encoding lipoate-protein ligase. Alignment of the sequence with the lipoate-protein ligases from *M. genitalium* (MG270 from the TIGR database²) and *Escherichia coli* (Genbank accession no. P32099) is shown in

² The *M. genitalium* sequence MG279 is listed in the TIGR database as unknown in function. The FASTA analysis reported here identifies the sequence as the gene encoding lipoate-protein ligase.

Figure 4. BESTFIT analysis of the *Mycoplasma* sequence (334 amino acid residues) with the lipoate-protein ligase from *E. coli* (337 amino acid residues) showed 55% similarity and 32% identity over a length of 335 residues (data not shown). A similar analysis with the enzyme from *M. genitalium* (336 residues) showed 62% similarity and 36% identity over a length of 335 residues.

After a gap of 19 bases, an open reading frame (from bases 1331 to 2443) was observed. FASTA analysis showed the protein product to be the α -subunit of pyruvate dehydrogenase Enzyme I. Figure 5 shows an alignment of the sequence from *M. capricolum* with those from other sources recovered from the Genbank and TIGR databases. It is clear that the regions of total amino acid conservation (shown in reverse shading) are

1		110
<i>Mycge</i>	MKKVIVIGIN HAGTSFIRTL LSKSKDFKVN AYDRNTNISF LGCGIALAVS GVVKNDDLF YSNPEELKQM GANIFMSHDV TNIDLIKKQV TVRDLTSNKE FTQFQDQLVI	
<i>Entfa</i>	.MKVIVLGSS HGYYEAVEEL LNLLHPDAEQ WYKEKGDFIS LSCGMQLYLE GKVKDVNSVR YMGEKMSR GVNVFSNTEI TAIQPKEHQV TVKDLVSGEE RVENYDKLII	
111		220
<i>Mycge</i>	ASGAWPICMN VENKVTHKPL EFNYTDKYCG NVKNLISCKL YQHALTLIDS FRKDFTIKSV AIVGSGYIGL ELAEAAWLCK KVTVTVIDLLD KPAGNNFDHE FTDELEKVMQ	
<i>Entfa</i>	SPGAVPFELD IPGK..... LDLDNIYLMRG RQWAIKLKQK .TVDPEVNNSV VVIGSGYIGI EAAEAFAKAG KKVTVIDILD RPLGVYLDK FTDVLTEEME	
221		330
<i>Mycge</i>	KDGLKLMGMC SVKGFVUDST NNVVKGVETD KGIVNADLN QSIGFRPSTK FVVKDQNPEF IHNGSIKUNE FLOALNHKD VYIGGCCAIY NAASEQYENI DLATNAVKG	
<i>Entfa</i>	ANNITIATGE TVERYEGRDR ..VQKVTD KNAYDADLVV VAVGVRPNIA WL .KGTTEL HPNLKIKTDE YMFT .SEPDVF FAVGDATLIK YNPADTEVNI ALATNARKKG	
331		440
<i>Mycce</i>	
<i>Mycge</i>	LVAAMHIIGS NQVKLQSIVG TNALHIFGLN LAACGLTEQR AKKLGFDPVGI SVVDDNDRPE FMOSYDKVRF KLVYDKKLR ILGAOLLENN TNHSEIIFYI ALAIQKQMLL	
<i>Entfa</i>	RFAVKNLEEP VK.PFFGVQG SSGLAVFVDYK FASTGNEVM AQKLGKETKA VTVVEDYLMF FNPDKQKAWF KLVYDPEQTO ILGAOLMS.K ADLTANAINI SLAIQAKMTI	
441	478	
<i>Mycce</i>	DEBPLVDDIFF LPHFNKPFPNF ISLAGEVVLG LNYFKKEK	
<i>Mycge</i>	TEBLGLVDYVF LPHYNKPFPNF VLATVHQALG FSYYIPKK	
<i>Entfa</i>	EDAYADFFF QPAFDPENNI INTAAAEAV.KQER	

Fig. 3. Alignment of sequences of characterized NADH oxidase proteins from various bacteria. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in the listed NADH oxidase proteins are shown in reverse shading. Abbreviations used and references to published sequences are: *M. capricolum* (*Mcapr*) (this work); *M. genitalium* (*Mgeni*) (MG275 from the TIGR database); *E. faecalis* (*Entfa*) (Ross & Claiborne, 1992).

```

1 .MINLLI .SK YHEDAMMII EYLTYHYKA KE .PIVFFF CHANLIVR NQAFAPLIL EAQAKNKKI VKNNTWYV YQHLSWYS LIVDNSTDV D .QKALQP 110
Mycca .MQTTIAP VFNLYFPAE EWLTLERPK NELVKVYFE CHANLIVR MDTYAVRL KELESKONL PRPSFSTAA PHDLSLFS IILPRTGKVM ENAEGTRNN
Mycce STLRLL SD SYDWFNAV EECIFFRQMPA TQ .RVLFLW RRADEWVIG AOPWPKCET RRMEBMRML ARISSEGRAY PHDLSLFS FTMAGKPE... .DQKTIST
Ecoli

111 .IITYLNQK NINAMFSGRN DMVI....DG YVNSCNOLQK TNEKTLVHG LDFDVLLKRM PKLVVPPEN LKHQQIREKP ARVRIKHF KDNIDIDL S TFINDVVS 220
Mycca ..VVPFLNSL NNPVAFPHRE ELEI....NN KEPNLQAYT AKDRLLVHG MFDATPLKRL AVKNDVTKT IASKGVDDVA KRWVWKRYL PNWT... A KFLEEMINPF
Mycce SIVLNALN GVSABASGRN DLVVKTVEGD RSVSOSCEA TKDRGFHGT DLNLDLRLR ANVLPKRR LAKGITVSRV SWVTEBLL PGITHEQVACE AITEFAAFFH
Ecoli

221 VNKNEKIKWIA LTDXQEKKYIQ SRKET .KFD QDVKLGEKNT ESKPLVKKQYL ESKGKFITNL DWDNIVITNI KIYGDFLGTQ GTEKLEAKHII GVFKD... K KDWVKVLFQF 330
Mycca VTIEKAEITV LTDXDALKWA KRKE .HFO SIEUHBLKTY EYNFKNKKRF NNAGLFCENV QWEKFTVVDI KFGYDFLSV DITPVTKKII QKYD... Y KTFKLNFEL
Mycce ..GERVEAEI ISPNNKTPDLP NFAETFARQS SIEUHBLQAP AFSHLLDERF TWGC VELHF DUEKCHITRA QVFTPSLNP APLEALAGRQ CLYRADMLQ QCEALLVDP
Ecoli

331 353
Mycca DLEIAFIKNT TSDDITNLLF KD.
Mycce DHFSDFYFGSL KPEQLGVIP DNK
Ecoli PPEKEKREL SAWMAGAVR. ...

```

Fig. 4. Alignment of sequences of characterized lipoate–protein ligase proteins from various bacteria. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in the listed lipoate–protein ligase proteins are shown in reverse shading. Abbreviations used and references to published sequences are: *M. capricolum* (*Mcapr*) (this work); *M. genitalium* (*Mgeni*) (MG270 from the TIGR database); *E. coli* (*Ecoli*) (Morris et al., 1994).

Fig. 5 (on next page). Alignment of sequences of characterized Enzyme I- α proteins of the pyruvate and α -ketoadic dehydrogenases and acetoain catabolism complexes from various bacteria. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in all listed Enzyme I- α proteins are shown in reverse shading. Abbreviations used and references to published sequences are: *M. capricolum* (odpa-mycca) (this work); *M. genitalium* (odpa-myce) (MG274 from the TIGR database); *A. laidlawii* (odpa-achla) (Wallbrandt et al., 1992); *B. subtilis* (odpa-bacsu) (Hemila et al., 1990) and (odba-bacsu) (Wang et al., 1993); *B. stearothermophilus* (odpa-bacst) (Borges et al., 1990); *A. thaliana* (odpa-arath) (Luethy et al., 1995); yeast (odpa-yeast) (Behal et al., 1989); human (odba-human) (McKean et al., 1992), (odpa-human) (Ho et al., 1989), and (odpt-human) (Dahl et al., 1987); pig (odpa-pig) (Sermon et al., 1990); mouse (odpa-mouse) (Fitzgerald et al., 1992) and (odpt-mouse) (Fitzgerald et al., 1992); rat (odba-rat) (Zhang et al., 1987), (odpt-rat) (Cullingford et al., 1993), and (odpa-rat) (Cullingford et al., 1994); *Ascaris suum* (odpt-ascu) (Johnson et al., 1992) and (odpa-ascu) (Johnson et al., 1992); *C. magnum* (acoa-cloma) (Kruger et al., 1994); *A. eutrophus* (acoa-aleuc) (Priefer et al., 1991); *Pelobacter carbinolicus* (acoa-pelca) (Oppermann & Steinbuchel, 1994); *K. pneumoniae* (acoa-klepN) (Deng et al., 1994); bovine (odba-bovin) (Hu et al., 1988); *Pseudomonas putida* (odba-psepu) (Burns et al., 1988). odpa and odpt designations correspond to pyruvate dehydrogenase, whereas odba designations correspond to α -ketoadic dehydrogenase complex gene products, respectively, and the acoa designation corresponds to acetoain catabolism complexes.

also shared by the sequence from *M. capricolum*. The signature sequence (shown as residues 250–280) G(D/E)(G/A)(X26)NN, characteristic of thiamine diphosphate-dependent enzymes, is also found in the *M. capricolum* sequence. A phylogenetic tree (Fig. 6) of the sequences of the α -subunit family shows that the *M. capricolum* protein is related closely to the corresponding proteins from *M. genitalium*, *Acholeplasma laidlawii*, and those from Gram-positive organisms (*Bacillus subtilis* and *B. stearothermophilus*).

The last base of the termination codon (TAA) of the sequence encoding the α -subunit of pyruvate dehydrogenase Enzyme I is also the first base of an open reading frame (from bases 2443 to 3432) that codes for the β -subunit of pyruvate dehydrogenase Enzyme I. Figure 7 shows an alignment of the sequence from *M. capricolum* with those from other sources recovered from the Genbank and TIGR databases. The alignment shows that residues that are totally conserved (shown as reverse shading) are also identical in the sequence from *M. capricolum*. It is noteworthy that all the bacterial sequences are from 40 to 70 residues shorter at the amino-terminal end than those sequences from eukaryotes. A phylogenetic tree (Fig. 8) shows a similar pattern to that observed for ODPA; the protein from *M. capricolum* is most closely related to those from *M. genitalium* (MG273, TIGR, database) and *A. laidlawii*.

After a gap of 28 bases, a new open reading frame (from bases 3461 to 4877), corresponding to the Enzyme II subunit of py-

ruvate dehydrogenase, was identified. An alignment of the *M. capricolum* sequence with other Enzyme II sequences (Fig. 9) shows clearly that the conserved residues (reverse shading) are found in the *M. capricolum* sequence. Boxes shaded in grey cor-

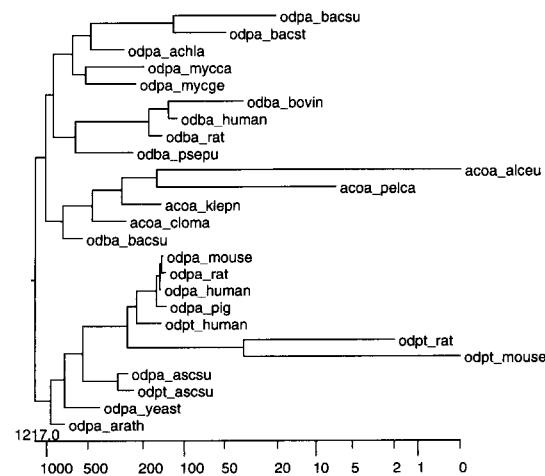


Fig. 6. Phylogenetic tree of sequenced proteins of the Enzyme I- α sub-unit family. Relative evolutionary distances are shown on the horizontal scale. Abbreviations are as in the legend to Figure 5.

1									
odpb-mycca									
odpb-mycke									
odpb-achlia									
odpb-bacsu									
odpb-bact									
odbb-human	MAVAAAAGW	LRLRRAAGAE	GHWRRLPGAG	LARGFLHPAA	TVEDAAQRQ	VAHFTFOPDP	EPRREVGOTOK	MMLFQSUTSA	MAI INNIKAVTDA
odbb-bovin	MAAVAAPAGW	LRLRRAAGAD	GFWRKLGAG	LSRGFLQAS	AYGAAGAQRQ	VAHFTFOPDP	EPEVYQGTQK	MNLFQAVTS	MSKIQ VNNIEALNNA
odbb-rat	CNSAD WP.	AA	LVOGLQPA	.VDDASQKRR	VAHFTFOPDP	ESLYQGOTQR	MNLFQSUTSA	LENSLADNPT	MAI ITLLEAINQA
odbb-psepu									
odpb-ascsu									
odpb-human	MAAVS G.	LVRPPI	REVSGLLKKR	FHWTPAQAQ	VTVRDALNQ	MTMVOAITDA	MTMVOAITDA	MTMVOAITDA	MAI ITLLEAINQA
odpb-yeast									
odpb-arath									
acob-alceu									
acob-peleca									
acob-klepn									
acob-cloma									
11									
odpb-mycca	CTEGVVFR	ATOGIAHVKG	NDFCFNABIS	BAMFACVGLC	MANNGKPKVL	EFOEGGLGLA	SQNIFNTIS	RMRNRTRGKY	TAPMVIRMP
odpb-mycke	GPEGVVFR	ATAGLJQKQYQ	SERWDQDIA	BNSMACJIGVC	AAICLKPIV	IQPSGSFSP	AMQQTIVHAA	KLRNRSRGVII	TAPLVNRP
odpb-achlia	GPEGVVFR	ATAGLJQKQYQ	ETRWDPLBIA	BSAIVGSAVE	MAINGLPKIV	IQFDGFIPF	GTYDLVTHAA	RMRNRSRGQF	MGCGIKALEH
odpb-bacsu	GVNGGVFR	ATEGLQKAEFG	EDRVFDPDIA	PSGIGGLAEL	LGLGFRPV	IQFFGPVYE	VMDVSQSGQNA	RMRYRSRGRW	TAPVMLRFLP
odpb-bact	GVNGGVFR	ATEGLQKAEFG	EDRVFDPDIA	ESGGIGGLAEL	LALOCPFRPV	IQFFGPVYE	VMDSICQCMQA	RIRYRTGGRY	HSEALEVLFQ
odbb-human	AF.GOVFR	CTVGLRDKYQ	KERFVN1BLC	EOGIVGFCIG	IVATICATAIA	IQFQADYIFF	AEDQIVNEAA	KYRYRSDFL	MTSDEGLVLA
odbb-bovin	AF.GOVFR	CTVGLRDKYQ	KERFVN1BLC	EOGIVGFCIG	IVATICATAIA	IQFQADYIFF	AEDQIVNEAA	KYRYRSDFL	MTSDEGLVLA
odbb-rat	AF.GOVFR	CTVGLRDKYQ	KERFVN1BLC	EOGIVGFCIG	IVATICATAIA	IQFQADYIFF	AEDQIVNEAA	KYRYRSDFL	MTSDEGLVLA
acob-alceu									
acob-peleca									
acob-klepn									
acob-cloma									
221									
odpb-mycca	HIPFVQIVCP	STPYDTKG	LAAIDSPDPV	IVVPTKLY		RAFKQ	EVPDEHYIVP	IGEYGIQE	NDLTIVTYGA
odpb-mycke	QIAGLKTVMP	SNPYDTG	LAAIESPDPV	IPFEPKLY		RAFRQ	EIPSDYYTVP	IGEANLISEP	QTVDCQKAIA
odpb-achlia	SIFPLKVTB	SNPYDTG	LAAINDPDV	VFLPEPKLY		RAGKQ	EVPAEYIEIP	IGKAGVVKQOS	LLKEKPHPNAT
odpb-bacsu	OOPF1KVV1Z	SNPYDTAG	LISAIRDNDPV	VFLPEPKLY		RSRFO	EPEEEYITP	LGKADVKRE	SELITIVSYGP
odpb-bact	OOPF1KVV1Z	SNPYDTAG	LISAIRDNDPV	VFLPEPKLY		RSRFO	EPEGEYITP	IGKADIKREG	TMFDLNLVY
odbb-human	HCPG1KVV1Z	SNPYDTAG	LSCIEDKNP	IPFEPKLY		RAAEF	EPIPEVNPIN	LSQAEVQIOP	MTKHMESLKAAD
odbb-bovin	HCPG1KVV1Z	SNPYDTAG	LSCIEDKNP	IPFEPKLY		RAAEF	EPIPEVNPIN	LSQAEVQIOP	MTKHMESLKAAD
odbb-rat	HCPG1KVV1Z	SNPYDTAG	LSCIEDKNP	IPFEPKLY		RAAEF	EPIPEVNPIN	LSQAEVQIOP	MTKHMESLKAAD
odbb-bacsu	NOPELK1KVV1Z	SNPYDTAG	LAAKRDNDPV	LEFPEPKLY		RLIKG	EPPADDVLP	IGKADVKRE	TDLSITYIYGA
odbb-psepu	QCQCLRTW	SNPYDTAG	LIAISCDP	IPFEPKLY		RLIKG	EPPADDVLP	IGKADVKRE	MTKHMESLKAAD
odbb-ascsu	HCPG1KVV1Z	YDCEDARGL	KAAVRDDNDPV	IGCLMELING	GPFDPGDHDRP	VTPWSKHPHS	ADIDHVNVSAA	KSNYMSAGQO	TDLSITYIYGA
odpb-human	HCPG1KVV1Z	YDCEDARGL	KAAVRDDNDPV	IGCLMELING	GPFDPGDHDRP	VTPWSKHPHS	ADIDHVNVSAA	KSNYMSAGQO	TDLSITYIYGA
odpb-yeast									
odpb-arath									
acob-alceu	HIPF1KVV1Z	SNPYDTAG	LAAKRDNDPV	IFCEN1KLYG	LEGEVPE				
acob-peleca	HIPF1KVV1Z	SNPYDTAG	LAASTADDPPC	VFFEPHQKLYG	MKEVPE				
acob-klepn	HIPF1KVV1Z	SNPYDTAG	LAASTADDPPC	VFFEPHQKLYG	MKEVPE				
acob-cloma	HIPF1KVV1Z	SNPYDTAG	LAASTADDPPC	VFFEPHQKLYG	MKEVPE				
331									
odpb-mycca	IDIDLRSIK	PCKKVMIES	VKIKTGRLLV	HEAVKSFSVS	AEIIATVNE	ECEFYIKAPL	SRCTGYDVIT	EFDRG	EGY FQVNPKKVLV
odpb-mycke	IIELDLRTIS	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIITSVTE	ELEFTYKLKAP	QRVTFIDIVV	ELARG	KMQEELLDFKF
odpb-achlia	VELIILRTIS	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIITSVTE	ELEFTYKLKAP	QRVTFIDIVV	ELARG	EKY QFEINARVID
odpb-bacsu	AEVUDLRTVS	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	KAFPHLEAP	VRFTGFDITV	ELARG	AVNOLLK..
odpb-bact	AEVUDLRTVS	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	RAILSLEAPV	LRVAAPD.T	PEFSQAA	EKY QFEINARVID
odbb-human	CEVIDLRTII	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	RAILSLEAPV	LRVAAPD.T	PEFSQAA	AVNOLLK..
odbb-bovin	CEVIDLRTII	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	RAILSLEAPV	LRVAAPD.T	PEFSQAA	AVNOLLK..
odbb-rat	CEVIDLRTII	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	RAILSLEAPV	LRVAAPD.T	PEFSQAA	AVNOLLK..
odbb-bacsu	CEVIDLRTII	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	RAILSLEAPV	LRVAAPD.T	PEFSQAA	AVNOLLK..
odbb-psepu	CEVIDLRTII	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	ECFLNLEAPI	SRVCCYDT	EPPHIFPFF	AVNOLLK..
odbb-ascsu	CEVINLRSIR	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	ECFLNLEAPI	SRVCCYDT	EPPHIFPFF	AVNOLLK..
odpb-human	CEVINLRSIR	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	ECFLNLEAPI	SRVCCYDT	EPPHIFPFF	AVNOLLK..
odpb-yeast									
odpb-arath									
acob-alceu	AEVINLRSIR	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	ECLFLDPLA	KRLGDPIDA	MYAPTMK	AVNOLLK..
acob-peleca	AEVINLRSIR	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	ECLFLDPLA	KRLGDPIDA	MYAPTMK	AVNOLLK..
acob-klepn	AEVINLRSIR	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	ECLFLDPLA	KRLGDPIDA	MYAPTMK	AVNOLLK..
acob-cloma	AEVINLRSIR	PDKOTVFNS	VKIKTGRLLV	TEAVKSPFTS	AEIILMVNE	ECLFLDPLA	KRLGDPIDA	MYAPTMK	AVNOLLK..
431									
odpb-mycca									
odpb-mycke									
odpb-achlia									
odpb-bacsu									
odpb-bact									
odbb-human									
odbb-bovin									
odbb-rat									
odbb-bacsu									
odbb-psepu									
odpb-ascsu									
odpb-human									
odpb-yeast									
odpb-arath									
acob-alceu									
acob-peleca									
acob-klepn									
acob-cloma									

Fig. 7. Alignment of sequences of members of the Enzyme I-β subunit family. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in the listed Enzyme I-β proteins are shown in reverse shading. Abbreviations used and references to published sequences are: *M. capricolum* (odpb-mycca) (this work); *M. genitalium* (odpb-mycke) (MG273 from the TIGR database); *A. laidlawii* (odpb-achlia) (Wallbrandt et al., 1992); *B. subtilis* (odpb-bacsu) (Hemila et al., 1990) and (odbb-bacsu) (Wang et al., 1993); *B. stearothermophilus* (odpb-bact) (Borges et al., 1990), human (odbb-human) (Nobukuni et al., 1990a) and (odpb-human) (Ho & Patel, 1990); bovine (odbb-bovin) (Nobukuni et al., 1990b); rat (odbb-rat) (Zhao et al., 1992); *P. putida* (odbb-psepu) (Burns et al., 1988); *A. suum* (odpb-ascu) (Wheelock et al., 1991); yeast (odpb-yeast) (Miran et al., 1993); *A. thaliana* (odpb-arath) (Luethy et al., 1994); *A. eutrophus* (acob-alceu) (Priefer et al., 1991); *P. carbinolicus* (acob-peleca) (Oppermann & Steinbuchel, 1994); *K. pneumoniae* (acob-klepn) (Deng et al., 1994); *C. magnum* (acob-cloma) (Kruger et al., 1994). odpb, odbb, acob refer to the genes encoding the β-subunits of Enzyme I of the pyruvate dehydrogenase, 2-oxoisovalerate dehydrogenase, and acetooin catabolism complexes.

respond to the lipoyl domains (approximately 70 residues). The proteins from *E. coli* and *Azotobacter vinelandii* contain three such domains; those from *E. faecalis*, *A. laidlawii*, *Alcaligenes eutrophus*, *Haemophilus influenzae*, *Pseudomonas aeruginosa*, *Dictyostelium discoideum*, *Arabidopsis thaliana*, and human contain two; and all the others shown, including the *Mycoplasma*

sequences, contain one lipoyl domain. Each one of the lipoyl domains contains a conserved lysine residue (the site of lipoylation), generally preceded by an aspartyl residue. In the case of *A. vinelandii*, the aspartate of the first lipoyl domain is replaced by alanine. For *Klebsiella pneumoniae*, the conserved aspartate is replaced by serine. All the lipoyl domains are also character-

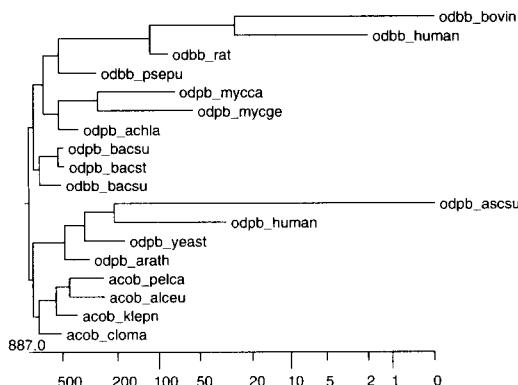


Fig. 8. Phylogenetic tree of sequenced proteins of the Enzyme I- β subunit family. Relative evolutionary distances are shown on the numerical scale. Abbreviations are as in the legend to Figure 7.

ized by the presence of a conserved GD pair. In the case of odp2 from *B. subtilis*, the G is replaced by N. For odo2 of *A. vinelandii*, odp2 of rat, and odp2 of human, the D is replaced by E. In approximately half of the proteins (odp2-psepu, odo2-azovi, odp2-rat, odo2-rat, odo2-human, odp2-entfa, odp2-achla, odp2-alceu, odp2-haein, odp2-pseae, odp2-dicdi, odp2-arath, odp2-human, odp2-ecoli, and odp2-azovi), the lipoyl domain is separated from the remainder of the protein by an A,P-rich linker. It is worth noting that the linker in odp2-dicdi is enriched in serine. The central portions of the proteins contain an E3 binding domain of approximately 50 residues and the carboxyl-terminal 250 residues corresponds to the catalytic domain, which effects the acyl transfer reaction. The conserved histidine (at residue 753) acts as the general base catalyst (Mattevi et al., 1992). A phylogenetic tree (Fig. 10) shows the preservation of the evolutionary relationship of the proteins from *M. capricolum*, *M. genitalium*, and *A. laidlawii*.

Following a gap of 18 bases, an open reading frame (from bases 4796 to 6686) was found; it corresponds to the gene encoding dihydrolipoamide dehydrogenase. The alignment shown in Figure 11 emphasizes a unique feature of the dihydrolipoamide dehydrogenase from *M. capricolum*. This protein contains an aminoterminal lipoyl domain (shown as a boxed grey area), a feature observed previously only in the DLDH proteins from *A. eutrophus* (Hein & Steinbüchel, 1994) and *Clostridium magnum* (Kruger et al., 1994). These lipoyl domains contains the characteristic conserved lysine residue (the site of lipoylation) (Mattevi et al., 1992), preceded by an aspartyl residue (residues 42 and 43). There are a number of totally conserved residues (shown in reverse shading) characteristic of dihydrolipoamide dehydrogenase sequences. The region from residues 171 to 180 corresponds to the motif GXGXXGYXXA, which is a possible nucleotide binding site. The conserved region from residues 206 to 218 corresponds to the signature GGXCLNXGCXP(S/T)K. In this region, the combination of the flavin ring with an adjacent disulphide bridge forms the redox center, which is involved with the transfer of the reducing equivalents from dihydrolipoamide to NAD⁺. The vicinal cysteines in this region may bind lipoic acid and undergo reversible oxidation-reduction. The active species (dimeric) of E3 has four domains: FAD binding,

NAD binding, central, and interface (Mattevi et al., 1992). A phylogenetic tree (Fig. 12) documents the consistency of the relationship of the *M. capricolum*, *M. genitalium*, and *A. laidlawii* sequences.

A space of 21 bases separates the coding sequence for dihydrolipoamide dehydrogenase from the next open reading frame. The region from bases 6707 to 7675 corresponds to the gene encoding phosphotransacetylase. The seven known sequences for phosphotransacetylase are aligned in Figure 13. The proteins from *E. coli* and *H. influenzae* are unique in that they have amino-terminal extensions of approximately 400 residues. All of the sequences show characteristic regions of total conservation, highlighted by reverse shading. A phylogenetic tree (Fig. 14) shows clearly a distinct branch for the proteins from *H. influenzae* and *E. coli* and a separate branch for the proteins from *M. genitalium* and *M. capricolum*.

A 13-base spacer separates the *pta* gene from the next open reading frame, encoding acetate kinase (from bases 7689 to 8866). The alignment shown in Figure 15 compares the six known acetate kinase sequences. The amino-terminal regions are characterized by a highly conserved sequence (residues 12–17, shown in reverse shading), GSSS(I/L)K, that may be involved in nucleotide binding. The phylogenetic tree shown in Figure 16 demonstrates the sequence relatedness of the *M. capricolum* and *M. genitalium* proteins, as well as the *E. coli*, *H. influenzae* similarities. Seventy-one bases separate the end of the sequence encoding acetate kinase from the one encoding *orfA*.

The sequence shown in Figure 2 was examined for regulatory features. Upstream of the *lplA* gene, within the coding sequence of *naox*, there is a possible –35 (TTGACA, bases 180–185, underlined) and –10 (TTTAAT, bases 204–209, boxed) promoter sequence. Upstream of the *odpA* gene, within the *lplA* coding sequence, there is a possible –35 (TTGACT, bases 1226–1231, underlined) and –10 (AAGAAT, bases 1249–1254, boxed) sequence. Upstream of *odpB*, within the *odpA* coding sequence, is found the potential –35 (TTGACA, bases 2328–2333, underlined) and –10 (CAAAAT, bases 2353–2358, boxed) sequence. Within the *odpB* coding sequence, upstream of *odp2*, potential –35 (TTGATA, bases 3350–3355, underlined) and –10 (TAAA AA, bases 3384–3389, boxed) sequences are located. Within the *dldH* coding region, upstream of *pta*, there are two potential –35 sequences (TTGGAA, bases 6523–6528 and TTGCTA, bases 6594–6599, underlined) and two potential –10 (AGAAAT, bases 6550–6555 and AACAAAT, bases 6619–6624, boxed) sequences. Near the end of the *ack* coding sequence is located a potential –35 (TTGTTT, bases 8841–8846, underlined) and –10 (TTAAAT, bases 8864–8869, boxed) sequence. The spacer between the end of the *ack* gene and the beginning of *orfA* (71 bases) is AT-rich (marked with diverging arrows) and may form a stem-loop structure that plays a role in transcription termination of the *ack*-containing message.

Probable ribosome-binding sites are recognizable at appropriate positions before coding sequences. AGGAAAA (bases 270–276, shaded) is found before *lplA*; AGAAAGG (bases 1315–1321, shaded) is found before *odpA*; AGGAGGA (bases 2425–2431, shaded) is localized preceding *odpB*; AGAAAGG (bases 3447–3453, shaded) precedes *odp2*; AGTGAAG (bases 4781–4787, shaded) is localized in front of *dldH*; AGAAAGA (bases 6696–6702, shaded) precedes *pta*; AGGAGAA (bases 7675–7681, shaded) is found in front of *ack*; AGAAAGC (bases 8927–8933, shaded) is found in front of *orfA*.

A

Fig. 9A. (*Figure continues on facing page.*)

B	551	
odp2-mycca	VRKATVKAMT KSHTEIAAFT GMKNTDITET HKMRTELEK, HAASAS.GIKL TYLAFLIIKAV AKSLRDMPNI	IVYRGDFANNK IQFMHNINIG IADTPNGCIV VPIVKGADHL
odp2-mycge	MRAKIAEAMT KSHAIIPPTV LTFYVYNATKL QYRESVNGY ALSKY. SMK SYFAFPFKVAK VNALKKFVVF	IVYGDPMQNE IVLNDDINVQ IVDTTEEGCI VPIVKQAQTK
odp2-bacst	IRRAIAKAMV HSXHTAPHRV LMDEADVTKL VAHRRKKFKA IAEEK. GIKL TFLPYVVKAL VSALREYPVLF	IVSISDTEETE IIQKHYNIG IADTDRCGQ VPIVKHADR
odp2-bacsu	IRKAIAKAMV NSXHTAPHRV LMDEVDTVNLS VAHRRKKFQKQ. VAADQ. GIKL TFLPYVVKAL TSALKKFVVL	IVSISDCTDE VIQKHYFNIG IADTDTEGQ VPIVKVNAADR
odc2-bacsu	RQQTIAKMLR EVQOTSAMLT FTVNEVDTMAV MNLRKRRKDQ FFE. QNEVKL GFMSFPFTKAV	VAALKKYPPL NAEIQQ.. DE LIVVKFYDID IVNAAVEGCV VPIVVRADRL
odb2-bacsu	VRKAIALSNMK RSKTEI PHAM TMMEVDVTNN VAYRNNSIKDQ FKKTE. GFNL TFFAFPVKAV AQLKEFPOI	NSMM.. AGDK IIQKQDINIS IVVATEDSF VPIVKINADEF
acoc-cloma	IRKIIASRHM ESWITSEPTVY YDIKVUDMTSL KRKFDAKDV CKV TTVDLIVKIV SKVLLQFPLL NCIN.. GNE LITRNYNVNG VVVAIDGGM VPIVVKYANEK	IVQFDEAQS IRRTEDADIS VVALPAGI TPIVRAERK
acoc-klepns	MRAIASRLQ TSKQSPHPF LSVLDLQLERL LALRQDINIE VPG. VPKI SVNDLVLKAV ALALVAVPDV	IVQFDEAQS IRRTEDADIS VVALPAGI TPIVRAERK
acoc-pelca	IGAAISNTVQ NSK TIPOPF VTMGEEA KEPFRAGLKAE GKA. V SMNDMVIRAL GKAIEQYPMV	IVATLGKEYG L.. NADVNI VVAGTDDAII MPVVKGCQAL
odb2-psepu	DLRKLIQRMQ DAKRRVAFHS YVEEIDVTAL SALRQQLNS. KHCDGS.RGKL TLLPFLVRLA VVALRDFPQI	IVATYDDEAQI ITRHGAHVRS ITQGDDNCI VPIVLRHAEAG
odc2-azovi	LRAKVAERL EAQSSMAMLS TFNEVNMPKV MELRAKYKDL FETHINGVRL GFMSFPVFKAA VEALKRQPGV	IVASIDG.. ND IVVHYGQDQG VVSSSDRGCV VPIVLRNAEFL
odc2-ecoli	LRAKVAERL EAKNSTAMLS TFNEVNMPKI MDLRLQVGEA FERKR. GIRL GFMSFPVVKAV VEALKRQPGV	IVASIDG.. DD IVVHNHYFDVQ MVVSTPRGV VPIVLRDVTDL
odp2-staur	MRAIAKAMV NSXHTAPHRV LMDEIDVQAL WDHRKKFKE, 1AAEQ. GTKL TFLPYVVKAL GIAASKRVPVY	IVVHKHYWNIG IADTDRCGQ VPIVVKHADR
odp2-neucr	MRTKIIARL ESVTENPHF VSTNLNSVSKL LKL7QALNNS ADGR. YKL SVNDLFLKAV GIAASKRVPVY	IVOFETFVDS VVATPNCQI TPIVVKYVEGK
odp2-yeast	MRSII1GERL OSTQGIPSYI VSSKSISISKL LKLRLQSLNAT AND. YKL SINDLVLKAV TVAAXKPVDPD	IVAYLPNPNENV IRKFKNVNDVS VVATPTGQI IVVKNEAK
odp2-rat	IRTAQIORM OSKOTI PHYI LSVDNVNGEV LLVRLKELNLB GKCK. GKT SVNDLFLKAVS ALACLKVPED	IVSSWM.. DTW IRONHVVDVS VVATPAGI TPIVNAHIC
odo2-rat	MQRQIAQRL BEAQNTAMLT TFNEVDSMSI QEMBRARKHQL FKLKHN LKL GLMSAVFKAS AFALQFQPVIL	IVAVIDATDQV VVATPAGI TPIVNAHIC
odb2-bovin	FHKAVKMTMS AA_LKIPHFG YCDEVDTLTEL VKLREELKPI AFAR. GIKL SMFPMFLKAA SLCLLQFPLV	IVASDENCON ITYKASHNIG IADTDTEGQI VPIVVKVNOQI
odc2-human	FOKAMVTTMS AA_LKIPHFG YCDEVDTLTEL VKLREELKPI AFAR. GIKL SMFPMFLKAA SLCLLQFPLV	IVASDENCON ITYKASHNIG IADTDTEGQI VPIVVKVNOQI
odc2-yeast	MQRQIAQRL BEAQNTAMLT TFNEVDSMSI QEMBRARKHQL FKLKHN LKL GMFSAVFKAS AFALQFQPVIL	IVAVIDDTKE VVVRDYDID VVATPAGI TPIVNAHEAM
odp2-entfa	TRKAIAKAMV NSXHTAPHRV LHDVEEVSKL NDHRKKFKE, 1AAEQ. GTKL TFLPYVVKAL TSTVOKFPLI	IVKAGIE, DD IVVYRDYDID VVATPAGI TPIVNAESEL
odp2-achla	LRAKVAERL ESVTENPHF VSTNLNSVSKL LKL7QALNNS ADGR. YKL SVNDLFLKAV GIAASKRVPVY	IVASIDAAQE IVVKYKNFNIS IADTDHGEY VPIVVKANTK
odp2-azovi	INKISGANLH RNWVMPHIV NHDEADTIL EAFRFLNLNEK NEXS. GIKV TMLAFMIFAT VAAALKFPFV	IVASIDAAQE IVVKYKNFNIS IADTDHGEY VPIVVKANTK
odp2-azcuv	INKISGANLH RNWVMPHIV NHDEADTIL EAFRFLNLNEK NEXS. GIKV TMLAFMIFAT VAAALKFPFV	IVASIDAAQE IVVKYKNFNIS IADTDHGEY VPIVVKANTK
odp2-haein	TFNEVDSMSI QEMBRARKHQL FKLKHN LKL GLMSAVFKAS AFALQFQPVIL	IVASIDGDNLW LKKYFNIG FRADTPNGCV VPIVKAADKK
odp2-psse	TFNEVDSMSI QEMBRARKHQL FKLKHN LKL GLMSAVFKAS AFALQFQPVIL	IVASIDGDNLW LKKYFNIG FRADTPNGCV VPIVKAADKK
odp2-dicdi	IRKVTAAART QFSKTPHIVY LTMECRVDSLK LTLAERLQKPLH LESKQKIPHL YLQSDQVULP LLAFLRKLQELQE NHG.. VVKV SVNDLFLKAV	IVTDFRDPV VVATPAGI TPIVNAHIC
odp2-arath	IRVIAQIRM QSKTQHIVY LSIDQVNLNGRV LLLRKELNKI LSCV. SKV SVNDLFLKAVS ALACLKVPED	IVSSWM.. DTW IRONHVVDVS VVATPAGI TPIVNAHIC
odp2-human	IRVIAQIRM QSKTQHIVY LSIDQVNLNGRV LLLRKELNKI LSCV. SKV SVNDLFLKAVS ALACLKVPED	IVSSWM.. DTW IRONHVVDVS VVATPAGI TPIVNAHIC
odp2-ecoli	IQKISGANLH RNWVMPHIV HFDKTDITEL AFTRKQONEE AAKRKLQDNL VTFVNFVPL	IVSSLSSEDQR TLTKYKINIG VVADTPNGCV VPIVFKDWNKX
odp2-azovi	LMQIGATNLH RSWLNVPVHT QFESADITEL AEFRVQAQKAV AEKA.. VVKV TVPLLKAV	IVSSLSSEDQR TLTKYKINIG VVADTPNGCV VPIVFKDWNKX
	661	
SVE1AIKIS	ELANKAKDGK LTRAEMTEAT FTVSNFQCSV. GLDYATPII SDE.SAILGV GTMSQPLYYI NGELOKRF.. IMPLSMT CIIHRIIDGADAGRFLVQD
odp2-mycca	SVE1AIKIS ELANKAKDGK LTRGDLNKGQ ISVINPFGSL. GAAVGTPIQ YEE. MCIVAT GNLERIEIVK VVGI. AVH.. TILPLTIA AIIHWRVQGAD VGRGKEIAK
odp2-mycge	SVE1AIKIS ELANKAKDGK LTPGEMKGAS CTITNIGCA. GGWFPTPVIN HEE.VAILGI GRIAEPKIV DGEIVAAPI.. MLLASLS FIERMIIGAT AQKALNHIIK
odp2-bacst	SVE1SEDEIN GLATKAREKG LAPAEMGKAS CTITNIGCA. GGWFPTPVIN HEE.VAILGI GRIAEPKIV DGEIVAAPI.. VLAISLS FIERMIIGAT AQNALNHIIK
odc2-bacsu	TFAGIEKEIG ELAKKARLNNK LTLSLEEGGS FTITNGCFT. GMSMTPILN SQQ. VGLGOM HIKIQLRPAV LDEERFENR.. PMMYIALS YIIRVICKE AVGFLVTTIKN
odp2-bacsu	TIGKIAKIDT GLAKKVRDQK LTADDMQGQT FTITNGCFT. GMSMTPILN SQQ. VGLGOM HIKIQLRPAV LDEERFENR.. DMVNLCLS LIIHVRLLGIVL CGRFLGRVQ
acoc-cloma	GLKESTEIVK LDKAKMSQN LKPENMFTGGT FTITNQCMF. GIEYFSPNIIQ EOE.VAILGV NKITETPVVQ NGEIVIKP.. LMNLSLS AMBRAVISV AAQFLSKVKE
acoc-klepns	SISDISENIEH SLVTRAKAGK LDKAKMSQN LKPENMFTGGT FTITNQCMF. GIEYFSPNIIQ EOE.VAILGV NKITETPVVQ NGEIVIKP.. LMNLSLS AMBRAVISV AAQFLSKVKE
acoc-pelca	SLEVASVASS AVIDVKVAGT CGPAEMAGGN FAISNLCM. GVRQFDAINI PEO. SAILAL GAGEVRVVV DGQIVARQ.. QMTVSLSL CIBRVIICAA GAFLRNSSD
odb2-psepu	SLWANAGEIS RLJANARAQNNS ASREELSGST ITLTSVPLM GVD. SGFALV PGNCMSLAVV GGIKDEVVQV GMEMPV.. VSTMKVTLSL AIIHWRVQGAD SAQFLVELKR
odc2-azovi	SLAEIEGGN EFGKGRDAGK LTIEEWTGQT FTISNGCFV. GSLLSPITVN PEO. TAIIQML HIKIQLRPAV NQO. VEIL.. MMNLSS FIBRVIQGMD ALALFIQAVRG
odc2-ecoli	GMDAIEKKK ELAVKGRDQK LTVEDLTTGGN FTITNGCFT. GSLLSPITVN PEO. TAIIQML HIKIQLRPAV NQO. VEIL.. PMMYIALS YIIRLIIQCE AVTFLVTTMD
odp2-stau	SIFQISDEIN ELAVKGRDQK LTADEMMGKAT CTITNIGCA. GGWFPTPVIN HEE.VAILGI GRIAQPKIV DGEIVAAPI.. PMMYIALS YIIRLIIQCE SVGLVTTIKE
odp2-neucr	GLESISAIAK LPEKXPEYOGGS ISISNCHMNP OASQFTAIIQ PEO. AIIALV GAPOKPVAVVW EMEDGTGVS WDEQIIVTAVS FIBKVKVIGAV GAEWIRELK	... VLAISLS FIBRQDIDGAT QONAMNHIIK
odp2-arath	GLSISQNEIK ELVKKARLNNK LAPEEFQFGGT ICISNCHMNP AVMFTPSIJN PEO. STILAL ATVERV.. V DAAAENGFS FDNOVITGQ FIBRITDAG GAEMKELKT	... RMMKLSSL FIBRIVIGAT QAKAMNNIKR
odp2-azcuv	GLETIASDVG SLASKARSKQ LQHPEFOFGGT FTISNLCMF. GIKNSFAINI PEO. STILAL ATVERV.. V DAAAENGFS FDNOVITGQ FIBRITDAG GAEMKELKT	... TLPPLSLA VVHRIIDGAD GGRFLMRVK
odp2-rat	NYADIERTIN ELGEKARNEK LAIEMDMDGT FTISNCGV. GSLGFTPIV PEO. SAILGM HGIFDRPVAVV GKK. VEVR.. IIELWM
odp2-bovin	SIEFATEILN RLQKLQSAQG LSTNDLIGGT FTISNCGV. GGTYAKPVLN PEO. VAIAGL GTIKALPFRN EXGEVKA.. PMMYVALT YIIRLIIQCE AVTFLVTTMD
odc2-human	SIFATEILN RLQKLQSAQG LSTNDLIGGT FTISNCGV. GGTYAKPVLN PEO. VAIAGL GTIKALPFRN EXGEVKA.. QIMNNVWSL AIIHWRVQGAD VSRFSNLWK
odc2-yeast	NFADIERTIN ELGEKARNEK LAIEMDMDGT FTISNCGV. GGTYAKPVLN PEO. VAIAGL GTIKALPFRN EXGEVKA.. QIMNNVWSL AIIHWRVQGAD VSRFSNLWK
odp2-entfa	SLVDIENEIV RLSHARKDQK LTLEDMTGGT FTISNCGV. GSLLGFTTIN PEO. SAILGM HGIFDRPVAVV GKK. VEVR.. PMMYVALT YIIRLIIQCE AVTFLVTTMD
odp2-achla	SMFAIDEIN EKAALAIKSL LTQADMRTGGT ITISNIGSV. GGWFPTPVIN HEE.VAILGV CTIAQEPPVVV ADGEIVVG.. VLAISLS FIBRQDIDGAT QONAMNHIIK
odp2-azcuv	SEFASLOVSLA SLADDITIARK ISMDQCTRGDT FTITNPFSA. GIAEFGPVIV YEE.LAILGI GKIDRKPWVW GNEIKIAH.. RMMKLSSL FIBRIVIGAT QAKAMNNIKR
odp2-haein	YLVLESOEMS ELAKKARDQK LXPDMQGGC FS1SSSLQGL GGTYTFPIV PEO. VAIIMGV CKSYQKPV WW DGKQFAPR.. TLPPLSLA VVHRIIDGAD GGRFLMRVK
odp2-psse	GYIELSERELM EVSKAREKGK LTASDMQGGC FTISSLQGL GGTYTFPIV PEO. VAIIMGV CKSYQKPV WW DGKQFAPR.. LTLPLSLA VVHRIIDGAD GGRFLMRVK
odp2-dicdi	SLLQLAEEAA ELADKARNK LXAADMQGAC FTISSLQGL GGTYTFPIV PEO. VAIIMGV CKSYQKPV WW DGKQFAPR.. LTLPLSLA VVHRIIDGAD GGRFLMRVK
odp2-arath	SISIAELEVK ELQAKRQNGK LHPSEFEGST FTISNLCM. GIKQFAAVIN PEO. AIIALV .. QKLVSF LSNKPDSPYE TATILSVTLA CIBRVIQGAV CAEWLKSFKD	... LMLPMSL SIBRVIQGAD AARTPKRLOE
odp2-human	GVETIANDVWV SLATKAREKGK LPHEFQGGT FTISNLCM. GIKQFAAVIN PEO. AIIALV .. QKLVSF LSNKPDSPYE TATILSVTLA CIBRVIQGAV CAEWLKSFKD	... LMLPMSL SIBRVIQGAD AARTPKRLOE
odp2-ecoli	GIIELSERELM TISKSKARDQK LTAGEMQGGC FTISSLQGL GTTHFAPIVW PEO. VAIIMGV SKSAMEP. VV NGKEFVPR.. LMLPMSL SIBRVIQGAD GARFITIINN
odp2-azovi	SLLQLAEEAA ELAEKARSKK LGADAMQGAC FTISSLQGL GTTHFAPIVW PEO. VAIIMGV SKSAMEP. VV NGKEFVPR.. LMLPMSL SIBRVIQGAD GARFITIINN
	771	
odp2-mycca	YLS.. PVPV LFM..	
odp2-mycge	QIEELIDLTV A..	
odp2-bacst	LLS.. DPEL LIMEA..	
odp2-bacsu	LLN.. DPQI ILMEA..	
odc2-bacsu	LLEDEPQ.. LLLLEG..	
odp2-bacsu	LLEDEPQ.. LLLLEG..	
acoc-cloma	YME.. KPEL LML..	
acoc-klepns	YME.. KPEL LML..	
acoc-pelca	LLENPEEL..	
odp2-psepu	LLENPEEL..	
odc2-azovi	LLE.. OPAC LFVE..	
odc2-ecoli	LLEDPAR.. LLLDV..	
odp2-stau	LLEDPTR.. LLLDV..	
odp2-bacsu	LLN.. NPEL LLMEG..	
odp2-bacsu	LLN.. NPEL LLMEG..	
acoc-klepns	LEIDESIKTS VY..	
acoc-pelca	LEIDESIKTS VY..	
odp2-yeast	VIEPLLELL..	
odp2-rat	VIEPLLELL..	
odp2-bovin	VIEPLLELL..	
odc2-human	YLENPAMFLM DLK..	
odc2-human	YLENPAMFLM DLK..	
odp2-human	AVEDPVR.. LLLDL..	
odc2-yeast	LIEDPRKCY GDLKFAAHTN LIS	
odp2-achla	LLA.. DPEL LLLMS..	
odp2-azcuv	LLA.. NPEL LLLMS..	
odp2-haein	LLADFRILL..	
odp2-psse	LLADFRILL..	
odp2-dicdi	YVENPIKLL..	
odp2-arath	NFEDVRLRLL..	
odp2-human	YLERPITML..	
odp2-ecoli	TLSDIRRLVM..	
odp2-azovi	LLADIRALL..	

Fig. 9. Alignment of sequences of members of the Enzyme II family. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in the listed Enzyme II proteins are shown in reverse shading. Boxed shaded regions correspond to lipoyl domains. Abbreviations used and references to published sequences are: *M. capricolum* (odp2-mycca) (this work); *M. genitalium* (odp2-mycge) (MG272 from the TIGR database); *B. stearothermophilus* (odp2-bacst) (Borges et al., 1990); *B. subtilis* (odp2-bacsu) (Wang et al., 1993), (odp2-bacsu) (Wang et al., 1993), and (odp2-bacsu) (Carlsson & Hednerstedt, 1989); *C. magnum* (acoc-cloma) (Kruger et al., 1994); *K. pneumoniae* (acoc-klepns) (Deg et al., 1994); *P. carbinolicus* (acoc-pelca) (Oppermann & Steinbuchel, 1994); *P. putida* (odb2-psepu) (Burns et al., 1988); *A. vinelandii* (odo2-azovi) (Westphal & de Kok, 1990) and (odp2-azovi) (Hanemaaijer et al., 1988); *E. coli* (odo2-ecoli) (Spencer et al., 1984) and (odp2-ecoli) (Guest, 1987); *Staphylococcus aureus* (odp2-staur) (Hermila, 1991); *Neurospora crassa* (odp2-neucr) (Kreader et al., 1989); yeast (odp2-yeast) (Niu et al., 1988) and (odo2-yeast) (Repetto & Tzagoloff, 1990); rat (odp2-rat) (Gershwin et al., 1987) and (odo2-rat) (Nakano et al., 1991); bovine (odb2-bovin) (Lau et al., 1988); human (odp2-human) (Lau et al., 1992), (odp2-human) (Thekkumkara et al., 1988), and (odo2-human) (Nakano et al., 1993); *E. faecalis* (odp2-entfa) (Allen & Perham, 1991); *A. laidlawii* (odp2-achla) (Wallbrandt et al., 1992); *A. eutrophus* (odp2-alecu) (Hein & Steinbuchel, 1994); *H. influenzae* (odp2-haein) (H1232 from the TIGR database); *P. aeruginosa* (odp2-psse) (Genbank accession no. U47920); *D. discoideum* (odp2-dicdi) (Genbank accession no. u06634); *A. thaliana* (odp2-arath) (Guan et al., 1995). odb2, odb2, and odo2 refer to the genes encoding the Enzyme IIs of the pyruvate dehydrogenase; α -oxo acid dehydrogenase and 2-oxoglutarate dehydrogenase complexes, respectively; and acoc refers to acetooin dehydrogenase EII.

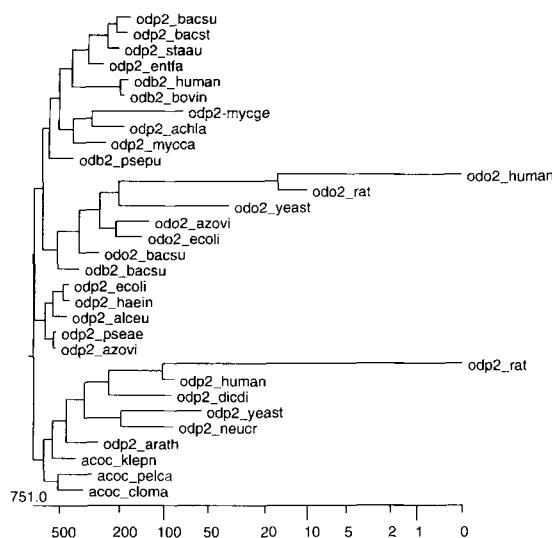


Fig. 10. Phylogenetic tree of sequenced proteins of the Enzyme II family. Relative evolutionary distances are shown on the numerical scale. Abbreviations are as in the legend to Figure 9.

Gene transcription analyses

The cloned sequence encodes 13 genes that might be transcribed in the same direction (see Fig. 1). The only significant gap between sequences occurs between the *ack* and *orfA* open reading frames. The program TERMINATOR was used to search the cloned sequence for potential transcription termination sequences (designated by diverging arrows in Fig. 2). Nineteen bases downstream of the translation termination codon for *ack*, there is a possible transcription termination region (from nucleotides 8886–8914). Using the program FOLDRNA, the stability of this stem-loop structure was calculated to be -11.8 kcal/mol . The stem-loop structure formed contains a perfectly matched stem 12 bases long. It is therefore possible that the region from *naox* to *ack* is transcribed as a single mRNA. Evidence was presented previously (Zhu et al., 1994) that the *ptsI* and *crr* genes constituted an operon. Consequently, it may be the case that the *orfA* and *kdtB* genes are cotranscribed.

Discussion

The sequence analysis presented here (see Figs. 1, 2) demonstrates the presence in *M. capricolum* of a unique arrangement of genes involved in the metabolism of pyruvate. In other bacterial species where gene mapping has been performed (*E. coli*, *H. influenzae*, *M. genitalium*), the genes are somewhat scattered throughout the genome. For example, in *E. coli*, the genes encoding the pyruvate dehydrogenase complex are located at 3 min on the genetic map, whereas the *ack*, *pta* and *ptsH*, *ptsI* and *crr* genes are in the 50–52-min region. In contrast, in *M. capricolum*, all the relevant genes are clustered in a single region. The positioning of these genes close to the *ptsI-crr* operon (whose products use PEP to form pyruvate) may also be of some regulatory significance. The scheme shown in Figure 17 indicates that all the enzymes required for the conversion of PEP to acetate and ATP, including those activities used for lipoylation of

the pyruvate dehydrogenase and regeneration of NAD from NADH, are accounted for in the region of the *M. capricolum* genome sequenced in this study (see Fig. 1).

We reported previously (Zhu et al., 1993) that the HPr protein from *M. capricolum* has an unusually high isoelectric point. Using the computer program PEPTIDESORT, we calculated the pIs of the protein sequences deduced in this study. The pIs of the E1 and EII proteins of the pyruvate dehydrogenase complex were in the range 5.3–6.75 and were similar for both *M. capricolum* and *M. genitalium*. However, several of the proteins from *M. genitalium* showed significantly higher pIs than the corresponding proteins from *M. capricolum*. For DLDH, the pIs for *M. capricolum* and *M. genitalium* were 5.35 and 7.08, respectively. For lipoate–protein ligase, the pIs for *M. capricolum* and *M. genitalium* were 6.97 and 9.29, respectively. For phosphotransacetylase, the pIs for *M. capricolum* and *M. genitalium* were 5.24 and 7.48, respectively. For acetate kinase, the pIs for *M. capricolum* and *M. genitalium* were 6.67 and 9.09, respectively. The significance of the widespread occurrence of proteins with high isoelectric points in *M. genitalium* remains to be clarified.

In order to evaluate the possibility that the genes encoding the enzymes involved in pyruvate metabolism are part of a polycistronic operon(s), northern blotting and primer extension experiments using probes derived from the plasmids described in Figure 1 were performed. No detectable mRNA species were found. Similar lack of success in detecting *ack*- or *pta*-specific mRNA species was reported for *Methanosaeca thermophila* (Latimer & Ferry, 1993) and it was suggested that the mRNAs may have a short half-life or be degraded rapidly during the RNA preparation. Consequently, the nature of the transcripts encoding the enzymes for pyruvate metabolism remains to be established.

The dihydrolipoamide dehydrogenase from *M. capricolum* is atypical, containing an amino-terminal lipoyl domain (see Fig. 11). This structure is also shared by the enzymes from *A. eutrophus* and *C. magnum*. This observation opens the possibility that effective function of the pyruvate dehydrogenase complex may be possible with the association of lipoate residues with either the E2 or E3 components. It is interesting to note that a recent description of an outer membrane protein from *Neisseria meningitidis* (de la Sierra et al., 1994) indicated that it contained an amino-terminal lipoyl domain and was otherwise homologous to lipoamide dehydrogenases.

The *M. capricolum* pyruvate dehydrogenase and *C. magnum* acetoin dehydrogenase complexes show a unique similarity. Each of these complexes contain an E2 with a single lipoyl domain, as well as an E3 with a single lipoyl domain. Comparison of the sequences of these lipoyl domains (see Figs. 9, 11) show that they are almost exact duplicates. This suggests that the lipoyl domains in the E3 proteins of these organisms arose by a duplication of the preexisting domain in the genes encoding the E2s.

In contrast, the pyruvate dehydrogenase complex of *A. eutrophus* is characterized by an E2 with two lipoyl domains, as well as an E3 with a single lipoyl domain. In this case, the two lipoyl domains of E2 are essentially identical, consistent with the idea that they arose by a duplication mechanism. However, the single lipoyl domain in the E3 diverges considerably from the sequences in the E2. Therefore, it seems unlikely that the lipoyl domain of the E2 in this organism arose by a simple duplication of the sequences in the gene encoding the E2.

A	1	MIFUKEPKADI GEGGLTECTVA EULVVKCDTV KECOPLYFVE DIAUNSEILPS PVAKCALIN ISTCOEIKVW DVVTEIDDS STTASTSKV EVVEENASVU GATPVNSDVL M:VIEWKVPD IGDPLDAEVI EULVKAQDTV EVEQSLIVLSDKASMDVPS SAACKVUVK VRKGDKVQGQ AVICTEAQO AAAAPAPAOA PAPAOAPAPA AAAPAPAPAA acol-cloma MAKIVVMPKL GLUTMTEGTLV TWKAEGDDQV KVGEILFEVS IDKUTNEVES SDHGIVRLL VNEGGVVVECL NPVAIIGSAU EDISLINGS SEGSGSQAEQS DTAKPKEVE	110
	111	PSRAPKPKAE AKVEVVEENA SVVGATPVSN DVLPSPRKPK VKEAPKVDQ IEDTPFVVCV CAGIGYVTA IKSAGQLGLKT LIIEK..... EYVGGCCL NWGCPKTL dldh-mycca .ASHSGGA..... DQICMELV..... PPGPGVSA..... PRAADLGMLT VLVER..... STGGCCL NWGCPKSL dldh-alceu AV..... MSKEYEII..... GPPGVWA..... IAAQOYAKV ALVEK..... ESTGGCCL NWGCPKVL acol-cloma M..... MDYDLII..... GPPGVWA..... IYAGHKKLKT LVIEK..... EYVGGCCL NWGCPKTF dldh-achla M..... MHDKYDVLLI..... GPPGVWA..... IRAGOLGLRT VLVER..... OHGGCCL NWGCPKAL dldh-pea M..... MAMA NLARRKGYSI LSSETLRYS FSLRSRAFAS GSDENDVII..... GPPGVWA..... IAAQOLGFKT CLIEK..... RGAIIGGCCL NWGCPKAL dldh-trryb M..... MRPRC FPI..... FNPFYDVV..... GPPGVWA..... IAAQOLGFKT ACVKE..... RGAIIGGCCL NWGCPKAL dldh-pig MOSWSRVYC TLAKRGHFNRI IAHGLOGVSA VP..... LRTYAD OPIDADTV..... GPPGVWA..... IAAQOLGFKT VCIEK..... NETGGCCL NWGCPKAL dldh-human MOSWSRVYC SLAKRGHFNRI IAHGLOGLSA VP..... LRTYAD OPIDADTV..... GPPGVWA..... IAAQOLGFKT VCIEK..... NETGGCCL NWGCPKAL dldh-yeast M..... MLR IRSLLNNKRA FSSSTVRTLTI NKSII DVII..... GPPGVWA..... IAAQOLGFKT ACVKE..... RGKGGCCL NWGCPKAL dld3-psepu M..... M..... MKSYDVII..... GPPGVWA..... IRAGOLGLTV ACVKE..... RSTGGCCL NWGCPKAL dldh-psf1 M..... SOKFDVWII..... GPPGVWA..... IAAQOLGFKT ACIEKYIGKE GKVAVLGGCCL NWGCPKAL dldh-azovi M..... MSQKFIVVII..... GPPGVWA..... IKSAGQLGLKT ALIEKYKGKE GKTAAVGGCCL NWGCPKAL dld2-psepu M..... TQKFIVVII..... GPPGVWA..... IAAQOLGFKT ACIEKYDAE GKLAVGGCCL NWGCPKAL acol-pelca M..... ADEEFLDVLV..... GPPGVWA..... IAAQALGMKV AVV..... SRPTGGCCL NEGGCCL NWGCPKAL dldh-ecoli M..... M..... SE TEIKTVVVL..... GPPGVWA..... FRCADLGLET VLVER..... NTGGCCL NWGCPKAL dldh-hasin M..... MS KEIKTVVVL..... GPPGVWA..... IAAQALGMKV VLVER..... STGGCCL NWGCPKAL dldh-bacst M..... MVVGD FAIETTETLV..... GPPGVWA..... IAAQALGMKV VLVER..... GNGGGCCL NWGCPKAL dldh-bacsu M..... MVVGD FPIETTETLV..... GPPGVWA..... IAAQALGMKV VLVER..... ATGGCCL NWGCPKAL dldh-stau M..... MVVGD FPIETTETLV..... GPPGVWA..... IAAQALGMKV VLVER..... GNGGGCCL NWGCPKAL dldh-halvo M..... MVVGD FPIETTETLV..... GPPGVWA..... IAAQALGMKV VLVER..... DAYGGCCL NWGCPKAL dld1-psepu M..... MQ QTITTLII..... GPPGVWA..... IAAQALGMKV VLVER..... QALGGCCL NWGCPKAL	220
	221	LKTSVHVYHD VHAKAKELGIV LQNTENVVID WAQALERKNG VVKRPTGCVK YLLDDKNKVTQ IKGEAIALDK NTISV NNKN..... YRVN NLVIAASGTP dldh-mycca LHNAAVIDEA KALAHH..... GILGEEKAKID NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-alceu LHNSSLQVTEM KEGD..... KLG DIEGGSIVVN WKHQIKRKKI VIKKVSQSVS GLTCKNCVW IKOTAFESK NWGCPKTL NWGCPKTL acol-cloma LKSAKVFNTV KKSMDFG VSTSGEVGFW WSKV1KSRKD VPKQDINGVVA FLLKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-achla LKRAKIDYL VHA..... DYG ITINGAKLD WKLQKQKQE VDKDQDQVCR ITIIGKAKVW IECAETVDAK NWGCPKTL NWGCPKTL dldh-pea LHLGAEVAHTH THAS..... QLG ISV GEVNVNQ LQKLVQPSRT VSQDLCVPA YLLKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-trryb LTHATHMYHDA HA..... NFEV RYV LMGGAGVMTD VAKMCKQKKE SVNGLSCVW YLLKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-pig LNNSHHYHMA HGTDPFA SRG IEM SEVRLN LAMKMEQKST AVKALGKJIA HLPKQNKVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-human LNNSHHYHMA HGTDPFA SRG IEM SEVRLN LDKMMEQKST AVKALGKJIA HLPKQNKVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-yeast LNNLSPHQM H. TEAQ. KRG IDVNGDIKIN VANPOKAADK AVKALGKJIA LLFQKNCVW YLNGFSPEDE NWGCPKTL NWGCPKTL dld3-psepu LHASRHYHEAN SGDEFA. HLG IEVKP..... TLA LAQMMGKQD SVTGLCQVW YLPRGLCVV NWGCPKTL NWGCPKTL NWGCPKTL dldh-psf1 LHASRHYHEAN KE. AFK. VHG IEA. KGD..... TLA LAQMMGKQD SVTGLCQVW YLPRGLCVV NWGCPKTL NWGCPKTL NWGCPKTL dldh-azovi LDSSYXPHEN HE. SEP. LHG IET. VAVDAD..... TLA LAQMMGKQD SVTGLCQVW YLPRGLCVV NWGCPKTL NWGCPKTL NWGCPKTL acol-pelca LHSWVQHNA RD. KFD. MNG IEL. PAPLNK LAKMRKKEC VSDP..... LPAFLLKQVW LDFKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-ecoli LHWAKVIEEA KALAHH..... GIVPGEPKTD IDKIRTMKKE VLGKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-hasin LHWAKVIEEA KHANHH..... GIVYPSERPIE LDEVRACKAE VKAQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-bacst LISHARHREQA KHSEH..... M..... GIKAENAVTID FAKVYEWKAS VPKLQKNCVW GLLKKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-bacst INAGRHENBNA KHSDD..... M..... GITARENVTVD FTKVYEWKAS VPKLQKNCVW GLLKKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-stau LHASHRHPVAA OHSEN..... L..... GIVIAESVSN POKVQEEPKSS VPKLQKNCVW GLLKKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-halvo ITGANLHAEA QNAEE..... M..... GIADPVG. WD MSQLRDWKSG VPKLQKNCVW GLLKKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dld1-psepu IHAEQFHQA SRKTFP. SPL GISVSPRLD IQQSVAWQHQD IVDRPQVWV ALLKKHGVK VPKLQKNCVW NWGCPKTL NWGCPKTL	330
	331	NHPLPLPGPDQ GRKGDIILDS TGLSVPKIE ETLVVII GCVV IGPFFSLPA SLGCKTVLWQ GLPTILE. ML DKDIDAMTK ELKNRNYNIQV ITNASVKEP DGSVYVQI.. dldh-mycca VKLPLPGPDQ ED. P.R. TDS TGALPQVWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-alceu P.FPEI EG. N. KLSGVIDS TGALSLESNP ESIAIIGVW IGPFFASIN SLGCKTVIIE MLPHILP. PM DREISEIKA LKIRDGI. NI NNNCKVTRIE QGEDGLKVSF acol-cloma RLDSPHBLA RD. KFD. MNG IEL. PAPLNK LAKMRKKEC VSDP..... LPAFLLKQVW LDFKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-achla RLDSPHBLA RD. KFD. MNG IEL. PAPLNK LAKMRKKEC VSDP..... LPAFLLKQVW LDFKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-pea RALPGLADPQ E. HIWYTY..... ALRPLKIP LKSLLNGCQK IGPFFASIN DLGCKTVLWQ LQASLQVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-trryb TALPFLPDPD K. VV..... TSS TGALQALQWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-pig TPFPGITIDE D. TV..... VSS TGALSLKKWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-human TPFPGITIDE D. TI..... VSS TGALSLKKWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-yeast TPFPGIEVDE E. KI..... VSS TGALSLKEDR KLTIIIGCQI IGPFFASVW LQGCKTVW NWGCPKTL NWGCPKTL NWGCPKTL dld3-psepu TPLPGVTDID Q. RI..... VSS TGALSLPQVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-psf1 VEIPPLPQDS D. VI..... VSS TGALLEQPAVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-azovi VEIPPLAPVQD D. VI..... VSS TGALDFQNVN KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dld2-psepu IDIPPAVQDQ D. VI..... VSS TGALDFQPAVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL acol-pelca AVQPGVITDID Q. VI..... VSS TGALDFQPAVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-ecoli IQLPFIPIED P.R. WDS TDALELKEDV ERLLVWNGCQI IGPFFASVW LQGCKTVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-hasin VOLPFIPIED P.R. WDS TDALELKEDV KLLMIGCQI IGPFFASVW LQGCKTVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-bacst IELPNPKFS. N.R. LDS TGALNLGVW SLKLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-bacsu IELPNPKFS. E.R. LDS TGALNLKEP KLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-stau IELPNPKFG. K.R. IDS TGALNLQEVW SLKLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-halvo IQPIGDFDQ E. PV. WSS RDALEADTVFVW SLKLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dld1-psepu VELPMLPGL. P.G. ISS TEALPAKLP QHLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL	440
	441	NHPLPLPGPDQ GRKGDIILDS TGLSVPKIE ETLVVII GCVV IGPFFSLPA SLGCKTVLWQ GLPTILE. ML DKDIDAMTK ELAKQGMF. VF LKGSVQTVQADT ASADG. VSL dldh-mycca VKLPLPGPDQ ED. P.R. TDS TGALPQVWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-alceu P.FPEI EG. N. KLSGVIDS TGALSLESNP ESIAIIGVW IGPFFASIN SLGCKTVIIE MLPHILP. PM DREISEIKA LKIRDGI. NI NNNCKVTRIE QGEDGLKVSF acol-cloma RLDSPHBLA RD. KFD. MNG IEL. PAPLNK LAKMRKKEC VSDP..... LPAFLLKQVW LDFKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-achla RLDSPHBLA RD. KFD. MNG IEL. PAPLNK LAKMRKKEC VSDP..... LPAFLLKQVW LDFKNCVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-pea RALPGLADPQ E. HIWYTY..... ALRPLKIP LKSLLNGCQK IGPFFASIN DLGCKTVLWQ LQASLQVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-trryb TALPFLPDPD K. VV..... TSS TGALQALQWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-pig TPFPGITIDE D. TV..... VSS TGALSLKKWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-human TPFPGITIDE D. TI..... VSS TGALSLKKWV NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-yeast TPFPGIEVDE E. KI..... VSS TGALSLKEDR KLTIIIGCQI IGPFFASVW LQGCKTVW NWGCPKTL NWGCPKTL NWGCPKTL dld3-psepu TPLPGVTDID Q. RI..... VSS TGALSLPQVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-psf1 VEIPPLPQDS D. VI..... VSS TGALLEQPAVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-azovi VEIPPLAPVQD D. VI..... VSS TGALDFQNVN KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dld2-psepu IDIPPAVQDQ D. VI..... VSS TGALDFQPAVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL acol-pelca AVQPGVITDID Q. VI..... VSS TGALDFQPAVW KHLVQKNCVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-ecoli IQLPFIPIED P.R. WDS TDALELKEDV ERLLVWNGCQI IGPFFASVW LQGCKTVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-hasin VOLPFIPIED P.R. WDS TDALELKEDV KLLMIGCQI IGPFFASVW LQGCKTVW NWGCPKTL NWGCPKTL NWGCPKTL dldh-bacst IELPNPKFS. N.R. LDS TGALNLGVW SLKLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-bacsu IELPNPKFS. E.R. LDS TGALNLKEP KLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-stau IELPNPKFG. K.R. IDS TGALNLQEVW SLKLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dldh-halvo IQPIGDFDQ E. PV. WSS RDALEADTVFVW SLKLLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL dld1-psepu VELPMLPGL. P.G. ISS TEALPAKLP QHLVQVW NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL NWGCPKTL	550
	551DGQD QMIKGEYVNE SVRKTSLTG F. ENIGLEL TPKRGVVNE YQETNLDGVY AIGD.VVGKS M..... TAVTKGA IVAANRIAK ANKAHAEDV MNYDKVPSI dldh-myccaEGEAAA EPQRDVLVWV SVRSPNGK..... ISAEGAKVAV SERGFINVDA QMRNTVPHF AIGD.VGP M..... LAKAVHEA HVAEEAAHG. EKAY.... FDAOQIAPSVA dldh-alceu I....DGKGA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC acol-clomaDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-achlaDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-peaDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-trrybDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-pigDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-humanDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-yeastDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dld3-psepuDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-psf1DGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-azoviDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dld2-psepuDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC acol-pelcaDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-ecoliDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-hasinDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-bacstDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-bacsuDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-stauDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dldh-halvoDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC dld1-psepuDGGKA ESDIVEKVLV AVGRSNTIE LDVERKIVQT E....GGSIVIEND KMETVANRQVW AIGD.CTQK VLA.....VNSDVSQ VAAEANIMG. ONNK.... MDYKTCVAC	660
	552	YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-mycca YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-alceu YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET acol-cloma YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-achla YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-pea YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-trryb YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-pig YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-human YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dldh-yeast YTHPEVMSIC KTEEQQLKQIN LEYKAKFPPK SAIGKALADD DTGCKTVLWQ CIIASPTAII HIIGRNQVW ISEITAVIEC EGTEITEIANT IHPHPTMSA IGEAAEALET dld3-psepu YTHPELPAVWQ KTEEQQLKQIN RAYKVKFPF TANSRKINH ETGCFAKVIA DADTEVFLWV HLGVPSVEM IGECPVAMEF SASAEDIALT CHPHPTMSA LRAQAMNDV dldh-psf1 YTHPELPAVWQ KTEEQQLKQIN RAYKVKFPF TANSRKINH ETGCFAKVIA DADTEVFLWV HLGVPSVEM IGECPVAMEF SASAEDIALT CHPHPTMSA LRAQAMNDV dldh-azovi YTHPELPAVWQ KTEEQQLKQIN RAYKVKFPF TANSRKINH ETGCFAKVIA DADTEVFLWV HLGVPSVEM IGECPVAMEF SASAEDIALT CHPHPTMSA LRAQAMNDV dld2-psepu YTHPELPAVWQ KTEEQQLKQIN RAYKVKFPF TANSRKINH ETGCFAKVIA DADTEVFLWV HLGVPSVEM IGECPVAMEF SASAEDIALT CHPHPTMSA LRAQAMNDV acol-pelca YTHPEVAVWQ KTEEQQLKQIN RAYKVKFPF TANSRKINH ETGCFAKVIA DADTEVFLWV HLGVPSVEM IGECPVAMEF SASAEDIALT CHPHPTMSA LRAQAMNDV dldh-ecoli YTHPEVAVWQ KTEEQQLKQIN RAYKVKFPF TANSRKINH ETGCFAKVIA DADTEVFLWV HLGVPSVEM IGECPVAMEF SASAEDIALT CHPHPTMSA LRAQAMNDV dldh-hasin YTHPEVAVWQ KTEEQQLKQIN RAYKVKFPF TANSRKINH ETGCFAKVIA DADTEVFLWV HLGVPSVEM IGECPVAMEF SASAEDIALT CHPHPTMSA LRAQAMNDV dldh-bacst FSPECAVSG YFEQQAKDQVW AIGRNLALT DTFGLKFLWV RKEQDVGIAA QIIGPNASDQ IAEGLAIEF GASAEDVARW CHAHPTLSEA FKEANMMAA.Y dldh-bacsu FSPECAVSG YFEQQAKDQVW AIGRNLALT DTFGLKFLWV RKEQDVGIAA QIIGPNASDQ IAEGLAIEF GASAEDVARW CHAHPTLSEA FKEANMMAA.Y dldh-stau FSPECAVSG YFEQQAKDQVW AIGRNLALT DTFGLKFLWV RKEQDVGIAA QIIGPNASDQ IAEGLAIEF GASAEDVARW CHAHPTLSEA FKEANMMAA.Y dldh-halvo FSPECAVSG YFEQQAKDQVW AIGRNLALT DTFGLKFLWV RKEQDVGIAA QIIGPNASDQ IAEGLAIEF GASAEDVARW CHAHPTLSEA FKEANMMAA.Y dld1-psepu FSPECAVSG YFEQQAKDQVW AIGRNLALT DTFGLKFLWV RKEQDVGIAA QIIGPNASDQ IAEGLAIEF GASAEDVARW CHAHPTLSEA FKEANMMAA.Y	660

Fig. 11A. (Figure continues on following page.)

	661	678
<i>dldh_mycce</i>	GKAIHP.....	
<i>dldh_alceu</i>	GTCCTDVPPR KR.....	
<i>acol_cloma</i>	NQAIHMNPK.....	
<i>dldh_achla</i>		
<i>dldh_mycke</i>	DVPS.....	
<i>acol_klepn</i>	DQPLHQ.....	
<i>dldh_pea</i>	DKPIHI.....	
<i>dldh_trybb</i>	AKTINF.....	
<i>dldh_pig</i>	GKAIFN.....	
<i>dldh_human</i>	GKSINF.....	
<i>dldh_yeast</i>	DKAIIC.....	
<i>dld3_psepu</i>	GMAMOI.....	
<i>dldh_psef1</i>	GHAIIHANRK KR.....	
<i>dldh_azovi</i>	GHAIIHVANRK K.....	
<i>dld2_psepu</i>	GGAIIHVANRK KR.....	
<i>acol_pelca</i>	GAAVHC.....	
<i>dldh_ecoli</i>	GSITDLPNPK AKKK.....	
<i>dldh_haein</i>	GSITDLPNAK AKEKIISI	
<i>dldh_bacst</i>	GTPIIHITK.....	
<i>dldh_bacsu</i>	GSPIIHVVK.....	
<i>dldh_staa</i>	GYPIHTM.....	
<i>dldh_halvo</i>	GQAIIHLNR.....	
<i>dld1_psepu</i>	GHALHI.....	

Fig. 11. Alignment of sequences of members of the dihydrolipoamide dehydrogenase family. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in the listed Enzyme II proteins are shown in reverse shading. Boxed shaded region corresponds to a lipoyl domain. Abbreviations used and references to published sequences are: *M. capricolum* (*dldh-mycce*) (this work); *A. eutrophus* (*dldh-alceu*) (Hein & Steinbuchel, 1994); *C. magnum* (*acol-cloma*) (Kruger et al., 1994); *A. laidlawii* (*dldh-achla*) (Wallbrandt et al., 1992); *M. genitalium* (*dldh-mycke*) (MG271 from the TIGR database); *K. pneumoniae* (*acol-klepn*) (Genbank accession no. U30887); pea (*dldh-pea*) (Bourguignon et al., 1992); *Trypanosoma brucei* (*dldh-trybb*) (Else et al., 1993); pig (*dldh-pig*) (Otulakowski & Robinson, 1987); human (*dldh-human*) (Pons et al., 1988); yeast (*dldh-yeast*) (Browning et al., 1988); *P. putida* (*dld2-psepu*) (Palmer et al., 1991a), (*dld1-psepu*) (Burns et al., 1989), and (*dld3-psepu*) (Palmer et al., 1991b); *Pseudomonas fluorescens* (*dldh-psef1*) (Benen et al., 1989); *A. vinelandii* (*dldh-azovi*) (Westphal & de Kok, 1988); *P. carbinolicus* (*acol-pelca*) (Oppermann & Steinbuchel, 1994); *E. coli* (*dldh-ecoli*) (Guest, 1987); *H. influenzae* (*dldh-haein*) (HI1231 from the TIGR database); *B. stearothermophilus* (*dldh-bacst*) (Borges et al., 1990); *B. subtilis* (*dldh-bacsu*) (Hemila et al., 1990); *S. aureus* (*dldh-staa*) (Hemila, 1991); *Halofexax volcanii*, (*dldh-halvo*) (Vettakkorumakankav & Stevenson, 1992). *dldh*, *dld1*, and *dld2* refer to the genes encoding dihydrolipoamide dehydrogenases from the pyruvate, 2-oxoglutarate, and branched chain α -ketoadic dehydrogenase complexes, respectively. *dld3* refers to the third dehydrogenase isolated from *P. putida* and *acol* refers to the dihydrolipoamide dehydrogenase of the acetoin dehydrogenase complex.

Mycoplasmas are generally believed to be descendants of Gram-positive bacteria. All the *M. capricolum* proteins involved with the metabolism of pyruvate described here show phylogenetic relatedness to the homologous proteins from Gram-positive bacteria and *M. genitalium*.

The complete genome of *M. genitalium* has been reported recently (Fraser et al., 1995). Because it might be expected that these two species are closely related, the question was posed concerning the genomic locations of the genes encoding enzymes of pyruvate metabolism. Figure 18 shows a comparison of the location of the genes of interest in *M. capricolum* and *M. genitalium*.

genitalium. It is most surprising to see that, whereas all the genes involved with pyruvate metabolism are clustered in *M. capricolum*, this is not the case in *M. genitalium*. In both organisms, the *ptsI* genes are located approximately 500 kb from the replication origin. Further, the *gyrA,B* complex is separated from the *naox*, *odpA*, *odpB*, *odp2*, *dldH*, *lplA* complex by approximately 300 kb in both organisms. Clearly, there have been extensive rearrangements in the genomes of these organisms during evolution, resulting in resolution of *gyrA,B* from the replication origin in the case of *M. capricolum* and scattering of genes involved with pyruvate metabolism in *M. genitalium*.

In summary, the present work has demonstrated the unique arrangement of the genes encoding enzymes involved with pyruvate metabolism in *M. capricolum* and described some unique properties of the products of these genes.

Materials and methods

Growth of cells

M. capricolum (kid strain) were grown in modified Edwards medium at pH 8, as described previously (Mugharbil & Cirillo, 1978). After harvesting, cells were stored as frozen pellets for future use.

Nucleic acid preparations

DNA was prepared from frozen cells as previously described (Ausubel et al., 1990). RNA was prepared by suspension of cells (100 mg) in buffer (1 mL) containing 50 mM Tris-HCl, pH 6.8, 2 mM EDTA, and 1% SDS; the suspension was mixed with 5 mL of 4 M guanidium thiocyanate homogenization buffer

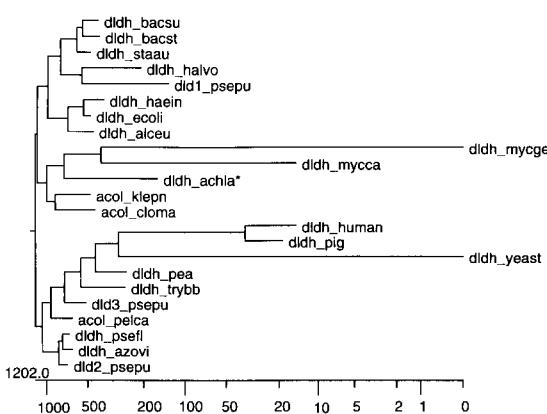


Fig. 12. Phylogenetic tree of sequenced proteins of the dihydrolipoamide dehydrogenase family. Relative evolutionary distances are shown on the numerical scale. Abbreviations are as in the legend to Figure 11. *, corresponds to a partial amino acid sequence (see Fig. 11).

pta-haein	1	MSRTIILIPV STGVGLTSIS LGLIHSLEQK GTKVAFMKPV SQPSTGDEKL DRTTSIIRTS TSLET.AEFP MLSVAESLIG QNQSDVLLKEK IVANHQQLTK NNDIVVLLK
pta-ecoli		SRIIMLIPT GTSGVGLTSV LGVIRAMERK GVRLSVFKPI AQPFRTGGDAP DQTTTIVRAN SSSTTAAEPL KMSYVEGLS SNQKDVLMLEE IVANHYHANTK DAEVVLVEGL
pta-haein	111	IPTRKHGYAN SINYEIAQAL DAEIVLVAAP ATETPTELKD RVEAAAALFG GKNNPNLGV VVNKFNAVPD ESGRTRPDLA EIFFDSFQHNH ISETEVNKLGF AGSAKILLC
pta-ecoli		VPTRKHQPAQ SLNEYIAKTL NAEIVFVMSQ GTDTPOLKE RIELTNNSPG GAKNTNTIVG IVNKLNAVPD EGQRTRPDLS EIFFDSSKAK VNNVDPANVQ ESSPLVGLA
pta-haein	220	IPWANANLIAT RAIDLVKHLD ASIINEGEIN .RRIRGITFC ARSLPNMVEH FRAGSLLVAS ADRPDVLVA ALAASNGIEI GGILLTGGYK IDAQINKLKR PTFEKAALP
pta-ecoli		PWPSFSDLIAT RAIDMARHLN ATIINEGDI TRRVKSFTPC ARSIPHMLEH FRAGSLLVTS ADRPDVLVA CLAAMNGVEI GALLLTGGYE MDARISKLC RAFATG. LPV
pta-haein	221	IPWANANLIAT RAIDLVKHLD ASIINEGEIN .RRIRGITFC ARSLPNMVEH FRAGSLLVAS ADRPDVLVA ALAASNGIEI GGILLTGGYK IDAQINKLKR PTFEKAALP
pta-ecoli		PWPSFSDLIAT RAIDMARHLN ATIINEGDI TRRVKSFTPC ARSIPHMLEH FRAGSLLVTS ADRPDVLVA CLAAMNGVEI GALLLTGGYE MDARISKLC RAFATG. LPV
pta-myccs	331	PRIEGNTIWQT ALSLQSFSNLE VPVDDHERIE NIQKYSIQHF NADFINNLVA DSSRLRPLSP PAFRQLTLEL ARAAKRKLVL PEGDEPRTK AVAHLCAERGI A. ECVLLADP
pta-myccs		FMTWNTNIWQT SLSLSFSNLE VPVDDHERIE KVQEYVANYI NADWIESLTA TSSRSRLSP PAFAQLTEL ARAKGAKRKLVL PEGDEPRTVK AAAICAERGI A. TCVLLGNP
pta-myccs	441	KEVPESEIKNN SSIIKTCILDE FD TK EFEEEFVKKI KGKA_TIEVA HQVQLPNYI GAMLVKLNQA DCLM_EGLNNT TADP_RPALD . IIGTCPGNYI ACSIYRMS..
pta-myccs		QXIFPANF . D KKHITYVIDE MD LT SYANFYVER KHKGDLKBA OKFVRDPSSL AATLVALKV DGEVCGKBEYA TADP_RPALD . LLATG INF VESVIME..
pta-parde	EI	PGAGR . D KKHITYVIDE MD LT SYANFYVER KHKGDLKBA OKFVRDPSSL AATLVALKV DGEVCGKBEYA TADP_RPALD . LLATG INF VESVIME..
pta-mette	ADIKALAGDL D. LS	KAKI PDVKTYEKED EYINAFLP KHKGDLKBA OKFVRDPSSL AATLVALKV DGEVCGKBEYA TADP_RPALD . LLATG INF VESVIME..
pta-bacsu	NEIQJAKAKEL NLTLCG	GKVIY DPDKTYEKGK DLVQDVAFVER KGKA_TIECA RKAALDNTYD GTMLVYKGLA DGLVGECAAHSS SDFT_RPAVO . IVKTAKGAAL ASAFAHISVP
pta-haein	AVSRVAQAE GVKLGGKGTI INPA	DVRE NYDVRDLEP KAKGMTEATA REQLEDITVU GLTMALNEAV DGLVGECAHHS SDFT_RPAVO . IVKTAKGAAL ASAFAHISVP
pta-ecoli	AETINRVAASQ GVELGAGIEI VPDE	SYVGRIVLVEF KNKGMTETVA REQLEDNVU GLTMLEQDEV DGLVGECAHHS SDFT_RPAVO . IVKTAKGAAL ASAFAHISVP
pta-myccs	551 KGNENY IPTDCALNIK PETSQOLVENT OMAVDFAKAL NVKNVEAALL SYSTNTCSKG EDVDRVHQAV EILLSKSKEDT VCEGEID . FDDAPDKKTB DNFKNPKV..
pta-myccs	 KGEERL YFTVDCAPAVB ENSOELAHT ENTPNPFKSL NEDEIKMPL SYSTNTCSKG EMVDVKVLAT KLFLEKHPBL HQSVC. GELO . DDAFVKEVK LOKAPQ. LTV
pta-parde	GPAPAVRGGM IFADCPGLVIO BDARELAABA LSAASCDRR LAAEPRPVLV SYSTNTCSKG PSLGRICREAL ALIRAAAPGL EVDGEMLO . DDALEADIR ARKAPAESL	
pta-mette	DCEVGSDGTF LDPLGDSMVM BSEVEDVANA VISAKTASL VQDVDPKVALM SYSTNTCSKG KLTJEATIASL KLAQEALEPDI AIDGELO . DDAFVTKPV ASKAGPSPV	
pta-bacsu RGEEOV VFAEDCAINIA BESDODLAESA IESENTAKMF DIE_PVAML SYSTNTCSKG DTEKEVADAV KIAKEKAPEL TLDGEFO . DDAFVPPVSA ESKGADPSEI	
pta-haein DQVL VYGDCAVNPD ETABQELAESA IQSADSAAK GFID_PVAMN SYSTNTCSKG ADVEKVKREAT RIAKEKRPD LIDGELO . YDAAMMEDVA RSKAPNSPV.	
pta-ecoli EQVY VYGDCAVNPD ETABQELAESA IQSADSAAAF GIE_PVAML SYSTNTCSKG SDVEKVKREAT RLQAEKRPD MIDGELO . YDAAMMDVA KSHAPNSPV.	
pta-myccs	661	KQTDPIDUFF DINAGNIGWY LAQPMGGFEA IG_FULGLNO FUNDLRSRGAT VSFVLTATAIM FLYLSY.
pta-myccs		KNSANVYUFP NDLGAGNIAWY IAQRLGGDYA IG_FULGLSS FUNDLRSRGAT VSFVLTATAIM FLYLSY.
pta-parde	TGRNPVNUFP DLAQIGRIGYV IAQRLGLTVA IG_FULGLAK FUNDLRSRACS VSFVLTATAIM FLYLSY.	
pta-mette	AGKANVYUFP LAQNLAKAAYA YGPFITVGLAK FUNDLRSRGS DEPIFIVGAVV FTVQAAAQDK	
pta-bacsu	KGDANVYUFP SLEAGIGWY IAQRLGNFPEA FGFLGLGMM FUNDLRSRGN AEVYNLALI TAQALQ.	
pta-haein	AGKATVYUFP DILQGTYWY AVQRSADLVS IGLQALCMIK FUNDLRSRGAL VSFVLTATAIM TAQOQATQ.	
pta-ecoli	AGKATVYUFP DILQGTYWY AVQRSADLVS IGLQALCMIK FUNDLRSRGAL VSFVLTATAIM TAQOQATQ.	

Fig. 13. Alignment of sequences of members of the phosphotransacetylase family. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in the listed phosphotransacetylase proteins are shown in reverse shading. Abbreviations used and references to published sequences are: *M. capricolum* (pta-mycca) (this work); *M. genitalium* (pta-mycge) (MG299 from the TIGR database); *Paracoccus denitrificans* (pta-parde) (Van Spanning et al., 1995); *M. thermophila* (pta-mette) (Latimer & Ferry, 1993); *B. subtilis* (pta-bacsu) (Glaser et al., 1993); *H. influenzae* (pta-haein) (H11203 from the TIGR database); *E. coli* (pta-ecoli) (Kakuda et al., 1994).

(Sambrook et al., 1989). The cell suspension was frozen in dry ice-ethanol, then warmed briefly in a 64 °C water bath (Salser et al., 1967). Total RNA was purified by ultracentrifugation through 5.7 M CsCl/10 mM EDTA (Sambrook et al., 1989).

Cloning and screening

Genomic DNA fragments produced by digestion of *M. capricolum* DNA with *Hind* III (9,542-bp fragment), *Xba* I (6,654-bp

fragment), or *Spe* I (4,384-bp fragment) were cloned into the vector pBluescript II KS⁺ (pKSII⁺). Recombinant plasmids were used to transform Epicurean coli XLI-Blue Competent cells (Stratagene). Colonies were lifted onto nylon membranes (NEN Research Products, NEF-978). [³²P]5'-end-labeled oligonucleotide probes (1×10^6 cpm/mL of hybridization solution) were used for selecting positive clones. Oligonucleotide probes, synthesized as tritlyl-off derivatives on an Applied Biosystems 380B DNA synthesizer, were labeled with [γ ³²P]ATP by the DNA 5'-end-labeling method (Sambrook et al., 1989). Prehybridization was performed at 40 °C for 4 h in 6× SSPE/0.1% SDS/10× Denhardt's solution containing 20 mg/mL tRNA and 50 mg/mL of denatured heterologous DNA. Hybridization was performed at 42 °C for 16 h in 6× SSPE/10% SDS solution containing 1.4×10^6 cpm of [³²P]-labeled oligonucleotide/mL. The membrane was finally washed in 0.5× SSPE/1% SDS solution at 40 °C for 30 min. Positive clones were detected by autoradiography.

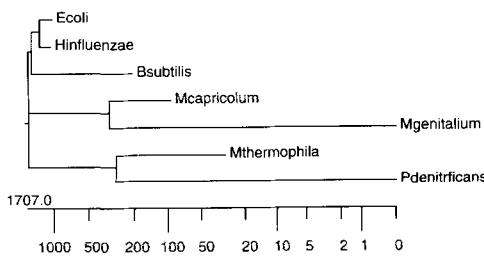


Fig. 14. Phylogenetic tree of sequenced proteins of the phosphotransacetylase family. Relative evolutionary distances are shown on the numerical scale. Abbreviations are as in the legend to Figure 13.

DNA sequencing

DNA sequencing on both strands of the DNA was performed by the dideoxy chain termination method of Sanger et al. (1977), with [α -³⁵S]dATP, using Sequenase 2.0 (United States Biochemicals) DNA sequencing kits. M13 forward or reverse primers or

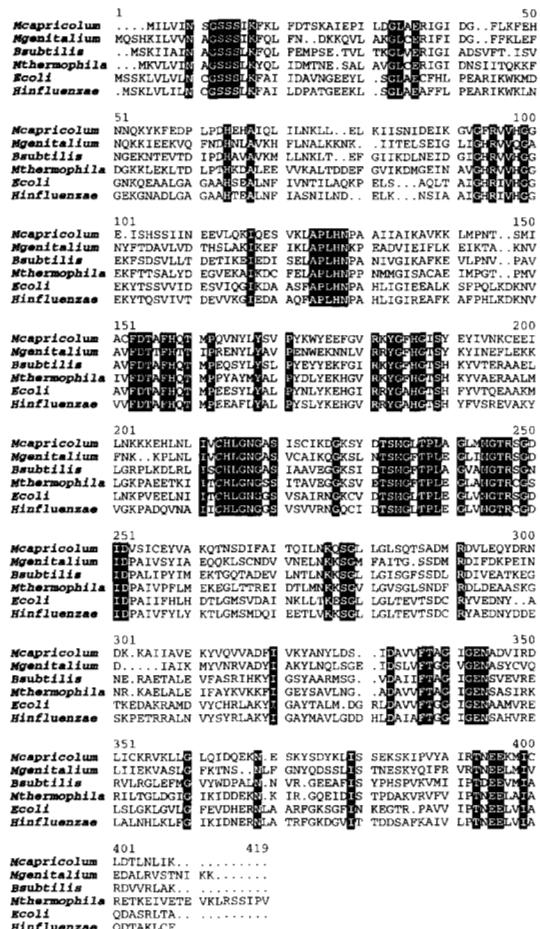


Fig. 15. Alignment of sequences of members of the acetate kinase family. Numbering above the aligned sequences corresponds only to the residue in the alignment rather than to a residue number in any of the aligned proteins. Residues that are conserved in the listed phosphotransacetylase proteins are shown in reverse shading. Abbreviations used and references to published sequences are: *M. capricolum* (*Mcapricolum*) (this work); *M. genitalium* (*Mgenitalium*) (MG357 from the TIGR database); *B. subtilis* (*Bsubtilis*) (Grundy et al., 1993); *M. thermophila* (*Mthermophila*) (Latimer & Ferry, 1993); *E. coli* (*Matsuyama et al., 1989*); *H. influenzae* (*Hinfluenzae*) (HI1204 from the TIGR database).

specific primers complementary to previously determined sequences were used.

Computer analyses

Analyses of DNA and protein sequence were performed using the GCG programs, version 7.2 (Devereux et al., 1984). Isoelectric points were calculated using the PEPTIDESORT program. A search for transcription termination sites used the TERMINATOR program. Stem-loop structures were analyzed using FOLDRNA. Translation frames were detected using the MAP program. Phylogenetic trees were constructed using the MEGALIGN module of the LaserGene program (DNAStar, Madison, Wisconsin) by the method of Hein (1990).

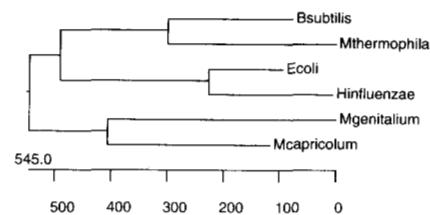


Fig. 16. Phylogenetic tree of sequenced proteins of the acetate kinase family. Relative evolutionary distances are shown on the numerical scale. Abbreviations are as in the legend to Figure 15.

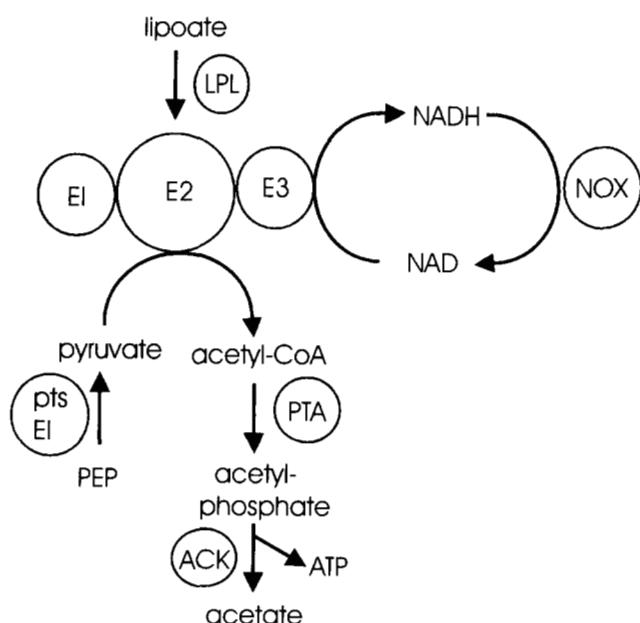


Fig. 17. Metabolic scheme for conversion of phosphoenolpyruvate to acetate in *M. capricolum*.

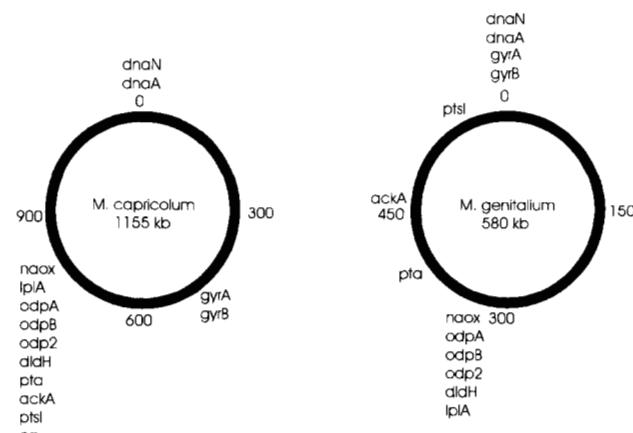


Fig. 18. Comparison of positions on the genetic map of various genes in *M. capricolum* and *M. genitalium*.

References

- Allen AG, Perham RN. 1991. Two lipoyl domains in the dihydrolipoamide acetyltransferase chain of the pyruvate dehydrogenase multienzyme complex of *Streptococcus faecalis*. *FEBS Lett* 287:206-210.
- Ausubel FA, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K, eds. 1990. *Current protocols in molecular biology*. New York: Greene Publishing and Wiley-Interscience. pp 2.4.1-2.4.5.
- Behal RH, Browning KS, Reed LJ. 1989. Nucleotide and deduced amino acid sequence of the alpha subunit of yeast pyruvate dehydrogenase. *Biochem Biophys Res Commun* 164:941-946.
- Benen JAE, van Berkel WJ, van Dongen WMAM, Mueller F, de Kok A. 1989. Molecular cloning and sequence determination of the lpd gene encoding lipoamide dehydrogenase from *Pseudomonas fluorescens*. *J Gen Microbiol* 135:1787-1797.
- Borges A, Hawkins CF, Packman LC, Perham RN. 1990. Cloning and sequence analysis of the genes encoding the dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase components of the pyruvate dehydrogenase multienzyme complex of *Bacillus stearothermophilus*. *Eur J Biochem* 194:95-102.
- Bourguignon J, Macherel D, Neuberger M, Douce R. 1992. Isolation, characterization, and sequence analysis of a cDNA clone encoding 1-protein, the dihydrolipoamide dehydrogenase component of the glycine cleavage system from pea-leaf mitochondria. *Eur J Biochem* 204:865-873.
- Browning KS, Uhlinger DJ, Reed LJ. 1988. Nucleotide sequence for yeast dihydrolipoamide dehydrogenase. *Proc Natl Acad Sci USA* 85:1831-1834.
- Burns G, Brown T, Hatter K, Sokatch JR. 1988. Comparison of the amino acid sequences of the transacylase components of branched chain oxoacid dehydrogenase of *Pseudomonas putida*, and the pyruvate and 2-oxoglutarate dehydrogenases of *Escherichia coli*. *Eur J Biochem* 176:165-169.
- Burns G, Brown T, Hatter K, Sokatch JR. 1989. Sequence analysis of the lpdV gene for lipoamide dehydrogenase of branched-chain-oxoacid dehydrogenase of *Pseudomonas putida*. *Eur J Biochem* 179:61-69.
- Carlsson P, Hederstedt L. 1989. Genetic characterization of *Bacillus subtilis* odhA and odhB, encoding 2-oxoglutarate dehydrogenase and dihydrolipoamide transsuccinylase, respectively. *J Bacteriol* 171:3667-3672.
- Cirillo VP. 1979. Transport systems. In: Barile MF, Razin S, eds. *The mycoplasmas*. New York: Academic Press. pp 323-350.
- Cullingford TE, Clark JB, Phillips IR. 1993. Characterization of cDNAs encoding the rat testis-specific E1 alpha subunit of the pyruvate dehydrogenase complex: Comparison of expression of the corresponding mRNA with that of the somatic E1 alpha subunit. *Biochim Biophys Acta* 1216:149-153.
- Cullingford TE, Clark JB, Phillips IR. 1994. The pyruvate dehydrogenase complex: Cloning of the rat somatic E1 alpha subunit and its coordinate expression with the mRNAs for the E1 beta, E2, and E3 catalytic subunits in developing rat brain. *J Neurochem* 62:1682-1690.
- Dahl HH, Hunt SM, Hutchison WM, Brown GK. 1987. The human pyruvate dehydrogenase complex. Isolation of cDNA clones for the E1 alpha subunit, sequence analysis, and characterization of the mRNA. *J Biol Chem* 262:7398-7403.
- de la Sierra IL, Prange T, Fourme R, Padron G, Fuentes P, Musacchio A, Madrazo J. 1994. Crystallization and preliminary X-ray investigation of a recombinant outer membrane protein from *Neisseria meningitidis*. *J Mol Biol* 235:1154-1155.
- Deng WL, Chang HY, Peng HL. 1994. Acetoin catabolic system of *Klebsiella pneumoniae* CG43: Sequence, expression, and organization of the aco operon. *J Bacteriol* 176:3527-3535.
- Devereux J, Haeblerli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387-395.
- Else AJ, Hough DW, Danson MJ. 1993. Cloning, sequencing, and expression of *Trypanosoma brucei* dihydrolipoamide dehydrogenase. *Eur J Biochem* 212:423-429.
- Fitzgerald J, Hutchison WM, Dahl HH. 1992. Isolation and characterization of the mouse pyruvate dehydrogenase E1 alpha genes. *Biochim Biophys Acta* 1131:83-90.
- Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, et al. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270:397-403.
- Gershwin ME, MacKay IR, Sturgess A, Coppel RL. 1987. Identification and specificity of a cDNA encoding the 70 kb mitochondrial antigen recognized in primary biliary cirrhosis. *J Immunol* 138:3525-3531.
- Glaser P, Kunst F, Arnaud M, Coudart MP, Gonzales W, Hullo MF, Ionescu M, Lubochinsky B, Marcelino L, Moszer I, Presecan E, Santana M, Schneider E, Schweizer J, Vertes A, Rapoport G, Danchin A. 1993. *Bacillus subtilis* genome project: Cloning and sequencing of the 97 kb region from 325 degrees to 333 degrees. *Mol Microbiol* 10:371-384.
- Grundy FJ, Waters DA, Allen SH, Henkin TM. 1993. Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA. *J Bacteriol* 175:7348-7355.
- Guan Y, Rawsthorne S, Scofield G, Shaw P, Doonan J. 1995. Cloning and characterization of a dihydrolipoamide acetyltransferase (E2) subunit of the pyruvate dehydrogenase complex from *Arabidopsis thaliana*. *J Biol Chem* 270:5412-5417.
- Guest JR. 1987. Functional implications of structural homologies between chloramphenicol acetyltransferase and dihydrolipoamide acetyltransferase. *FEMS Microbiol Lett* 44:417-422.
- Hanemaaijer R, Janssen A, de Kok A, Veeger C. 1988. The dihydrolipoyl-transacetylase component of the pyruvate dehydrogenase complex from *Azotobacter vinelandii*. Molecular cloning and sequence analysis. *Eur J Biochem* 174:593-599.
- Hein J. 1990. Unified approach to alignment and phylogenies. *Methods Enzymol* 183:626-644.
- Hein S, Steinbuchel A. 1994. Biochemical and molecular characterization of the *Alcaligenes eutrophus* pyruvate dehydrogenase complex and identification of a new type of dihydrolipoamide dehydrogenase. *J Bacteriol* 176:4394-4408.
- Hemila H, Palva A, Paulin L, Arvidson S, Palva I. 1990. Secretory S complex of *Bacillus subtilis*: Sequence analysis and identity to pyruvate dehydrogenase. *J Bacteriol* 172:5052-5063.
- Hemila H. 1991. Lipoamide dehydrogenase of *Staphylococcus aureus*: Nucleotide sequence and sequence analysis. *Biochim Biophys Acta* 1129:119-123.
- Ho L, Patel MS. 1990. Cloning and cDNA sequence of the beta-subunit component of human pyruvate dehydrogenase complex. *Gene* 86:297-302.
- Ho L, Wexler ID, Liu TC, Thekkumkara TJ, Patel MS. 1989. Characterization of cDNAs encoding human pyruvate dehydrogenase alpha-subunit. *Proc Natl Acad Sci USA* 86:5330-5334.
- Hu CWC, Lau KS, Griffin TA, Chuang JL, Fisher CW, Cox RP, Chuang DT. 1988. Isolation and sequencing of a cDNA encoding the decarboxylase (E1)-alpha precursor of bovine branched-chain alpha-keto acid dehydrogenase complex: Expression of E1-alpha mRNA and subunit in maple-syrup-urine-disease and 3T3-L1 cells. *J Biol Chem* 263:9007-9014.
- Johnson RK, Komuniecki R, Sun Y, Wheelock MJ. 1992. Characterization of cDNA clones for the alpha subunit of pyruvate dehydrogenase from *Ascaris suum*. *Mol Biochem Parasitol* 51:37-48.
- Kakuda H, Hosono K, Shiroishi K, Ichihara S. 1994. Identification and characterization of the ackA (acetate kinase A)-pta (phosphotransacetylase) operon and complementation analysis of acetate utilization by an ackA-pta deletion mutant of *Escherichia coli*. *J Biochem* 116:916-922.
- Kreader CA, Langer CS, Heckman JE. 1989. A mitochondrial protein from *Neurospora crassa* detected both on ribosomes and in membrane fractions. Analysis of the gene, the message and the protein. *J Biol Chem* 264:317-327.
- Kruger N, Oppermann FB, Lorenzl H, Steinbuchel A. 1994. Biochemical and molecular characterization of the *Clostridium magnum* acetoin dehydrogenase enzyme system. *J Bacteriol* 176:3614-3630.
- Latimer MT, Ferry JG. 1993. Cloning, sequence analysis, and hyperexpression of the genes encoding phosphotransacetylase and acetate kinase from *Methanosarcina thermophila*. *J Bacteriol* 175:6822-6829.
- Lau KS, Chuang JL, Herring WJ, Danner DJ, Cox RP, Chuang DT. 1992. The complete cDNA sequence for dihydrolipoyl transacetylase (E2) of human branched-chain alpha-keto acid dehydrogenase complex. *Biochim Biophys Acta* 1132:319-321.
- Lau KS, Griffin TA, Hu CWC, Chuang DT. 1988. Conservation of primary structure in the lipoyl-bearing and dihydrolipoyl dehydrogenase binding domains of mammalian branched-chain alpha-keto acid dehydrogenase complex: Molecular cloning of human and bovine transacetylase (E2) cDNAs. *Biochemistry* 27:1972-1981.
- Luethy MH, Miernyk JA, Randall DD. 1994. The nucleotide and deduced amino acid sequences of a cDNA encoding the E1 beta-subunit of the *Arabidopsis thaliana* mitochondrial pyruvate dehydrogenase complex. *Biochim Biophys Acta* 1187:95-98.
- Luethy MH, Miernyk JA, Randall DD. 1995. The mitochondrial pyruvate dehydrogenase complex: Nucleotide and deduced amino-acid sequences of a cDNA encoding the *Arabidopsis thaliana* E1 alpha-subunit. *Gene* 164:251-254.
- Maniloff J. 1983. Evolution of wall-less prokaryotes. *Annu Rev Microbiol* 37:477-499.
- Matsuyama A, Yamamoto H, Nakano E. 1989. Cloning, expression, and nucleotide sequence of the *Escherichia coli* K-12 ackA gene. *J Bacteriol* 171:577-580.
- Mattevi A, deKok A, Perham RN. 1992. The pyruvate dehydrogenase multi-enzyme complex. *Curr Opin Struct Biol* 2:877-887.
- McKean MC, Winkeler KA, Danner DJ. 1992. Nucleotide sequence of the 5' end including the initiation codon of cDNA for the E1 alpha subunit

- of the human branched chain alpha-ketoacid dehydrogenase complex. *Biochim Biophys Acta* 1171:109–112.
- Miles RJ. 1992. Catabolism in mollicutes. *J Gen Microbiol* 138:1773–1783.
- Miran SG, Lawson JE, Reed LJ. 1993. Characterization of PDH beta 1, the structural gene for the pyruvate dehydrogenase beta subunit from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 90:1252–1256.
- Miyata M, Wang L, Fukumura T. 1991. Physical mapping of the *Mycoplasma capricolum* genome. *FEMS Microbiol Lett* 79:329.
- Morris TW, Reed KE, Cronan JE Jr. 1994. Identification of the gene encoding lipoate-protein ligase A of *Escherichia coli*. Molecular cloning and characterization of the *lplA* gene and gene product. *J Biol Chem* 269:16091–16100.
- Mughal U, Cirillo VP. 1978. *Mycoplasma* phosphoenolpyruvate-dependent sugar phosphotransferase system: Glucose-negative mutant and regulation of intracellular cyclic AMP. *J Bacteriol* 133:203–209.
- Nakano K, Matsuda S, Sakamoto T, Takase C, Nakagawa S, Ohta S, Ariyama T, Inazawa J, Abe T, Miyata T. 1993. Human dihydrolipoamide succinyltransferase: cDNA cloning and localization on chromosome 14q24.2-q24.3. *Biochim Biophys Acta* 1216:360–368.
- Nakano K, Matuda S, Yamanaka T, Tubouchi H, Nakagawa S, Titani K, Ohta S, Miyata T. 1991. Purification and molecular cloning of succinyltransferase of rat alpha-ketoglutarate dehydrogenase complex. Absence of a sequence motif of the putative E3 and/or E1 binding site. *J Biol Chem* 266:19013–19017.
- Niu XD, Browning KS, Behal RH, Reed LJ. 1988. Cloning and nucleotide sequence of the gene for dihydrolipoamide acetyltransferase from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 85:7546–7550.
- Nobukuni Y, Mitsubuchi H, Endo F, Akaboshi I, Asaka J, Matsuda I. 1990a. Maple syrup urine disease. Complete primary structure of the E1 beta subunit of human branched chain alpha-ketoacid dehydrogenase complex deduced from the nucleotide sequence and a gene analysis of patients with this disease. *J Clinical Investigation* 86:242–247.
- Nobukuni Y, Mitsubuchi H, Endo F, Asaka Y, Oyama R, Titani K, Matsuda I. 1990b. Isolation and characterization of a complementary cDNA clone coding for the E1-beta subunit of the bovine branched-chain alpha-ketoacid dehydrogenase complex: Complete amino acid sequence of the precursor protein and its proteolytic processing. *Biochemistry* 29:1154–1160.
- Oppermann FB, Steinbuchel A. 1994. Identification and molecular characterization of the *aco* genes encoding the *Pelobacter carbinolicus* acetoin dehydrogenase enzyme system. *J Bacteriol* 176:469–485.
- Otulakowski G, Robinson BH. 1987. Isolation and sequence determination of cDNA clones for porcine and human lipoamide dehydrogenase: Homology to other disulfide oxidoreductases. *J Biol Chem* 262:17313–17318.
- Palmer JA, Hatter K, Sokatch JR. 1991a. Cloning and sequence analysis of the LPD-glc structural gene of *Pseudomonas putida*. *J Bacteriol* 173:3109–3116.
- Palmer JA, Madhusudhan KT, Hatter K, Sokatch JR. 1991b. Cloning, sequence and transcriptional analysis of the structural gene for LPD-3, the third lipoamide dehydrogenase of *Pseudomonas putida*. *Eur J Biochem* 202:231–240.
- Pons G, Raefsky-Estrin C, Carothers DJ, Pepin RA, Javed AA, Jesse BW, Ganapathi MK, Samols D, Patel MS. 1988. Cloning and cDNA sequence of the dihydrolipoamide dehydrogenase component of human alpha-ketoacid dehydrogenase complexes. *Proc Natl Acad Sci USA* 85:1422–1426.
- Postma PW, Lengeler JW, Jacobson GR. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol Rev* 57:543–594.
- Priefert H, Hein S, Kreuger N, Zeh K, Schmidt B, Steinbuchel A. 1991. Identification and molecular characterization of the *Alcaligenes eutrophus* H16 *aco* operon genes involved in acetoin catabolism. *J Bacteriol* 173:4056–4071.
- Razin S. 1985. Molecular biology and genetics of *Mycoplasmas* (mollicutes). *Microbiol Rev* 49:419–455.
- Razin S. 1992. Peculiar properties of *Mycoplasmas*: The smallest self-replicating prokaryotes. *FEMS Microbiol Lett* 100:423–432.
- Repetto B, Tzagoloff A. 1990. Structure and regulation of KGD2, the structural gene for yeast dihydrolipoyl transsuccinylase. *MCB* 10:4221–4232.
- Ross RP, Claiborne A. 1992. Molecular cloning and analysis of the gene encoding NADH oxidase from *Streptococcus faecalis* 10C1. Comparison with NADH peroxidase and the flavoprotein disulfide reductases. *J Mol Biol* 227:658–671.
- Salser W, Gesteland RF, Bolle A. 1967. In vitro synthesis of bacteriophage lysozyme. *Nature* 215:588–591.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: A laboratory manual*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
- Sanger F, Nicklen S, Coulson AR. 1977. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467.
- Sermon K, DeMeirlier L, Elpers I, Lissens W, Liebaers I. 1990. Characterization of cDNA for porcine PDHE1-alpha and comparison with the human cDNA. *Nucleic Acids Res* 18:4925.
- Spencer ME, Darlison MG, Stephens PE, Duckenfield IK, Guest JR. 1984. Nucleotide sequence of the *sucB* gene encoding the dihydrolipoamide succinyltransferase of *Escherichia coli* K12 and homology with the corresponding acetyltransferase. *Eur J Biochem* 141:361–374.
- Thekkumkara TJ, Ho L, Wexler ID, Pons G, Liu TC, Patel MS. 1988. Nucleotide sequence of a cDNA for the dihydrolipoamide acetyltransferase component of human pyruvate dehydrogenase complex. *FEBS Lett* 240:45–48.
- Van Spanning RJM, De Boer APN, Slotboom DJ, Reijnders WNM, Stouthamer AH. 1995. Isolation and characterization of a novel insertion sequence element, IS1248, in *Paracoccus denitrificans*. *Plasmid* 34:11–21.
- Vettakkumakankav NN, Stevenson KJ. 1992. Dihydrolipoamide dehydrogenase from *Haloflexax volcanii*: Gene cloning, complete primary structure and comparison to other dihydrolipoamide dehydrogenases. *Biochemical Cellular Biol* 70:656–663.
- Wallbrandt PHM, Tegman V, Jonsson BH, Wieslander A. 1992. Identification and analysis of the genes coding for the putative pyruvate dehydrogenase enzyme complex in *Acholeplasma laidlawii*. *J Bacteriol* 174:1388–1396.
- Wang GF, Kuriki T, Roy KL, Kaneda T. 1993. The primary structure of branched-chain alpha-oxo acid dehydrogenase from *Bacillus subtilis* and its similarity to other alpha-oxo acid dehydrogenases. *Eur J Biochem* 213:1091–1099.
- Westphal AH, de Kok A. 1988. Lipoamide dehydrogenase from *Azotobacter vinelandii*: Molecular cloning, organization and sequence analysis of the gene. *Eur J Biochem* 172:299–305.
- Westphal AH, de Kok A. 1990. The 2-oxoglutarate dehydrogenase complex from *Azotobacter vinelandii*. 2. Molecular cloning and sequence analysis of the gene encoding the succinyltransferase component. *Eur J Biochem* 187:235–239.
- Wheelock MJ, Komuniecki R, Duran E, Johnson KR. 1991. Characterization of cDNA clones for the beta subunit of pyruvate dehydrogenase from *Ascaris suum*. *Mol Biochem Parasitol* 45:9–18.
- Zhang B, Kuntz M, Goodwin GW, Harris RA, Crabb DW. 1987. Molecular cloning of a cDNA for the E1-alpha subunit of rat liver branched chain alpha-ketoacid dehydrogenase. *J Biol Chem* 262:15220–15224.
- Zhao Y, Kuntz MJ, Harris RA, Crabb DW. 1992. Molecular cloning of the E1 beta subunit of the rat branched chain alpha-ketoacid dehydrogenase. *Biochim Biophys Acta* 1132:207–210.
- Zhu PP, Reizer J, Peterkofsky A. 1994. Unique dicistronic operon (*ptsI-crr*) in *Mycoplasma capricolum* encoding enzyme I and the glucose-specific enzyme IIa of the phosphoenolpyruvate:sugar phosphotransferase system: Cloning, sequencing, promoter analysis, and protein characterization. *Protein Sci* 3:2115–2128.
- Zhu PP, Reizer J, Reizer A, Peterkofsky A. 1993. Unique monocistronic operon (*ptsH*) in *Mycoplasma capricolum* encoding the phosphocarrier protein, HPr, of the phosphoenolpyruvate:sugar phosphotransferase system. *J Biol Chem* 268:26531–26540.