Introduction

The genus *Vibrio* are among the most numerous and ubiquitous marine bacteria and commonly found as etiologic agents of diseases in both vertebrates and invertebrates throughout the world. *Vibrio salmonicida*, a psychrophilic pathogen marine gram-negative microorganism, the etiological agent of cold-water vibriosis (also called hemorrhagic syndrome or Hitra disease) in Atlantic salmon and Atlantic cod, has been studies for some years. Although efficient vaccine has been developed, the molecular mechanisms of host invasion and specificity, colonization and growth are largely unknown.

In order to gain insight into these mechanisms, the genome of *V. salmonicida* have been sequenced, and all genes are being identified and analyzed. Technological advances developed over the last decade, makes genome sequencing a feasible task also for small laboratories.

Problem

You are given the linear sequence a random part of the *V. salmonicida* genome. Your task is to identify and predict the open reading frames (ORF's) encoding protein coding sequences (CDS's), if any in this region of the genome. Then you will annotate the CDS's, that is to assign the possible functions to the proteins encoded by these hypothetical genes. Collect all this information and make an overview where you present all your "evidences".

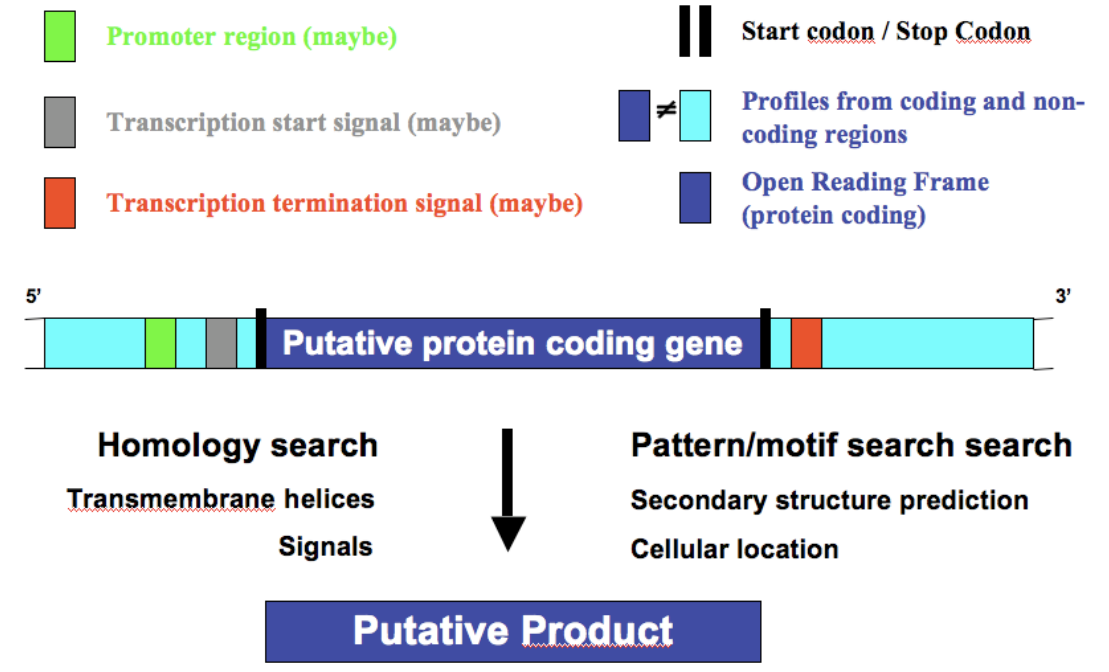


Figure 1. General features associated with a prokaryotic gene, and typical methods used to assign function to a gene product.

Methods

Identification and prediction of ORF`s and CDS`s can be performed using the academic available automated gene prediction programs GeneMark, Glimmer and FGENESB. These programs use different algorithms, so if you run all of these different gene prediction programs on the same piece of DNA they will all come up with different solutions. We recommend that you use more than one program and compare the results to get the best possible identification of the potential CDS's.

GeneMark:	http://opal.biology.gatech.edu/GeneMark/
Glimmer:	http://www.ncbi.nlm.nih.gov/genomes/MICROBES/glimmer_3.cgi
FGENESB (Softberry):	http://www.softberry.com/berry.phtml?topic=gfindb
ORF finder (NCBI)	http://www.ncbi.nlm.nih.gov/gorf/gorf.html

Each predicted CDS must be studied more in detail to find out more about the possible function of these hypothetical genes and their products. Protein function is an open ended issue; there are different levels of functional annotation of proteins, ranging from biochemical functions such as kinase activity, to physiological function such as a role in immunological signalling cascade. Information from various sources needs to be collected and compared to piece together the most likely functions of the protein of interest. Because of the nature of this type of analysis, it is restricted to information that is already "known" and is dependent on regular updates of the relevant databases. In addition, ab-initio analysis of novel sequences can provide clues as to function of a protein through local or global homologies to other proteins for which some functional information is available or from discovery of conserved domains within the amino acid sequence. This type of analysis will increase in accuracy as further genome sequence data is tied in with the ongoing functional studies.

You should start with simple BLASTp and FASTA searches to see if there are any similarities between your hypothetical gene products, and the sequences in the databases you search. Remember that not all hits are necessary good hits. When you search a database you will almost always get a hit, but you have to be critical to what the search tools returns to you. You must take into consideration the similarity score and the E-value. Is it likely that your protein should be similar to a protein from the organism you find matches to?

From the lectures in this course you learned that the structure of proteins are more conserved during evolution than the sequences. So if you can't find any good matches with sequence searches, the story doesn't end here. There are several public databases and resources available, which you can use to analyse your sequences in more detail and maybe be able to assign some functions to these hypothetical proteins.

Take a look at the information available at various web resources in order to assign possible functions to the gene products. Use your protein sequences to search against motif/pattern databases, predict secondary structure, look for transmembrane helices and so on. Collect and compare information and try to find out as much as you can about these hypothetical gene products.

Text searches:

SRS	http://srs.ebi.ac.uk/srsbin/cgi-bin/wgetz?-page+srsq2+-noSession
Entrez	http://www.ncbi.nlm.nih.gov/Entrez/

Sequence similarity searches:

BLAST:	http://www.ncbi.nlm.nih.gov/BLAST/
FASTA:	http://www.ebi.ac.uk/fasta33/

Motif/pattern databases:

InterPro	http://www.ebi.ac.uk/interpro/
Pfam	http://www.sanger.ac.uk/Software/Pfam/
SMART	http://smart.embl-heidelberg.de/
PRINTS	http://www.bioinf.man.ac.uk/dbbrowser/PRINTS/
BLOCKS	http://www.blocks.fhcrc.org/
ProDom	http://protein.toulouse.inra.fr/prodom/current/html/home.php
TIGRfam	http://www.tigr.org/TIGRFAMs/
PROSITE	http://us.expasy.org/prosite/

Other important resources:

t-RNA Scan	http://www.genetics.wustl.edu/eddy/tRNAscan-SE/
PSORT	http://www.psort.org/psortb/index.html
SignalP	http://www.cbs.dtu.dk/services/SignalP-1.1/
TMPRED	http://www.ch.embnet.org/software/TMPRED_form.html
TMHMM	http://www.cbs.dtu.dk/services/TMHMM/
COILS	http://www.ch.embnet.org/software/COILS_form.html
Coiled-coil	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_lupas.html
HTH	http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hth.html
DOTLET	http://www.isrec.isb-sib.ch/java/dotlet/Dotlet.html

Secondary structure prediction:

PredictProtein	http://www.embl-heidelberg.de/predictprotein/
PSIPRED	http://bioinf.cs.ucl.ac.uk/psipred/index.html
PSA	http://bmerc-www.bu.edu/psa/

Alignment:

ClustalW	http://clustalw.genome.ad.jp/
T-Coffee	http://www.cmbi.kun.nl/bioinf/tools/T_COFFEE/

BioEdit (alignment editor) you can download and install the program on your computer http://www.mbio.ncsu.edu/BioEdit/bioedit.html

>Seq_1
ccttattatgctgccctcaattctgactgaactatcgcatgggtgactaaatcattgacc
gaatttccgactcggaaattcgtaaacaccagacgactttcacataaaatacattcgtac
ggatctacttttacatatccttttaacattgcggcataccctggcattttcggctcagct
tctattgtcatacctaaagctgcataaactctaggcagagcttctccacgacgacgcatt
gataaaaaaccgtagtatcggatcatcttgaaatgtttatcaggatagtgctctactatc
cgtcttatcatctcttctggtgataatgttaggctgtctgttgttcctgttcgatggtct
aaataattaaacgttatcattccgcctttggcgtaatgacttaaacgtgacgctgaaatc
gggggccgttttaaataccgaccaagatagttcatcgtcggttttacattatttgtcttt
ttagcaaaatgaagcttccaacgacgattatattgactgcttaaaaagcgtgaccaatcc
gttttattacggatataggggcattcttcgcttgataaatcaagctcataataagcctta
cccaataaactgacgatagccgctctccaacaaggctctgtcgttttcatttggaagtaa
atgggtttccataaaccggtacgttcacaaattccccccacgagtgaccgataaatgtaag
tgcgtattccaattcagttttcgaccgtaagtatgaagagcacaaaagataccgacatct
attcctttatcttttgcccatcccagcagaatgtttgcagcacatttgaataatttattt
aacagccaacggttatgacgaaagataggccatagcgtgtttggaagggtaaaggtgatg
tgttgatattcgcattcagggaagacatgttgttgcttttgtatccatcgctctgtggct
ttcatgccacagctactgcacgctcgagatttacaggtttggtgaatatatttgatatgg
gtacagtcagggttgcaacaatgatattcgcgagagccaaaagccgctgtcccacaggac
agcatctttgtgacattttcaatcacgaccgctcttaggttagctttgttattatgaaga
aatttaagccagttattttgactattaaataattgtttcaggggtttatatgcgtgcatg
gcgataggataaacgatttattggcaacaatggaagaggttgagtccttttgatttgcgc
cgtaggcgcctattgttccaggtttttgccacgctttgattaaacggtttgtgtcttggt
gatgcgtaaagtttgcaccgcccgcccaccagatcattttcaattcagggtaggtaagtg
tatggccgttgtgctgatattctgtgcctggtttttctaaacattcaacaatacgagcaa
cagggaaagcagtaacagaaccttgaacggcccaatcgttaccaccggcagagctcccac
ctaacgatggggacatcgcaggtaatacacccgcatcacgcgttgggttaccgccatttg
agtaatggtatgaataaccaaagccaccgcccggtagaccaacttgacccaacattgaag
ctagaactgttagcatccagtgacgttgctcaccgtgttgttgacgttgcataccccaac
cagccattagcatggtgcggttttctttgaagatcttagccagtagttccatttgttttg
ccggaacaccacagatcgcttctgcccattgtggggttttctcaatgccgtcttctttac
ccattaagtacgcttcaaattgatcaaaaccaaccgtgtatttagcaaggaattcagtat
cgtgttgttctgttttgattaactggtaagcaatacccatcatcatcgcgacatcagatt
gcgggttaggggcaatccattgtgctttatcgccaaagaactcagccgtttctgaacgca
ttgggtcaataataatgactgtttttcctgatttttcagttgatggaagaactcaagtc
cttgcccatcacttgaagaccacgcaatttttaatgtatttaatgggtttaaaccccata
gaacaaccacatcactgctttctaaaataacagggtaagtcgtttgttgttcatacactt
caatagaaccaaccacgtgcggcataattacttgagcagcacctgttgagtaatcaccta
agtgtcctgcataaccaccagacatgctcatgtagcgttgaagtaaagtctgagctttat
gtaatacaccacttgagcgccaaccataagaacctgcgaagatcgattctgggccatacg
cgttacgaattcgtgtatgttgctcatggatcagttgatacgcttgttcccacgatactt
gaacaaattcgtcactgccgcgaacccctgagggttacttgggttttctagataccctt
tacgaaccattgggtattttacacgtgctttagtatgaacctgagaaggtcctgtttgtt
gcaagctgtttggtacggtttgagctaacccactttttgtcgcgataagctgaccatctt
tgacttcacaaagtaatggtcccatacggcttgaggttagaattaatgaatctgtacggc
caagcgtagaggcattcacagataacggagcgatagacgtgatcgctaatgcacccgctg
atgcacctgcgccttttaaaaagccgcgacgagaaatagtagtcatagtcactcccttat
tgatgagttgccacgtctttcgcgtggttttggaagaatttagtcaagatctcaaggtta
agctctgagatatcggtgcgagcgcccatgcttttagcgacaggtccccaagaattcaca
gtgaagtgatttggtgggattttggcgtgacaggttgcacaatagacgttgtctagctct
tctgcgtaatcccataatggttggtttgaagctaaaacaggtgcatcgattgtgccagtt
agctcagcggtgcgccattcgttaccataagcatctgtgttcgcttcacctaatgtcagt
gctttgatgcctgtatcagtaaagagtagcgataattgcacgttggccttctcccatgta
gataaccgattctgcgcctttcatttggtaaccagaaatagtcactgtacgtttgccttc
atcgttactaaccgttgctagttcaaccgttgggttaattgtcgcaagatcggccatttt
aatgcttttaatcgggtaaacttttctgcatccgctttggtgtttttcgccatgtcaaa
aaggtggtcaaaagcagacgtatctaatgtggcttgtggtgcaaagtgagcaacgccttt
atgacaatcagtacaggtttgattgttttctcttccgtatgcgtgcatttttgtgcact
agcagattggtcgtaagtttccattgcatcaaagctgtgacaagaacgacaagtagcaga
atcgttttcgcgcatttgttcccagacggtttctgccatcgctaaacgatgttcttcata
tttctcgtcagtgtctatttttccagtaacaaactcgtgataaatgtctttagacgcacg

aatttttgtaattaggtaatcaataggttcttctggaatatggcaatcggcacattctgc
acggatgccttttggttactgaaatgaactgaaccttggtattcatgcaacggtttttc
cattgtatgacatgaaacacagaattcagtggttgatgtcgcatgaagcacggcttgcgt
ccccccccaaagttatccacccgatggcaataccaacggcagcgattaaagcgatataacg
ttttttcataatatgatccttgatatagccatattctttgtatggtcaatcgcaaatcct
accacttttgtattatttttctatgaaataagtggtatttatgtcaataaatttaatttt
accctctaaattgacatttctgttaataaggtttcatcttaagacgtttttgatgttag
tttcttctttatttgttgttgggttacttttttgttgcaccctatttttttgatttaca
aaactttacgaatgcaaagaatgtaaatttaatacaaataaaaacgattatcatttatat
tccatctcagaatttaatcacccgttttattaatgaaagatttggagataatggaatatg
ttttctaaaagtcagttggctcttgttattggcgccgttttatctatccctttgcttac
gctgaaacagatgcatcaactactgatgagcacatggtcgttactggtcgtgattacggt
tataaagcagacacaaactcaacgtcaatgcgtatggaagcaacgcaactagaaactcct
gggcaagtaacggttattgatgagcaaatcattgatgagcaacgtgcgagtacactaggt
gaagtgctgaagaatgattcaagtatttctgaaggctctaaatcaacgaaccgtgagcgt
ttttacttgcgtggttttaaattaagcagcagttcaagtttcttacgtgatggtgttcaa
cactggtctcattatcgtcaaccaattgagttattagaacgtgtagaagtaatgaaaggc
cctgcaggtttactttacggtaagtccacgcctggtggcttagtgaacatgatttctaaa
aaaccgacttatgaaacacaagtaaccatcagtcaagatatcggttcaaataattatact
cgttcaattgctgacgtaagtggttcattgaatgaggaccaaacattacgtgctcgcgtg
attttatctcaagaaaatcaagattcttggcgtacacgttttgatggcactgctgttgaa
actgatcgtactgttggtggtctttttgttgattatgatattaacgaagacatcatgctt
tctgctcattacgatagaacaaatgaaatcggtgatttggataatggttcaaaggttgat
acaacaaccggtgaagtcattgacccaaatacggttaaagaccaaaactgggcgaagaca
gacaatgatgtggctaattacggtgtaggcttaactgcaaacttgacggattcatggtct
gttaaaacgggtattagccgtcagtattatgagcgtcagcgtactgaatcggatattaat
aaagatggtagtagctacaaatcatcagatcgtcatgatgaatggacatttgatactgct
tatgttgatttcatgggtgatgttgatgccttaggcgttaaccaccgtttattagtgggt
ggtaatggcctgaattatgattataagcgtcaatacgatagctctaaatcatgcataaca
agcactgatacgtgcagtgacggttttaatcgccctagtgatttaaattcaagtaatgat
acggttgggccccactctaagagccaacattacggtatttaccttcaagatttagtgaca
tttaatgactcttggcagttacttgctggcgcacgttttgctgtagataaaacagaagac
agcaaaggaaaaaaagaagattacactaacattctttctaagcttggtgttatttatcat
cctgctgaaaatggttctatttatgccgtatattctgaaagttttgaaccagtagcaacg
ataaatgatgctgatgatgtgaatgatggtcaaaaacaagatccaatgaaaggaaactta
tacgagttaggatctaagtgggaactattagataacagcttgtatgtgtctggtgcttta
ttccaaattgtacagtcaaacatgcaagtgtctgaagatattaacggtggttcaaataca
cgtacttcacaggtcggtgagcaagttcatactggtattgaattagcggcgacgggttac
ctgacagaagcgttctcattaagcgcatcaacgacatttcttgatgctaaatatgaaaat
gatccagcgttagaaggtaaaacacctgcggatgtacctgaatttagcgcaagtatttgg
tctcagtacgatttcaataacggcacaggtattaaccttggtgtttatcacgtaggcgag
cgttatggtgattctaaaaacgaattcaaaaaagacgcttatacacgagtagatatgggt
atctcacataccttaaagtacgacgacagtttagattttgtcgctcgatttaatgtagaa
aacctattagatactgattatttagctggtggtagtgcaagtgatgtggttgttggtgaa
ggccgaaactacatggcgacactgcaagttaaatactaagttagcgtagcaaatataagg
gagaattcggtctccctttaatgcgtttatttgaatgtgaattttatcctttaaacatga
taattaagaaatcgacaatatgaatattttaaaaaccagcattggtatcgccctagctac
catgatcacaaccgctcaagcctctacgcttgatcaagcattatctatcgaaaataaatc
gaatgcagcctctgctgtgagccaaacaaaaatagataagagtgctgaagtgacattaac
gtataaagctgaaattgagcagctacaagaagaagtcaaaaatcttgccgtttatcgtga
tcatttgactggtttagtcgatagccaacaacaagaaatgacgaacttaaataatcaaat
cgagactattaaagacaccagacaaggcgtagtgccattgatgtataagatgctgtcagg
tcttacggtattagttgaacaagataaaccaatcaagattgcagaacgtcagcaacgcat
cgaaaaacttgatcacatgatgacaagagcggatgtgagcgatgcagaaaaataccgccg
tattctggaagcgtatcaaatcgaaatggattacggctcaaaactggggtcataccaaac
tcaaattatggttggaaatgacgaagtgatcgatgcggatatgctgcatttaggtcgtgt
ttctttgcttgctcgtcgtttggattcttcacaatactggagctggaatacctaccgtaa
tcagtggacattattggattcatcatttaatgacgacttagcaaaagcgtattctgtcgc
ttacaaacaagtggcaccgagcttattgactctacccgtttctttatcactgaaagaggt
gaaataatatgtcaatcaaaccacttattttagcgatgtctttattcggtttttccctaa
acacgttagcggcaagtagtgaattagttcaagaaacaaagcaagcaagccatgttcaaa

aacaacacaatgccgaacgtgaagctgactttatgctgacagagaaagaatttaaagcac
aaaagaacgccttgattgctgagcgtaaaaaactgcaaaaagagagcgatgaactgagtt
ctttattcagtagtaatgagcaatcacttgccacgatagagcagcaactacatttagaaa
caggcagtttaggtgagctgtttggtgtggtaagacaaacggcaaaagatcttaattctg
aattaaaaaactcagtgactggttcaattgatgctaataatgcacaagtgattaatgaca
ttgtagaggctaaatcattgccatcaatgactcagttaacagggttatggcaaaccatgg
ttcaacaagtgcaagcgagttcagaatcggcgttaatgccggtgacggtaatcaatgggg
cagggcaccgctctgaaatcaatgcttaccgtttaggcaatatggcgcttgttggtgatg
aaggcttcctacgttgggataataaaaagagcgtagcaacgtattattcacaacagccag
accttgcgccatcagtcgctacattatctgatatgcaacagcagcaagtaatgatgatgg
cagtggatccatcgcgtggcatcatgctagagcaacttgcaaattcgccaacgctaaaag
atcgccttcaagcgggtggtgtggttggacaaatcattatcgctctattggcaattggcc
tgcttatttcgttatttaatggtgcgagattgttggttattcgtcagcagatccgcactc
aattaaaagatcctgcacacatcggtaataatcctcttggtcgtgttttatcagtttatg
atgcagaaaaaactcgttcggtagaggcattagaacttcgcttattagaaacaattgttg
atgagcagcaaggcttagaaaaaggcctatcaatgcttaagcttctcgctgcattagcac
ctatgttaggtctgttaggtacggttactggcatgatcgaaacgttccaagtgatcacgc
aatttggtaatggcgatcctaaagtgatggcgggtggtatttctatggcattggtaacaa
ccgtattgggcttagttgcagcgatgcccctgttattaactcacaatattttaagctctc
aagcagaaagcattcgcactatccttgaaaagcaaggcttgagtttggtggcagagcaag
cagaaaaagaaaccaataatacaccagtaggaatggcagcgtaatgacggcttttctctc
tcattttatggcacaaggcggccccatttatggtggttgggcgcagttgttgcggtatg
ttggttattggtggttgagcgtgtgtttttcttattattcacttttcctgcacaacgaaa
agtatggttagataaatggcaacagcgagacgatcatgattcgtggtatgcgaaaacgat
ccgtgatggttgggtaagcgaagctcatcttcaattaaatcaatatctgaattttatcaa
agttcttgtggcaatttgcccaatgctcggtttgctaggcacagtaacaggcatgatctc
tgtgtttgatgtaatggcaaatcaaggcagcagtcaacctaaattgatggcgtctggtat
ttcattagccacattgcctacgatggctggaatggtagcagcattagcaggaatgtttgt
tcatgctcgtttagccaaagtctgccgccgtaaagaacatcatttagaaaaagcattaag
gagtcagtcatgagacttggccgtcgttcttcaaatcaagaagaagcacagatcgattta
acatcaatgctcgatatcgtatttatcatgctgatatttttattgtgactagctcattt
gtgcgtgaatcaggggtagaagtgaatcgccctcaagcaagcaatgtatcgagtcaaaaa
gacgcgggtatctttgttgctattacggcttcaaacgatattttcattgataaacgcatg
gtggacgccgagcgagttcaagccatattagagcatttattattagaacaacctgatgcg
tcgctcgtcattcaagctgatgaacatgcgtataacggcacggttgtgaaagtgatggat
gcggcaaaaggggctggtgttaaaggcatcgcattggctgcggagaaaaactaatgtggc
gattattgttggcagtgcctattgcgctttctatgacggtagggctattcacgtttatgg
catggatggtagacaacggtcataacggcaaaccagaaagtggtgagacgattcgctttg
acatggtgatggtagagcaagaacaagatgttcagcgccgtcaacgtcgagtgcctgaaa
aaccagaaacaccagagccgccaccagaggcaaaaccaatggcgtctcaaacacaaacgt
cgatggcaacgccgcaaagtaatatgccgctaatcggtcttgattcaagtatttcaggct
tagcgataaatgcaccgagtttcaccgattttacaggtaaccaacaagcgatgccgttgt
atcgtgttgagcctcgttatccttcaaaagcgatgaaaagaggtgctgaggggtttgttg
ttatgtcatttaccattgatgagcaaggtcgcccaaccaatgtaaaagtgacggatgcaa
aaccacgccgaatgtttgaaagagaagcgttaaaagcgctggaaaaatggaaatatcagc
caaaagtggtcgatgggaaagccattgctcaagttggacaaacggtgaaattggagttta
cgttagcaaaatgatgaaattaaaattacattcttctgttcttattcttagcttcgcttt
atttttttctggggcagtgcaagctcagcaacttagtcaatacaatgcgggtaaagtgca
gcgagcgttgcaattacaacagcaagaaaaaatgcccgaggcaattgctgttttggccga
attaacccccttctcaggattatgataaagcctatattcaacgtatgttgggtgtgtttca
ttggcaggccgggagtagtaagtcggccgttaaatatctcactaaagcggtaaatagcca
gcaattagccgatgagcaagcatggataacgcaacgtatgcttgccgatttattattaac
agagcagcgctttagtgatgcgttaccgcattattatcaactagcgaactcagtgcctga
aaaacaaaaaataggcgatttatggttccgtatttctcaaacccattaccagttggcgca
atggggaaaggtactttctgcattaaataagcatgaacgtgttgcgatatataagaaaca
aagtgaaagaagccaatcgctagtgctgaagttaggggcacaactacaattaaaacaatg
gaaatcggcgatccctaccttgaaccgtttaattgtacttgaaccaaataaagcgacatg
gtggcagcagttaacgagtattcagcttcgtttgaatcaacctaaaaatgcgttagatac
tcttattcttgcagagcgacaaggaattgaattgagtcaatcagagttgaaattgttggc
tcaactttatgctcagcgcggtatcccagagaaagccgcagaagtattaagtcagttaga
tgaagcgcaaagtgatattgatttactgacaatgcaggcaagttattggcaagcagcaaa

agagtgggataaagccactgtcacttggtcaaaggctgcaacgctagattcacgttatta
ttggcaagtggcccaattgcagatccaagagggtcatcatcaacaaggcttgaagacttt
agataaggttaaagattacaaaaagacagatgatattgcattagctaaagtgagagccta
ttacaagctaaaccaaatagattcagccttactttatgcgaaaagagccaatgattttac
gccatcaaatgaagcgaaagggtggataaagtatttgtctcaattgagaaaaacagccaa
tacttcagcgtaaactccatttaaggtaaggaatttgaatgttgctgagttagcattggc
ccgtctaaaaaagccaatttcaccatcgctgtttttttaaaaggtatcgaaaatataatt
atcggggattttataactgttttagccctatttcatgagttttagatacacttatcctac
aaaccttattatgctgccctcaattctgactgaactatcgcatgggtgactaaatcattg
accgaatttccgactcggaaattcgtaaacaccagacgactttcacataaaatacattcg
tacggatctacttttacatatccttttaacattgcggcataccctggcattttcggctca
gcttctattgtcatacctaaagctgcataaactctaggcagagcttctccacgacgacgc
attgataaaaaaccgtagtatcggatcatcttgaaatgtttatcaggatagtgctctact
atccgtcttatcatctcttctggtgataatgttaggctgtctgttgttcctgttcgatgg
tctaaataattaaacgttatcattccgcctttggcgtaatgacttaaacgtgacgctgaa
atcgggggccgttttaaataccgaccaagatagttcatcgtcggttttacattatttgtc
tttttagcaaaatgaagcttccaacgacgattatattgactgcttaaaaagcgtgaccaa
tccgttttattacggatataggggcattcttcgcttgataaatcaagctcataataagcc
ttacccaataaactgacgatagccgctctccaacaaggctctgtcgttttcatttggaag
taaatgggtttccataaaccggtacgttcacaaattcccccacgagtgaccgataaatgt
aagtgcgtattccaattcagttttcgaccgtaagtatgaagagcacaaaagataccgaca
tctattcctttatcttttgcccatcccagcagaatgtttgcagcacatttgaataattta
tttaacagccaacggttatgacgaaagataggccatagcgtgtttggaagggtaaaggtg
atgtgttgatattcgcattcagggaagacatgttgttgcttttgtatccatcgctctgtg
gctttcatgccacagctactgcacgctcgagatttacaggtttggtgaatatatttgata
tgggtacagtcagggttgcaacaatgatattcgcgagagccaaaagccgctgtcccacag
gacagcatctttgtgacattttcaatcacgaccgctcttaggttagctttgttattatga
agaaatttaagccagttattttgactattaaataattgtttcaggggtttatatgcgtgc
atggcgataggataaacgatttattggcaacaatggaagaggttgagtccttttgatttg
cgccgtaggcgcctattgttcaatgcccttatcgaatgatgagggcttttctattgctcg
taaaatagacagtaaacaatccccaaacgtaaacaatgtaaattgaatgtcaacgcaaat
cattatcaattacattaccatccattattcttaaatgatcctgtcttattcatgttgctg
aaaaacaaacaagtacaattgtgggcgcgtcgcttacatgtttacgtatcaatggcgctt
ctactcgttgtcttatttttctcgatcacagggatcacgctaaatcgccctgattggttt
gtcagcagctctccagacatcaaaaacaccactctttctgttcctaatgatgtgcttttt
tctctagatgaaaaacaccacgttctttggaataaaccaaacaccgtcgctttgcttgat
tacttaaatcaacataccaacctttctggaacgccttctaatgtcgatattttaccgac
gttgaagacggtgaattagtcgagggtgaattgtcattaaactataaaggccccggttat
aacgcgtctgtttatattgatttaacgacaggtatagcagacatcgaaagctcaaattat
ggtgttgttgcgttattaaatgatttacacaaaggacgaaacagcggtgaggtctggaag
gcattcattgatattaccgcgctattaatggtctttttgtactcacaggtgtgtgctta
attttaccgaaaaaaagaacactgatgacatcaatgaagtggatggcatttggctcgtca
atcacgcttgcattgtattttattgccgtaccttaacccttttaatcttgtatccataag
gacatcacaatggcttcactctctacctatttacgtcattcaattgctgcaatggtctta
actacacctcttttggctaattcagcggctatttctgagagtgcgaaagtcgatgttgct
tttgagcttccgacgattacgaccaccatgtacgcgcgtccttatgtggcggtgtggatt
gaaaacagcgaacgtaaatcagtaaaaacgattgagctgtgggtaggtaaagatgagtgg
ctaaaagatctaagaagttggtggcgcaaagtcggtcgttacgatcgtgaattagttgat
gctgttacttctgcaactcgaccagcaggaaaatataaattcagttgggatggaaaagac
gactcaggccagacgctaccacaaggggattatacgttacatatcgaagtggttcgtgag
catggtggtcgtaattatttacgccaaaaaatgaaattgaccgacactgatatgtcatac
caacttgctccaactgaagagactggcgtaatcacaatcgaatataaaaattaatctaga
tgtccatttatcagggaattaaacgaattatgaaaacgacaataatgaacaataaagtaa
aggccgcaatattaagtggcgtaatggcagcaggtttactggttgctacgactgcgaacg
ctcaccctcgttgggtattgccttctcattttacggtttctaaagaaggtggtgattggc
ttacgtttgatgttaccgcttcacacggtacgtttgtatttgataaaccggcaggcagtg
atcaagcacaagtagtgatgcctgatggtcgtactgagcgtccaaactttgtcgttaaag
gcaaacgtcgttcaatgtttgacttctattttgaagaagaaggcacgcataaagttcgaa
ttaataacgagccaacgtattacacccaatataaagcaggccgtcgcgatacggtgaaat
ggattaaagcaactaaggcggatcgtgcattagtgcttcctgagaaagcgcgtgatgtgg
taagccaggtaagttttactcgtgctgaaagttacgtgacagtaggcaagcctacaacag

aagcaatgaagcttgataatacgctgttagagcttaagccattaacacacccatcagata
ttcttgaaggtgagccagttacttttcagttcttctttaatggtaaacctcaagcaggtg
taacggcggatattactcgtgaaggaacgttataccgcaatcatcaagagcaaatagacg
tagtgagtgataaagagggaaaaatcacttttactcctgaaactgcgggtcgttacctga
tgaaatcaaattataaaggtgagctgaaagaccacccacttgcagaacaagcaagcgtaa
atgttcacttaacgtttgaagcactacttcaataatgcagttagcaccactttccgtaaa
ggtgggtggtgctttttatcgagatttattcatgactaaagtaaacctgaaaacccttt
ttgtatcttctcttctttttttcttctgcgagttttgcgcattttcctttaatgagctgtt
ggtttgaatcgaatcaagtgctgtgcgaaacaggctacagtgatggcagcaaagccgttg
attatgatgtgaatatgtacgattacgatgacaacctgattgcaaaagtcaccacagata
aacgttcaattgcggaatttgcgaatcctgaaacagacttttatttggtgtttgattctg
gacatgaaagcccagttgaagtcgatgttgtcgagatctcagagaaatgattattcaatc
tgtcactccagataatggaagcgcagaaggcaaatggatcgcactcattattgggtgcat
tttgcttattgcggtggtcgggttgccatatcatcaatcgaacgatgaaggcgtaaagct
tgaatcacaccaagttcttatcaccgatttagcgcaagacgatttagctatgattgctga
gctacgtttagcctatgaagaaatttacgatattcaggctgaatcaacagaagaaaatgc
atggccagagattgtttacttagcggatgaatggttaccgccttttgtgcaagaccaaag
ctgggaacgtcgcggaaaacaccaatggcagcatttaggttctggagtttatctgggttt
aaaagtgaaaaaagaaggcgctgcgtcattcatttttaatgtgaaaggcgatgatattga
gctttggctgcaccctgaattgatgtcgacctctatctcctctgattcaatctcttggaa
tgaaagcgaactcattcaacagcaatgggtacaagttgtatttcctgctcatgatactca
tttaaattgataccaattaaagataaataataatgatgaaatctattttatacgtacgtt
cttttgttctgcttagcttactctctttgttttctgcgtcggcgatgtcggcagaaaaac
ctttggtgattggcattaccttacaaccttattacagctatgtgaaagcagtagtaggtg
ataaagcagaaattcttcctttggttgatgcagggtttaatccccacaattatttacccc
aaccaaacgatttaaaacgtttaaataaaatggatgtgattgttgttaatggcattggtc
atgatgactttgcgctaaaagtgattgcggcggcaaacaacagcgagttaatggtattag
aagccaataaagacgttccgttgttaccagcaatggggcaatctgtaggtaaaggggcgg
tgaacccacacacatttgttggtctatctacaacaattcaaaaggtttacaccattgcca
atgaactatctcgaattgaccgtgaaaatgcgtcgttctaccgtaaaaatgcacgtaaat
acgcgaaaaaattccgattaatgaagcgtgatgccatgttatctctaggtgaattagata
ccgcaggaatgaaagtagcaacaacgcacaatgcgtatggttatattcttcaagaatttg
gtgttgatgtcgctgcagttattgaacctgctcacggcgtagaaccaagcgcgagccaac
tgcaagggacgattgaaaagatccgtgaatctgacattgatgttttatttatgaattga
acatgccaaaccgttttgtcgacacgattgaagaggcgacaggcgttcagttgtatcgct
tttcgcacatgacgcacggtccttatgaagacaataaagttgaaattgaaatgaaatcaa
atttagacaccctaatcgaagccatgaaatttgccgctgataaggacgctaaataatgtc
aggaccagcaatttcgttaaccaatgttgggctgacttatggtgaaaataccattttaga
aggggtgaaccaacgctttgaagccggtcaatgccatgtggtgatggggcctaatggcgg
aggtaaaacctcgttattgcgttcaattttaggattaaccgcgtttcaaggtgaaattaa
tattgaatggccagagcagaaaggggtgattggttatgtgccacaaaaagccgtttttga
atctagcttgccactaacggtgaaagatttcattttattgaatcaaacgcgagtgcctct
gtttttaagtgcgttatggcgtaagaaaagcagccaagaagcatttgcaatggcgcagct
agagcgtgttggtatgtctaaccgtgccgatcgtcgcatggggcaactttctggtggtga
gcaacaacgtgtgttattcgcgcaagcgttactggataatcccgatttacttgttcttga
tgaaccaacaacaggaatggatgagcacggggttcgttatcttgaatccttaatgcatga
atttatcgcgcaaggtaaaacggtattagccgttcaccacgatgtgaccgctgttcgccg
tcttaaagcacatgttcatgttatcaatcgtaatttggttgattctggttgttattcaga
agtattaaaacctgagaaaattgaacgtttattcaatcattatacgggtgcaagtgctca
gtacgatgcatcaaaaaccaacaatggagaggctgcgtaatggaactgtttcgtcagtgg
gcgttatatggcgtcgatcaaggtttcttaagtgacagttttcgtatgcctttatggtt
aatgctattgttgccgctatccttttagggccattgttaggtggtttggggacgttagtg
attgcaaagcgtttagccttttctctgaagcggttggtcatgccgccttaacgggtatc
gcgttaggtgtgttattgggagagccagcagagaacccaatcatcggcttatttagtttc
tgtatgatctttgctttgttactgcattttgtaagaaaccgaaccaatgtgccttacgat
acgttagttggggtgttccttgccttggcattggctgtgggcgctgcattactgatgtac
gtagcacgtaaaattaatattcacatgctagagaacgtattgtttggttctatcttaacg
gttacggaccaagacattgcgattttagcaggcagttgcttattgattgtcgtggtatta
attccgatgtttaatcgtattttattaacctgtatcagccctgatattgctagagttcgt
ggctacaaaaccaatttttatgattacctatttgtgatgatgattacccttgtgacgatt
gcatcggtaaaaattgtcggtgcggttttagttggcgcgttattgctcatacctggtgca

acggctcgattgttaactaagaatatgggcagttttgtcttgatgtctgcactattagcg
acaatatcgtgcttgattggcaccgtattaccgatggaattagaaatgccagtgccttct
ggggctgcgattattattgtttcagctttcttcttttacttgcgaccacctaccgaatt
cttcgtaagaattaacagcgtagacctcaataataatcgaagaatataggataagaaatg
aaacagaattcaatgaaatggatagcgattgctagtatatgtgtagcctcactatttagc
atcaaaacagtaatagcggcagatgtattaacggcaacccccgtgacgtatatgctatca
agcgagttattaaaagaaacgtcattaacgacaaagtatttaccgccaaatcgttatggg
attgagcgtttacccaattggtttgccaataaaggggcgaaaaccgttcaagactcagta
aaaacagcccaagcggtggtgacgattggggcggtatgggctgatgatcctctgtatgtg
aacgcaagatacagcaatattcagattattgaagtcgatgcctctcaagcgataactcct
cgtgcacaaggtgtggcagcgttacgtttatctgatggaacgacctcaaaatacgcatgg
ttaaacccaacaaacttaatcagaatggtagcgattgtgagtgatgattttcaacgtatc
tggcctgataaatcagagcagatccaaaaaaaccaagagcaatttatgttgtcggtacgt
caattaatcaaccagcagcaaaaagtgattttagagaatgatatcgactcggttattttg
ttgtctgagcagttagaagactttgcctctggcaatcaattgtttgtggttgatcgctta
acaaaaccagctcttgaatggacagaagatgataaagcacagttgaaagacttaattgaa
gaagatgagtcagtttggattatcacaactcaaaaaccaaacaagcagatgcaagcattt
ctgccaatggaacgtgtattggttgttgatgcggttgatcgttggggcagagcaggaata
aacagtaaagcgccattacagagatggcaattagcgttatagccttatctagattgggcg
agtttactctaaaaataaaagcagtaggggcttattgttacgatctagcaccgactgctt
ttttttatggttttatttagtgtagtttaaataccaagggattatgacgatgttttccct
aaagataggcggattagattattatcgtatacataatttaagtaagtcgtatttttgat
taacatgagagtcagaaaagcactaaactatccatgttcttatggtttttaacaaatagg
gttgtgtcgtttttgtgaggatcgaggaattgttttatttattattttgaatcgctgt
cattgtttataaaaaggaaaaactatgtcttatgtagaaccgattaaaacacgtaaacat
ctgaaacatgcagaaaaatatcttaaaaaaaatcatgaccttgcttttacccttgtttgg
aatattggtatagaaagtggattaaggattagtgatattcttaaattacagtatcgcaat
attaatttatcaaatggacagtgtaaaatcattgaaagtaaaggcacattggctcgtaga
gcaaaagcaaagatgagggtgttaaaaaaagtaaaggaagaattgctttttcattatcaa
gctaggggttgtccaaaaaaatgagtattatctatattaccccgcctcagctgattgtt
tcttttatccctaaatcttgggaatcgacagtgaatgaaagagtaaagcacgcaatagac
tctgcttcacctgtgactcgaactttctatttgagtgaatctattctaataactcttaaa
caacgaaaacaagagtataaagagctagctaatgactatatttttaatcgtcagacactc
tcaagtaatcgtgcaaaaggaggggagggactgttgactcgacaagcttgttggtatgtt
ttttctaaattgacagcgatattgaaagagattggtgttgatgtaaaggtagggtgtcat
acattaaggaagagcttcgctcgacatttgtattttgctacgggtaaagatattggattg
ttgatgaccgttattggacatcaatctgaaagcatgagtttacgctatataggcttgtcc
aaagatgagactgagttggctcaaaaaaaattaattgtctatttatcgagaaaaactcgc
atatgataactaggccgcaaataatgaccttaatattaac