

Western equine encephalomyelitis virus, complete genome

NCBI Reference Sequence: NC_003908.1

FASTA Graphics

Go to:

LOCUS NC_003908 11484 bp ss-RNA linear VRL 13-AUG-2018

DEFINITION Western equine encephalomyelitis virus, complete genome.

ACCESSION NC_003908

VERSION NC_003908.1

DBLINK BioProject: PRJNA485481

KEYWORDS RefSeq.

SOURCE Western equine encephalitis virus

ORGANISM Western equine encephalitis virus
Viruses; Riboviria; Orthornavirae; Kitrinoviricota; Alsuviricetes;
Martellivirales; Togaviridae; Alphavirus.

REFERENCE 1 (bases 1 to 11484)

AUTHORS Netolitzky,D.J., Schmaltz,F.L., Parker,M.D., Rayner,G.A.,
Fisher,G.R., Trent,D.W., Bader,D.E. and Nagata,L.P.

TITLE Complete genomic RNA sequence of western equine encephalitis virus
and expression of the structural genes

JOURNAL J. Gen. Virol. 81 (Pt 1), 151-159 (2000)

PUBMED 10640553

REFERENCE 2 (bases 1 to 11484)

CONSRTM NCBI Genome Project

TITLE Direct Submission

JOURNAL Submitted (01-AUG-2000) National Center for Biotechnology
Information, NIH, Bethesda, MD 20894, USA

REFERENCE 3 (bases 1 to 11484)

AUTHORS Netolitzky,D.J., Schmaltz,F.L., Parker,M.D., Rayner,G.A.,
Fisher,G.R., Trent,D.W., Bader,D.E. and Nagata,L.P.

TITLE Direct Submission

JOURNAL Submitted (08-DEC-1999) Medical Countermeasures Section, Defence
Research Establishment Suffield, P.O. Box 4000, Stn Main, Medicine
Hat, Alberta T1A 8K6, Canada

COMMENT VALIDATED REFSEQ: This record has undergone validation or
preliminary review. The reference sequence was derived from
AF214040.

The mature peptides were added to the annotation at the NCBI.

COMPLETENESS: full length.

FEATURES Location/Qualifiers

source 1..11484
/organism="Western equine encephalitis virus"
/mol_type="genomic RNA"

```

/strain="71V-1658"
/db_xref="taxon:11039"
misc_feature 1..11484
    /note="sequence of this entry added to the sequence of
    AF214039, strain CBA87, constitutes the Western equine
    encephalomyelitis virus complete genome"
5'UTR <1..24
gene 25..7428
    /locus_tag="Weevgp1"
    /db_xref="GenelD:944530"
CDS 25..7428
    /locus_tag="Weevgp1"
    /note="The mature peptides were added to the annotation at
    the NCBI; encodes nsP1, nsP2, nsP3 and nsP4 proteins"
    /codon_start=1
    /transl_except=(pos:5581..5583,aa:OTHER)
    /product="nonstructural polyprotein"
    /protein_id="NP_640330.1"
    /db_xref="GenelD:944530"

/translation="MERIHVDLDADSPYVKSLQRTFPQFEIARQVTDNDHANARAFS
HVATKLIESEVDRDQVILDIGSAPVRHAHSNHRYHCICPMISAEDPDRLQRYAERLKK
SDITDKNIASKAADLLEVMSTPDAETPSLCMHTDATCRYFGSVAVYQDVYAVHAPTSI
YHQALKGVRTIYWIGFDTPPFMYKNMAGSYPTYNTNWADERVLEARNIGLGNSDLQES
RLGKLSILRKKRLQPTNKIIFSVGSTIYTEDRSLLRSWHLPNVFHLKGKSNFTGRCGT
IVSCEGYVIKKITISPGLYGKVENLASTMHREGFLSCKVTDTLRGERVSFAVCTYVPA
TLCDQMTGILATDVSVDDAQKLLVGLNQRIVVNGRTQRNTNTMQNYLLPVVAQAFSRW
AREHRADLDDEKELGVRERTLTMGCCWAFKTQKITSYKKPGTQTIKKVPAVFDSFVI
PRLTSHGLDMGFRRLKLLLEPTVKPAPAITMADVEHLRGLQQEAEVAAAEIREAL
PPLLPEIEKETVEAEVDLIMQEAGAGSVETPRGHIRVTSYPGEEKIGSYAILSPQAVL
NSEKLACIHPLAEQVLVMT HKGRAGRYKVEPYHGKVIVPEGTAVPVQDFQALSESATI
VFNEREFVNRYLHHIAINGGALNTDEEYKTVKTQD TDSEYVFDIDARKCVKREDAGP

```

LCLTGDLVDPPFHEFAYESLKTRPAAPHKVPTIGVYGVPGSGKSGIISAVTKKDLVV
SAKKENCAEIIRDVRRMRRMDVAARTVDSVLLNGVKHPVNTLYIDEAFACHAGTLLAL
IAIVKPKKVVLCGDPKQCGFFNMMCLKVHFNHDICTEVYHKSISRCTQTVTAIVSTL
FYDKRMKTVNPCADKIIIDTTGTTKPHKDDLILTCFRGWVKQLQIDYKNHEIMTAAAS
QGLTRKGVYAVRYKVNENPLYSQTSEHVNVLTRTEKRIVWKTLAGDPWIKTLTAKYP
GDFTASLDDWQREHDAIMARVLDPQTADVFNKVNVCWAKALEPVLATANIVLTRQQ
WETLHPFKHDRAYSPEMALNFFCTRFFGVDLDSGLFSAPTVALTYRDQHWDNSPGKNM
YGLNREVAKELSRRYPCITKAVDTGRVADIRNNTIKDYSPTINVVPLNRRLPHSLIVD
HKGQGTDDHSGFLSKMKGKSVLVIGDPISIPGKKVESMGPLPTNTIRCDLDLGIPSHV
GKYDIIFVNVRTPYRNHHYQQCEDHAIHHSMLTCKAVHHLNTGGTCVAIGYGLADRAT
ENIITAVARSFRFTRVCQPKNTAENTEVLFVFFGKDNGNHTHDQDRLGVVLDNIYQGS
TRYEAGRAPAYRVIRGDISKADQAIVNAANSKGQPGSGVCGALYRKWPAAAFDRQPIA
VGTARLVKHEPLIIHAVGPNFSKMPEPEGDLKLAAYMSIASIVNAERITKISVPLLS
TGIYSGGKDRVMQSLHHLFTAFDTTDADVTIYCLDKQWETRIIEAIHRKESVEILDDD
KPVDidLVRVHPNSSLAGRPGYSVNEGKLYSYLEGTRFHQTAKDIAEIHAMWPNKSEA
NEQICLYILGESMSSIRSKCPVEESEASAPPHTLPCLCNyamtaervyrlrsakkeqf
AVCSSFLLPKYRITGVQKLQCSKPVLFSGVVPPAVHPRKYAEIILETPPPPATTTVIC
EPTVPERIPSPVISRAPSAESLLSLGGVSFSSSATRSSTAWSYDRRFVVTADVHQAN
TSTWSIPSAPGLDVQLPSDVTDSHWSIPSASGFVVRTPSVQDLTAECAKPRGLAEIMQ
DFNTAPFQFLSDYRPVPAPRRRPIPSRSTASAPPVPKPRRTKYQQPPGVARAISEAE
LDEYIRQHSNXRYEAGAYIFSSETGQGHLQQKSVRQCKLQEPILDRAVHEKYYAPRLD
LEREKMLQKKLQLCASEGNRSRYQSRKVENMKAITAERLISGLGTYLSSEVNPVECYR

VNYPVPIYSSTVINRFTSAEVAVKTCNLVIQENYPTVASYCITDEYDAYLDMVDGASC
 CLDTATFCPAKLRSYPKKHSYLQPEIRSAVPSPIQNTLQNVLAAATKRNCNVTQMREL
 PVLDSAAFNVDCFKKYACNDEYWDTRDNPIRLTTENVVTQYVTKLKGPKAAALFANTH
 NLKPLQEIPMDQFVMDLKRDVKVTPGTHTEERPKVQVIQAADPLATAYLCGIHRELV
 RRLNAVLLPNIHTLFDMSAEDFDAIIAEHFHHGDPVLETDIASFDKSEDDAIAISALM
 ILEDLGVDQPLLDLIEAAFGNITSVHLPTGTRFKFGAMMKSGMFLTLFVNTLVNIMIA
 SRVLRERLTTSACAASIGDDNIVHGVVSDTLMAERCATWLNMEVKIIDAVIGIKAPYF

CGGFILVDQITGTACRVADPLKRLFKLGKPLPVDDTQDCDRRRALHDEAMRWNRIGIT

	DELVKAVESRYEIIAGLIITSLSTLAESVKNFKSIRGSPITLYG"
mat_peptide	25..5601
	/locus_tag="Weevgp1"
	/product="nonstructural protein P123"
	/note="together with nsP4, constitutes the minus-strand polymerase"
	/protein_id="NP_818937.1"
mat_peptide	25..1623
	/locus_tag="Weevgp1"
	/product="guanine-7N-methyltransferase"
	/note="capping enzyme and polymerase complex component; involved in initiation of negative strand RNA synthesis; in SFV nsP1, residues 245-264 are responsible for binding with anionic membrane phospholipids; probably mediates association of viral replicative complex with membranes; guanylyltransferase; nonstructural protein nsP1"
	/protein_id="NP_818933.1"
mat_peptide	1624..4005
	/locus_tag="Weevgp1"
	/product="NTPase"
	/note="viral RNA capping enzyme and polymerase complex component; the NTPase/RNA helicase/5'-triphosphatase domain is located in the N-terminal half of the protein and the proteinase at the C-terminus; the proteinase domain mediates nonstructural polyprotein processing; RNA 5'-triphosphatase; nonstructural protein nsP2; papain-like proteinase"
	/protein_id="NP_818934.1"
mat_peptide	4006..5601

	/locus_tag="Weevgp1" /product="nonstructural protein nsP3" /note="phosphoprotein, polymerase complex component; predicted phosphoesterase (similar to the Appr-1'-p processing enzyme) formerly known as 'X-domain'; together with the membrane binding nsP1 domain, SFV nsP3 mediates targeting the non-structural polyprotein to the endosome." /protein_id="NP_818935.1"
mat_peptide	5602..7425 /locus_tag="Weevgp1" /product="RNA-dependent RNA polymerase catalytic subunit" /note="nonstructural protein nsP4" /protein_id="NP_818936.1"
gene	7473..11183 /locus_tag="Weevgp2" /db_xref="GenelD:944529"
CDS	7473..11183 /locus_tag="Weevgp2" /note="The mature peptides were added to the annotation at the NCBI; encodes nucleocapsid, E3, E2, 6K and E1 proteins" /codon_start=1 /product="structural polyprotein" /protein_id="NP_640331.1" /db_xref="GenelD:944529"

/translation="MFPYPQLNFPPVYPTNPMA YRDPNPPRRRWRPFRPPLAAQIEDL
 RRSIVNLTFKQRSPNPPPGPPPKKKKSAPKPKPTQPKKKKQQA KRTKRKPKPGKRQRM
 CMKLESDKTFPIMLNGQVNGYACVVGGRMLKPLHVEGKIDNEQLAAVKLKKASMYDLE
 YGDVPQNMKSDLQYTSKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNR
 GRVVAIVLGGANEGTRTALS VVTWNQKGV TIRDTP EGSEPWSLVTALCVLSNVT FPCD
 KPPVCYSLTPERTLDVLEENV DNP NYDTLLENVLKCP SRRPKRSITDDFTLTSPY LGF
 CPYCRHSTPCFSPIKIENVWDESDDGSIRIQVSAQFGYNQAGTADVTKFRYMSFDHDH
 DIKEDSMEKIAISTSGPCRRLGHKG YFLLAQCPPGDSVTVSITSGASENSCTVEKKIR
 RKFVGREEYLFPPVHGKLVKCHVYDHLKETSAGYITMHRPGPHAYKSYLEEASGEVYI

KPPSGKNVTYECKCGDYSTGIVSTRTKMNGCTKAKQCIAYKSDQTKWVFNSPDLIRHT
 DHSVQGLHIPFRLTPTVCPVPLAHTPTVTKWFKGITLHLTAMRPTLLTTRKLGLRAD
 ATAEWITGSTSRNFSVREGLEYVWGNHEPVRVWAQESAPGDPHGWPHEIIHHYYHRH
 PVYTVIVLCGVALAILVGTASSAACIAKARRDCLTPYALAPNATVPTALAVLCCIRPT
 NAETFGETLNHLWFNNQPFLWAQLCIPLAALVILFRFCSCCMPFLLVAGVCLGKVDAF
 EHATTVPNVPGIPYKALVERAGYAPLNLEITVVSELTPSTNKEYVTCKFHTVIPSPQ
 VKCCGSLECKASSKADYTCRVFGGVYPFMWGGAQCFCDSSENTQLSEAYVEFAPDCTID
 HAVALKVHTAALKVGLRIVYGNTTAHLDTFVNGVTPGSSRDLKVIAGPISAAAFSPFDH
 KVVIRKGLVYNYDFPEYGAMKPGAFGDIQASSLDATDIVARTDIRLLKPSVKNIHVPY
 TQAVSGYEMWKNNSGRPLQETAPFGCKIEVEPLRASNCAYGHIPISIDIPDAAFVRSS
 ESPTILEVSCTVADCIYSADFGGSLTLQYKADREGHCPVHSHSTTAVLKEATTHVTAV
 GSITLHFSTSSPQANFIVSLCGKKSTCNAECKPPADHIIGEPHKVDQEFQAAVSKTSW
 NWLLALFGGASSLIVGLIVLVCSSMLINTRR"

mat_peptide	7473..8249	
	/locus_tag="Weevgp2"	
	/product="C protein"	
	/note="the capsid protein is a serine proteinase that releases itself from the precursor"	
	/protein_id="NP_818938.1"	
mat_peptide	8250..8429	
	/locus_tag="Weevgp2"	
	/product="E3 protein"	
	/note="In some other alphaviruses, E3 was not detected as a separate protein"	
	/protein_id="NP_818939.1"	
mat_peptide	8430..9698	
	/locus_tag="Weevgp2"	
	/product="E2 protein"	
	/protein_id="NP_818940.1"	
mat_peptide	9699..9863	
	/locus_tag="Weevgp2"	
	/product="6K protein"	
	/protein_id="NP_818941.1"	

mat_peptide	9864..11180 /locus_tag="Weevgp2" /product="E1 protein" /protein_id="NP_818942.1"
gene	7473..9910 /locus_tag="Weevgp3" /db_xref="GenelD:13165415"
CDS	join(7473..9827,9827..9910) /locus_tag="Weevgp3" /note="Truncated version of structural polyprotein that will be produced when frameshifting occurs at nt 9827; Truncated version of structural polyprotein and transframe fusion protein have been added by NCBI Staff with the kind help of Dr. David Karlin (Department of Zoology, University of Oxford, UK) and Dr. Andrew Firth (Department of Pathology, University of Cambridge, UK)" /codon_start=1 /product="truncated polyprotein" /protein_id="YP_006491227.1" /db_xref="GenelD:13165415"

/translation="MFPYPQLNFPPVYPTNP MAYRDPNPPRRRWRPFRPPLAAQIEDL

RRSIVNLTFKQRSPNPPPGPPPKKKKSAPKPKPTQPKKKKQQA KRTKRKPKPGKRQRM

CMKLESDKTFPIMLNGQVNGYACVVGGRMLKPLHVEGKIDNEQLAAVKLKKASMYDLE

YGDVPQNMKSDTLQYTS DKPPGFYNWHHGAVQYENGRFTVPRGVGGKGDSGRPILDNR

GRVVAIVLGGANEGTRTALS VVTWNQKGV TIRDTP EGSEPWSLVTALCVLSNVT FPCD

KPPVCYSLTPERTLDVLEENVDNPNYDTLLENVLKCP SRRPKRSITDDFTLTSPYLG F

CPYCRHSTPCFSPIKIENVWDESDDGSIRIQVSAQFGYNQAGTADVTKFRYMSFDH DH

DIKEDSMEKIAISTSGPCRR LGHKGYFLLAQCPPGDSVTVSITSGASENSCTVEKKIR

RKFVGREEYLFPPVHGKLVKCHVYDHLKETSAGYITMHRPGPHAYKSYLEEASGEVYI

KPPSGKNVTYECKG DYSTGIVSTRTKMNGCTKAKQCIAYKSDQTKWVFNSPD LIRHT

DHSVQGKLHIFRLTPTVCPVPLAHTPTVTKWFKGITLHLTAMRPTLLTTRKLGLRAD

ATAEWITGSTSRNFSV GREGLEYVWGNHEPVRVWAQESAPGDPHGWPHEIIHHYHRH

PVYTVIVLCGVALAILVGTASSAACIAKARRDCLTPYALAPNATVPTALAVLCCIRPT

NAETFGETLNHLWFNNQPFLWAQLCIPLAALVILFRFCSCCMPFLIGCRRLPGEGRRL

RTCDHCAKCSGDPV"

mat_peptide join(9699..9827,9827..9907)

/locus_tag="Weevgp3"

/product="transframe fusion protein"

/note="Transframe fusion (TF) protein expressed via
programmed ribosomal frameshifting. TF protein presumably
plays a stabilizing role in the virion structure."

/protein_id="YP_006491228.1"

3'UTR 11184..11484

ORIGIN

1 accctacaaa ctaatcgatc caatatggaa agaattcacg ttgacttaga tgctgacagc
61 ccgatgtgca agtcgttaca gcgagcggtt ccacaatttg agatcgaagc aaggcagggtc
121 actgacaatg accatgccaa tgccagagcg ttttcgatg tggcaacaaa gtcattgag
181 agcgaagtgc accgggacca agttatcttg gacattggaa gtgcgcccgt cagacatgca
241 cattccaatc accgctatca ttgatctgc cctatgataa gcgctgaaga cccggacaga
301 ctacaacggt atgcagaaag acttaagaaa agtgacatta ccgacaagaa catagcctct
361 aaggcggcag acctgctgga agtcatgtca acaccagacg cagagactcc atctctgtgt
421 atgcacacag acgccacgtg taggtacttt ggaagtgtag cagtatacca agatgtgtac
481 gcagtccatg caccgacatc aatctaccac caggcgctta aaggagttag gacaatttac
541 tggataggct ttgacacgac cccttttatg taaaaaaca tggcagggtc ctaccctact
601 tacaacacga actgggctga cgagagagta ttggaagcac gtaacattgg cctcggtaac
661 tcagatcttc aggagagcag gcttggaanaa ctctcaatcc ttaggaagaa gagggtccaa
721 cctactaata agatcatatt ctcggttggt tcaacaatct acacagaaga tagatcactg
781 ttacgtagct ggcattctcc aaacgtgttc cacttgaaag gaaagtctaa cttcacaggt
841 agatgtggga ccattgtcag ctgtgaaggg tacgtcatca aaaagataac gatcagccca
901 ggactatacg gtaaagtga gaacttggcg tccacaatgc atcgcgaggg tttcttgagt
961 tgcaaagtca cagatacgct gcgcggcgag agggtttctt ttgctgtgtg tacgtatgta
1021 ccagccacac tttcgatca gatgacaggg attctggcaa ctgacgttag tgtggatgac
1081 gcacaaaaac tattggttgg gctcaaccaa aggattgtcg tcaatggtag gacgcaaaga
1141 aatactaaca caatgcagaa ctatctatta ccagtggctg cccaggcggt ttccagggtg
1201 gcgcgtgaac atcgtgccga cttggacgac gagaaagaac taggggtgcg ggagcgcact
1261 ctactatgg gctgctgctg ggctttcaag acccagaaaa tcacatccat ctacaagaag
1321 cctggtacgc aaacaattaa gaaagtacct gccgtctttg actcatttgt gattccagc
1381 cttaccagcc acgggctcga tatgggcttc cgccgtaggc tcaagctgct gcttgaacca
1441 actgtcaaac ccgcaccggc tattacaatg gccgatgtgg agcatctgcg tggcttacag
1501 caagaagctg aagaagtggc tgcagcggaa gagatcagag aagccctgcc acccttgctc
1561 cctgaaatag aaaaagagac cgtagaggca gaagtagacc tcattatgca agaggcagga
1621 gcaggtagcg tggagacacc acgaggacac atcagggtga caagttacc aggccaagag
1681 aagattgggt cttacgctat actttaccc caggcggtat tgaatagtga aaaactggcg
1741 tgtatccacc cattggcgga acaagtactg gtaatgactc acaaaggtag ggcagggaga

1801 tacaaagtcg agccatacca cggtaaggtc attgtaccag aagggacggc ggtccctgtt
1861 caagacttcc aggcattgag tgagagcgct acgatcggtt tcaacgagag ggagttcgta
1921 aacagatacc tgcaccacat cgcaatcaac ggaggagcgc taaactga cgaagagtag
1981 tataagactg taaagactca ggacacagac tcagaatacg tcttcgatat tgacgcacga
2041 aagtgtgtta agcgagaaga cgcagggtccc ttgtgcctaa ccggtgatct ggtagatcca
2101 ccatttcacg agtttgcgta cgagagtctc aagacacgac cagcagcacc tcacaaagtc
2161 ccaaccatcg gagtctatgg agtgccaggc tcaggtaaatt ctggaatcat caaaagcgt
2221 gtgactaaga aagatctggt tgtgagtgcg aagaaggaaa actgcgaga aatcatcagg
2281 gatgtaagga ggatgagacg tatggatgtt gctgctagga ctgtcgattc agtgcttcta
2341 aatgggggtta agcaccctgt taactctgt tacattgatg aggcatttgc ctgccatga
2401 gggacgctgc tggcactgat tgccatcgtc aaacctaaaga aagtggattt gtgcggggac
2461 ccaaaacaat gcggcttctt taacatgatg tgcctgaaag tacattttaa ccatgacata
2521 tgcactgaag tgtaccataa aagcatctct aggagggtgca cacagactgt aaccgccatc
2581 gtctccacgc tctctacga caagcgaatg aagacgggtta acctatgtgc tgataaaatc
2641 atcatagata ccacagggac cacaagccg cacaagatg atctgattct aacctgttcc
2701 agaggatggg tgaacagct acagattgac tacaaaaatc acgaaatcat gactgcggct
2761 gcatcgcaag gacttacgcg gaaaggcgtt tatgtgtca ggtacaaagt caacgagaat
2821 ccactctact cgcagacttc tgagcacgtg aacgtgttac ttacacgcac agaaaaacgc
2881 attgtctgga agacgctagc tggatgccccc tggataaaga cacttacagc taaatatccc
2941 ggggatttca cggcttcatt ggacgactgg cagcgcgaac acgacccat tatggcacgc
3001 gttcttgata agccgcagac agctgatgtg ttccagaata aggtgaacgt ctgtggggc
3061 aaggcttag agccagtctt ggccacggcc aacattgtgc tgacgagaca gcagtgggag
3121 acgttgacc cattcaagca tgacagagcg tactcacctg aaatggcact gaacttctt
3181 tgcaccaggt tcttggagt agacctggac agtgggttat ttccgctcc taccgtcgca
3241 ctacttaca gggatcagca ctgggataac tcgccaggga agaacaatga tgggcttaat
3301 agagaggtag caaaggagtt gtcacggcga tatccgtgca tcacaaaagc ggttgacaca
3361 ggcagggtag ctgatataag gaataatacc atcaaggact actctccaac aattaatgtg
3421 gttccattaa atcgccggtt gcccactcg ttgatcgtt accacaaagg acagggatca
3481 actgatcaca gcggttcct atctaagatg aagggcaaatt ctgtgttggt gatcggcgat
3541 cctatcagca ttccaggga gaaagtagag tccatgggtc cattgccac taataccatc
3601 aggtgtgatc tcgatttggg aatacttagc catgtcggta aatatgacat tatcttgtc
3661 aatgttagga ccccgtagc gaaccatcac taccaacagt gcgaggatca cgctatccac
3721 cacagcatgc taacgtgtaa gggtgtccac cacctgaaca ctggcggaac atgtgtggct
3781 atagggtatg ggcttgctga tcgcgaacc gagaatatca tactgcggt ggcacgtca
3841 ttaggttta cccgtgtctg tcagcctaag aactgccc aaaactga ggttctctc
3901 gtgttctcg gcaaggacaa cggcaaccac acacatgacc aggacagact cgggttagtg
3961 cttgacaaca tctatcaagg gtcaaccagg tacgaggcag ggagagctcc agcgtacaga
4021 gtgatcagag gtgacattag caagagcgct gaccaagcta tcgttaatgc tgctaatagc
4081 aaaggtaac caggttccg agtgtcgggt gactgtacc gaaaatggcc ggctgcttt
4141 gatagacagc caatagctgt cgggacggct agacttgtga agcacgaacc gctcatcata
4201 catgctgtag gacccaattt ttctaagatg ccggaaccgg agggcgacct taagctcgca
4261 gctgcctaca tgagcatagc gtccatcgtc aacgtgagc ggattacaaa aatatcagta
4321 ccgctactgt caaccggcat ctattctggt ggcaaagatc gagtgtatga atcattgcat
4381 cacctgttca ctgcttcga cactacggat gccgatgta ccatatattg ctggataaaa

4441 caatgggaga ccaggataat cgaggccatt caccgcaaag aaagcgtcga aattctggat
4501 gatgacaagc cagtagacat tgacttgggtc aggggtccacc caaacagctc ttggcaggc
4561 agaccagggtt actccgtcaa tgagggcaag ttgtattcat acctggaagg tacacgattc
4621 catcagaccg ccaaggacat tgccgaaatc catgcaatgt ggccaacaa atctgaggct
4681 aatgagcaga ttgtcttga catcctgggg gagagtatgt ccagcatccg ctccaaatgc
4741 ccagtagagg agtcagaggc gtctgtcca cctcacacac ttccatgcct gtgtaattac
4801 gctatgacgg ctgagcgcgt atacaggttg cgctctgcga agaaagaaca gttcgccga
4861 tgctcatcat tcctgttggc gaagtacagg atcacaggcg tgcagaagct acagtgcagc
4921 aaaccagtcc tgttttcagg cgctgtacca cgggctgtac accccaggaa gtacgcggaa
4981 ataattctag aaacgccacc accgccagca acgacaaccg taatatgtga accaactgtg
5041 ccagaacgta taccagtcc ggtgatttct agagcaccaa gtgcggaatc actgctatcg
5101 cttggcggcg tctcgttctc tagctctgcc acacgctcgt caaccgctg gagcgactat
5161 gacaggcggt ttgtggttac agctgatgtg catcaagcga acacgtctac gtggagcatc
5221 cctagtgtc ctggttga cgccagctg cttctgacg tcaactgattc cactggagt
5281 attccaagt catcaggctt tgaagtgaga acaccatctg tacaggacct aactgaggag
5341 tgtgcgaagc ctcgtggact ggccgaaata atgcaagact tcaatactgc tcctttcag
5401 tttcttcgg actacagacc agtaccggca ccacggagac gccccatccc atcacctaga
5461 tcgacggctt ccgcacctcc agttccaaag ccacgcagga ctaagtacca acaaccacca
5521 ggagtcgcta gagcgatctc agaagcggag ttggacgagt acatccgtca acactccaac
5581 tgacgggatg aagcgggagc gtatatttc tcacggaaa caggccaagg tcacctcaa
5641 cagaaatcag tacgtcaatg taaactacaa gaacctatat tggatcgggc cgtccatgag
5701 aagtattacg cccgcgcct cgatctcgaa agagagaaaa ttttacagaa gaaactgcaa
5761 ttatgcgcct ctgaaggaaa tagaagcagg tatcaatcac gaaaagtaga aaatatgaaa
5821 gcaattacag cggagcgact catttctgga ttgggcacat atctatcatc agaagtgaat
5881 cctgtcgagt gttacagagt caattatcct gtaccaatct actcgtcaac ggtaattaac
5941 aggtttacat ctgcagaggt cgcggttaa acgtgcaact tagttatcca agagaattac
6001 cctacagtag ccagtattg tataacagat gaatacagat cgatatctga catggtggac
6061 ggcgcatcgt gctgtctaga tacagccact tttgtccgg ctactgag aagctaccca
6121 aagaagcata gctatttga gccagagata agatcagccg tccatcgcc tatacagaat
6181 acattacaaa atgtattggc tgcagctact aaaaggaatt gcaacgttac ccaaatgcga
6241 gaattacctg tcttagattc ggcggcattt aatgttgatt gtttaagaa atacgcatgc
6301 aatgatgagt actgggatac cttcgcgat aaccctattc ggctaactac agagaacgtt
6361 acgcaatatg tgacaaagct gaaagggccg aaagcagcag cattgtttgc gaatactcat
6421 aatctaaaac cgttcagga gataccaatg gatcaattcg tcatggatct aaagagagat
6481 gtcaaagtta ccccggcac gaaacataca gaggagcggc ctaagggtgca ggtattcag
6541 gctgcagatc ccctgtctac cgcttacctt tgcgggatcc atcggaatt agtccgtaga
6601 ctgaatgcgg tgcttctgcc aaatatccat actctctcg acatgtcagc ggaagatttt
6661 gatgcgatta ttgctgaaca tttccaccac ggcgaccag tattggaaac ggacatcgcg
6721 tcgtttgata aaagcgaaga cgacgctatc gccatttcgg cgttgatgat ccttgaggac
6781 ttaggtgtcg accaaccgct ctagatttg atagaggcgg cgttcggcaa tatcacatct
6841 gtgcacctac ctacaggaac gaggtttaa tttggtgcca tgatgaaatc cggtatgttc
6901 ttaacgctgt ttgtcaacac actagtcaat atcatgattg ctagcagagt actacgtgaa
6961 cgtttaacca cgtcagcgtg cgcgccctct atcggcgacg ataacatagt gcatggtgtc
7021 gtctccgaca cttgatggc ggagagatgc gccactggc tgaacatgga agtaaaaatt

7081 attgatgcag ttattggtat caaagcacc tacttctgtg ggggatttat cctggtggac
7141 cagataacag gcacagcctg cagagtcgca gaccctctaa aaaggctttt taagcttga
7201 aaaccattgc cagtcgatga tacccaagac tgcgaccgcc gccgggact gcgatgatga
7261 gcaatgcat ggaacagaat tggaattacg gacgagttag tgaaggccgt agaattcaga
7321 tacgagatca tactggcagg cctgatcatc acgtctctgt ccacgttagc cgaaagcgtt
7381 aagaacttca agagcataag agggagccca atcacctct acggctgacc taaatagggtg
7441 acgtagtaga cacgcaccta cccaccggca gaatgtttcc ataccctcag ctgaactttc
7501 caccagtta cctacaaat ccgatggctt accgagatcc aaaccctct aggcgccgt
7561 ggaggccgtt tcggcccccg ctggctgctc aaatcgaaga tctaggagg tcgatagtca
7621 acttgacttt caaacaacga tcacctaate cgccgccagg tccaccgcca aagaagaaga
7681 agagtgtcc taagccaaaa cctactcagc ctaaaaagaa gaagcagcaa gccaagaggaa
7741 cgaaacgcaa gcctaaacca gggaaacgac aacgtatgtg tatgaagttg gagtcggaca
7801 agacatttcc gatcatgctg aacggccaag tgaatggata tgcctgcgtt gtcggaggaa
7861 ggctgatgaa accactcac gttgaaggaa aaattgataa tgagcaatta gcggccgtga
7921 aattgaagaa ggctagcatg tacgacttgg agtacggcga cgttcccag aacatgaaat
7981 cagacacgct gcagtaacc agcgacaaac caccgggctt ctacaactgg caccacggcg
8041 cagtccagta tgagaatggg agatttaccg taccgagagg agtggggcgg aaaggcgaca
8101 gcggaagacc gatcctggac aacagaggca gattgtggc tattgttcta ggagggtcaa
8161 atgagggcac gcgtacggcg cttcagtgg tcaactggaa ccagaaagg gtgaccatta
8221 gggatacccc cgaaggttct gaaccgtggt cactagtac agcgtatgc gtgcttctga
8281 atgtcacgtt cccatgcgac aaaccacccg tgtgctattc actgacgcca gaacgaacac
8341 tcgacgtgct cgaagagaac gtcgacaatc caaattacga cagctgctg gagaacgtct
8401 tgaaatgtcc atcacgccg cccaaacgaa gcattaccga tgactcaca ctgaccagtc
8461 cctacctggg gttctgccc tattgcagac actcaacgcc gtgttccagc ccaataaaaa
8521 ttgagaacgt gtgggacgaa tctgatgatg gatcgattag aatccaggtc tcggcacaat
8581 tcggctacaa tcaggcaggc actgccgatg tcaccaaatt ccgttacatg tcttccgacc
8641 acgaccatga catcaaggaa gacagtatgg agaaaatagc tatcagcaca tctggacct
8701 gccgtctct tggccacaaa gggacttcc tgttagctca atgtctcca ggtgacagt
8761 taaccgtcag tatcacgagc ggagcatctg agaattcatg caccgtggag aaaaagatca
8821 ggagggaagt gtgcgtaga gaggagtact tgtcccacc cgtccatgga aagctggtaa
8881 agtgccacgt ttacgatcac ttgaaggaga cgtctgccg gtacataacc atgcacaggc
8941 caggcccaca cggtataag tctatctgg aggaagcgtc aggcgaagtg tacattaac
9001 caccttctgg caagaacgtc acctacgaat gtaagtgtgg cgactacagc acaggtatcg
9061 tgagcacgac aacgaagatg aacggctgca cttaaagcaa acagtgcatt gcctacaaga
9121 gcgaccaaac gaaatgggtc ttaactcgc cggatcttat taggcacaca gaccactcag
9181 tgcaaggtaa attgcacatt ccattccgct tgacaccgac agtctgccc gttccgttag
9241 ctacacgcc tacagtcacg aagtgggtca aaggcatcac cctccactg actgcaatgc
9301 gaccaacatt gctgacaacg agaaaattgg ggctgcgagc agacgcaaca gcagaatgga
9361 ttacagggtc tacatccagg aattttctg tggggcgaga agggctggag tacgtatggg
9421 gtaaccatga accagtcaga gtctgggccc aggagtcggc accaggcgac ccacatggat
9481 ggccgcatga gatcatcatc cactattatc atcgcatcc agtctacact gtcattgtgc
9541 tgtgtggtgt cgctcttct atcctggtag gactgcac atcagcagct tgcacgcca
9601 aagcaagaag agactgcctg acgcatacgc cgctgcacc gaacgcaacg gtaccacag
9661 cattagcggg ttgtgctgc attcgccaa ccaacgctga aacattgga gaaacttga

9721 accatctgtg gtttaacaac caaccgttc tctgggcaca gttgtgcatt cctctggcag
9781 cgcttggtat tctgtccgc tgcttttcat gctgcatgcc tttttattg gttgcaggcg
9841 tctgcctggg gaaggtagac gccttcgaac atgcgaccac tgtgccaaat gttccgggga
9901 tcccgataa ggcgttggtc gaacgcgcag gttacgcgc acttaacctg gagatcacgg
9961 tcgtctcatc ggaattaaca cttcaacta acaaggagta cgtgacctgc aaattccaca
10021 cagtcattcc ttcaccacaa gttaaatgct gcgggtccct cgagtgcag gcatcctcaa
10081 aggcggatta cacatgccgc gtttttggcg gtgtgtaccc tttcatgtgg ggaggcgcac
10141 aatgcttctg tgacagttag aacacacaac tgagttaggc gtacgtcgag ttcgtccag
10201 actgcactat agatcacgca gtcgcactaa aagttcacac agctgctctg aaagtggcc
10261 tgcgtatagt atacggcaac accaccgcgc acctggatac gttgtcaat ggcgtcacgc
10321 caggttctc acgggacctg aaggtcatag caggggccgat atcagccgct tttcacct
10381 ttgaccataa ggtcgtcatc agaaaggggc ttgttataa ctacgacttc cctgagtatg
10441 gagtatgaa accaggagcg ttcggcgata ttcaagcatc ctcgcttgat gctacagaca
10501 tagtagcccg cactgacata cggctgctga agccttctgt caagaacatc cacgtcccct
10561 acaccaagc agtatcaggg tatgaaatgt ggaagaaca ctcaggacga ccctgcaag
10621 aaacagcacc atttgatgt aaaattgaag tggagcctct gcgagcgtct aactgtgctt
10681 acgggcacat ccctatctcg attgacatcc ctgatgcagc tttgtgaga tcatcagaat
10741 caccaacaat ttagaagtt agctgcacag tagcagactg cattattct gcagactttg
10801 gtgttctct aacattacag tacaagctg acaggaggag acattgtcca gttactccc
10861 actccacgac agctgtttg aaggaagcga ccacacatgt gactgccgta ggcagcataa
10921 cactacattt tagcacatcg agcccacaag caaattttat agtttcgcta tgcggcaaga
10981 agtccacctg caatgctgaa tgtaaaccac cggccgacca cataattgga gaaccacata
11041 aagtcgacca agaattccag gcggcagttt ccaaaacatc ttggaactgg ctgctgcac
11101 tgtttggggg agcatcatcc ctcatgttg taggacttat agtgttggtc tgcagctcta
11161 tgcttataaa cacacgtaga tgactgagcg cggacactga catagcggta aaactcgatg
11221 tacttccgag gaagcgtggt gcataatgcc acgcgccgct tgacactaaa actcgatgta
11281 tttccgagga agcacagtgc ataagtctgt gcagtgtcac attaatcgta tatcacata
11341 catattaaca aactatatc acttttatga gactcactat gggctcttaa tatacactac
11401 acatatttta cttaaaaaca ctatacacac ttataaatt cttttataat tttcttttg
11461 tttttattt gtttttaaaa tttc

//