

Identification of Bacterial Unknowns by rRNA Sequence Similarity

Resource Type: Curriculum: Classroom

Publication Date: 8/2/1999

Authors

Richard Roller

Department of Microbiology

The University of Iowa

Iowa City, Iowa

USA

Email: richard-roller@uiowa.edu

Abstract

This exercise allows students to share a common and exciting scientific experience; use of a sequence database to analyze an unknown sequence. Specifically, the students learn to use database searching for identification of an organism and its close relatives. Students are provided via e-mail with ribosomal RNA sequences from organisms unknown to the students. They are directed to access the GenBank nucleotide sequence database, search it using their sequences, and then analyze the data to identify the organism from which the sequence was derived. Students are also asked to identify close relatives of the unknown and evaluate the degree of divergence between their unknown and its relatives.

Activity

Invitation for User Feedback. If you have used the activity and would like to provide feedback, please send an e-mail to MicrobeLibrary@asmusa.org. Feedback can include ideas which complement the activity and new approaches for implementing the activity. Your comments will be added to the activity under a separate section labeled "Feedback." Comments may be edited.

Editor's Note (2008): This Curriculum Resource was published prior to establishment of current criteria of submission, and as such, does not contain all criteria required of current publications. However, the Editorial Committee felt that the activity itself remained worthwhile and relevant, and encourages potential users to contact the authors for clarification as needed. If you do update this activity for use with your students, and are interested in updating the resource for distribution in the library, please contact ASM at MicrobeLibrary@asmusa.org.

INTRODUCTION

Time Required.

30 minutes to 1 hour, depending upon student experience with computers and the Internet.

Pedagogical Function.

The purpose of this activity is to introduce students to the concept and mechanics of sequence database searching, to reinforce lecture- or reading-based introductions to organism identification and classification based on nucleotide sequence divergence, and to provide a concrete example of the use of PCR in the lab.

Background.

Students should have basic understanding of mechanisms of DNA replication (sufficient to understand the conceptual basis for PCR) and have been introduced to the principle of the PCR itself. Students should also have basic familiarity with ribosome structure and function (sufficient to understand that ribosomes have RNA and protein components and that all cellular organisms encode rRNAs). Students should have received an introductory lecture or reading to introduce them to the use of sequence divergence for phylogenetic classification and the particular utility of 16S rRNA sequences for this purpose. Students may also have greater appreciation for the exercise if they have first been through an exercise for identification of an unknown using biochemical and functional tests.

PROCEDURE

Materials.

Computer with access to the World Wide Web

Electronic copies of two unknown rRNA sequences. These can be provided on disk or sent to the student in the text of an e-mail message. We have used the following organisms as unknowns:

Bacillus megaterium
Bdellovibrio bacteriovorus
Bordetella pertussis
Borrelia burgdorferi
Chlamydia trachomatis
Clostridium botulinum
Escherichia coli
Helicobacter pylori
Legionella pneumophila

Listeria monocytogenes
Mycobacterium leprae
Neisseria gonorrhoea
Rickettsia rickettsii
Salmonella typhi
Staphylococcus aureus
Streptococcus pyogenes
Thermus aquaticus
Treponema pallidum
Vibrio cholerae
Yersinia pestis

These sequences or those from other organisms of interest can be collected from GenBank by searching for the organism name as described in the Possible Modifications section below. Alternatively, sequences for the organisms named above are available in the appendix for cutting and pasting into the students' e-mail messages.

Safety Issues. Not applicable.

Instructor Version.

Students are provided with an activity handout (see below) and with two unknown rRNA sequences in an e-mail message (text reproduced below). Most students required no assistance in accessing the Internet and cutting and pasting sequences into the database query form, but individual assistance was provided as required.

Activity Handout.

E-mail message text:

Student name,

Below this note you will find your two unknown 16S rRNA sequences. You may want to first save the content of this message to a disk before trying anything else. Then use the copy and paste functions with your e-mail reader or a word processor to put the sequences into the box on the database search page. When you have identified your unknowns, you might find it interesting to do an Internet search using the names of the organisms. Almost all of them are of importance for human health or for biotechnology. Good luck!

Sequence #1

Sequence #2
<Insert sequence 2>

ASSESSMENT and OUTCOMES

Suggestions for Assessment.

The activity handout includes a set of questions that students should be able to answer if they understand how to interpret the search output.

Problems and Caveats.

Preparation time may be as long as a full day if the instructor decides to acquire all of his/her own rRNA sequence examples. Care must be taken to choose only full-length rRNA sequences since smaller sequence fragments (of which there are many in the database) may provide perfect matches with nonidentical organisms. With a preexisting collection of sequences in hand, preparation is confined to sending e-mail messages. Depending upon class size, this may require between 30 minutes and 2 hours.

SUPPLEMENTARY MATERIALS

Possible Modifications. (Optional)

Different rRNA sequences may be chosen at the instructor's discretion. Individual sequences should be test run by the instructor. Some 16S rRNAs will return only 100% identical sequences in the top 100 matches, providing no information about close relatives.

NCBI taxonomy browser

rRNA sequences may be found by searching GenBank for the desired organism name. I will outline the procedure using *Bacillus anthracis* as an example. Direct your Web browser to the National Center for Biotechnology Information (NCBI) taxonomy browser at: <http://www.ncbi.nlm.nih.gov/Taxonomy/tax.html>. Roughly in the middle of the Web page, there is a query box for entering a search. Click in the query box and enter the name of the organism of interest (in this case, *Bacillus anthracis*), then click on the "Start Search" button. The program will return a page that indicates the number of nucleotide (in this case 54) and protein (in this case 228) sequences from this organism deposited in GenBank. If the list of DNA sequences is very short, it is most convenient to browse for the 16S rRNA sequences simply by clicking on the "Get Sequences" button. If the list is long, it is most convenient to select "Yes" next to the "Refine Query in Entrez?" button and then click on the "Get Sequences" button. The program will return a page in which you can add terms to the query. Click in the "Enter Terms" box and type in "16S". Using the pull down menu just to the right of "Search Field" select TextWord. Then click the "Search" button. You will be returned to an updated version of the same page. Near the top of the page, click on the "Retrieve 4 Documents" button to get a list of the matching sequences. For *Bacillus anthracis*, four sequences are returned, the first of which is the 16S rRNA sequence (the other three are intergenic spacer sequences from the region of the 16S gene). The sequences can each be viewed in a number of formats. To obtain a sequence for further use, click on the link for FASTA report. The program will return a page in which the sequence is presented in a simple text format. This sequence can then be selected and copied and pasted to a word processing document.

Note that the text of the e-mail message suggests an Internet browser search on the name of the identified organism. We did not make this mandatory in our first experience with this exercise, but I would do so in the future. This represents an

opportunity to introduce students to the variety of information (good and bad) available on the Internet about organisms and diseases.

Instructors may also choose to include the source of the amplified DNA in each student's e-mail message. For example, students receiving the *Legionella pneumophila* sequence might be told that the organism from which the DNA was amplified was obtained from the sputum of a patient suffering from fever and severe pneumonia. Searching out appropriate sources for each specimen and pasting these into the e-mail messages will add to preparation time, but this additional information will help reinforce the basic concept that rRNA sequence comparisons are a means of identifying organisms from "real-world" unknown specimens.

References.[GenBank Overview](#)

Information about GenBank is derived from the National Center for Biotechnology Information GenBank Overview page at: <http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html>.

Acknowledgments.

The author is grateful to Dr. Marcia Cordts for suggesting submission of this activity and for critical reading of the manuscript.

Appendix.[16S rRNA sequences for organisms listed in Activity section](#)

Contents of
this page

 Curriculum Resources

[Bacillus megaterium](#)
[Bdellovibrio bacteriovorus](#)
[Bordetella pertussis](#)
[Borrelia burgdorferi](#)
[Chlamydia trachomatis](#)
[Clostridium botulinum](#)
[Escherichia coli](#)
[Helicobacter pylori](#)
[Legionella pneumophila](#)
[Listeria monocytogenes](#)
[Mycobacterium leprae](#)
[Neisseria gonorrhoea](#)
[Rickettsia rickettsii](#)
[Salmonella typhi](#)
[Staphylococcus aureus](#)
[Streptococcus pyogenes](#)
[Thermus aquaticus](#)
[Treponema pallidum](#)
[Vibrio cholerae](#)
[Yersinia pestis](#)

Appendix. 16S rRNA sequences for organisms listed in Activity section

Bacillus megaterium

GATGAACGCTGGCGGCGTGCCTAATACATGCAAGTCGAGCGAACTGATTAGAA
 G CTTGCTTCTATGACGTTAGCGGGCGGACGGGTGAGTAACACGTGGGCAACCTG
 CC TGTAAGACTGGGATAACTTCGGGAAACCGAGGCTAATACCGGATAGGATCT
 TCTC CTTCATGGGAGATGATTGAAAGATGGTTTCGGCTATCACTTACAGATGGG
 CCCGC GGTGCATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCAACGATGC
 GTAGCC GACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAG
 ACTCCTAC GGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAAAGTCTGA
 CGGAGCAACG CCGCGTGAGTGATGAAGGCTTTCGGGTTCGTAAACTCTGTTGT
 TAGGGAAGAACA AGTACAAGAGTAACTGCTTGTACCTTGACGGTACCTAACCA
 GAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAA
 GCGTTATCCGGAATTATT GGGCGTAAAGCGCGCGCAGGCGGTTTCTTAAAGTCT
 GATGTGAAAGCCCACGGCT CAACCGTGGAGGGTTCATTGGAACTGGGGAACT
 TGAGTGCAGAAGAGAAAAGCG GAATTCCACGTGTAGCGGTGAAATGCGTAGA
 GATGTGGAGGAACACCAGTGGCGAAGGCGGCTTTTTGGTCTGTAAGTACGCT
 GAGGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATAACCCTGGTAGTCCACG
 CCGTAAACGATGAGTGCTAAGTGTTAGAGGGTTTCCGCCCTTTAGTGCTGCAG
 CTAACGCATTAAGCACTCCGCCCTGGGGAGTACGGTTCGCAAGACTGAAACTCA
 AAGGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGGTTTAATTCGAA
 GCAACGCGAAGAACCTTACCAGGTCTTGACATCCTCTGACAACCTCTAGAGAT
 AGAGCGTTCCTTTCGGGGGACAGAGTGACAGGTGGTGCATGGTTGTCGTC
 GCTCGTGTCTGAGATGTTGGGTTAAGTCCCAGCAACGAGCGCAACCCTTGATC
 TTAGTTGCCAGCATTTAGTTGGGCACTCTAAGGTGACTGCCGGTGACAAACC
 GGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTATGACCTGGGCTA
 CACACGTGCTACAATGGATGGTACAAAGGGCTGCAAGACCGCGAGGTCAA
 GCCAATCCCATAAAACCATTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTA
 CATGAAGCTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAATAC
 GTTCCCGGGCCTTGTACACACCGCCCGTCACACCACGAGAGTTTGTAAACAC
 CCGAAGTCGGTGGAGTAACCGTAAGGACGTAGCCGCCTAAGGTGGGACAG
 ATGATTGGGGTGAAGTC GTAACAA

Bdellovibrio bacteriovorus

TAAACTGGAGAGTTTGAATCTGGCTCAGAACNNACGCTGGCGGCGTGCCTAATACA
 TGCAAGTCGACGNNAGCTTTCGGGTGAGTACTAGTGTNGCGCACGGGTGAGGAA
 CGCGTGGATAATCTGCCTTAGAGTGGGGGATAACTAGTCGAAAGATTAGCTAATACC
 GCATAAGACCACAGGAGCTGCGGCTCTAGGGGTCAAAGGTTTTTTCGCTCTAAGATG
 AGTCCGCGTAAGATTAGCTAGTTGGTGAGGTAACGGCTCACCAAGGCGACGATCTT
 TAACTGGTCTGAGAGGATGATCAGTCACACTGGAAGTACGACACGGTCCAGACTCC
 TACGGGAGGCAGCAGTAGGGAATATTGCACAATGGAGGAACTCTGATGCAGCGA
 CGCCGCGTGAGTGATGAAGGCCTTCGGGTTCGTAAGCTCTGTTCGACAGGGGAATAA
 CACAATGAATGTACCCTGTAAGAAAGGATCGGCTAACTTCGTNCCANNAGCCGCGN
 NNAGACGAGNGATCCTAGCGTTGTTTCGGAATTATTGGGCGTAAAGCGGATGTAGG
 TGGCTTTGTAAGTCAGATGTGAAAGCCCAGGGCTCAACCCTGGTAGTGCATTTGAT

ACTGCGAAGCTTGAGTGTCGGAGAGGTTACTAGAATTGTTGGTGTAGTGGTGAAAT
 ACGTAGATATCAACAGGAATACCGGAGGCGAAGGCGGGTAAGTGGCCGAACACTG
 AACTGAGATCCGNAAGCGTGGGGATCAAACAGGATTAGATACCCTGGTAGTCCAC
 GCCGTAAACGATGGATACTTGTGTTGGAGGTATTGACCCCTTCAGTGACGAAGCT
 AACGCGTTAAGTATCCCGCCTGGGGAGTACGGTCGCAAGATTAAACTCAAAGAAA
 TTGNNNNNNNNNNNGCACAAGNNNNNNNNNNNNNNNGTTTAATTCGATGCAACGC
 GAAGAACCTTACCTAGGCTTGACATGTAAGTGGAAAGTGGCAGAAATGTCGTCGCC
 GCAAGGGTTCGGTACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTGTCGTGAGATGT
 TGGGTTAAGTCCCGCAACGAGCGCAACCCCTGCATTTAGTTGCCAGCATTTCAGTTG
 GGCACCTTAGATGGACTGCCGGTGTAAACCGGAGGAAGGTGGGGATGACGTCAA
 GTCCCTCATGGCCCTTATGCCTAGGGCTACACACGTGCTACAATGGTAGTCACAGAG
 CGAAGCTAAGCCGCGAGGTAGAGCAAATCGCTTAAAAGCTATCTAAGTTCAGATTG
 ATCTCTGCAACTCGAGATCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGAATG
 CCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGAAAGT
 TGGCTGTACCAGAAGTCGNTGCGCTAACCGCAAGGAGGCAGGCGCCCAAGGTAT
 GGTNNACGANNNNNNNNNNNNNGTAACAAGNNNNNNNNNNNNNGAACCTGNNN
 NNNGATCACCTCCTTTCT

Bordetella pertussis

AACTGAAGAGTTTGATCCTGGCTCAGATTGAACGCTGGCGGGATGCTTTACACATG
 CAAGTCGGACGGCAGCACGGGCTTCGGCCTNGTGGCGAGTGGCGAACGGGTGAG
 TAATGTATCGGAACGTGCCAGTAGCGGGGGATAACTACGCGAAAGCGTAGCTAAT
 ACCGCATACGCCCTACGGGGGAAAGCGGGGGACCTTCGGGCCTCGCACTATTGGA
 GCGGCCGATATCGGATTAGCTNGTTGGTGGGGTAACGGCCTACCAAGGCGACGAT
 CCGTAGCTGGTTTGAGAGGACGACCAGCCACACTGGGACTGAGACACGGCCCAGN
 CTCTACGGGAGGCAGCAGTGGGGAATTTTGGACAATGGGGGCAACCCCTGATCCA
 GCCATCCCGCGTGTGCGATGAAGGCCTTCGGGTTGTAAAGCACTTTTGGCAGGAAA
 GAAACGGCACGGGCTAATATCCTGTGCAACTGACGGTACCTGCAGAATAAGCACCG
 GCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGCAAGCGTTAATCGGAAT
 TACTGGGCGTAAAGCGTGCAGGCGGTTTCGGAAAGAAAGATGTGAAATCCCAGG
 GCTTAACCTTGGAAGTGCATTTTAACTACCGGGCTAGAGTGTGTCAGAGGGAGGT
 GGAATTCCGCGTGTAGCAGTGAATGCGTAGATATGCGGAGGAACACCGATGGCG
 AAGGCAGCCTCCTGGGATAACACTGACGCTCATGCACGAAAGTGTGGGGAGCAA
 CAGGATTAGATACCCTGGTAGTCCACGCCCTAACGATGTCAACTAGCTGTTGGGG
 CCTTCGGGCCTTGGTAGCGCAGCTAACGCGTGAAGTTGACCGCCTGGGGAGTACG
 GTCGCAAGATTAAACTCAAAGGAATTGACGGGGACCCGCACAAGCGGTGGATGAT
 GTGGATTAATTCGATGCAACGCGAAAAACCTTACCTACCCTTGACATGTCTGGAATC
 CCGAAGAGATTTGGGAGTGCTCGCAAGAGAACCAGAACACAGGTGCTGCATGGCTG
 TCGTCAGCTCGTGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTTGT
 CATTAGTTGCTACGAAAGGGCACTCTAATGAGACTGCCGGTGACAAACCGGAGGAA
 GGTGGGGATGACGTGAAGTCCCTCATGGCCCTTATGGGTAGGGCTTCACACGTCATA
 CAATGGTCGGGACAGAGGGTTGNCAACCCGCGAGGGGGAGCCAATCCCAGAAAC
 CCGGTGCTNGTCCGGATCGCAGTCTGCAACTCGACTGCGTGAAGTCGGAATCGCT
 AGTAATCGCGGATCAGCATGTCGCGGTGAATACGTTCCCGGGTCTTGTACACACCG
 CNCGTACACCATGGGAGTGGGTTTTACCAGAAGTAGTTAGCCTAACCGCAAGGGG
 GCGGATTACCACGGTAGGA

Borrelia burgdorferi

TGATCCTGGCTTAGAACTAACGCTGGCAGTGCGTCTTAAGCATGCAAGTCAAACGG
GATGTAGTAATACATTCAGTGGCGAACGGGTGAGTAACGCGTGGATGATCTACCTA
CGAGATGGGGATAACTATTAGAAATAGTAGCTAATACCGAATAAGGTCAATTAATTT
GTAAAGTGATGAAAGGAAGCCTTTAAAGCTTCGCTTGTAGATGAGTCTGCGTCTTAT
TAGCTAGTTGGTAGGGTAAATGCCTACCAAGGCGATGATAAGTAACCGGCCTGAGA
GGGTGAACGGTCACACTGGAAGTGAAGTACGGTCCAGACTCCTACGGGAGGCAGC
AGCTAAGAATCTTCCGCAATGGGCGAAAGCCTGACGGAGCGACTGCGTGAATGA
AGAAGGTCGAAAGATTGTAAAATTTCTTTTATAAATGAGGAATAAGCTTTGTAGGAAAT
GACAAAGTGATGACGTTAATTTATGAATAAGCCCCGGCTAATTACGTGCCAGCAGCC
GCGGTAATACGTAAGGGGCGAGCGTTGTTCCGGGATTATTGGGCGTAAAGGGTGAG
TAGGCCGATATATAAGTCTATGCATAAAATACCACAGCTCAACTGTGGACCTATGTT
GGAACTATATGTCTAGAGTCTGATAGAGGAAGTTAGAATTTCTGGTGTAAAGGGTGG
AATCTGTTGATATCAGAAAGAATACCGGAGGGCGAAGGCGAACTTCTGGGTCAAGACT
GACGCTGAGTCACGAAAGCGTAGGGAGCAAACAGGATTAGATACCCTGGTAGTCTA
CGCTGTAAACGATGCACACTTGGTGTAACTAAAAGTTAGTACCGAAGCTAACGTGT
TAAGTGTGCCGCTGGGGAGTATGCTCGCAAGAGTGAAACTCAAAGGAATTGACGG
GGGCCCCGACAAGCGGTGGAGCATGTGGTTTAATTCGATGATACGCGAGGAACCTT
ACCAGGGCTTGACATATATAGGATATAGTTAGAGATAATTATTCCCCGTTTGGGGCCT
ATATACAGGTGCTGCATGGTTGTCGTCAGCTCGTGCTGTGAGGTGTTGGGTAAAGTC
CCGCAACGAGCGCAACCCTTGTTATCTGTTACCAGCATGTAATGGTGGGGACTCAGA
TAAGACTGCCGGTGATAAGTCGGAGGAAGGTGAGGATGACGTCAAATCATCATGGC
CCTTATGTCTGGGCTACACACGTGCTACAATGGCCTGTACAAAGCGAAGCGAAGC
AGTGATGTGAAGCAAACGCATAAAGCAGGTCTCAGTCCGGATTGAAGTCTGAAAC
TCGACTTCATGAAGTTGGAATCGCTAGTAATCGTATATCAGAATGATACGGTGAATAC
GTTCTCGGGCCTTGTAACACACCGCCCGTCACACCACCCGAGTTGAGGATACCCGAA
GCTATTATTCTAACCCGTAAGGGAGGAAGGTATTTAAGGTATGTTTAGTGAGGGGGG
TGAAGTCGTAACAAGGTAGCCGTAAGTGGAAAGTGCGGCTGGATCAC

Chlamydia trachomatis

CTGAGAATTTGATCTTGGTTCAGATTGAACGCTGGCAGGCGTGGATGAGGCATGCAA
GTCGAACGGAACAATTGCTTCGGTGATTGTTTAGTGGCGGAAGGGTTAGTAATGCA
TAGATAATTTGTCCCTAACTTGGGAATAACGGTTGGAAACGGCCGCTAATACCGAAT
GTGGCGTAGATTAGGCATCTAAATTACGTAAAGAAGGGGATCTTCGGACCTTTCGG
TTAAGGGAGAGTCTATGTGATATCAGCTAGTTGGTGGGGTAAAGGCCTACCAAGGC
TATGACGTCTAGGCGGATTGAGAGATTGGCCGCCAACACTGGGACTGAGACTGTC
CCAGACTCCTACGGGAGGCTGCAGTCGAGAATCTTTCGCAATGGACGGAAGTCTGA
CGAAGCGACGCCGCGTGTGTGATGAAGGCTCTAGGGTTGTAAAGCACTTTCGCTTG
GGAATAAGAGAAGACGGTTAATACCCGTTGGATTTGAGCGTACCAGGTAAAGAAGCA
CCGGCTAACTCCGTGCCAGCAGCTGCGGTAATACGGAGGGTGCTAGCGTTAATCGG
ATTTATTGGGCGTAAAGGGCGTGTAGGCGGAAAGGTAAAGTTAGTTGTCAAATCTCGG
GGCTCAACCCCGAATCGGCATCTAAAACCTATTTTTCTAGAGGGTAGATGGAGAAAAG
GGAATTTACGTGTAGCGGTGAAATGCGTAGATATGTGGAAGAACACCAGTGGCGA
AGGCGCTTTTCTAATTTATACCTGACGCTAAGGCGCGAAAGCAAGGGGAGCAAACAG
GATTAGATACCCTGGTAGTCCTTGCCGTAACGATGCATACTTGATGTGGATGGTCTC
AACCCCATCCGTGTGCGGAGCTAACGCGTTAAGTATGCCGCTGAGGAGTACACTCG
CAAGGGTGAAACTCAAAGAATTGACGGGGGCCGACAAGCAGTGGAGCATGTG
GTTAATTCGATGCAACGCGAAGGACCTTACCTGGGTTTGACATGCATATGACCCGG
GCAGAAATGTCGTTTTCCGCAAGGACATATGCACAGGTGCTGCATGGCTGTCGTCAG

CTCGTGCCGTGAGGTGTTGGGTAAAGTCCCGCAACGAGCGCAACCCTTATCGTTAGT
 TGCCAGCACTTAGGGTGGGAACCTAACGAGACTGCCTGGGTAAACCAGGAGGAAG
 GCGAGGATGACGTCAAGTCAGCATGGCCCTTATGCCAGGGCGACACACGTGCTAC
 AATGGCCAGTACAGAAGGTAGCAAGATCGCGAGATGGAGCAAATCCTCAAAGCTGG
 CCCCAGTTCGGATTGTAGTCTGCAACTCGACTACATGAAGTCGGAATTGCTAGTAAT
 GGCGTGTACAGCCATAACGCCGTGAATACGTTCCCGGGCCTTGTACACACCGGCCGT
 CACATCATGGGAGTTGGTTTTACCTTAAGTCGTTGACTCAACCCGCAAGGGAGAGAG
 GCGCCCAAGGTGAGGCTGATGACTGGGATGAAGTCGTAACAAGGTAGCCCTACCGG
 AAGGTGGGGCTGGATCACCTCCTTT

Clostridium botulinum

GCTGGCGGCGTGCTTAACACATGCAAGTCGAGCGATGAAGCTTCCTTCGGGGAGTG
 GATTAGCGGCGGACGGGTGAGTAACACGTGGGTAACTGCCTCAAAGTGGGGGATA
 GCCTTCCGAAAGGAAGATTAATACCGCATAATATAAGAGAATCGCATGATTTTCTTA
 TCAAAGATTTATTGCTTTGAGATGGACCCGCGGCGCATTAGCTAGTTGGTAAGGTAA
 CGGCTTACCAAGGCAACGATGCGTAGCCGACCTGAGAGGGTGATCGGCCACCATTG
 GAACTGAGACACGGTCCAGACTCCTACGGGGAGGCAGCAGTGGGGAATATTGCGCA
 ATGGGGGAAACCCTGACGCAGCAACGCCGTTCGTGGGTGATGAAGGTCTTCGGATTGT
 AAAGCCCTGTTTTCTAGGACGATAATGACGGTACTAGAGGAGGAAGCCACGGCTAA
 CTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCGAGCGTTGTCCGGATTTACTG
 GGCGTAAAGGGTTCGTAGGCGGATGTTAAGTGGGATGTGAAATCCCCGGGCTTAA
 CCTGGGGGCTGCATTCCAACTGGATATCTAGAGTGCAGGAGAGGAAAGCGGAATT
 CCTAGTGTAGCGGTGAAATGCGTAGAGATTAGGAAGAACACCAGTGBCGAAGGCGG
 CTTTCTGGACTGTAAGTACGCTGAGGCACGAAAGCGTGGGTAGCAMACAGGATTA
 GATACCCTGGTAGNCCACGCCGTAACGATGGATACTAGGTGTAGGGGGTATCAAC
 TCCCCCTGTGCCGCAGTTAACACAATAAGTATCCCGCCTGGGGAGTNCGGTCCGAA
 GATTAAACTCAAAGGAATTGACGGGGGCCCGCACAAAGCAGCGGAGCATGTGGTTT
 AATTCGAAGCAACGCGAAGAACCCTTACCTGGACTTGACATCCCTTGCATAGCCTAG
 AGATAGGTGAAGCCCTTCGGGGCAAGGAGACAGGTGGTGCATGGTTGTTCGTCAGCT
 CGTGTTCGTGAGATGTTAGGTTAAGTCCTGCAACGAGCGCAACCCTTGTATTAGTTG
 CTACCATTAAGTTGAGCACTCTAATGGAGACTGCCTGGGTAACCAGGAGGAAGGTG
 GGGATGACGTCAAATCATCATGCCCTTATGTCCAGGGCTACACACGTGCTACAAT
 GGTAGGTACAATAAGACGCAAGACCGTGAGGTGGAGCAAACTTATAAAACCTATC
 TCAGTTCGGATTGTAGGCTGCAACTCGCCTACATGAAGCTGGAGTTGCTAGTAATCG
 CGAATCAGAATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCAC
 ACCATGAGAGCTGGTAACACCCGAAGTCCGTGAGGTAACCGTAAGGAGCCAGCGG
 CCGAAGGTGGGATTAGTGATTGGGGTGAAGTCGTAACAAGGT

Escherichia coli

AAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACAT
 GCAAGTCGAACGGTAACAGGAAGAAGCTTGCTTCTTTGCTGACGAGTGGCGGACG
 GGTGAGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGTA
 GCTAATACCGCATAACGTCGCAAGACCAAAGAGGGGGACCTTCGGGCCCTTTGCCA
 TCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGGCG
 ACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAAGTACTGAGACACGGT
 CCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTG
 ATGCAGCCATGCCGCGTGTATGAAGAAGGCCCTTCGGGTTGTAAAGTACTTTCAGCG
 GGGAGGAAGGGAGTAAAGTTAATACCTTTGCTCATTGACGTTACCCGCAGAAGAAG

CACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCAAGCGTTAAT
 CGGAATTACTGGGCGTAAAGCGCACGCAGGCGGTTTGTAAAGTCAGATGTGAAATC
 CCCGGGCTCAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCGTAGAGG
 GGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGT
 GGCGAAGGCGGCCCCCTGGACGAAGACTGACGCTCAGGTGCGAAAGCGTGGGGA
 GCAAACAGGATTAGATAACCTGGTAGTCCACGCCGTAAACGATGTCGACTTGGAGG
 TTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGGGG
 AGTACGGCCGCAAGGTTAAACTCAAATGAATTGACGGGGGCCCGCACAAAGCGGT
 GGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGGTCTTGACATCCA
 CGGAAGTTTTTCAGAGATGAGAATGTGCCTTCGGGAACCGTGAGACAGGTGCTGCAT
 GGCTGTCGTCAGCTCGTGTGTGAAATGTTGGGTAAAGTCCCGCAACGAGCGCAAC
 CTTATCCTTTGTTGCCAGCGGTCCGGCCGGAACTCAAAGGAGACTGCCAGTGAT
 AAAGTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTACGACCAGGGC
 TACACACGTGCTACAATGGCGCATACAAAGAGAAGCGACCTCGCGAGAGCAAGCG
 GACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGT
 CGGAATCGCTAGTAATCGTGGATCAGAATGCCACGGTGAATACGTTCCCGGGCCTT
 GTACACACCGCCCGTACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTTAGCTTAA
 CCTTCGGGAGGGCGCTTACCACCTTTGTGATTGACTGGGGTGAAGTCGTAACAA
 GGTAACCGTAGGGGAACCTGCGGTTGGATCACCTCCTTA

Helicobacter pylori

TTTATGGAGAGTTTGATCCTGGCTCAGAGTGAACGCTGGCGGCGTGCCTAATACATG
 CAAGTCGAACGATGAAGCTTCTAGCTTGCTAGAAATGCTGATTAGTGGCGCACGGGT
 GAGTAACGCATAGGTCATGTGCCTCTTAGTTTGGGATAGCCATTGGAAACGATGAT
 TAATACCAGATACTCCCTACGGGGGAAAGATTTATCGCTAAGAGATCAGCCTATG
 TCCTATCAGCTTGTGGTAAGGTAATGGCTTACCAAGGCTATGACGGGTATCCGG
 CCTGAGAGGGTGAACGGACACACTGGAAGTGAAGACACGGTCCAGACTCCTACG
 GGAGGCAGCAGTAGGGAATATTGCTCAATGGGGGAAACCCTGAAGCAGCAACGC
 CGCGTGGAGGATGAAGGTTTTAGGATTGTAAACTCCTTTTGTAGAGAAGATAAT
 GACGGTATCTAACGAATAAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAAT
 ACGGAGGGTGAAGCGTTACTCGGAATCACTGGGCGTAAAGAGCGCGTAGGGC
 GGATAGTCAGTCAGGTGTGAAATCCTATGGCTTAAACCATAGAAGTGCATTTGAA
 ACTACTATTCTAGAGTGTGGGAGAGGTAGGTGGAATTCTTGGTGTAGGGGTAAA
 ATCCGTAGAGATCAAGAGGAATACTCATTGCGAAGGCGACCTGCTGGAACATT
 ACTGACGCTGATTGCGCGAAAGCGTGGGGAGCAAACAGGATTAGATAACCTGG
 TAGTCCACGCCCTAACGATGGATGCTAGTTGTTGGAGGGCTTAGTCTCTCCAG
 TAATGCAGCTAACGCATTAAGCATCCCGCCTGGGGAGTACGGTTCGCAAGATTA
 AAAGTCAAAGGAATAGACGGGGACCCGCACAAGCGGTGGAGCATGTGGTTTAA
 TTCGAAGATACACGAAGAACCTTACCTAGGCTTGACATTGAGAGAATCCGCTAG
 AAATAGTGGAGTGTCTAGCTTGCTAGACCTTGAAAACAGGTGCTGCACGGCTGT
 CGTCAGCTCGTGTGCTGAGATGTTGGGTAAAGTCCCGCAACGAGCGCAACCCC
 CTTTCTTAGTTGCTAACAGGTTATGCTGAGAACTCTAAGGATACTGCCTCCGTAAG
 GAGGAGGAAGGTGGGGACGACGTCAAGTCATCATGGCCCTTACGCCTAGGGCTA
 CACACGTGCTACAATGGGGTGCACAAAGAGAAGCAATACTGCGAAGTGGAGCCA
 ATCTTCAAACACCTCTCAGTTCGGATTGTAGGCTGCAACTCGCCTGCATGAAGC
 TGGAATCGCTAGTAATCGCAAATCAGCCATGTTGCGGTGAATACGTTCCCGGGTC
 TTGTAATCACCGCCCGTACACCATGGGAGTTGTGTTTGCCTTAAAGTCAGGATAC
 TAAATTGGCTACTGCCACGGCACACACAGCGACTGGGGTGAAGTCGTAACAAG
 GTAACCGTAGGTTGAACCTGCGGCTGGATCACCTCCTT

Legionella pneumophila

NAACTGAAGAGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCATGCTTAACACATG
 CAAGTCGAACGGCAGCATTGTCTAGCTTGGCTAGACAGATGGCGAGTGGCGAACGGG
 TGAGTAACCGCTAGGAATATGCCTTGAAGAGGGGGACAACCTTGGGGAAACTCAAGC
 TAATACCGCATAATGTCTGAGGACGAAAGCTGGGGACCTTCGGGCCTGGCGCTTTA
 AGATTAGCCTGCGTCCGATTAGCTAGTTGGTGGGGTAAGGGCCTACCAAGGCGACG
 ATCGGNAGCTNGTCTGAGAGGATGACCAGCCACACTGGAAGTGGAGACACGGTCCAG
 ACTCCTACGGGAGGCAGCAGTGGGGAATATTGGACAATGGGGGCAACCCTGATCCA
 GCAATGCCCGTGTGTGAAGAAGGCCTNAGGGTTGTAAAGCACTTTCAGTGGGGAG
 GAGGGTTGATAGGTTAAGAGCTGATTAAGTGGACGTTACCCACAGAAGAAGCACCG
 GCTAACTCCGTGCCAGCAGCCNCGGTAATACGGAGNGTGCNAGCGTTAATCGGAAT
 TACTGGGCGTAAAGGGTGCCTAGGTGGTTGATTAAGTTATCTGTGAAATTCCTGGGC
 TTAACCTGGGACGGTCAGATAATACTGGTTGACTCGAGTATGGGAGAGGGTGTAGTGA
 ATTTCCGGTGTAGCGGTGAAATGCGTAGAGATCGGAAGGAACACCAGTGGCGAAGG
 CGGCTACCTGGCCTAATACTGACACTGAGGCACGAAAGCGTGGGGAGCAAACAGGA
 TTAGATACCTTGGTAGTCCACGCTGTAAACGATGTCAACTAGCTGTTGGTTATATGAA
 AATAATTAGTGGCGCAGCAAACGCGATAAGTTGACCGCCTGGGGAGTACGGTTCGCA
 AGATTA AAACTCAAAGGAATTGACGGGGGCNCGCACAAGCGGTGGAGCATGTGGTT
 TAATTCGATGCAACGCGAAGAACCTTACCTACCCTTGACATACAGTGAATTTGCAG
 AGATGCATTAGTGCCTTCGGGAACACTGATACAGGTGCTGCATGGCTGTCGTCAGCT
 CGTGTGCTGAGATGTTGGGTTAAGTCCCGTAACGAGCGCAACCCTTATCCTTAGTTG
 CCAGCATGTGATGGTGGGGACTCTAAGGAGACTGCCGGTGACAAACCGGAGGAAGG
 CGGGGATGACGTCAAGTCATCATGGCCCTTACGGGTAGGGCTACACACGTGCTACA
 ATGGCCGATACAGAGGGCGGCGAAGGGGCGACCTGGAGCAAATCCTTAAAAGTCG
 GTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATCGCTAGTAA
 TCGCGAATCAGCATGTCGCGGTGAATACGTTCCCGGGCCTTGTACACACCGCCCGT
 CACACCATGGGAGTGGGTTGCACCAGAAGTAGATAGTCTAACCTTCGGGGGGACGT
 TTACCACGGTGTGGTTCATGNCNNGGGNNNNNNNGTAACAAGNNNNNNNNNNNN
 NNNNNNNNNNNNNNGATCACCTCCTTA

Listeria monocytogenes

NTTTTAAAGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGCGTGCCTAATACA
 TGCAAGTCGAACGAACGGAGGAAGAGCTTGTCTTCCAAAGTTAGTGGCGGACGGG
 TGAGTAACACGTGGGCAACCTGCCTGTAAGTTGGGGATAACTCCGGGAAACCGGAG
 CTAATACCGAATGATARARTGTGGCGCATGCCACGCTTTTGAAGATGGTTTCGGCT
 ATCGCTTACAGATGGGCCCNCGGTGCATTAGCTAGTTGGTAGGGTAATGGCCTACC
 AAGGCAACGATGCATAGCCGACCTGAGAGGGTNATCGGCCACACTGGGACTGAGA
 CACGGCCNNGACTCCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGACGAA
 AGTCTGACGGAGCAACGCCGCGTGTATGAAGAAGGTTTTCCGGATCGTAAAGTACTG
 TTGTTAGAGAAGAACAAGGATAAGAGTAACTGCTTGTCCCTTGACGGTATCTAACC
 AGAAAGCCACGGCTAACTACGTGCCAGCAGCCGCGTAATACGTAGGTGGCNAG
 CGTTGTCCGGATTTATTGGGCGTAAAGCGCGCGCAGGCGGTCTTTTAAAGTCTGATG
 TGAAAGCCCCCGGCTTAACCGGGNNGGGTCATTGGAAACTGGAAGACTGGAGTG
 CAGAAGAGGAGAGTGAATTCCACGTGTAGCGGTGAAATGCGTAGATATGTGGAG
 GAACACCAGTGGCGAAGGCGACTCTCTGGTCTGTAAGTACGCTGAGGCGCNNA
 AGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAACGATG
 AGTGCTAAGTGTTAGGGGGTTTCCGCCCTTAGTGCTGCAGCTAACGCATTAAGC

ACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGG
 CCNGCACAAGCGGTGGAGCATGTGGTTTAATTCGAANNAACGCGAAGAACCTTA
 CCAGGTCTTGACATCCTTTGACCACTCTGGAGACAGAGCTTCCCTTCGGGGACA
 AAGTGACAGGTGGNGCATGGTTGTCGTCAGCTCGTGTCTGTGAGATGTTGGGTAA
 GTCCCGCAACGAGCGCAACCCTTGATTTTAGTTGCCAGCATTAGTTGGGCACTC
 TAAAGTGACTGCCGGTGCAAGCCGGAGGAAGGTGGGGATGACGTCAAATCATCA
 TGCCCCTTATGACCTGGGCTACACACGTGCTACAATGGATAGTACAAAGGGTNGC
 GAAGCCGCGAGGTGGAGCTAATCCATAAACTATTCTCAGTTCGGATTGTAGGC
 TGCAACTCGCCTACATGAAGCCGGAATCGCTAGTAATCGTGGATCAGCATGCCAC
 GGTGAATACGTTCCCGGGCCTTGACACACCGCCCGTCACACCACGAGAGTTTGT
 AACACCCGAAGTCGGTAGGGTAACCTTTATGGAGCCAGCCGCCGAAGGTGGGAC
 AGATAATNGGGGNNNNNNNGTAACAAGNNNNNNNNNNNNNNNNNNNNNNNNNN
 NNGATCACCTCCTTTCT

Mycobacterium leprae

TTGTTTGGAGAGTTTGATCCTGGCTCAGGACGAACGCTGGCGGGCGTGCTTAACAC
 ATGCAAGTTCGAACGGAAAGGTCTCTAAAAAATCTTTTTTAGAGATACTCGAGTGG
 CGAACGGGTGAGTAACACGTGGGTAATCTGCCCTGCACTTCAGGGATAAGCTTG
 GGAAACTGGGTCTAATACCGGATAGGACTTCAAGGCGCATGTCTTGTGGTGGAA
 AGCTTTTTGCGGTGCAGGATGGGCCC GCGGCCTATCAGCTTGTGGTGGGGTGA
 CGGCCTACCAAGGCGACGACGGGTAGCCGGCCTGAGAGGGTGTCCGGCCACA
 CTGGGACTGAGATACGGCCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATT
 GCACAATGGGCGCAAGCCTGATGCAGCGACGCCCGCTGGGGGATGACGGCCTT
 CGGGTTGTAAACCTCTTTCACCATCGACGAAGGTCTGGGTTTTCTCGGATTGACG
 GTAGGTGGAGAAGAAGCACCGGCCAACTACGTGCCAGCAGCCGCGGTAATACG
 TAGGGTGCAGCGTGTGTCGGAATTAAGTGGGCGTAAAGAGCTCGTAGGTGGTTG
 TCGCGTTGTTTCGTGAAATCTCACGGCTTAACTGTGAGCGTGCGGGCGATACGGG
 CAGACTAGAGTACTGCAGGGGAGACTGGAATTCCTGGTGTAGCGGTGGAATGCG
 CAGATATCAGGAGGAACACCAGTGGCGAAGGCGGGTCTCTGGGCAGTAAGTAC
 GCTGAGGAGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCA
 CGCCGTAAACGGTGGGTACTAGGTGTGGGTTTCCTTCCTTGGGATCCGTGCCGT
 AGCTAACGCATTAAGTACCCCGCCTGGGGAGTACGGCCGCAAGGCTAAACTC
 AAAGGAATTGACGGGGGCCCGCACAAGCGGCGGAGCATGTGGATTAATTCGAT
 GCAACGCGAAGAACCTTACCTGGGTTTGACATGCACAGGATGCGTCTAGAGATA
 GGCACCTCCCTTGTGGCCTGTGTGCAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCTG
 TGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTGTCTCATGTTGCCAGCAC
 GTAATGGTGGGGACTCGTGAGAGACTGCCGCGGTCAACTCGGAGGAAGGTGGGGA
 TGACGTCAAGTCATCATGCCCTTATGTCCAGGGCTTCACACATGCTACAATGGCCG
 GTACAAAGGGCTGCGATGCCGCAAGGTAAAGCGAATCCTTTAAAGCCGGTCTCAG
 TTCGGATCGGGTCTGCAACTCGACCCCGTGAAGTCGGAGTCGCTAGTAATCGCA
 GATCAGCAACGCTGCGGTGAATACGTTCCCGGGCCTTGACACACCGCCCGTCAC
 GTCATGAAAGTCGGTAACACCCGAAGCCAGTGGCCTAACCTCGGGAGGGAGCT
 GTCCAAGGTGGGATCGGCGATTGGGACGAAGTCGTAACAAGGTAGCCGTACCGG
 AAGGTGCGGCTGGATCACCTCCTTTCT

Neisseria gonorrhoea

TGAACATAAGAGTTTGATCCTGGCTCAGATTGAACGCTGGCGGCATGCTTTACACA
 TGCAAGTCGGACGGCAGCACAGGGAAGCTTGCTTCTCGGGTGGCGAGTGGCGAA

CGGGTGAGTAACATATCGGAACGTACCGGGTAGCGGGGGATAACTGATCGAAAGA
TCAGCTAATACCGCATACTGCTTGTGAGAGGGAAAGCAGGGGACCTTCGGGCCTTGC
GCTATCCGAGCGGCCGATATCTGATTAGCTGGTTGGCGGGGTAAAGGCCACCAA
GGCGACGATCAGTAGCGGGTCTGAGAGGATGATCCGCCACACTGGGACTGAGAC
ACGGCCAGACTCCTACGGGAGGCAGCAGTGGGGAATTTGGACAATGGGCGCA
AGCCTGATCCAGCCATGCCGCGTGTCTGAAGAAGGCCTTCGGGTTGTAAAGGACT
TTTGTACAGGGAAGAAAAGGCTGTTGCCAATATCGGCGGCCGATGACGGTACCTGA
AGAATAAGCACCGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGGTGC GA
GCGTTAATCGGAATTACTGGGCGTAAAGCGGGCGCAGACGGTTACTTAAGCAGGA
TGTGAAATCCCCGGGCTCAACCCGGGAAGTGC GTTCTGAACTGGGTGACTCGAGT
GTGTCAGAGGGAGGTGGAATTCCACGTGTAGCAGTGAAATGCGTAGAGATGTGGA
GGAATACCGATGGCGAAGGCAGCCTCCTGGGATAACACTGACGTTTCATGTCCGAA
AGCGTGGGTAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCCTAACGATGT
CAATTAGCTGTTGGGCAACTTGATTGCTTGGTAGCGTAGCTAACGCGTGAAATTGA
CCGCCTGGGGAGTACGGTCGCAAGATTA AAACTCAAAGGAATTGACGGGGACCC
GCACAAGCGGTGGATGATGTGGATTAATTCGATGCAACGCGAAGAACCTTACCTG
GTTTTGACATGTGCGGAATCCTCCGGAGACGGAGGAGTGCCTTCGGGAGCCGTA
ACACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTCTGTGAGATGTTGGGTAAAGTC
CCGCAACGAGCGCAACCCTTGTCAATTAGTTGCCATCATTTCGGTTGGGCACTCTAAT
GAGACTGCCGGTGACAAGCCGGAGGAAGGTGGGGATGACGTCAAGTCCTCATGG
CCCTTATGACCAGGGCTTCACACGTCATAACAATGGTTCGGTACAGAGGGTAGCCAA
GCCGCGAGGCGGAGCCAATCTCACAAAACCGATCGTAGTCCGGATTGCACTCTGC
AACTCGAGTGCATGAAGTCGGAATCGCTAGTAATCGCAGGTCAGCATACTGCCGT
GAATACGTTCCCGGTCTTGTACACACCCGCCGTCACACCATGGGAGTGGGGGAT
ACCAGAAGTAGGTAGGGTAACCGCAAGGAGTCCGCTTACCACGGTATGCTTCATGA
CTGGGGTGAAGTCGTAACAAGGTAGCCGTAGGGGAACCTGCGGCTGGATCACCTC
CTTTCT

Rickettsia rickettsii

CCTGGCTCAAACGGAACGCTATCGGTATGCTTAACACATGCAAGTCGAACGGACT
AATTGGGGCTTGCTCCAATTAGTTAGTGGCAGACGGGTGAGTAACACGTGGGAAT
CTACCCATCAGTACGGAATAACTTTTAGAAATAAAAGCTAATACCGTATATTCTCTG
CGGAGGAAAGATTTATCGCTGATGGATGAGCCC GCGTCAGATTAGGTAGTTGGT
GAGGTAATGGCTCACCAAGCCGACGATCTGTAGCTGGTCTGAGAGGATGATCAG
CCACACTGGGACTGAGACACGGCCAGACTCCTACGGGAGGCAGCAGTGGGGA
ATATTGGACAATGGGCGAAAGCCTGATCCAGCAATACCGAGTGAGTGATGAAGGC
CTTAGGGTTGTAAAGCTCTTTTAGCAAGGAAGATAATGACGTTACTTGCAGAAAAA
GCCCCGGCTAACTCCGTGCCAGCAGCCGCGGTAAGACGGAGGGGGCTAGCGTT
GTTTCGGAATTACTGGGCGTAAAGAGTGC GTAGGCGGTTTAGTAAGTTGGAAGTGA
AAGCCC GGGGCTTAACCTCGGAATTGCTTTCAAACTACTAATCTAGAGTGTAGTA
GGGGATGATGGAATTCCTAGTGTAGAGGTGAAATTCTTAGATATTAGGAGGAACA
CCGGTGGCGAAGGCGGTCATCTGGGCTACCACTGACGCTGATGCACGAAAGCG
TGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTG
CTAGATATCGGAAGATTCTCTTTTCGGTTTCGCAGCTAACGCATTAAGCACTCCGCC
TGGGGAGTACGGTCGCAAGATTA AAACTCAAAGGAATTGACGGGGGCTCGCACAA
GCGGTGGAGCATGCGGTTTAATTCGATGTTACGCGAAAAACCTTACCAACCTTGA
CATGGTGGTTGCGGATCGCAGAGATGCTTTTCTTCAGCTCGGCTGGACCACACAC
AGGTGTTGCATGGCTGTCGTCAGCTCGTGTCTGTGAGATGTTGGGTAAAGTCCCGC
AACGAGCGCAACCCTCATTCTTATTTGCCAGCGGGTAATGCCGGGAACCTATAAGA

AAACTGCCGGTGATAAGCCGGAGGAAGGTGGGGACGACGTCAAGTCATCATGGC
 CCTTACGGGTTGGGCTACACGCGTGCTACAATGGTGTTTACAGAGGGAAGCAAGA
 CGGCGACGTGGAGCAAATCCCTAAAAGACATCTCAGTTCGGATTGTTCTCTGCAAC
 TCGAGAGCATGAAGTTGGAATCGCTAGTAATCGCGGATCAGCATGCCGCGGTGAA
 TACGTTCTCGGGCCTTGTACACACTGCCCGTCACGCCATGGGAGTTAGTTTTACCT
 GAAGGTGGTGAGCTAACGCAAGAGGCAGCCAACCACGGTAAAATTAGCGACTGGG
 GTGAAGTCGTAA

Salmonella typhi

AAATTGAAGAGTTTGATCATGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACA
 TGCAAGTCGAACGGTAACAGGAAGCAGCTTGCTCTTTGCTGACGAGTGGCGGACG
 GGTGAGTAATGTCTGGGAAACTGCCTGATGGAGGGGGATAACTACTGGAAACGGT
 GGCTAATACCGCATAACGTCGCAAGACCAAGAGGGGGACCTTCGGGCCTCTTGC
 CATCGGATGTGCCAGATGGGATTAGCTAGTAGGTGGGGTAACGGCTCACCTAGG
 CGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGAAGCTGAAGCACG
 GTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCC
 TGATGCAGCCATGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAG
 CGGGGAGGAAGGTGTTGTGGTTAATAACCGCAGCAATTGACGTTACCCGCAGAAG
 AAGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGAAGCGT
 TAATCGGAATTACTGGGCGTAAAGCGCACGCAGGCCGGTTTGTAAAGTCAGATGTG
 AAATCCCCGGGCTAACCTGGGAACTGCATCTGATACTGGCAAGCTTGAGTCTCG
 TAGAGGGGGGTAGAATTCCAGGTGTAGCGGTGAAATGCGTAGAGATCTGGAGGA
 ATACCGGTGGCGAAGGCGGCCCTGGACGAAGACTGACGCTCAGGTGCGAAA
 GCGTGGGGAGCAAACAGGATTAGATAACCCTGGTAGTCCACGCCGTAAACGATCTC
 TACTTGGAGGTTGTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTAGA
 GTGCTTGGGGAGTACGGCCGCAAGGTTAAACTCAAATGAATTGACGGGGGCCCG
 CACAAGCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTGG
 TCTTGACATCCACAGAACTTCCAGAGATGAGATTGTGCCTTCGGGAACTGTGAGA
 CAGGTGCTGCATGGCTGTCGTCAGCTCGTGTGTTGTGAAATGTTGGGTTAAGTCCCG
 CAACGAGCGCAACCCTTATCCTTTGTTGCCAGCGGTCCGGCCGGGAACTCAAAG
 GAGACTGCCAGTGATAAACTGGAGGAAGGTGGGGATGACGTCAAGTCATCATGG
 CCTTACGACCAGGGCTACACACGTGCTACAATGGCGCATACAAAGAGAAGCGA
 CCTCGCGAGAGCAAGCGGACCTCATAAAGTGCGTCGTAGTCCGGATTGGAGTCT
 GCAACTCGACTCCATGAAGTCGGAATCGCTAGTAATCGTGGATCAGAATGCCACG
 GTGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAGTGGGT
 TGCAAAGAAGTAGGTAGCTTAACCTTCGGGAGGGCGCTTACCCTTTGTGATTCA
 TGACTGGGGTGAAGTCGTAACAAGGTAACCGTAGGGGAACCTGCGGTTGGATCA
 CCTCCTTA

Staphylococcus aureus

CTGGCTCAGGATGAACGCTGGCGGCGTGCCATAACATGCAAGTCGAGCGAACG
 GACGAGAAGCTTGCTTCTCTGATGTTAGCGGCGGACGGGTGAGTAACACGTGGA
 TAACCTACCTATAAGACTGGGATAACTTCGGGAAACCGGAGCTAATACCAGATAAT
 ATTTTGAACCGCATGGTTCAAAGGTGAAAGACGGTCTTGCTGTCACTTATAGATGG
 ATCCGCGCTGCATTAGCTAGTTGGTAAGGTAACGGCTTACCAAGGCAACGATGCA
 TAGCCGACCTGAGAGGGTGATCGKCCACACTGGAACCTGAGACACGGTCCAGACT
 CCTACGGGAGGCAGCAGTAGGGAATCTTCCGCAATGGGCGAAAGCCTGACGGAG
 CAACGCCGCGTGAGTGATGAAGGTCTTCGGATCGTAAAACCTCTGTTATTAGGGAA

GAACATATGTGTAAGTAACTGTGCACATCTTGACGGTACCTAATCAGAAAGCCACG
GCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTGGCAAGCGTTATCCGGA
ATTATTGGGCGTAAAGCGCGCGTAGGCGGTTTTTYAAGTCTGATGTGAAAGCCCA
CGGCTCAACCGTGGAGGGTCATTGGAAACTGGAAAACCTTGAGTGCAGAAGAGGA
AAGTGAATTCCATGTGTAGCGGTGAAATGCGCAGAGATATGGAGGAACACCAGT
GGCGAAGGCGACTTTCTGGTCTGTAAGTACGCTGATGTGCGAAAGCGTGGGGA
TCAAACAGGATTAGATACCCTGGTAGTCCACGCCGTAAACGATGAGTGCTARGTG
TTAGGGGGTTTTCCGCCCTTAGTGCTGCAGCTAACGCATTAAGCACTCCGCCTGG
GGAGTACGACCGCAAGGTTGAAACTCAAAGGAATTGACGGGGACCCGCACAAGC
GGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAAATCTTGACA
TCCTTTGACAACCTCTAGAGATAGAGCCTTCCCCTTCGGGGGACAAAGTGACAGGT
GGTGCATGGTTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAAC
GAGCGCAACCCCTTAAGCTTAGTTGCCATCATTAAAGTTGGGCACTCTAAGTTGACT
GCCGGTGACAAACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTT
ATGATTTGGGCTACACACGTGCTACAATGGACAATACAAAGGGCAGCGAAACCG
CGAGGTCAAGCAAATCCCATAAAGTTGTTCTCAGTTCGGATTGTAGTCTGCAACTC
GACTACATGAAGCTGGAATCGCTAGTAATCGTAGATCAGCATTCTACGGTGAATAC
GTTCCCGGGTCTTGTACACACCGCCCGTACACCACGAGAGTTTGTAAACCCCG
AAGCCGGTGGAGTAACTTTTAGGAGCTAGCCGTCGAAGGTGGGACAAATGATTG
GGGTGA

Streptococcus pyogenes

GACGAACGCTGGCGGCGTGCCTAATACATGCAAGTAGAACGCTGAGAACTGGTG
CTTGCACCGGTTCAAGGAGTTGCGAACGGGTGAGTAACGCGTAGGTAACCTACC
TCATAGCGGGGATAACTATTGGAAACGATAGCTAATACCGCATAAGAGAGACT
AACGCATGTTAGTAATTTAAAAGGGGCAATTGCTCCACTATGAGATGGACCTGC
GTTGTATTAGCTAGTTGGTGAGGTAAAGGCTACCAAGGCGACGATACATAGC
CGACCTGAGAGGGTGATCGGCCACACTGGGACTGAGACACGGCCCAGACTC
CTACGGGAGGCAGCAGTAGGGAATCTTCGGCAATGGGGGCAACCCCTGACCG
AGCAACGCCGCGTGAGTGAAGAAGGTTTTTCGGATCGTAAAGCTCTGTTGTTA
GAGAAGAATGATGGTGGGAGTGGAAAATCCACCAAGTGACGGTAACTAACC
AGAAAGGGACGGCTAACTACGTGCCAGCAGCCGCGGTAATACGTAGGTCCC
GAGCGTTGTCCGATTTATTGGGCGTAAAGCGAGCGCAGGCGGTTTTTTAAG
TCTGAAGTTAAAGGCATTGGCTCAACCAATGTACGCTTTGGAAACTGGAGAAC
TTGAGTGCAGAAGGGGAGAGTGGAAATCCATGTGTAGCGGTGAAATGCGTAG
ATATATGGAGGAACACCGGTGGCGAAAGCGGCTCTCTGGTCTGTAAGTACG
CTGAGGCTCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGTCCA
CGCCGTAAACGATGAGTGCTAGGTGTTAGGCCCTTTCGGGGCTTAGTGCCG
GAGCTAACGCATTAAGCACTCCGCCTGGGGAGTACGACCGCAAGGTTGAAAC
TCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGAGCATGTGGTTTAATTC
GAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCCCGATGCCCGCTCTAG
AGATAGAGTTTTACTTCGGTACATCGGTGACAGGTGGTGCATGGTTGTCGTC
AGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTA
TTGTTAGTTGCCATCATTAAAGTTGGGCACTCTAGCGGAGACTGCCGGTAATAA
ACCGGAGGAAGGTGGGGATGACGTCAAATCATCATGCCCTTATGACCTG
GGCTACACACGTGCTACAATGGTTGGTACAACGAGTCGCAAGCCGGTGAC
GGCAAGCTAATCTCTTAAAGCCAATCTCAGTTCGGATTGTAGGCTGCAACTC
GCCTACATGAAGTCGGAATCGCTAGTAATCGCGGATCAGCACGCCGCGGT
GAATACGTTCCCGGGCCTTGTACACACCGCCCGTACACCACGAGAGTTTG

TAACACCCGAAGTCGGTGAGGTAACCTATTAGGAGCCAGCCGCCTAAGGTG
GGATAGATGATTGGGGTGAAGTCGTAACAAGGTAGCCGTATCGGAAGG

Thermus aquaticus

GCTCAGGGTGAACGCTGGCGGCGTGCCTAAGACATGCAAGTCGTGCGGG
CCGTGGGGTATCTCACGGTCAGCGGCGGACGGGTGAGTAACGCGTGGGTG
ACCTACCCGGAAGAGGGGGACAACATGGGGAAACCCAGGCTAATCCCC
ATGTGGACACATCCTGTGGGGTGTGTTTAAAGGGTTTTGCCCGCTTCCGGA
TGGGCCCGCGTCCCATCAGCTAGTTGGTGGGGTAAGAGCCCACCAAGGCG
ACGACGGGTAGCCGGTCTGAGAGGACGGCCGGCCACAGGGGCACTGAGA
CACGGGCCCCACTCCTACGGGAGGCAGCAGTTAGGAATCTTCCGCAATGG
GCGCAAGCCTGACGGAGCGACGCCGCTTGGAGGAGGAAGCCCTTCGGGGT
GTAACCTCCTGAACCCGGGACGAAACCCCCGATGAGGGGACTGACGGTAC
CGGGGTAATAGCGCCGGCCAACCTCCGTGCCAGCAGCCGCGGTAATACGGA
GGGCGCGAGCGTTACCCGGATTTACTGGGCGTAAAGGGCGTGTAGGCGGCT
TGGGGCGTCCCATGTGAAAGGCCACGGCTCAACCGTGGAGGAGCGTGGGAT
ACGCTCAGGCTAGACGGTGGGAGAGGGTGGTGGAAATTCGCCGAGTAGCGGT
GAAATGCGCAGATAACGGGAGGAACGCCGATGGCGAAGGCAGCCACCTGG
TCCACTCGTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACCGGATTAGAT
ACCCGGGTAGTCCACGCCCTAACGATGCGCGCTAGGTCTCTGGGTTATCTG
GGGGCCGAAGCTAACCGGTTAAGCGCGCCGCCTGGGGAGTACGGCCGCAAG
GCTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAAGCGGTGGAGCATGTGG
TTTAATTCGAAGCAACGCGAAGAACCTTACCAGGCCTTGACATGCTAGGGAA
CCTGGGTGAAAGCCTGGGGTGCCCCGCGAGGGGAGCCCTAGCACAGGTGCT
GCATGGCCGTCGTCAGCTCGTGTGCTGAGATGTTGGGTTAAGTCCCGCAACG
AGCGCAACCCCTGCCGTTAGTTGCCAGCGGGTGAAGCCGGGCACTCTAACG
GGACTGCCTGCGAAAGCAGGAGGAAGGCGGGGACGACGTCTGGTCATCATG
GCCCTTACGGCCTGGGCGACACACGTGCTACAATGCCACTACAGAGCGAG
GCGACCTGGCAACAGGGAGCGAATCGCAAAAAGGTGGGGCGTAGTTCGGATT
GGGGTCTGCAACCCGACCCCATGAAGCCGGAATCGCTAGTAATCGCGGATC
AGCCATGCCGCGGTGAATACGTTCCCGGGCCTTGACACACCCGCCGTCAC
GCCATGGGAGCGGGTTCTACCCGAAGTCGCCGGGAGCCTTAGGGCAGGCGC
CGAGGGTAGGGCCCGTGACTGGGGCGAAGTCGTAACAAGGTAGCTGTACCG

Treponema pallidum

TCGAACGGCAAGGAAGCGAATTTTCGTTTCTCTAGAGTGGCGGACTGGTGA
GTAACGCGTGGGTAAATCTGCCTTTGAGATGGGGATAGCCTCTAGAAATAGG
GGGTAATACCGAATACGCTCTTTTGGACGTAGTCTTTGAGAGGAAAGGGG
CTGCGGCCTCGCTCAGAGATGAGCCTGCGACCCATTAGCTTGTTGGTGGGG
TAATGGCCTACCAAGGCGTCGATGGGTATCCGACCTGAGAGGGTGACCGGA
CACACTGGGACTGAGATACGGCCCAGACTCCTACGGGAGGCAGCAGCTAAG
AATATTCGCAATGGGCGAAAGCCTGACGGAGCGACACCGCGTGGATGAGG
AAGGTGCAAAGATTGTAAGTTCTTTTGGCGACGAAGAATGAGGACGGGAGG
GAATGCCCGTTTGATGACGGTAGTCGTGCGAATAAGCCCCGGCTAATTACGT
GCCAGCAGCCGCGGTAACACGTAAGGGGCGAGCGTTGTTTCGGAATTATTGG
GCGTAAAGGGCATGCAGGCGGACTGGTAAGCCTGGTGTGAAATCCCCGAGC
TCAACTTGGGAACCTGCACTGGGTACTGCTGGTCTAGAATCACGGAGGGGAA
ACCGGAATTCCAAGTGTAGGGGTGGAATCTGTAGATATTTGGAAGAACCCG

GTGGCGAAGGCGGGTTTCTGGCCGATGATTGACGCTGAGGTGCGAAGGTG
TGGGGAGCGAACAGGATTAGATACCCTGGTAGTCCACACAGTAAACGATGT
ACACTAGGTGTTGGGGCATGAGTCTCGGCGCCGACGCGAACGCATTAAGT
GTACCGCCTGGGGAGTATGCTCGCAAGAGTGAAACTCAAAGGAATTGACG
GGGGCCCCGCACAAGCGGTGGAGCATGTGGTTTAATTCGATGATACGCGAG
GAACCTTACCCGGGTTTGACATCAAGAGGAGCGCCGTAGAAATGCGGTGG
CGTAGCGATACGCCCTTGACAGGTGCTGCATGGCTGTCGTCAGCTCGTG
CCGTGAGGTGTTGGGTAAAGTCCCACAACGAGCGCAACCCCTACTGCCAG
TTGCCAGCAAGTGGTGTGGGGACTCTGGCGGAACTGCCGGTGACAAACC
GGAGGAAGGTGGGGATGACGTCAAGTCATCATGGCCCTTATGTCCGGGG
CTACACACGTGCTACAATGGTTGCTACAGAGCGATGCGAGGTTGTGAAGT
GGAGCAAACCGCAAAAAGGCAATCGTAGTCCGGATTGAAGTCTGAAACTC
GACTTCATGAAGTTGGAATCGCTAGTAATCGCACATCAGCATGGTGCGGT
GAATGTGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATCCGAGTT
GGAGATACCCGAAGTCACTAGCCTAACCCGCAAGGGAGGGCGGTGCC
GAAGGTATGTTGGTAAGGAGGGTGAAGTCGTAACAAGGTAGCCGTAC
CGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGAGA

Vibrio cholerae

AGAGTTTGATNNTGGCTCAGATTGAACGCTGGCGGCAGGCCTAACACAT
GCAAGTCGAGCGGCAGCACAGAGGAACTTGTTCCTTGGGTGGCGAGCG
GCGGACGGGTGAGTAATGCCTGGGAAATTGCCCGGTAGAGGGGGATAA
CCATTGAAACGATGGCTAATACCGCATAACCTCGCAAGAGCAAAGCAG
GGGACCTTCGGGCCTTGCCTACCGGATATGCCAGGTGGGATTAGCTA
GTTGGTGAGGTAAGGGCTACCAAGGCGACGATCCCTAGCTGGTCTGAG
AGGATGATCAGCCACACTGGAAGTGAACACGGTCCAGACTCCTACGGG
AGGCAGCAGTGGGGAATATTGCACAATGGGCGCAAGCCTGATGCAGCCA
TGCCGCGTGTATGAAGAAGGCCTTCGGGTTGTAAAGTACTTTCAGTAGGG
AGGAAGGTGGTTAAGTTAATACCTTAATCATTGACGTTACCTACAGAAGA
AGCACCGGCTAACTCCGTGCCAGCAGCCGCGGTAATACGGAGGGTGCA
AGCGTTAATCGGAATTACTGGGCGTAAAGCGCATGCAGGTGGTTTGTAA
GTCAGATGTGAAAGCCCTGGGCTCAACCTAGGAATCGCATTTGAAACTGA
CAAGCTAGAGTACTGTAGAGGGGGGTAGAATTCAGGTGTAGCGGTGAA
ATGCGTAGAGATCTGAAGGAATACCGGTGGCGAAGGCGGCCCCCTGGA
CAGATACTGACACTCAGATGCGAAAGCGTGGGGAGCAAACAGGATTAGA
TACCCTGGTAGTCCACGCCGTAAACGATGTCTACTTGGAGGTTGTGCC
TAGAGGCGTGGCTTTCGGAGCTAACGCGTTAAGTAGACCGCCTGGGGA
GTACGGTCGCAAGATTAAACTCAAATGAATTGACGGGGGCCCGCACAA
GCGGTGGAGCATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCTA
CTCTTGACATCCAGAGAATCTAGCGGAGACGCTGGAGTGCCTTCGGGA
GCTCTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATG
TTGGGTAAAGTCCCACAACGAGCGCAACCCCTTATCCTTGTGGCCAGCA
CGTAATGGTGGGAACTCCAGGGAGACTGCCGGTGATAAACCGGAGGA
AGGTGGGGACGACGTCAAGTCATCATGGCCCTTACGAGTAGGGCTACA
CACGTGCTACAATGGCGTATACAGAGGGCAGCGAATACCGCGAAGGTG
GAGCGAATCTCACAAGTACGTCGTAGTCCGGATTGGAGTCTGCAACTC
GACTCCATGAAGTCGGAATCGCTAGTAATCGCAAATCAGAATGTTGCGG
TGAATACGTTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGGAG
TGGGCTGCAAAAAGAAGCAGGTAGTTAACCTTCGGGAGGACGCTTGCC

ACTTTGTGGTTCATGACTGGGGTGAAGTCGTAACAAGGTAGCGCTAGG
GGAACCTGGCGCTGGATCACCTCCTTT

Yersinia pestis

ATTGAACGCTGGCGGCAGGCCTAACACATGCAAGTCGAGCGGCAGCGGG
AAGTAGTTTACTACTTTGCCGGCGAGCGGCGGACGGGTGAGTAATGTCTG
GGGATCTGCCTGATGGAGGGGGATAACTACTGGAAACGGTAGCTAATACC
GCATGACCTCGCAAGAGCAAAGTGGGGGACCTTAGGGCCTCACGCCATC
GGATGAACCCAGATGGGATTAGCTAGTAGGTGGGGTAATGGCTCACCTAG
GCGACGATCCCTAGCTGGTCTGAGAGGATGACCAGCCACACTGGAECTGA
GACACGGTCCAGACTCCTACGGGAGGCAGCAGTGGGGAATATTGCACAAT
GGGCGCAAGCCTGATGCAGCCATGCCGCGTGTGTGAAGAAGGCCTTCGGGT
TGTAAGCACTTTCAGCGAGGAGGAAGGGGTTGAGTTTAATACGCTCAATCA
TTGACGTTACTCGCAGAAGAAGCACCCGGCTAACTCCGTGCCAGCAGCCGCG
GTAATACGGAGGGTGCAAGCGTTAATCGGAATTAAGTGGGCGTAAAGCGCAC
GCAGGCGGTTTGTAAAGTCAGATGTGAAATCCCCGCGCTTAACGTGGGAAC
TGCAITTTGAAACTGGCAAGCTAGAGTCTTGTAGAGGGGGGTAGAATTCCAG
GTGTAGCGGTGAAATGCGTAGAGATCTGGAGGAATACCGGTGGCGAAGGCG
GCCCCCTGGACAAAGACTGACGCTCAGGTGCGAAAGCGTGGGGAGCAAAC
AGGATTAGATACCCTGGTAGTCCACGCTGTAAACGATGTCGACTTGGAGGTT
GTGCCCTTGAGGCGTGGCTTCCGGAGCTAACGCGTTAAGTCGACCGCCTGG
GGAGTACGGCCGCAAGGTTAAACTCAAATGAATTGACGGGGGCCCGCAC
AAGCGGTGGAGCATGTGGTTTAATTGATGCAACGCGAAGAACCTTACCTA
CTCTTGACATCCACAGAATTTGGCAGAGATGCTAAAGTGCCTTCGGGAAC
GTGAGACAGGTGCTGCATGGCTGTCGTCAGCTCGTGTGTGAAATGTTGGGT
TAAGTCCCGCAACGAGCGCAACCCCTTATCCTTTGTTGCCAGCACGTAATGG
TGGGAACTCAAGGGAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGATG
ACGTCAAGTCATCATGGCCCTTACGAGTAGGGCTACACACGTGCTACAATG
GCAGATACAAAGTGAAGCGAACTCGCGAGAGCCAGCGGACCACATAAAGT
CTGTCGTAGTCCGGATTGGAGTCTGCAACTCGACTCCATGAAGTCGGAATC
GCTAGTAATCGTAGATCAGAATGCTACGGTGAATACGTTCCCGGGCCTTGTA
CACACCGCCCGTCACACCATGGGAGTGGGTTGCAAAAGAAGTAGGTAGCTT
AACCTTCGGGAGGGCGCTTACCCTTTGTGATTCATGACTGGGGTGAAGTCGT

Contents of this page

[Background](#)
[Procedure](#)
[Output and Interpretation](#)
[Graphical representation of the results](#)
[List of sequences that match the query](#)
[Base-to-base alignment of the query sequence with two of its matches](#)
[Questions](#)

Curriculum Resources

Appendix II. Activity handout

Identification of Bacterial Unknowns by rRNA Sequence Similarity

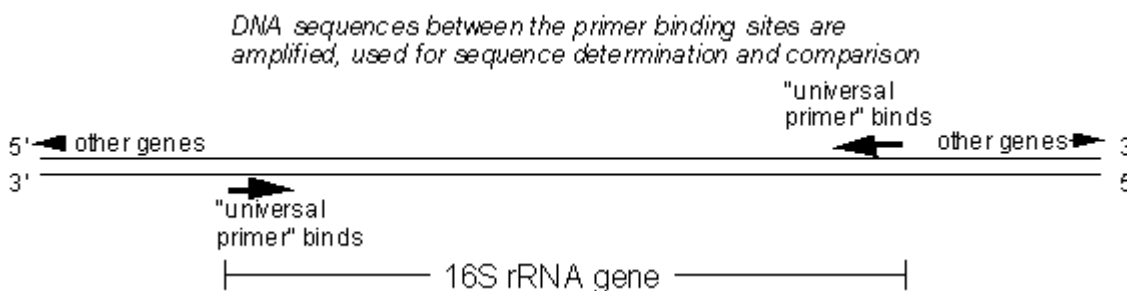
Background

You have been introduced in lecture to the idea that phylogenetic relatedness between organisms is reflected by similarity between nucleic acid sequences in their genomes. It has also been explained why ribosomal RNA sequences are particularly suitable for such comparisons.

In addition to providing a powerful tool for determining the relationships between well-characterized organisms, rRNA sequence data provide an equally powerful way of rapidly identifying and characterizing samples of unknowns. It is now extremely convenient, rapid and relatively inexpensive to obtain rRNA sequences from an unknown sample, to compare the obtained sequence to a database of sequences and to learn whether your unknown corresponds to a previously characterized species or is truly something new. If it is something new, you can rapidly determine which organisms are its closest relatives, and perhaps predict some of the physiological properties of your unknown.

The most commonly used method for obtaining rRNA sequences from an unknown involves the use of the polymerase chain reaction (PCR) to amplify rRNA gene sequences from the genome of your organism. In the polymerase chain reaction a pair of short synthetic DNAs that can base pair with opposite ends of the gene of interest (and on opposite strands) are mixed with a DNA sample from the organism of interest along with dNTPs and a DNA polymerase. After heating to make the sample DNA single-stranded, the synthetic DNAs, called primers, can base pair with the sample DNA and be used as primers by the DNA polymerase. After one such reaction, if you began with one copy of your gene of interest, you will now have two. If the reaction is repeated, you will have four, and so on. In general, the reaction is repeated twenty to thirty times.

For organism identification, the primers used are derived from opposite ends of the 16S rRNA gene. Because the sequence of the 16S rRNA gene is highly conserved in many different organisms, you don't need to know the exact sequence of the particular gene you are trying to amplify. Investigators use "universal primers" that can base pair with any 16S rRNA sequence well enough to allow PCR amplification.



The PCR product (many millions of copies of the rRNA gene from your organism) can be used as substrate in other reactions that determine the sequence of base pairs in the DNA. One can then use a computer to compare that sequence to a database of sequences

such as GenBank.

GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. Investigators who determine the sequences of genes (or sometimes of entire genomes) deposit the sequence information with GenBank. Aside from doing this as an altruistic service to the scientific community, many journals require submission of new sequence data to GenBank as a precondition to publishing any manuscript referring to the sequence. There are approximately 2,162,000,000 nucleotide bases in 3,044,000 sequence records as of December 1998. Many of these records are individual genes or gene fragments. An increasing number are derived from complete genome sequences of viruses, bacteria, and complex eukaryotic organisms. For example, the complete genomes of bacteria such as *E. coli*, of the yeast *Saccharomyces cerevisiae*, and the fruit fly *Drosophila melanogaster* have been determined and deposited in GenBank. A new release is made every two months. GenBank is part of the International Nucleotide Sequence Database Collaboration, which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at the National Center for Biotechnology Information (NCBI). These three organizations exchange data on a daily basis. The GenBank database contains 16S rRNA sequences from many thousands of Bacteria, Archaea and Eukaryotes.

Procedure

You will each be sent an e-mail message containing two DNA sequences that might have been amplified from unknown organisms. Using whatever web browser makes you happy, go to the following URL:

<http://www.ncbi.nlm.nih.gov/blast/blast.cgi>


The page you will arrive at, "Basic BLAST", is maintained by the National Center for Biotechnology Information (a part of the National Institutes of Health - your tax dollars at work). It is maintained as a service to the life science community, and provides a mechanism by which an investigator can submit a DNA (or protein) sequence and have a sophisticated program, called BLAST, compare it to all of the sequences deposited in the GenBank database. If the program finds that the submitted sequence is similar to sequences in the database, it will deliver the names and a brief description of the matching sequences in the output for the program.

On the web page (see the figure below) you should see a box in which you can paste your sequence (marked with an asterisk on the figure). You may first need to click inside the box in order to activate it. There are also several pull-down menus (numbered on the figure) that allow you to (1) choose the program used for the match search, (2) choose the database to be compared to and (3) specify the format that your sequence is in. The program should be `blastn`. This program allows you to compare a submitted nucleotide sequence to a nucleotide sequence database. The database should be `nr`. The format menu should say "Sequence in FASTA format." All of these are the default choices, so you should not need to make any changes.

After you have pasted your sequence into the box and confirmed that the program, database and sequence format options are correct, click on the "Search" button. The time required for the search will vary depending upon the time of day, but in any case should take no longer than 10 minutes.

Output and Interpretation

As an example, this handout shows the output from a GenBank BLAST search using the *E. coli* 16S rRNA sequence as the query. A successful search such as this one will generate many pages of data. The output screen (below) allows you to select the way you would like the results formatted. This example shows results in the "Graphical Overview" format. Click on the "Format results" button when you are ready to view the results.

 RollerAdvancedBlastreduced.gif (35566 bytes)

The results of the database search are presented in three different formats.


1. The first format is a graphical representation of the results in which your input sequence (called the **query**) is represented as a numbered bar. Underneath the query bar, the matching sequences are also represented as colored bars (you can't see the colors on this copy, but you will see them on your computer screen). The color of the bar indicates the degree of similarity between the query and the match, and the length of the bar shows the length of the sequence that was noticeably similar to the query.
2. Below the graphical output is a list of sequences that match the query, giving their reference number in the database, a very brief description, a **score** (the higher the score, the more similar the sequences), and an **E value** (which gives the odds that the similarity between the query sequence and the match might have occurred purely by chance). You will notice that for all of the matches listed for *E. coli* rRNA, the odds that the similarity is due to chance is zero. The program has returned only the 100 best matches. In fact, *E. coli* 16S rRNA has significant similarity to each of the thousands of small subunit rRNAs ever sequenced. This is because small subunit rRNA is well conserved across all living creatures. This is one of the reasons that it can be used for identification and classification purposes.
3. Finally, the third data format provides the most information. It is a base-to-base alignment of the query sequence with each of its matches, starting from the "best match" and progressing down the list (in this case, all 100 sequence alignments are available for viewing, but we have chosen only


two examples). Each alignment has a score, and also a measurement of the identity between the sequences, presented as a fraction and a percent. The best match (the one at the top of the list) is, as you might expect, *E. coli* 16S rRNA. The identities value is 100%, meaning that there are no differences between the query and match sequences. If our query had been an unknown sequence and we had gotten this result, we could have inferred that our unknown organism was *E. coli* or a close relative.

Before answering the questions (below) for the two DNA sequences that we have e-mailed to you, familiarize yourself with the output formats. The matches denoted by arrows in the "List of sequences that match the query" are the two alignments that we have reproduced in the "Base-to-base alignment of the query sequence with two of its matches".

You will notice that for our example query (*E. coli* 16S rRNA), many *E. coli* matching sequences were returned. The 16S rRNA gene from *E. coli* has been sequenced many times from many different strains over the years, and most or all of these results have been deposited in GenBank. Some of these are partial sequences and so get lower scores than the full sequence (the score obtained is affected both by the percent of identical nucleotides and by the length of the match). As a result you can see some *E. coli* matches that have scores lower than other species. A rigorous comparison between two species should always involve the total length of the rRNA. You should also notice that some non-*E. coli* sequences were returned. This allows us to take a look at bacteria that are closely related to this species. In this example all non-*E. coli* species are distinguishable from *E. coli*. That is, there are no perfect matches that are not *E. coli*. This is not always the case. *Salmonella* species appear high on the list and often. Farther down the list, you will see *Citrobacter*, *Klebsiella*, and *Erwinia*. A few of the alignments have been reproduced for you to examine.

Graphical representation of the results

 RollerBlastResults.gif (6643 bytes)

 Rollerresults.gif (3166 bytes)

 RollerResultsgraph.gif (3008 bytes)

List of sequences that match the query

Sequences producing significant alignments:	Score (bits)	E Value
gb M87049 ECOUW85 E. coli genomic sequence of the region from 8...	3057	0.0
gb J01695 ECORGNB E.coli rRNA operon (rrnB) coding for Glu-tRNA...	3057	0.0
gb U72488 CVU72488 Cloning vector pRNA8, complete sequence	3057	0.0
gb U00006 ECOUW89 E. coli chromosomal region from 89.2 to 92.8 ...	3057	0.0
emb V00348 ECRRNBZ E. coli ribosomal operon rrnB encoding the 1...	3057	0.0
gb AE000471 AE000471 Escherichia coli K-12 MG1655 section 361 o...	3057	0.0
gb AE000474 AE000474 Escherichia coli K-12 MG1655 section 364 o...	3057	0.0
gb AE000460 AE000460 Escherichia coli K-12 MG1655 section 350 o...	3057	0.0

gb J01859 ECORRD E.coli 16S ribosomal RNA. >gi 640886 emb A1456...	3041	0.0
gb AE000452 AE000452 Escherichia coli K-12 MG1655 section 342 o...	2993	0.0
dbj D90887 D90887 E.coli genomic DNA, Kohara clone #437(58.6-59...	2993	0.0
gb L10328 ECOUW82 E. coli; the region from 81.5 to 84.5 minutes	2993	0.0
gb U18997 ECOUW67 Escherichia coli K-12 chromosomal region from...	2985	0.0
gb AE000406 AE000406 Escherichia coli K-12 MG1655 section 296 o...	2985	0.0
gb AE000129 AE000129 Escherichia coli K-12 MG1655 section 19 of...	2979	0.0
gb U70214 ECU70214 Escherichia coli chromosome minutes 4-6	2979	0.0
gb AE000345 AE000345 Escherichia coli K-12 MG1655 section 235 o...	2972	0.0
dbj D90886 D90886 E.coli genomic DNA, Kohara clone #436(58.4-58...	2966	0.0
emb Z83204 EC16SRRN E.coli 16S ribosomal RNA;	2948	0.0
emb Z83205 EC16SRRN1 E.coli 16S ribosomal RNA (EHEC Strain ATCC...	2932	0.0
gb U92196 SBU92196 Salmonella bareilly 16S ribosomal RNA gene, ...	2922	0.0
emb X96963 SF16SRRGE S.flexneri 16S rRNA gene	2918	0.0
emb X96964 SS16SRRGE S.sonnei 16S rRNA gene	2910	0.0
dbj D83536.2 ECOTSF Escherichia coli genomic DNA. (4.1 - 6.1 min)	2908	0.0
emb Z83203 HA16SRRN H.alvei 16S ribosomal RNA;	2898	0.0
emb X96965 SB16SRRGE S.boydii 16S rRNA gene	2886	0.0
gb U90317 SWU90317 Salmonella weltevreden sw1 16S ribosomal RNA...	2874	0.0
gb U92192 SCU92192 Salmonella chingola 16S ribosomal RNA gene, ...	2853	0.0
gb U92195 SHU92195 Salmonella houten 16S ribosomal RNA gene, co...	2851	0.0
emb X80721 ECRRNAPK3 E.coli rrnA gene	2831	0.0
gb U90316 STU90316 Salmonella typhimurium Stm1 16S ribosomal RN...	2827	0.0
emb X80722 ECRRNBPK3 E.coli rrnB gene	2817	0.0
emb X96966 SD16SRRGE S.dysenteriae 16S rRNA gene	2815	0.0
emb X80731 EC16SRPK3 E.coli (pk3) gene for 16S rRNA	2811	0.0
emb X80724 ECATC16SR E.coli (ATCC 25922) gene for 16S rRNA	2797	0.0
emb X80725 ECAT1177T E.coli (ATCC 11775T) gene for 16S rRNA	2781	0.0
emb X80679 SF16SRD S.flexneri 16S rRNA gene	2777	0.0
emb X80726 SSG16SR S.sonnei gene for 16S rRNA	2765	0.0
emb X80732 ECMC4100 E.coli (MC4100) gene for 16S rRNA	2759	0.0

emb X80723 ECRRNCPK3 E.coli rrnC gene	2757	0.0
emb X80729 ECRRNGPK3 E.coli rrnG gene	2726	0.0
emb X80727 ECRRNDPK3 E.coli rrnD gene	2708	0.0
emb X80728 ECRRNEPK3 E.coli rrnE gene	2690	0.0
emb X80730 ECRRNHPK3 E.coli rrnH gene	2682	0.0
emb X80733 ESP16SRRN Escherichia sp. gene for 16S rRNA	2678	0.0
gb AF029226 AF029226 Salmonella bongori strain JEO 4162 16S rib...	2676	0.0
gb AF025364 AF025364 Citrobacter sedlakii 16S ribosomal RNA gen...	2664	0.0
emb AJ010485.1 EAM010485 Erwinia amylovora 16S rRNA gene, tRNA-...	2545	0.0
gb U78184 KOU78184 Klebsiella oxytoca 16S ribosomal RNA gene, p...	2522	0.0
gb U78183 KOU78183 Klebsiella oxytoca 16S ribosomal RNA gene, p...	2522	0.0
emb X83265 EA16SRR E.amylovora 16S rRNA gene	2502	0.0
emb AJ233414.1 EMA233414 Erwinia mallotivora 16S rRNA gene (str...	2498	0.0
emb AJ233410.1 EAM233410 Erwinia amylovora 16S rRNA gene (strai...	2482	0.0
emb Y17661 KOY17661 Klebsiella oxytoca 16S rRNA gene, strain 58...	2482	0.0
emb Y17660 KOY17660 Klebsiella oxytoca 16S rRNA gene, strain 57...	2478	0.0
gb AF075271 AF075271 Alterococcus agarolyticus 16S ribosomal RN...	2476	0.0
gb U78182 KOU78182 Klebsiella ornithinolytica 16S ribosomal RNA...	2458	0.0
gb U39556 ESU39556 Enterobacter sp. 16S rRNA gene, partial sequ...	2450	0.0
emb Y17664 KSY17664 Klebsiella sp. 16S rRNA gene, strain BEC441...	2450	0.0
emb Y17655 KOY17655 Klebsiella oxytoca 16S rRNA gene, strain AT...	2428	0.0
emb Z96088 EAZ96088 Erwinia amylovora LMG 2024 16S ribosomal RNA	2424	0.0
emb Y17666 KOY17666 Klebsiella ornithinolytica 16S rRNA gene, s...	2407	0.0
emb Z96093 ECZ96093 Erwinia chrysanthemi LMG 2804 16S ribosomal...	2403	0.0
emb Y17662 KOY17662 Klebsiella ornithinolytica 16S rRNA gene, s...	2403	0.0
emb AJ233412.1 ECH233412 Erwinia chrysanthemi 16S rRNA gene (st...	2399	0.0
gb U80195 EAU80195 Erwinia amylovora 16S ribosomal RNA gene, pa...	2393	0.0
emb AJ233416.1 EQU233416 Erwinia quercina 16S rRNA gene (strain...	2387	0.0
gb AF015258 AF015258 Aquamonas haywardensis 16S ribosomal RNA g...	2385	0.0
emb Z96084 EMZ96084 Erwinia mallotivora LMG 2708 16S ribosomal RNA	2385	0.0
dbj AB004756 AB004756 Klebsiella ornithinolytica gene for 16S r...	2371	0.0

emb Z96090 ECZ96090 <i>Erwinia carotovora</i> LMG 2386 16S ribosomal RNA	2365	0.0
gb U20275 BAU20275 Unidentified bacteria of host <i>D. paulistorum</i> ...	2359	0.0
gb U80198 ECU80198 <i>Erwinia carotovora</i> subsp. <i>betavasculorum</i> 16S...	2359	0.0
emb AJ223469 BQ16SRRN <i>Brenneria quercina</i> ex <i>Erwinia quercina</i> 1...	2357	0.0
emb X93215 KP16SRRN1 <i>K. planticola</i> 16S rRNA gene (strain DSM 306...	2337	0.0
emb Z96092 ECZ96092 <i>Erwinia cacticida</i> LMG 2720 16S ribosomal RNA	2331	0.0
gb U80199 ECU80199 <i>Erwinia carotovora</i> subsp. <i>wasabiae</i> 16S ribos...	2323	0.0
gb U80200 ECU80200 <i>Erwinia chrysanthemi</i> 16S ribosomal RNA gene,...	2309	0.0
emb Z96094 ECZ96094 <i>Erwinia cyripedii</i> LMG 2657 16S ribosomal RNA	2254	0.0
gb AF076038 AF076038 <i>Serratia marcescens</i> 16S ribosomal RNA gene...	2198	0.0
emb AJ233419.1 ESA233419 <i>Erwinia salicis</i> 16S rRNA gene (strain ...	2189	0.0
gb AF076037 AF076037 <i>Escherichia coli</i> 16S ribosomal RNA gene, p...	2187	0.0
emb Z96097 ESZ96097 <i>Erwinia salicis</i> LMG 2698 16S ribosomal RNA	2155	0.0
gb U88546 SPU88546 <i>Salmonella paratyphi</i> A 16S ribosomal RNA gen...	2095	0.0
gb U92197 SAU92197 <i>Salmonella agona</i> 16S ribosomal RNA gene, com...	2066	0.0
gb U90314 SBU90314 <i>Salmonella blockley</i> Sb1 16S ribosomal RNA ge...	2042	0.0
gb U88545 STU88545 <i>Salmonella typhi</i> 16S ribosomal RNA gene, com...	2024	0.0
gb U88547 SPU88547 <i>Salmonella paratyphi</i> B 16S ribosomal RNA gen...	2016	0.0
gb U92193 SBU92193 <i>Salmonella bovis morbificans</i> 16S ribosomal R...	2016	0.0
gb U92194 SWU92194 <i>Salmonella waycross</i> 16S ribosomal RNA gene, ...	2016	0.0
gb U90315 SMU90315 <i>Salmonella matopeni</i> Sm1 16S ribosomal RNA ge...	2008	0.0
gb AF025366 AF025366 <i>Citrobacter diversus</i> strain CDC 8132-86 16...	1992	0.0
gb U88548 SPU88548 <i>Salmonella paratyphi</i> C 16S ribosomal RNA gen...	1968	0.0
gb AF025372 AF025372 <i>Citrobacter diversus</i> strain CDC 3613-63 16...	1953	0.0
dbj AB004755 AB004755 <i>Klebsiella planticola</i> gene for 16S riboso...	1933	0.0
emb AJ233432.1 SOD233432 <i>Serratia odorifera</i> 16S rRNA gene (stra...	1820	0.0
gb U65654 CLU65654 <i>Cimex lectularius</i> endosymbiont 16S rRNA gene...	1804	0.0
gb AF130915.1 AF130915 <i>Enterobacter agglomerans</i> strain A44 16S ...	1768	0.0
gb U31490 UGU31490 Unidentified gamma proteobacterium S11 16S r...	1610	0.0
emb AJ223408 PCA223408 <i>Pectobacterium carotovorum</i> subsp. <i>wasabi</i> ...	1576	0.0

Base-to-base alignment of the query sequence with two of its matches

gb|M87049|ECOUW85 E. coli genomic sequence of the region from 84.5 to 86.5 minutes.
 Length = 91414
 Score = 3057 bits (1542), Expect = 0.0
 Identities = 1542/1542 (100%)
 Strand = Plus / Plus

```

Query: 1  aaattgaagagtttgatcatggctcagattgaacgctggcggcaggcctaacacatgcaa60
          |||
Sbjct:88591 aaattgaagagtttgatcatggctcagattgaacgctggcggcaggcctaacacatgcaa 88650
Query: 61  gtcgaacggtaacaggaagaagcttgcttctttgctgacgagtgccggacgggtgagtaa120
          |||
Sbjct:88651 gtcgaacggtaacaggaagaagcttgcttctttgctgacgagtgccggacgggtgagtaa 88710
Query: 121  tgtctgggaaactgcctgatggagggggataactactggaaacggtagctaataaccgcat180
          |||
Sbjct:88711 tgtctgggaaactgcctgatggagggggataactactggaaacggtagctaataaccgcat 88770
Query: 181  aacgtcgcaagaccaaagagggggaccttcgggcctcttgccatcggatgtgccagatg240
          |||
Sbjct:88771 aacgtcgcaagaccaaagagggggaccttcgggcctcttgccatcggatgtgccagatg 88830
Query: 241  ggattagctagtagtggtgggtaacggctcacctaggcgacgatccctagctggtctgaga300
          |||
Sbjct:88831 ggattagctagtagtggtgggtaacggctcacctaggcgacgatccctagctggtctgaga 88890
Query: 301  ggatgaccagccacactggaactgagacacggctccagactcctacgggagggcagcagtg360
          |||
Sbjct:88891 ggatgaccagccacactggaactgagacacggctccagactcctacgggagggcagcagtg 88950
Query: 361  ggaatattgcacaatggcgcaagcctgatgcagccatgcccgctgtatgaagaaggcct420
          |||
Sbjct:88951 ggaatattgcacaatggcgcaagcctgatgcagccatgcccgctgtatgaagaaggcct 89010
Query: 421  tcgggttgtaaaagtactttcagcggggaggaaggagtaaaagttaatacctttgctcatt480
          |||
Sbjct:89011 tcgggttgtaaaagtactttcagcggggaggaaggagtaaaagttaatacctttgctcatt 89070
Query: 481  gacgttaccgcgagaagaagcaccggctaactccgtgccagcagccggtaatacggag540
          |||
Sbjct:89071 gacgttaccgcgagaagaagcaccggctaactccgtgccagcagccggtaatacggag 89130
Query: 541  ggtgcaagcgttaactcggaattactggcgtaaaagcgcacgcaggcggttgtaagtca600
          |||
Sbjct:89131 ggtgcaagcgttaactcggaattactggcgtaaaagcgcacgcaggcggttgtaagtca 89190
Query: 601  gatgtgaaatccccgggctcaacctgggaactgcatctgatactggcaagccttgagtctc660
          |||
Sbjct:89191 gatgtgaaatccccgggctcaacctgggaactgcatctgatactggcaagccttgagtctc 89250
Query: 661  gtagaggggggtagaattccagggtgtagcggtagaaatgcgtagagatctggaggaatacc720
          |||
Sbjct:89251 gtagaggggggtagaattccagggtgtagcggtagaaatgcgtagagatctggaggaatacc 89310
Query: 721  ggtggcgaaggcggccccctggacgaagactgacgctcaggtgcaaaagcgtggggagca780
          |||
Sbjct:89311 ggtggcgaaggcggccccctggacgaagactgacgctcaggtgcaaaagcgtggggagca 89370
Query: 781  aacaggattagataccctggtagtccacgcgtaaacgatgtcgacttggaggttggtgcc840
          |||
Sbjct:89371 aacaggattagataccctggtagtccacgcgtaaacgatgtcgacttggaggttggtgcc 89430
Query: 841  cttgagggcgtggcttccggagctaacgcgttaagtgcaccgctggggagtagcggccgca900
          |||
Sbjct:89431 cttgagggcgtggcttccggagctaacgcgttaagtgcaccgctggggagtagcggccgca 89490
Query: 901  aggttaaaactcaaatgaattgacggggggcccgacaaagcgtggagcatgtggtttaa960
          |||
Sbjct:89491 aggttaaaactcaaatgaattgacggggggcccgacaaagcgtggagcatgtggtttaa 89550
Query: 961  tcgatgcaacgcgaagaaccttacctggctcttgacatccacggaagtttccagagatgag1020
          |||
Sbjct:89551 tcgatgcaacgcgaagaaccttacctggctcttgacatccacggaagtttccagagatgag 89610
Query: 1021  aatgtgccttcgggaaccgtgagacaggtgctgcatggctgtcgtcagctcgtgttgta1080
          |||
Sbjct:89611 aatgtgccttcgggaaccgtgagacaggtgctgcatggctgtcgtcagctcgtgttgta 89670
Query: 1081  aatgttgggttaagtcccgcacagcgcacaccttatcctttggtgccagcgggtccggc1140
          |||
Sbjct:89671 aatgttgggttaagtcccgcacagcgcacaccttatcctttggtgccagcgggtccggc 89730
Query: 1141  cgggaactcaaggagactgccagtgataaactggaggaaggtggggatgacgtcaagtc1200
          |||
Sbjct:89731 cgggaactcaaggagactgccagtgataaactggaggaaggtggggatgacgtcaagtc 89790
Query: 1201  atcatggcccttacgaccagggtcacacagtgctacaatggcgcatacaagagaagcg1260
          |||
Sbjct:89791 atcatggcccttacgaccagggtcacacagtgctacaatggcgcatacaagagaagcg 89850

```

```

Query:1261acctcgcgagagcaagcggacctcataaagtgcgtcgtagtcggatggagctgcaac1320
      |||
Sbjct:89851acctcgcgagagcaagcggacctcataaagtgcgtcgtagtcggatggagctgcaac89910
Query:1321tcgactccatgaagtcggaatcgctagtaatcgtggatcagaatgccacggtgaatacgt1380
      |||
Sbjct:89911tcgactccatgaagtcggaatcgctagtaatcgtggatcagaatgccacggtgaatacgt89970
Query:1381tccccggccttgtacacaccgcccgtcacaccatgggagtggggtgcaaaagaagtaggt1440
      |||
Sbjct:89971tccccggccttgtacacaccgcccgtcacaccatgggagtggggtgcaaaagaagtaggt90030
Query:1441agcttaaccttcgggagggcgcttaccactttgtgattcatgactggggtgaagtcgtaa1500
      |||
Sbjct:90031agcttaaccttcgggagggcgcttaccactttgtgattcatgactggggtgaagtcgtaa90090
Query:1501caaggtaacctgtaggggaacctgcggttgatcacctcctta1542
      |||
Sbjct:90091caaggtaacctgtaggggaacctgcggttgatcacctcctta90132

```

gb|U92196|SBU92196 *Salmonella bareilly* 16S ribosomal RNA gene, complete sequence

Length = 1541

Score = 2922 bits (1474), Expect = 0.0

Identities = 1526/1542 (98%), Gaps = 1/1542 (0%)

Strand = Plus / Plus

```

Query: 1 aaattgaagagtttgatcatggctcagattgaacgctggcggcaggcctaacacatgcaa60
      |||
Sbjct: 1 aaattgaagagtttgatcatggctcagattgaacgctggcggcaggcctaacacatgcaa60
Query: 61 gtcgaacggtaacaggaagaagcttgcttcttggctgacgagtgccggacgggtgagtaa120
      |||
Sbjct: 61 gtcgaacggtaacaggaagcagcttgct-cttggctgacgagtgccggacgggtgagtaa119
Query: 121 tgtctgggaaactgcctgatggagggggataaactactggaacggtagctaataaccgcat180
      |||
Sbjct: 120 tgtctgggaaactgcctgatggagggggataaactactggaacggtagctaataaccgcat179
Query: 181 aacgtcgcaagaccaagagggggaccttcgggcctcctgccatcggatgtgccagatg240
      |||
Sbjct: 180 aacgtcgcaagaccaagagggggaccttcgggcctcctgccatcggatgtgccagatg239
Query: 241 ggattagctagtaggtggggtaacggctcacctaggcgcagatccctagctggtctgaga300
      |||
Sbjct: 240 ggattagctagtaggtggggtaacggctcacctaggcgcagatccctagctggtctgaga299
Query: 301 ggatgaccagccacactggaactgagacacggtccagactcctacgggaggcagcagtg360
      |||
Sbjct: 300 ggatgaccagccacactggaactgagacacggtccagactcctacgggaggcagcagtg359
Query: 361 ggaatatgcaacaatgggcgcaagcctgatgcagccatgccgcgtgtatgaagaaggcct420
      |||
Sbjct: 360 ggaatatgcaacaatgggcgcaagcctgatgcagccatgccgcgtgtatgaagaaggcct419
Query: 421 tcgggttgtaaagtactttcagcggggaggaaggagtaagtttaaccttgctcatt480
      |||
Sbjct: 420 tcgggttgtaaagtactttcagcggggaggaagggtgtgtggttaacacccagctaatt479
Query: 481 gacgttacccgcagaagaagcaccggctaactccgtgccagcagcccggttaatacggag540
      |||
Sbjct: 480 gacgttacccgcagaagaagcaccggctaactccgtgccagcagcccggttaatacggag539
Query: 541 ggtgcaagcgttaactcggaattactgggcgtaaacgcgcagcagggcgtttgttaagtc600
      |||
Sbjct: 540 ggtgcaagcgttaactcggaattactgggcgtaaacgcgcagcagggcgtttgttaagtc599
Query: 601 gatgtgaaatccccgggctcaacctgggaactgcatctgatactggcaagcttgagtctc660
      |||
Sbjct: 600 gatgtgaaatccccgggctcaacctgggaactgcatctgatactggcaagcttgagtctc659
Query: 661 gtagaggggggtagaattccaggtgtagcgggtgaaatgcgtagagatctggaggaatacc720
      |||
Sbjct: 660 gtagaggggggtagaattccaggtgtagcgggtgaaatgcgtagagatctggaggaatacc719
Query: 721 ggtggcgaaggcggccccctggacgaagactgacgctcaggtgcaaaagcgtggggagca780
      |||
Sbjct: 720 ggtggcgaaggcggccccctggacgaagactgacgctcaggtgcaaaagcgtggggagca779
Query: 781 aacaggattagataccctggtagtccacgcccgtaaacgatgctgactggaggttggtgcc840
      |||
Sbjct: 780 aacaggattagataccctggtagtccacgcccgtaaacgatgctgactggaggttggtgcc839
Query: 841 cttgaggcgtggcctccggagctaaccgcttaagtcgaccgctggggagtagcggccgca900
      |||
Sbjct: 840 cttgaggcgtggcctccggagctaaccgcttaagtcgaccgctggggagtagcggccgca899

```

```

Query: 901 aggttaaaactcaaataaatgacgggggcccgcacaagcgggtggagcatgtggtttaat960
          |||
Sbjct: 900 aggttaaaactcaaataaatgacgggggcccgcacaagcgggtggagcatgtggtttaat959
Query: 961 tcgatgcaacgcgaagaaccttacctgggtcttgacatccacggaagtttccagagatgag1020
          |||
Sbjct: 960 tcgatgcaacgcgaagaaccttacctgggtcttgacatccacggaagtttccagagatgag1019
Query:1021 aatgtgccttcgggaaccgtgagacaggtgctgcatggctgctcgtcagctcgtgtgtga1080
          |||
Sbjct: 1020 aatgtgccttcgggaaccgtgagacaggtgctgcatggctgctcgtcagctcgtgtgtga1079
Query:1081 aatgttgggttaagtcccgcacagcgcaacccttatccttgggtgcccagcgggtccggc1140
          |||
Sbjct: 1080 aatgttgggttaagtcccgcacagcgcaacccttatccttgggtgcccagcgggtccggc1139
Query:1141 cgggaactcaaaggagactgccagtgataaactggaggaaggtggggatgacgtcaagt1200
          |||
Sbjct: 1140 cgggaactcaaaggagactgccagtgataaactggaggaaggtggggatgacgtcaagt1199
Query:1201 atcatggcccttacgaccagggctacacacgtgctacaatggcgcatacaaagagaagcg1260
          |||
Sbjct: 1200 atcatggcccttacgaccagggctacacacgtgctacaatggcgcatacaaagagaagcg1259
Query:1261 acctcgcgagagcaagcggacctcataaagtgcgtcgtagtccggattggagctgcaac1320
          |||
Sbjct: 1260 acctcgcgagagcaagcggacctcataaagtgcgtcgtagtccggattggagctgcaac1319
Query:1321 tcgactccatgaagtcggaatcgctagtaaatcggtgatcagaatgccacgggtgaatacgt1380
          |||
Sbjct: 1320 tcgactccatgaagtcggaatcgctagtaaatcggtgatcagaatgccacgggtgaatacgt1379
Query:1381 tcccgggccttgtagcacaccgcccgtcacaccatgggagtggggtgcaaaagaagtaggt1440
          |||
Sbjct: 1380 tcccgggccttgtagcacaccgcccgtcacaccatgggagtggggtgcaaaagaagtaggt1439
Query:1441 agcttaaccttcgggagggcgcttaccactttgtgattcatgactggggtgaagtcgtaa1500
          |||
Sbjct: 1440 agcttaaccttcgggagggcgcttaccactttgtgattcatgactggggtgaagtcgtaa1499
Query:1501 caaggtaaccgtaggggaacctgcggttggatcacctcctta1542
          |||
Sbjct:1500 caaggtaaccgtaggggaacctgcggttggatcacctcctta1541

```

Questions

Answer the following questions.

Take a look at the sample results returned using *E. coli* rRNA as the query. In your text or lab manual, find a diagram in which the gram negative organisms and specifically the enterics are grouped according to physiological criteria.

From these criteria, which organisms would you have expected to be most closely related to *E. coli*?

Is there a perfect correspondence between the groupings based on physiological markers and the results of the sequence comparison?

Judging from these results, are physiological criteria at all predictive of phylogenetic relatedness? If so, give examples of structural or physiological properties that *E. coli* does share with its close phylogenetic relatives.

Take a look at the alignment between the *E. coli* and *Salmonella bareilly* (the best non-*E. coli* match). How many nucleotide differences in 16S rRNA differentiate these two species?

Examine the results returned for your unknown sequence 1. What is the top match returned for your query sequence?

Is there 100% identity between your query sequence and the top match?

Is there 100% identity between your query sequence and any of the other matches? If so, are these multiple entries for the same species or entries for different species? If different species, do you have any basis for identifying your unknown with one species?

Do you have matches in which there is less than 100% identity between your query and the match? If so, what are the closest phylogenetic relatives to your unknown predicted by this method?

Examine the results returned for your unknown sequence 2. What is the top match returned for your query sequence?

Is there 100% identity between your query sequence and the top match?

Is there 100% identity between your query sequence and any of the other matches? If so, are these multiple entries for the same species or entries for different species? If different species, do you have any basis for identifying your unknown with one species?

Do you have matches in which there is less than 100% identity between your query and the match? If so, what are the closest phylogenetic relatives to your unknown predicted by this method?