UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING or 371(c) DATE	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
61/652,086	05/25/2012		305	BERK-187PRV		

CONFIRMATION NO. 2189

REPLACEMENT FILING RECEIPT

84220
UC Berkeley - OTL
Bozicevic, Field & Francis LLP
1900 University Avenue, Suite 200
East Palo Alto, CA 94303

OC000000060169402

Date Mailed: 04/01/2013

Receipt is acknowledged of this provisional patent application. It will not be examined for patentability and will become abandoned not later than twelve months after its filing date. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF APPLICANT, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection. Please verify the accuracy of the data presented on this receipt. **If an error is noted on this Filing Receipt, please submit a written request for a Filing Receipt Correction. Please provide a copy of this Filing Receipt with the changes noted thereon. If you received a "Notice to File Missing Parts" for this application, please submit any corrections to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections**

Inventor(s)

Martin Jinek, Berkeley, CA;
Jennifer A. Doudna, Berkeley, CA;
Emmanuelle Charpentier, Berkeley, CA;
Krzysztof Chylinski, Berkeley, CA;

Applicant(s)

Martin Jinek, Berkeley, CA;
Jennifer A. Doudna, Berkeley, CA;
Emmanuelle Charpentier, Berkeley, CA;
Krzysztof Chylinski, Berkeley, CA;

Power of Attorney:
Paula Borden--42344

Permission to Access - A proper **Authorization to Permit Access to Application by Participating Offices** (PTO/SB/39 or its equivalent) has been received by the USPTO.

If Required, Foreign Filing License Granted: 06/11/2012

The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 61/652,086**

Projected Publication Date: None, application is not eligible for pre-grant publication

Non-Publication Request: No

Early Publication Request: No
**** SMALL ENTITY ****

Title

METHODS AND COMPOSITIONS FOR RNA-DIRECTED SITE-SPECIFIC DNA MODIFICATION

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications:

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4158).

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier

license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

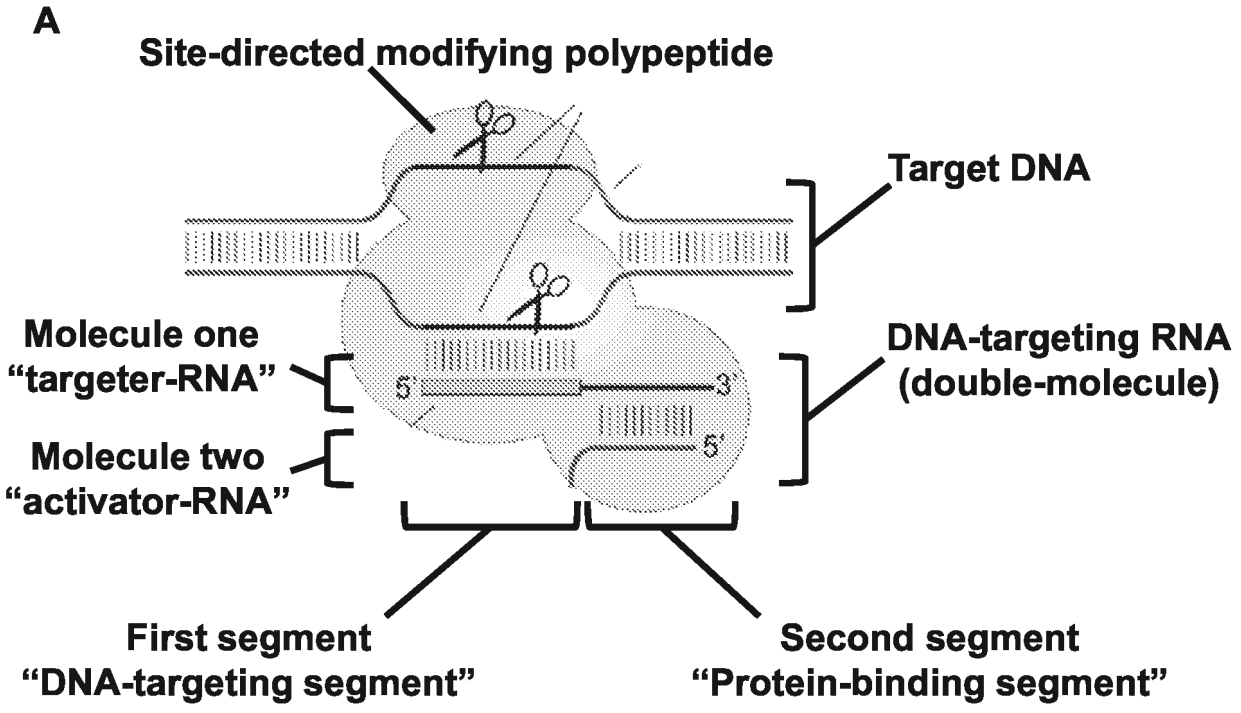
NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

FIGURE 1


A
Site-directed modifying polypeptide
Target DNA
Molecule one
"targeter-RNA"
Molecule two
"activator-RNA"
5'
3'
5'
DNA-targeting RNA
(double-molecule)
First segment
"DNA-targeting segment"
Second segment
"Protein-binding segment"

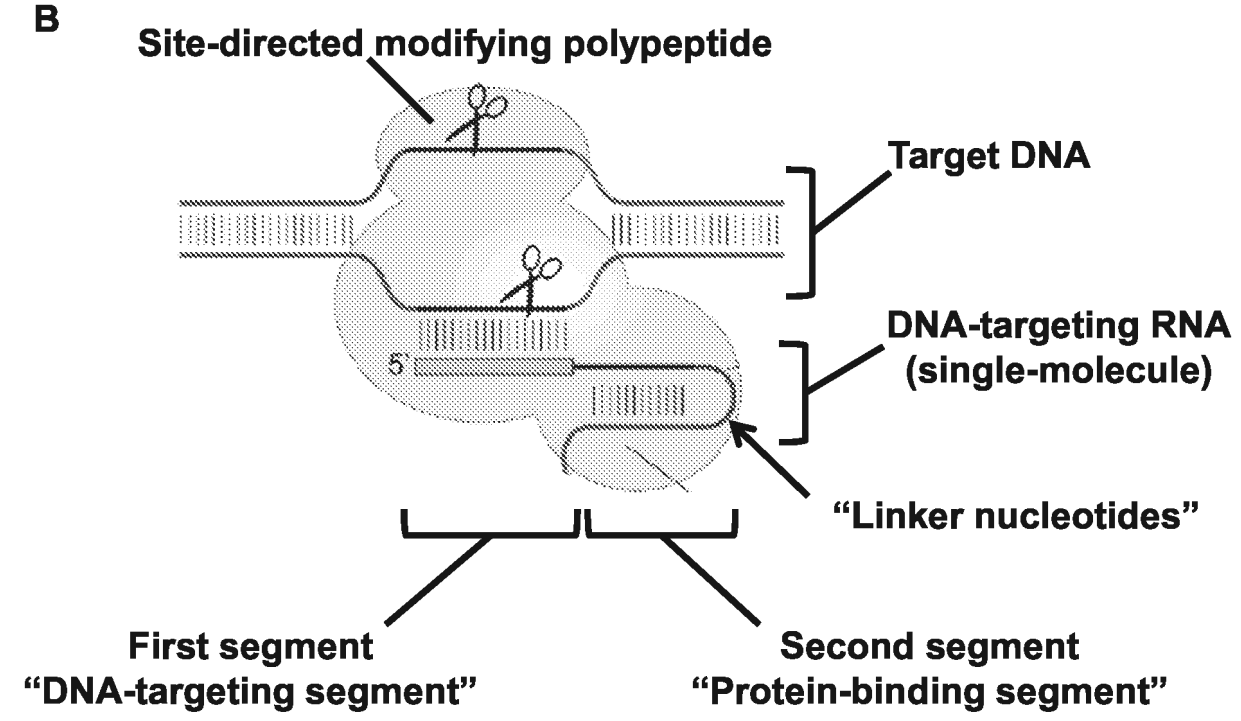

B
Site-directed modifying polypeptide
Target DNA
5'
DNA-targeting RNA
(single-molecule)
"Linker nucleotides"
First segment
"DNA-targeting segment"
Second segment
"Protein-binding segment"

FIGURE 2

Cas9/Csn1 Streptococcus pyogenes (Total length = 1368)

Domain 1: RuvC endonuclease domain motif 1 (7-166)

*MDKKYS**IGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALL***
FDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE
SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIY
***LALAHMIKFRG**HFLIEGDLNPDNDVDKLFIQLVQTYNQLFEENPINASGVDA*
KAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAED
AKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITK
APLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGG
ASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGEL
HAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEE
TITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYN
ELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIEC
FDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFED
REMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTI
LDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS
PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRER
MKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDIN
RLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKN
YWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA
QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHH
***AHDAYLNAVVGTALIKKYPK**LESEFVYGDYKVYDVRKMIAKSEQEIGKATA*
KYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL
SMPQVNIVKKTEVQTGGFSKESILPKRNDKLIARKKDWDPKKYGGFDSPTV
AYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKK
DLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEK
LKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNK
HRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQ
SITGLYETRIDLSQLGGD

Domain 2: RuvC/HNH/RuvC endonuclease domain motifs (731-1003)

FIGURE 3

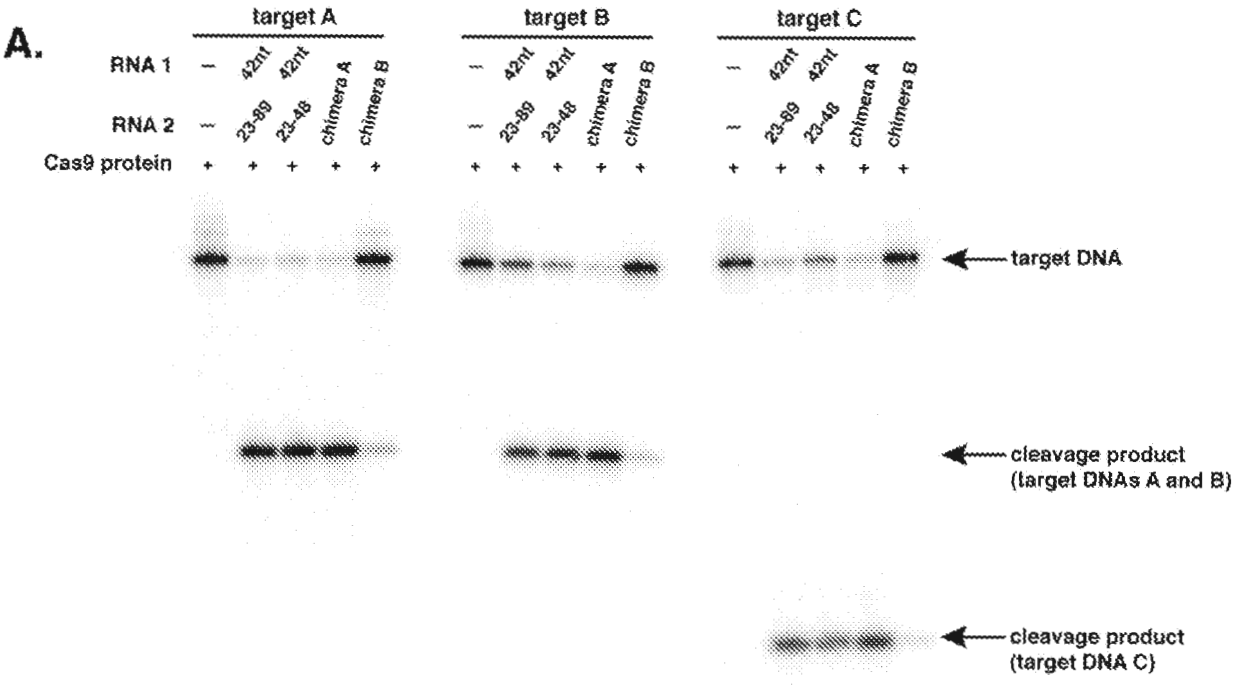

A.
target A
target B
target C
RNA 1
RNA 2
Cas9 protein
–
42nt
42nt
chimera A
chimera B
23-89
23-48
+
target DNA
cleavage product
(target DNAs A and B)
cleavage product
(target DNA C)

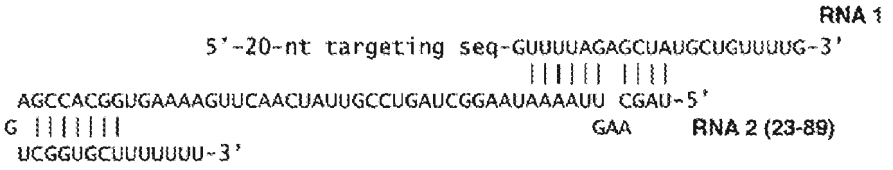
FIGURE 3

B.

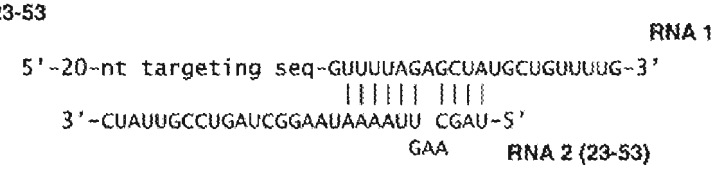
Structures of targeting RNAs

Two-molecule DNA-targeting RNAs

RNA 1: 42 nts
RNA 2: nts 23-89

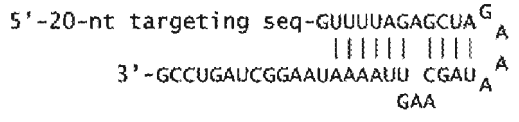


RNA1: 42 nt
RNA2: nts 23-53



Single-molecule DNA-targeting RNAs

chimera A



chimera B

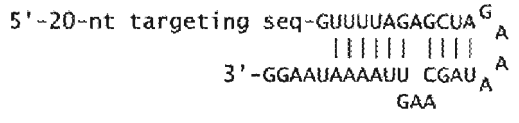
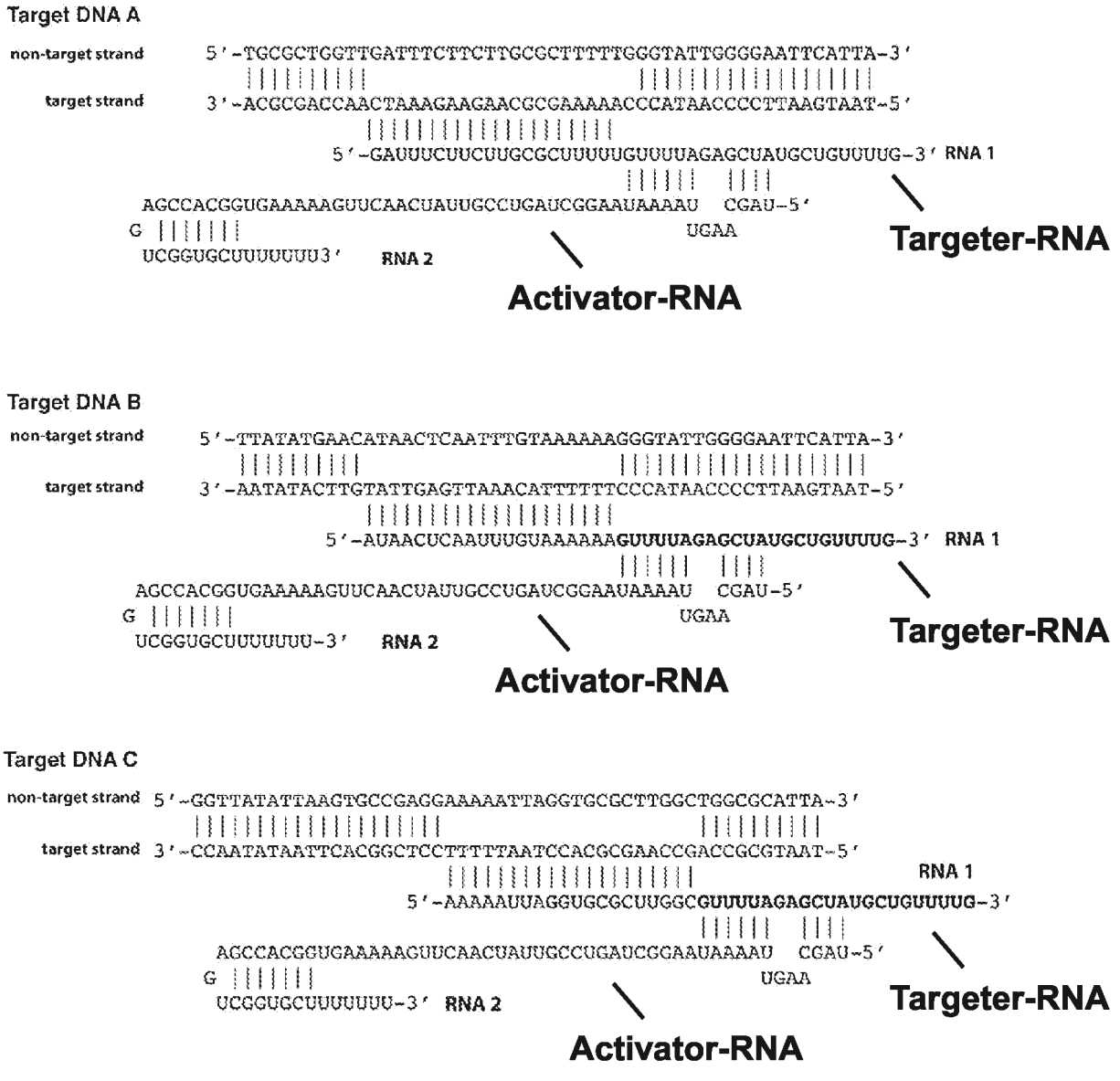


FIGURE 3

C.


Target DNA A
non-target strand
5'-TGCGCTGGTTGATTTCTTCTTGCGCTTTTTGGGTATTGGGGAATTCATTA-3'
target strand
3'-ACGCGACCAACTAAAGAAGAACGCGAAAAACCCATAACCCCTTAAGTAAT-5'
5'-GAUUUCUUCUUGCGCUUUUUGUUUUAGAGCUAUGCUGUUUUG-3' RNA 1
AGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUCGGAAUAAAAU CGAU-5'
G
UCGGUGCUUUUUUU3'
RNA 2
UGAA
Targeter-RNA
Activator-RNA
Target DNA B
non-target strand
5'-TTATATGAACATAACTCAATTTGTAAAAAAGGGTATTGGGGAATTCATTA-3'
target strand
3'-AATATACTTGTATTGAGTTAAACATTTTTTCCCATAACCCCTTAAGTAAT-5'
5'-AUAACUCAAUUUGUAAAAAAGUUUUAGAGCUAUGCUGUUUUG-3' RNA 1
AGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUCGGAAUAAAAU CGAU-5'
G
UCGGUGCUUUUUUU-3'
RNA 2
UGAA
Targeter-RNA
Activator-RNA
Target DNA C
non-target strand
5'-GGTTATATTAAGTGCCGAGGAAAAATTAGGTGCGCTTGGCTGGCGCATTA-3'
target strand
3'-CCAATATAATTCACGGCTCCTTTTTAATCCACGCGAACCGACCGCGTAAT-5'
RNA 1
5'-AAAAAUUAGGUGCGCUUGGCGUUUUAGAGCUAUGCUGUUUUG-3'
AGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUCGGAAUAAAAU CGAU-5'
G
UCGGUGCUUUUUUU-3'
RNA 2
UGAA
Targeter-RNA
Activator-RNA

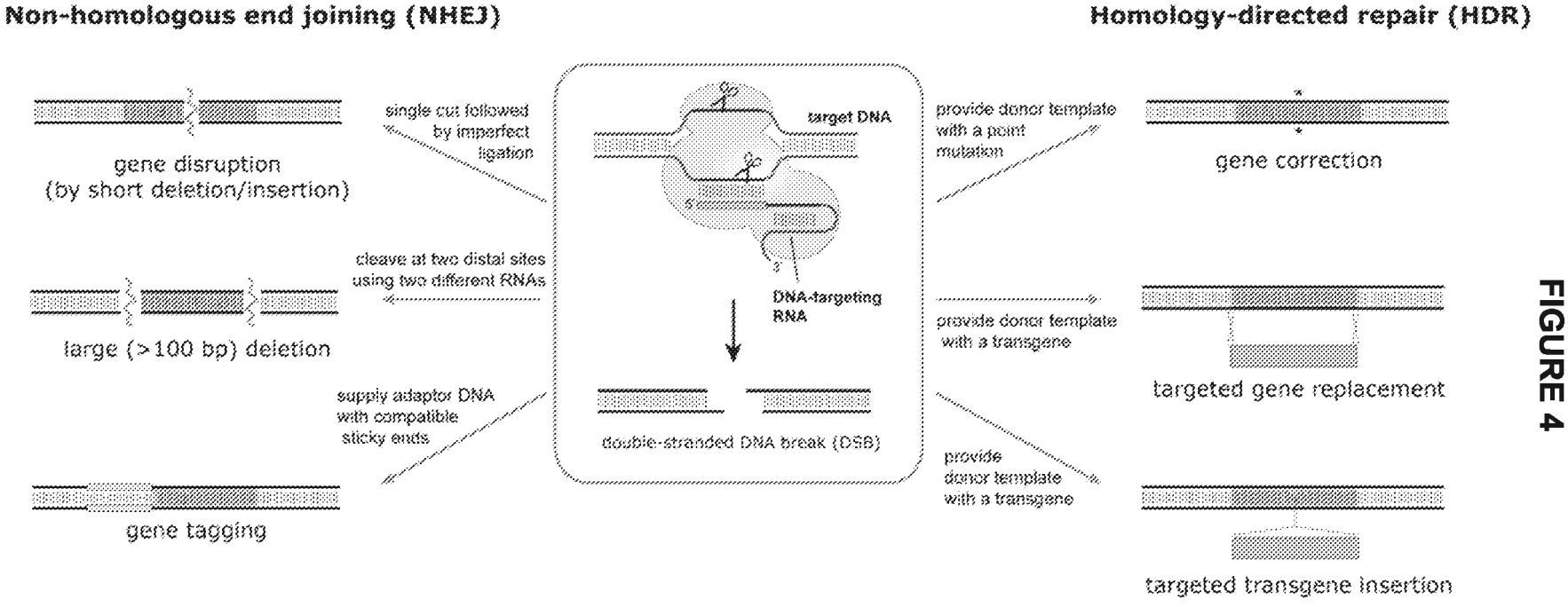
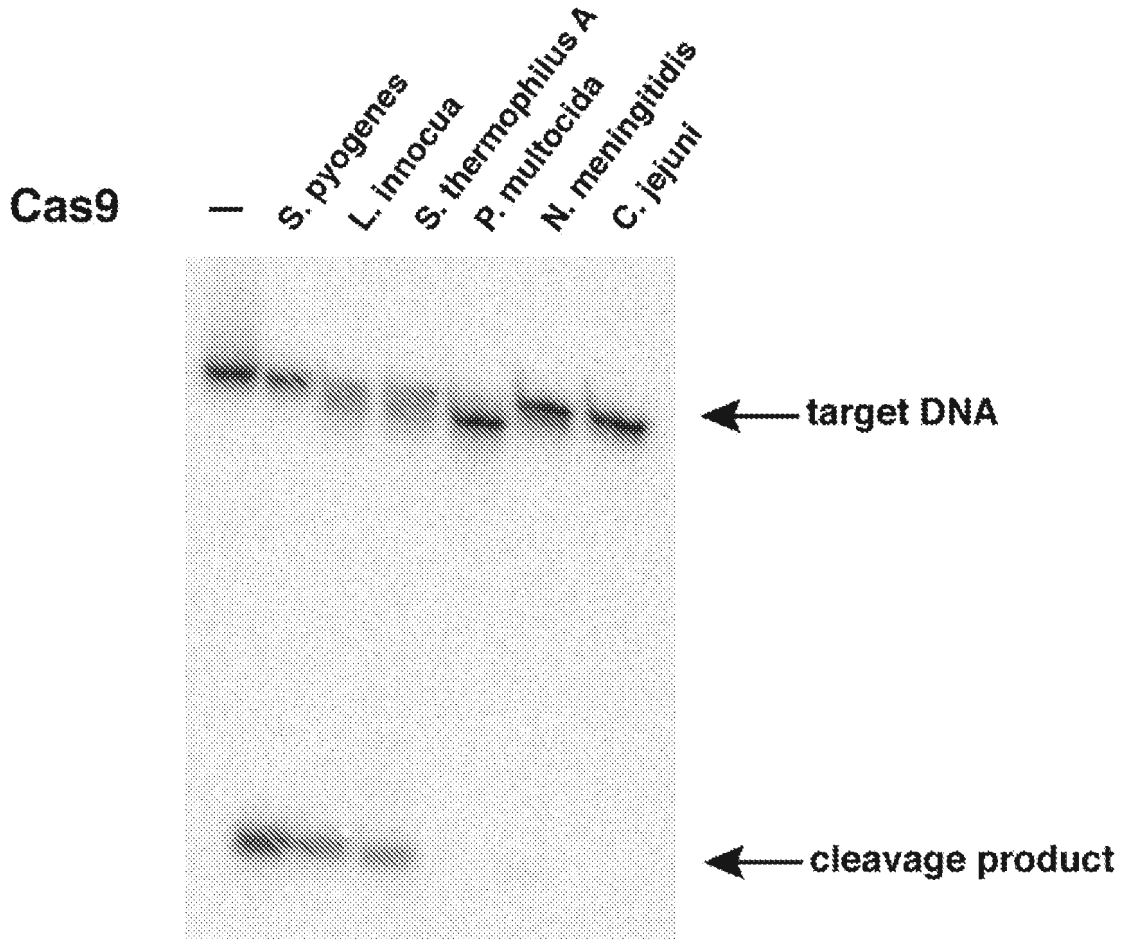

Non-homologous end joining (NHEJ)
Homology-directed repair (HDR)
gene disruption
(by short deletion/insertion)
single cut followed
by imperfect
ligation
cleave at two distal sites
using two different RNAs
large (>100 bp) deletion
supply adaptor DNA
with compatible
sticky ends
gene tagging
target DNA
DNA-targeting
RNA
double-stranded DNA break (DSB)
provide donor template
with a point
mutation
provide donor template
with a transgene
provide
donor template
with a transgene
gene correction
targeted gene replacement
targeted transgene insertion

FIGURE 4

FIGURE 5

A.


Cas9
—
S. pyogenes
L. innocua
S. thermophilus A
P. multocida
N. meningitidis
C. jejuni
target DNA
cleavage product

B.

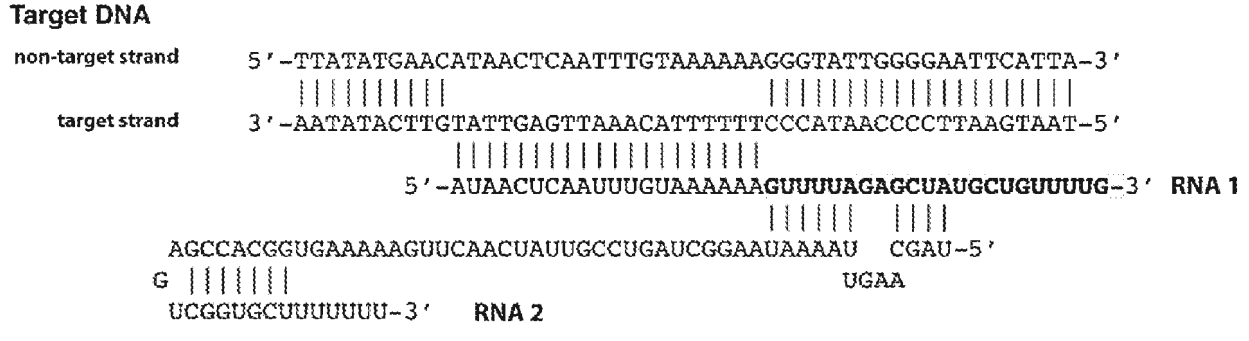

Target DNA
non-target strand
5′-TTATATGAACATAACTCAATTTGTAAAAAAGGGTATTGGGGAATTCATTA-3′
target strand
3′-AATATACTTGTATTGAGTTAAACATTTTTTCCCATAACCCCTTAAGTAAT-5′
5′-AUAACUCAAUUUGUAAAAAAGUUUUAGAGCUAUGCUGUUUUG-3′ RNA 1
AGCCACGGUGAAAAAGUUCAACUAUUGCCUGAUCGGAAUAAAAU CGAU-5′
G
UGAA
UCGGUGCUUUUUUU-3′ RNA 2

FIGURE 6

A. tracrRNA sequences ("activator-RNA")

>*S. pyogenes* SF370
GUUGGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCG
UUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU

>*L. innocua* Clip11262
AUAUUGUUAGUAUUCAAAAUAACAUAGCAAGUUAAAAUAAGGCUUUGU
CCGUUAUCAACUUUUAAUUAAGUAGCGCUGUUUCGGCGCUUUUUU

>*S. mutans* UA159
UGUUGGAAUCAUUCGAAACAACACAGCAAGUUAAAAUAAGGCAGUGAU
UUUUAAUCCAGUCCGUACACAACUUGAAAAAGUGCGCACCGAUUCGG
UGCUUUUUU

>*S. thermophilus* LMD-9 (locus 1)
UUGUGGUUUGAAACCAUUCGAAACAACACAGCGAGUUAAAAUAAGGC
UUAGUCCGUACUCAACUUGAAAAGGUGGCACCGAUUCGGUGUUUUUU
U

>*S. thermophilus* LMD-9 (locus 2)
UAAUAAUAGUGUAAGGGACGCCUUACACAGUUACUUAAAUCUUGCAG
AAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUU
AUGGCAGGGUGUUUUCGUUAUUU

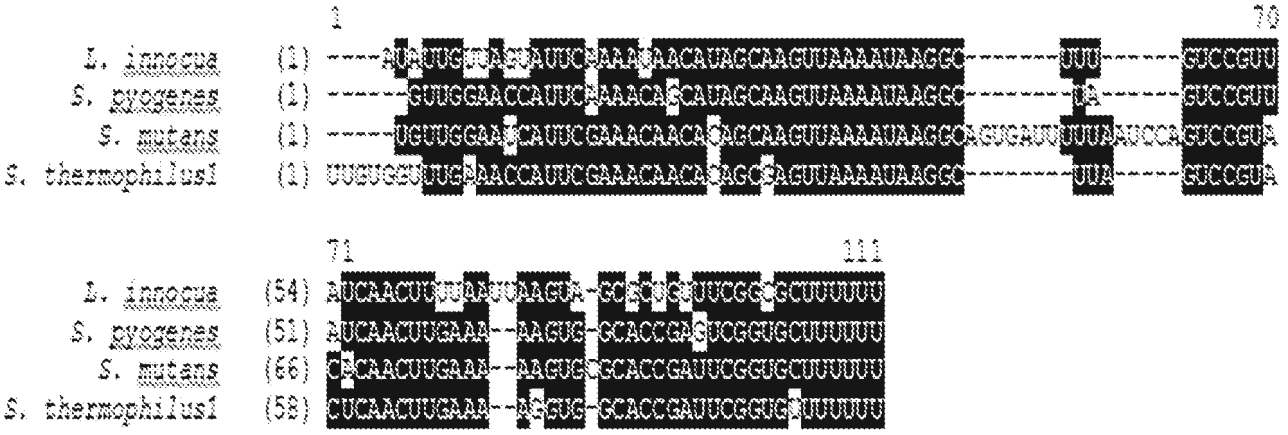
>*N. meningitidis* Z2491
CAUAUUGUCGCACUGCGAAAUGAGAACCGUUGCUACAAUAAGGCCGU
CUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUAAAGCUUCUGCUUUA
AGGGGCAUCGUUUAUUUC

>*P. multocida* Pm70
GCAUAUUGUUGCACUGCGAAAUGAGAGACGUUGCUACAAUAAGGCUU
CUGAAAAGAAUGACCGUAACGCUCUGCCCCUUGUGAUUCUUAAUUGC
AAGGGGCAUCGUUUUU

>*M. mobile* 163K
UAUUAUGUAUUUCGAAAUACAGAUGUACAGUUAAGAAUACAUAAGAAU
GAUACAUCACUAAAAAAAGGCUUUAUGCCGUAACUACUACUUAUUUUC
AAAAUAAGUAGUUUUUUUU

FIGURE 6

B.


1
70
L. innocua (1)
S. pyogenes (1)
S. mutans (1)
S. thermophilus1 (1)
71
111
L. innocua (54)
S. pyogenes (51)
S. mutans (66)
S. thermophilus1 (58)

C.

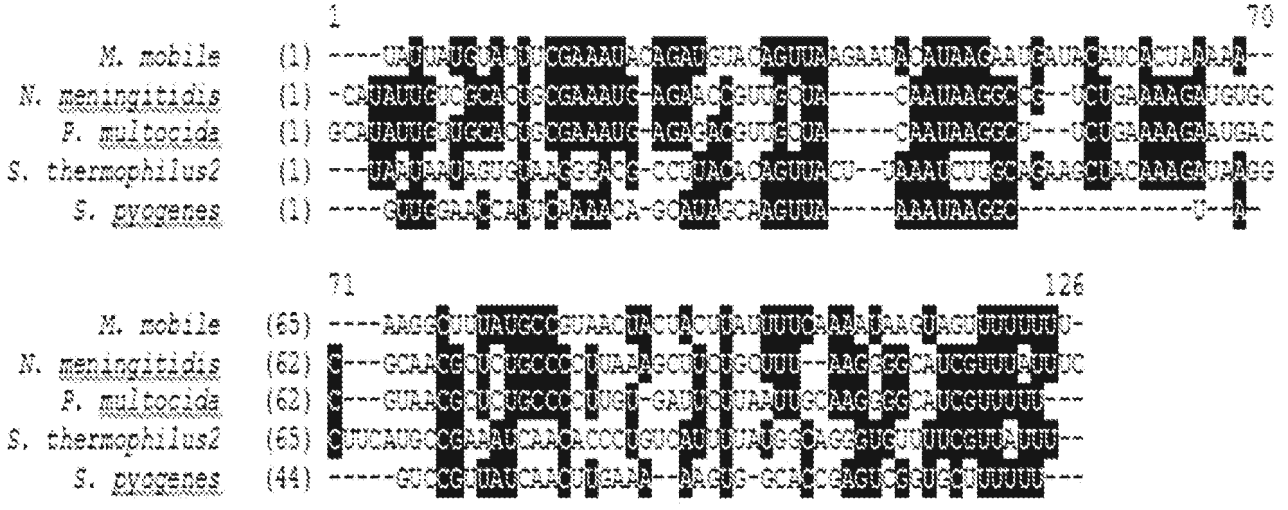

1
70
M. mobile (1)
N. meningitidis (1)
P. multocida (1)
S. thermophilus2 (1)
S. pyogenes (1)
71
126
M. mobile (65)
N. meningitidis (62)
P. multocida (62)
S. thermophilus2 (65)
S. pyogenes (44)

FIGURE 7

A. DNA comprising the sequences that encode the "Duplex-forming segment" of the crRNA repeat ("targeter-RNA")

>S. pyogenes SF370
GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC

>L. innocua Clip11262
GTTTTAGAGCTATGTTATTTTGAATGCTAACAAAAC

>S. mutans UA159
GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC

>S. thermophilus LMD-9 (locus 1)
GTTTTAGAGCTGTGTTGTTTCGAATGGTTCCAAAAC

>S. thermophilus LMD-9 (locus 2)
GTTTTTGTACTCTCAAGATTTAAGTAACTGTACAAC

>N. meningitidis Z2491
GTTGTAGCTCCCTTTCTCATTTCGCAGTGCTACAAT

>P. multocida Pm70
GTTGTAGTTCCCTCTCTCATTTCGCAGTGCTACAAT

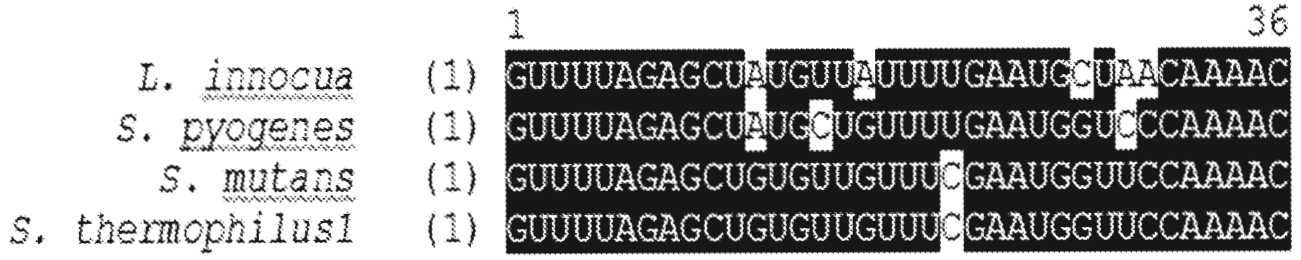
>M. mobile 163K
GTTTTGGTGTAGTATCATTCTTATGTATTCTTAAAC

>F. novicida U112
GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAG

>C. jejuni NCTC11168
ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC

FIGURE 7

B.


1
36
L. innocua (1) GUUUUAGAGCUAUGUUAUUUUGAAUGCUAACAAAAC
S. pyogenes (1) GUUUUAGAGCUAUGCUGUUUUGAAUGGUCCCAAAAC
S. mutans (1) GUUUUAGAGCUGUGUUGUUUCGAAUGGUUCCAAAAC
S. thermophilus1 (1) GUUUUAGAGCUGUGUUGUUUCGAAUGGUUCCAAAAC

C.

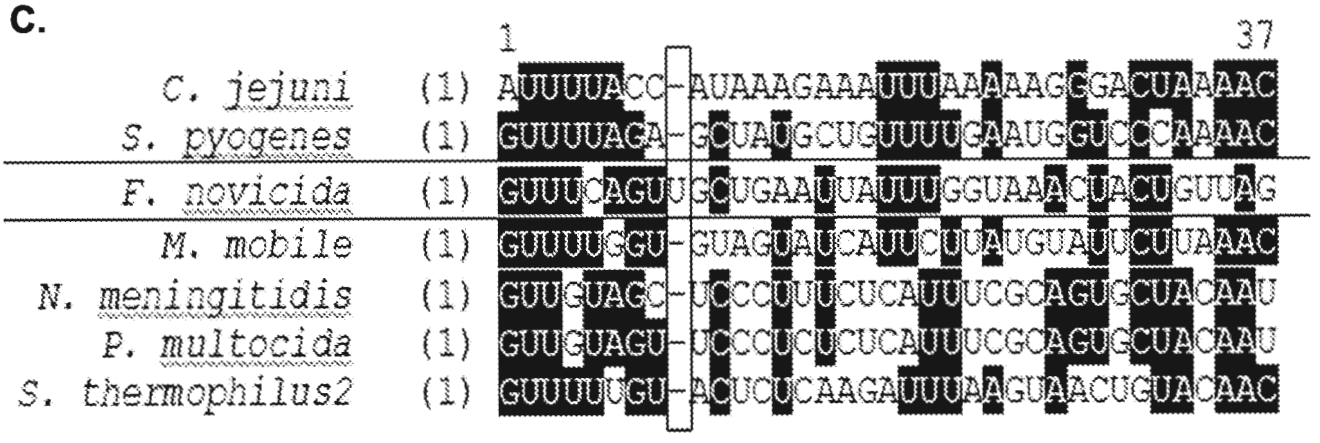

1
37
C. jejuni (1) AUUUUACC-AUAAAGAAAUUUAAAAAGGGACUAAAAC
S. pyogenes (1) GUUUUAGA-GCUAUGCUGUUUUGAAUGGUCCCAAAAC
F. novicida (1) GUUUCAGUUGCUGAAUUAUUUGGUAAACUACUGUUAG
M. mobile (1) GUUUUGGU-GUAGUAUCAUUCUUAUGUAUUCUUAAAC
N. meningitidis (1) GUUGUAGC-UCCCUUUCUCAUUUCGCAGUGCUACAAU
P. multocida (1) GUUGUAGU-UCCCUCUCUCAUUUCGCAGUGCUACAAU
S. thermophilus2 (1) GUUUUUGU-ACUCUCAAGAUUUAAGUAACUGUACAAC

FIGURE 8

A.

STRAIN	NUMBER OF CRISPRs	CAS54 CRISPR identifier[a]	CRISPR REPEAT:tracrRNA BASEPAIRING[b]
Streptococcus pyogenes SF370	2	NC_002737_1	5' GUUUUAG--AGCUAUGCUGUUUUGAAUGGUCCCAAAAC 3' ||||• |||||||||||||||||||| |||| | 3' AAAUUGAACGAUACGACAAAACUUACCAAGGUUGUU 5'
Streptococcus mutans UA159	1	NC_004350_1	GUUUUAG--AGCUGUGUUGUUUCGAAUGGUUCCAAAAC ||||• ||||||||||||||||||•|||||| || AAAUUGAACGACACAACAAAGCUUACUAAGGUUGUG
Streptococcus thermophilus LMD-9	3	NC_008532_5	GUUUUAG--AGCUGUGUUGUUUCGAAUGGUUCCAAAAC ||||• |||||||||||||||||||| |||| | AAAUUGAGCGACACAACAAAGCUUACCAAAGUUUGG
		NC_008532_2	GUUUUUGUACUCU-CAAGAUUUAAGUAACUGUACAAC ||| |• ||| ||||||||||||||||| || AAACAUCGAAGACGUUCUAAAUUCAUUGACACAUUC
Listeria innocua Clip11262	1	NC_003212_2	GUUUUAG--AGCUAUGUUAUUUUGAAUGCUAACAAAAC ||||• |||||||||||||||||•||||||| | AAAUUGAACGAUACAAUAAAACUUAUGAUUGUUAUA
Treponema denticola ATCC35405	1	NC_002967_1	GUUUGAG--AGUUGUGUAAUUUAAGAUGGAUCUCAAAC ||||• |||||||||||||||||||||| ||| AACUUGAGCAACACAUUAAAUUCUACCUAGAAUUUA
Neisseria meningitidis Z2491	2	NC_003116_10	GUUGUAGCUCCCUUUCUCAUUUCGCAGUGCUACAAU •||||||| | |||||||||||||||| ||||| UAACAUCGUUGCCAAGAGUAAAGCGUCACGCUGUUA
Streptococcus gordonii str. Challis substr. CH1	1	NC_009785_2	GUUUUUGUACUCU-CAAGAUUUAAGUAACUGUACAAC ||| |• |||| |||||||||||||||| || AAACAUCGAAGACGUUCUAAAUUCAUUGACACAUUC
Bifidobacterium longum DJO10A	1	NC_010816_1	No significant antirepeat found
Bifidobacterium bifidum S17	1	NC_014616_1	GUUUCA-AUGCCUGUCAGAUCAAUGACUUUGACCAC • ||||||||||| ||||||||•|||| || AGUUAAAUACGGACAGUCCAGUUACUGGAACUAGUA
Lactobacillus salivarius UCC118	1	NC_007929_1	GUUUCAGAAGUAUGUUAAAUCAAUAAGGUUAAGACC |• || ||||||||||||| ||||•||•||| AAGUUGAGUCUUACAAUUUAGUUACUCCAGUUUUGG

FIGURE 8

B.

STRAIN	NUMBER OF CRISPRs	CAS54 CRISPR identifier[a]	CRISPR REPEAT:tracrRNA BASEPAIRING[b]
Francisella tularensis subsp. *novicida* U112	2	NC_008601_1[c]	CUAACAGUAGUUUACCAAAUAAUUCAGCAACUGAAAC UUGUGUUCAUGUAUGGUUUAUUAGAUUGUUGACUUUG CUAACAGUAGUUUACCAAAUAAUUCA-GCAACUGAAAC UUUAAUAUUUACAUGGUUUAUUAAUUACG-AGACAUUA
Legionella pneumophila str. Paris	1	NC_006368_1	CCAAUAAUCCCUCAUCUAAAAAUCCA-ACCACUGAAAC AUUUAAUCUUUAGUAGAUUUAAAGCUAUGG-GACUUUA
Nitratifractor salsuginis DSM 16511	2	NC_014935_3	GUUUUAAGACCCCUCAAAA-CCCCACCCUGUUACAAU UAAAAUUCUGGGGAGUUUUUUGGGAUGGGACAAUGU
Butyrivibrio fibrisolvens 16/4	1	ND	GUUUUAGUACCCGGGAAAAUUAAGUGAUUGGAAAAC AAAUCAUCCAGGCUUUUUUAAUUCACUAAUCUUUGA
Streptobacillus moniliformis DSM 12112	1	NC_013515_1	GAUACA-GAUUCUUGGUAAAUUUGAAUAAUGUGAAUG UAUGUCUUAGACAUCAUUUGAACUUAUUACACUUAG
Mycoplasma mobile 163K	1	NC_006908_1	GUUUUGGUGUAGUAUCAUUCUUAUGUAUUCUUAAAC AAAAAUCACUACAUAGUAAGAAUACAUAAGAAUUGA
Capnocytophaga ochracea DSM7271	1	NC_013162_2	UGUUAUCAAUCGCAAAACUACAAAAUUUGAAAGCAAUUCACAAC AUAAUAUGGAUGAAAACAGCGUGUUAAACUUUCGUUAAGUGUUA
Elusimicrobium minutum Pei191	1	NC_010644_1	No significant antirepeat found
Bacteroides fragilis NCTC 9343	2	NC_003228_4	No significant antirepeat found
Acidovorax ebreus TPSY	1	ND	AGUGUAGCUAUCCGGGGUGAGAGAGGGAGCUACAAC UAACACCAAUAGGCCCCACUCUCGGCAACGACGUUA
Wolinella succinogenes DSM 1740	2	NC_005090_1	AUUAUAGCGAGGAAUGGCUGAGUAGGCGGCUAUAAC CAGCGACGCGAUUUACCGACUCAUCCGUCGCCGAUA
		NC_005090_2[c]	GCAACACUUUAUAGCAAAUCCGCUUAGCCUGUGAAAC CGUUGUGAGAUAUUAUUAGGGUAGGUCGAAAUUGUUU GCAACACUUUAUAGCAAAUCCGCUUAGCCUGUGAAAC CGUUGUGAAAUAUCGUUUAGGCAAGCUACGGAACUUU

FIGURE 8

C.

STRAIN	NUMBER OF CRISPRs	CAS54 CRISPR identifier*	CRISPR REPEAT:tracrRNA BASEPAIRING*
Prevotella ruminicola 23	1	NC_014033_4	GUUGUGGGUUGAUGUAGGAAAGGAAUG AUAUACAAC •:::::•: ::::::::::: :•: : •: :: UAACACUAUUCUACAUCCUUAGAAUGCACACGUUUU
Fibrobacter succinogenes subsp. *succinogenes* S85	1	NC_013410_1	GUUGUUAUUGCACUAGUAAAACUAGAAAAUCUGAAAGCAAUUCACAAC • • : ::• :::: : : : :::::::::::::::::::::: GUUUUUAAAAU-UGAUAUAACUAAAGUAUUAGACUUUCGUUAAGUGUUA
Campylobacter jejuni NCTC11168	1	NC_002163_2*	AUUUUACCAUAAAGAAAUUUAAAAAGGGACUAAAAC ::: : :: : :::::: :•: ::::: UAACUUUAAAUCCCAACUAAUUUAGCUUUAUUUUA AUUUUACCAUAAAGAAAUUUAAAAAGGGACUAAAAC : : :::::::::::::::::::::::::: AUAUUAGAGUAUUCUUUAAAUUUUUCCCUGAUUUUA
Pasteurella multocida Pm70	3	NC_002663_3	AUUGUAGCACUGCGAAAUGAGAGAGGGAACUACAAC ::::: ::::::::::::::::::: : ::::::: UAACAACGUGACGCUUUACUCUCUGCAACGAUGUUA
Rhodopseudomonas palustris BisB18	1	NC_007925_4	AGCCUACCACGGGGAAAUCGGUAGGGAAGCCACGGC :• : : ::::::::::::::•::::::::::: UUAGUACGCCACCCUUUAGCCAUCCUUUCGGUGCCG
Arthrobacter polytropus DSM 2926 plasmid pLYOP01	1	NC_014633_2	No significant antirepeat found@
Citrobacter oli DSM 7084	1	NC_014363_7	GUUUUGG-G-GCAGUGUCGUUUUGACUGGUAAU-CAAAC ::::• ::::::•:::::•::::::::::: :: : AAACUGAGCGUCAUAGCAAGACUGACCAUUAAGUCU
Rhodospirillum rubrum ATCC 11170	11	NC_007643_6*	ACUCUACCAUGGCGGUGUGGGACGGGGCCAUGGAAC ::::: :::::: :::::::::::::::::::::::: GCGAGAUAGUACCG-CACACCCUGCCCCGGUACCUUG ACUCUACCAUGGCGGUGUGGGACGGGGCCAUGGAAC :::::::::::::: :::::::::: : : : UGAGAUGGUACCGGCACACCCUGCCAAGAGGGCAAA
Dinoroseobacter shibae DFL 12	2	NC_009952_2	AGUUUAGCUGUUCAGAAUUCGGGGUCCAGCCG-CAAC ::•: •:::::::::::::: :::: ::::: :: ACAGAAAUGCAAGUCUUAAGCGCCAGCUCGGCAAUU
Verminephrobacter eiseniae EF01-2	3	NC_008786_1	AUUGUAGCUCACUGGGAUAUGCGCACUGGCCG-GAAC •:: :::::::::::::::::: :::: :: :: GCCUAUAGAGUGACCCUAUACGCGAGACCUGCGAUU

FIGURE 9

A. *Streptococcus pyogenes*

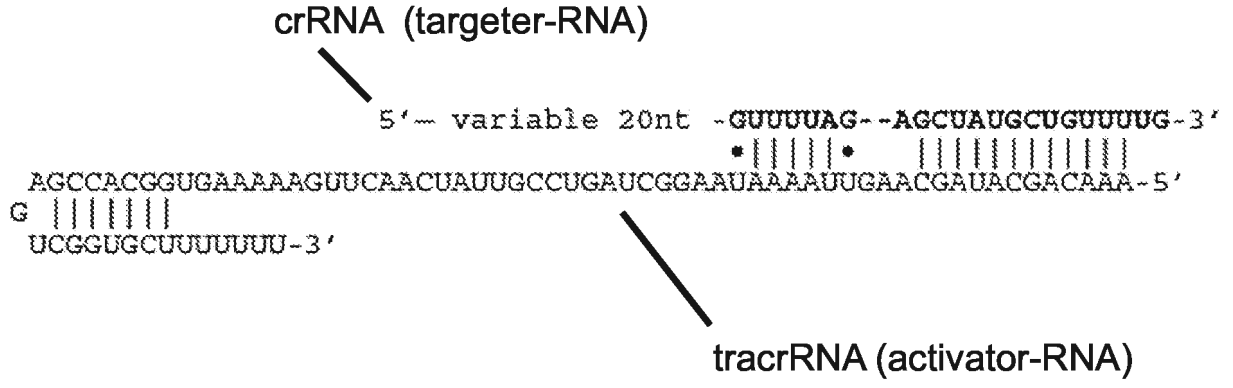
Sequence comprising the "duplex-forming segment" of the crRNA repeat ("targeter-RNA")

5' GUUUUAGAGCUAUGCUGUUUUGAAUGGUCCCAAAAC 3'

tracrRNA sequence ("activator-RNA")

5' GUUGGAACCAUUCAAAACAGCAUAGCAAGUUAAAAUAAGGCUAGUCC
GUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU 3'

Base-pairing as thought to occur *in vivo* (crRNA /tracrRNA)



Example of a single-molecule DNA-targeting RNA

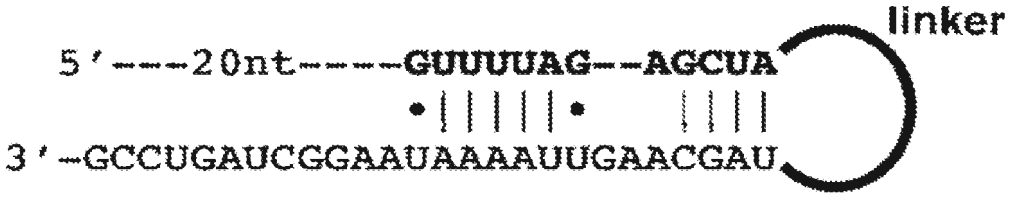


FIGURE 9

B. *Listeria innocua*

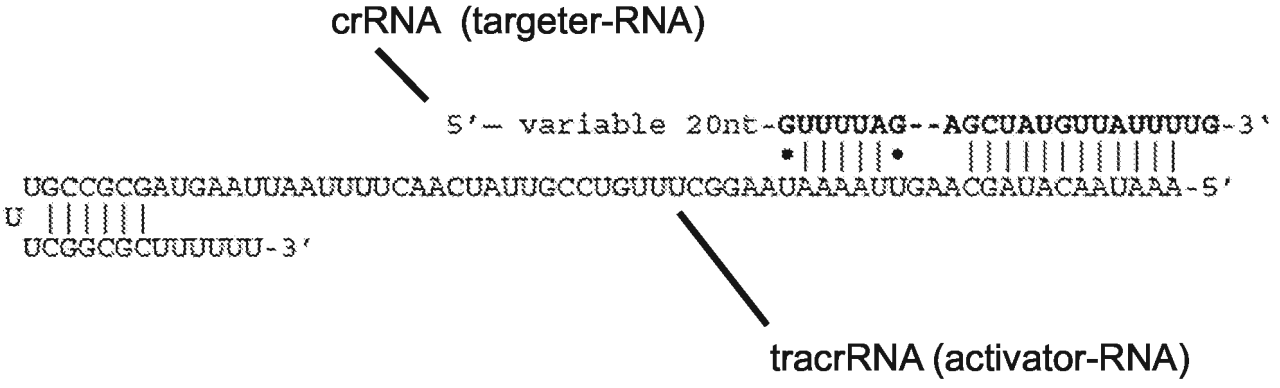
Sequence comprising the "duplex-forming segment" of the crRNA repeat ("targeter-RNA")

5' GUUUUAGAGCUAUGUUAUUUUGAAUGCUAACAAAAC 3'

tracrRNA sequence ("activator-RNA")

5' AUAUUGUUAGUAUUCAAAAUAACAUAGCAAGUUAAAAUAAGGCUUUGU
CCGUUAUCAACUUUUAAUUAAGUAGCGCUGUUUCGGCGCUUUUUU 3'

Base-pairing as thought to occur *in vivo* (crRNA /tracrRNA)



Example of a single-molecule DNA-targeting RNA

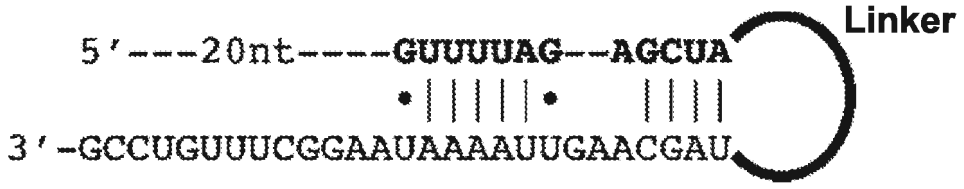


FIGURE 9

C. *Streptococcus mutans*

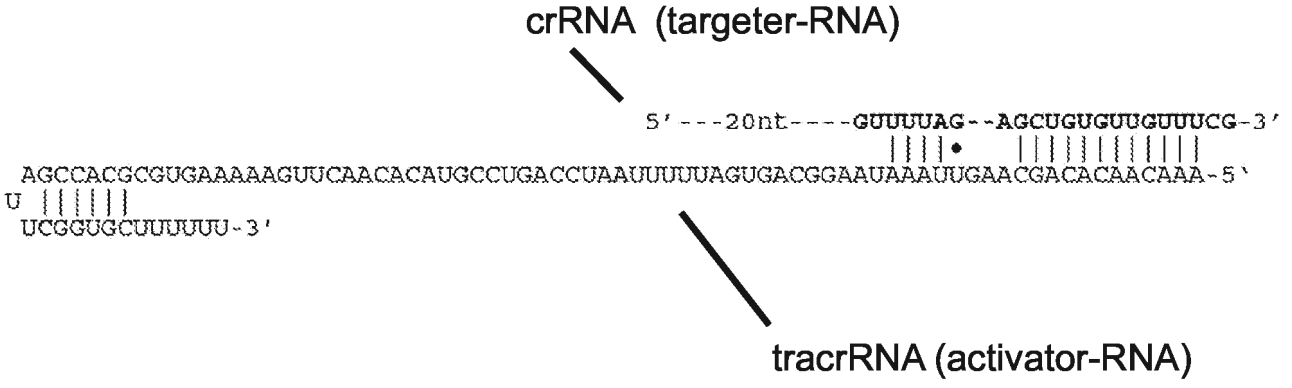
Sequence comprising the "duplex-forming segment" of the crRNA repeat ("targeter-RNA")

5' GUUUUAGAGCUGUGUUGUUUCGAAUGGUUCCAAAAC 3'

tracrRNA sequence ("activator-RNA")

5' UGUUGGAAUCAUUCGAAACAACACAGCAAGUUAAAAUAAGGCAGUGAU UUUUAAUCCAGUCCGUACACAACUUGAAAAAGUGCGCACCGAUUCGGUG CUUUUUU 3'

Base-pairing as thought to occur *in vivo* (crRNA /tracrRNA)



Example of a single-molecule DNA-targeting RNA

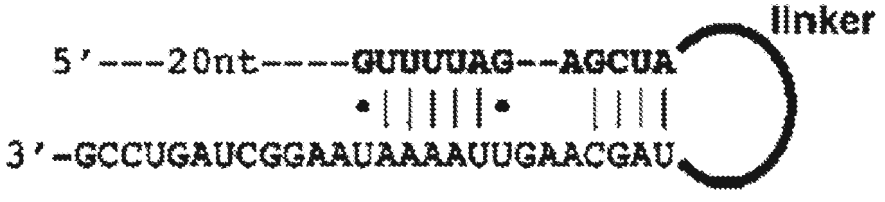


FIGURE 9

D. *Streptococcus thermophilus – locus 1 (or A)*

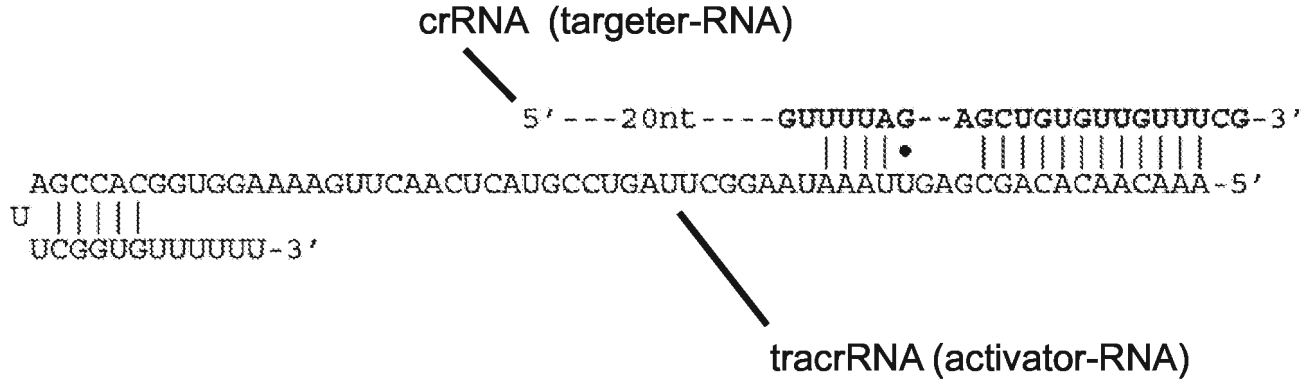
Sequence comprising the "duplex-forming segment" of the crRNA repeat ("targeter-RNA")

5' GUUUUAGAGCUGUGUUGUUUCGAAUGGUUCCAAAAC 3'

tracrRNA sequence ("activator-RNA")

5' UUGUGGUUUGAAACCAUUCGAAACAACACAGCGAGUUAAAAUAAGGCU
UAGUCCGUACUCAACUUGAAAAGGUGGCACCGAUUCGGUGUUUUUUU 3'

Base-pairing as thought to occur *in vivo* (crRNA /tracrRNA)



Example of a single-molecule DNA-targeting RNA

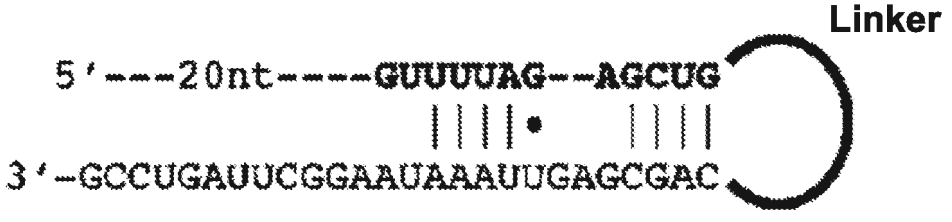


FIGURE 9

E. *Streptococcus thermophilus – locus 2 (or B)*

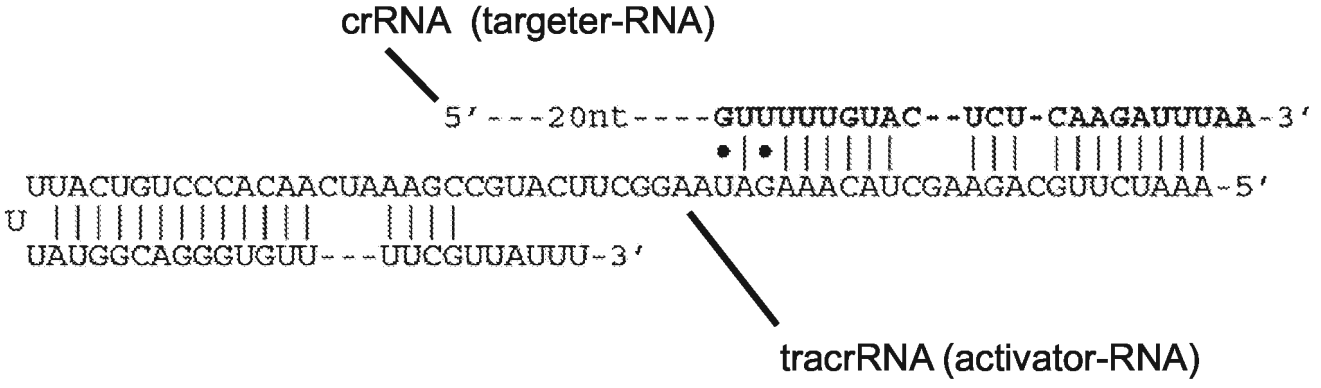
Sequence comprising the "duplex-forming segment" of the crRNA repeat ("targeter-RNA")

5' GUUUUAGAGCUGUGUUGUUUCGAAUGGUUCCAAAAC 3'

tracrRNA sequence ("activator-RNA")

5' UAAUAAUAGUGUAAGGGACGCCUUACACAGUUACUUAAAUCUUGCAG
AAGCUACAAAGAUAAGGCUUCAUGCCGAAAUCAACACCCUGUCAUUUUA
UGGCAGGGUGUUUUCGUUAUUU 3'

Base-pairing as thought to occur *in vivo* (crRNA /tracrRNA)



Example of a single-molecule DNA-targeting RNA

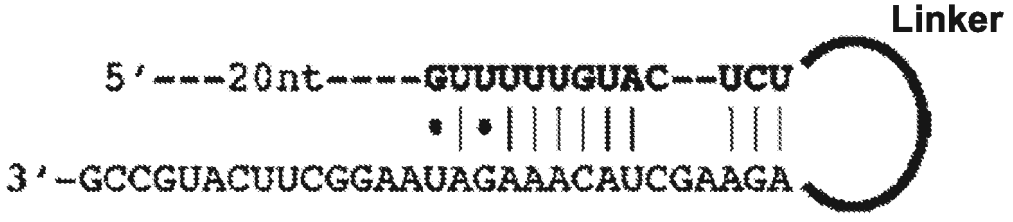


FIGURE 9

F. *Neisseria meningitidis*

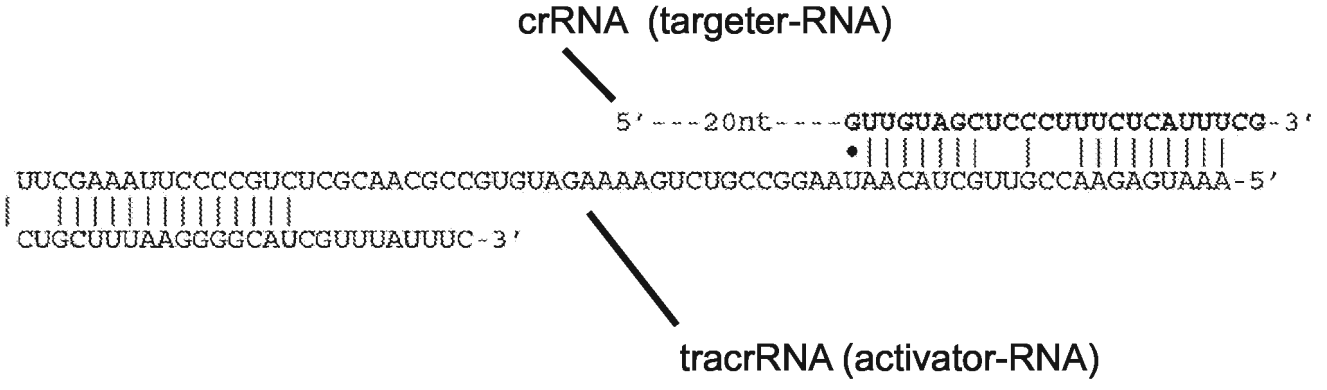
Sequence comprising the "duplex-forming segment" of the crRNA repeat ("targeter-RNA")

5' GUUGUAGCUCCCUUUCUCAUUUCGCAGUGCUACAAU 3'

tracrRNA sequence ("activator-RNA")

5' CAUAUUGUCGCACUGCGAAAUGAGAACCGUUGCUACAAUAAGGCCGU
CUGAAAAGAUGUGCCGCAACGCUCUGCCCCUUAAAGCUUCUGCUUUAA
GGGGCAUCGUUUAUUUC 3'

Base-pairing as thought to occur *in vivo* (crRNA /tracrRNA)



Example of a single-molecule DNA-targeting RNA

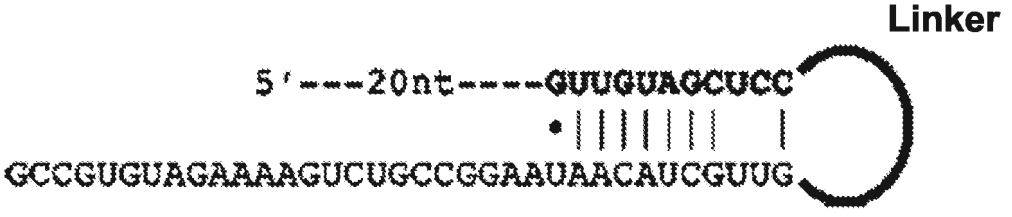


FIGURE 9

G. *Pasteurella multocida*

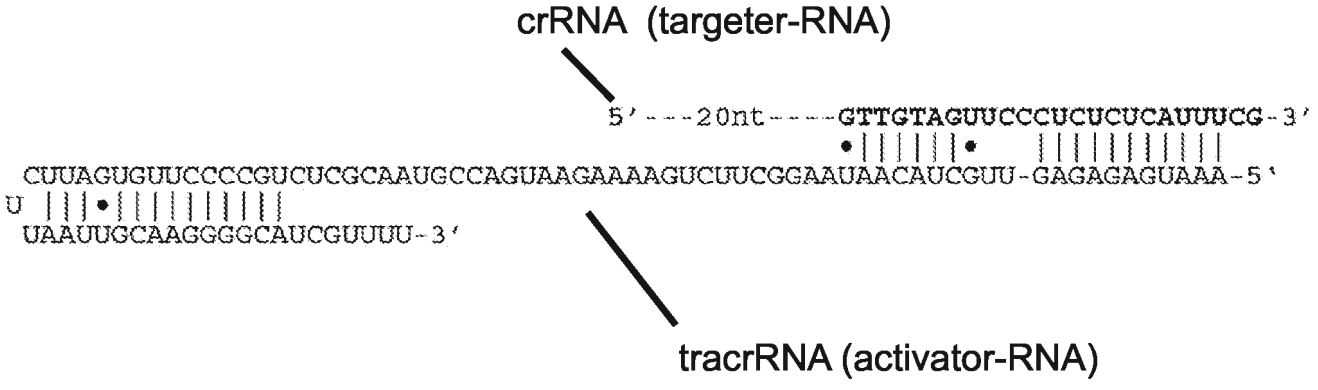
Sequence comprising the "duplex-forming segment" of the crRNA repeat ("targeter-RNA")

5' GUUGUAGUUCCCUCUCUCAUUUCGCAGUGCUACAAU 3'

tracrRNA sequence ("activator-RNA")

5' GCAUAUUGUUGCACUGCGAAAUGAGAGACGUUGCUACAAUAAGGCUU
CUGAAAAGAAUGACCGUAACGCUCUGCCCCUUGUGAUUCUUAAUUGCA
AGGGGCAUCGUUUUU 3'

Base-pairing as thought to occur *in vivo* (crRNA /tracrRNA)



Example of a single-molecule DNA-targeting RNA

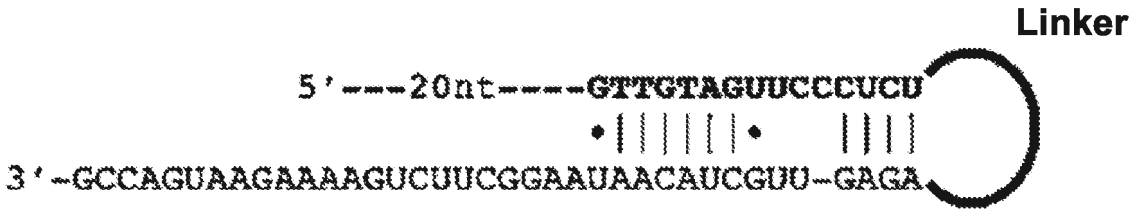


FIGURE 10

A. Sequence identities relative to: *S. pyogenes* Cas9/Csn1

		Sequence identities - MUSCLE alignment		
		Full-length	*Domain 1*	*Domain 2*
species	*reference sequence*	% indentity	% indentity	% identity
Streptococcus pyogenes M1 GAS	NP_269215	100.0	100.0	100.0
Streptococcus pyogenes MGAS5005	YP_282132.1	99.9	99.4	100.0
Listeria innocua Clip11262	NP_472073	54.3	60.0	64.9
Pasteurella multocida subsp. multocida str. Pm70	NP_246064.1	19.7	29.0	25.9
Streptococcus thermophilus LMD-9 Csn1-A	YP_820832	59.2	75.6	72.4
Streptococcus thermophilus LMD-9 Csn1-B	YP_820161.1	20.6	27.3	26.8
Neisseria meningitidis Z2491	YP_002342100.1	20.2	33.6	28.1
Streptococcus mutans UA159	NP_721764	64.9	78.1	74.1
Streptococcus gordonii str. Challis substr. CH1	YP_001450662.1	19.8	28.2	27.0
Campylobacter jejuni subsp. jejuni NCTC 11168	YP_002344900.1	19.6	30.3	26.3
Treponema denticola ATCC 35405	NP_970941	32.5	47.3	38.8

B. Sequence identities relative to: *N. meningitidis* Cas9/Csn1

		Sequence identities - MUSCLE alignment		
		Full-length	*Domain 1*	*Domain 2*
species	reference sequence	% indentity	% indentity	% identity
Streptococcus pyogenes M1 GAS	NP_269215	20.2	33.6	28.1
Streptococcus pyogenes MGAS5005	YP_282132.1	20.3	34.5	28.1
Listeria innocua Clip11262	NP_472073	18.8	33.6	25.9
Pasteurella multocida subsp. multocida str. Pm70	NP_246064.1	64.3	72.1	69.0
Streptococcus thermophilus LMD-9 Csn1-A	YP_820832	19.6	35.3	25.9
Streptococcus thermophilus LMD-9 Csn1-B	YP_820161.1	25.8	35.7	35.1
Neisseria meningitidis Z2491	YP_002342100.1	100.0	100.0	100.0
Streptococcus mutans UA159	NP_721764	19.2	36.1	25.5
Streptococcus gordonii str. Challis substr. CH1	YP_001450662.1	25.3	37.5	35.8
Campylobacter jejuni subsp. jejuni NCTC 11168	YP_002344900.1	34.7	45.0	41.0
Treponema denticola ATCC 35405	NP_970941	18.8	31.5	25.5

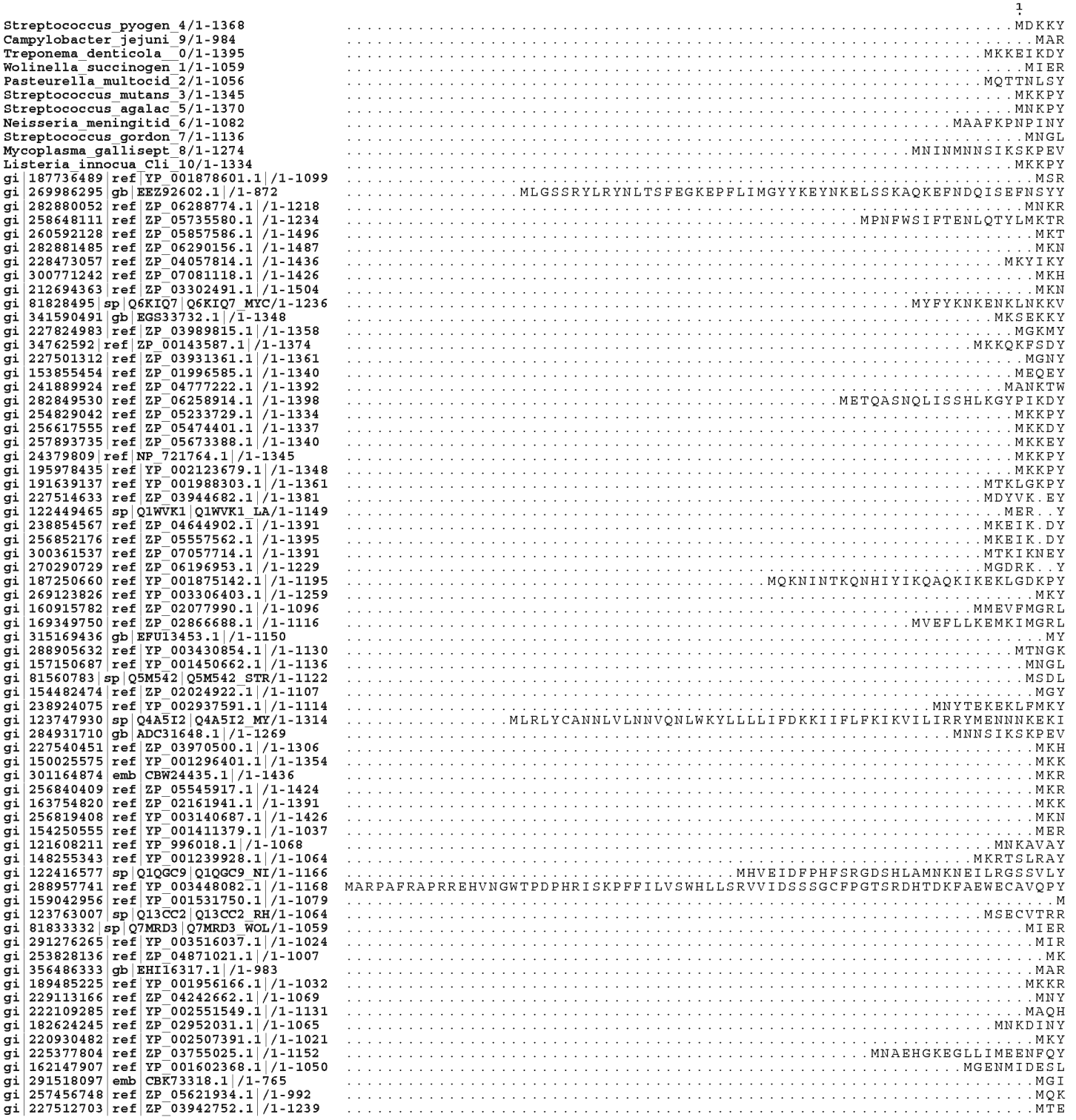

1
Streptococcus_pyogen_4/1-1368 MDKKY
Campylobacter_jejuni_9/1-984 MAR
Treponema_denticola_0/1-1395 MKKEIKDY
Wolinella_succinogen_1/1-1059 MIER
Pasteurella_multocid_2/1-1056 MQTTNLSY
Streptococcus_mutans_3/1-1345 MKKPY
Streptococcus_agalac_5/1-1370 MNKPY
Neisseria_meningitid_6/1-1082 MAAFKPNPINY
Streptococcus_gordon_7/1-1136 MNGL
Mycoplasma_gallisept_8/1-1274 MNINMNNSIKSKPEV
Listeria_innocua_Cli_10/1-1334 MKKPY
gi|187736489|ref|YP_001878601.1|/1-1099 MSR
gi|269986295|gb|EEZ92602.1|/1-872 MLGSSRYLRYNLTSFEGKEPFLIMGYYKEYNKELSSKAQKEFNDQISEFNSYY
gi|282880052|ref|ZP_06288774.1|/1-1218 MNKR
gi|258648111|ref|ZP_05735580.1|/1-1234 MPNFWSIFTENLQTYLMKTR
gi|260592128|ref|ZP_05857586.1|/1-1496 MKT
gi|282881485|ref|ZP_06290156.1|/1-1487 MKN
gi|228473057|ref|ZP_04057814.1|/1-1436 MKYIKY
gi|300771242|ref|ZP_07081118.1|/1-1426 MKH
gi|212694363|ref|ZP_03302491.1|/1-1504 MKN
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236 MYFYKNKENKLNKKV
gi|341590491|gb|EGS33732.1|/1-1348 MKSEKKY
gi|227824983|ref|ZP_03989815.1|/1-1358 MGKMY
gi|34762592|ref|ZP_00143587.1|/1-1374 MKKQKFSDY
gi|227501312|ref|ZP_03931361.1|/1-1361 MGNY
gi|153855454|ref|ZP_01996585.1|/1-1340 MEQEY
gi|241889924|ref|ZP_04777222.1|/1-1392 MANKTW
gi|282849530|ref|ZP_06258914.1|/1-1398 METQASNQLISSHLKGYPIKDY
gi|254829042|ref|ZP_05233729.1|/1-1334 MKKPY
gi|256617555|ref|ZP_05474401.1|/1-1337 MKKDY
gi|257893735|ref|ZP_05673388.1|/1-1340 MKKEY
gi|24379809|ref|NP_721764.1|/1-1345 MKKPY
gi|195978435|ref|YP_002123679.1|/1-1348 MKKPY
gi|191639137|ref|YP_001988303.1|/1-1361 MTKLGKPY
gi|227514633|ref|ZP_03944682.1|/1-1381 MDYVK.EY
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149 MER..Y
gi|238854567|ref|ZP_04644902.1|/1-1391 MKEIK.DY
gi|256852176|ref|ZP_05557562.1|/1-1395 MKEIK.DY
gi|300361537|ref|ZP_07057714.1|/1-1391 MTKIKNEY
gi|270290729|ref|ZP_06196953.1|/1-1229 MGDRK..Y
gi|187250660|ref|YP_001875142.1|/1-1195 MQKNINTKQNHIYIKQAQKIKEKLGDKPY
gi|269123826|ref|YP_003306403.1|/1-1259 MKY
gi|160915782|ref|ZP_02077990.1|/1-1096 MMEVFMGRL
gi|169349750|ref|ZP_02866688.1|/1-1116 MVEFLLKEMKIMGRL
gi|315169436|gb|EFU13453.1|/1-1150 MY
gi|288905632|ref|YP_003430854.1|/1-1130 MTNGK
gi|157150687|ref|YP_001450662.1|/1-1136 MNGL
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122 MSDL
gi|154482474|ref|ZP_02024922.1|/1-1107 MGY
gi|238924075|ref|YP_002937591.1|/1-1114 MNYTEKEKLFMKY
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314 MLRLYCANNLVLNNVQNLWKYLLLLIFDKKIIFLFKIKVILIRRYMENNNKEKI
gi|284931710|gb|ADC31648.1|/1-1269 MNNSIKSKPEV
gi|227540451|ref|ZP_03970500.1|/1-1306 MKH
gi|150025575|ref|YP_001296401.1|/1-1354 MKK
gi|301164874|emb|CBW24435.1|/1-1436 MKR
gi|256840409|ref|ZP_05545917.1|/1-1424 MKR
gi|163754820|ref|ZP_02161941.1|/1-1391 MKK
gi|256819408|ref|YP_003140687.1|/1-1426 MKN
gi|154250555|ref|YP_001411379.1|/1-1037 MER
gi|121608211|ref|YP_996018.1|/1-1068 MNKAVAY
gi|148255343|ref|YP_001239928.1|/1-1064 MKRTSLRAY
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166 MHVEIDFPHFSRGDSHLAMNKNEILRGSSVLY
gi|288957741|ref|YP_003448082.1|/1-1168 MARPAFRAPRREHVNGWTPDPHRISKPFFILVSWHLLSRVVIDSSSGCFPGTSRDHTDKFAEWECAVQPY
gi|159042956|ref|YP_001531750.1|/1-1079 M
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064 MSECVTRR
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059 MIER
gi|291276265|ref|YP_003516037.1|/1-1024 MIR
gi|253828136|ref|ZP_04871021.1|/1-1007 MK
gi|356486333|gb|EHI16317.1|/1-983 MAR
gi|189485225|ref|YP_001956166.1|/1-1032 MKKR
gi|229113166|ref|ZP_04242662.1|/1-1069 MNY
gi|222109285|ref|YP_002551549.1|/1-1131 MAQH
gi|182624245|ref|ZP_02952031.1|/1-1065 MNKDINY
gi|220930482|ref|YP_002507391.1|/1-1021 MKY
gi|225377804|ref|ZP_03755025.1|/1-1152 MNAEHGKEGLLIMEENFQY
gi|162147907|ref|YP_001602368.1|/1-1050 MGENMIDESL
gi|291518097|emb|CBK73318.1|/1-765 MGI
gi|257456748|ref|ZP_05621934.1|/1-992 MQK
gi|227512703|ref|ZP_03942752.1|/1-1239 MTE

FIGURE 11 A.

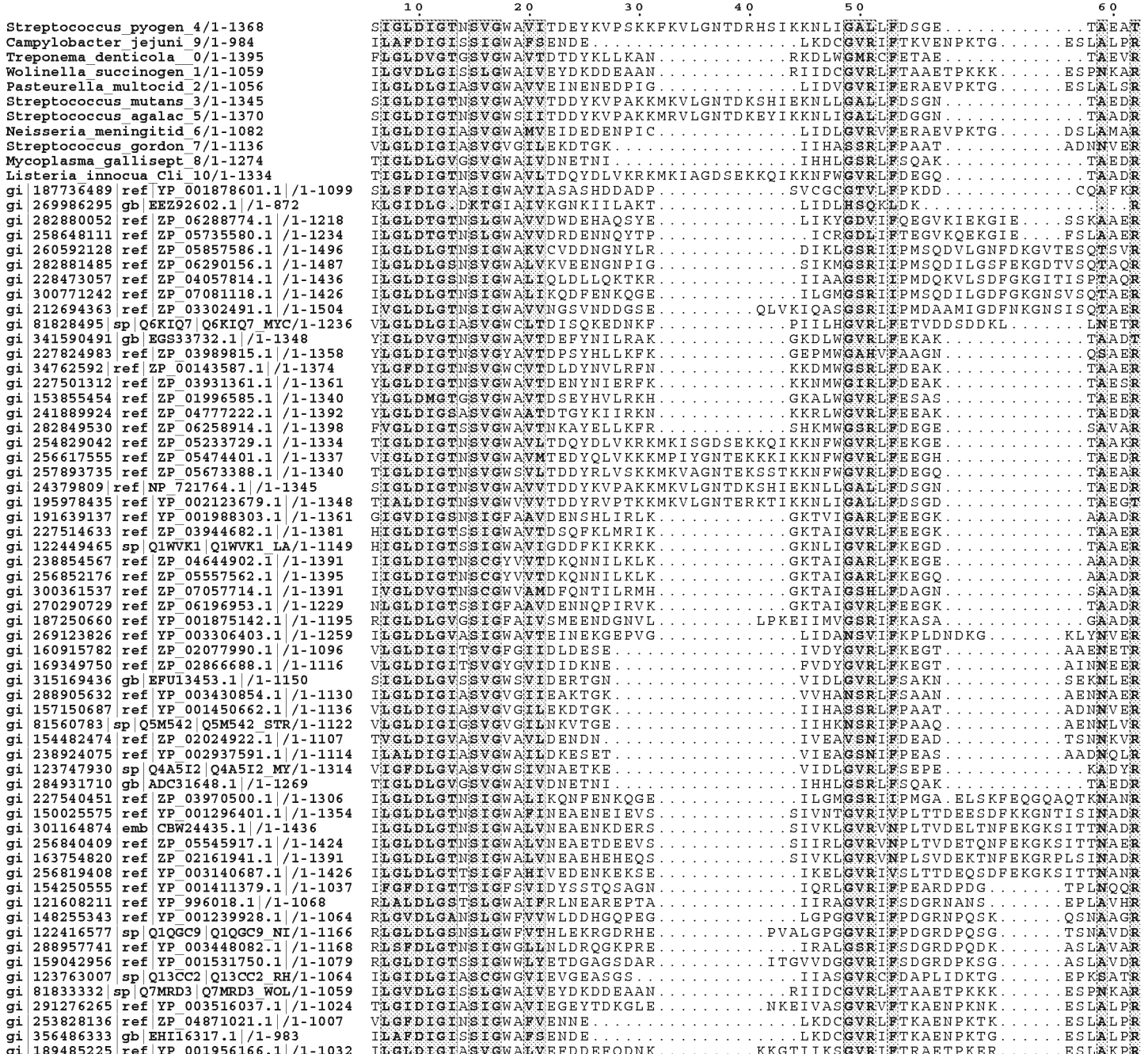

10 20 30 40 50 60
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 B.

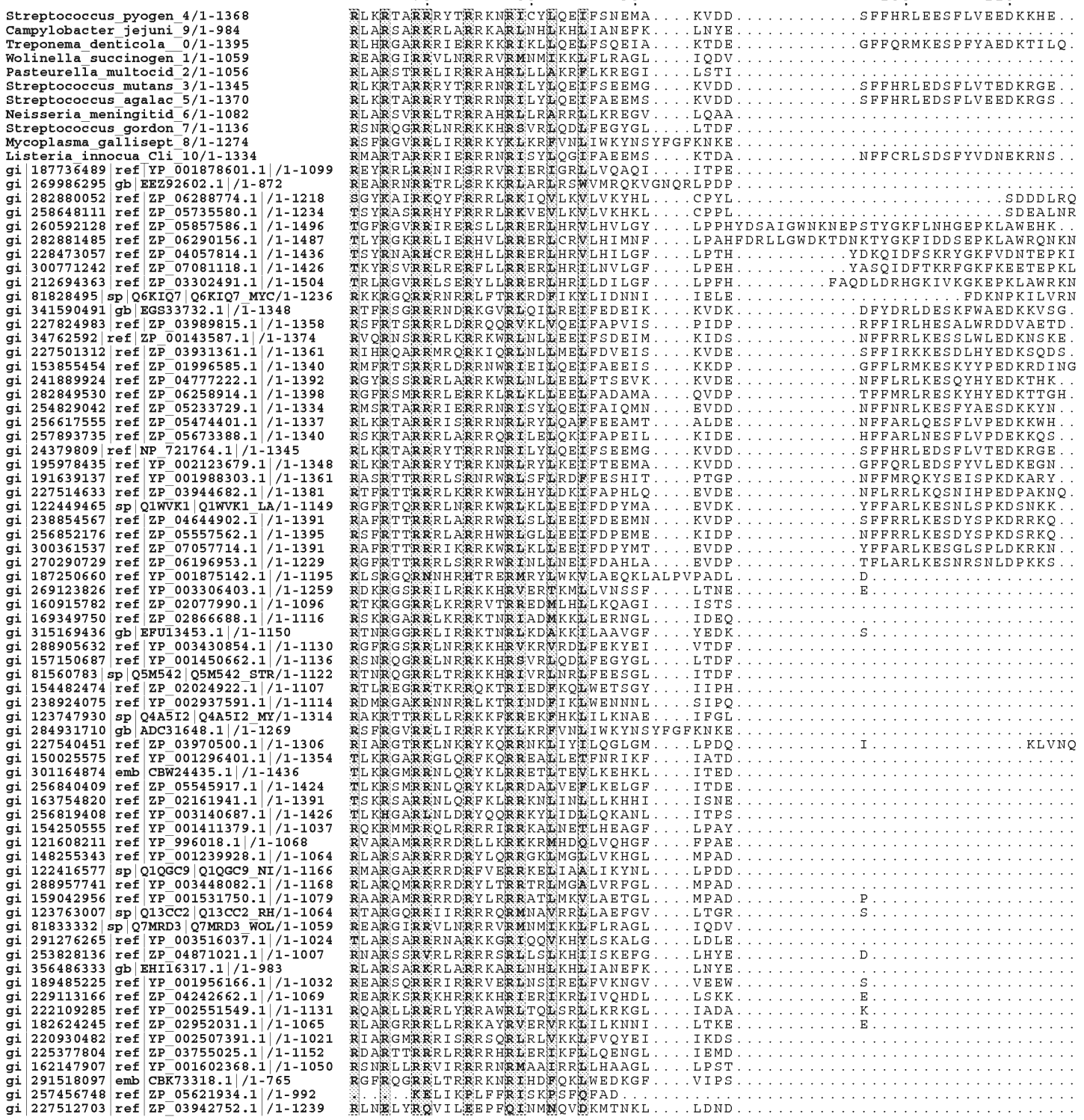

70
80
90
100
110
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 C.

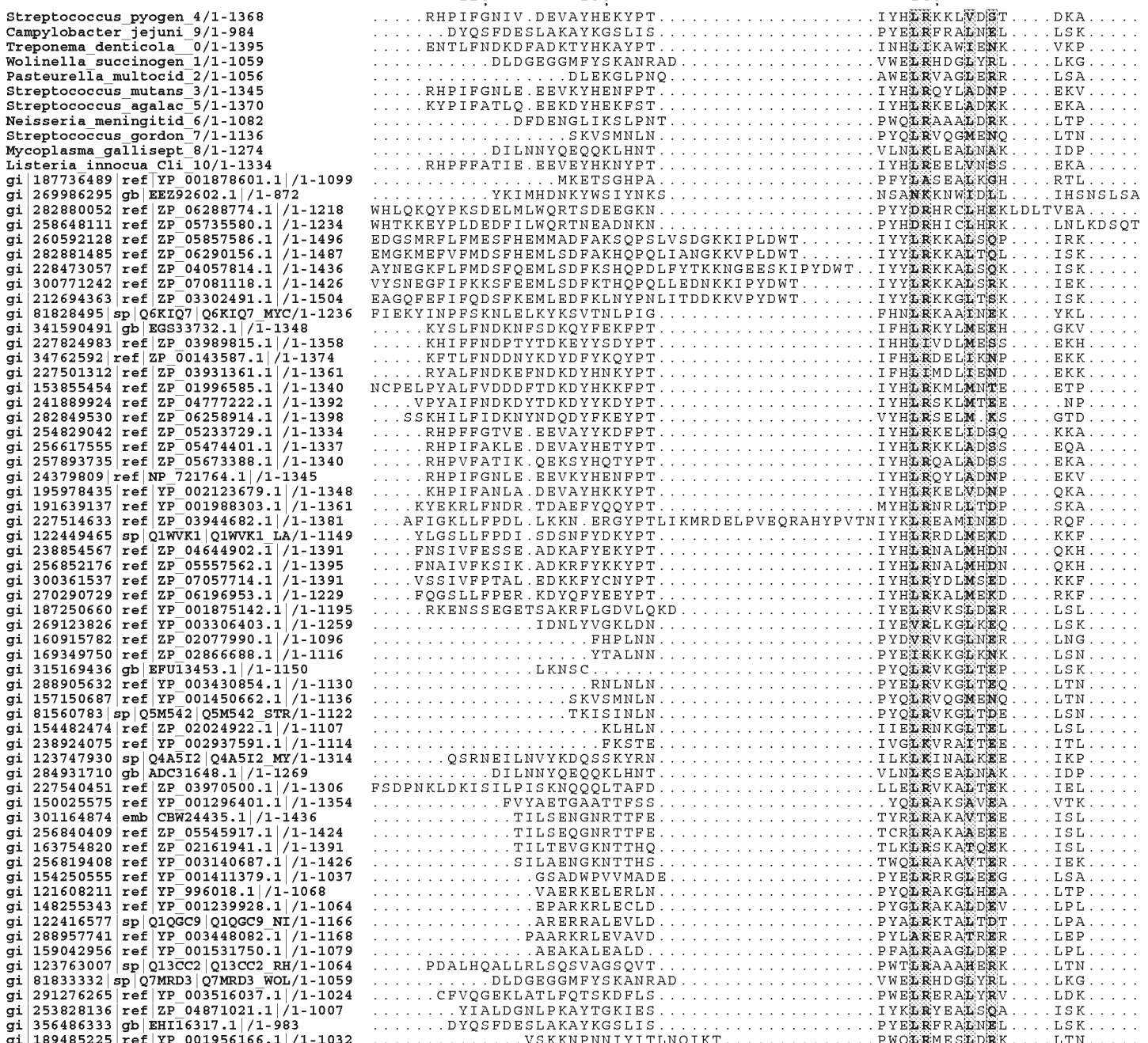

120
130
140
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 D.

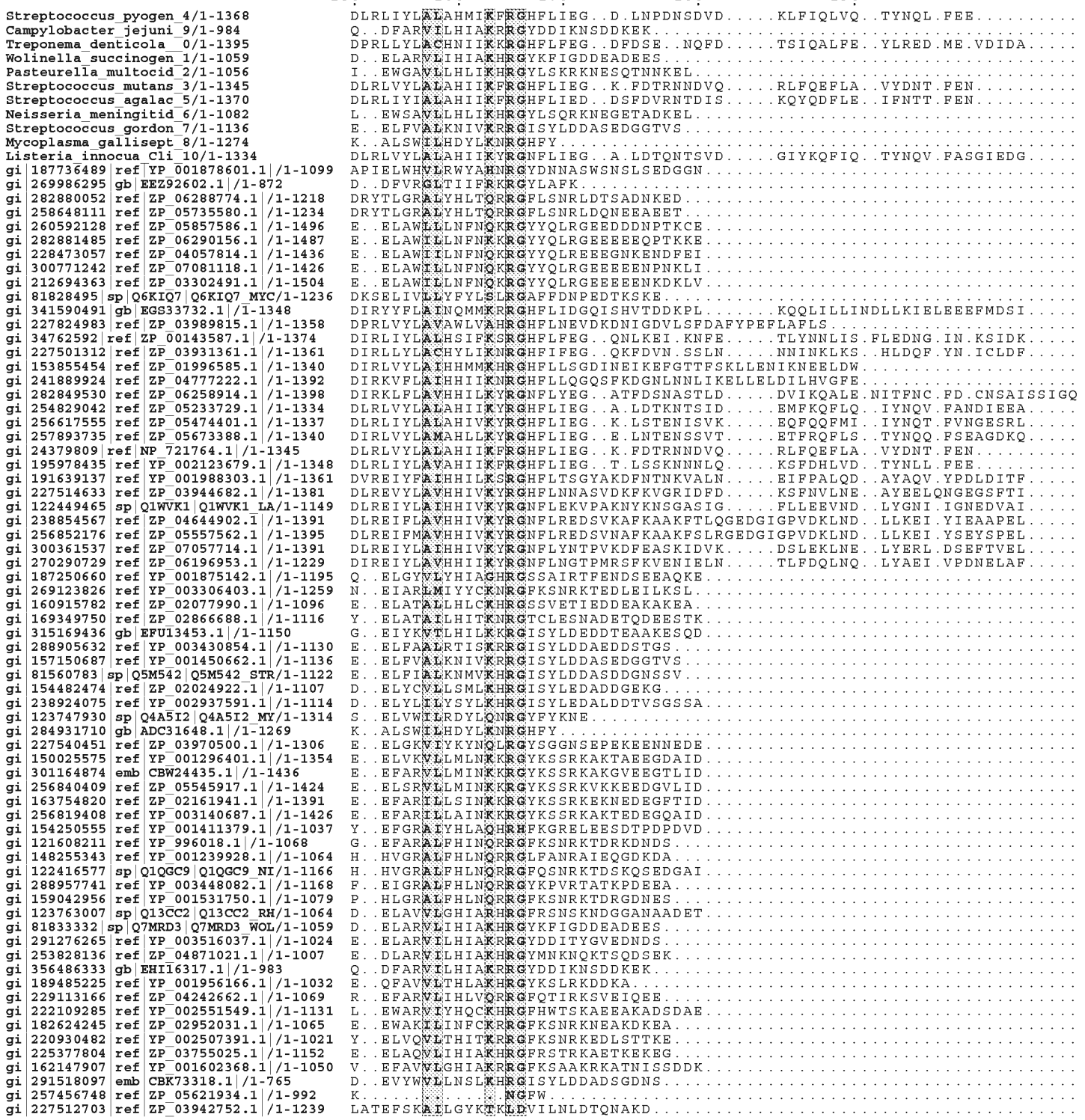

150
160
170
180
190
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 E.

200 210 220 230 240 250
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 F.

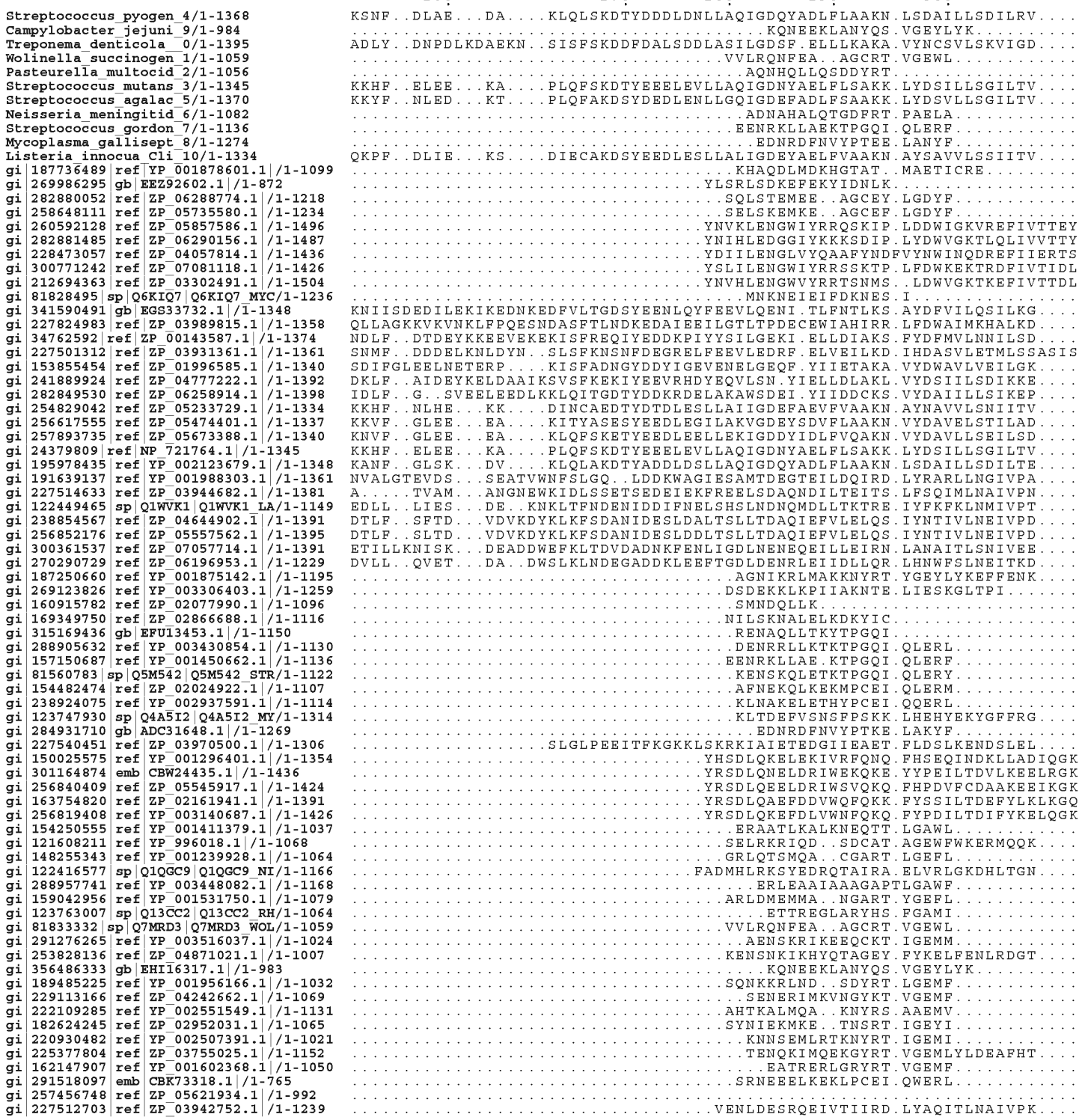

260
270
280
290
300
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 G.

310
320
330
340
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 H.

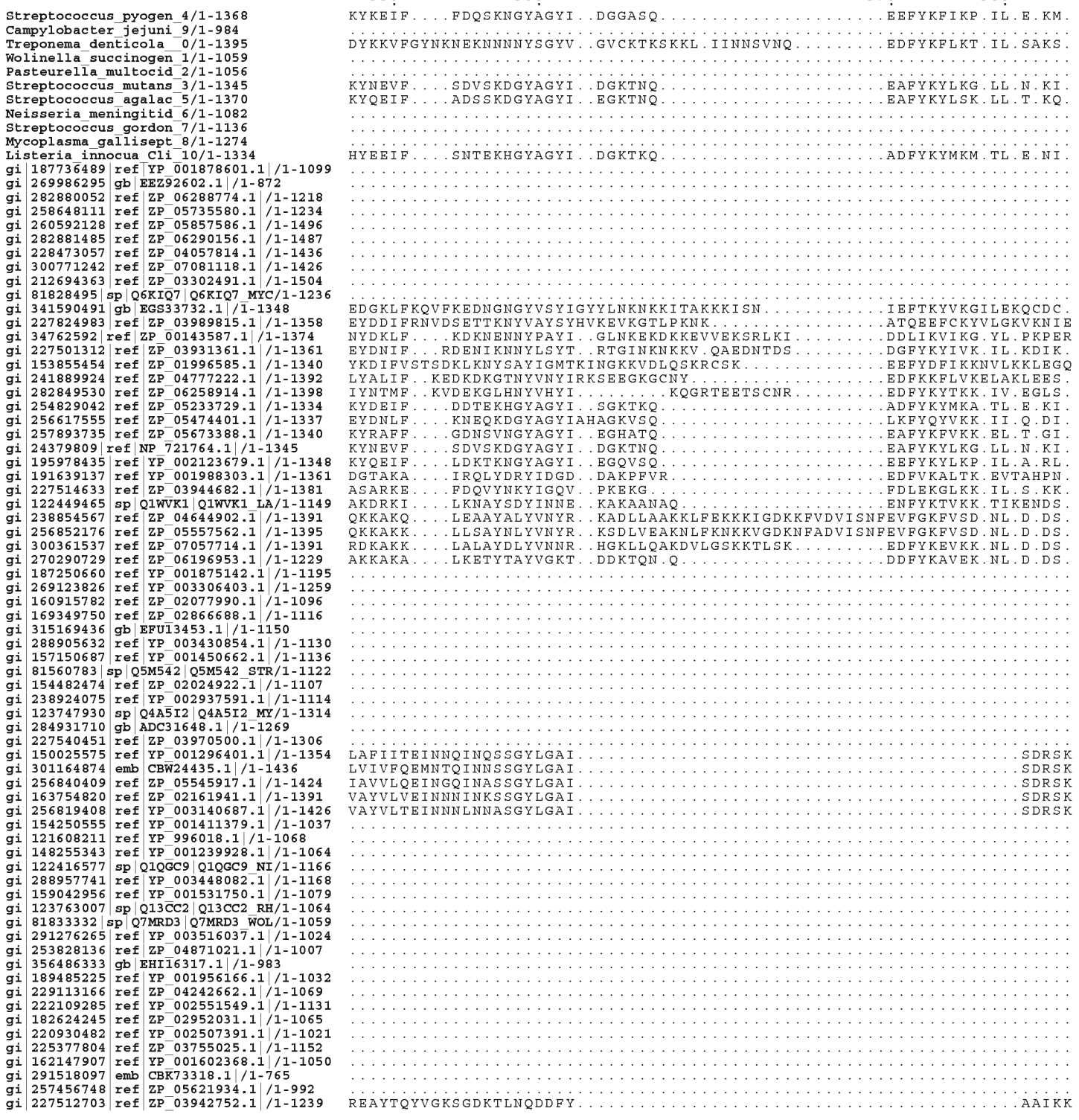

350 360 370 380
Streptococcus_pyogen_4/1-1368 KYKEIF....FDQSKNGYAGYI..DGGASQ.....EEFYKFIKP.IL.E.KM.
Campylobacter_jejuni_9/1-984 .....
Treponema_denticola_0/1-1395 DYKKVFGYNKNEKNNNNYSGYV..GVCKTKSKKL.IINNSVNQ.....EDFYKFLKT.IL.SAKS.
Wolinella_succinogen_1/1-1059 .....
Pasteurella_multocid_2/1-1056 .....
Streptococcus_mutans_3/1-1345 KYNEVF....SDVSKDGYAGYI..DGKTNQ.....EAFYKYLKG.LL.N.KI.
Streptococcus_agalac_5/1-1370 KYQEIF....ADSSKDGYAGYI..EGKTNQ.....EAFYKYLSK.LL.T.KQ.
Neisseria_meningitid_6/1-1082 .....
Streptococcus_gordon_7/1-1136 .....
Mycoplasma_gallisept_8/1-1274 .....
Listeria_innocua_Cli_10/1-1334 HYEEIF....SNTEKHGYAGYI..DGKTKQ.....ADFYKYMKM.TL.E.NI.
gi|187736489|ref|YP_001878601.1|/1-1099 .....
gi|269986295|gb|EEZ92602.1|/1-872 .....
gi|282880052|ref|ZP_06288774.1|/1-1218 .....
gi|258648111|ref|ZP_05735580.1|/1-1234 .....
gi|260592128|ref|ZP_05857586.1|/1-1496 .....
gi|282881485|ref|ZP_06290156.1|/1-1487 .....
gi|228473057|ref|ZP_04057814.1|/1-1436 .....
gi|300771242|ref|ZP_07081118.1|/1-1426 .....
gi|212694363|ref|ZP_03302491.1|/1-1504 .....
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236 .....
gi|341590491|gb|EGS33732.1|/1-1348 EDGKLFKQVFKEDNGNGYVSYIGYYLNKNKKITAKKKISN.....IEFTKYVKGILEKQCDC.
gi|227824983|ref|ZP_03989815.1|/1-1358 EYDDIFRNVDSETTKNYVAYSYHVKEVKGTLPKNK.....ATQEEFCKYVLGKVKNIE
gi|34762592|ref|ZP_00143587.1|/1-1374 NYDKLF....KDKNENNYPAYI..GLNKEKDKKEVVEKSRLKI.....DDLIKVIKG.YL.PKPER
gi|227501312|ref|ZP_03931361.1|/1-1361 EYDNIF....RDENIKNNYLSYT..RTGINKNKKV.QAEDNTDS.....DGFYKYIVK.IL.KDIK.
gi|153855454|ref|ZP_01996585.1|/1-1340 YKDIFVSTSDKLKNYSAYIGMTKINGKKVDLQSKRCSK.....EEFYDFIKKNVLKKLEGQ
gi|241889924|ref|ZP_04777222.1|/1-1392 LYALIF..KEDKDKGTNYVNYIRKSEEGKGCNY.....EDFKKFLVKELAKLEES.
gi|282849530|ref|ZP_06258914.1|/1-1398 IYNTMF..KVDEKGLHNYVHYI.....KQGRTEETSCNR.....EDFYKYTKK.IV.EGLS.
gi|254829042|ref|ZP_05233729.1|/1-1334 KYDEIF....DDTEKHGYAGYI..SGKTKQ.....ADFYKYMKA.TL.E.KI.
gi|256617555|ref|ZP_05474401.1|/1-1337 EYDNLF....KNEQKDGYAGYIAHAGKVSQ.....LKFYQYVKK.II.Q.DI.
gi|257893735|ref|ZP_05673388.1|/1-1340 KYRAFF....GDNSVNGYAGYI..EGHATQ.....EAFYKFVKK.EL.T.GI.
gi|24379809|ref|NP_721764.1|/1-1345 KYNEVF....SDVSKDGYAGYI..DGKTNQ.....EAFYKYLKG.LL.N.KI.
gi|195978435|ref|YP_002123679.1|/1-1348 KYQEIF....LDKTKNGYAGYI..EGQVSQ.....EEFYKYLKP.IL.A.RL.
gi|191639137|ref|YP_001988303.1|/1-1361 DGTAKA....IRQLYDRYIDGD..DAKPFVR.....EDFVKALTK.EVTAHPN.
gi|227514633|ref|ZP_03944682.1|/1-1381 ASARKE....FDQVYNKYIGQV..PKEKG.....FDLEKGLKK.IL.S.KK.
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149 AKDRKI....LKNAYSDYINNE..KAKAANAQ.....ENFYKTVKK.TIKENDS.
gi|238854567|ref|ZP_04644902.1|/1-1391 QKKAKQ....LEAAYALYVNYR..KADLLAAKKLFEKKKIGDKKFVDVISNFEVFGKFVSD.NL.D.DS.
gi|256852176|ref|ZP_05557562.1|/1-1395 QKKAKK....LLSAYNLYVNYR..KSDLVEAKNLFKNKKVGDKNFADVISNFEVFGKFVSD.NL.D.DS.
gi|300361537|ref|ZP_07057714.1|/1-1391 RDKAKK....LALAYDLYVNNR..HGKLLQAKDVLGSKKTLSK.....EDFYKEVKK.NL.D.DS.
gi|270290729|ref|ZP_06196953.1|/1-1229 AKKAKA....LKETYTAYVGKT..DDKTQN.Q.....DDFYKAVEK.NL.D.DS.
gi|187250660|ref|YP_001875142.1|/1-1195 .....
gi|269123826|ref|YP_003306403.1|/1-1259 .....
gi|160915782|ref|ZP_02077990.1|/1-1096 .....
gi|169349750|ref|ZP_02866688.1|/1-1116 .....
gi|315169436|gb|EFU13453.1|/1-1150 .....
gi|288905632|ref|YP_003430854.1|/1-1130 .....
gi|157150687|ref|YP_001450662.1|/1-1136 .....
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122 .....
gi|154482474|ref|ZP_02024922.1|/1-1107 .....
gi|238924075|ref|YP_002937591.1|/1-1114 .....
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314 .....
gi|284931710|gb|ADC31648.1|/1-1269 .....
gi|227540451|ref|ZP_03970500.1|/1-1306 .....
gi|150025575|ref|YP_001296401.1|/1-1354 LAFIITEINNQINQSSGYLGAI.....SDRSK
gi|301164874|emb|CBW24435.1|/1-1436 LVIVFQEMNTQINNSSGYLGAI.....SDRSK
gi|256840409|ref|ZP_05545917.1|/1-1424 IAVVLQEINGQINASSGYLGAI.....SDRSK
gi|163754820|ref|ZP_02161941.1|/1-1391 VAYVLVEINNNNINKSSGYLGAI.....SDRSK
gi|256819408|ref|YP_003140687.1|/1-1426 VAYVLTEINNNLNNASGYLGAI.....SDRSK
gi|154250555|ref|YP_001411379.1|/1-1037 .....
gi|121608211|ref|YP_996018.1|/1-1068 .....
gi|148255343|ref|YP_001239928.1|/1-1064 .....
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166 .....
gi|288957741|ref|YP_003448082.1|/1-1168 .....
gi|159042956|ref|YP_001531750.1|/1-1079 .....
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064 .....
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059 .....
gi|291276265|ref|YP_003516037.1|/1-1024 .....
gi|253828136|ref|ZP_04871021.1|/1-1007 .....
gi|356486333|gb|EHI16317.1|/1-983 .....
gi|189485225|ref|YP_001956166.1|/1-1032 .....
gi|229113166|ref|ZP_04242662.1|/1-1069 .....
gi|222109285|ref|YP_002551549.1|/1-1131 .....
gi|182624245|ref|ZP_02952031.1|/1-1065 .....
gi|220930482|ref|YP_002507391.1|/1-1021 .....
gi|225377804|ref|ZP_03755025.1|/1-1152 .....
gi|162147907|ref|YP_001602368.1|/1-1050 .....
gi|291518097|emb|CBK73318.1|/1-765 .....
gi|257456748|ref|ZP_05621934.1|/1-992 .....
gi|227512703|ref|ZP_03942752.1|/1-1239 REAYTQYVGKSGDKTLNQDDFY.....AAIKK

FIGURE 11 I.

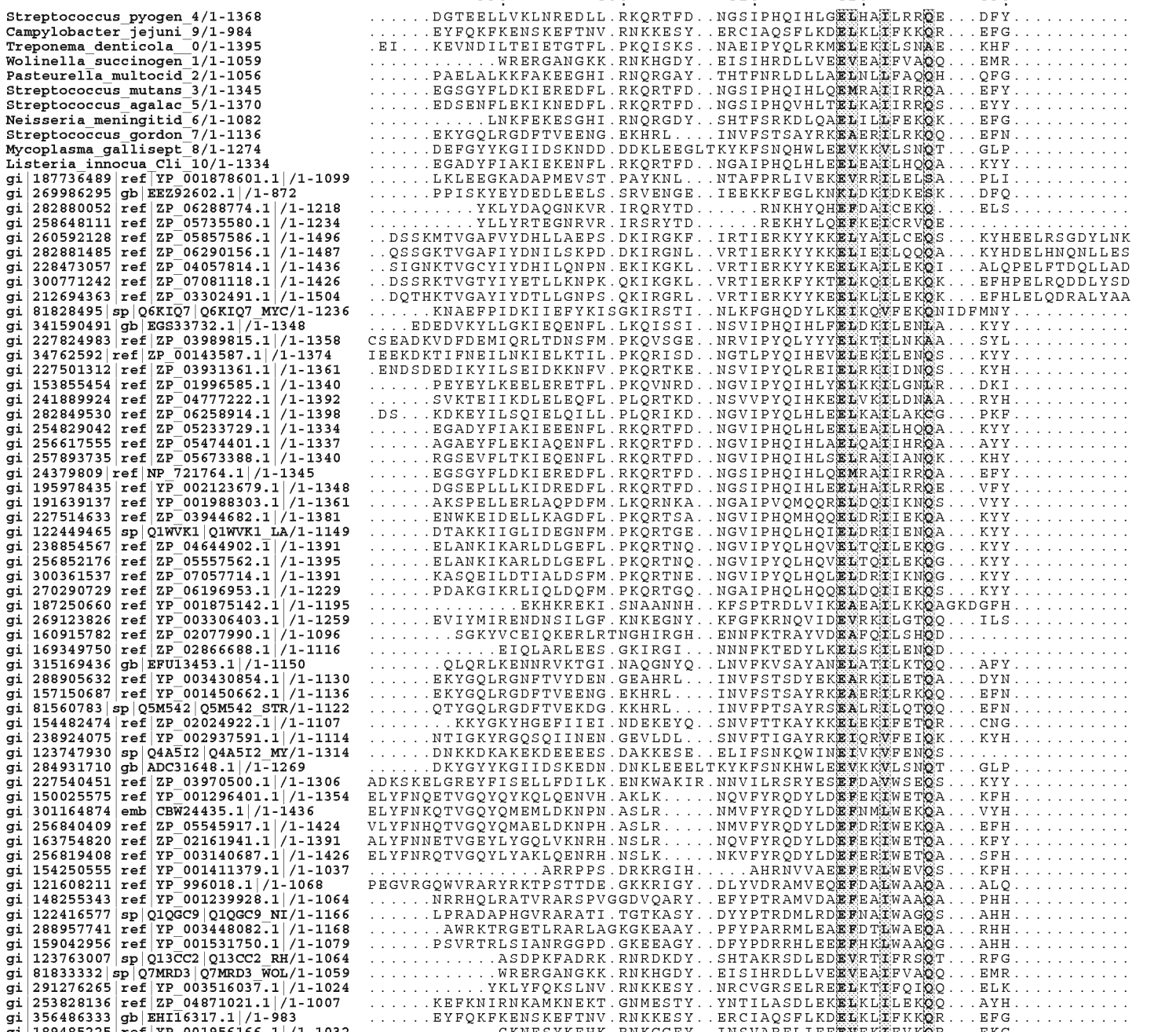

390
400
410
420
430
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 J.

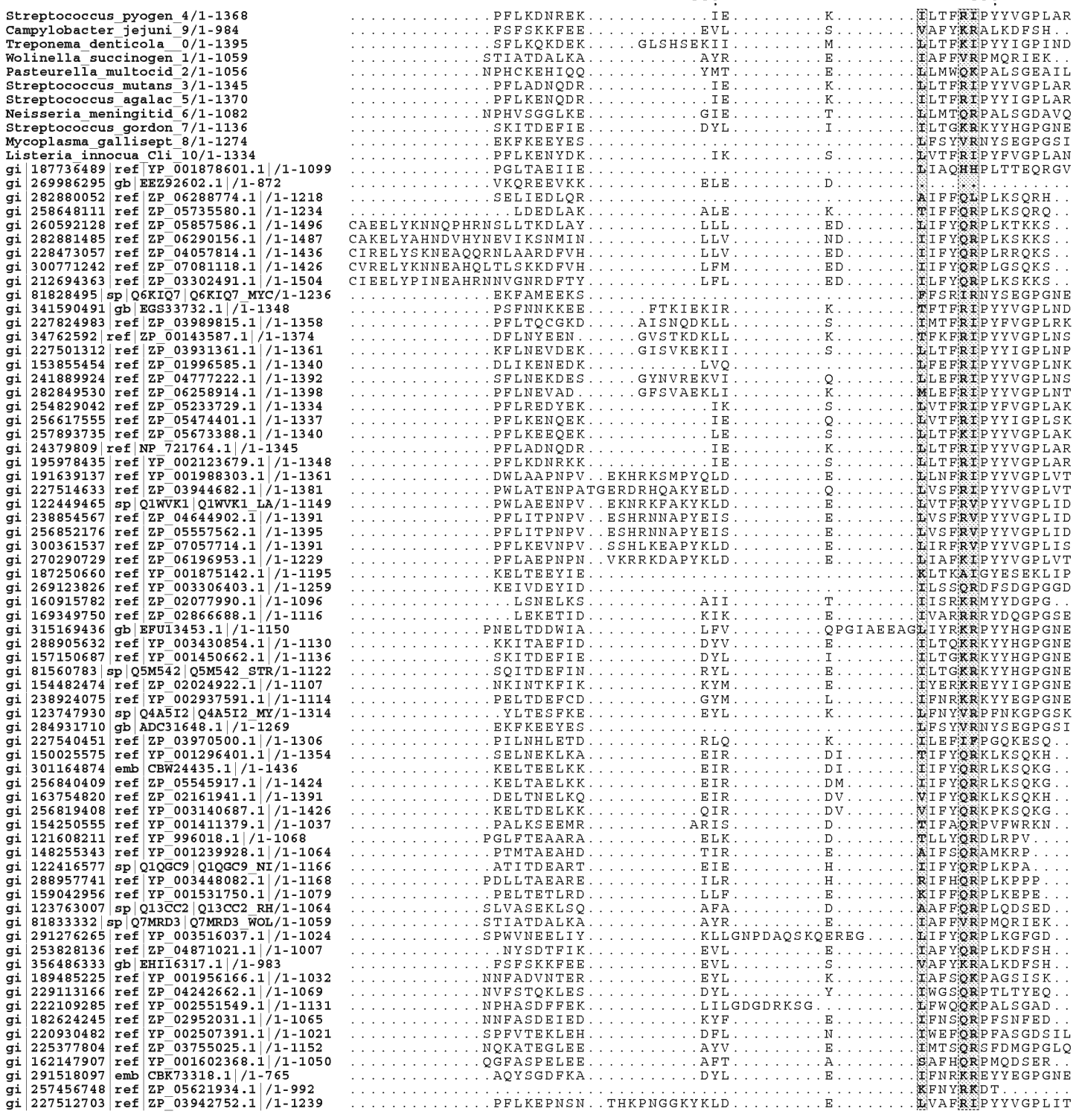

440
450
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 K.

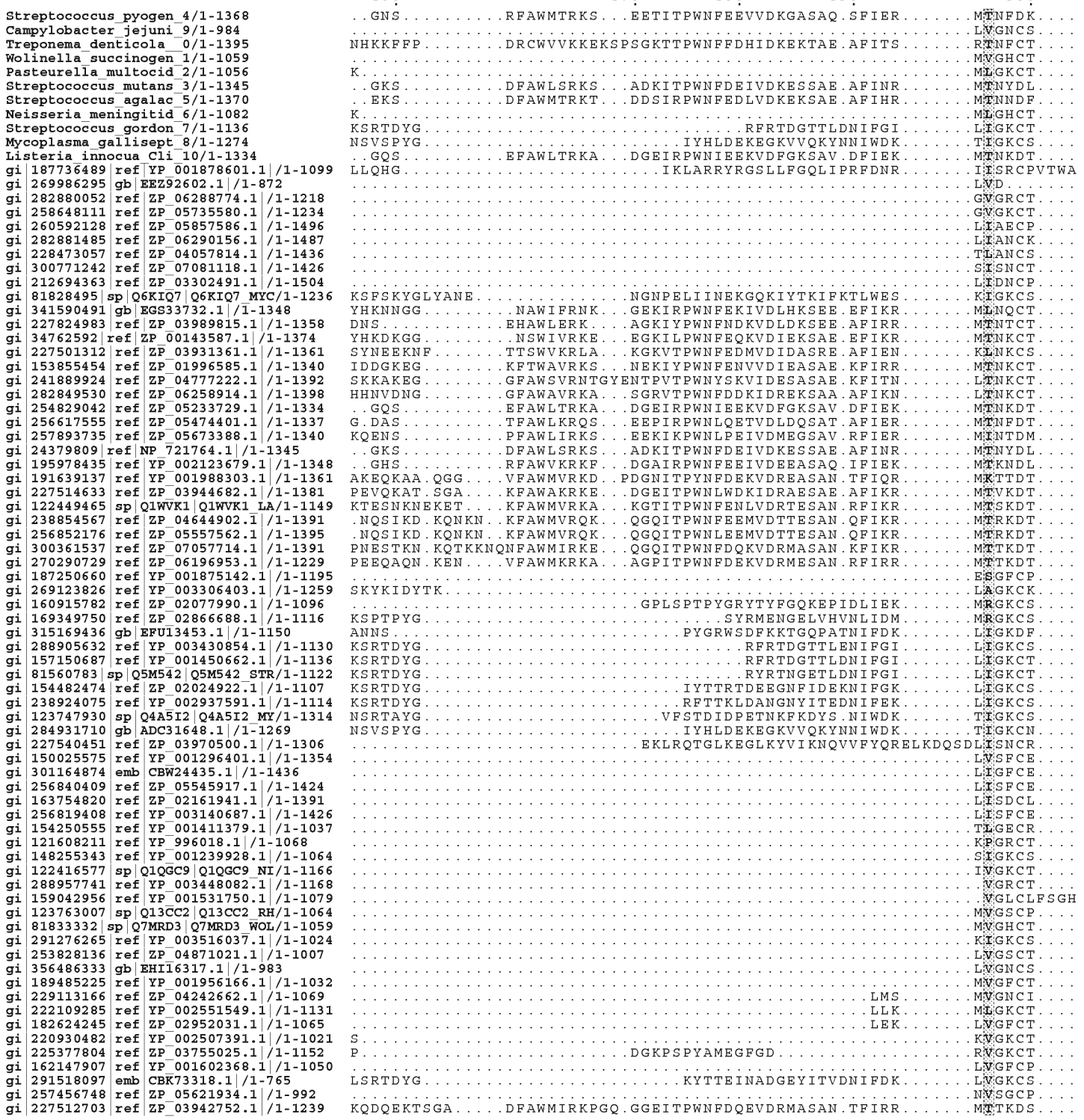

460
470
480
490
500
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 L.

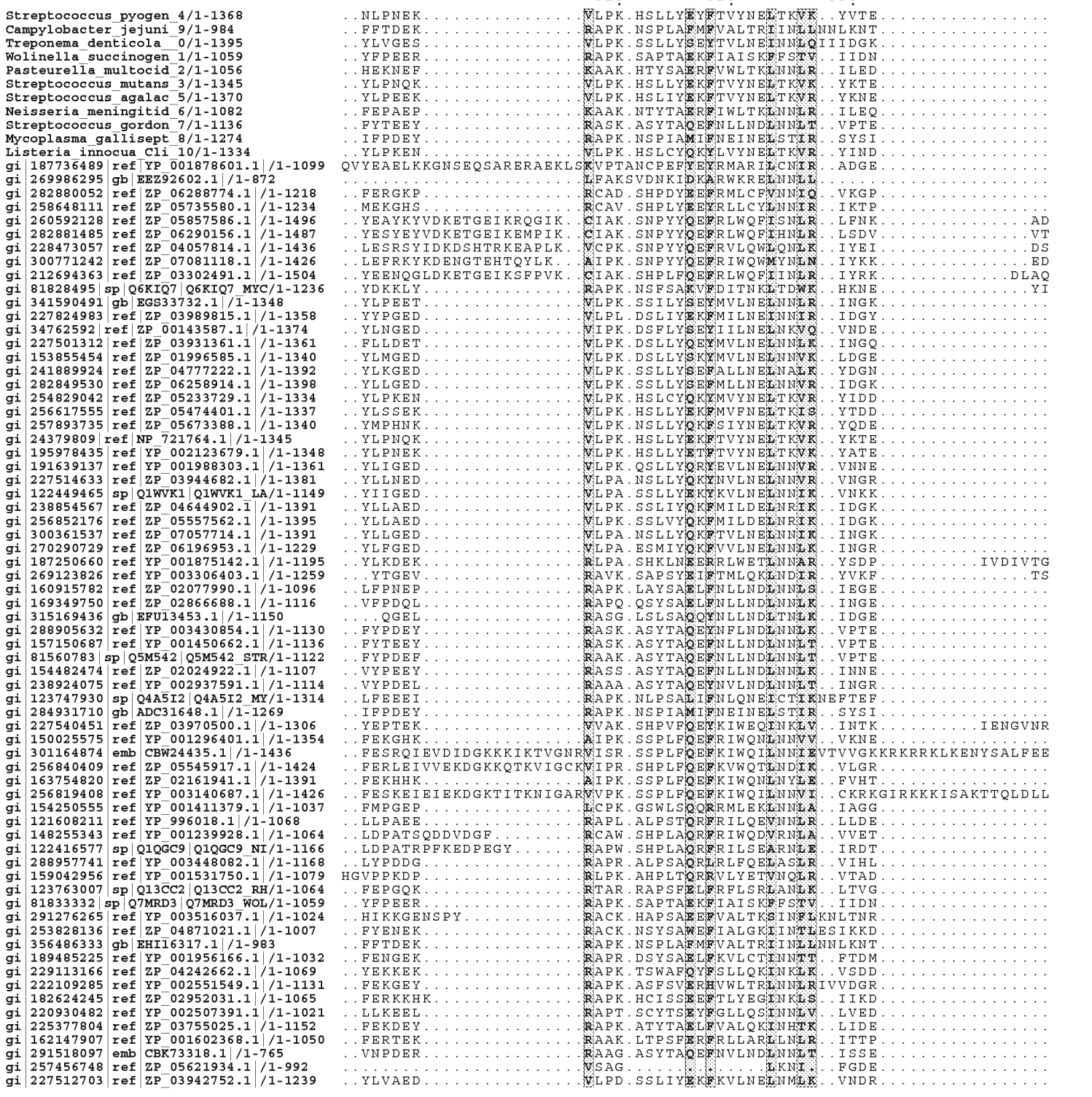

510
520
530
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 M.

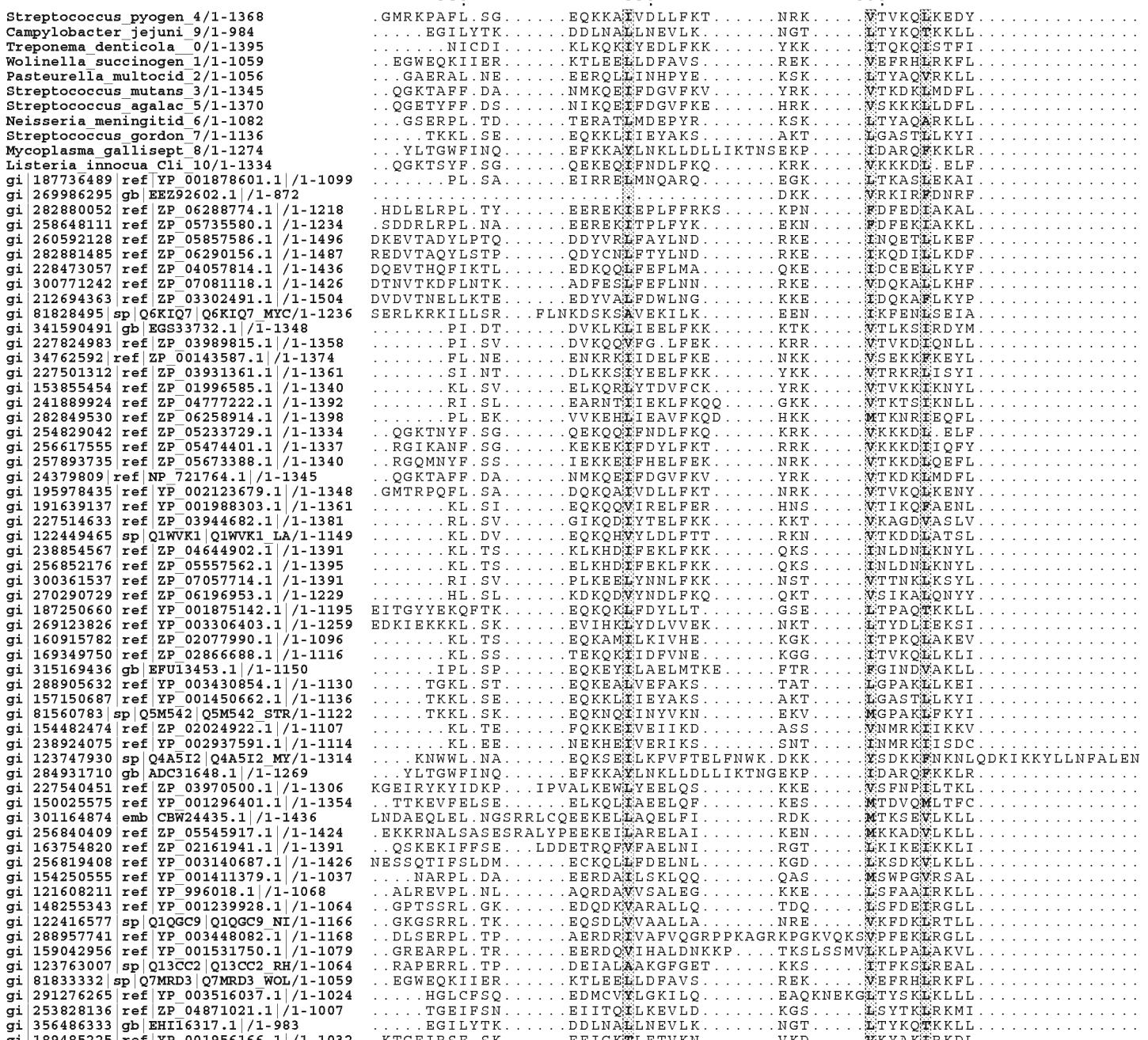

540
550
560
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 N.

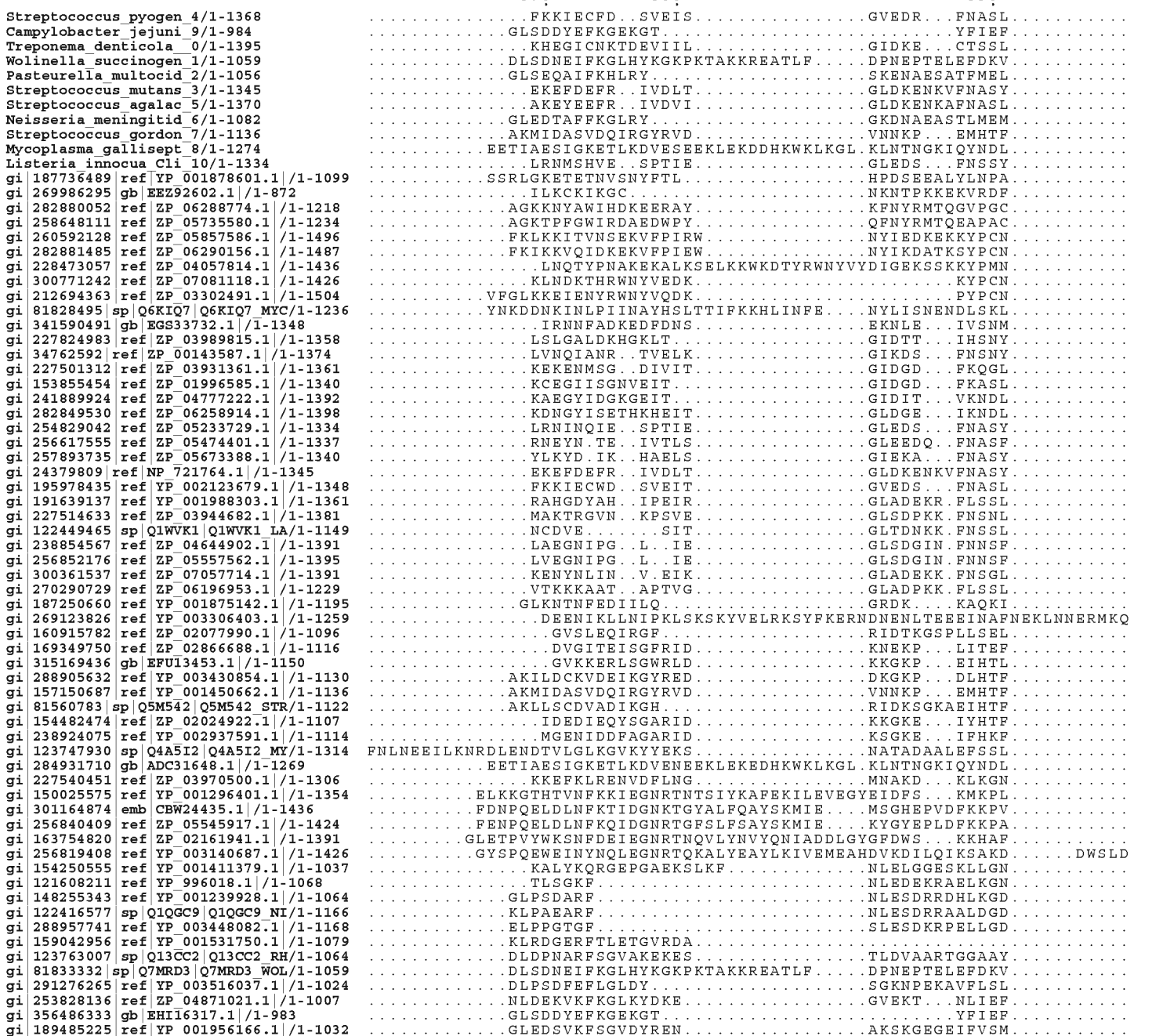

570
580
590
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 O.

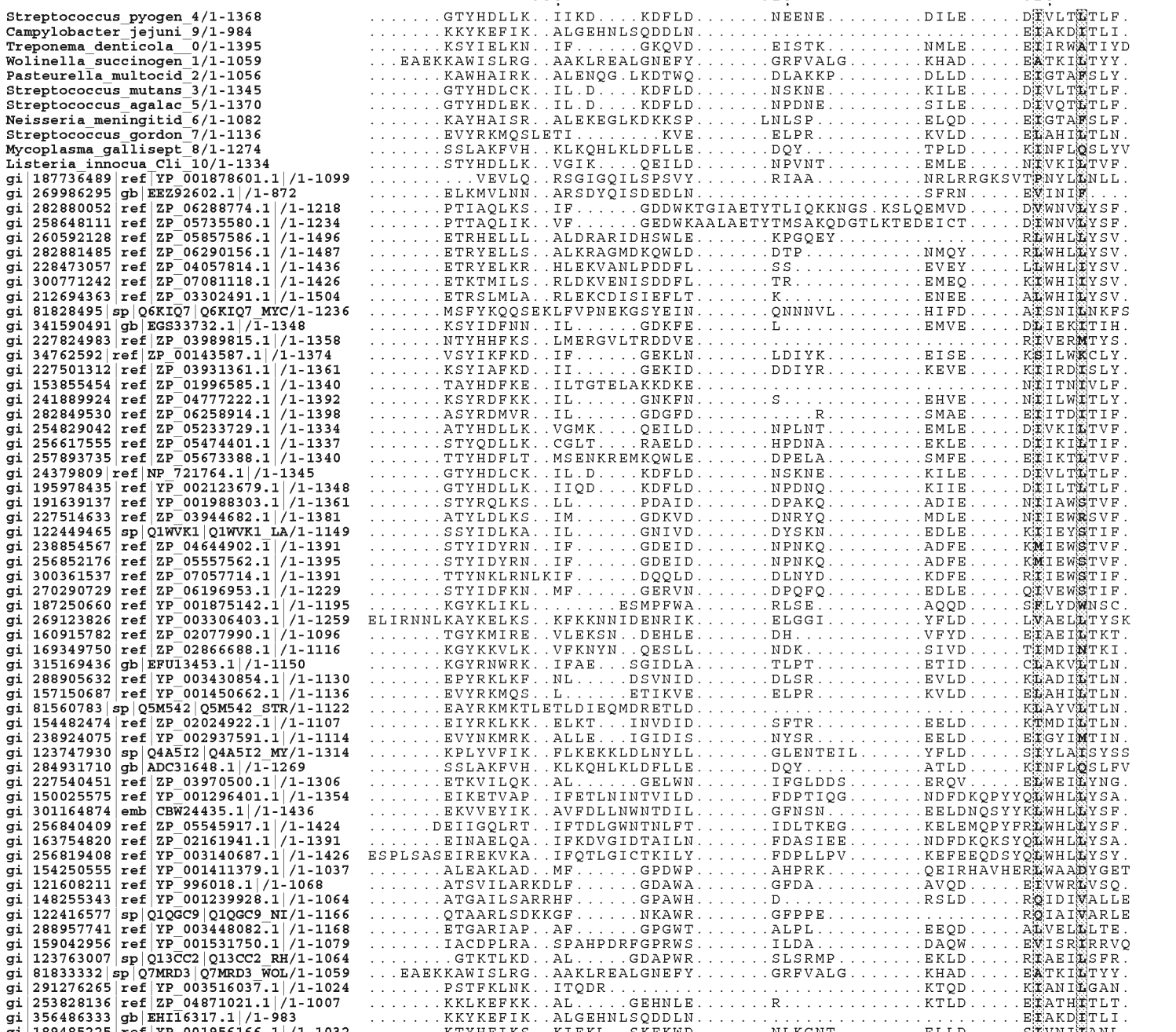

600
610
620
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 P.

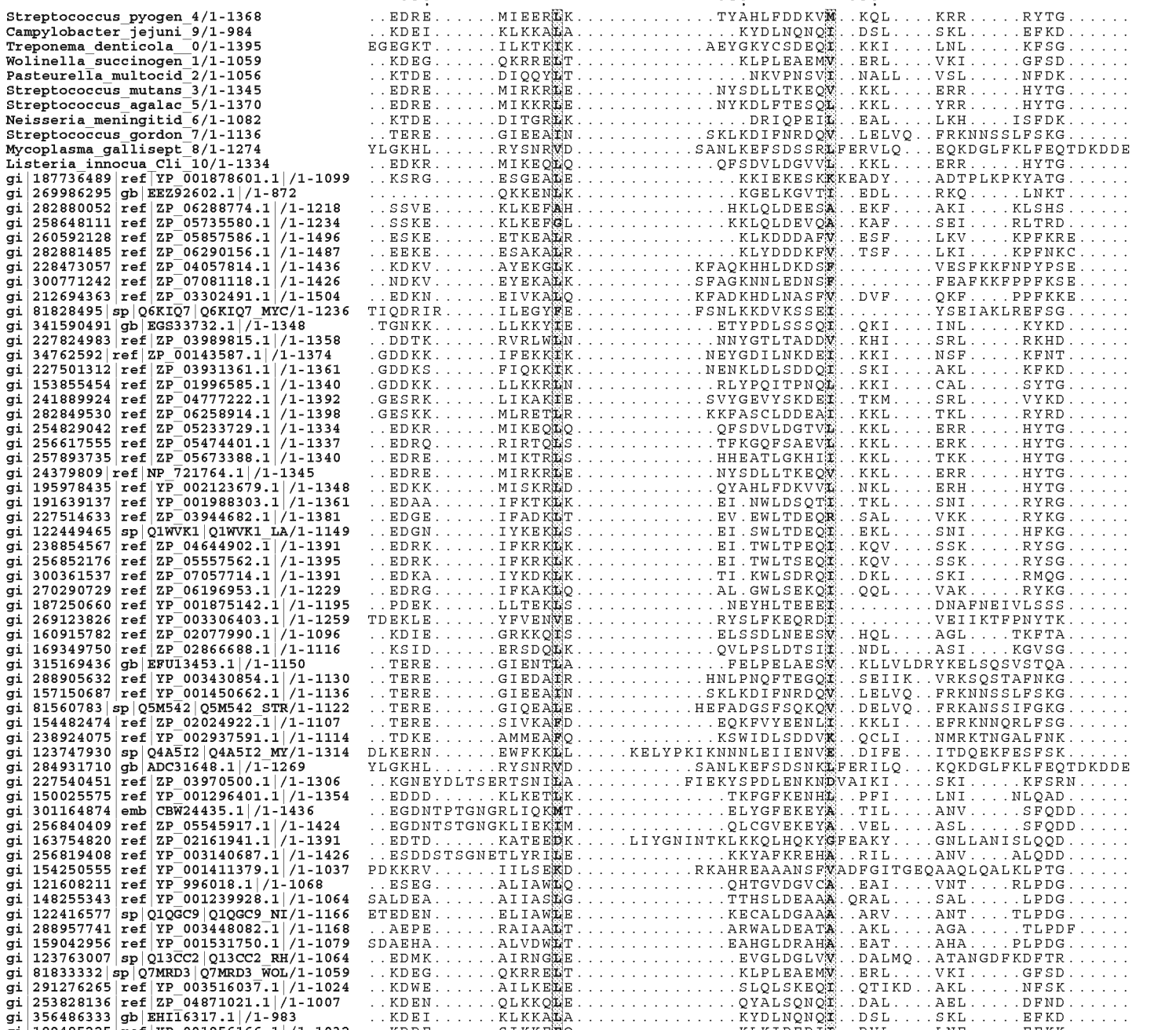

630
640
650
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 Q.

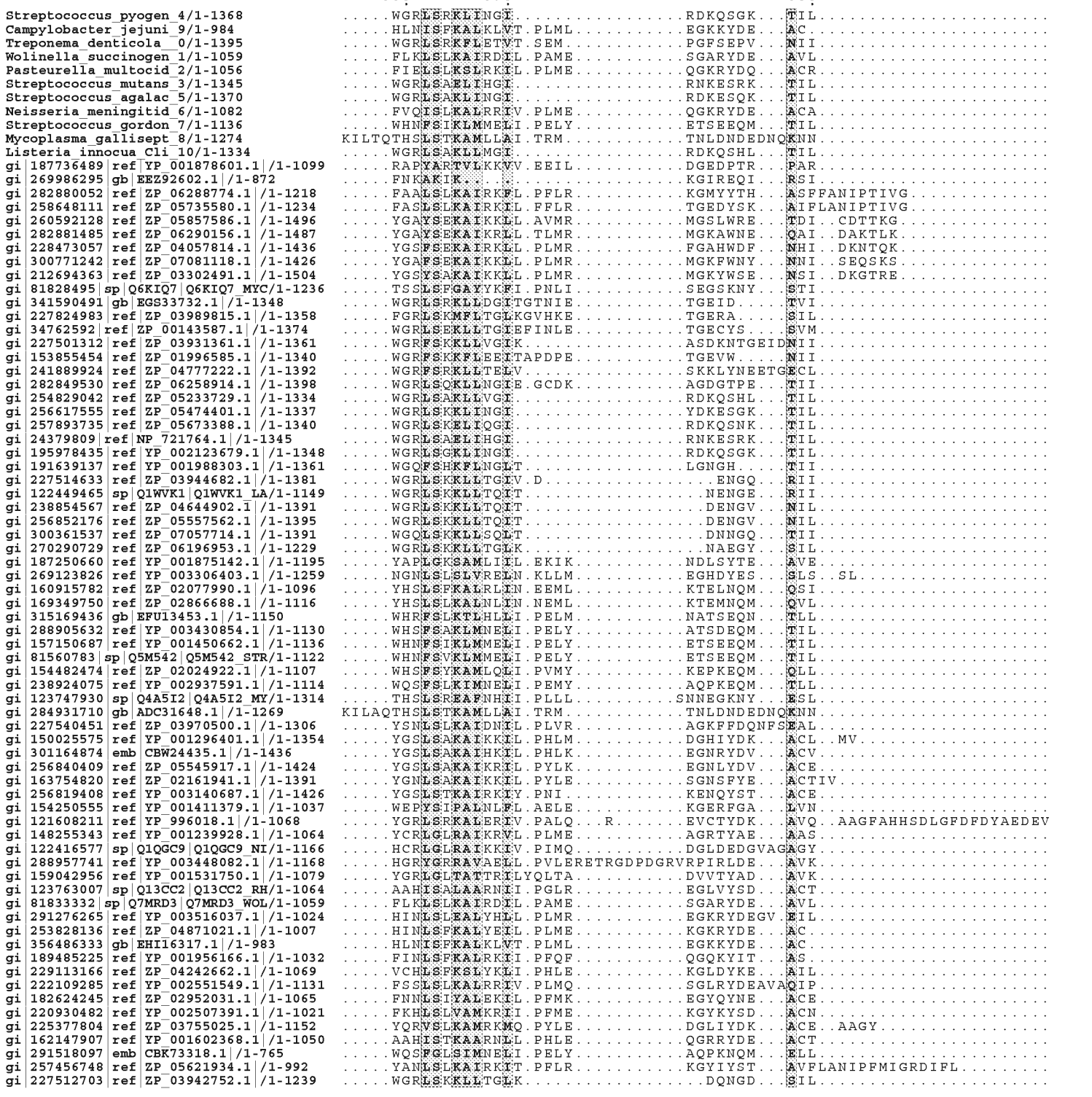

660
670
680
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 R.

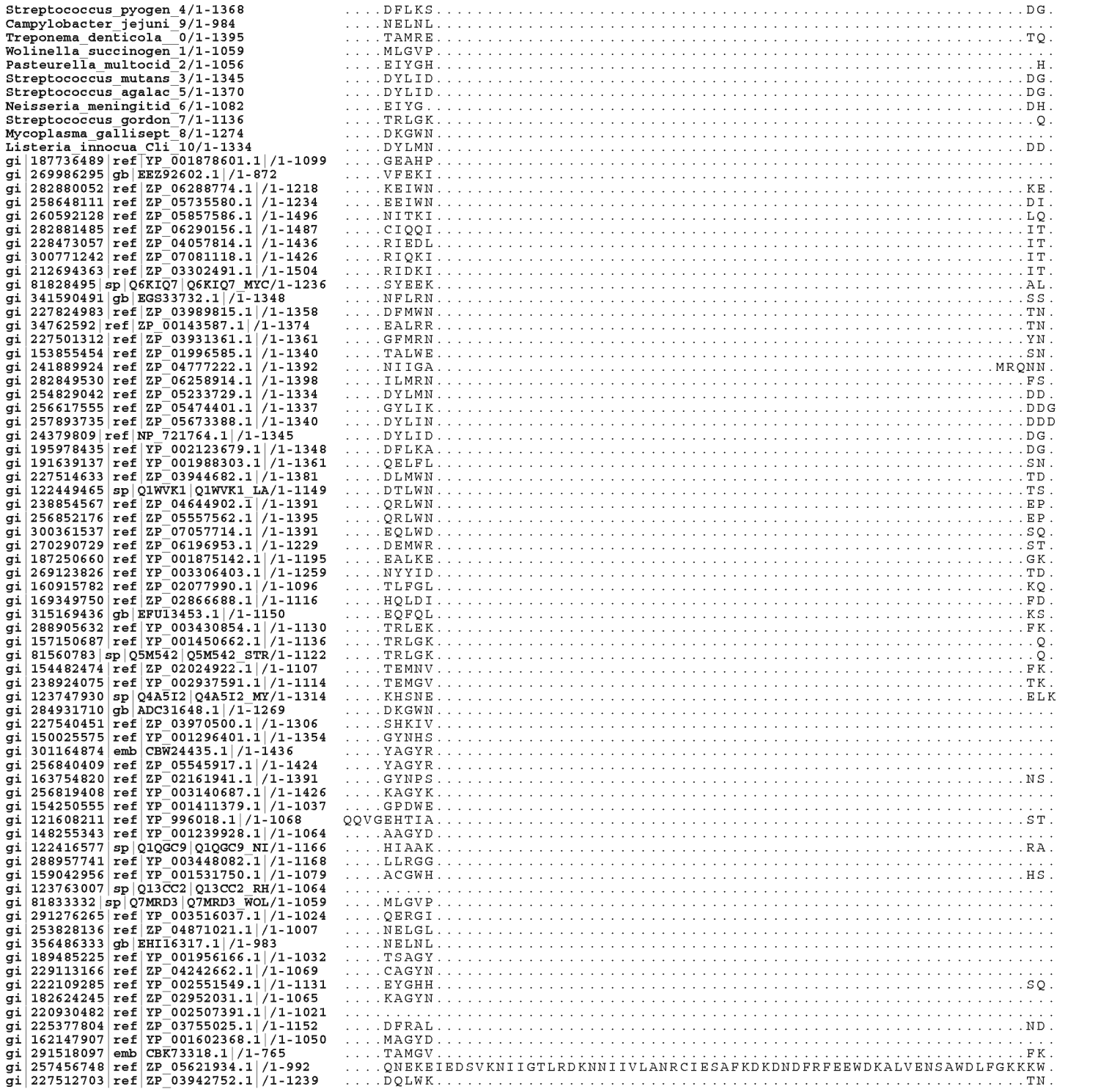

Streptococcus_pyogen_4/1-1368 ....DFLKS....DG.
Campylobacter_jejuni_9/1-984 ....NELNL.....
Treponema_denticola_0/1-1395 ....TAMRE....TQ.
Wolinella_succinogen_1/1-1059 ....MLGVP.....
Pasteurella_multocid_2/1-1056 ....EIYGH....H.
Streptococcus_mutans_3/1-1345 ....DYLID....DG.
Streptococcus_agalac_5/1-1370 ....DYLID....DG.
Neisseria_meningitid_6/1-1082 ....EIYG....DH.
Streptococcus_gordon_7/1-1136 ....TRLGK....Q.
Mycoplasma_gallisept_8/1-1274 ....DKGWN.....
Listeria_innocua_Cli_10/1-1334 ....DYLMN....DD.
gi|187736489|ref|YP_001878601.1|/1-1099 ....GEAHP.....
gi|269986295|gb|EEZ92602.1|/1-872 ....VFEKI.....
gi|282880052|ref|ZP_06288774.1|/1-1218 ....KEIWN....KE.
gi|258648111|ref|ZP_05735580.1|/1-1234 ....EEIWN....DI.
gi|260592128|ref|ZP_05857586.1|/1-1496 ....NITKI....LQ.
gi|282881485|ref|ZP_06290156.1|/1-1487 ....CIQQI....IT.
gi|228473057|ref|ZP_04057814.1|/1-1436 ....RIEDL....IT.
gi|300771242|ref|ZP_07081118.1|/1-1426 ....RIQKI....IT.
gi|212694363|ref|ZP_03302491.1|/1-1504 ....RIDKI....IT.
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236 ....SYEEK....AL.
gi|341590491|gb|EGS33732.1|/1-1348 ....NFLRN....SS.
gi|227824983|ref|ZP_03989815.1|/1-1358 ....DFMWN....TN.
gi|34762592|ref|ZP_00143587.1|/1-1374 ....EALRR....TN.
gi|227501312|ref|ZP_03931361.1|/1-1361 ....GFMRN....YN.
gi|153855454|ref|ZP_01996585.1|/1-1340 ....TALWE....SN.
gi|241889924|ref|ZP_04777222.1|/1-1392 ....NIIGA....MRQNN.
gi|282849530|ref|ZP_06258914.1|/1-1398 ....ILMRN....FS.
gi|254829042|ref|ZP_05233729.1|/1-1334 ....DYLMN....DD.
gi|256617555|ref|ZP_05474401.1|/1-1337 ....GYLIK....DDG
gi|257893735|ref|ZP_05673388.1|/1-1340 ....DYLIN....DDD
gi|24379809|ref|NP_721764.1|/1-1345 ....DYLID....DG.
gi|195978435|ref|YP_002123679.1|/1-1348 ....DFLKA....DG.
gi|191639137|ref|YP_001988303.1|/1-1361 ....QELFL....SN.
gi|227514633|ref|ZP_03944682.1|/1-1381 ....DLMWN....TD.
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149 ....DTLWN....TS.
gi|238854567|ref|ZP_04644902.1|/1-1391 ....QRLWN....EP.
gi|256852176|ref|ZP_05557562.1|/1-1395 ....QRLWN....EP.
gi|300361537|ref|ZP_07057714.1|/1-1391 ....EQLWD....SQ.
gi|270290729|ref|ZP_06196953.1|/1-1229 ....DEMWR....ST.
gi|187250660|ref|YP_001875142.1|/1-1195 ....EALKE....GK.
gi|269123826|ref|YP_003306403.1|/1-1259 ....NYYID....TD.
gi|160915782|ref|ZP_02077990.1|/1-1096 ....TLFGL....KQ.
gi|169349750|ref|ZP_02866688.1|/1-1116 ....HQLDI....FD.
gi|315169436|gb|EFU13453.1|/1-1150 ....EQFQL....KS.
gi|288905632|ref|YP_003430854.1|/1-1130 ....TRLEK....FK.
gi|157150687|ref|YP_001450662.1|/1-1136 ....TRLGK....Q.
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122 ....TRLGK....Q.
gi|154482474|ref|ZP_02024922.1|/1-1107 ....TEMNV....FK.
gi|238924075|ref|YP_002937591.1|/1-1114 ....TEMGV....TK.
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314 ....KHSNE....ELK
gi|284931710|gb|ADC31648.1|/1-1269 ....DKGWN.....
gi|227540451|ref|ZP_03970500.1|/1-1306 ....SHKIV.....
gi|150025575|ref|YP_001296401.1|/1-1354 ....GYNHS.....
gi|301164874|emb|CBW24435.1|/1-1436 ....YAGYR.....
gi|256840409|ref|ZP_05545917.1|/1-1424 ....YAGYR.....
gi|163754820|ref|ZP_02161941.1|/1-1391 ....GYNPS....NS.
gi|256819408|ref|YP_003140687.1|/1-1426 ....KAGYK.....
gi|154250555|ref|YP_001411379.1|/1-1037 ....GPDWE.....
gi|121608211|ref|YP_996018.1|/1-1068 QQVGEHTIA....ST.
gi|148255343|ref|YP_001239928.1|/1-1064 ....AAGYD.....
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166 ....HIAAK....RA.
gi|288957741|ref|YP_003448082.1|/1-1168 ....LLRGG.....
gi|159042956|ref|YP_001531750.1|/1-1079 ....ACGWH....HS.
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064 .....
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059 ....MLGVP.....
gi|291276265|ref|YP_003516037.1|/1-1024 ....QERGI.....
gi|253828136|ref|ZP_04871021.1|/1-1007 ....NELGL.....
gi|356486333|gb|EHI16317.1|/1-983 ....NELNL.....
gi|189485225|ref|YP_001956166.1|/1-1032 ....TSAGY.....
gi|229113166|ref|ZP_04242662.1|/1-1069 ....CAGYN.....
gi|222109285|ref|YP_002551549.1|/1-1131 ....EYGHH....SQ.
gi|182624245|ref|ZP_02952031.1|/1-1065 ....KAGYN.....
gi|220930482|ref|YP_002507391.1|/1-1021 .....
gi|225377804|ref|ZP_03755025.1|/1-1152 ....DFRAL....ND.
gi|162147907|ref|YP_001602368.1|/1-1050 ....MAGYD.....
gi|291518097|emb|CBK73318.1|/1-765 ....TAMGV....FK.
gi|257456748|ref|ZP_05621934.1|/1-992 ....QNEKEIEDSVKNIIGTLRDKNNIIVLANRCIESAFKDKDNDFRFEEWDKALVENSAWDLFGKKKW.
gi|227512703|ref|ZP_03942752.1|/1-1239 ....DQLWK....TN.

FIGURE 11 S.

```
                                                690                 700       710
Streptococcus_pyogen_4/1-1368                ...FANRNFM...QLIHD....DSLTFKEDIQKAQV
Campylobacter_jejuni_9/1-984                 .....KVAI...NEDKK....DFLPAFN
Treponema_denticola_0/1-1395                 .....NNLM...ELLSS....EFTFTENIKKINSG
Wolinella_succinogen_1/1-1059                ..............HKEKS....AILPPL
Pasteurella_multocid_2/1-1056                .....YGEA...NQKTS....QLLP
Streptococcus_mutans_3/1-1345                ...NSNRNFM...QLIND....DALSFKEEIAKAQV
Streptococcus_agalac_5/1-1370                ...RSNRNFM...QLIND....DGLSFKSIISKAQA
Neisseria_meningitid_6/1-1082                .....YGKK...NTEEK....IYLPPI
Streptococcus_gordon_7/1-1136                .....RSKE...TSKRT....KYID
Mycoplasma_gallisept_8/1-1274                .....FEAI...KNFDQ....KFIDITKTNNNLSL
Listeria_innocua_Cli_10/1-1334               ...GLNRNLM...QLIND....SNLSFKSIIEKEQV
gi|187736489|ref|YP_001878601.1|/1-1099      .....DGEL...KAHDG....CLYCLLDTDSSVNQ
gi|269986295|gb|EEZ92602.1|/1-872            .....SGRS...KFCKE....HLKEFSE
gi|282880052|ref|ZP_06288774.1|/1-1218       .....QNRKYIMENVGE....LVFNYQPKHREVQGTIEMLIKDFLANNFELPAGAT
gi|258648111|ref|ZP_05735580.1|/1-1234       .....ERRS...RLLYK....VDLALRPQEERDAEEHRTIAERIKAAVAAEVPDAPID
gi|260592128|ref|ZP_05857586.1|/1-1496       .....GEID...EKLKE....KISASV..SSYNQLSDFQGLPVWLACY
gi|282881485|ref|ZP_06290156.1|/1-1487       .....DNVD...SKLKE....KIDDSR..YVLHHVSDFQGLPTYLASY
gi|228473057|ref|ZP_04057814.1|/1-1436       .....GVEN...EEIRT....ILREKAEKYQLEKETDFQDLPLWLAQY
gi|300771242|ref|ZP_07081118.1|/1-1426       .....GEYD...ENIKD....KVREKA..IHLTTETDFQGLQLWLVQY
gi|212694363|ref|ZP_03302491.1|/1-1504       .....GEYD...EKIEN....RVREKA..IHLADVSDFKGLPLWLVCY
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236      .....QNQK...NNFSH....SNLFE
gi|341590491|gb|EGS33732.1|/1-1348           .....DNLM...QIIGS....QNYSFNNDIDKLRK
gi|227824983|ref|ZP_03989815.1|/1-1358       .....DNLM...QLLSE....CYTFSDEITKLQEA
gi|34762592|ref|ZP_00143587.1|/1-1374        .....YNLM...ELLSS....KFTLQESIDNENKE
gi|227501312|ref|ZP_03931361.1|/1-1361       .....YNIM...QLMAN....VFTFKDEIEKLNSK
gi|153855454|ref|ZP_01996585.1|/1-1340       .....NNLM...QLLSN....EYRFMEEVETYNMG
gi|241889924|ref|ZP_04777222.1|/1-1392       .....ILFM...ELLAD....RFDYSKQIVEFNKE
gi|282849530|ref|ZP_06258914.1|/1-1398       .....YNLM...ELLGD....KFSFMERIQEINAK
gi|254829042|ref|ZP_05233729.1|/1-1334       ...GLNRNLM...QLIND....SNLSFKSIIEKEQV
gi|256617555|ref|ZP_05474401.1|/1-1337       VSKHYNRNFM...QLIND....SQLSFKNAIQKAQS
gi|257893735|ref|ZP_05673388.1|/1-1340       FPHHRNRNFM...QLIND....DSLSFKKEIKKAQM
gi|24379809|ref|NP_721764.1|/1-1345          ...NSNRNFM...QLIND....DALSFKEEIAKAQV
gi|195978435|ref|YP_002123679.1|/1-1348       ...FANRNFM...QLIHD....SELSFIDEIAKAQV
gi|191639137|ref|YP_001988303.1|/1-1361       .....NNLM...QILTD....ETLQKKMTELNADK
gi|227514633|ref|ZP_03944682.1|/1-1381        .....QNFM...QIVNQ....PVFKEQIDQLNQKA
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149       .....NNFI...QVISD....ESIQAKLAEINGEY
gi|238854567|ref|ZP_04644902.1|/1-1391        .....ETLT...EVLAN....PVIKRKISEANSLF
gi|256852176|ref|ZP_05557562.1|/1-1395        .....ETLT...EVLAN....PVIKRKIVEANSSF
gi|300361537|ref|ZP_07057714.1|/1-1391        .....NNFM...QIVNQ....SDFKDAIAVANQNL
gi|270290729|ref|ZP_06196953.1|/1-1229        .....ENFM...QIQSR....PEFAALIQQANEKQ
gi|187250660|ref|YP_001875142.1|/1-1195       .....LTKE...KQAIK....DRLPYYGAVLQESTQKIIAKGF
gi|269123826|ref|YP_003306403.1|/1-1259       .....VDWE...KFPTI....SEIE
gi|160915782|ref|ZP_02077990.1|/1-1096        .....NNEL...SVKGM....KNIQAD
gi|169349750|ref|ZP_02866688.1|/1-1116        .....KNRV...SLKGQ....RQIYAD
gi|315169436|gb|EFU13453.1|/1-1150            .....DVRK...RYSEY....KKLP
gi|288905632|ref|YP_003430854.1|/1-1130       .....VNKK...SSKNT....KTID
gi|157150687|ref|YP_001450662.1|/1-1136       .....RSKE...TSKRT....KYID
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122       .....KTTS...SSNKT....KYID
gi|154482474|ref|ZP_02024922.1|/1-1107        .....SKKE...KYVNY....KYIP
gi|238924075|ref|YP_002937591.1|/1-1114       .....GTQE...EFAGL....KYIP
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314       KRTEK.AELK...AQQNQ....KYLK
gi|284931710|gb|ADC31648.1|/1-1269            .....FEAI...KNFDQ....KFIDITKKNNNLSL
gi|227540451|ref|ZP_03970500.1|/1-1306        .....KLLN...ENVED....PFEKSAQEFLDKNQDVLSDGGIVNAIA
gi|150025575|ref|YP_001296401.1|/1-1354       .....SSLT...TAQNN....DRVLKD
gi|301164874|emb|CBW24435.1|/1-1436           .....HSES...SLTRE....EIANKV
gi|256840409|ref|ZP_05545917.1|/1-1424        .....HSKS...SLTKE....EIENKV
gi|163754820|ref|ZP_02161941.1|/1-1391        .....LTKE...QQQNR....KLKD
gi|256819408|ref|YP_003140687.1|/1-1426       .....HSKL...SLTTE....ELEARE
gi|154250555|ref|YP_001411379.1|/1-1037       .....GWRR...TNFPH....RNQPTGEILDKLPS
gi|121608211|ref|YP_996018.1|/1-1068          .....GEVQ...AQYAF....KQLPYYGKALQRHV
gi|148255343|ref|YP_001239928.1|/1-1064       .....HALLPGGKLSPT....GYLPYYGQWLQNDV
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166       .....GYDH...AKLPTGEQLGRLPYYGQWLQDAV
gi|288957741|ref|YP_003448082.1|/1-1168       .....KDHS...DFSREGALLDALPYYGAVLERHV
gi|159042956|ref|YP_001531750.1|/1-1079       .....DGRT...GECFD.....RLPYYGEVLERHV
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064       .....RVGY...DHAAR....PAVP
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059       ..............HKEKS....AILPPL
gi|291276265|ref|YP_003516037.1|/1-1024       .....FSKP...QPKNR....QLLPPLSELA
gi|253828136|ref|ZP_04871021.1|/1-1007        .....KAKT...SNQKS....EFLPAF
gi|356486333|gb|EHI16317.1|/1-983             .....KVAI...NEDKK....DFLPAFN
gi|189485225|ref|YP_001956166.1|/1-1032       .....SLKQ...EGIKN....KLVPPA
gi|229113166|ref|ZP_04242662.1|/1-1069        .....FRDEN...KRDKK....LFLP
gi|222109285|ref|YP_002551549.1|/1-1131       .....RIEP...GAAKH....LYLPPFYEAQRKYA
gi|182624245|ref|ZP_02952031.1|/1-1065        .....FKAIY...EGVRT....KKLP
gi|220930482|ref|YP_002507391.1|/1-1021       .....MAEL...DFTGS....SKLEKCNKLT
gi|225377804|ref|ZP_03755025.1|/1-1152        .....GNKK...HLLKG....EEINAI
gi|162147907|ref|YP_001602368.1|/1-1050       .....HAAS...RLSHH....GQIVAKTQFN
gi|291518097|emb|CBK73318.1|/1-765            .....SRGD...RFLEC....KEIP
gi|257456748|ref|ZP_05621934.1|/1-992         .....NEYD...EEKRK....TIISQVSEKVEDNLKIAVGKNPNDYKYPLLRTDDLIMDYLNQKGFVVK
gi|227512703|ref|ZP_03942752.1|/1-1239        .....DNFM...QIQSR....DEFAKQIHDENAKQ
```

FIGURE 11 T.

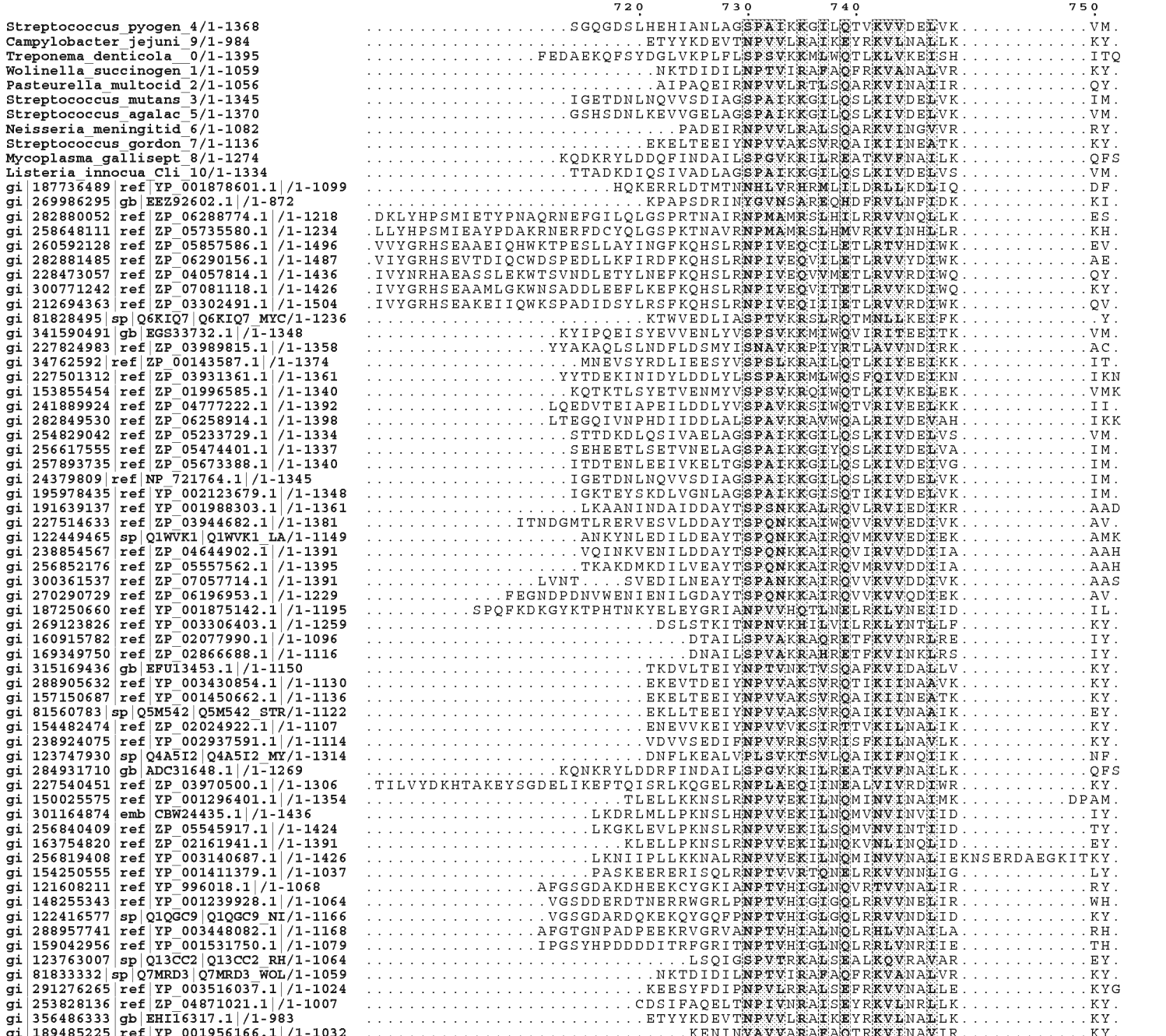

720 730 740 750
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 U.

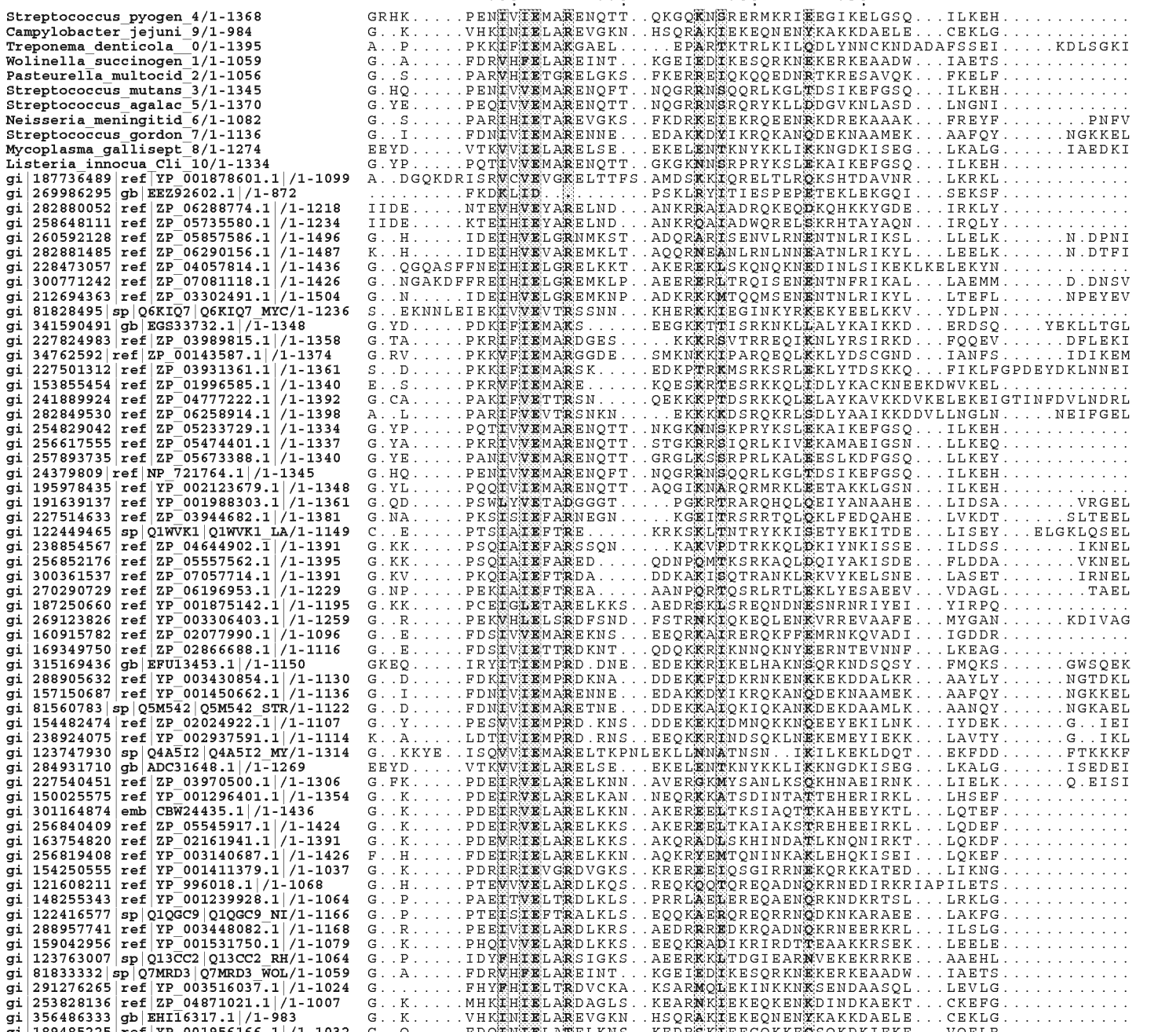

760
770
780
790
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 V.

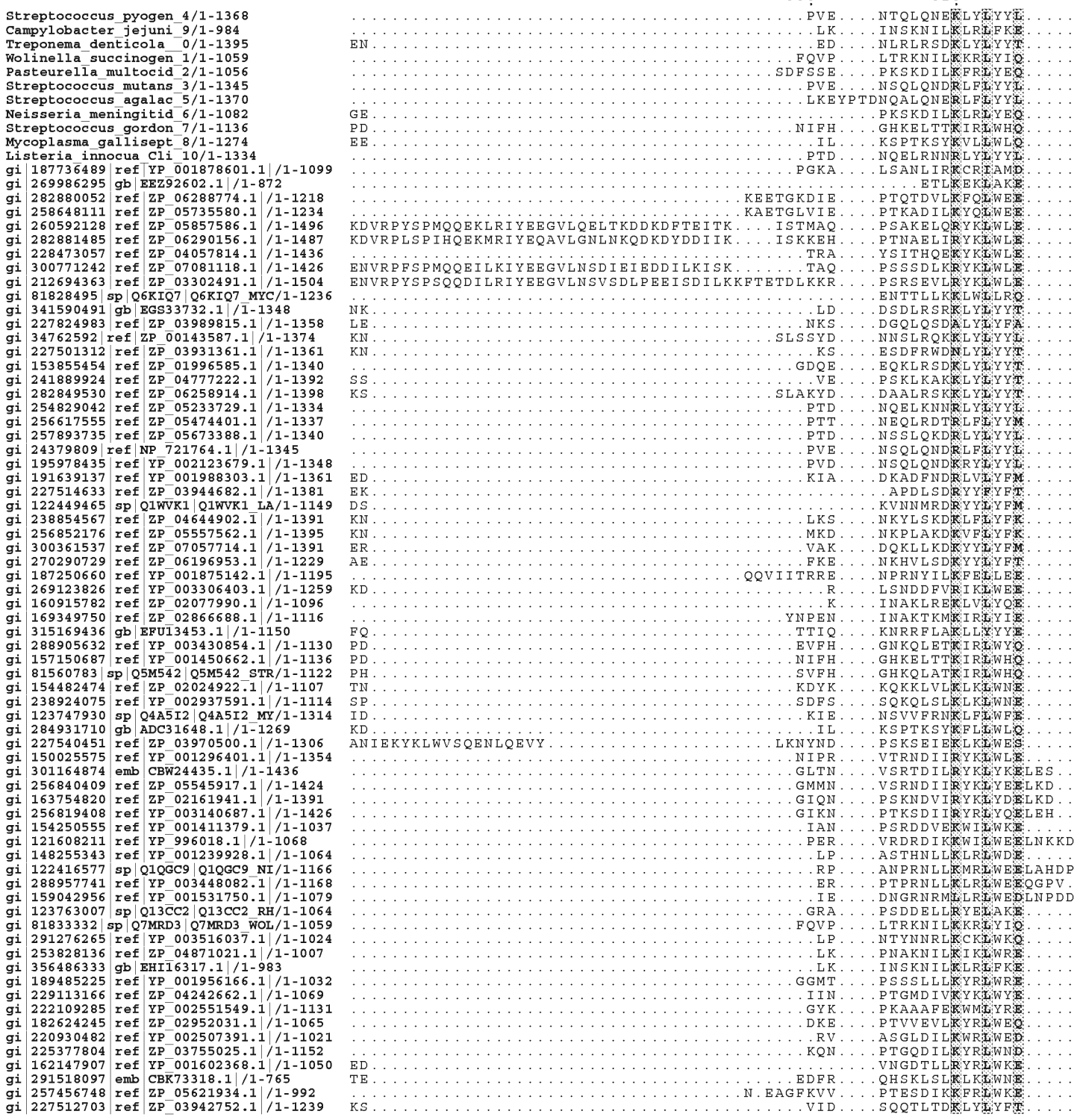

800
810
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 W.

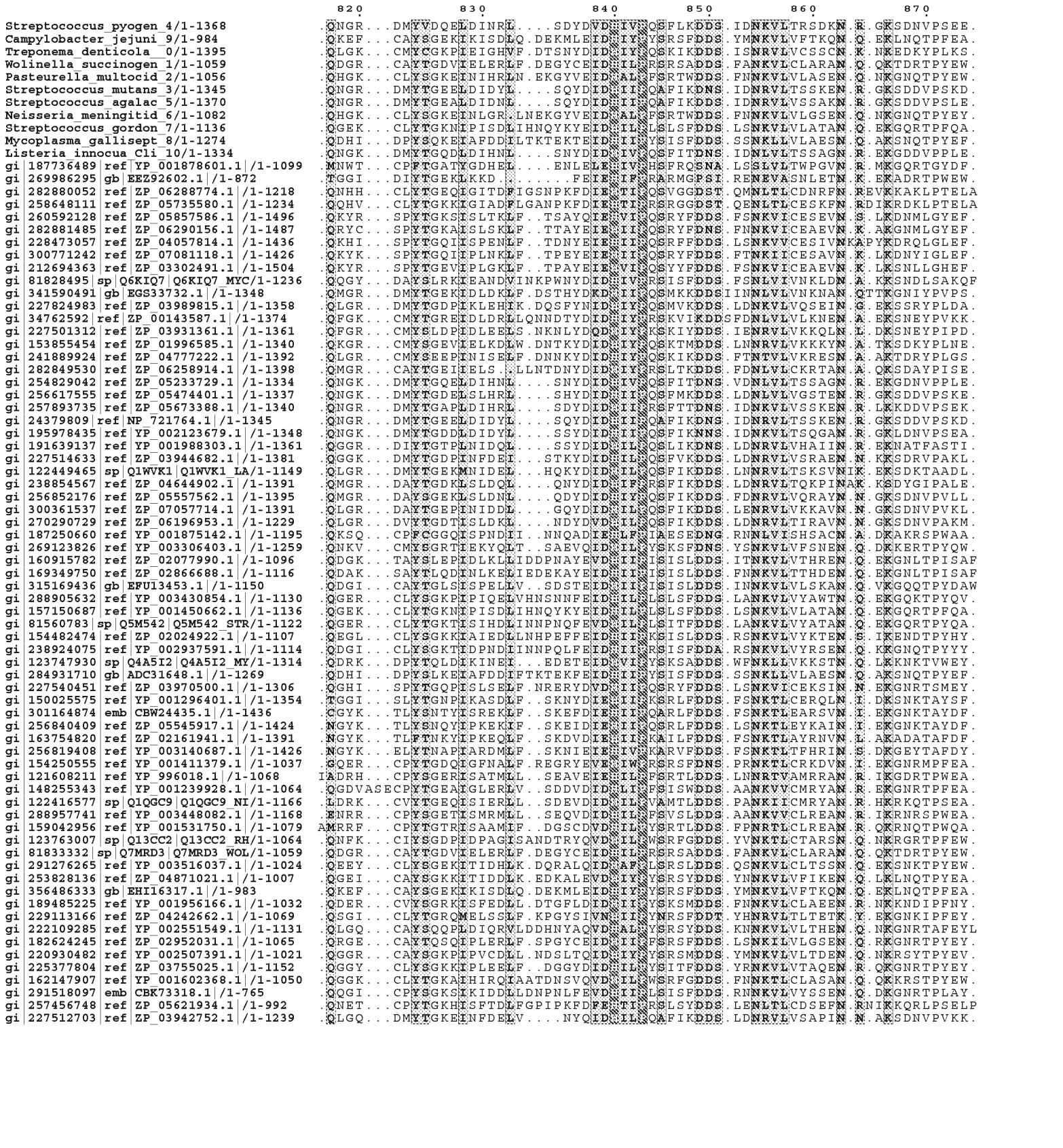

820
830
840
850
860
870
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 X.

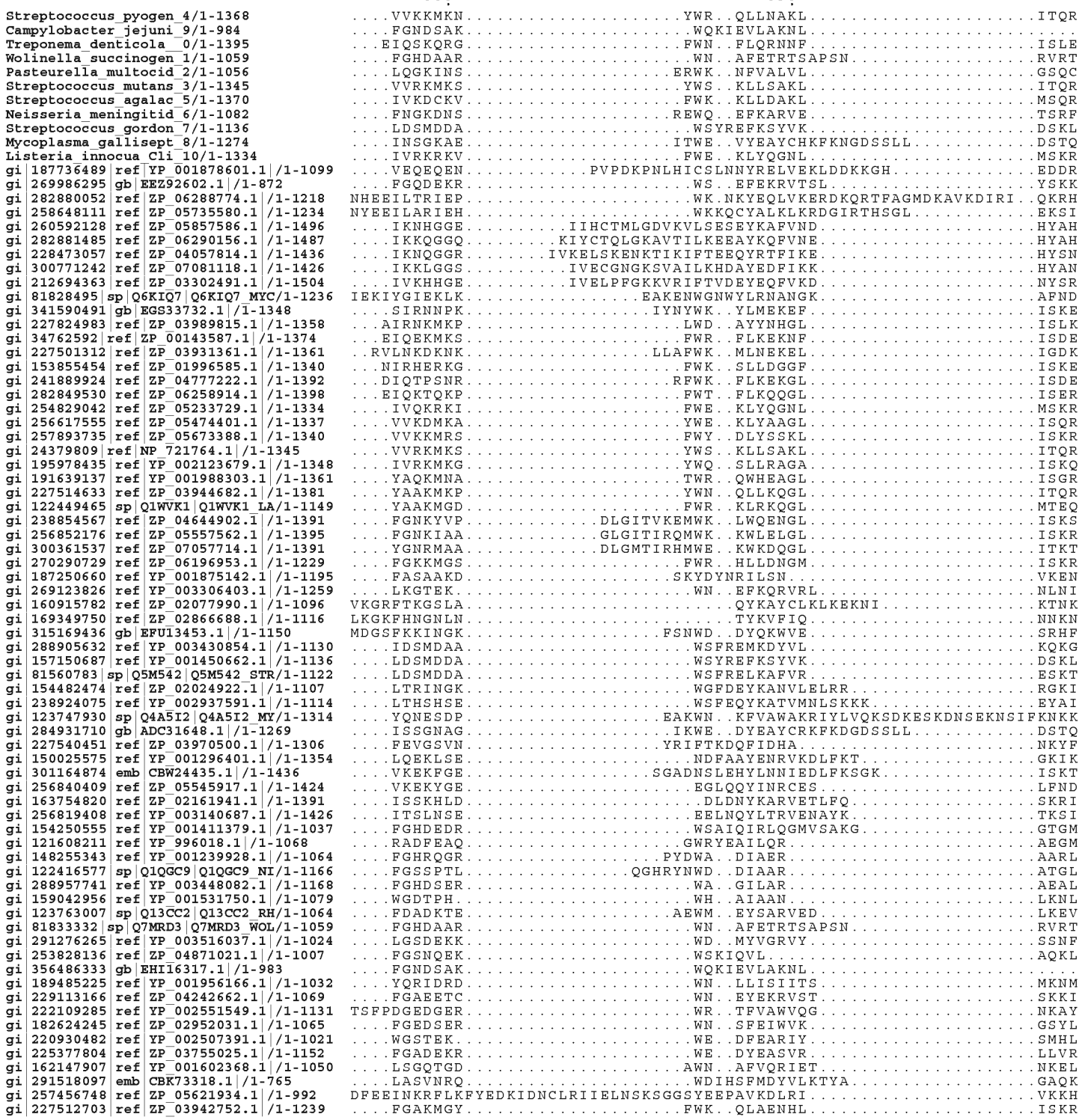

880
890
Streptococcus_pyogen_4/1-1368 ....VVKKMKN.....YWR..QLLNAKL.....ITQR
Campylobacter_jejuni_9/1-984 ....FGNDSAK.....WQKIEVLAKNL.....
Treponema_denticola_0/1-1395 ...EIQSKQRG.....FWN..FLQRNNF.....ISLE
Wolinella_succinogen_1/1-1059 ....FGHDAAR.....WN..AFETRTSAPSN.....RVRT
Pasteurella_multocid_2/1-1056 ....LQGKINS.....ERWK..NFVALVL.....GSQC
Streptococcus_mutans_3/1-1345 ....VVRKMKS.....YWS..KLLSAKL.....ITQR
Streptococcus_agalac_5/1-1370 ....IVKDCKV.....FWK..KLLDAKL.....MSQR
Neisseria_meningitid_6/1-1082 ....FNGKDNS.....REWQ..EFKARVE.....TSRF
Streptococcus_gordon_7/1-1136 ....LDSMDDA.....WSYREFKSYVK.....DSKL
Mycoplasma_gallisept_8/1-1274 ....INSGKAE.....ITWE..VYEAYCHKFKNGDSSLL.....DSTQ
Listeria_innocua_Cli_10/1-1334 ....IVRKRKV.....FWE..KLYQGNL.....MSKR
gi|187736489|ref|YP_001878601.1|/1-1099 ....VEQEQEN.....PVPDKPNLHICSLNNYRELVEKLDDKKGH.....EDDR
gi|269986295|gb|EEZ92602.1|/1-872 ....FGQDEKR.....WS..EFEKRVTSL.....YSKK
gi|282880052|ref|ZP_06288774.1|/1-1218 NHEEILTRIEP.....WK.NKYEQLVKERDKQRTFAGMDKAVKDIRI..QKRH
gi|258648111|ref|ZP_05735580.1|/1-1234 NYEEILARIEH.....WKKQCYALKLKRDGIRTHSGL.....EKSI
gi|260592128|ref|ZP_05857586.1|/1-1496 ....IKNHGGE.....IIHCTMLGDVKVLSESEYKAFVND.....HYAH
gi|282881485|ref|ZP_06290156.1|/1-1487 ....IKKQGGQ.....KIYCTQLGKAVTILKEEAYKQFVNE.....HYAH
gi|228473057|ref|ZP_04057814.1|/1-1436 ....IKNQGGR.....IVKELSKENKTIKIFTEEQYRTFIKE.....HYSN
gi|300771242|ref|ZP_07081118.1|/1-1426 ....IKKLGGS.....IVECGNGKSVAILKHDAYEDFIKK.....HYAN
gi|212694363|ref|ZP_03302491.1|/1-1504 ....IVKHHGE.....IVELPFGKKVRIFTVDEYEQFVKD.....NYSR
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236 IEKIYGIEKLK.....EAKENWGNWYLRNANGK.....AFND
gi|341590491|gb|EGS33732.1|/1-1348 ....SIRNNPK.....IYNYWK..YLMEKEF.....ISKE
gi|227824983|ref|ZP_03989815.1|/1-1358 ...AIRNKMKP.....LWD..AYYNHGL.....ISLK
gi|34762592|ref|ZP_00143587.1|/1-1374 ...EIQEKMKS.....FWR..FLKEKNF.....ISDE
gi|227501312|ref|ZP_03931361.1|/1-1361 ..RVLNKDKNK.....LLAFWK..MLNEKEL.....IGDK
gi|153855454|ref|ZP_01996585.1|/1-1340 ...NIRHERKG.....FWK..SLLDGGF.....ISKE
gi|241889924|ref|ZP_04777222.1|/1-1392 ...DIQTPSNR.....RFWK..FLKEKGL.....ISDE
gi|282849530|ref|ZP_06258914.1|/1-1398 ...EIQKTQKP.....FWT..FLKQQGL.....ISER
gi|254829042|ref|ZP_05233729.1|/1-1334 ....IVQKRKI.....FWE..KLYQGNL.....MSKR
gi|256617555|ref|ZP_05474401.1|/1-1337 ....VVKDMKA.....YWE..KLYAAGL.....ISQR
gi|257893735|ref|ZP_05673388.1|/1-1340 ....VVKKMRS.....FWY..DLYSSKL.....ISKR
gi|24379809|ref|NP_721764.1|/1-1345 ....VVRKMKS.....YWS..KLLSAKL.....ITQR
gi|195978435|ref|YP_002123679.1|/1-1348 ....IVRKMKG.....YWQ..SLLRAGA.....ISKQ
gi|191639137|ref|YP_001988303.1|/1-1361 ....YAQKMNA.....TWR..QWHEAGL.....ISGR
gi|227514633|ref|ZP_03944682.1|/1-1381 ....YAAKMKP.....YWN..QLLKQGL.....ITQR
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149 ....YAAKMGD.....FWR..KLRKQGL.....MTEQ
gi|238854567|ref|ZP_04644902.1|/1-1391 ....FGNKYVP.....DLGITVKEMWK..LWQENGL.....ISKS
gi|256852176|ref|ZP_05557562.1|/1-1395 ....FGNKIAA.....GLGITIRQMWK..KWLELGL.....ISKR
gi|300361537|ref|ZP_07057714.1|/1-1391 ....YGNRMAA.....DLGMTIRHMWE..KWKDQGL.....ITKT
gi|270290729|ref|ZP_06196953.1|/1-1229 ....FGKKMGS.....FWR..HLLDNGM.....ISKR
gi|187250660|ref|YP_001875142.1|/1-1195 ....FASAAKD.....SKYDYNRILSN.....VKEN
gi|269123826|ref|YP_003306403.1|/1-1259 ....LKGTEK.....WN..EFKQRVRL.....NLNI
gi|160915782|ref|ZP_02077990.1|/1-1096 VKGRFTKGSLA.....QYKAYCLKLKEKNI.....KTNK
gi|169349750|ref|ZP_02866688.1|/1-1116 LKGKFHNGNLN.....TYKVFIQ.....NNKN
gi|315169436|gb|EFU13453.1|/1-1150 MDGSFKKINGK.....FSNWD..DYQKWVE.....SRHF
gi|288905632|ref|YP_003430854.1|/1-1130 ....IDSMDAA.....WSFREMKDYVL.....KQKG
gi|157150687|ref|YP_001450662.1|/1-1136 ....LDSMDDA.....WSYREFKSYVK.....DSKL
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122 ....LDSMDDA.....WSFRELKAFVR.....ESKT
gi|154482474|ref|ZP_02024922.1|/1-1107 ....LTRINGK.....WGFDEYKANVLELRR.....RGKI
gi|238924075|ref|YP_002937591.1|/1-1114 ....LTHSHSE.....WSFEQYKATVMNLSKKK.....EYAI
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314 ....YQNESDP.....EAKWN..KFVAWAKRIYLVQKSDKESKDNSEKNSIFKNKK
gi|284931710|gb|ADC31648.1|/1-1269 ....ISSGNAG.....IKWE..DYEAYCRKFKDGDSSLL.....DSTQ
gi|227540451|ref|ZP_03970500.1|/1-1306 ....FEVGSVN.....YRIFTKDQFIDHA.....NKYF
gi|150025575|ref|YP_001296401.1|/1-1354 ....LQEKLSE.....NDFAAYENRVKDLFKT.....GKIK
gi|301164874|emb|CBW24435.1|/1-1436 ....VKEKFGE.....SGADNSLEHYLNNIEDLFKSGK.....ISKT
gi|256840409|ref|ZP_05545917.1|/1-1424 ....VKEKYGE.....EGLQQYINRCES.....LFND
gi|163754820|ref|ZP_02161941.1|/1-1391 ....ISSKHLD.....DLDNYKARVETLFQ.....SKRI
gi|256819408|ref|YP_003140687.1|/1-1426 ....ITSLNSE.....EELNQYLTRVENAYK.....TKSI
gi|154250555|ref|YP_001411379.1|/1-1037 ....FGHDEDR.....WSAIQIRLQGMVSAKG.....GTGM
gi|121608211|ref|YP_996018.1|/1-1068 ....RADFEAQ.....GWRYEAILQR.....AEGM
gi|148255343|ref|YP_001239928.1|/1-1064 ....FGHRQGR.....PYDWA..DIAER.....AARL
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166 ....FGSSPTL.....QGHRYNWD..DIAAR.....ATGL
gi|288957741|ref|YP_003448082.1|/1-1168 ....FGHDSER.....WA..GILAR.....AEAL
gi|159042956|ref|YP_001531750.1|/1-1079 ....WGDTPH.....WH..AIAAN.....LKNL
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064 ....FDADKTE.....AEWM..EYSARVED.....LKEV
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059 ....FGHDAAR.....WN..AFETRTSAPSN.....RVRT
gi|291276265|ref|YP_003516037.1|/1-1024 ....LGSDEKK.....WD..MYVGRVY.....SSNF
gi|253828136|ref|ZP_04871021.1|/1-1007 ....FGSNQEK.....WSKIQVL.....AQKL
gi|356486333|gb|EHI16317.1|/1-983 ....FGNDSAK.....WQKIEVLAKNL.....
gi|189485225|ref|YP_001956166.1|/1-1032 ....YQRIDRD.....WN..LLISIITS.....MKNM
gi|229113166|ref|ZP_04242662.1|/1-1069 ....FGAEETC.....WN..EYEKRVST.....SKKI
gi|222109285|ref|YP_002551549.1|/1-1131 TSFPDGEDGER.....WR..TFVAWVQG.....NKAY
gi|182624245|ref|ZP_02952031.1|/1-1065 ....FGEDSER.....WN..SFEIWVK.....GSYL
gi|220930482|ref|YP_002507391.1|/1-1021 ....WGSTEK.....WE..DFEARIY.....SMHL
gi|225377804|ref|ZP_03755025.1|/1-1152 ....FGADEKR.....WE..DYEASVR.....LLVR
gi|162147907|ref|YP_001602368.1|/1-1050 ....LSGQTGD.....AWN..AFVQRIET.....NKEL
gi|291518097|emb|CBK73318.1|/1-765 ....LASVNRQ.....WDIHSFMDYVLKTYA.....GAQK
gi|257456748|ref|ZP_05621934.1|/1-992 DFEEINKRFLKFYEDKIDNCLRIIELNSKSGGSYEEPAVKDLRI.....VKKH
gi|227512703|ref|ZP_03942752.1|/1-1239 ....FGAKMGY.....FWK..QLAENHL.....ISKR

FIGURE 11 Y.

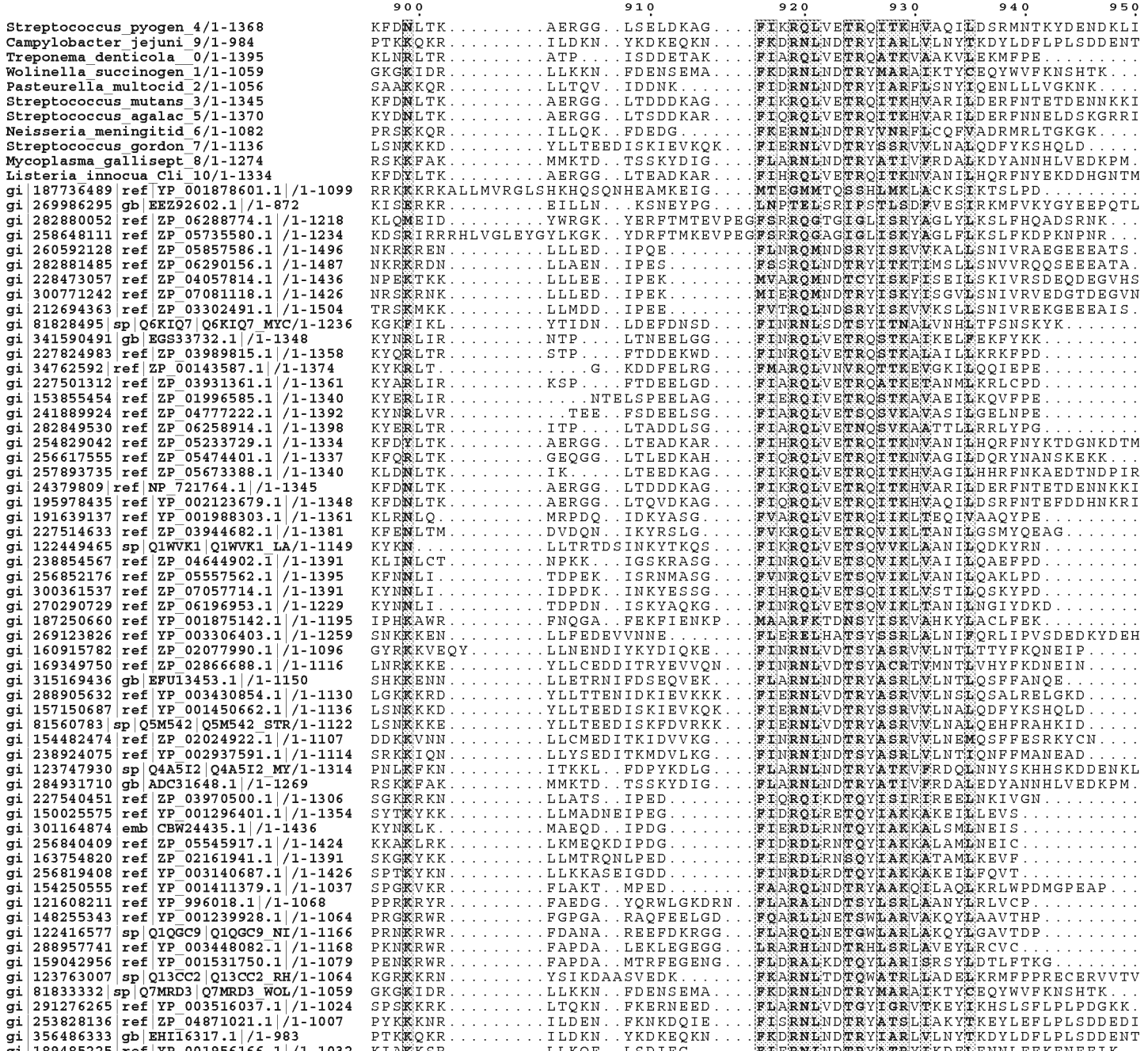

900
910
920
930
940
950
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 Z.

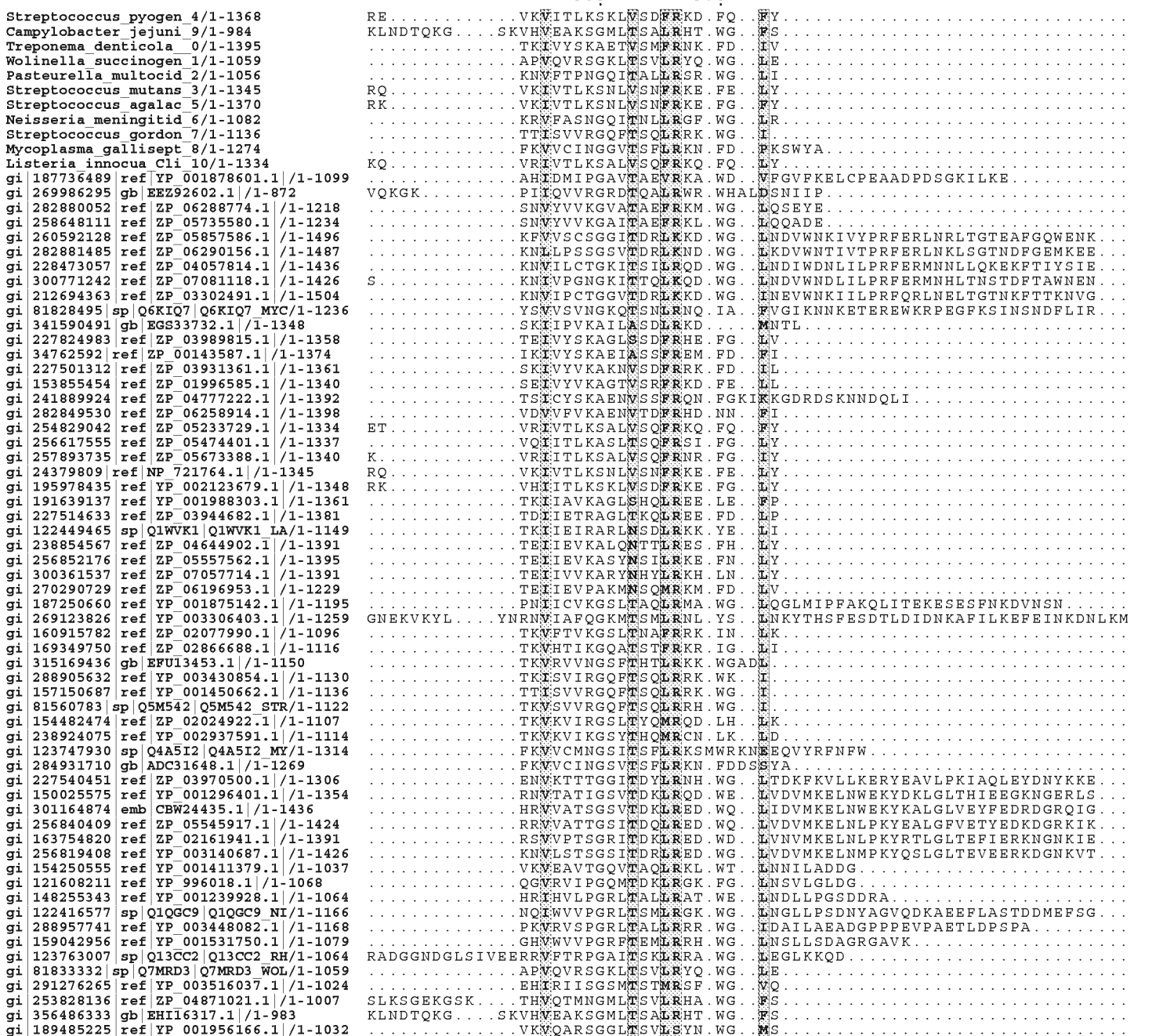

960
970
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AA.

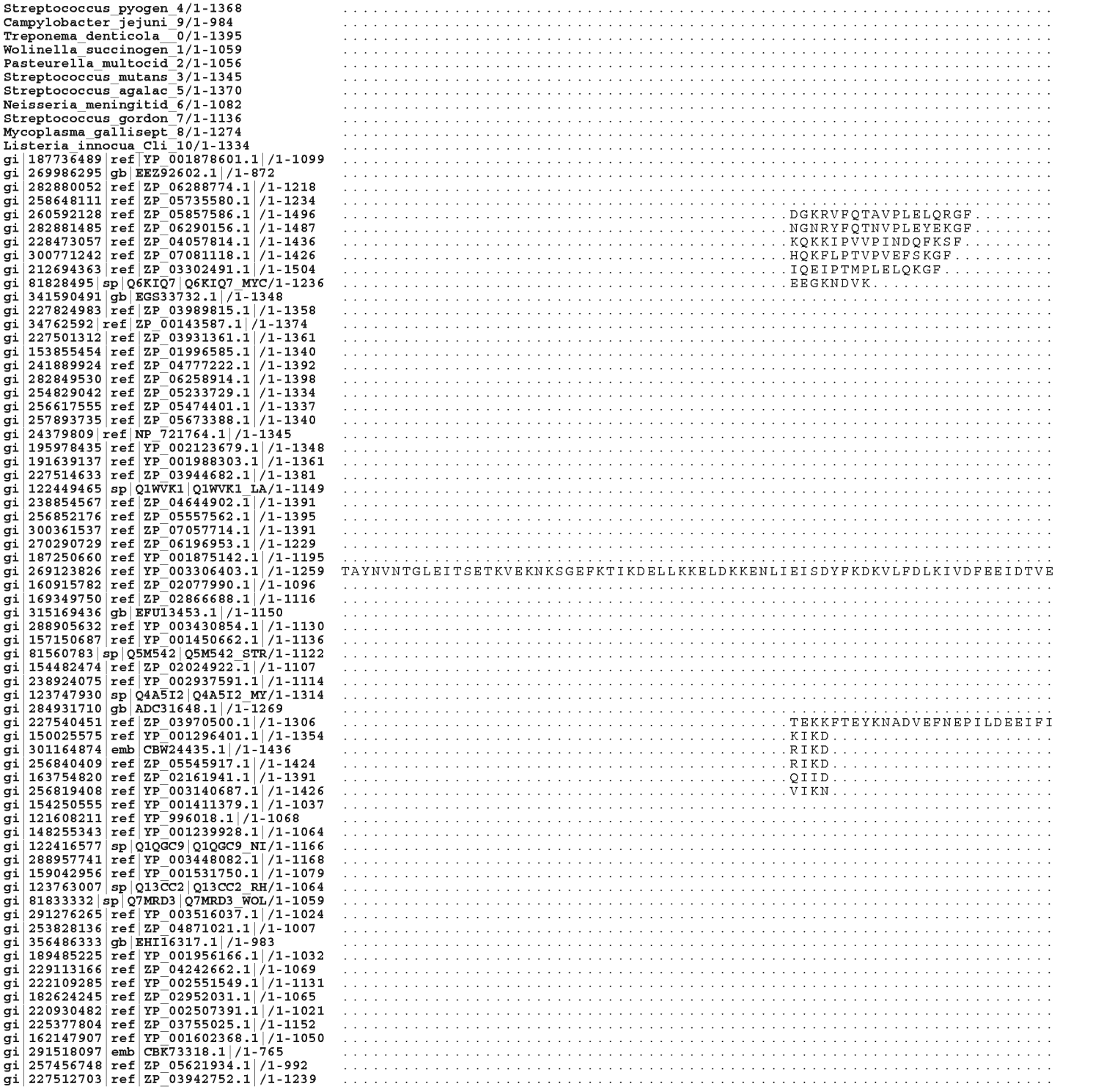

Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496 DGKRVFQTAVPLELQRGF
gi|282881485|ref|ZP_06290156.1|/1-1487 NGNRYFQTNVPLEYEKGF
gi|228473057|ref|ZP_04057814.1|/1-1436 KQKKIPVVPINDQFKSF
gi|300771242|ref|ZP_07081118.1|/1-1426 HQKFLPTVPVEFSKGF
gi|212694363|ref|ZP_03302491.1|/1-1504 IQEIPTMPLELQKGF
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236 EEGKNDVK
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259 TAYNVNTGLEITSETKVEKNKSGEFKTIKDELLKKELDKKENLIEISDYFKDKVLFDLKIVDFEEIDTVE
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306 TEKKFTEYKNADVEFNEPILDEEIFI
gi|150025575|ref|YP_001296401.1|/1-1354 KIKD
gi|301164874|emb|CBW24435.1|/1-1436 RIKD
gi|256840409|ref|ZP_05545917.1|/1-1424 RIKD
gi|163754820|ref|ZP_02161941.1|/1-1391 QIID
gi|256819408|ref|YP_003140687.1|/1-1426 VIKN
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AB.

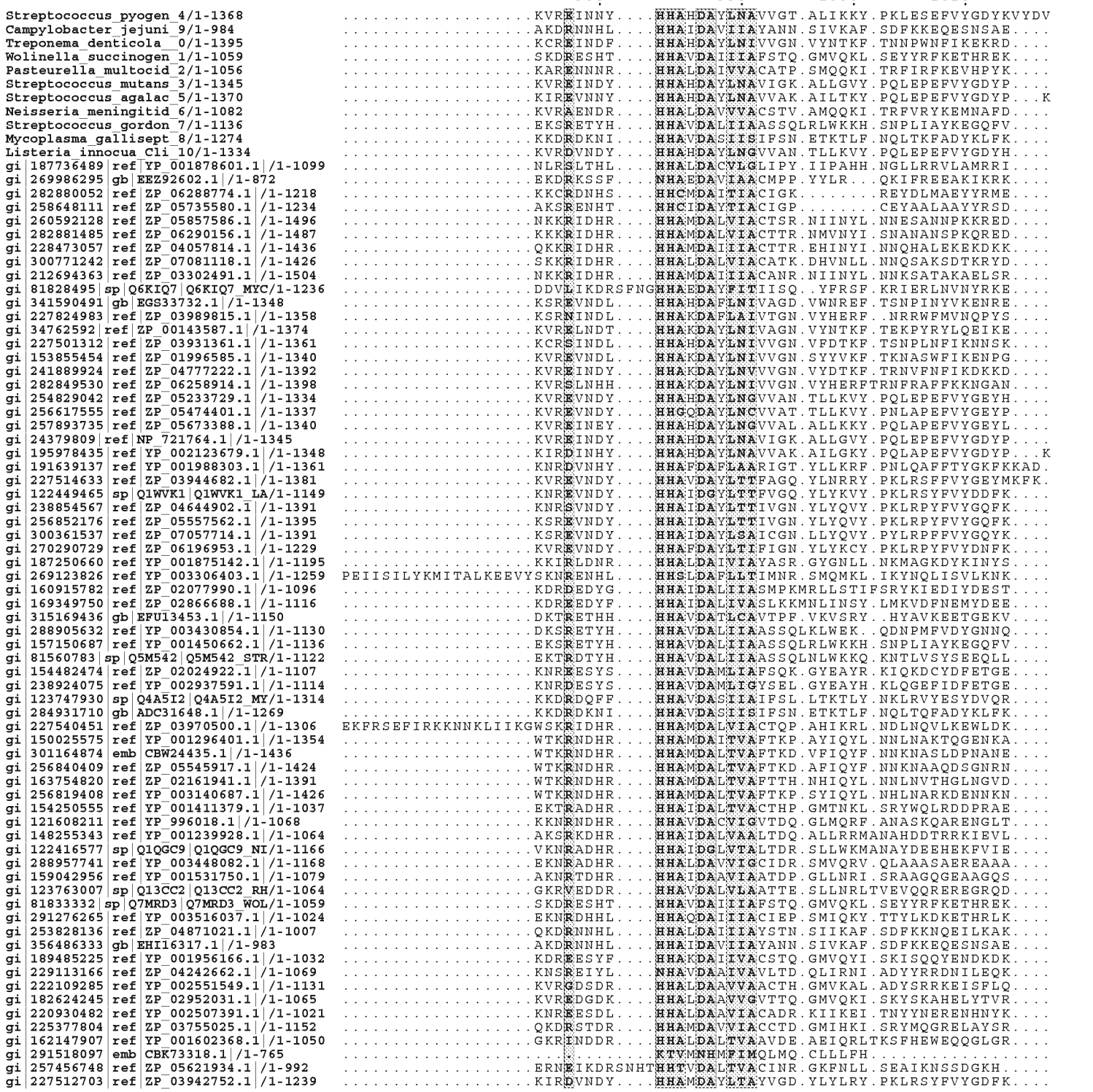

980
990
1000
1010
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AC.

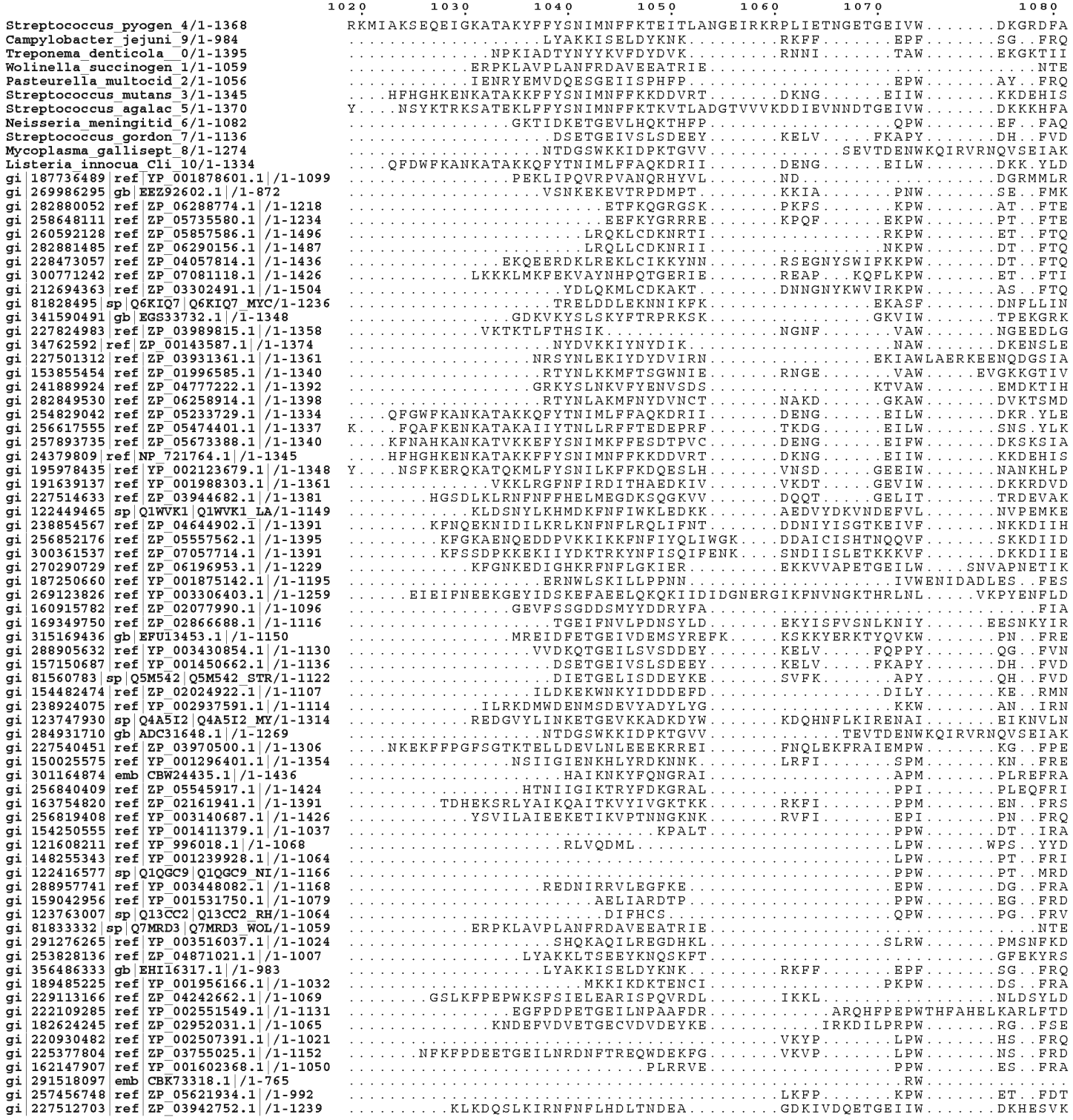

1020 1030 1040 1050 1060 1070 1080
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AD.

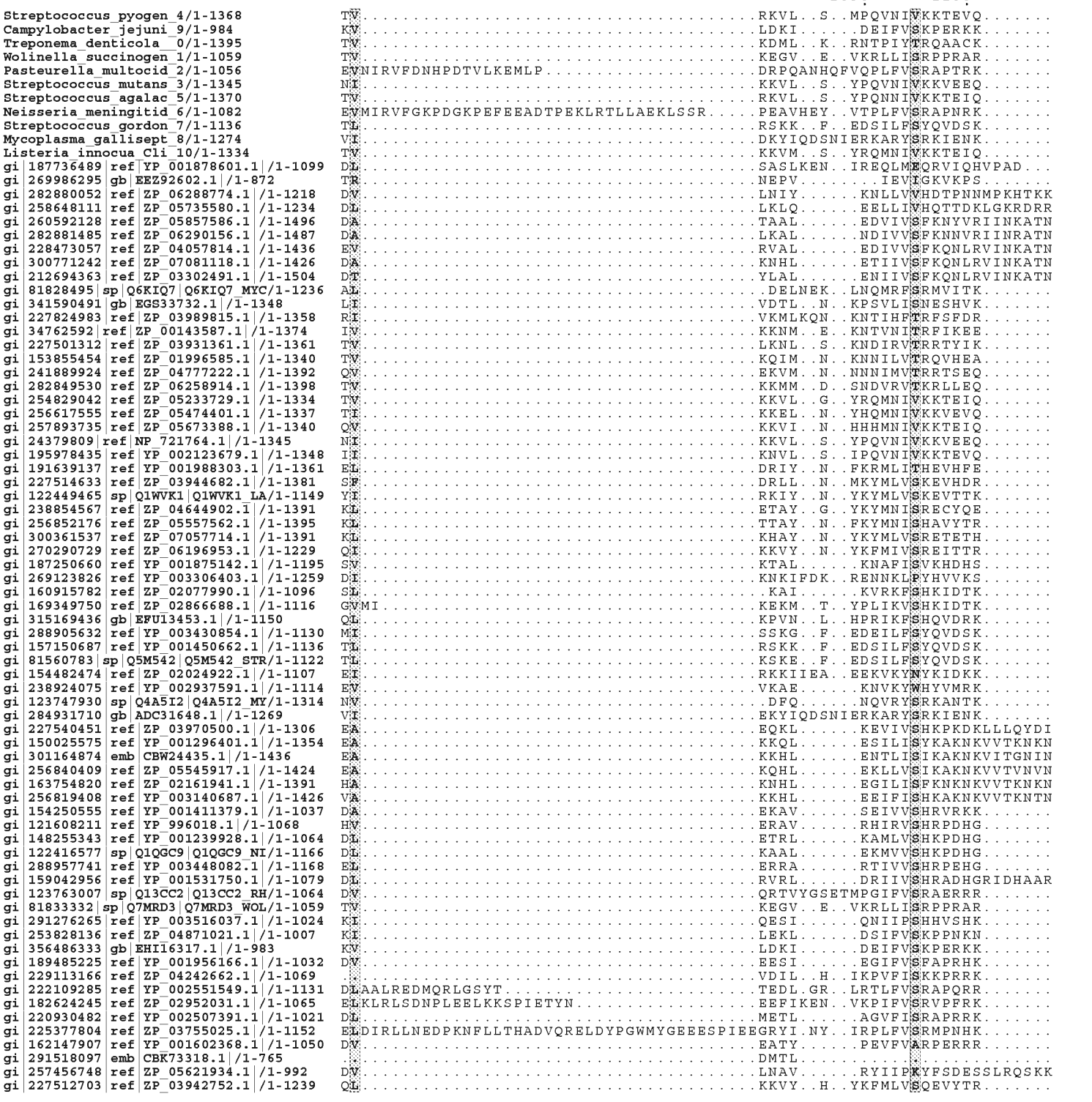

1090
1100
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AE.

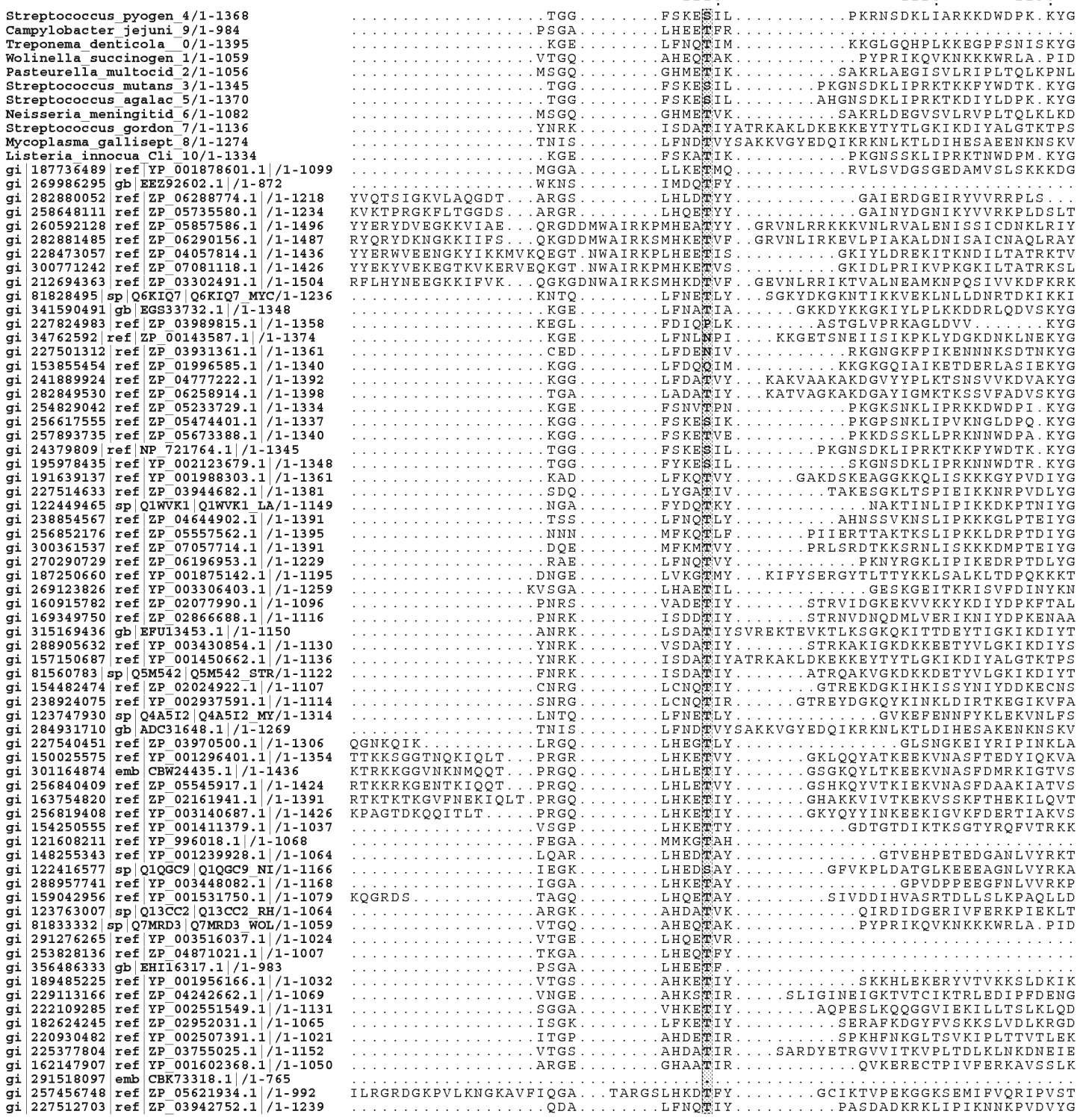

1110
1120
1130
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola__0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AF.

```
Streptococcus_pyogen_4/1-1368                ..GFDSPTVAYSVLVVAK
Campylobacter_jejuni_9/1-984                 
Treponema_denticola_0/1-1395                 ..GYNKVSAAYYTLIEYEE
Wolinella_succinogen_1/1-1059                EEKFESFKADRVASA
Pasteurella_multocid_2/1-1056                LENMVNKE
Streptococcus_mutans_3/1-1345                ..GFDSPIVAYSILVIAD
Streptococcus_agalac_5/1-1370                ..GFDSPIVAYSVLVVAD
Neisseria_meningitid_6/1-1082                LEKMVNRE
Streptococcus_gordon_7/1-1136                KTGFYKFLDLYKTDKSQFLMYQ
Mycoplasma_gallisept_8/1-1274                KKQFVYRKLVNVSLLNNDKLADLFAEKEDILMYRANPWVINL
Listeria_innocua_Cli_10/1-1334               ..GLDSPNMAYAVVIE
gi|187736489|ref|YP_001878601.1|/1-1099      KKEKNQVKASKLVGV
gi|269986295|gb|EEZ92602.1|/1-872            ..KYLLKP
gi|282880052|ref|ZP_06288774.1|/1-1218       ..SFTKPEELENI
gi|258648111|ref|ZP_05735580.1|/1-1234       EKDLDKI
gi|260592128|ref|ZP_05857586.1|/1-1496       INGLVEKGFNTKQILDKFKNTNDHFSMQKIDKVEIWQFTNEEEKLVATRKPLDTSFDTKKIASI
gi|282881485|ref|ZP_06290156.1|/1-1487       VNDLVDKHFNKKQLVAHFKSLNYKWNRQDVSKVEVWISSDNKTPMVAVRKPLDTSFDKKKIASI
gi|228473057|ref|ZP_04057814.1|/1-1436       DSSFDDKRIKKI
gi|300771242|ref|ZP_07081118.1|/1-1426       DTSFNLKTIGSI
gi|212694363|ref|ZP_03302491.1|/1-1504       LLELWNLGFDAKRIKKYFEDNRETWSDINLSKIEVYYFSKDTKDRFFATRKPLDTSFDRKKIENNI
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236      EEFFDEDKLKENELTKLHIFN
gi|341590491|gb|EGS33732.1|/1-1348           ..GYKAINGAFFFLVEHTK
gi|227824983|ref|ZP_03989815.1|/1-1358       ..GYDKSTAAYYLLVRFTL
gi|34762592|ref|ZP_00143587.1|/1-1374        ..YYTSLKAAYFIYVE
gi|227501312|ref|ZP_03931361.1|/1-1361       ..GYSGIKNAYFVLLEDE
gi|153855454|ref|ZP_01996585.1|/1-1340       ..GYNKAAGAYFMLVES
gi|241889924|ref|ZP_04777222.1|/1-1392       ..GYTSIGIAYYSIFEYTIVNKK
gi|282849530|ref|ZP_06258914.1|/1-1398       ..GMTKIKNAYSIIVQYTG
gi|254829042|ref|ZP_05233729.1|/1-1334       ..GFDGSKMAYAIIIE
gi|256617555|ref|ZP_05474401.1|/1-1337       ..GFDSPVVAYTVLFT
gi|257893735|ref|ZP_05673388.1|/1-1340       ..GLGSPNVAYTVAFT
gi|24379809|ref|NP_721764.1|/1-1345          ..GFDSPIVAYSILVIAD
gi|195978435|ref|YP_002123679.1|/1-1348      ..GFDSPTVAYSVLVIAK
gi|191639137|ref|YP_001988303.1|/1-1361      ..GYTQETGSYLSVVRL
gi|227514633|ref|ZP_03944682.1|/1-1381       ..AYTNGTSAFMTIIKFT
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149      ..GYKGKVSSYMMLVKIQK
gi|238854567|ref|ZP_04644902.1|/1-1391       ..GYSGNKDSFFVLVKIVK
gi|256852176|ref|ZP_05557562.1|/1-1395       ..GYSNNNDAYLALVKVD
gi|300361537|ref|ZP_07057714.1|/1-1391       ..GYTNNSDAYMVIARIN
gi|270290729|ref|ZP_06196953.1|/1-1229       ..GYSGNTDAYLAIVALED
gi|187250660|ref|YP_001875142.1|/1-1195      PKDFLETALLKFKGRESEMKNEKIKSAIE
gi|269123826|ref|YP_003306403.1|/1-1259      LGKIFDKDG
gi|160915782|ref|ZP_02077990.1|/1-1096       AEDILNNAYQEKYLMAL
gi|169349750|ref|ZP_02866688.1|/1-1116       NTLINDIINDNTDKYIMYHKDPQTFDKIKAIVL
gi|315169436|gb|EFU13453.1|/1-1150           VDGWEAFKKKQDKLLMKD
gi|288905632|ref|YP_003430854.1|/1-1130      QNGFDTFIKKYNKDKTQFLMYQ
gi|157150687|ref|YP_001450662.1|/1-1136      KTGFYKFLDLYKTDKSQFLMYQ
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122      QDGYDAFMKIYKKDKSKFLMYR
gi|154482474|ref|ZP_02024922.1|/1-1107       LKKMINSGKGSDLLMYN
gi|238924075|ref|YP_002937591.1|/1-1114      KLAFSKKDSDRERLLVYL
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314      ..RKDLRKFILEDLNEESEKNKK
gi|284931710|gb|ADC31648.1|/1-1269           KRQFVYRKLVNVSLLNNDKLADLFAEKEDILMYRANPWVINL
gi|227540451|ref|ZP_03970500.1|/1-1306       GKQFATEKTIEKI
gi|150025575|ref|YP_001296401.1|/1-1354      ..KKEYREALLKRLEENDNDPKKAFTGKNALNKTPIYISLKDNIIVPEKVKTVWLETDYTIRKDITPDLK
gi|301164874|emb|CBW24435.1|/1-1436          ..KLAYRDALLKRLYENDNDPKKAFAGKNSLDKQPIWLDKEQMRKVPEKVKIVTLEAIYTIRKEISPDLK
gi|256840409|ref|ZP_05545917.1|/1-1424       ..KRAYRNALLKRLETFGNDPKKAFTGKNALEKNPLYADKYQTIRVPEKVQTVEFETIYTIRKPVDPALN
gi|163754820|ref|ZP_02161941.1|/1-1391       ..KPAYREALLKRLASFKNDPKKAFSGKNALSKNPLYFDITSGKKVPDTVKLSYFENSYVIRKEVNSDNF
gi|256819408|ref|YP_003140687.1|/1-1426      ..NPVYREALLKRLQANDNDPKKAFAGKNALSKNPIYLDESKTKTLPEKVNLTYLEEDFSIRKDISPDNF
gi|154250555|ref|YP_001411379.1|/1-1037      IESLSKGELDEI
gi|121608211|ref|YP_996018.1|/1-1068         ..GI
gi|148255343|ref|YP_001239928.1|/1-1064      FVDISEKEIDRI
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166      IESLNENEVDRI
gi|288957741|ref|YP_003448082.1|/1-1168      IDGLSKDEINSV
gi|159042956|ref|YP_001531750.1|/1-1079      EPGRSGQV
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064      DKDLERIPVPEPYGKAA
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059      EEKFESFKADRVASA
gi|291276265|ref|YP_003516037.1|/1-1024      
gi|253828136|ref|ZP_04871021.1|/1-1007       
gi|356486333|gb|EHI16317.1|/1-983            
gi|189485225|ref|YP_001956166.1|/1-1032      LSDLENIPCN
gi|229113166|ref|ZP_04242662.1|/1-1069       DFPMYGKD
gi|222109285|ref|YP_002551549.1|/1-1131      FDKLLNPESNDHFVEPH
gi|182624245|ref|ZP_02952031.1|/1-1065       LDNFYNYE
gi|220930482|ref|YP_002507391.1|/1-1021      LETMVKNTKGGI
gi|225377804|ref|ZP_03755025.1|/1-1152       ..GYYDKD
gi|162147907|ref|YP_001602368.1|/1-1050      EADLERI
gi|291518097|emb|CBK73318.1|/1-765           
gi|257456748|ref|ZP_05621934.1|/1-992        LDEKVAEKIIDKRIRKT
gi|227512703|ref|ZP_03942752.1|/1-1239       ..GYSGNVDAYMAIVRIHG
```

FIGURE 11 AG.

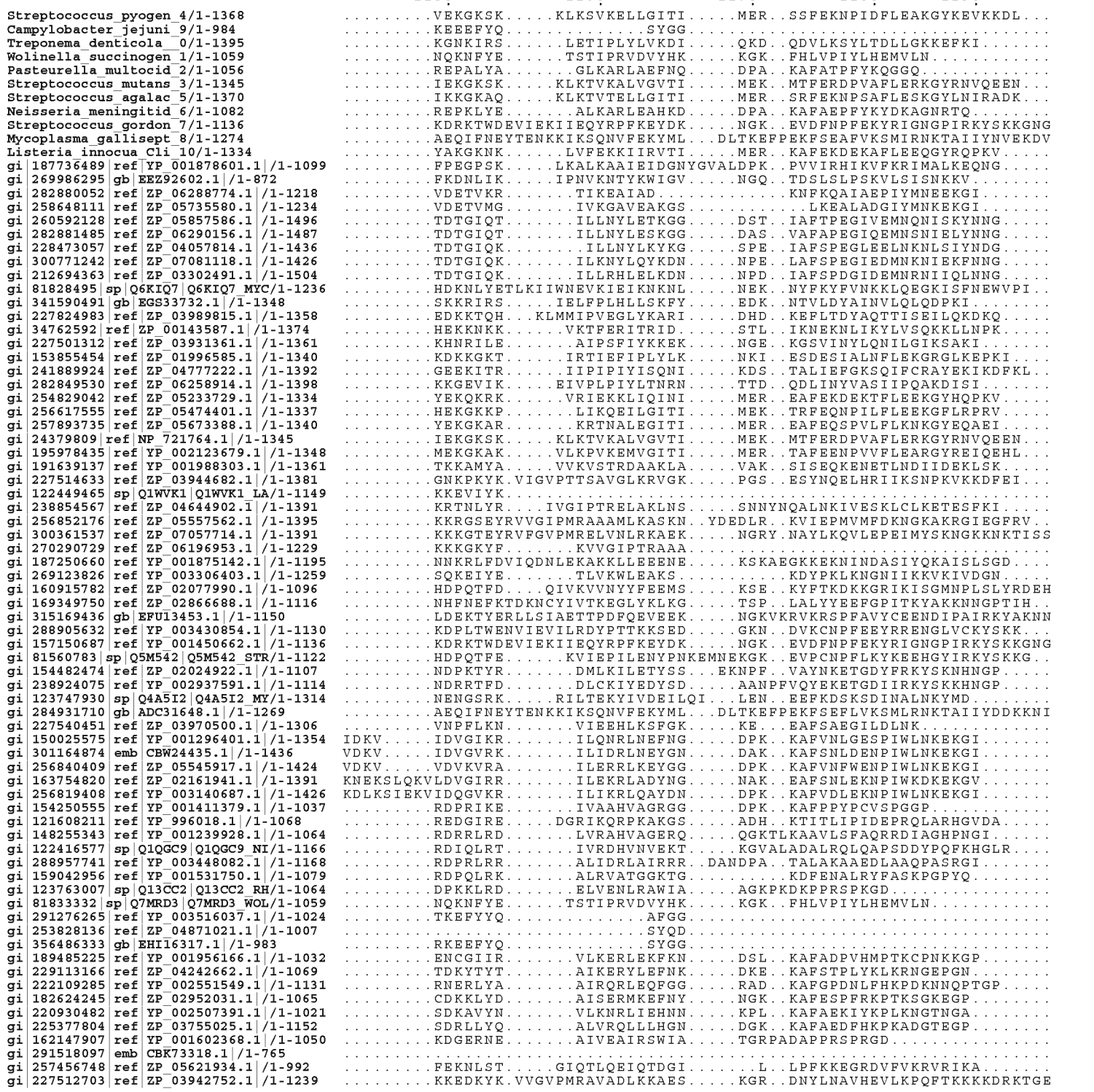

1150
1160
1170
1180
1190
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AH.

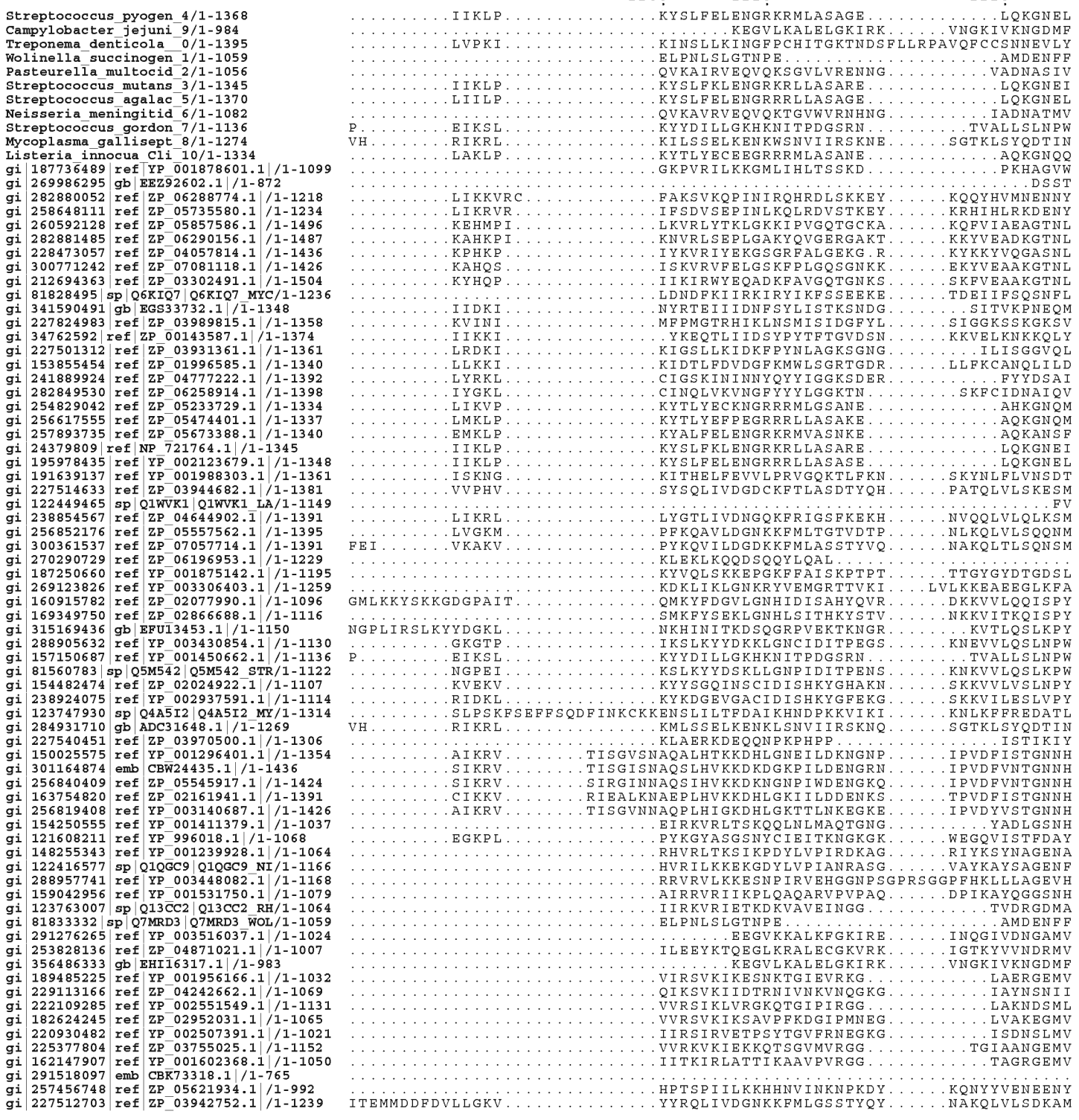

1200
1210
1220
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AI.

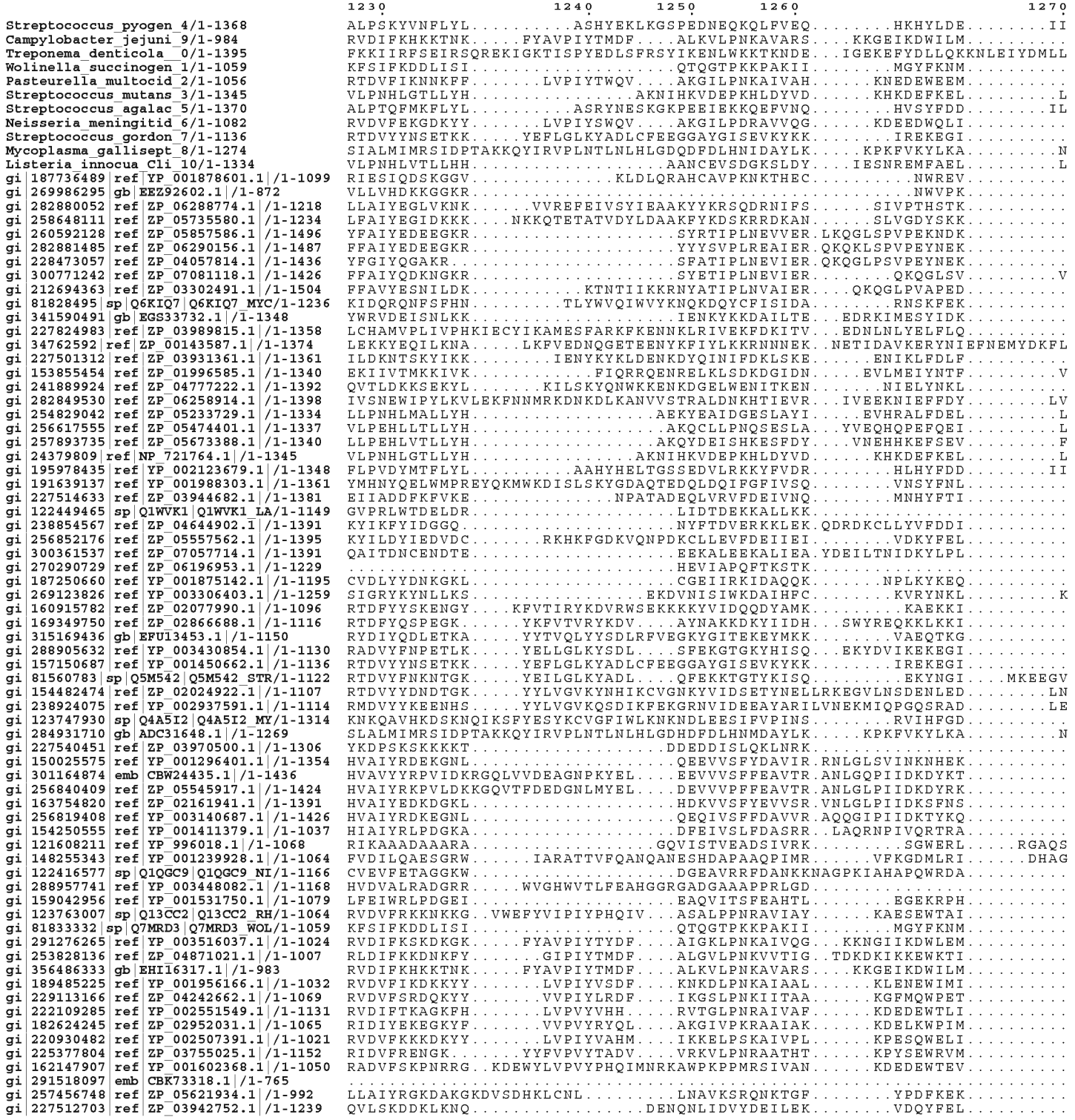

1230
1240
1250
1260
1270
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AJ.

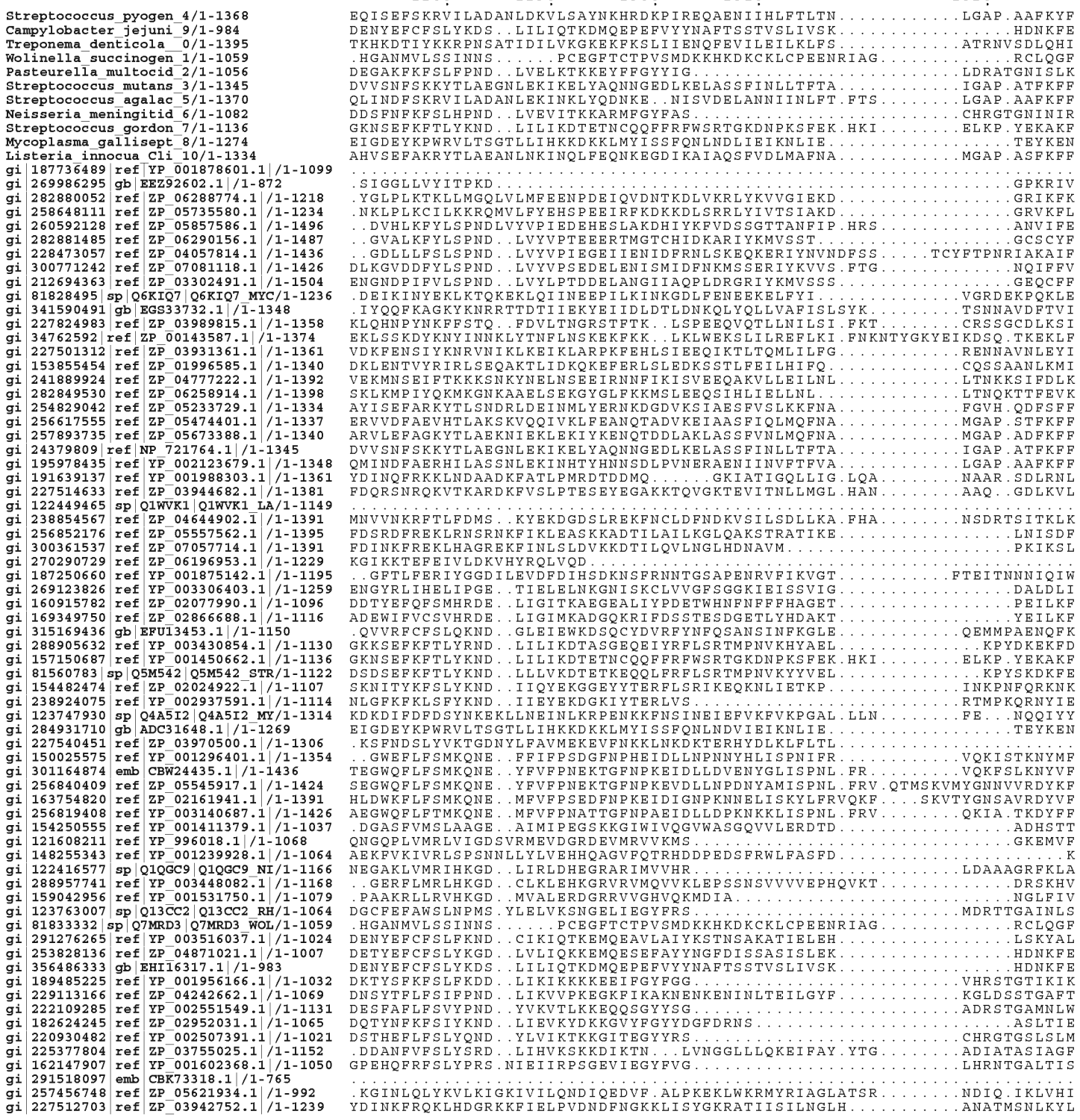

1280 1290 1300 1310 1320
Streptococcus_pyogen_4/1-1368 EQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTN LGAP.AAFKYF
Campylobacter_jejuni_9/1-984 DENYEFCFSLYKDS..LILIQTKDMQEPEFVYYNAFTSSTVSLIVSK HDNKFE
Treponema_denticola_0/1-1395 TKHKDTIYKKRPNSATIDILVKGKEKFKSLIIENQFEVILEILKLFS ATRNVSDLQHI
Wolinella_succinogen_1/1-1059 .HGANMVLSSINNS PCEGFTCTPVSMDKKHKDKCKLCPEENRIAG RCLQGF
Pasteurella_multocid_2/1-1056 DEGAKFKFSLFPND..LVELKTKKEYFFGYYIG LDRATGNISLK
Streptococcus_mutans_3/1-1345 DVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTA IGAP.ATFKFF
Streptococcus_agalac_5/1-1370 QLINDFSKRVILADANLEKINKLYQDNKE..NISVDELANNIINLFT.FTS LGAP.AAFKFF
Neisseria_meningitid_6/1-1082 DDSFNFKFSLHPND..LVEVITKKARMFGYFAS CHRGTGNININR
Streptococcus_gordon_7/1-1136 GKNSEFKFTLYKND..LILIKDTETNCQQFFRFWSRTGKDNPKSFEK.HKI ELKP.YEKAKF
Mycoplasma_gallisept_8/1-1274 EIGDEYKPWRVLTSGTLLIHKKDKKLMYISSFQNLNDLIEIKNLIE TEYKEN
Listeria_innocua_Cli_10/1-1334 AHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIAQSFVDLMAFNA MGAP.ASFKFF
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872 .SIGGLLVYITPKD GPKRIV
gi|282880052|ref|ZP_06288774.1|/1-1218 .YGLPLKTKLLMGQLVLMFEENPDEIQVDNTKDLVKRLYKVVGIEKD GRIKFK
gi|258648111|ref|ZP_05735580.1|/1-1234 .NKLPLKCILKKRQMVLFYEHSPEEIRFKDKKDLSRRLYIVTSIAKD GRVKFL
gi|260592128|ref|ZP_05857586.1|/1-1496 ..DVHLKFYLSPNDLVYVPIEDEHESLAKDHIYKFVDSSGTTANFIP.HRS ANVIFE
gi|282881485|ref|ZP_06290156.1|/1-1487 ..GVALKFYLSPND..LVYVPTEEERTMGTCHIDKARIYKMVSST GCSCYF
gi|228473057|ref|ZP_04057814.1|/1-1436 ..GDLLLFSLSPND..LVYVPIEGEIIENIDFRNLSKEQKERIYNVNDFSS TCYFTPNRIAKAIF
gi|300771242|ref|ZP_07081118.1|/1-1426 DLKGVDDFYLSPND..LVYVPSEDELENISMIDFNKMSSERIYKVVS.FTG NQIFFV
gi|212694363|ref|ZP_03302491.1|/1-1504 ENGNDPIFVLSPND..LVYLPTDDELANGIIAQPLDRGRIYKMVSSS GEQCFF
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236 .DEIKINYEKLKTQKEKLQIINEEPILKINKGDLFENEEKELFYI VGRDEKPQKLE
gi|341590491|gb|EGS33732.1|/1-1348 .IYQQFKAGKYKNRRTTDTIIEKYEIIDLDTLDNKQLYQLLVAFISLSYK TSNNAVDFTVI
gi|227824983|ref|ZP_03989815.1|/1-1358 KLQHNPYNKFFSTQ..FDVLTNGRSTFTK..LSPEEQVQTLLNILSI.FKT CRSSGCDLKSI
gi|34762592|ref|ZP_00143587.1|/1-1374 EKLSSKDYKNYINNKLYTNFLNSKEKFKK..LKLWEKSLILREFLKI.FNKNTYGKYEIKDSQ.TKEKLF
gi|227501312|ref|ZP_03931361.1|/1-1361 VDKFENSIYKNRVNIKLKEIKLARPKFEHLSIEEQIKTLTQMLILFG RENNAVNLEYI
gi|153855454|ref|ZP_01996585.1|/1-1340 DKLENTVYRIRLSEQAKTLIDKQKEFERLSLEDKSSTLFEILHIFQ CQSSAANLKMI
gi|241889924|ref|ZP_04777222.1|/1-1392 VEKMNSEIFTKKKSNKYNELNSEEIRNNFIKISVEEQAKVLLEILNL LTNKKSIFDLK
gi|282849530|ref|ZP_06258914.1|/1-1398 SKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNL LTNQKTTFEVK
gi|254829042|ref|ZP_05233729.1|/1-1334 AYISEFARKYTLSNDRLDEINMLYERNKDGDVKSIAESFVSLKKFNA FGVH.QDFSFF
gi|256617555|ref|ZP_05474401.1|/1-1337 ERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQLMQFNA MGAP.STFKFF
gi|257893735|ref|ZP_05673388.1|/1-1340 ARVLEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLASSFVNLMQFNA MGAP.ADFKFF
gi|24379809|ref|NP_721764.1|/1-1345 DVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELASSFINLLTFTA IGAP.ATFKFF
gi|195978435|ref|YP_002123679.1|/1-1348 QMINDFAERHILASSNLEKINHTYHNNSDLPVNERAENIINVFTFVA LGAP.AAFKFF
gi|191639137|ref|YP_001988303.1|/1-1361 YDINQFRKKLNDAADKFATLPMRDTDDMQ GKIATIGQLLIG.LQA NAAR.SDLRNL
gi|227514633|ref|ZP_03944682.1|/1-1381 FDQRSNRQKVTKARDKFVSLPTESEYEGAKKTQVGKTEVITNLLMGL.HAN AAQ..GDLKVL
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391 MNVVNKRFTLFDMS..KYEKDGDSLREKFNCLDFNDKVSILSDLLKA.FHA NSDRTSITKLK
gi|256852176|ref|ZP_05557562.1|/1-1395 FDSRDFREKLRNSRNKFIKLEASKKADTILAILKGLQAKSTRATIKE LNISDF
gi|300361537|ref|ZP_07057714.1|/1-1391 FDINKFREKLHAGREKFINLSLDVKKDTILQVLNGLHDNAVM PKIKSL
gi|270290729|ref|ZP_06196953.1|/1-1229 KGIKKTEFEIVLDKVHYRQLVQD
gi|187250660|ref|YP_001875142.1|/1-1195 ..GFTLFERIYGGDILEVDFDIHSDKNSFRNNTGSAPENRVFIKVGT FTEITNNNIQIW
gi|269123826|ref|YP_003306403.1|/1-1259 ENGYRLIHELIPGE..TIELELNKGNISKCLVVGFSGGKIEISSVIG DALDLI
gi|160915782|ref|ZP_02077990.1|/1-1096 DDTYEFQFSMHRDE..LIGITKAEGEALIYPDETWHNFNFFFHAGET PEILKF
gi|169349750|ref|ZP_02866688.1|/1-1116 ADEWIFVCSVHRDE..LIGIMKADGQKRIFDSSTESDGETLYHDAKT YEILKF
gi|315169436|gb|EFU13453.1|/1-1150 .QVVRFCFSLQKND..GLEIEWKDSQCYDVRFYNFQSANSINFKGLE QEMMPAENQFK
gi|288905632|ref|YP_003430854.1|/1-1130 GKKSEFKFTLYRND..LILIKDTASGEQEIYRFLSRTMPNVKHYAEL KPYDKEKFD
gi|157150687|ref|YP_001450662.1|/1-1136 GKNSEFKFTLYKND..LILIKDTETNCQQFFRFWSRTGKDNPKSFEK.HKI ELKP.YEKAKF
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122 DSDSEFKFTLYKND..LLLVKDTETKEQQLFRFLSRTMPNVKYYVEL KPYSKDKFE
gi|154482474|ref|ZP_02024922.1|/1-1107 SKNITYKFSLYKND..IIQYEKGGEYYTERFLSRIKEQKNLIETKP INKPNFQRKNK
gi|238924075|ref|YP_002937591.1|/1-1114 NLGFKFKLSFYKND..IIEYEKDGKIYTERLVS RTMPKQRNYIE
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314 KDKDIFDFDSYNKEKLLNEINLKRPENKKFNSINEIEFVKFVKPGAL.LLN FE...NQQIYY
gi|284931710|gb|ADC31648.1|/1-1269 EIGDEYKPWRVLTSGTLLIHKKDKKLMYISSFQNLNDVIEIKNLIE TEYKEN
gi|227540451|ref|ZP_03970500.1|/1-1306 .KSFNDSLYVKTGDNYLFAVMEKEVFNKKLNKDKTERHYDLKLFLTL
gi|150025575|ref|YP_001296401.1|/1-1354 ..GWEFLFSMKQNE..FFIFPSDGFNPHEIDLLNPNNYHLISPNIFR VQKISTKNYMF
gi|301164874|emb|CBW24435.1|/1-1436 TEGWQFLFSMKQNE..YFVFPNEKTGFNPKEIDLLDVENYGLISPNL.FR VQKFSLKNYVF
gi|256840409|ref|ZP_05545917.1|/1-1424 SEGWQFLFSMKQNE..YFVFPNEKTGFNPKEVDLLNPDNYAMISPNL.FRV.QTMSKVMYGNNVVRDYKF
gi|163754820|ref|ZP_02161941.1|/1-1391 HLDWKFLFSMKQNE..MFVFPSEDFNPKEIDIGNPKNNELISKYLFRVQKF SKVTYGNSAVRDYVF
gi|256819408|ref|YP_003140687.1|/1-1426 AEGWQFLFTMKQNE..MFVFPNATTGFNPAEIDLLDPKNKKLISPNL.FRV QKIA.TKDYFF
gi|154250555|ref|YP_001411379.1|/1-1037 .DGASFVMSLAAGE..AIMIPEGSKKGIWIVQGVWASGQVVLERDTD ADHSTT
gi|121608211|ref|YP_996018.1|/1-1068 QNGQPLVMRLVIGDSVRMEVDGRDEVMRVVKMS GKEMVF
gi|148255343|ref|YP_001239928.1|/1-1064 AEKFVKIVRLSPSNNLLYLVEHHQAGVFQTRHDDPEDSFRWLFASFD K
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166 NEGAKLVMRIHKGD..LIRLDHEGRARIMVVHR LDAAAGRFKLA
gi|288957741|ref|YP_003448082.1|/1-1168 ..GERFLMRLHKGD..CLKLEHKGRVRVMQVVKLEPSSNSVVVVEPHQVKT DRSKHV
gi|159042956|ref|YP_001531750.1|/1-1079 .PAAKRLLRVHKGD..MVALERDGRRVVGHVQKMDIA NGLFIV
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064 DGCFEFAWSLNPMS.YLELVKSNGELIEGYFRS MDRTTGAINLS
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059 .HGANMVLSSINNS PCEGFTCTPVSMDKKHKDKCKLCPEENRIAG RCLQGF
gi|291276265|ref|YP_003516037.1|/1-1024 DENYEFCFSLFKND..CIKIQTKEMQEAVLAIYKSTNSAKATIELEH LSKYAL
gi|253828136|ref|ZP_04871021.1|/1-1007 DETYEFCFSLYKGD..LVLIQKKEMQESEFAYYNGFDISSASISLEK HDNKFE
gi|356486333|gb|EHI16317.1|/1-983 DENYEFCFSLYKDS..LILIQTKDMQEPEFVYYNAFTSSTVSLIVSK HDNKFE
gi|189485225|ref|YP_001956166.1|/1-1032 DKTYSFKFSLFKDD..LIKIKKKKEEIFGYFGG VHRSTGTIKIK
gi|229113166|ref|ZP_04242662.1|/1-1069 DNSYTFLFSIFPND..LIKVVPKEGKFIKAKNENKENINLTEILGYF KGLDSSTGAFT
gi|222109285|ref|YP_002551549.1|/1-1131 DESFAFLFSVYPND..YVKVTLKKEQQSGYYSG ADRSTGAMNLW
gi|182624245|ref|ZP_02952031.1|/1-1065 DQTYNFKFSIYKND..LIEVKYDKKGVYFGYYDGFDRNS ASLTIE
gi|220930482|ref|YP_002507391.1|/1-1021 DSTHEFLFSLYQND..YLVIKTKKGITEGYYRS CHRGTGSLSLM
gi|225377804|ref|ZP_03755025.1|/1-1152 .DDANFVFSLYSRD..LIHVKSKKDIKTN...LVNGGLLLQKEIFAY.YTG ADIATASIAGF
gi|162147907|ref|YP_001602368.1|/1-1050 GPEHQFRFSLYPRS.NIEIIRPSGEVIEGYFVG LHRNTGALTIS
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992 .KGINLQLYKVLKIGKIVILQNDIQEDVF.ALPKEKLWKRMYRIAGLATSR NDIQ.IKLVHI
gi|227512703|ref|ZP_03942752.1|/1-1239 YDINKFRQKLHDGRKKFIELPVDNDFNGKKLISYGKRATIISILNGLH ANATMSNLKYL

FIGURE 11 AK.

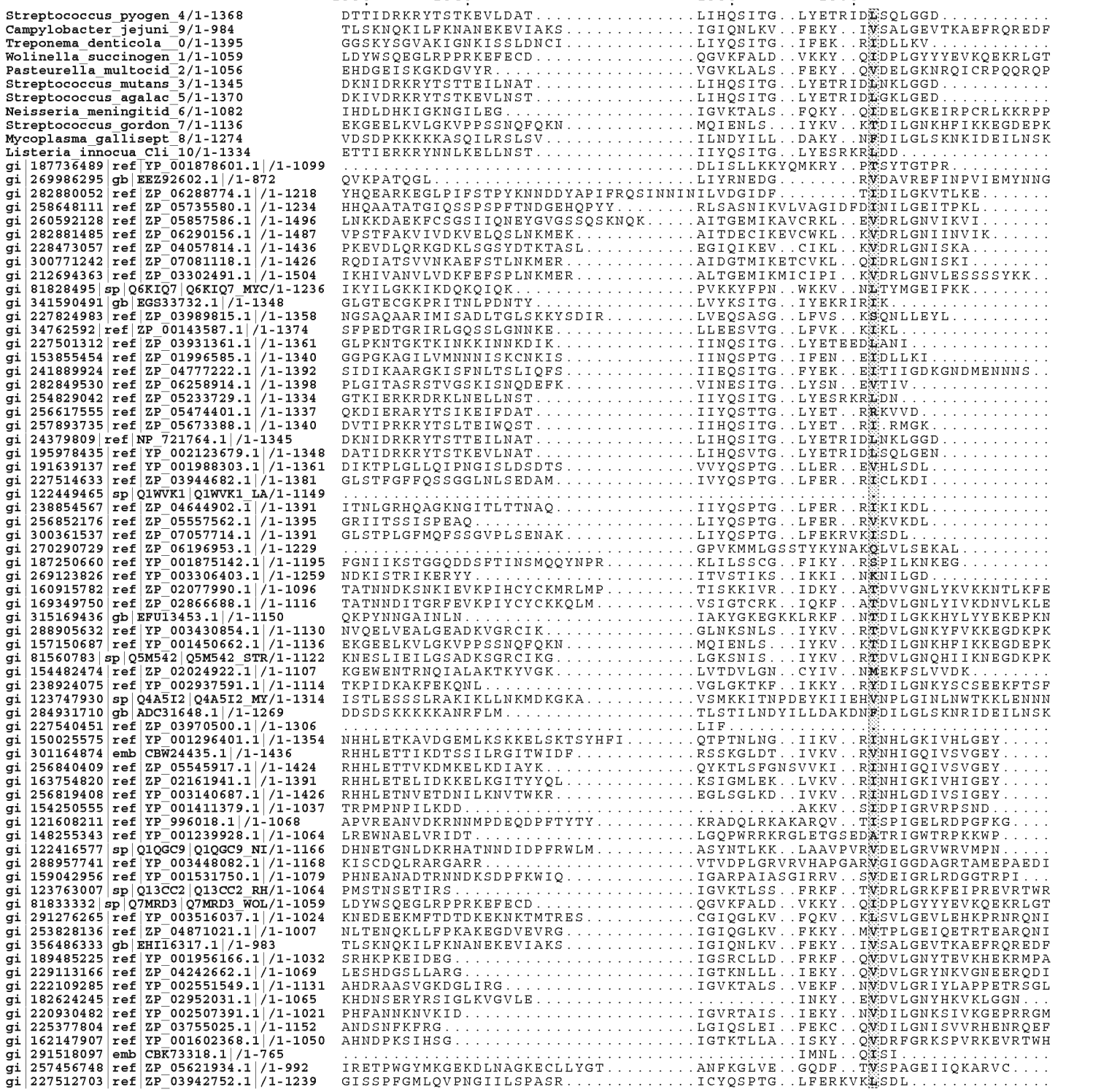

1330
1340
1350
1360
Streptococcus_pyogen_4/1-1368
Campylobacter_jejuni_9/1-984
Treponema_denticola_0/1-1395
Wolinella_succinogen_1/1-1059
Pasteurella_multocid_2/1-1056
Streptococcus_mutans_3/1-1345
Streptococcus_agalac_5/1-1370
Neisseria_meningitid_6/1-1082
Streptococcus_gordon_7/1-1136
Mycoplasma_gallisept_8/1-1274
Listeria_innocua_Cli_10/1-1334
gi|187736489|ref|YP_001878601.1|/1-1099
gi|269986295|gb|EEZ92602.1|/1-872
gi|282880052|ref|ZP_06288774.1|/1-1218
gi|258648111|ref|ZP_05735580.1|/1-1234
gi|260592128|ref|ZP_05857586.1|/1-1496
gi|282881485|ref|ZP_06290156.1|/1-1487
gi|228473057|ref|ZP_04057814.1|/1-1436
gi|300771242|ref|ZP_07081118.1|/1-1426
gi|212694363|ref|ZP_03302491.1|/1-1504
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
gi|341590491|gb|EGS33732.1|/1-1348
gi|227824983|ref|ZP_03989815.1|/1-1358
gi|34762592|ref|ZP_00143587.1|/1-1374
gi|227501312|ref|ZP_03931361.1|/1-1361
gi|153855454|ref|ZP_01996585.1|/1-1340
gi|241889924|ref|ZP_04777222.1|/1-1392
gi|282849530|ref|ZP_06258914.1|/1-1398
gi|254829042|ref|ZP_05233729.1|/1-1334
gi|256617555|ref|ZP_05474401.1|/1-1337
gi|257893735|ref|ZP_05673388.1|/1-1340
gi|24379809|ref|NP_721764.1|/1-1345
gi|195978435|ref|YP_002123679.1|/1-1348
gi|191639137|ref|YP_001988303.1|/1-1361
gi|227514633|ref|ZP_03944682.1|/1-1381
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
gi|238854567|ref|ZP_04644902.1|/1-1391
gi|256852176|ref|ZP_05557562.1|/1-1395
gi|300361537|ref|ZP_07057714.1|/1-1391
gi|270290729|ref|ZP_06196953.1|/1-1229
gi|187250660|ref|YP_001875142.1|/1-1195
gi|269123826|ref|YP_003306403.1|/1-1259
gi|160915782|ref|ZP_02077990.1|/1-1096
gi|169349750|ref|ZP_02866688.1|/1-1116
gi|315169436|gb|EFU13453.1|/1-1150
gi|288905632|ref|YP_003430854.1|/1-1130
gi|157150687|ref|YP_001450662.1|/1-1136
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
gi|154482474|ref|ZP_02024922.1|/1-1107
gi|238924075|ref|YP_002937591.1|/1-1114
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
gi|284931710|gb|ADC31648.1|/1-1269
gi|227540451|ref|ZP_03970500.1|/1-1306
gi|150025575|ref|YP_001296401.1|/1-1354
gi|301164874|emb|CBW24435.1|/1-1436
gi|256840409|ref|ZP_05545917.1|/1-1424
gi|163754820|ref|ZP_02161941.1|/1-1391
gi|256819408|ref|YP_003140687.1|/1-1426
gi|154250555|ref|YP_001411379.1|/1-1037
gi|121608211|ref|YP_996018.1|/1-1068
gi|148255343|ref|YP_001239928.1|/1-1064
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
gi|288957741|ref|YP_003448082.1|/1-1168
gi|159042956|ref|YP_001531750.1|/1-1079
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
gi|291276265|ref|YP_003516037.1|/1-1024
gi|253828136|ref|ZP_04871021.1|/1-1007
gi|356486333|gb|EHI16317.1|/1-983
gi|189485225|ref|YP_001956166.1|/1-1032
gi|229113166|ref|ZP_04242662.1|/1-1069
gi|222109285|ref|YP_002551549.1|/1-1131
gi|182624245|ref|ZP_02952031.1|/1-1065
gi|220930482|ref|YP_002507391.1|/1-1021
gi|225377804|ref|ZP_03755025.1|/1-1152
gi|162147907|ref|YP_001602368.1|/1-1050
gi|291518097|emb|CBK73318.1|/1-765
gi|257456748|ref|ZP_05621934.1|/1-992
gi|227512703|ref|ZP_03942752.1|/1-1239

FIGURE 11 AL.

```
Streptococcus_pyogen_4/1-1368            .............
Campylobacter_jejuni_9/1-984             KK..........
Treponema_denticola__0/1-1395            .............
Wolinella_succinogen_1/1-1059            IPQMRSAKKLVKK
Pasteurella_multocid_2/1-1056            VR..........
Streptococcus_mutans_3/1-1345            .............
Streptococcus_agalac_5/1-1370            .............
Neisseria_meningitid_6/1-1082            VR..........
Streptococcus_gordon_7/1-1136            LKFKK.......
Mycoplasma_gallisept_8/1-1274            LDLDKIAK....
Listeria_innocua_Cli_10/1-1334           .............
gi|187736489|ref|YP_001878601.1|/1-1099  .............
gi|269986295|gb|EEZ92602.1|/1-872        .............
gi|282880052|ref|ZP_06288774.1|/1-1218   .............
gi|258648111|ref|ZP_05735580.1|/1-1234   .............
gi|260592128|ref|ZP_05857586.1|/1-1496   .............
gi|282881485|ref|ZP_06290156.1|/1-1487   .............
gi|228473057|ref|ZP_04057814.1|/1-1436   .............
gi|300771242|ref|ZP_07081118.1|/1-1426   .............
gi|212694363|ref|ZP_03302491.1|/1-1504   .............
gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236  .............
gi|341590491|gb|EGS33732.1|/1-1348       .............
gi|227824983|ref|ZP_03989815.1|/1-1358   .............
gi|34762592|ref|ZP_00143587.1|/1-1374    .............
gi|227501312|ref|ZP_03931361.1|/1-1361   .............
gi|153855454|ref|ZP_01996585.1|/1-1340   .............
gi|241889924|ref|ZP_04777222.1|/1-1392   .............
gi|282849530|ref|ZP_06258914.1|/1-1398   .............
gi|254829042|ref|ZP_05233729.1|/1-1334   .............
gi|256617555|ref|ZP_05474401.1|/1-1337   .............
gi|257893735|ref|ZP_05673388.1|/1-1340   .............
gi|24379809|ref|NP_721764.1|/1-1345      .............
gi|195978435|ref|YP_002123679.1|/1-1348  .............
gi|191639137|ref|YP_001988303.1|/1-1361  .............
gi|227514633|ref|ZP_03944682.1|/1-1381   .............
gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149  .............
gi|238854567|ref|ZP_04644902.1|/1-1391   .............
gi|256852176|ref|ZP_05557562.1|/1-1395   .............
gi|300361537|ref|ZP_07057714.1|/1-1391   .............
gi|270290729|ref|ZP_06196953.1|/1-1229   .............
gi|187250660|ref|YP_001875142.1|/1-1195  .............
gi|269123826|ref|YP_003306403.1|/1-1259  .............
gi|160915782|ref|ZP_02077990.1|/1-1096   FD..........
gi|169349750|ref|ZP_02866688.1|/1-1116   FK..........
gi|315169436|gb|EFU13453.1|/1-1150       IIK.........
gi|288905632|ref|YP_003430854.1|/1-1130  LDFKNNKK....
gi|157150687|ref|YP_001450662.1|/1-1136  LKFKK.......
gi|81560783|sp|Q5M542|Q5M542_STR/1-1122  LDF.........
gi|154482474|ref|ZP_02024922.1|/1-1107   ............
gi|238924075|ref|YP_002937591.1|/1-1114  C...........
gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314  ............
gi|284931710|gb|ADC31648.1|/1-1269       LGLDKIVK....
gi|227540451|ref|ZP_03970500.1|/1-1306   .............
gi|150025575|ref|YP_001296401.1|/1-1354  .............
gi|301164874|emb|CBW24435.1|/1-1436      .............
gi|256840409|ref|ZP_05545917.1|/1-1424   .............
gi|163754820|ref|ZP_02161941.1|/1-1391   .............
gi|256819408|ref|YP_003140687.1|/1-1426  .............
gi|154250555|ref|YP_001411379.1|/1-1037  .............
gi|121608211|ref|YP_996018.1|/1-1068     .............
gi|148255343|ref|YP_001239928.1|/1-1064  .............
gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166  .............
gi|288957741|ref|YP_003448082.1|/1-1168  S............
gi|159042956|ref|YP_001531750.1|/1-1079  .............
gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064  GEACT........
gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059  IPQMRSAKKLVKK
gi|291276265|ref|YP_003516037.1|/1-1024  ALKTTPKHV....
gi|253828136|ref|ZP_04871021.1|/1-1007   QPRKKHGLR....
gi|356486333|gb|EHI16317.1|/1-983        KK..........
gi|189485225|ref|YP_001956166.1|/1-1032  YITRKDKKH....
gi|229113166|ref|ZP_04242662.1|/1-1069   KIQQN........
gi|222109285|ref|YP_002551549.1|/1-1131  A...........
gi|182624245|ref|ZP_02952031.1|/1-1065   .............
gi|220930482|ref|YP_002507391.1|/1-1021  EKYNSFKSN....
gi|225377804|ref|ZP_03755025.1|/1-1152   H...........
gi|162147907|ref|YP_001602368.1|/1-1050  GEACISPTPPG..
gi|291518097|emb|CBK73318.1|/1-765       .............
gi|257456748|ref|ZP_05621934.1|/1-992    .............
gi|227512703|ref|ZP_03942752.1|/1-1239   .............
```

FIGURE 11 AM.

FIGURE 12 A.

>Streptococcus_pyogenes
MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAK
VDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLF
IQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDN
LLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQE
EFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFA
WMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLF
KTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKV
MKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGIL
QTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQE
LDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIK
RQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVY
GDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQT
GGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKD
LIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAN
LDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

>Listeria_innocua
MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKIAGDSEKKQIKKNFWGVRLFDEGQTAADRRMARTARRRIERRRNRISYLQGIFAEEM
SKTDANFFCRLSDSFYVDNEKRNSRHPFFATIEEEVEYHKNYPTIYHLREELVNSSEKADLRLVYLALAHIIKYRGNFLIEGALDTQNTSVD
GIYKQFIQTYNQVFASGIEDGSLKKLEDNKDVAKILVEKVTRKEKLERILKLYPGEKSAGMFAQFISLIVGSKGNFQKPFDLIEKSDIECAKD
SYEEDLESLLALIGDEYAELFVAAKNAYSAVVLSSIITVAETETNAKLSASMIERFDTHEEDLGELKAFIKLHLPKHYEEIFSNTEKHGYAGYI
DGKTKQADFYKYMKMTLENIEGADYFIAKIEKENFLRKQRTFDNGAIPHQLHLEELEAILHQQAKYYPFLKENYDKIKSLVTFRIPYFVGPLA
NGQSEFAWLTRKADGEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYLVYNELTKVRYINDQGKTSYFSGQEKE
QIFNDLFKQKRKVKKKDLELFLRNMSHVESPTIEGLEDSFNSSYSTYHDLLKVGIKQEILDNPVNTEMLENIVKILTVFEDKRMIKEQLQQFS
DVLDGVVLKKLERRHYTGWGRLSAKLLMGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEKEQVTTADKDIQSIVADLAGSPA
IKKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTGKGKNNSRPRYKSLEKAIKEFGSQILKEHPTDNQELRNNRLYLYYLQNGKDMYT
GQDLDIHNLSNYDIDHIVPQSFITDNSIDNLVLTSSAGNREKGDDVPPLEIVRKRKVFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKAR
FIHRQLVETRQITKNVANILHQRFNYEKDDHGNTMKQVRIVTLKSALVSQFRKQFQLYKVRDVNDYHHAHDAYLNGVVANTLLKVYPQLE
PEFVYGDYHQFDWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKKYLDTVKKVMSYRQMNIVKKTEIQKGEFSKATIKPKGNSSK
LIPRKTNWDPMKYGGLDSPNMAYAVVIEYAKGKNKLVFEKKIIRVTIMERKAFEKDEKAFLEEQGYRQPKVLAKLPKYTLYECEEGRRRM
LASANEAQKGNQQVLPNHLVTLLHHAANCEVSDGKSLDYIESNREMFAELLAHVSEFAKRYTLAEANLNKINQLFEQNKEGDIKAIAQSFV
DLMAFNAMGAPASFKFFETTIERKRYNNLKELLNSTIIYQSITGLYESRKRLDD

>Streptococcus_mutans
MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRRYTRRRNRILYLQEIFSEEM
GKVDDSFFHRLEDSFLVTEDKRGERHPIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQ
RLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEE
ELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKDGYAGYIDG
KTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLARG
KSDFAWLSRKSADKITPWNFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFD
GVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYS
DLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQVIGETDNLNQVVSDIAGSPAI
KKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYT
GEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGF
IKRQLVETRQITKHVARILDERFNTETDENNKKIRQVKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFV
YGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQTGGFSKESILPKGNSDKLIPR
KTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLENGRK
RLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELAS
SFINLLTFTAIGAPATFKFFDKNIDRKRYTSTTEILNATLIHQSITGLYETRIDLNKLGGD

>Streptococcus thermophilus LMD-9 Csn1-A
MTKPYSIGLDIGTNSVGWAVTTDNYKVPSKKMKVLGNTSKKYIKKNLLGVLLFDSGITAEGRRLKRTARRRYTRRRNRILYLQEIFSTEMA
TLDDAFFQRLDDSFLVPDDKRDSKYPIFGNLVEEKAYHDEFPTIYHLRKYLADSTKKADLRLVYLALAHMIKYRGHFLIEGEFNSKNNDIQK
NFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESYDEDLE
TLLGYIGDDYSDVFLKAKKLYDAILLSGFLTVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQ
EDFYVYLKKLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFLAKNKERIEKILTFRIPYYVGPLARGNSDF
AWSIRKRNEKITPWNFEDVIDKESSAEAFINRMTSFDLYLPEEKVLPKHSLLYETFNVYNELTKVRFIAESMRDYQFLDSKQKKDIVRLYFK
DKRKVTDKDIIEYLHAIYGYDGIELKGIEKQFNSSLSTYHDLLNIINDKEFLDDSSNEAIIEEIIHTLTIFEDREMIKQRLSKFENIFDKSVLKKLS
RRHYTGWGKLSAKLINGIRDEKSGNTILDYLIDDGISNRNFMQLIHDDALSFKKKIQKAQIIGDEDKGNIKEVVKSLPGSPAIKKGILQSIKIVD
ELVKVMGGRKPESIVVEMARENQYTNQGKSNSQQRLKRLEKSLKELGSKILKENIPAKLSKIDNNALQNDRLYLYYLQNGKDMYTGDDLD
IDRLSNYDIDHIIPQAFLKDNSIDNKVLVSSASNRGKSDDVPSLEVVKKRKTFWYQLLKSKLISQRKFDNLTKAERGGLSPEDKAGFIQRQL
VETRQITKHVARLLDEKFNNKKDENNRAVRTVKIITLKSTLVSQFRKDFELYKVREINDFHHAHDAYLNAVVASALLKKYPKLEPEFVYGDY
PKYNSFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETGESVWNKESDLATVRRVLSYPQVNVVKKVEEQNHGLDRGKPK
GLFNANLSSKPKPNSNENLVGAKEYLDPKKYGGYAGISNSFTVLVKGTIEKGAKKKITNVLEFQGISILDRINYRKDKLNFLLEKGYKDIELI
ELPKYSLFELSDGSRRMLASILSTNNKRGEIHKGNQIFLSQKFVKLLYHAKRISNTINENHRKYVENHKKEFEELFYYILEFNENYVGAKKN
GKLLNSAFQSWQNHSIDELCSSFIGPTGSERKGLFELTSRGSAADFEFLGVKIPRYRDYTPSSLLKDATLIHQSVTGLYETRIDLAKLGEG

FIGURE 12 B.

>Streptococcus thermophilus LMD-9 Csn1-B
MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRV
KGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINV
FPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTA
QEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMD
RETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIF
GKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAI
KIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTG
KTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISK
FDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVS
YSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQD
GYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDI
TPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKE
QQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF

>Neisseria_meningitidis
MAAFKPNPINYILGLDIGIASVGWAMVEIDEDENPICLIDLGVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVL
QAADFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVADNAHALQTGDFRTPAELAL
NKFEKESGHIRNQRGDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTY
TAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKE
GLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKK
NTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEP
KSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKA
RVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHAL
DAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTPEKLRTLLAEKL
SSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAF
AEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWQLI
DDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNINIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR

>Pasteurella_multocida
MQTTNLSYILGLDLGIASVGWAVVEINENEDPIGLIDVGVRIFERAEVPKTGESLALSRRLARSTRRLIRRRAHRLLLAKRFLKREGILSTIDL
EKGLPNQAWELRVAGLERRLSAIEWGAVLLHLIKHRGYLSKRKNESQTNNKELGALLSGVAQNHQLLQSDDYRTPAELALKKFAKEEGHI
RNQRGAYTHTFNRLDLLAELNLLFAQQHQFGNPHCKEHIQQYMTELLMWQKPALSGEAILKMLGKCTHEKNEFKAAKHTYSAERFVWLT
KLNNLRILEDGAERALNEEERQLLINHPYEKSKLTYAQVRKLLGLSEQAIFKHLRYSKENAESATFMELKAWHAIRKALENQGLKDTWQDL
AKKPDLLDEIGTAFSLYKTDEDIQQYLTNKVPNSVINALLVSLNFDKFIELSLKSLRKILPLMEQGKRYDQACREIYGHHYGEANQKTSQLLP
AIPAQEIRNPVVLRTLSQARKVINAIIRQYGSPARVHIETGRELGKSFKERREIQKQQEDNRTKRESAVQKFKELFSDFSSEPKSKDILKFRL
YEQQHGKCLYSGKEINIHRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLASENQNKGNQTPYEWLQGKINSERWKNFVALVLGSQCSAA
KKQRLLTQVIDDNKFIDRNLNDTRYIARFLSNYIQENLLLVGKNKKNVFTPNGQITALLRSRWGLIKARENNNRHHALDAIVVACATPSMQQ
KITRFIRFKEVHPYKIENRYEMVDQESGEIISPHFPEPWAYFRQEVNIRVFDNHPDTVLKEMLPDRPQANHQFVQPLFVSRAPTRKMSGQ
GHMETIKSAKRLAEGISVLRIPLTQLKPNLLENMVNKEREPALYAGLKARLAEFNQDPAKAFATPFYKQGGQQVKAIRVEQVQKSGVLVR
ENNGVADNASIVRTDVFIKNNKFFLVPIYTWQVAKGILPNKAIVAHKNEDEWEEMDEGAKFKFSLFPNDLVELKTKKEYFFGYYIGLDRAT
GNISLKEHDGEISKGKDGVYRVGVKLALSFEKYQVDELGKNRQICRPQQRQPVR

>Campylobacter_jejuni
MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAK
AYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNV
RNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLN
NLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKK
ALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKV
LNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKML
EIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARL
VLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNS
AELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNG
DMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLIV
SKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK

FIGURE 12 C.

>Treponema_denticola
MKKEIKDYFLGLDVGTGSVGWAVTDTDYKLLKANRKDLWGMRCFETAETAEVRRLHRGARRRIERRKKRIKLLQELFSQEIAKTDEGFFQ
RMKESPFYAEDKTILQENTLFNDKDFADKTYHKAYPTINHLIKAWIENKVKPDPRLLYLACHNIIKKRGHFLFEGDFDSENQFDTSIQALFEY
LREDMEVDIDADSQKVKEILKDSSLKNSEKQSRLNKILGLKPSDKQKKAITNLISGNKINFADLYDNPDLKDAEKNSISFSKDDFDALSDDLA
SILGDSFELLLKAKAVYNCSVLSKVIGDEQYLSFAKVKIYEKHKTDLTKLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSGYVGVCKTKSKKL
IINNSVNQEDFYKFLKTILSAKSEIKEVNDILTEIETGTFLPKQISKSNAEIPYQLRKMELEKILSNAEKHFSFLKQKDEKGLSHSEKIIMLLTFKI
PYYIGPINDNHKKFFPDRCWVVKKEKSPSGKTTPWNFFDHIDKEKTAEAFITSRTNFCTYLVGESVLPKSSLLYSEYTVLNEINNLQIIIDGK
NICDIKLKQKIYEDLFKKYKKITQKQISTFIKHEGICNKTDEVIILGIDKECTSSLKSYIELKNIFGKQVDEISTKNMLEEIIRWATIYDEGEGKTIL
KTKIKAEYGKYCSDEQIKKILNLKFSGWGRLSRKFLETVTSEMPGFSEPVNIITAMRETQNNLMELLSSEFTFTENIKKINSGFEDAEKQFS
YDGLVKPLFLSPSVKKMLWQTLKLVKEISHITQAPPKKIFIEMAKGAELEPARTKTRLKILQDLYNNCKNDADAFSSEIKDLSGKIENEDNLR
LRSDKLYLYYTQLGKCMYCGKPIEIGHVFDTSNYDIDHIYPQSKIKDDSISNRVLVCSSCNKNKEDKYPLKSEIQSKQRGFWNFLQRNNFIS
LEKLNRLTRATPISDDETAKFIARQLVETRQATKVAAKVLEKMFPETKIVYSKAETVSMFRNKFDIVKCREINDFHHAHDAYLNIVVGNVYN
TKFTNNPWNFIKEKRDNPKIADTYNYYKVFDYDVKRNNITAWEKGKTIITVKDMLKRNTPIYTRQAACKKGELFNQTIMKKGLGQHPLKKE
GPFSNISKYGGYNKVSAAYYTLIEYEEKGNKIRSLETIPLYLVKDIQKDQDVLKSYLTDLLGKKEFKILVPKIKINSLLKINGFPCHITGKTNDS
FLLRPAVQFCCSNNEVLYFKKIIRFSEIRSQREKIGKTISPYEDLSFRSYIKENLWKKTKNDEIGEKEFYDLLQKKNLEIYDMLLTKHKDTIYK
KRPNSATIDILVKGKEKFKSLIIENQFEVILEILKLFSATRNVSDLQHIGGSKYSGVAKIGNKISSLDNCILIYQSITGIFEKRIDLLKV

>Wolinella_succinogen
MIERILGVDLGISSLGWAIVEYDKDDEAANRIIDCGVRLFTAAETPKKKESPNKARREARGIRRVLNRRRVRMNMIKKLFLRAGLIQDVDLD
GEGGMFYSKANRADVWELRHDGLYRLLKGDELARVLIHIAKHRGYKFIGDDEADEESGKVKKAGVVLRQNFEAAGCRTVGEWLWRER
GANGKKRNKHGDYEISIHRDLLVEEVEAIFVAQQEMRSTIATDALKAAYREIAFFVRPMQRIEKMVGHCTYFPEERRAPKSAPTAEKFIAIS
KFFSTVIIDNEGWEQKIIERKTLEELLDFAVSREKVEFRHLRKFLDLSDNEIFKGLHYKGKPKTAKKREATLFDPNEPTELEFDKVEAEKKA
WISLRGAAKLREALGNEFYGRFVALGKHADEATKILTYYKDEGQKRRELTKLPLEAEMVERLVKIGFSDFLKLSLKAIRDILPAMESGARYD
EAVLMLGVPHKEKSAILPPLNKTDIDILNPTVIRAFAQFRKVANALVRKYGAFDRVHFELAREINTKGEIEDIKESQRKNEKERKEAADWIAE
TSFQVPLTRKNILKKRLYIQQDGRCAYTGDVIELERLFDEGYCEIDHILPRSRSADDSFANKVLCLARANQQKTDRTPYEWFGHDAARWN
AFETRTSAPSNRVRTGKGKIDRLLKKNFDENSEMAFKDRNLNDTRYMARAIKTYCEQYWVFKNSHTKAPVQVRSGKLTSVLRYQWGLE
SKDRESHTHHAVDAIIIAFSTQGMVQKLSEYYRFKETHREKERPKLAVPLANFRDAVEEATRIENTETVKEGVEVKRLLISRPPRARVTGQ
AHEQTAKPYPRIKQVKNKKKWRLAPIDEEKFESFKADRVASANQKNFYETSTIPRVDVYHKKGKFHLVPIYLHEMVLNELPNLSLGTNPEA
MDENFFKFSIFKDDLISIQTQGTPKKPAKIIMGYFKNMHGANMVLSSINNSPCEGFTCTPVSMDKKHKDKCKLCPEENRIAGRCLQGFLDY
WSQEGLRPPRKEFECDQGVKFALDVKKYQIDPLGYYYEVKQEKRLGTIPQMRSAKKLVKK

>Streptococcus_agalac
MNKPYSIGLDIGTNSVGWSIITDDYKVPAKKMRVLGNTDKEYIKKNLIGALLFDGGNTAADRRLKRTARRRYTRRRNRILYLQEIFAEEMSK
VDDSFFHRLEDSFLVEEDKRGSKYPIFATLQEEKDYHEKFSTIYHLRKELADKKEKADLRLIYIALAHIIKFRGHFLIEDDSFDVRNTDISKQY
QDFLEIFNTTFENNDLLSQNVDVEAILTDKISKSAKKDRILAQYPNQKSTGIFAEFLKLIVGNQADFKKYFNLEDKTPLQFAKDSYDEDLENL
LGQIGDEFADLFSAAKKLYDSVLLSGILTVIDLSTKAPLSASMIQRYDEHREDLKQLKQFVKASLPEKYQEIFADSSKDGYAGYIEGKTNQE
AFYKYLSKLLTKQEDSENFLEKIKNEDFLRKQRTFDNGSIPHQVHLTELKAIIRRQSEYYPFLKENQDRIEKILTFRIPYYIGPLAREKSDFAW
MTRKTDDSIRPWNFEDLVDKEKSAEAFIHRMTNNDFYLPEEKVLPKHSLIYEKFTVYNELTKVRYKNEQGETYFFDSNIKQEIFDGVFKEH
RKVSKKKLLDFLAKEYEEFRIVDVIGLDKENKAFNASLGTYHDLEKILDKDFLDNPDNESILEDIVQTLTLFEDREMIKKRLENYKDLFTESQ
LKKLYRRHYTGWGRLSAKLINGIRDKESQKTILDYLIDDGRSNRNFMQLINDDGLSFKSIISKAQAGSHSDNLKEVVGELAGSPAIKKGILQ
SLKIVDELVKVMGYEPEQIVVEMARENQTTNQGRRNSRQRYKLLDDGVKNLASDLNGNILKEYPTDNQALQNERLFLYYLQNGRDMYTG
EALDIDNLSQYDIDHIIPQAFIKDDSIDNRVLVSSAKNRGKSDDVPSLEIVKDCKVFWKKLLDAKLMSQRKYDNLTKAERGGLTSDDKARFI
QRQLVETRQITKHVARILDERFNNELDSKGRRIRKVKIVTLKSNLVSNFRKEFGFYKIREVNNYHHAHDAYLNAVVAKAILTKYPQLEPEFV
YGDYPKYNSYKTRKSATEKLFFYSNIMNFFKTKVTLADGTVVVKDDIEVNNDTGEIVWDKKKHFATVRKVLSYPQNNIVKKTEIQTGGFSK
ESILAHGNSDKLIPRKTKDIYLDPKKYGGFDSPIVAYSVLVVADIKKGKAQKLKTVTELLGITIMERSRFEKNPSAFLESKGYLNIRADKLIILP
KYSLFELENGRRRLLASAGELQKGNELALPTQFMKFLYLASRYNESKGKPEEIEKKQEFVNQHVSYFDDILQLINDFSKRVILADANLEKIN
KLYQDNKENISVDELANNIINLFTFTSLGAPAAFKFFDKIVDRKRYTSTKEVLNSTLIHQSITGLYETRIDLGKLGED

>Streptococcus_gordon
MNGLVLGLDIGIASVGVGILEKDTGKIIHASSRLFPAATADNNVERRSNRQGRRLNRRKKHRSVRLQDLFEGYGLLTDFSKVSMNLNPYQ
LRVQGMENQLTNEELFVALKNIVKRRGISYLDDASEDGGTVSSDYGKAVEENRKLLAEKTPGQIQLERFEKYGQLRGDFTVEENGEKHR
LINVFSTSAYRKEAERILRKQQEFNSKITDEFIEDYLIILTGKRKYYHGPGNEKSRTDYGRFRTDGTTLDNIFGILIGKCTFYTEEYRASKASY
TAQEFNLLNDLNNLTVPTETKKLSEEQKKLIIEYAKSAKTLGASTLLKYIAKMIDASVDQIRGYRVDVNNKPEMHTFEVYRKMQSLETIKVEE
LPRKVLDELAHILTLNTEREGIEEAINSKLKDIFNRDQVLELVQFRKNNSSLFSKGWHNFSIKLMMELIPELYETSEEQMTILTRLGKQRSKE
TSKRTKYIDEKELTEEIYNPVVAKSVRQAIKIINEATKKYGIFDNIVIEMARENNEEDAKKDYIKRQKANQDEKNAAMEKAAFQYNGKKELPD
NIFHGHKELTTKIRLWHQQGEKCLYTGKNIPISDLIHNQYKYEIDHILPLSLSFDDSLSNKVLVLATANQEKGQRTPFQALDSMDDAWSYRE
FKSYVKDSKLLSNKKKDYLLTEEDISKIEVKQKFIERNLVDTRYSSRVVLNALQDFYKSHQLDTTISVVRGQFTSQLRRKWGIEKSRETYHH
HAVDALIIAASSQLRLWKKHSNPLIAYKEGQFVDSETGEIVSLSDEEYKELVFKAPYDHFVDTLRSKKFEDSILFSYQVDSKYNRKISDATIY
ATRKAKLDKEKKEYTYTLGKIKDIYALGTKTPSKTGFYKFLDLYKTDKSQFLMYQKDRKTWDEVIEKIIEQYRPFKEYDKNGKEVDFNPFEK
YRIGNGPIRKYSKKGNGPEIKSLKYYDILLGKHKNITPDGSRNTVALLSLNPWRTDVYYNSETKKYEFLGLKYADLCFEEGGAYGISEVKYK
KIREKEGIGKNSEFKFTLYKNDLILIKDTETNCQQFFRFWSRTGKDNPKSFEKHKIELKPYEKAKFEKGEELKVLGKVPPSSNQFQKNMQI
ENLSIYKVKTDILGNKHFIKKEGDEPKLKFKK

FIGURE 12 D.

>Mycoplasma_gallisept
MNINMNNSIKSKPEVTIGLDLGVGSVGWAIVDNETNIIHHLGSRLFSQAKTAEDRRSFRGVRRLIRRRKYKLKRFVNLIWKYNSYFGFKNK
EDILNNYQEQQKLHNTVLNLKLEALNAKIDPKALSWILHDYLKNRGHFYEDNRDFNVYPTEELANYFDEFGYYKGIIDSKNDDDDKLEEGL
TKYKFSNQHWLEEVKKVLSNQTGLPEKFKEEYESLFSYVRNYSEGPGSINSVSPYGIYHLDEKEGKVVQKYNNIWDKTIGKCSIFPDEYR
APKNSPIAMIFNEINELSTIRSYSIYLTGWFINQEFKKAYLNKLLDLLIKTNSEKPIDARQFKKLREETIAESIGKETLKDVESEEKLEKDDHKW
KLKGLKLNTNGKIQYNDLSSLAKFVHKLKQHLKLDFLLEDQYTPLDKINFLQSLYVYLGKHLRYSNRVDSANLKEFSDSSRLFERVLQEQK
DGLFKLFEQTDKDDEKILTQTHSLSTKAMLLAITRMTNLDNDEDNQKNNDKGWNFEAIKNFDQKFIDITKTNNNLSLKQDKRYLDDQFIND
AILSPGVKRILREATKVFNAILKQFSEEYDVTKVVIELARELSEEKELENTKNYKKLIKKNGDKISEGLKALGIAEDKIEEILKSPTKSYKVLLW
LQQDHIDPYSQKEIAFDDILTKTEKTEIDHIIPYSISFDDSSSNKLLVLAESNQAKSNQTPYEFINSGKAEITWEVYEAYCHKFKNGDSSLLDS
TQRSKKFAKMMKTDTSSKYDIGFLARNLNDTRYATIVFRDALKDYANNHLVEDKPMFKVVCINGGVTSFLRKNFDPKSWYAKKDRDKNIH
HAVDASIISIFSNETKTLFNQLTKFADYKLFKNTDGSWKKIDPKTGVVSEVTDENWKQIRVRNQVSEIAKVIDKYIQDSNIERKARYSRKIEN
KTNISLFNDTVYSAKKVGYEDQIKRKNLKTLDIHESAEENKNSKVKKQFVYRKLVNVSLLNNDKLADLFAEKEDILMYRANPWVINLAEQIF
NEYTENKKIKSQNVFEKYMLDLTKEFPEKFSEAFVKSMIRNKTAIIYNVEKDVVHRIKRLKILSSELKENKWSNVIIRSKNESGTKLSYQDTIN
SIALMIMRSIDPTAKKQYIRVPLNTLNLHLGDQDFDLHNIDAYLKKPKFVKYLKANEIGDEYKPWRVLTSGTLLIHKKDKKLMYISSFQNLND
LIEIKNLIETEYKENVDSDPKKKKKASQILRSLSVILNDYILLDAKYNFDILGLSKNKIDEILNSKLDLDKIAK

>gi|187736489|ref|YP_001878601.1|/1-1099
MSRSLSFDIGYASIGWAVIASASHDDADPSVCGCGTVLFPKDDCQAFKRREYRRLRRNIRSRRVRIERIGRLLVQAQIITPEMKETSGHPA
PFYLASEALKGHRTLAPIELWHVLRWYAHNRGYDNNASWSNSLSEDGGNGEDTERVKHAQDLMDKHGTATMAETICRELKLEEGKADA
PMEVSTPAYKNLNTAFPRLIVEKEVRRILELSAPLIPGLTAEIIELIAQHHPLTTEQRGVLLQHGIKLARRYRGSLLFGQLIPRFDNRIISRCPV
TWAQVYEAELKKGNSEQSARERAEKLSKVPTANCPEFYEYRMARILCNIRADGEPLSAEIRRELMNQARQEGKLTKASLEKAISSRLGKE
TETNVSNYFTLHPDSEEALYLNPAVEVLQRSGIGQILSPSVYRIAANRLRRGKSVTPNYLLNLLKSRGESGEALEKKIEKESKKKEADYADT
PLKPKYATGRAPYARTVLKKVVEEILDGEDPTRPARGEAHPDGELKAHDGCLYCLLDTDSSVNQHQKERRLDTMTNNHLVRHRMLILDR
LLKDLIQDFADGQKDRISRVCVEVGKELTTFSAMDSKKIQRELTLRQKSHTDAVNRLKRKLPGKALSANLIRKCRIAMDMNWTCPFTGATY
GDHELENLELEHIVPHSFRQSNALSSLVLTWPGVNRMKGQRTGYDFVEQEQENPVPDKPNLHICSLNNYRELVEKLDDKKGHEDDRRR
KKKRKALLMVRGLSHKHQSQNHEAMKEIGMTEGMMTQSSHLMKLACKSIKTSLPDAHIDMIPGAVTAEVRKAWDVFGVFKELCPEAADP
DSGKILKENLRSLTHLHHALDACVLGLIPYIIPAHHNGLLRRVLAMRRIPEKLIPQVRPVANQRHYVLNDDGRMMLRDLSASLKENIREQLM
EQRVIQHVPADMGGALLKETMQRVLSVDGSGEDAMVSLSKKKDGKKEKNQVKASKLVGVFPEGPSKLKALKAAIEIDGNYGVALDPKPV
VIRHIKVFKRIMALKEQNGGKPVRILKKGMLIHLTSSKDPKHAGVWRIESIQDSKGGVKLDLQRAHCAVPKNKTHECNWREVDLISLLKKY
QMKRYPTSYTGTPR

>gi|269986295|gb|EEZ92602.1|/1-872
MLGSSRYLRYNLTSFEGKEPFLIMGYYKEYNKELSSKAQKEFNDQISEFNSYYKLGIDLGDKTGIAIVKGNKIILAKTLIDLHSQKLDKRREA
RRNRRTRLSRKKRLARLRSWVMRQKVGNQRLPDPYKIMHDNKYWSIYNKSNSANKKNWIDLLIHSNSLSADDFVRGLTIIFRKRGYLAFK
YLSRLSDKEFEKYIDNLKPPISKYEYDEDLEELSSRVENGEIEEKKFEGLKNKLDKIDKESKDFQVKQREEVKKELEDLVDLFAKSVDNKID
KARWKRELNNLLDKKVRKIRFDNRFILKCKIKGCNKNTPKKEKVRDFELKMVLNNARSDYQISDEDLNSFRNEVINIFQKKENLKKGELKG
VTIEDLRKQLNKTFNKAKIKKGIREQIRSIVFEKISGRSKFCKEHLKEFSEKPAPSDRINYGVNSAREQHDFRVLNFIDKKIFKDKLIDPSKLR
YITIESPEPETEKLEKGQISEKSFETLKEKLAKETGGIDIYTGEKLKKDFEIEHIFPRARMGPSIRENEVASNLETNKEKADRTPWEWFGQD
EKRWSEFEKRVTSLYSKKKISERKREILLNKSNEYPGLNPTELSRIPSTLSDFVESIRKMFVKYGYEEPQTLVQKGKPIIQVVRGRDTQALR
WRWHALDSNIIPEKDRKSSFNHAEDAVIAACMPPYYLRQKIFREEAKIKRKVSNKEKEVTRPDMPTKKIAPNWSEFMKTRNEPVIEVIGKV
KPSWKNSIMDQTFYKYLLKPFKDNLIKIPNVKNTYKWIGVNGQTDSLSLPSKVLSISNKKVDSSTVLLVHDKKGGKRNWVPKSIGGLLVYIT
PKDGPKRIVQVKPATQGLLIYRNEDGRVDAVREFINPVIEMYNNG

>gi|282880052|ref|ZP_06288774.1|/1-1218
MNKRILGLDTGTNSLGWAVVDWDEHAQSYELIKYGDVIFQEGVKIEKGIESSKAAERSGYKAIRKQYFRRRLRKIQVLKVLVKYHLCPYLS
DDDLRQWHLQKQYPKSDELMLWQRTSDEEGKNPYYDRHRCLHEKLDLTVEADRYTLGRALYHLTQRRGFLSNRLDTSADNKEDGVVK
SGISQLSTEMEEAGCEYLGDYFYKLYDAQGNKVRIRQRYTDRNKHYQHEFDAICEKQELSSELIEDLQRAIFFQLPLKSQRHGVGRCTFE
RGKPRCADSHPDYEEFRMLCFVNNIQVKGPHDLELRPLTYEEREKIEPLFFRKSKPNFDFEDIAKALAGKKNYAWIHDKEERAYKFNYRM
TQGVPGCPTIAQLKSIFGDDWKTGIAETYTLIQKKNGSKSLQEMVDDVWNVLYSFSSVEKLKEFAHHKLQLDEESAEKFAKIKLSHSFAAL
SLKAIRKFLPFLRKGMYYTHASFFANIPTIVGKEIWNKEQNRKYIMENVGELVFNYQPKHREVQGTIEMLIKDFLANNFELPAGATDKLYHP
SMIETYPNAQRNEFGILQLGSPRTNAIRNPMAMRSLHILRRVVNQLLKESIIDENTEVHVEYARELNDANKRRAIADRQKEQDKQHKKYGD
EIRKLYKEETGKDIEPTQTDVLKFQLWEEQNHHCLYTGEQIGITDFIGSNPKFDIEHTIPQSVGGDSTQMNLTLCDNRFNREVKKAKLPTEL
ANHEEILTRIEPWKNKYEQLVKERDKQRTFAGMDKAVKDIRIQKRHKLQMEIDYWRGKYERFTMTEVPEGFSRRQGTGIGLISRYAGLYL
KSLFHQADSRNKSNVYVVVKGVATAEFRKMWGLQSEYEKKCRDNHSHHCMDAITIACIGKREYDLMAEYYRMEETFKQGRGSKPKFSKP
WATFTEDVLNIYKNLLVVHDTPNNMPKHTKKYVQTSIGKVLAQGDTARGSLHLDTYYGAIERDGEIRYVVRRPLSSFTKPEELENIVDETV
KRTIKEAIADKNFKQAIAEPIYMNEEKGILIKKVRCFAKSVKQPINIRQHRDLSKKEYKQQYHVMNENNYLLAIYEGLVKNKVVREFEIVSYIE
AAKYYKRSQDRNIFSSIVPTHSTKYGLPLKTKLLMGQLVLMFEENPDEIQVDNTKDLVKRLYKVVGIEKDGRIKFKYHQEARKEGLPIFSTP
YKNNDDYAPIFRQSINNINILVDGIDFTIDILGKVTLKE

FIGURE 12 E.

>gi|258648111|ref|ZP_05735580.1|/1-1234
MPNFWSIFTENLQTYLMKTRILGLDTGTNSLGWAVVDRDENNQYTPICRGDLIFTEGVKQEKGIEFSLAAERTSYRASRRHYFRRRLRKV
EVLKVLVKHKLCPPLSDEALNRWHTKKEYPLDEDFILWQRTNEADNKNPYHDRHICLHRKLNLKDSQTDRYTLGRALYHLTQRRGFLSN
RLDQNEEAEETGKVKSAISELSKEMKEAGCEFLGDYFYLLYRTEGNRVRIRSRYTDREKHYLQEFKEICRVQELDEDLAKALEKTIFFQRP
LKSQRQGVGKCTMEKGHSRCAVSHPLYEEYRLLCYLNNIRIKTPSDDRLRPLNAEEREKITPLFYKEKNFDFEKIAKKLAGKTPFGWIRDA
EDWPYQFNYRMTQEAPACPTTAQLIKVFGEDWKAALAETYTMSAKQDGTLKTEDEICTDIWNVLYSFSSKEKLKEFGLKKLQLDEVQAK
AFSEIRLTRDFASLSLKAIRKILFFLRTGEDYSKAIFLANIPTIVGEEIWNDIERRSRLLYKVDLALRPQEERDAEEHRTIAERIKAAVAAEVPD
APIDLLYHPSMIEAYPDAKRNERFDCYQLGSPKTNAVRNPMAMRSLHMVRKVINHLLRKHIIDEKTEIHIEYARELNDANKRQAIADWQRE
LSKRHTAYAQNIRQLYKAETGLVIEPTKADILKYQLWEEQQHVCLYTGKKIGIADFLGANPKFDIEHTIPRSRGGDSTQENLTLCESKFNRD
IKRDKLPTELANYEEILARIEHWKKQCYALKLKRDGIRTHSGLEKSIKDSRIRRRHLVGLEYGYLKGKYDRFTMKEVPEGFSRRQGAGIGLI
SKYAGLFLKSLFKDPKNPNRSNVYVVKGAITAEFRKLWGLQQADEAKSRENHTHHCIDAYTIACIGPCEYAALAAYYRSDEEFKYGRRRE
KPQFEKPWPTFTEDLLKLQEELLIVHQTTDKLGKRDRRKVKTPRGKFLTGGDSARGRLHQETYYGAINYDGNIKYVVRKPLDSLTEKDLD
KIVDETVMGIVKGAVEAKGSLKEALADGIYMNKEKGILIKRVRIFSDVSEPINLKQLRDVSTKEYKRHIHLRKDENYLFAIYEGIDKKKNKKQT
ETATVDYLDAAKFYKDSKRRDKANSLVGDYSKKNKLPLKCILKKRQMVLFYEHSPEEIRFKDKKDLSRRLYIVTSIAKDGRVKFLHHQAAT
ATGIQSSPSPFTNDGEHQPYYRLSASNIKVLVAGIDFDINILGEITPKL

>gi|260592128|ref|ZP_05857586.1|/1-1496
MKTILGLDLGTNSIGWAKVCVDDNGNYLRDIKLGSRIIPMSQDVLGNFDKGVTESQTSVRTGFRGVRRIRERSLLRRERLHRVLHVLGYLP
PHYDSAIGWNKNEPSTYGKFLNHGEPKLAWEHKEDGSMRFLFMESFHEMMADFAKSQPSLVSDGKKIPLDWTIYYLRKKALSQPIRKEE
LAWLLLNFNQKRGYYQLRGEEDDDNPTKCEEYYELKVVGIEADEEGKNGNTWYNVKLENGWIYRRQSKIPLDDWIGKVREFIVTTEYEK
DGITLKKDKNGDVKRSFRSPSEDDWGLRKKRTEKLLDSSKMTVGAFVYDHLLAEPSDKIRGKFIRTIERKYYKKELYAILCEQSKYHEELR
SGDYLNKCAEELYKNNQPHRNSLLTKDLAYLLLEDLIFYQRPLKTKKSLIAECPYEAYKYVDKETGEIKRQGIKCIAKSNPYYQEFRLWQFI
SNLRLFNKADDKEVTADYLPTQDDYVRLFAYLNDRKEINQETLLKEFFKLKKITVNSEKVFPIRWNYIEDKEKKYPCNETRHELLLALDRAR
IDHSWLEKPGQEYRLWHLLYSVESKEETKEALRKLKDDDAFVESFLKVKPFKREYGAYSEKAIKKLLAVMRMGSLWRETDICDTTKGNIT
KILQGEIDEKLKEKISASVSSYNQLSDFQGLPVWLACYVVYGRHSEAAEIQHWKTPESLLAYINGFKQHSLRNPIVEQCILETLRTVHDIWK
EVGHIDEIHVELGRNMKSTADQRARISENVLRNENTNLRIKSLLLELKNDPNIKDVRPYSPMQQEKLRIYEEGVLQELTKDDKDFTEITKIST
MAQPSAKELQRYKLWLEQKYRSPYTGKSISLTKLFTSAYQIEHVIPQSRYFDDSFSNKVICESEVNSLKDNMLGYEFIKNHGGEIIHCTML
GDVKVLSESEYKAFVNDHYAHNKRKRENLLLEDIPQEFLNRQMNDSRYISKVVKALLSNIVRAEGEEEATSKFVVSCSGGITDRLKKDWG
LNDVWNKIVYPRFERLNRLTGTEAFGQWENKDGKRVFQTAVPLELQRGFNKKRIDHRHHAMDALVIACTSRNIINYLNNESANNPKKRED
LRQKLCDKNRTIRKPWETFTQDATAALEDVIVSFKNYVRIINKATNYYERYDVEGKKVIAEQRGDDMWAIRKPMHEATYYGRVNLRRKKK
VNLRVALENISSICDNKLRIYINGLVEKGFNTKQILDKFKNTNDHFSMQKIDKVEIWQFTNEEEKLVATRKPLDTSFDTKKIASITDTGIQTILL
NYLETKGGDSTIAFTPEGIVEMNQNISKYNNGKEHMPILKVRLYTKLGKKIPVGQTGCKAKQFVIAEAGTNLYFAIYEDEEGKRSYRTIPLN
EVVERLKQGLSPVPEKNDKDVHLKFYLSPNDLVYVPIEDEHESLAKDHIYKFVDSSGTTANFIPHRSANVIFELNKKDAEKFCSGSIIQNEY
GVGSSQSKNQKAITGEMIKAVCRKLEVDRLGNVIKVI

>gi|282881485|ref|ZP_06290156.1|/1-1487
MKNILGLDLGSNSVGWALVKVEENGNPIGSIKMGSRIIPMSQDILGSFEKGDTVSQTAQRTLYRGKRRLIERHVLRRERLCRVLHIMNFLP
AHFDRLLGWDKTDNKTYGKFIDDSEPKLAWRQNKNEMGKMEFVFMDSFHEMLSDFAKHQPQLIANGKKVPLDWTIYYLRKKALTQLISK
EELAWILLNFNKKRGYYQLRGEEEEEQPTKKEEYKVLKVISVDADEGQKGNGIWYNIHLEDGGIYKKKSDIPLYDWVGKTLQLIVVTTYEK
DGKTPKIFDSGKQNPSYRLPKEDDWGLVKKRTESQLQSSGKTVGAFIYDNILSKPDDKIRGNLVRTIERKYYKKELIEILQQQAKYHDELH
NQNLLESCAKELYAHNDVHYNEVIKSNMINLLVNDIIFYQRPLKSKKSLIANCKYESYEYVDKETGEIKEMPIKCIAKSNPYYQEFRLWQFIH
NLRLSDVVTREDVTAQYLSTPQDYCNLFTYLNDRKEIKQDILLKDFFKIKKVQIDKEKVFPIEWNYIKDATKSYPCNETRYELLSALKRAGM
DKQWLDDTPNMQYRLWHLLYSVEEKEESAKALRKLYDDDKFVTSFLKIKPFNKCYGAYSEKAIKRLLTLMRMGKAWNEQAIDAKTLKCIQ
QIITDNVDSKLKEKIDDSRYVLHHVSDFQGLPTYLASYVIYGRHSEVTDIQCWDSPEDLLKFIRDFKQHSLRNPIVEQVILETLRVVYDIWKA
EKHIDEIHVEVAREMKLTAQQRNEANLRNLNNEATNLRIKYLLEELKNDTFIKDVRPLSPIHQEKMRIYEQAVLGNLNKQDKDYDDIIKISKK
EHPTNAELIRYKLWLEQRYCSPYTGKAISLSKLFTTAYEIEHIIPQSRYFDNSFNNKVICEAEVNKAKGNMLGYEFIKKQGGQKIYCTQLGK
AVTILKEEAYKQFVNEHYAHNKRKRDNLLAENIPESFSSRQLNDTRYITKTIMSLLSNVVRQQSEEEATAKNLLPSSGSVTDRLKNDWGLK
DVWNTIVTPRFERLNKLSGTNDFGEMKEENGNRYFQTNVPLEYEKGFKKKRIDHRHHAMDALVIACTTRNMVNYISNANANSPKQREDL
RQLLCDKNRIINKPWDTFTQDALKALNDIVVSFKNNVRIINRATNRYQRYDKNGKKIIFSQKGDDMWAIRKSMHKETVFGRVNLIRKEVLPI
AKALDNISAICNAQLRAYVNDLVDKHFNKKQLVAHFKSLNYKWNRQDVSKVEVWISSDNKTPMVAVRKPLDTSFDKKKIASITDTGIQTILL
NYLESKGGDASVAFAPEGIQEMNSNIELYNNGKAHKPIKNVRLSEPLGAKYQVGERGAKTKKYVEADKGTNLFFAIYEDEEGKRYYYSVP
LREAIERQKQKLSPVPEYNEKGVALKFYLSPNDLVYVPTEEERTMGTCHIDKARIYKMVSSTGCSCYFVPSTFAKVIVDKVELQSLNKMEK
AITDECIKEVCWKLKVDRLGNIINVIK

FIGURE 12 F.

>gi|228473057|ref|ZP_04057814.1|/1-1436
MKYIKYILGLDIGSNSIGWALIQLDLLQKTKRIIAAGSRIIPMDQKVLSDFGKGITISPTAQRTSYRNARHCRERHLLRRERLHRVLHILGFLPT
HYDKQIDFSKRYGKFVDNTEPKIAYNEGKFLFMDSFQEMLSDFKSHQPDLFYTKKNGEESKIPYDWTIYYLRKKALSQKISKEELAWIILNF
NQKRGYYQLREEEGNKENDFEIISSKIISVIKKDKDKKYNKYWYDIILENGLVYQAAFYNDFVYNWINQDREFIIERTSLKNGAEKVKLSYLP
TFDEIDKMDKKEQDKYYKKIKIKTENSIGNKTVGCYIYDHILQNPNEKIKGKLVRTIERKYYKEELKAILEKQIALQPELFTDQLLADCIRELYS
KNEAQQRNLAARDFVHLLVEDIIFYQRPLRRQKSTLANCSLESRSYIDKDSHTRKEAPLKVCPKSNPYYQEFRVLQWLQNLKIYEIDSDQE
VTHQFIKTLEDKQQLFEFLMAQKEIDCEELLKYFLNQTYPNAKEKALKSELKKWKDTYRWNYVYDIGEKSSKKYPMNETRYELKRHLEKV
ANLPDDFLSSEVEYLLWHLIYSVKDKVAYEKGLKKFAQKHHLDKDSFVESFKKFNPYPSEYGSFSEKAIRKLLPLMRFGAHWDFNHIDKN
TQKRIEDLITGVENEEIRTILREKAEKYQLEKETDFQDLPLWLAQYIVYNRHAEASSLEKWTSVNDLETYLNEFKQHSLRNPIVEQVVMETL
RVVRDIWQQYGQGQASFFNEIHIELGRELKKTAKEREKLSKQNQKNEDINLSIKEKLKELEKYNTRAYSITHQEKYKLWLEQKHISPYTGQI
ISPENLFTDNYEIEHIIPQSRFFDDSLSNKVVCESIVNKAPYKDRQLGLEFIKNQGGRIVKELSKENKTIKIFTEEQYRTFIKEHYSNNPEKTK
KLLLEEIPEKMVARQMNDTCYISKFISEILSKIVRSDEQDEGVHSKNVILCTGKITSILRQDWGLNDIWDNLILPRFERMNNLLQKEKFTIYSI
EKQKKIPVVPINDQFKSFQKKRIDHRHHAMDAIIIACTTREHINYINNQHALEKEKDKKEKQEERDKLREKLCIKKYNNRSEGNYSWIFKKP
WDTFTQEVRVALEDIVVSFKQNLRVINKATNYYERWVEENGKYIKKMVKQEGTNWAIRKPLHEETISGKIYLDREKITKNDILTATRKTVDS
SFDDKRIKKITDTGIQKILLNYLKYKGSPEIAFSPEGLEELNKNLSIYNDGKPHKPIYKVRIYEKGSGRFALGEKGRKYKKYVQGASNLYFGI
YQGAKRSFATIPLNEVIERQKQGLPSVPEYNEKGDLLLFSLSPNDLVYVPIEGEIIENIDFRNLSKEQKERIYNVNDFSSTCYFTPNRIAKAIF
PKEVDLQRKGDKLSGSYDTKTASLEGIQIKEVCIKLKVDRLGNISKA

>gi|300771242|ref|ZP_07081118.1|/1-1426
MKHILGLDLGTNSIGWALIKQDFENKQGEILGMGSRIIPMSQDILGDFGKGNSVSQTAERTKYRSVRRLRERFLLRRERLHRILNVLGFLPE
HYASQIDFTKRFGKFKEETEPKLVYSNEGFIFKKSFEEMLSDFKTHQPQLLEDNKKIPYDWTIYYLRKKALSRKIEKEELAWILLNFNQKRG
YYQLRGEEEEENPNKLIEFYSLKIVDVLADEPQKGKSDIWYSLILENGWIYRRSSKTPLFDWKEKTRDFIVTIDLNDDGSVKTDKEGNEKR
SFRAPGENDWTLVKKKTEQEIDSSRKTVGTYIYETLLKNPKQKIKGKLVRTIERKFYKTELKQILEKQKEFHPELRQDDLYSDCVRELYKNN
EAHQLTLSKKDFVHLFMEDIIFYQRPLGSQKSSISNCTLEFRKYKDENGTEHTQYLKAIPKSNPYYQEFRIWQWMYNLNIYKKEDDTNVTK
DFLNTKADFESLFEFLNNRKEVDQKALLKHFKLNDKTHRWNYVEDKKYPCNETKTMILSRLDKVENISDDFLTREMEQKIWHIIYSVNDKV
EYEKALKSFAGKNNLEDNSFFEAFKKFPPFKSEYGAFSEKAIKKLLPLMRMGKFWNYNNISEQSKSRIQKIITGEYDENIKDKVREKAIHLT
TETDFQGLQLWLVQYIVYGRHSEAAMLGKWNSADDLEEFLKEFKQHSLRNPIVEQVITETLRVVKDIWQKYGNGAKDFFREIHIELGREM
KLPAEERERLTRQISENENTNFRIKALLAEMMDDNSVENVRPFSPMQQEILKIYEEGVLNSDIEIEDDILKISKTAQPSSSDLKRYKLWLEQK
YKSPYTGQIIPLNKLFTPEYEIEHIIPQSRYFDDSFTNKIICESAVNKLKDNYIGLEFIKKLGGSIVECGNGKSVAILKHDAYEDFIKKHYANNR
SKRNKLLLEDIPEKMIERQMNDTRYISKYISGVLSNIVRVEDGTDEGVNSKNIVPGNGKITTQLKQDWGLNDVWNDLILPRFERMNHLTNS
TDFTAWNENHQKFLPTVPVEFSKGFSKKRIDHRHHALDALVIACATKDHVNLLNNQSAKSDTKRYDLKKKLMKFEKVAYNHPQTGERIER
EAPKQFLKPWETFTIDAKNHLETIIVSFKQNLRVINKATNYYEKYVEKEGTKVKERVEQKGTNWAIRKPMHKETVSGKIDLPRIKVPKGKILT
ATRKSLDTSFNLKTIGSITDTGIQKILKNYLQYKDNNPELAFSPEGIEDMNKNIEKFNNGKAHQSISKVRVFELGSKFPLGQSGNKKEKYVE
AAKGTNLFFAIYQDKNGKRSYETIPLNEVIERQKQGLSVVDLKGVDDFYLSPNDLVYVPSEDELENISMIDFNKMSSERIYKVVSFTGNQIF
FVRQDIATSVVNKAEFSTLNKMERAIDGTMIKETCVKLQIDRLGNISKI

>gi|212694363|ref|ZP_03302491.1|/1-1504
MKNIVGLDLGTNSIGWAVVNGSVNDDGSEQLVKIQASGSRIIPMDAAMIGDFNKGNSISQTAERTRLRGVRRLSERYLLRRERLHRILDIL
GFLPFHFAQDLDRHGKIVKGKEPKLAWRKNEAGQFEFIFQDSFKEMLEDFKLNYPNLITDDKKVPYDWTIYYLRKKGLTSKISKEELAWILL
NFNQKRGYYQLRGEEEEENKDKLVEFYALKVVAVEDSGEKKGKDIWYNVHLENGWVYRRTSNMSLDWVGKTKEFIVTTDLEKDGTPKK
DKEGNVKRSFRAPKEDDWSLVKKKTEADIDQTHKTVGAYIYDTLLGNPSQKIRGRLVRTIERKYYKEELKLILEKQKEFHLELQDRALYAA
CIEELYPINEAHRNNVGNRDFTYLFLEDILFYQRPLKSKKSLIDNCPYEENQGLDKETGEIKSFPVKCIAKSHPLFQEFRLWQFIINLRIYRKD
LAQDVDVTNELLKTEEDYVALFDWLNGKKEIDQKAFLKYPVFGLKKEIENYRWNYVQDKPYPCNETRSLMLARLEKCDISIEFLTKENEEA
LWHILYSVEDKNEIVKALQKFADKHDLNASFVDVFQKFPPFKKEYGSYSAKAIKKLLPLMRMGKYWSENSIDKGTRERIDKIITGEYDEKIE
NRVREKAIHLADVSDFKGLPLWLVCYIVYGRHSEAKEIIQWKSPADIDSYLRSFKQHSLRNPIVEQIIIETLRVVRDIWKQVGNIDEIHVELGR
EMKNPADKRKKMTQQMSENENTNLRIKYLLTEFLNPEYEVENVRPYSPSQQDILRIYEEGVLNSVSDLPEEISDILKKFTETDLKKRPSRS
EVLRYKLWLEQKYRSPYTGEVIPLGKLFTPAYEIEHVIPQSYYFDDSFSNKVICEAEVNKLKSNLLGHEFIVKHHGEIVELPFGKKVRIFTVD
EYEQFVKDNYSRTRSKMKKLLMDDIPEEFVTRQLNDSRYISKVVKSLLSNIVREKGEEEAISKNVIPCTGGVTDRLKKDWGINEVWNKIILP
RFQRLNELTGTNKFTTKNVGIQEIPTMPLELQKGFNKKRIDHRHHAMDAIIIACANRNIINYLNNKSATAKAELSRYDLQKMLCDKAKTDNN
GNYKWVIRKPWASFTQDTYLALENIIVSFKQNLRVINKATNRFLHYNEEGKKIFVKQGKGDNWAIRKSMHKDTVFGEVNLRRIKTVALNEA
MKNPQSIVVKDFKRKLLELWNLGFDAKRIKKYFEDNRETWSDINLSKIEVYYFSKDTKDRFFATRKPLDTSFDRKKIENNITDTGIQKILLRH
LELKDNNPDIAFSPDGIDEMNRNIIQLNNGKYHQPIIKIRWYEQADKFAVGQTGNKSSKFVEAAKGTNLFFAVYESNILDKKTNTIIKKRNYA
TIPLNVAIERQKQGLPVAPEDENGNDPIFVLSPNDLVYLPTDDELANGIIAQPLDRGRIYKMVSSSGEQCFFIKHIVANVLVDKFEFSPLNKM
ERALTGEMIKMICIPIKVDRLGNVLESSSSYKK

FIGURE 12 G.

>gi|81828495|sp|Q6KIQ7|Q6KIQ7_MYC/1-1236
MYFYKNKENKLNKKVVLGLDLGIASVGWCLTDISQKEDNKFPIILHGVRLFETVDDSDDKLLNETRRKKRGQRRRNRRLFTRKRDFIKYLI
DNNIIELEFDKNPKILVRNFIEKYINPFSKNLELKYKSVTNLPIGFHNLRKAAINEKYKLDKSELIVLLYFYLSLRGAFFDNPEDTKSKEMNKNE
IEIFDKNESIKNAEFPIDKIIEFYKISGKIRSTINLKFGHQDYLKEIKQVFEKQNIDFMNYEKFAMEEKSFFSRIRNYSEGPGNEKSFSKYGLYA
NENGNPELIINEKGQKIYTKIFKTLWESKIGKCSYDKKLYRAPKNSFSAKVFDITNKLTDWKHKNEYISERLKRKILLSRFLNKDSKSAVEKIL
KEENIKFENLSEIAYNKDDNKINLPIINAYHSLTTIFKKHLINFENYLISNENDLSKLMSFYKQQSEKLFVPNEKGSYEINQNNNVLHIFDAISNI
LNKFSTIQDRIRILEGYFEFSNLKKDVKSSEIYSEIAKLREFSGTSSLSFGAYYKFIPNLISEGSKNYSTISYEEKALQNQKNNFSHSNLFEKT
WVEDLIASPTVKRSLRQTMNLLKEIFKYSEKNNLEIEKIVVEVTRSSNNKHERKKIEGINKYRKEKYEELKKVYDLPNENTTLLKKLWLLRQ
QQGYDAYSLRKIEANDVINKPWNYDIDHIVPRSISFDDSFSNLVIVNKLDNAKKSNDLSAKQFIEKIYGIEKLKEAKENWGNWYLRNANGKA
FNDKGKFIKLYTIDNLDEFDNSDFINRNLSDTSYITNALVNHLTFSNSKYKYSVVSVNGKQTSNLRNQIAFVGIKNNKETEREWKRPEGFKS
INSNDFLIREEGKNDVKDDVLIKDRSFNGHHAEDAYFITIISQYFRSFKRIERLNVNYRKETRELDDLEKNNIKFKEKASFDNFLLINALDELN
EKLNQMRFSRMVITKKNTQLFNETLYSGKYDKGKNTIKKVEKLNLLDNRTDKIKKIEEFFDEDKLKENELTKLHIFNHDKNLYETLKIIWNEV
KIEIKNKNLNEKNYFKYFVNKKLQEGKISFNEWVPILDNDFKIIRKIRYIKFSSEEKETDEIIFSQSNFLKIDQRQNFSFHNTLYWVQIWVYKN
QKDQYCFISIDARNSKFEKDEIKINYEKLKTQKEKLQIINEEPILKINKGDLFENEEKELFYIVGRDEKPQKLEIKYILGKKIKDQKQIQKPVKKY
FPNWKKVNLTYMGEIFKK

>gi|341590491|gb|EGS33732.1|/1-1348
MKSEKKYYIGLDVGTNSVGWAVTDEFYNILRAKGKDLWGVRLFEKAKTAADTRTFRSGRRRNDRKGVRLQILREIFEDEIKKVDKDFYDR
LDESKFWAEDKKVSGKYSLFNDKNFSDKQYFEKFPTIFHLRKYLMEEHGKVDIRYYFLAINQMMKRRGHFLIDGQISHVTDDKPLKQQLIL
LINDLLKIELEEEFMDSIFEILADVNEKKTEKKNILKELINAQDFNKQEGKILYSIFESIVTGKAKIKNIISDEDILEKIKEDNKEDFVLTGDSYEE
NLQYFEEVLQENITLFNTLKSAYDFVILQSILKGKSTLSDAQVERYDEHKKDLEILKKVIKKYDEDGKLFKQVFKEDNGNGYVSYIGYYLNK
NKKITAKKKISNIEFTKYVKGILEKQCDCEDEDVKYLLGKIEQENFLLKQISSINSVIPHQIHLFELDKILENLAKYYPSFNNKKEEFTKIEKIRK
TFTFRIPYYVGPLNDYHKNNGGNAWIFRNKGEKIRPWNFEKIVDLHKSEEEFIKRMLNQCTYLPEETVLPKSSILYSEYMVLNELNNLRING
KPIDTDVKLKLIEELFKKKTKVTLKSIRDYMIRNNFADKEDFDNSEKNLEIVSNMKSYIDFNNILGDKFELEMVEDLIEKITIHTGNKKLLKKYIE
ETYPDLSSSQIQKIINLKYKDWGRLSRKLLDGITGTNIETGEIDTVINFLRNSSDNLMQIIGSQNYSFNNDIDKLRKKYIPQEISYEVVENLYVS
PSVKKMIWQVIRITEEITKVMGYDPDKIFIEMAKSEEGKKTTISRKNKLLALYKAIKKDERDSQYEKLLTGLNKLDDSDLRSRKLYLYYTQMG
RDMYTGEKIDLDKLFDSTHYDKDHIIPQSMKKDDSIINNLVLVNKNANQTTKGNIYPVPSSIRNNPKIYNYWKYLMEKEFISKEKYNRLIRNT
PLTNEELGGFINRQLVETRQSTKAIKELFEKFYKKSKIIPVKAILASDLRKDMNTLKSREVNDLHHAHDAFLNIVAGDVWNREFTSNPINYVK
ENREGDKVKYSLSKYFTRPRKSKGKVIWTPEKGRKLIVDTLNKPSVLISNESHVKKGELFNATIAGKKDYKKGKIYLPLKKDDRLQDVSKY
GGYKAINGAFFFLVEHTKSKKRIRSIELFPLHLLSKFYEDKNTVLDYAINVLQLQDPKIIIDKINYRTEIIIDNFSYLISTKSNDGSITVKPNEQMY
WRVDEISNLKKIENKYKKDAILTEEDRKIMESYIDKIYQQFKAGKYKNRRTTDTIIEKYEIIDLDTLDNKQLYQLLVAFISLSYKTSNNAVDFTVI
GLGTECGKPRITNLPDNTYLVYKSITGIYEKRIRIK

>gi|227824983|ref|ZP_03989815.1|/1-1358
MGKMYYLGLDIGTNSVGYAVTDPSYHLLKFKGEPMWGAHVFAAGNQSAERRSFRTSRRRLDRRQQRVKLVQEIFAPVISPIDPRFFIRLH
ESALWRDDVAETDKHIFFNDPTYTDKEYYSDYPTIHHLIVDLMESSEKHDPRLVYLAVAWLVAHRGHFLNEVDKDNIGDVLSFDAFYPEFL
AFLSDNGVSPWVCESKALQATLLSRNSVNDKYKALKSLIFGSQKPEDNFDANISEDGLIQLLAGKKVKVNKLFPQESNDASFTLNDKEDAI
EEILGTLTPDECEWIAHIRRLFDWAIMKHALKDGRTISESKVKLYEQHHHDLTQLKYFVKTYLAKEYDDIFRNVDSETTKNYVAYSYHVKEV
KGTLPKNKATQEEFCKYVLGKVKNIECSEADKVDFDEMIQRLTDNSFMPKQVSGENRVIPYQLYYYELKTILNKAASYLPFLTQCGKDAIS
NQDKLLSIMTFRIPYFVGPLRKDNSEHAWLERKAGKIYPWNFNDKVDLDKSEEAFIRRMTNTCTYYPGEDVLPLDSLIYEKFMILNEINNIRI
DGYPISVDVKQQVFGLFEKKRRVTVKDIQNLLLSLGALDKHGKLTGIDTTIHSNYNTYHHFKSLMERGVLTRDDVERIVERMTYSDDTKRV
RLWLNNNYGTLTADDVKHISRLRKHDFGRLSKMFLTGLKGVHKETGERASILDFMWNTNDNLMQLLSECYTFSDEITKLQEAYYAKAQLS
LNDFLDSMYISNAVKRPIYRTLAVVNDIRKACGTAPKRIFIEMARDGESKKKRSVTRREQIKNLYRSIRKDFQQEVDFLEKILENKSDGQLQ
SDALYLYFAQLGRDMYTGDPIKLEHIKDQSFYNIDHIYPQSMVKDDSLDNKVLVQSEINGEKSSRYPLDAAIRNKMKPLWDAYYNHGLISL
KKYQRLTRSTPFTDDEKWDFINRQLVETRQSTKALAILLKRKFPDTEIVYSKAGLSSDFRHEFGLVKSRNINDLHHAKDAFLAIVTGNVYHE
RFNRRWFMVNQPYSVKTKTLFTHSIKNGNFVAWNGEEDLGRIVKMLKQNKNTIHFTRFSFDRKEGLFDIQPLKASTGLVPRKAGLDVVKY
GGYDKSTAAYYLLVRFTLEDKKTQHKLMMIPVEGLYKARIDHDKEFLTDYAQTTISEILQKDKQKVINIMFPMGTRHIKLNSMISIDGFYLSIG
GKSSKGKSVLCHAMVPLIVPHKIECYIKAMESFARKFKENNKLRIVEKFDKITVEDNLNLYELFLQKLQHNPYNKFFSTQFDVLTNGRSTFT
KLSPEEQVQTLLNILSIFKTCRSSGCDLKSINGSAQAARIMISADLTGLSKKYSDIRLVEQSASGLFVSKSQNLLEYL

>gi|34762592|ref|ZP_00143587.1|/1-1374
MKKQKFSDYYLGFDIGTNSVGWCVTDLDYNVLRFNKKDMWGSRLFDEAKTAAERRVQRNSRRRLKRRKWRLNLLEEIFSDEIMKIDSNF
FRRLKESSLWLEDKNSKEKFTLFNDDNYKDYDFYKQYPTIFHLRDELIKNPEKKDIRLIYLALHSIFKSRGHFLFEGQNLKEIKNFETLYNNLI
SFLEDNGINKSIDKDNIEKLEKIICDSGKGLKDKEKEFKGIFNSDKQLVAIFKLSVGSSVSLNDLFDTDEYKKEEVEKEKISFREQIYEDDKPI
YYSILGEKIELLDIAKSFYDFMVLNNILSDSNYISEAKVKLYEEHKKDLKNLKYIIRKYNKENYDKLFKDKNENNYPAYIGLNKEKDKKEVVEK
SRLKIDDLIKVIKGYLPKPERIEEKDKTIFNEILNKIELKTILPKQRISDNGTLPYQIHEVELEKILENQSKYYDFLNYEENGVSTKDKLLKTFKF
RIPYYVGPLNSYHKDKGGNSWIVRKEEGKILPWNFEQKVDIEKSAEEFIKRMTNKCTYLNGEDVIPKDSFLYSEYIILNELNKVQVNDEFLN
EENKRKIIDELFKENKKVSEKKFKEYLLVNQIANRTVELKGIKDSFNSNYVSYIKFKDIFGEKLNLDIYKEISEKSILWKCLYGDDKKIFEKKIK
NEYGDILNKDEIKKINSFKFNTWGRLSEKLLTGIEFINLETGECYSSVMEALRRTNYNLMELLSSKFTLQESIDNENKEMNEVSYRDLIEESY
VSPSLKRAILQTLKIYEEIKKITGRVPKKVFIEMARGGDESMKNKKIPARQEQLKKLYDSCGNDIANFSIDIKEMKNSLSSYDNNSLRQKKLY
LYYLQFGKCMYTGREIDLDRLLQNNDTYDIDHIYPRSKVIKDDSFDNLVLVLKNENAEKSNEYPVKKEIQEKMKSFWRFLKEKNFISDEKY
KRLTGKDDFELRGFMARQLVNVRQTTKEVGKILQQIEPEIKIVYSKAEIASSFREMFDFIKVRELNDTHHAKDAYLNIVAGNVYNTKFTEKP
YRYLQEIKENYDVKKIYNYDIKNAWDKENSLEIVKKNMEKNTVNITRFIKEEKGELFNLNPIKKGETSNEIISIKPKLYDGKDNKLNEKYGYYT
SLKAAYFIYVEHEKKNKKVKTFERITRIDSTLIKNEKNLIKYLVSQKKLLNPKIIKKIYKEQTLIIDSYPYTFTGVDSNKKVELKNKKQLYLEKKY
EQILKNALKFVEDNQGETEENYKFIYLKKRNNNEKNETIDAVKERYNIEFNEMYDKFLEKLSSKDYKNYINNKLYTNFLNSKEKFKKLKLWE
KSLILREFLKIFNKNTYGKYEIKDSQTKEKLFSFPEDTGRIRLGQSSLGNNKELLEESVTGLFVKKIKL

FIGURE 12 H.

>gi|227501312|ref|ZP_03931361.1|/1-1361
MGNYYLGLDIGTNSVGWAVTDENYNIERFKKKNMWGIRLFDEAKTAESRRIHRQARRMRQRKIQRLNLLMELFDVEISKVDESFFIRKKE
SDLHYEDKSQDSRYALFNDKEFNDKDYHNKYPTIFHLIMDLIENDEKKDIRLLYLACHYLIKNRGHFIFEGQKFDVNSSLNNNINKLKSHLD
QFYNICLDFDNDKLIEILVNQNLNISEKSKKIKELLRKTNLEVGIANLIAGGSIKLSNMFDDDELKNLDYNSLSFKNSNFDEGRELFEEVLEDR
FELVEILKDIHDASVLETMLSSASISPDGGKYISSSMINSYNKHKRDLKLLKKIIKKYKKSEYDNIFRDENIKNNYLSYTRTGINKNKKVQAED
NTDSDGFYKYIVKILKDIKENDSDEDIKYILSEIDKKNFVPKQRTKENSVIPYQLREIELRKIIDNQSKYHKFLNEVDEKGISVKEKIISLLTFRIP
YYIGPLNPSYNEEKNFTTSWVKRLAKGKVTPWNFEDMVDIDASREAFIENKLNKCSFLLDETVLPKDSLLYQEYMVLNELNNLKINGQSIN
TDLKKSIYEELFKKYKKVTRKRLISYIKEKENMSGDIVITGIDGDFKQGLKSYIAFKDIIGEKIDDDIYRKEVEKIIRDISLYGDDKSFIQKKIKNE
NKLDLSDDQISKIAKLKFKDWGRFSKKLLVGIKASDKNTGEIDNIIGFMRNYNYNIMQLMANVFTFKDEIEKLNSKYYTDEKINIDYLDDLYLS
SPAKRMLWQSFQIVDEIKNIKNSDPKKIFIEMARSKEDKPTRKMSRKSRLEKLYTDSKKQFIKLFGPDEYDKLNNEIKNKSESDFRWDNLY
LYYTQFGRCMYSLDPIDLEELSNKNLYDQDHIYPKSKIYDDSIENRVLVKKQLNLDKSNEYPIPDRVLNKDKNKLLAFWKMLNEKELIGDKK
YARLIRKSPFTDEELGDFIARQLVETRQATKETANMLKRLCPDSKIVYVKAKNVSDFRRKFDILKCRSINDLHHAHDAYLNIVVGNVFDTKF
TSNPLNFIKNNSKNRSYNLEKIYDYDVIRNEKIAWLAERKEENQDGSIATVLKNLSKNDIRVTRRTYIKCEDLFDENIVRKGNGKFPIKENNN
KSDTNKYGGYSGIKNAYFVLLEDEKHNRILEAIPSFIYKKEKNGEKGSVINYLQNILGIKSAKILRDKIKIGSLLKIDKFPYNLAGKSGNGILISG
GVQLILDKNTSKYIKKIENYKYKLDENKDYQINIFDKLSKEENIKLFDLFVDKFENSIYKNRVNIKLKEIKLARPKFEHLSIEEQIKTLTQMLILFG
RENNAVNLEYIGLPKNTGKTKINKKINNKDIKIINQSITGLYETEEDLANI

>gi|153855454|ref|ZP_01996585.1|/1-1340
MEQEYYLGLDMGTGSVGWAVTDSEYHVLRKHGKALWGVRLFESASTAEERRMFRTSRRRLDRRNWRIEILQEIFAEEISKKDPGFFLRM
KESKYYPEDKRDINGNCPELPYALFVDDDFTDKDYHKKFPTIYHLRKMLMNTEETPDIRLVYLAIHHMMKHRGHFLLSGDINEIKEFGTTFS
KLLENIKNEELDWNLELGKEEYAVVESILKDNMLNRSTKKTRLIKALKAKSICEKAVLNLLAGGTVKLSDIFGLEELNETERPKISFADNGYD
DYIGEVENELGEQFYIIETAKAVYDWAVLVEILGKYTSISEAKVATYEKHKSDLQFLKKIVRKYLTKEEYKDIFVSTSDKLKNYSAYIGMTKIN
GKKVDLQSKRCSKEEFYDFIKKNVLKKLEGQPEYEYLKEELERETFLPKQVNRDNGVIPYQIHLYELKKILGNLRDKIDLIKENEDKLVQLFE
FRIPYYVGPLNKIDDGKEGKFTWAVRKSNEKIYPWNFENVVDIEASAEKFIRRMTNKCTYLMGEDVLPKDSLLYSKYMVLNELNNVKLDG
EKLSVELKQRLYTDVFCKYRKVTVKKIKNYLKCEGIISGNVEITGIDGDFKASLTAYHDFKEILTGTELAKKDKENIITNIVLFGDDKKLLKKRL
NRLYPQITPNQLKKICALSYTGWGRFSKKFLEEITAPDPETGEVWNIITALWESNNNLMQLLSNEYRFMEEVETYNMGKQTKTLSYETVE
NMYVSPSVKRQIWQTLKIVKELEKVMKESPKRVFIEMAREKQESKRTESRKKQLIDLYKACKNEEKDWVKELGDQEEQKLRSDKLYLYYT
QKGRCMYSGEVIELKDLWDNTKYDIDHIYPQSKTMDDSLNNRVLVKKKYNATKSDKYPLNENIRHERKGFWKSLLDGGFISKEKYERLIR
NTELSPEELAGFIERQIVETRQSTKAVAEILKQVFPESEIVYVKAGTVSRFRKDFELLKVREVNDLHHAKDAYLNIVVGNSYYVKFTKNASW
FIKENPGRTYNLKKMFTSGWNIERNGEVAWEVGKKGTIVTVKQIMNKNNILVTRQVHEAKGGLFDQQIMKKGKGQIAIKETDERLASIEKY
GGYNKAAGAYFMLVESKDKKGKTIRTIEFIPLYLKNKIESDESIALNFLEKGRGLKEPKILLKKIKIDTLFDVDGFKMWLSGRTGDRLLFKCA
NQLILDEKIIVTMKKIVKFIQRRQENRELKLSDKDGIDNEVLMEIYNTFVDKLENTVYRIRLSEQAKTLIDKQKEFERLSLEDKSSTLFEILHIF
QCQSSAANLKMIGGPGKAGILVMNNNISKCNKISIINQSPTGIFENEIDLLKI

>gi|241889924|ref|ZP_04777222.1|/1-1392
MANKTWYLGLDIGSASVGWAATDTGYKIIRKNKKRLWGVRLFEEAKTAEDRRGYRSSRRRLARRKWRLNLLEELFTSEVKKVDENFFLR
LKESQYHYEDKTHKVPYAIFNDKDYTDKDYYKDYPTIYHLRSKLMTEENPDIRKVFLAIHHIIKNRGHFLLQGQSFKDGNLNNLIKELLELDIL
HVGFEVTEEVVDKIADISLEKKTSKDKLNDIKELYPKEKQLLEVFRLIFGGKTSLDKLFAIDEYKELDAAIKSVSFKEKIYEEVRHDYEQVLSN
YIELLDLAKLVYDSIILSDIKKEGKTLSESKVELFEKHRDDLLKLKNLVKNDSKLSEDKKVELYALIFKEDKDKGTNYVNYIRKSEEGKGCNY
EDFKKFLVKELAKLEESSVKTEIIKDLELEQFLPLQRTKDNSVVPYQIHKEELVKILDNAARYHSFLNEKDESGYNVREKVIQLLEFRIPYYV
GPLNSSKKAKEGGFAWSVRNTGYENTPVTPWNYSKVIDESASAEKFITNLTNKCTYLKGEDVLPKSSLLYSEFALLNELNALKYDGNRISL
EARNTIIEKLFKQQGKKVTKTSIKNLLKAEGYIDGKGEITGIDITVKNDLKSYRDFKKILGNKFNSEHVENIILWITLYGESRKLIKAKIESVYGE
VYSKDEITKMSRLVYKDWGRFSRKLLTELVSKKLYNEETGECLNIIGAMRQNNILFMELLADRFDYSKQIVEFNKELQEDVTEIAPEILDDLY
VSPAVKRSIWQTVRIVEELKKIIGCAPAKIFVETTRSNQEKKKPTDSRKKQLELAYKAVKKDVKELEKEIGTINFDVLNDRLSSVEPSKLKAK
KLYLYYTQLGRCMYSEEPINISELFDNNKYDIDHIYPQSKIKDDSFTNTVLVKRESNAAKTDRYPLGSDIQTPSNRRFWKFLKEKGLISDEK
YNRLVRTEEFSDEELSGFIARQLVETSQSVKAVASILGELNPETSICYSKAENVSSFRQNFGKIKKGDRDSKNNDQLIKVREINDYHHAKD
AYLNVVVGNVYDTKFTRNVFNFIKDKKDGRKYSLNKVFYENVSDSKTVAWEMDKTIHQVEKVMNNNNIMVTRRTSEQKGGLFDATVYKA
KVAAKAKDGVYYPLKTSNSVVKDVAKYGGYTSIGIAYYSIFEYTIVNKKGEEKITRIIPIPIYISQNIKDSTALIEFGKSQIFCRAYEKIKDFKLLY
RKLCIGSKININNYQYYIGGKSDERFYYDSAIQVTLDKKSEKYLKILSKYQNWKKENKDGELWENITKENNIELYNKLVEKMNSEIFTKKKSN
KYNELNSEEIRNNFIKISVEEQAKVLLEILNLLTNKKSIFDLKSIDIKAARGKISFNLTSLIQFSIIEQSITGFYEKEITIIGDKGNDMENNNS

FIGURE 12 I.

>gi|282849530|ref|ZP_06258914.1|/1-1398
METQASNQLISSHLKGYPIKDYFVGLDIGTSSVGWAVTNKAYELLKFRSHKMWGSRLFDEGESAVARRGFRSMRRRLERRKLRLKLLEE
LFADAMAQVDPTFFMRLRESKYHYEDKTTGHSSKHILFIDKNYNDQDYFKEYPTVYHLRSELMKSGTDDIRKLFLAVHHILKYRGNFLYEG
ATFDSNASTLDDVIKQALENITFNCFDCNSAISSIGQILMEAGKTKSDKAKAIEHLVDTYIATDTVDTSSKTQKDQVKEDKKRLKAFANLVLG
LNASLIDLFGSVEELEEDLKKLQITGDTYDDKRDELAKAWSDEIYIIDDCKSVYDAIILLSIKEPGLTISESKVKAFNKHKDDLAILKSLLKSDR
SIYNTMFKVDEKGLHNYVHYIKQGRTEETSCNREDFYKYTKKIVEGLSDSKDKEYILSQIELQILLPLQRIKDNGVIPYQLHLEELKAILAKCG
PKFPFLNEVADGFSVAEKLIKMLEFRIPYYVGPLNTHHNVDNGGFAWAVRKASGRVTPWNFDDKIDREKSAAAFIKNLTNKCTYLLGEDV
LPKSSLLYSEFMLLNELNNVRIDGKPLEKVVKEHLIEAVFKQDHKKMTKNRIEQFLKDNGYISETHKHEITGLDGEIKNDLASYRDMVRILG
DGFDRSMAEEIITDITIFGESKKMLRETLRKKFASCLDDEAIKKLTKLRYRDWGRLSQKLLNGIEGCDKAGDGTPETIIILMRNFSYNLMELL
GDKFSFMERIQEINAKLTEGQIVNPHDIIDDLALSPAVKRAVWQALRIVDEVAHIKKALPARIFVEVTRSNKNEKKKKDSRQKRLSDLYAAIK
KDDVLLNGLNNEIFGELKSSLAKYDDAALRSKKLYLYYTQMGRCAYTGEIIELSLLNTDNYDIDHIYPRSLTKDDSFDNLVLCKRTANAQKS
DAYPISEEIQKTQKPFWTFLKQQGLISERKYERLTRITPLTADDLSGFIARQLVETNQSVKAATTLLRRLYPGVDVVFVKAENVTDFRHDNN
FIKVRSLNHHHHAKDAYLNIVVGNVYHERFTRNFRAFFKKNGANRTYNLAKMFNYDVNCTNAKDGKAWDVKTSMDTVKKMMDSNDVRV
TKRLLEQTGALADATIYKATVAGKAKDGAYIGMKTKSSVFADVSKYGGMTKIKNAYSIIVQYTGKKGEVIKEIVPLPIYLTNRNTTDQDLINY
VASIIPQAKDISIIYGKLCINQLVKVNGFYYYLGGKTNSKFCIDNAIQVIVSNEWIPYLKVLEKFNNMRKDNKDLKANVVSTRALDNKHTIEVRI
VEEKNIEFFDYLVSKLKMPIYQKMKGNKAAELSEKGYGLFKKMSLEEQSIHLIELLNLLTNQKTTFEVKPLGITASRSTVGSKISNQDEFKVI
NESITGLYSNEVTIV

>gi|254829042|ref|ZP_05233729.1|/1-1334
MKKPYTIGLDIGTNSVGWAVLTDQYDLVKRKMKISGDSEKKQIKKNFWGVRLFEKGETAAKRRMSRTARRRIERRRNRISYLQEIFAIQM
NEVDDNFFNRLKESFYAESDKKYNRHPFFGTVEEEVAYYKDFPTIYHLRKELIDSQKKADLRLVYLALAHIIKYRGHFLIEGALDTKNTSIDE
MFKQFLQIYNQVFANDIEEASLKKTEKNQEVAQILAEKFTRKDKLDKILSLYPGEKTTGVFAQFVNIIVGSTGKFKKHFNLHEKKDINCAEDT
YDTDLESLLAIIGDEFAEVFVAAKNAYNAVVLSNIITVTDSTTRAKLSASLIERFENHKEDLKKMKRFVRTYLPEKYDEIFDDTEKHGYAGYIS
GKTKQADFYKYMKATLEKIEGADYFIAKIEEENFLRKQRTFDNGVIPHQLHLEELEAILHQQAKYYPFLREDYEKIKSLVTFRIPYFVGPLAK
GQSEFAWLTRKADGEIRPWNIEEKVDFGKSAVDFIEKMTNKDTYLPKENVLPKHSLCYQKYMVYNELTKVRYIDDQGKTNYFSGQEKQQ
IFNDLFKQKRKVKKKDLELFLRNINQIESPTIEGLEDSFNASYATYHDLLKVGMKQEILDNPLNTEMLEDIVKILTVFEDKRMIKEQLQQFSD
VLDGTVLKKLERRHYTGWGRLSAKLLVGIRDKQSHLTILDYLMNDDGLNRNLMQLINDSNLSFKSIIEKEQVSTTDKDLQSIVAELAGSPAI
KKGILQSLKIVDELVSVMGYPPQTIVVEMARENQTTNKGKNNSKPRYKSLEKAIKEFGSQILKEHPTDNQELKNNRLYLYYLQNGKDMYT
GQELDIHNLSNYDIDHIVPQSFITDNSVDNLVLTSSAGNREKGDNVPPLEIVQKRKIFWEKLYQGNLMSKRKFDYLTKAERGGLTEADKAR
FIHRQLVETRQITKNVANILHQRFNYKTDGNKDTMETVRIVTLKSALVSQFRKQFQFYKVREVNDYHHAHDAYLNGVVANTLLKVYPQLEP
EFVYGEYHQFGWFKANKATAKKQFYTNIMLFFAQKDRIIDENGEILWDKRYLETVKKVLGYRQMNIVKKTEIQKGEFSNVTPNPKGKSNK
LIPRKKDWDPIKYGGFDGSKMAYAIIIEYEKQKRKVRIEKKLIQINIMEREAFEKDEKTFLEEKGYHQPKVLIKVPKYTLYECKNGRRRMLGS
ANEAHKGNQMLLPNHLMALLYHAEKYEAIDGESLAYIEVHRALFDELLAYISEFARKYTLSNDRLDEINMLYERNKDGDVKSIAESFVSLKK
FNAFGVHQDFSFFGTKIERKRDRKLNELLNSTIIYQSITGLYESRKRLDN

>gi|256617555|ref|ZP_05474401.1|/1-1337
MKKDYVIGLDIGTNSVGWAVMTEDYQLVKKKMPIYGNTEKKKIKKNFWGVRLFEEGHTAEDRRLKRTARRRISRRRNRLRYLQAFFEEA
MTALDENFFARLQESFLVPEDKKWHRHPIFAKLEDEVAYHETYPTIYHLRKKLADSSEQADLRLIYLALAHIVKYRGHFLIEGKLSTENISVK
EQFQQFMIIYNQTFVNGESRLVSAPLPESVLIEEELTEKASRTKKSEKVLQQFPQEKANGLFGQFLKLMVGNKADFKKVFGLEEEAKITYA
SESYEEDLEGILAKVGDEYSDVFLAAKNVYDAVELSTILADSDKKSHAKLSSSMIVRFTEHQEDLKKFKRFIRENCPDEYDNLFKNEQKDG
YAGYIAHAGKVSQLKFYQYVKKIIQDIAGAEYFLEKIAQENFLRKQRTFDNGVIPHQIHLAELQAIIHRQAAYYPFLKENQEKIEQLVTFRIPY
YIGPLSKGDASTFAWLKRQSEEPIRPWNLQETVDLDQSATAFIERMTNFDTYLSSEKVLPKHSLLYEKFMVFNELTKISYTDDRGIKANFS
GKEKEKIFDYLFKTRRKVKKKDIIQFYRNEYNTEIVTLSGLEEDQFNASFSTYQDLLKCGLTRAELDHPDNAEKLEDIIKILTIFEDRQRIRTQ
LSTFKGQFSAEVLKKLERKHYTGWGRLSKKLINGIYDKESGKTILGYLIKDDGVSKHYNRNFMQLINDSQLSFKNAIQKAQSSEHEETLSE
TVNELAGSPAIKKGIYQSLKIVDELVAIMGYAPKRIVVEMARENQTTSTGKRRSIQRLKIVEKAMAEIGSNLLKEQPTTNEQLRDTRLFLYYM
QNGKDMYTGDELSLHRLSHYDIDHIIPQSFMKDDSLDNLVLVGSTENRGKSDDVPSKEVVKDMKAYWEKLYAAGLISQRKFQRLTKGEQ
GGLTLEDKAHFIQRQLVETRQITKNVAGILDQRYNANSKEKKVQIITLKASLTSQFRSIFGLYKVREVNDYHHGQDAYLNCVVATTLLKVYP
NLAPEFVYGEYPKFQAFKENKATAKAIIYTNLLRFFTEDEPRFTKDGEILWSNSYLKTIKKELNYHQMNIVKKVEVQKGGFSKESIKPKGPS
NKLIPVKNGLDPQKYGGFDSPVVAYTVLFTHEKGKKPLIKQEILGITIMEKTRFEQNPILFLEEKGFLRPRVLMKLPKYTLYEFPEGRRRLLA
SAKEAQKGNQMVLPEHLLTLLYHAKQCLLPNQSESLAYVEQHQPEFQEILERVVDFAEVHTLAKSKVQQIVKLFEANQTADVKEIAASFIQ
LMQFNAMGAPSTFKFFQKDIERARYTSIKEIFDATIIYQSTTGLYETRRKVVD

FIGURE 12 J.

>gi|257893735|ref|ZP_05673388.1|/1-1340
MKKEYTIGLDIGTNSVGWSVLTDDYRLVSKKMKVAGNTEKSSTKKNFWGVRLFDEGQTAEARRSKRTARRRLARRRQRILELQKIFAPEI
LKIDEHFFARLNESFLVPDEKKQSRHPVFATIKQEKSYHQTYPTIYHLRQALADSSEKADIRLVYLAMAHLLKYRGHFLIEGELNTENSSVT
ETFRQFLSTYNQQFSEAGDKQTEKLDEAVDCSFVFTEKMSKTKKAETLLKYFPHEKSNGYLSQFIKLMVGNQGNFKNVFGLEEEAKLQF
SKETYEEDLEELLEKIGDDYIDLFVQAKNVYDAVLLSEILSDSTKNTRAKLSAGMIRRYDAHKEDLVLLKRFVKENLPKKYRAFFGDNSVNG
YAGYIEGHATQEAFYKFVKKELTGIRGSEVFLTKIEQENFLRKQRTFDNGVIPHQIHLSELRAIIANQKKHYPFLKEEQEKLESLLTFKIPYYV
GPLAKKQENSPFAWLIRKSEEKIKPWNLPEIVDMEGSAVRFIERMINTDMYMPHNKVLPKNSLLYQKFSIYNELTKVRYQDERGQMNYFS
SIEKKEIFHELFEKNRKVTKKDLQEFLYLKYDIKHAELSGIEKAFNASYTTYHDFLTMSENKREMKQWLEDPELASMFEEIIKTLTVFEDRE
MIKTRLSHHEATLGKHIIKKLTKKHYTGWGRLSKELIQGIRDKQSNKTILDYLINDDDFPHHRNRNFMQLINDDSLSFKKEIKKAQMITDTEN
LEEIVKELTGSPAIKKGILQSLKIVDEIVGIMGYEPANIVVEMARENQTTGRGLKSSRPRLKALEESLKDFGSQLLKEYPTDNSSLQKDRLYL
YYLQNGRDMYTGAPLDIHRLSDYDIDHIIPRSFTTDNSIDNKVLVSSKENRLKKDDVPSEKVVKKMRSFWYDLYSSKLISKRKLDNLTKIKL
TEEDKAGFIKRQLVETRQITKHVAGILHHRFNKAEDTNDPIRKVRIITLKSALVSQFRNRFGIYKVREINEYHHAHDAYLNGVVALALLKKYP
QLAPEFVYGEYLKFNAHKANKATVKKEFYSNIMKFFESDTPVCDENGEIFWDKSKSIAQVKKVINHHHMNIVKKTEIQKGGFSKETVEPKK
DSSKLLPRKNNWDPAKYGGLGSPNVAYTVAFTYEKGKARKRTNALEGITIMEREAFEQSPVLFLKNKGYEQAEIEMKLPKYALFELENGR
KRMVASNKEAQKANSFLLPEHLVTLLYHAKQYDEISHKESFDYVNEHHKEFSEVFARVLEFAGKYTLAEKNIEKLEKIYKENQTDDLAKLA
SSFVNLMQFNAMGAPADFKFFDVTIPRKRYTSLTEIWQSTIIHQSITGLYETRIRMGK

>gi|24379809|ref|NP_721764.1|/1-1345
MKKPYSIGLDIGTNSVGWAVVTDDYKVPAKKMKVLGNTDKSHIEKNLLGALLFDSGNTAEDRRLKRTARRRYTRRRNRILYLQEIFSEEM
GKVDDSFFHRLEDSFLVTEDKRGERHPIFGNLEEEVKYHENFPTIYHLRQYLADNPEKVDLRLVYLALAHIIKFRGHFLIEGKFDTRNNDVQ
RLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTYEE
ELEVLLAQIGDNYAELFLSAKKLYDSILLSGILTVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKDGYAGYIDG
KTNQEAFYKYLKGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFLADNQDRIEKLLTFRIPYYVGPLARG
KSDFAWLSRKSADKITPWNFDEIVDKESSAEAFINRMTNYDLYLPNQKVLPKHSLLYEKFTVYNELTKVKYKTEQGKTAFFDANMKQEIFD
GVFKVYRKVTKDKLMDFLEKEFDEFRIVDLTGLDKENKVFNASYGTYHDLCKILDKDFLDNSKNEKILEDIVLTLTLFEDREMIRKRLENYS
DLLTKEQVKKLERRHYTGWGRLSAELIHGIRNKESRKTILDYLIDDGNSNRNFMQLINDDALSFKEEIAKAQVIGETDNLNQVVSDIAGSPAI
KKGILQSLKIVDELVKIMGHQPENIVVEMARENQFTNQGRRNSQQRLKGLTDSIKEFGSQILKEHPVENSQLQNDRLFLYYLQNGRDMYT
GEELDIDYLSQYDIDHIIPQAFIKDNSIDNRVLTSSKENRGKSDDVPSKDVVRKMKSYWSKLLSAKLITQRKFDNLTKAERGGLTDDDKAGF
IKRQLVETRQITKHVARILDERFNTETDENNKKIRQVKIVTLKSNLVSNFRKEFELYKVREINDYHHAHDAYLNAVIGKALLGVYPQLEPEFV
YGDYPHFHGHKENKATAKKFFYSNIMNFFKKDDVRTDKNGEIIWKKDEHISNIKKVLSYPQVNIVKKVEEQTGGFSKESILPKGNSDKLIPR
KTKKFYWDTKKYGGFDSPIVAYSILVIADIEKGKSKKLKTVKALVGVTIMEKMTFERDPVAFLERKGYRNVQEENIIKLPKYSLFKLENGRK
RLLASARELQKGNEIVLPNHLGTLLYHAKNIHKVDEPKHLDYVDKHKDEFKELLDVVSNFSKKYTLAEGNLEKIKELYAQNNGEDLKELAS
SFINLLTFTAIGAPATFKFFDKNIDRKRYTSTTEILNATLIHQSITGLYETRIDLNKLGGD

>gi|195978435|ref|YP_002123679.1|/1-1348
MKKPYTIALDIGTNSVGWVVVTDDYRVPTKKMKVLGNTERKTIKKNLIGALLFDSGDTAEGTRLKRTARRRYTRRKNRLRYLKEIFTEEMA
KVDDGFFQRLEDSFYVLEDKEGNKHPIFANLADEVAYHKKYPTIYHLRKELVDNPQKADLRLIYLAVAHIIKFRGHFLIEGTLSSKNNNLQK
SFDHLVDTYNLLFEEQRLLTEGINAKELLSAALSKSKRLENLISLIPGQKKTGIFGNIIALSLGLTPNFKANFGLSKDVKLQLAKDTYADDLDS
LLAQIGDQYADLFLAAKNLSDAILLSDILTESDEITRAPLSASMVKRYREHHKDLVTLKTLIKDQLPEKYQEIFLDKTKNGYAGYIEGQVSQE
EFYKYLKPILARLDGSEPLLLKIDREDFLRKQRTFDNGSIPHQIHLEELHAILRRQEVFYPFLKDNRKKIESLLTFRIPYYVGPLARGHSRFA
WVKRKFDGAIRPWNFEEIVDEEASAQIFIEKMTKNDLYLPNEKVLPKHSLLYETFTVYNELTKVKYATEGMTRPQFLSADQKQAIVDLLFKT
NRKVTVKQLKENYFKKIECWDSVEITGVEDSFNASLGTYHDLLKIIQDKDFLDNPDNQKIIEDIILTLTLFEDKKMISKRLDQYAHLFDKVVLN
KLERHHYTGWGRLSGKLINGIRDKQSGKTILDFLKADGFANRNFMQLIHDSELSFIDEIAKAQVIGKTEYSKDLVGNLAGSPAIKKGISQTIKI
VDELVKIMGYLPQQIVIEMARENQTTAQGIKNARQRMRKLEETAKKLGSNILKEHPVDNSQLQNDKRYLYYLQNGKDMYTGDDLDIDYLS
SYDIDHIIPQSFIKNNSIDNKVLTSQGANRGKLDNVPSEAIVRKMKGYWQSLLRAGAISKQKFDNLTKAERGGLTQVDKAGFIQRQLVETR
QITKHVAQILDSRFNTEFDDHNKRIRKVHIITLKSKLVSDFRKEFGLYKIRDINHYHHAHDAYLNAVVAKAILGKYPQLAPEFVYGDYPKYNS
FKERQKATQKMLFYSNILKFFKDQESLHVNSDGEEIWNANKHLPIIKNVLSIPQVNIVKKTEVQTGGFYKESILSKGNSDKLIPRKNNWDTR
KYGGFDSPTVAYSVLVIAKMEKGKAKVLKPVKEMVGITIMERTAFEENPVVFLEARGYREIQEHLIIKLPKYSLFELENGRRRLLASASELQ
KGNELFLPVDYMTFLYLAAHYHELTGSSEDVLRKKYFVDRHLHYFDDIIQMINDFAERHILASSNLEKINHTYHNNSDLPVNERAENIINVFT
FVALGAPAAFKFFDATIDRKRYTSTKEVLNATLIHQSVTGLYETRIDLSQLGEN

>gi|191639137|ref|YP_001988303.1|/1-1361
MTKLGKPYGIGVDIGSNSIGFAAVDENSHLIRLKGKTVIGARLFEEGKAAADRRASRTTRRRLSRNRWRLSFLRDFFESHITPTGPNFFMR
QKYSEISPKDKARYKYEKRLFNDRTDAEFYQQYPTMYHLRNRLLTDPSKADVREIYFAIHHILKSRGHFLTSGYAKDFNTNKVALNEIFPAL
QDAYAQVYPDLDITFDENKVNEFKTVLLNEKATPSDTQRALVNLLLAEDGDKDILKQQKQVLTEFAKAVVGLKTKLNVALGTEVDSSEATV
WNFSLGQLDDKWAGIESAMTDEGTEILDQIRDLYRARLLNGIVPAGKTLSQAKVDDYAQHHEDLELFKDYLKQLEDDGTAKAIRQLYDRYI
DGDDAKPFVREDFVKALTKEVTAHPNAKSPELLERLAQPDFMLKQRNKANGAIPVQMQQRELDQIIKNQSVYYDWLAAPNPVEKHRKSM
PYQLDELLNFRIPYYVGPLVTAKEQKAAQGGVFAWMVRKDPDGNITPYNFDEKVDREASANTFIQRMKTTDTYLIGEDVLPKQSLLYQRY
EVLNELNNVRVNNEKLSIEQKQQVIRELFERHNSVTIKQFAENLRAHGDYAHIPEIRGLADEKRFLSSLSTYRQLKSLLPDAIDDPAKQADIE
NIIAWSTVFEDAAIFKTKLKEINWLDSQTITKLSNIRYRGWGQFSHKFLNGLTLGNGHTIIQELFLSNNNLMQILTDETLQKKMTELNADKLK
AANINDAIDDAYTSPSNKKALRQVLRVIEDIKRAADGQDPSWLYVETADGGTPGKRTRARQHQLQEIYANAAHELIDSAVRGELEDKIAD
KADFNDRLVLYFMQGGRDIYTGTPLNIDQLSSYDIDHILPQSLIKDNSLDNRVLVHAIINREKNATFASTIYAQKMNATWRQWHEAGLISGR
KLRNLQMRPDQIDKYASGFVARQLVETRQIIKLTEQIVAAQYPETKIIAVKAGLSHQLREELEFPKNRDVNHYHHAFDAFLAARIGTYLLKR
FPNLQAFFTYGKFKKADVKKLRGFNFIRDITHAEDKIVVKDTGEVIWDKKRDVDELDRIYNFKRMLITHEVHFEKADLFKQTVYGAKDSKEA
GGKKQLISKKKGYPVDIYGGYTQETGSYLSVVRLTKKAMYAVVKVSTRDAAKLAVAKSISEQKENETLNDIIDEKLSKISKNGKITHELFEVV
LPRVGQKTLFKNSKYNLFLVNSDTYMHNYQELWMPREYQKMWKDISLSKYGDAQTEDQLDQIFGFIVSQVNSYFNLYDINQFRKKLNDA
ADKFATLPMRDTDDMQGKIATIGQLLIGLQANAARSDLRNLDIKTPLGLLQIPNGISLDSDTSVVYQSPTGLLEREVHLSDL

FIGURE 12 K.

>gi|227514633|ref|ZP_03944682.1|/1-1381
MDYVKEYHIGLDIGTSSIGWAVTDSQFKLMRIKGKTAIGVRLFEEGKTAAERRTFRTTRRRLKRRKWRLHYLDKIFAPHLQEVDENFLRRL
KQSNIHPEDPAKNQAFIGKLLFPDLLKKNERGYPTLIKMRDELPVEQRAHYPVTNIYKLREAMINEDRQFDLREVYLAVHHIVKYRGHFLN
NASVDKFKVGRIDFDKSFNVLNEAYEELQNGEGSFTIEPSKVEKIGQLLLDTKMRKLDRQKAVAKLLEVKVADKEETKRNKQIATAMSKLV
LGYKADFATVAMANGNEWKIDLSSETSEDEIEKFREELSDAQNDILTEITSLFSQIMLNAIVPNGMSISESMMDRYWTHERQLAEVKEYLA
TQPASARKEFDQVYNKYIGQVPKEKGFDLEKGLKKILSKKENWKEIDELLKAGDFLPKQRTSANGVIPHQMHQQELDRIIEKQAKYYPWL
ATENPATGERDRHQAKYELDQLVSFRIPYYVGPLVTPEVQKATSGAKFAWAKRKEDGEITPWNLWDKIDRAESAEAFIKRMTVKDTYLLN
EDVLPANSLLYQKYNVLNELNNVRVNGRRLSVGIKQDIYTELFKKKKTVKAGDVASLVMAKTRGVNKPSVEGLSDPKKFNSNLATYLDLK
SIMGDKVDDNRYQMDLENIIEWRSVFEDGEIFADKLTEVEWLTDEQRSALVKKRYKGWGRLSKKLLTGIVDENGQRIIDLMWNTDQNFM
QIVNQPVFKEQIDQLNQKAITNDGMTLRERVESVLDDAYTSPQNKKAIWQVVRVVEDIVKAVGNAPKSISIEFARNEGNKGEITRSRRTQL
QKLFEDQAHELVKDTSLTEELEKAPDLSDRYYFYFTQGGKDMYTGDPINFDEISTKYDIDHILPQSFVKDDSLDNRVLVSRAENNKKSDRV
PAKLYAAKMKPYWNQLLKQGLITQRKFENLTMDVDQNIKYRSLGFVKRQLVETRQVIKLTANILGSMYQEAGTDIIETRAGLTKQLREEFD
LPKVREVNDYHHAVDAYLTTFAGQYLNRRYPKLRSFFVYGEYMKFKHGSDLKLRNFNFFHELMEGDKSQGKVVDQQTGELITTRDEVAK
SFDRLLNMKYMLVSKEVHDRSDQLYGATIVTAKESGKLTSPIEIKKNRPVDLYGAYTNGTSAFMTIIKFTGNKPKYKVIGVPTTSAVGLKRV
GKPGSESYNQELHRIIKSNPKVKKDFEIVVPHVSYSQLIVDGDCKFTLASDTYQHPATQLVLSKESMEIIADDFKFVKENPATADEQLVRVF
DEIVNQMNHYFTIFDQRSNRQKVTKARDKFVSLPTESEYEGAKKTQVGKTEVITNLLMGLHANAAQGDLKVLGLSTFGFFQSSGGLNLSE
DAMIVYQSPTGLFERRICLKDI

>gi|122449465|sp|Q1WVK1|Q1WVK1_LA/1-1149
MERYHIGLDIGTSSIGWAVIGDDFKIKRKKGKNLIGVRLFKEGDTAAERRGFRTQRRRLNRRKWRLKLLEEIFDPYMAEVDKYFFARLKES
NLSPKDSNKKYLGSLLFPDISDSNFYDKYPTIYHLRRDLMEKDKKFDLREIYLAIHHIVKYRGNFLEKVPAKNYKNSGASIGFLLEEVNDLYG
NIIGNEDVAILDNDKFEDVEKIILNDEIRNIDKQKNVGRLLVKDKKEKNIVTAFSKAIFGYKFNLEDLLLIESDEKNKLTFNDENIDDIFNELSHS
LNDNQMDLLTKTREIYFKFKLNMIVPTGYTISESMIEKYEMHKAHLKMYKEFINTLNAKDRKILKNAYSDYINNEKAKAANAQENFYKTVKK
TIKENDSDTAKKIIGLIDEGNFMPKQRTGENGVIPHQLHQIELDRIIENQAKYYPWLAEENPVEKNRKFAKYKLDELVTFRVPYYVGPLIDKT
ESNKNEKETKFAWMVRKAKGTITPWNFENLVDRTESANRFIKRMTSKDTYIIGEDVLPASSLLYEKYKVLNELNNIKVNKKKLDVEQKQHV
YLDLFTTRKNVTKDDLATSLNCDVESITGLTDNKKFNSSLSSYIDLKAILGNIVDDYSKNEDLEKIIEYSTIFEDGNIYKEKLSEISWLTDEQIE
KLSNIHFKGWGRLSKKLLTQITNENGERIIDTLWNTSNNFIQVISDESIQAKLAEINGEYANKYNLEDILDEAYTSPQNKKAIRQVMKVVEDIE
KAMKCEPTSIAIEFTREKRKSKLTNTRYKKISETYEKITDELISEYELGKLQSELDSKVNNMRDRYYLYFMQLGRDMYTGEKMNIDELHQK
YDIDHILPQSFIKDDSLNNRVLTSKSVNIKEKSDKTAADLYAAKMGDFWRKLRKQGLMTEQKYKNLLTRTDSINKYTKQSFIKRQLVETSQ
VVKLAANILQDKYRNTKIIEIRARLNSDLRKKYELIKNREVNDYHHAIDGYLTTFVGQYLYKVYPKLRSYFVYDDFKKLDSNYLKHMDKFNFI
WKLEDKKAEDVYDKVNDEFVLNVPEMKEYIRKIYNYKYMLVSKEVTTKNGAFYDQTKYNAKTINLIPIKKDKPTNIYGGYKGKVSSYMMLV
KIQKKKEVIYKFVGVPRLWTDELDRLIDTDEKKALLKK

>gi|238854567|ref|ZP_04644902.1|/1-1391
MKEIKDYIIGLDIGTNSCGYVVTDKQNNILKLKGKTAIGARLFKEGEAAADRRAFRTTRRRLARRRWRLSLLEEIFDEEMNKVDPSFFRRLK
ESDYSPKDRRKQFNSIVFESSEADKAFYEKYPTIYHLRNALMHDNQKHDLREIFLAVHHIVKYRGNFLREDSVKAFKAAKFTLQGEDGIGP
VDKLNDLLKEIYIEAAPELATDNLSKIEDIIKDKKLYKQDKLKQIANLLPKAVDSKDKAKLNKDIAKQVANALMGYMFRLDTLFSFTDVDVKD
YKLKFSDANIDESLDALTSLLTDAQIEFVLELQSIYNTIVLNEIVPDGMSLSESMVKKYDDHKKDLKLYKEYIDSLSDQKKAKQLEAAYALYV
NYRKADLLAAKKLFEKKKIGDKKFVDVISNFEVFGKFVSDNLDDSELANKIKARLDLGEFLPKQRTNQNGVIPYQLHQVELTQILEKQGKYY
PFLITPNPVESHRNNAPYEISELVSFRVPYYVGPLIDNQSIKDKQNKNKFAWMVRQKQGQITPWNFEEMVDTTESANQFIKRMTRKDTYL
LAEDVLPKSSLIYQKFMILDELNRIKIDGKKLTSKLKHDIFEKLFKKQKSINLDNLKNYLLAEGNIPGLIEGLSDGINFNNSFSTYIDYRNIFGDE
IDNPNKQADFEKMIEWSTVFEDRKIFKRKLKEITWLTPEQIKQVSSKRYSGWGRLSKKLLTQITDENGVNILQRLWNEPETLTEVLANPVIK
RKISEANSLFVQINKVENILDDAYTSPQNKKAIRQVIRVVDDIIAAAHGKKPSQIAIEFARSSQNKAKVPDTRKKQLDKIYNKISSEILDSSIKN
ELKNLKSNKYLSKDKLFLYFKQMGRDAYTGDKLSLDQLQNYDIDHIFPRSFIKDDSLDNRVLTQKPINAKKSDYGIPALEFGNKYVPDLGIT
VKEMWKLWQENGLISKSKLINLCTNPKKIGSKRASGFINRQLVETSQVIKLVAIILQAEFPDTEIIEVKALQNTTLRESFHLYKNRSVNDYHH
AIDAYLTTIVGNYLYQVYPKLRPYFVYGQFKKFNQEKNIDILKRLKNFNFLRQLIFNTDDNIYISGTKEIVFNKKDIIHKLETAYGYKYMNISRE
CYQETSSLFNQTLYAHNSSVKNSLIPKKKGLPTEIYGGYSGNKDSFFVLVKIVKKRTNLYRIVGIPTRELAKLNSSNNYNQALNKIVESKLCL
KETESFKILIKRLLYGTLIVDNGQKFRIGSFKEKHNVQQLVLQLKSMKYIKFYIDGGQNYFTDVERKKLEKQDRDKCLLYVFDDIMNVVNKR
FTLFDMSKYEKDGDSLREKFNCLDFNDKVSILSDLLKAFHANSDRTSITKLKITNLGRHQAGKNGITLTTNAQIIYQSPTGLFERRIKIKDL

FIGURE 12 L.

>gi|256852176|ref|ZP_05557562.1|/1-1395
MKEIKDYIIGLDIGTNSCGYVVTDKQNNILKLKGKTAIGARLFKEGQAAADRRSFRTTRRRLARRHWRLGLLEEIFDPEMEKIDPNFFRRLK
ESDYSPKDSRKQFNAIVFKSIKADKRFYKKYPTIYHLRNALMHDNQKHDLREIFMAVHHIVKYRGNFLREDSVNAFKAAKFSLRGEDGIGP
VDKLNDLLKEIYSEYSPELATNNLSKIEDIIKDKNLYKQDKLKQVATLLIKEADSKDKAKLNKDIAKQVANSFMGYMFRLDTLFSLTDVDVKD
YKLKFSDANIDESLDDLTSLLTDAQIEFVLELQSIYNTIVLNEIVPDGMSLSESMIKKYDDHKEDLKIFKEYIDSLSDQKKAKKLLSAYNLYVN
YRKSDLVEAKNLFKNKKVGDKNFADVISNFEVFGKFVSDNLDDSELANKIKARLDLGEFLPKQRTNQNGVIPYQLHQVELTQILEKQGKYY
PFLITPNPVESHRNNAPYEISELVSFRVPYYVGPLIDNQSIKDKQNKNKFAWMVRQKQGQITPWNLEEMVDTTESANQFIKRMTRKDTYL
LAEDVLPKSSLVYQKFMILDELNRIKIDGKKLTSELKHDIFEKLFKKQKSINLDNLKNYLLVEGNIPGLIEGLSDGINFNNSFSTYIDYRNIFGD
EIDNPNKQADFEKMIEWSTVFEDRKIFKRKLKEITWLTSEQIKQVSSKRYSGWGRLSKKLLTQITDENGVNILQRLWNEPETLTEVLANPVI
KRKIVEANSSFTKAKDMKDILVEAYTSPQNKKAIRQVMRVVDDIIAAAHGKKPSQIAIEFAREDQDNPQMTKSRKAQLDQIYAKISDEFLDD
AVKNELKNMKDNKPLAKDKVFLYFKQMGRDAYSGEKLSLDNLQDYDIDHIYPQSFIKDDSFDNRVLVQRAYNNGKSDNVPVLLFGNKIAA
GLGITIRQMWKKWLELGLISKRKFNNLITDPEKISRNMASGFVNRQLVETSQVIKLVANILQAKLPDTEIIEVKASYNSILRKEFNLYKSREVN
DYHHAIDAYLTTIVGNYLYQVYPKLRPYFVYGQFKKFGKAENQEDDPVKKIKKFNFIYQLIWGKDDAICISHTNQQVFSKKDIIDKLTTAYNF
KYMNISHAVYTRNNNMFKQTLFPIIERTTAKTKSLIPKKLDRPTDIYGGYSNNNDAYLALVKVDKKRGSEYRVVGIPMRAAAMLKASKNYD
EDLRKVIEPMVMFDKNGKAKRGIEGFRVLVGKMPFKQAVLDGNKKFMLTGTVDTPNLKQLVLSQQNMKYILDYIEDVDCRKHKFGDKVQ
NPDKCLLEVFDEIIEIVDKYFELFDSRDFREKLRNSRNKFIKLEASKKADTILAILKGLQAKSTRATIKELNISDFGRIITSSISPEAQLIYQSPTG
LFERRVKVKDL

>gi|300361537|ref|ZP_07057714.1|/1-1391
MTKIKNEYIVGLDVGTNSCGWVAMDFQNTILRMHGKTAIGSHLFDAGNSAADRRAFRTTRRRIKRRKWRLKLLEEIFDPYMTEVDPYFFA
RLKESGLSPLDKRKNVSSIVFPTALEDKKFYCNYPTIYHLRYDLMSEDKKFDLREIYLAIHHIVKYRGNFLYNTPVKDFEASKIDVKDSLEKL
NELYERLDSEFTVELDSSNALEIEKIIRDKKVFKINKVKSIHQLLSLKTENKERTKLIKDVSKQIINAILGYKTKFETILLKNISKDEADDWEFKL
TDVDADNKFENLIGDLNENEQEILLEIRNLANAITLSNIVEEGKTLSESMIDKYNKHSDDLELLKQVISDHPDRDKAKKLALAYDLYVNNRHG
KLLQAKDVLGSKKTLSKEDFYKEVKKNLDDSKASQEILDTIALDSFMPKQRTNENGVIPYQLHQLELDRIIKNQGKYYPFLKEVNPVSSHLK
EAPYKLDELIRFRVPYYVGPLISPNESTKNKQTKKNQNFAWMIRKEQGQITPWNFDQKVDRMASANKFIKRMTTKDTYLLGEDVLPANSL
LYQKFTVLNELNNIKINGKRISVPLKEELYNNLFKKNSTVTTNKLKSYLKENYNLINVEIKGLADEKKFNSGLTTYNKLRNLKIFDQQLDDLNY
DKDFERIIEWSTIFEDKAIYKDKLKTIKWLSDRQIDKLSKIRMQGWGQLSKKLLSQLTDNNGQTIIEQLWDSQNNFMQIVNQSDFKDAIAVA
NQNLLVNTSVEDILNEAYTSPANKKAIRQVVKVVDDIVKAASGKVPKQIAIEFTRDADDKAKISQTRANKLRKVYKELSNELASETIRNELER
VAKDQKLLKDKYYLYFMQLGRDAYTGEPINIDDLGQYDIDHILPQSFIKDDSLENRVLVKKAVNNGKSDNVPVKLYGNRMAADLGMTIRH
MWEKWKDQGLITKTKYNNLIIDPDKINKYESSGFIHRQLVETSQIIKLVSTILQSKYPDTEIIVVKARYNHYLRKHLNLYKSREVNDYHHAIDA
YLSAICGNLLYQVYPYLRPFFVYGQYKKFSSDPKKEKIIYDKTRKYNFISQIFENKSNDIISLETKKKVFDKKDIIEKLKHAYNYKYMLVSRET
ETHDQEMFKMTVYPRLSRDTKKSRNLISKKKDMPTEIYGGYTNNSDAYMVIARINKKKGTEYRVFGVPMRELVNLRKAEKNGRYNAYLK
QVLEPEIMYSKNGKKNKTISSFEIVKAKVPYKQVILDGDKKFMLASSTYVQNAKQLTLSQNSMQAITDNCENDTEEEKALEEKALIEAYDEI
LTNIDKYLPLFDINKFREKLHAGREKFINLSLDVKKDTILQVLNGLHDNAVMPKIKSLGLSTPLGFMQFSSGVPLSENAKLIYQSPTGLFEKR
VKISDL

>gi|270290729|ref|ZP_06196953.1|/1-1229
MGDRKYNLGLDIGTSSIGFAAVDENNQPIRVKGKTAIGVRLFEEGKTAADRRGFRTTRRRLSRRRWRINLLNEIFDAHLAEVDPTFLARLK
ESNRSNLDPKKSFQGSLLFPERKDYQFYEEYPTIYHLRKALMEKDRKFDIREIYLAVHHIIKYRGNFLNGTPMRSFKVENIELNTLFDQLNQ
LYAEIVPDNELAFDLAQVADVKDVLSSTTIYKMDKKKQLVKMMLLPASNKALQSENKKIVTQFVNAILNYKFKLDVLLQVETDADWSLKLN
DEGADDKLEEFTGDLDENRLEIIDLLQRLHNWFSLNEITKDGNSLSAAMVEKYENHHHHLGLLKKVIENHPDAKKAKALKETYTAYVGKTD
DKTQNQDDFYKAVEKNLDDSPDAKGIKRLIQLDQFMPKQRTGQNGAIPHQLHQQELDQIIEKQSKYYPFLAEPNPNVKRRKDAPYKLDEL
IAFKIPYYVGPLVTPEEQAQNKENVFAWMKRKAAGPITPWNFDEKVDRMESANRFIRRMTTKDTYLFGEDVLPAESMIYQKFVVLNELNN
LKINGRHLSLKDKQDVYNDLFKQQKTVSIKALQNYYVTKKKAATAPTVGGLADPKKFLSSLSTYIDFKNMFGERVNDPQFQEDLEQIVEW
STIFEDRGIFKAKLQALGWLSEKQIQQLVAKRYKGWGRLSKKLLTGLKNAEGYSILDEMWRSTENFMQIQSRPEFAALIQQANEKQFEGN
DPDNVWENIENILGDAYTSPQNKKAIRQVVKVVQDIEKAVGNPPEKIAIEFTREAAANPQRTQSRLRTLEKLYESAEEVVDAGLTAELAEF
KENKHVLSDKYYLYFTQLGRDVYTGDTISLDKLNDYDVDHILPQSFIKDDSLDNRVLTIRAVNNGKSDNVPAKMFGKKMGSFWRHLLDNG
MISKRKYNNLITDPDNISKYAQKGFINRQLVETSQVIKLTANILNGIYDKDTEIIEVPAKMNSQMRKMFDLVKVREVNDYHHAFDAYLTIFIGN
YLYKCYPKLRPYFVYDNFKKFGNKEDIGHKRFNFLGKIEREKKVVAPETGEILWSNVAPNETIKQIKKVYNYKFMIVSREITTRRAELFNQT
VYPKNYRGKLIPIKEDRPTDLYGGYSGNTDAYLAIVALEDKKKGKYFKVVGIPTRAAAKLEKLKQQDSQQYLQALHEVIAPQFTKSTKKGIK
KTEFEIVLDKVHYRQLVQDGPVKMMLGSSTYKYNAKQLVLSEKAL

>gi|187250660|ref|YP_001875142.1|/1-1195
MQKNINTKQNHIYIKQAQKIKEKLGDKPYRIGLDLGVGSIGFAIVSMEENDGNVLLPKEIIMVGSRIFKASAGAADRKLSRGQRNNHRHTRE
RMRYLWKVLAEQKLALPVPADLDRKENSSEGETSAKRFLGDVLQKDIYELRVKSLDERLSLQELGYVLYHIAGHRGSSAIRTFENDSEEA
QKENTENKKIAGNIKRLMAKKNYRTYGEYLYKEFFENKEKHKREKISNAANNHKFSPTRDLVIKEAEAILKKQAGKDGFHKELTEEYIEKLT
KAIGYESEKLIPESGFCPYLKDEKRLPASHKLNEERRLWETLNNARYSDPIVDIVTGEITGYYEKQFTKEQKQKLFDYLLTGSELTPAQTKK
LLGLKNTNFEDIILQGRDKKAQKIKGYKLIKLESMPFWARLSEAQQDSFLYDWNSCPDEKLLTEKLSNEYHLTEEEIDNAFNEIVLSSSYAP
LGKSAMLIILEKIKNDLSYTEAVEEALKEGKLTKEKQAIKDRLPYYGAVLQESTQKIIAKGFSPQFKDKGYKTPHTNKYELEYGRIANPVVHQ
TLNELRKLVNEIIDILGKKPCEIGLETARELKKSAEDRSKLSREQNDNESNRNRIYEIYIRPQQQVIITRRENPRNYILKFELLEEQKSQCPFC
GGQISPNDIINNQADIEHLFPIAESEDNGRNNLVISHSACNADKAKRSPWAAFASAAKDSKYDYNRILSNVKENIPHKAWRFNQGAFEKFIE
NKPMAARFKTDNSYISKVAHKYLACLFEKPNIICVKGSLTAQLRMAWGLQGLMIPFAKQLITEKESESFNKDVNSNKKIRLDNRHHALDAIV
IAYASRGYGNLLNKMAGKDYKINYSERNWLSKILLPPNNIVWENIDADLESFESSVKTALKNAFISVKHDHSDNGELVKGTMYKIFYSERG
YTLTTYKKLSALKLTDPQKKKTPKDFLETALLKFKGRESEMKNEKIKSAIENNKRLFDVIQDNLEKAKKLLEEENEKSKAEGKKEKNINDASI
YQKAISLSGDKYVQLSKKEPGKFFAISKPTPTTTGYGYDTGDSLCVDLYYDNKGKLCGEIIRKIDAQQKNPLKYKEQGFTLFERIYGGDILE
VDFDIHSDKNSFRNNTGSAPENRVFIKVGTFTEITNNNIQIWFGNIIKSTGGQDDSFTINSMQQYNPRKLILSSCGFIKYRSPILKNKEG

FIGURE 12 M.

>gi|269123826|ref|YP_003306403.1|/1-1259
MKYILGLDLGVASIGWAVTEINEKGEPVGLIDANSVIFKPLDNDKGKLYNVERRDKRGSRRILRRKKHRVERTKMLLVNSSFLTNEEIDNLY
VGKLDNIYEVRLKGLKEQLSKNEIARLMIYYCKNRGFKSNRKTEDLEILKSLGEKISEKDSDEKKLKPIIAKNTELIESKGLTPIEVIYMIREND
NSILGFKNKEGNYKFGFKRNQVIDEVRKILGTQQILSKEIVDEYIDILSSQRDFSDGPGGDSKYKIDYTKLAGKCKYTGEVRAVKSAPSYEIF
TMLQKLNDIRYVKFTSEDKIEKKKLSKEVIHKLYDLVVEKNKTLTYDLIEKSIDEENIKLLNIPKLSKSKYVELRKSYFKERNDNENLTEEEINA
FNEKLNNERMKQELIRNNLKAYKELKSKFKKNNIDENRIKELGGIYFLDLVAELLTYSKTDEKLEYFVENVERYSLFKEQRDIVEIIKTFPNYT
KNGNLSLSLVRELNKLLMEGHDYESSLSSLNYYIDTDVDWEKFPTISEIEDSLSTKITNPNVKHILVILRKLYNTLLFKYGRPEKVHLELSRD
FSNDFSTRNKIQKEQLENKVRREVAAFEMYGANKDIVAGKDRLSNDDFVRIKLWEEQNKVCMYSGRTIEKYQLTSAEVQIDHILPYSKSF
DNSYSNKVLVFSNENQDKKERTPYQWLKGTEKWNEFKQRVRLNLNISNKKKENLLFEDEVVNNEFLERELHATSYSSRLALNIFQRLIPV
SDEDKYDEHGNEKVKYLYNRNVIAFQGKMTSMLRNLYSLNKYTHSFESDTLDIDNKAFILKEFEINKDNLKMTAYNVNTGLEITSETKVEK
NKSGEFKTIKDELLKKELDKKENLIEISDYFKDKVLFDLKIVDFEEIDTVEPEIISILYKMITALKEEVYSKNRENHLHHSLDAFLLTIMNRSMQ
MKLIKYNQLISVLKNKEIEIFNEEKGEYIDSKEFAEELQKQKIIDIDGNERGIKFNVNGKTHRLNLVKPYENFLDDIKNKIFDKRENNKLPYHV
VKSKVSGALHAETILGESKGEITKRISVFDINYKNLGKIFDKDGSQKEIYETLVKWLEAKSKDYPKLKNGNIIKKVKIVDGNKDKLIKLGNKRY
VEMGRTTVKILVLKKEAEEGLKFASIGRYKYNLLKSEKDVNISIWKDAIHFCKVRYNKLKENGYRLIHELIPGETIELELNKGNISKCLVVGFS
GGKIEISSVIGDALDLINDKISTRIKERYYITVSTIKSIKKINKNILGD

>gi|160915782|ref|ZP_02077990.1|/1-1096
MMEVFMGRLVLGLDIGITSVGFGIIDLDESEIVDYGVRLFKEGTAAENETRRTKRGGRRLKRRRVTRREDMLHLLKQAGIISTSFHPLNNP
YDVRVKGLNERLNGEELATALLHLCKHRGSSVETIEDDEAKAKEAGETKKVLSMNDQLLKSGKYVCEIQKERLRTNGHIRGHENNFKTRA
YVDEAFQILSHQDLSNELKSAIITIISRKRMYYDGPGGPLSPTPYGRYTYFGQKEPIDLIEKMRGKCSLFPNEPRAPKLAYSAELFNLLNDLN
NLSIEGEKLTSEQKAMILKIVHEKGKITPKQLAKEVGVSLEQIRGFRIDTKGSPLLSELTGYKMIREVLEKSNDEHLEDHVFYDEIAEILTKTK
DIEGRKKQISELSSDLNEESVHQLAGLTKFTAYHSLSFKALRLINEEMLKTELNQMQSITLFGLKQNNELSVKGMKNIQADDTAILSPVAKR
AQRETFKVVNRLREIYGEFDSIVVEMAREKNSEEQRKAIRERQKFFEMRNKQVADIIGDDRKINAKLREKLVLYQEQDGKTAYSLEPIDLKL
LIDDPNAYEVDHIIPISISLDDSITNKVLVTHRENQEKGNLTPISAFVKGRFTKGSLAQYKAYCLKLKEKNIKTNKGYRKKVEQYLLNENDIYK
YDIQKEFINRNLVDTSYASRVVLNTLTTYFKQNEIPTKVFTVKGSLTNAFRRKINLKKDRDEDYGHHAIDALIIASMPKMRLLSTIFSRYKIEDI
YDESTGEVFSSGDDSMYYDDRYFAFIASLKAIKVRKFSHKIDTKPNRSVADETIYSTRVIDGKEKVVKKYKDIYDPKFTALAEDILNNAYQE
KYLMALHDPQTFDQIVKVVNYYFEEMSKSEKYFTKDKKGRIKISGMNPLSLYRDEHGMLKKYSKKGDGPAITQMKYFDGVLGNHIDISAH
YQVRDKKVVLQQISPYRTDFYYSKENGYKFVTIRYKDVRWSEKKKKYVIDQQDYAMKKAEKKIDDTYEFQFSMHRDELIGITKAEGEALIY
PDETWHNFNFFFHAGETPEILKFTATNNDKSNKIEVKPIHCYCKMRLMPTISKKIVRIDKYATDVVGNLYKVKKNTLKFEFD

>gi|169349750|ref|ZP_02866688.1|/1-1116
MVEFLLKEMKIMGRLVLGLDIGITSVGYGVIDIDKNEFVDYGVRLFKEGTAINNEERRSKRGARRLKRRKTNRIADMKKLLERNGLIDEQYT
ALNNPYEIRKKGLKNKLSNYELATAILHITKNRGTCLESNADETQDEESTKNILSKNALELKDKYICEIQLARLEESGKIRGINNNFKTEDYLK
ELSKILENQDLEKETIDKIKEIVARRRRYDQGPGSEKSPTPYGSYRMENGELVHVNLIDMMRGKCSVFPDQLRAPQQSYSAELFNLLNDL
NNLKINGEKLSSTEKQKIIDFVNEKGGITVKQLLKLIDVGITEISGFRIDKNEKPLITEFKGYKKVLKVFKNYNQESLLNDKSIVDTIMDINTKIK
SIDERSDQLKQVLPSLDTSIINDLASIKGVSGYHSLSLKALNLINNEMLKTEMNQMQVLHQLDIFDKNRVSLKGQRQIYADDNAILSPVAKR
AHRETFKVINKLRSIYGEFDSIVIETTRDKNTQDQKKRIKNNQKNYEERNTEVNNFLKEAGYNPENINAKTKMKIRLYIEQDAKSAYTLQDIN
LKELIEDEKAYEIDHIIPISISLDDSFNNKVLTTHDENQEKGNLTPISAFLKGKFHNGNLNTYKVFIQNNKNLNRKKKEYLLCEDDITRYEVVQ
NFINRNLVDTSYACRTVMNTLVHYFKDNEINTKVHTIKGQATSTFRKRIGLIKDREEDYFHHAIDALIVASLKKMNLINSYLMKVDFNEMYDE
ETGEIFNVLPDNSYLDEKYISFVSNLKNIYEESNKYIRGVMIKEKMTYPLIKVSHKIDTKPNRKISDDTIYSTRNVDNQDMLVERIKNIYDPKE
NAANTLINDIINDNTDKYIMYHKDPQTFDKIKAIVLNHFNEFKTDKNCYIVTKEGLYKLKGTSPLALYYEEFGPITKYAKKNNGPTIHSMKFYS
EKLGNHLSITHKYSTVNKKVITKQISPYRTDFYQSPEGKYKFVTVRYKDVAYNAKKDKYIIDHSWYREQKKLKKIADEWIFVCSVHRDELIGI
MKADGQKRIFDSSTESDGETLYHDAKTYEILKFTATNNDITGRFEVKPIYCYCKKQLMVSIGTCRKIQKFATDVLGNLYIVKDNVLKLEFK

>gi|315169436|gb|EFU13453.1|/1-1150
MYSIGLDLGVSSVGWSVIDERTGNVIDLGVRLFSAKNSEKNLERRTNRGGRRLIRRKTNRLKDAKKILAAVGFYEDKSLKNSCPYQLRVK
GLTEPLSKGEIYKVTLHILKKRGISYLDEDDTEAAKESQDYKEQVRENAQLLTKYTPGQIQLQRLKENNRVKTGINAQGNYQLNVFKVSAY
ANELATILKTQQAFYPNELTDDWIALFVQPGIAEEAGLIYRKRPYYHGPGNEANNSPYGRWSDFKKTGQPATNIFDKLIGKDFQGELRAS
GLSLSAQQYNLLNDLTNLKIDGEIPLSPEQKEYILAELMTKEFTRFGINDVAKLLGVKKERLSGWRLDKKGKPEIHTLKGYRNWRKIFAESG
IDLATLPTETIDCLAKVLTLNTEREGIENTLAFELPELAESVKLLVLDRYKELSQSVSTQAWHRFSLKTLHLLIPELMNATSEQNTLLEQFQL
KSDVRKRYSEYKKLPTKDVLTEIYNPTVNKTVSQAFKVIDALLVKYGKEQIRYITIEMPRDDNEEDEKKRIKELHAKNSQRKNDSQSYFMQ
KSGWSQEKFQTTIQKNRRFLAKLLYYYEQDGICAYTGLSISPELLVSDSTEIDHIIPISISLDDSINNKVLVLSKANQVKGQQTPYDAWMDG
SFKKINGKFSNWDDYQKWVESRHFSHKKENNLLETRNIFDSEQVEKFLARNLNDTRYASRLVLNTLQSFFANQETKVRVVNGSFTHTLR
KKWGADLDKTRETHHHHAVDATLCAVTPFVKVSRYHYAVKEETGEKVMREIDFETGEIVDEMSYREFKKSKKYERKTYQVKWPNFREQ
LKPVNLHPRIKFSHQVDRKANRKLSDATIYSVREKTEVKTLKSGKQKITTDEYTIGKIKDIYTVDGWEAFKKKQDKLLMKDLDEKTYERLLSI
AETTPDFQEVEEKNGKVKRVKRSPFAVYCEENDIPAIRKYAKNNNGPLIRSLKYYDGKLNKHINITKDSQGRPVEKTKNGRKVTLQSLKPY
RYDIYQDLETKAYYTVQLYYSDLRFVEGKYGITEKEYMKKVAEQTKGQVVRFCFSLQKNDGLEIEWKDSQCYDVRFYNFQSANSINFKGL
EQEMMPAENQFKQKPYNNGAINLNIAKYGKEGKKLRKFNTDILGKKHYLYYEKEPKNIIK

FIGURE 12 N.

>gi|288905632|ref|YP_003430854.1|/1-1130
MTNGKILGLDIGIASVGVGIIEAKTGKVVHANSRLFSAANAENNAERRGFRGSRRLNRRKKHRVKRVRDLFEKYEIVTDFRNLNLNPYELR
VKGLTEQLTNEELFAALRTISKRRGISYLDDAEDDSTGSTDYAKSIDENRRLLKTKTPGQIQLERLEKYGQLRGNFTVYDENGEAHRLINVF
STSDYEKEARKILETQADYNKKITAEFIDDYVEILTQKRKYYHGPGNEKSRTDYGRFRTDGTTLENIFGILIGKCSFYPDEYRASKASYTAQ
EYNFLNDLNNLKVPTETGKLSTEQKEALVEFAKSTATLGPAKLLKEIAKILDCKVDEIKGYREDDKGKPDLHTFEPYRKLKFNLDSVNIDDL
SREVLDKLADILTLNTEREGIEDAIRHNLPNQFTEGQISEIIKVRKSQSTAFNKGWHSFSAKLMNELIPELYATSDEQMTILTRLEKFKVNKK
SSKNTKTIDEKEVTDEIYNPVVAKSVRQTIKIINAAVKKYGDFDKIVIEMPRDKNADDEKKFIDKRNKENKKEKDDALKRAAYLYNGTDKLPD
EVFHGNKQLETKIRLWYQQGERCLYSGKPIPIQELVHNSNNFEIDHILPLSLSFDDSLANKVLVYAWTNQEKGQKTPYQVIDSMDAAWSF
REMKDYVLKQKGLGKKKRDYLLTTENIDKIEVKKKFIERNLVDTRYASRVVLNSLQSALRELGKDTKISVIRGQFTSQLRRKWKIDKSRETY
HHHAVDALIIAASSQLKLWEKQDNPMFVDYGNNQVVDKQTGEILSVSDDEYKELVFQPPYQGFVNMISSKGFEDEILFSYQVDSKYNRKV
SDATIYSTRKAKIGKDKKEETYVLGKIKDIYSQNGFDTFIKKYNKDKTQFLMYQKDPLTWENVIEVILRDYPTTKKSEDGKNDVKCNPFEEY
RRENGLVCKYSKKGKGTPIKSLKYYDKKLGNCIDITPEGSKNEVVLQSLNPWRADVYFNPETLKYELLGLKYSDLSFEKGTGKYHISQEKY
DVIKEKEGIGKKSEFKFTLYRNDLILIKDTASGEQEIYRFLSRTMPNVKHYAELKPYDKEKFDNVQELVEALGEADKVGRCIKGLNKSNLSIY
KVRTDVLGNKYFVKKEGDKPKLDFKNNKK

>gi|157150687|ref|YP_001450662.1|/1-1136
MNGLVLGLDIGIASVGVGILEKDTGKIIHASSRLFPAATADNNVERRSNRQGRRLNRRKKHRSVRLQDLFEGYGLLTDFSKVSMNLNPYQ
LRVQGMENQLTNEELFVALKNIVKRRGISYLDDASEDGGTVSSDYGKAVEENRKLLAEKTPGQIQLERFEKYGQLRGDFTVEENGEKHR
LINVFSTSAYRKEAERILRKQQEFNSKITDEFIEDYLIILTGKRKYYHGPGNEKSRTDYGRFRTDGTTLDNIFGILIGKCTFYTEEYRASKASY
TAQEFNLLNDLNNLTVPTETKKLSEEQKKLIIEYAKSAKTLGASTLLKYIAKMIDASVDQIRGYRVDVNNKPEMHTFEVYRKMQSLETIKVEE
LPRKVLDELAHILTLNTEREGIEEAINSKLKDIFNRDQVLELVQFRKNNSSLFSKGWHNFSIKLMMELIPELYETSEEQMTILTRLGKQRSKE
TSKRTKYIDEKELTEEIYNPVVAKSVRQAIKIINEATKKYGIFDNIVIEMARENNEEDAKKDYIKRQKANQDEKNAAMEKAAFQYNGKKELPD
NIFHGHKELTTKIRLWHQQGEKCLYTGKNIPISDLIHNQYKYEIDHILPLSLSFDDSLSNKVLVLATANQEKGQRTPFQALDSMDDAWSYRE
FKSYVKDSKLLSNKKKDYLLTEEDISKIEVKQKFIERNLVDTRYSSRVVLNALQDFYKSHQLDTTISVVRGQFTSQLRRKWGIEKSRETYHH
HAVDALIIAASSQLRLWKKHSNPLIAYKEGQFVDSETGEIVSLSDEEYKELVFKAPYDHFVDTLRSKKFEDSILFSYQVDSKYNRKISDATIY
ATRKAKLDKEKKEYTYTLGKIKDIYALGTKTPSKTGFYKFLDLYKTDKSQFLMYQKDRKTWDEVIEKIIEQYRPFKEYDKNGKEVDFNPFEK
YRIGNGPIRKYSKKGNGPEIKSLKYYDILLGKHKNITPDGSRNTVALLSLNPWRTDVYYNSETKKYEFLGLKYADLCFEEGGAYGISEVKYK
KIREKEGIGKNSEFKFTLYKNDLILIKDTETNCQQFFRFWSRTGKDNPKSFEKHKIELKPYEKAKFEKGEELKVLGKVPPSSNQFQKNMQI
ENLSIYKVKTDILGNKHFIKKEGDEPKLKFKK

>gi|81560783|sp|Q5M542|Q5M542_STR/1-1122
MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLTRRKKHRIVRLNRLFEESGLITDFTKISINLNPYQLRVK
GLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVF
PTSAYRSEALRILQTQQEFNSQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTNGETLDNIFGILIGKCTFYPDEFRAAKASYTAQ
EFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKVMGPAKLFKYIAKLLSCDVADIKGHRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDR
ETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTS
SSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPH
SVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNPNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFR
ELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTY
HHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIY
ATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKEMNEKGKEVPCNPFLKYKEEHG
YIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYADLQFEKKTGTYKISQEKYNGIMK
EEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPNVKYYVELKPYSKDKFEKNESLIEILGSADKSGRCIKGLGKSNISIYKVRT
DVLGNQHIIKNEGDKPKLDF

>gi|154482474|ref|ZP_02024922.1|/1-1107
MGYTVGLDIGVASVGVAVLDENDNIVEAVSNIFDEADTSNNKVRRTLREGRRTKRRQKTRIEDFKQLWETSGYIIPHKLHLNIIELRNKGLT
ELLSLDELYCVLLSMLKHRGISYLEDADDGEKGNAYKKGLAFNEKQLKEKMPCEIQLERMKKYGKYHGEFIIEINDEKEYQSNVFTTKAYK
KELEKIFETQRCNGNKINTKFIKKYMEIYERKREYYIGPGNEKSRTDYGIYTTRTDEEGNFIDEKNIFGKLIGKCSVYPEEYRASSASYTAQE
FNLLNDLNNLKINNEKLTEFQKKEIVEIIKDASSVNMRKIIKKVIDEDIEQYSGARIDKKGKEIYHTFEIYRKLKKELKTINVDIDSFTREELDKT
MDILTLNTERESIVKAFDEQKFVYEENLIKKLIEFRKNNQRLFSGWHSFSYKAMLQLIPVMYKEPKEQMQLLTEMNVFKSKKEKYVNYKYIP
ENEVVKEIYNPVVVKSIRTTVKILNALIKKYGYPESVVIEMPRDKNSDDEKEKIDMNQKKNQEEYEKILNKIYDEKGIEITNKDYKKQKKLVLK
LKLWNEQEGLCLYSGKKIAIEDLLNHPEFFEIDHIIPKSISLDDSRSNKVLVYKTENSIKENDTPYHYLTRINGKWGFDEYKANVLELRRRGK
IDDKKVNNLLCMEDITKIDVVKGFINRNLNDTRYASRVVLNEMQSFFESRKYCNTKVKVIRGSLTYQMRQDLHLKKNREESYSHHAVDAM
LIAFSQKGYEAYRKIQKDCYDFETGEILDKEKWNKYIDDDEFDDILYKERMNEIRKKIIEAEEKVKYNYKIDKKCNRGLCNQTIYGTREKDGK
IHKISSYNIYDDKECNSLKKMINSGKGSDLLMYNNDPKTYRDMLKILETYSSEKNPFVAYNKETGDYFRKYSKNHNGPKVEKVKYYSGQIN
SCIDISHKYGHAKNSKKVVLVSLNPYRTDVYYDNDTGKYYLVGVKYNHIKCVGNKYVIDSETYNELLRKEGVLNSDENLEDLNSKNITYKF
SLYKNDIIQYEKGGEYYTERFLSRIKEQKNLIETKPINKPNFQRKNKKGEWENTRNQIALAKTKYVGKLVTDVLGNCYIVNMEKFSLVVDK

FIGURE 12 O.

>gi|238924075|ref|YP_002937591.1|/1-1114
MNYTEKEKLFMKYILALDIGIASVGWAILDKESETVIEAGSNIFPEASAADNQLRRDMRGAKRNNRRLKTRINDFIKLWENNNLSIPQFKST
EIVGLKVRAITEEITLDELYLILYSYLKHRGISYLEDALDDTVSGSSAYANGLKLNAKELETHYPCEIQQERLNTIGKYRGQSQIINENGEVLD
LSNVFTIGAYRKEIQRVFEIQKKYHPELTDEFCDGYMLIFNRKRKYYEGPGNEKSRTDYGRFTTKLDANGNYITEDNIFEKLIGKCSVYPDE
LRAAAASYTAQEYNVLNDLNNLTINGRKLEENEKHEIVERIKSSNTINMRKIISDCMGENIDDFAGARIDKSGKEIFHKFEVYNKMRKALLEI
GIDISNYSREELDEIGYIMTINTDKEAMMEAFQKSWIDLSDDVKQCLINMRKTNGALFNKWQSFSLKIMNELIPEMYAQPKEQMTLLTEMG
VTKGTQEEFAGLKYIPVDVVSEDIFNPVVRRSVRISFKILNAVLKKYKALDTIVIEMPRDRNSEEQKKRINDSQKLNEKEMEYIEKKLAVTYG
IKLSPSDFSSQKQLSLKLKLWNEQDGICLYSGKTIDPNDIINNPQLFEIDHIIPRSISFDDARSNKVLVYRSENQKKGNQTPYYYLTHSHSEW
SFEQYKATVMNLSKKKEYAISRKKIQNLLYSEDITKMDVLKGFINRNINDTSYASRLVLNTIQNFFMANEADTKVKVIKGSYTHQMRCNLKL
DKNRDESYSHHAVDAMLIGYSELGYEAYHKLQGEFIDFETGEILRKDMWDENMSDEVYADYLYGKKWANIRNEVVKAEKNVKYWHYVM
RKSNRGLCNQTIRGTREYDGKQYKINKLDIRTKEGIKVFAKLAFSKKDSDRERLLVYLNDRRTFDDLCKIYEDYSDAANPFVQYEKETGDII
RKYSKKHNGPRIDKLKYKDGEVGACIDISHKYGFEKGSKKVILESLVPYRMDVYYKEENHSYYLVGVKQSDIKFEKGRNVIDEEAYARILV
NEKMIQPGQSRADLENLGFKFKLSFYKNDIIEYEKDGKIYTERLVSRTMPKQRNYIETKPIDKAKFEKQNLVGLGKTKFIKKYRYDILGNKYS
CSEEKFTSFC

>gi|123747930|sp|Q4A5I2|Q4A5I2_MY/1-1314
MLRLYCANNLVLNNVQNLWKYLLLLIFDKKIIFLFKIKVILIRRYMENNNKEKIVIGFDLGVASVGWSIVNAETKEVIDLGVRLFSEPEKADYR
RAKRTTRRLLRRKKFKREKFHKLILKNAEIFGLQSRNEILNVYKDQSSKYRNILKLKINALKEEIKPSELVWILRDYLQNRGYFYKNEKLTDE
FVSNSFPSKKLHEHYEKYGFFRGSVKLDNKLDNKKDKAKEKDEEEESDAKKESEELIFSNKQWINEIVKVFENQSYLTESFKEEYLKLFNY
VRPFNKGPGSKNSRTAYGVFSTDIDPETNKFKDYSNIWDKTIGKCSLFEEEIRAPKNLPSALIFNLQNEICTIKNEFTEFKNWWLNAEQKSE
ILKFVFTELFNWKDKKYSDKKFNKNLQDKIKKYLLNFALENFNLNEEILKNRDLENDTVLGLKGVKYYEKSNATADAALEFSSLKPLYVFIKF
LKEKKLDLNYLLGLENTEILYFLDSIYLAISYSSDLKERNEWFKKLLKELYPKIKNNNLEIIENVEDIFEITDQEKFESFSKTHSLSREAFNHIIPL
LLSNNEGKNYESLKHSNEELKKRTEKAELKAQQNQKYLKDNFLKEALVPLSVKTSVLQAIKIFNQIIKNFGKKYEISQVVIEMARELTKPNLE
KLLNNATNSNIKILKEKLDQTEKFDDFTKKKFIDKIENSVVFRNKLFLWFEQDRKDPYTQLDIKINEIEDETEIDHVIPYSKSADDSWFNKLLV
KKSTNQLKKNKTVWEYYQNESDPEAKWNKFVAWAKRIYLVQKSDKESKDNSEKNSIFKNKKPNLKFKNITKKLFDPYKDLGFLARNLNDT
RYATKVFRDQLNNYSKHHSKDDENKLFKVVCMNGSITSFLRKSMWRKNEEQVYRFNFWKKDRDQFFHHAVDASIIAIFSLLTKTLYNKLR
VYESYDVQRREDGVYLINKETGEVKKADKDYWKDQHNFLKIRENAIEIKNVLNNVDFQNQVRYSRKANTKLNTQLFNETLYGVKEFENNF
YKLEKVNLFSRKDLRKFILEDLNEESEKNKKNENGSRKRILTEKYIVDEILQILENEEFKDSKSDINALNKYMDSLPSKFSEFFSQDFINKCK
KENSLILTFDAIKHNDPKKVIKIKNLKFFREDATLKNKQAVHKDSKNQIKSFYESYKCVGFIWLKNKNDLEESIFVPINSRVIHFGDKDKDIFD
FDSYNKEKLLNEINLKRPENKKFNSINEIEFVKFVKPGALLLNFENQQIYYISTLESSSLRAKIKLLNKMDKGKAVSMKKITNPDEYKIIEHVN
PLGINLNWTKKLENNN

>gi|284931710|gb|ADC31648.1|/1-1269
MNNSIKSKPEVTIGLDLGVGSVGWAIVDNETNIIHHLGSRLFSQAKTAEDRRSFRGVRRLIRRRKYKLKRFVNLIWKYNSYFGFKNKEDILN
NYQEQQKLHNTVLNLKSEALNAKIDPKALSWILHDYLKNRGHFYEDNRDFNVYPTKELAKYFDKYGYYKGIIDSKEDNDNKLEEELTKYKF
SNKHWLEEVKKVLSNQTGLPEKFKEEYESLFSYVRNYSEGPGSINSVSPYGIYHLDEKEGKVVQKYNNIWDKTIGKCNIFPDEYRAPKNS
PIAMIFNEINELSTIRSYSIYLTGWFINQEFKKAYLNKLLDLLIKTNGEKPIDARQFKKLREETIAESIGKETLKDVENEEKLEKEDHKWKLKGL
KLNTNGKIQYNDLSSLAKFVHKLKQHLKLDFLLEDQYATLDKINFLQSLFVYLGKHLRYSNRVDSANLKEFSDSNKLFERILQKQKDGLFKL
FEQTDKDDEKILAQTHSLSTKAMLLAITRMTNLDNDEDNQKNNDKGWNFEAIKNFDQKFIDITKKNNNLSLKQNKRYLDDRFINDAILSPGV
KRILREATKVFNAILKQFSEEYDVTKVVIELARELSEEKELENTKNYKKLIKKNGDKISEGLKALGISEDEIKDILKSPTKSYKFLLWLQQDHID
PYSLKEIAFDDIFTKTEKFEIDHIIPYSISFDDSSSNKLLVLAESNQAKSNQTPYEFISSGNAGIKWEDYEAYCRKFKDGDSSLLDSTQRSKK
FAKMMKTDTSSKYDIGFLARNLNDTRYATIVFRDALEDYANNHLVEDKPMFKVVCINGSVTSFLRKNFDDSSYAKKDRDKNIHHAVDASII
SIFSNETKTLFNQLTQFADYKLFKNTDGSWKKIDPKTGVVTEVTDENWKQIRVRNQVSEIAKVIEKYIQDSNIERKARYSRKIENKTNISLFN
DTVYSAKKVGYEDQIKRKNLKTLDIHESAKENKNSKVKRQFVYRKLVNVSLLNNDKLADLFAEKEDILMYRANPWVINLAEQIFNEYTENK
KIKSQNVFEKYMLDLTKEFPEKFSEFLVKSMLRNKTAIIYDDKKNIVHRIKRLKMLSSELKENKLSNVIIRSKNQSGTKLSYQDTINSLALMIM
RSIDPTAKKQYIRVPLNTLNLHLGDHDFDLHNMDAYLKKPKFVKYLKANEIGDEYKPWRVLTSGTLLIHKKDKKLMYISSFQNLNDVIEIKNL
IETEYKENDDSDSKKKKKANRFLMTLSTILNDYILLDAKDNFDILGLSKNRIDEILNSKLGLDKIVK

>gi|227540451|ref|ZP_03970500.1|/1-1306
MKHILGLDLGTNSIGWALIKQNFENKQGEILGMGSRIIPMGAELSKFEQGQAQTKNANRRIARGTRKLNKRYKQRRNKLIYILQGLGMLPD
QIKLVNQFSDPNKLDKISILPISKNQQQLTAFDLLELRVKALTEKIELEELGKVIYKYNQLRGYSGGNSEPEKEENNEDESEENEKTNDSFV
VFSKVLSLGLPEEITFKGKKLSKRKIAIETEDGIIEAETFLDSLKENDSLELHVNIRRAKSGDTITIKLPSKTSWRKKMENIEKQLADKSKELG
REYFISELLFDILKENKWAKIRNNVILRSRYESEFDAVWSEQSKYYPILNHLETDRLQKILEFIFPGQKESQEKLRQTGLKEGLKYVIKNQVV
FYQRELKDQSDLISNCRYEPTEKVVAKSHPVFQEYKIWEQINKLVINTKIENGVNRKGEIRYKYIDKPIPVALKEWLYEELQSKKEVSFNPIL
TKLKKEFKLRENVDFLNGMNAKDKLKGNETKVILQKALGELWNIFGLDDSERQVELWEILYNGKGNEYDLTSERTSNILAFIEKYSPDLEN
KNDVAIKISKIKFSRNYSNLSLKAIDNILPLVRAGKFFDQNFSEALSHKIVKLLNENVEDPFEKSAQEFLDKNQDVLSDGGIVNAIATILVYDK
HTAKEYSGDELIKEFTQISRLKQGELRNPLAEQIINEALVIVRDIWRKYGFKPDEIRVELARELKNNAVERGKMYSANLKSQKHNAEIRNKLI
ELKQEISIANIEKYKLWVSQENLQEVYLKNYNDPSKSEIEKLKLWESQGHISPYTGQPISLSELFNRERYDVDHIIPQSRYFDDSLSNKVICE
KSINNEKGNRTSMEYFEVGSVNYRIFTKDQFIDHANKYFSGKKRKNLLATSIPEDPIQRQIKDTQYISIRIREELNKIVGNENVKTTTGGITDY
LRNHWGLTDKFKVLLKERYEAVLPKIAQLEYDNYKKETEKKFTEYKNADVEFNEPILDEEIFIEKFRSEFIRKKNNKLIIKGWSKRIDHRHHA
MDALVIACTQPAHIKRLNDLNQVLKEWLDKNKEKFFPGFSGTKTELLDEVLNLEEEKRREIFNQLEKFRAIEMPWKGFPEEAEQKLKEVIV
SHKPKDKLLLQYDIQGNKQIKLRGQLHEGTLYGLSNGKEIYRIPINKLAGKQFATEKTIEKIVNPFLKNVIEEHLKSFGKKEEAFSAEGILDLN
KKLAERKDEQQNPKPHPPISTIKIYYKDPSKSKKKKTDDEDDISLQKLNRKKSFNDSLYVKTGDNYLFAVMEKEVFNKKLNKDKTERHYDL
KLFLTLLIF

FIGURE 12 P.

>gi|150025575|ref|YP_001296401.1|/1-1354
MKKILGLDLGTNSIGWAFINEAENEIEVSSIVNTGVRIVPLTTDEESDFKKGNTISINADRTLKRGARRGLQRFKQRREALLETFNRIKFIATD
FVYAETGAATTFSSYQLRAKSAVEAVTKEELVKVLLMLNKKRGYKSSRKAKTAEEGDAIDGMKIAKEIFENNLTPGQWVYNSLSKGGKFIP
DFYHSDLQKELEKIVRFQNQFHSEQINDKLLADIQGKSRTQTSQHFSKTLNIPLAENKGTREETKLQHYQWRNEALANELDMKILAFIITEIN
NQINQSSGYLGAISDRSKELYFNQETVGQYQYKQLQENVHAKLKNQVFYRQDYLDEFEKIWETQAKFHSELNEKLKAEIRDITIFYQRKLK
SQKHLVSFCEFEKGHKAIPKSSPLFQEFRIWQNLNNVVVKNETTKEVFELSEELKQLIAEELQFKESMTDVQMLTFCELKKGTHTVNFKKI
EGNRTNTSIYKAFEKILEVEGYEIDFSKMKPLEIKETVAPIFETLNINTVILDFDPTIQGNDFDKQPYYQLWHLLYSAEDDDKLKETLKTKFGF
KENHLPFILNINLQADYGSLSAKAIKKILPHLMDGHIYDKACLMVGYNHSSSLTTAQNNDRVLKDTLELLKKNSLRNPVVEKILNQMINVINAI
MKDPAMGKPDEIRVELARELKANNEQRKKATSDINTATTEHERIRKLLHSEFNIPRVTRNDIIRYKLWLETGGISLYTGNPIKASDLFSKNYD
IEHIIPKSRLFDDSFSNKTLCERQLNIDKSNKTAYSFLQEKLSENDFAAYENRVKDLFKTGKIKSYTKYKKLLMADNEIPEGFIDRQLRETQYI
AKKAKEILLEVSRNVTATIGSVTDKLRQDWELVDVMKELNWEKYDKLGLTHIEEGKNGERLSKIKDWTKRNDHRHHAMDAITVAFTKPAYI
QYLNNLNAKTQGENKANSIIGIENKHLYRDKNNKLRFISPMKNFREEAKKQLESILISYKAKNKVVTKNKNTTKKSGGTNQKIQLTPRGRLH
KETVYGKLQQYATKEEKVNASFTEDYIQKVAKKEYREALLKRLEENDNDPKKAFTGKNALNKTPIYISLKDNIIVPEKVKTVWLETDYTIRK
DITPDLKIDKVIDVGIKRILQNRLNEFNGDPKKAFVNLGESPIWLNKEKGIAIKRVTISGVSNAQALHTKKDHLGNEILDKNGNPIPVDFISTG
NNHHVAIYRDEKGNLQEEVVSFYDAVIRRNLGLSVINKNHEKGWEFLFSMKQNEFFIFPSDGFNPHEIDLLNPNNYHLISPNIFRVQKISTK
NYMFNHHLETKAVDGEMLKSKKELSKTSYHFIQTPTNLNGIIKVRINHLGKIVHLGEY

>gi|301164874|emb|CBW24435.1|/1-1436
MKRILGLDLGTNSIGWALVNEAENKDERSSIVKLGVRVNPLTVDELTNFEKGKSITTNADRTLKRGMRRNLQRYKLRRETLTEVLKEHKLI
TEDTILSENGNRTTFETYRLRAKAVTEEISLEEFARVLLMINKKRGYKSSRKAKGVEEGTLIDGMDIARELYNNNLTPGELCLQLLDAGKKF
LPDFYRSDLQNELDRIWEKQKEYYPEILTDVLKEELRGKKRDAVWAICAKYFVWKENYTEWNKEKGKTEQQEREHKLEGIYSKRKRDEA
KRENLQWRVNGLKEKLSLEQLVIVFQEMNTQINNSSGYLGAISDRSKELYFNKQTVGQYQMEMLDKNPNASLRNMVFYRQDYLDEFNM
LWEKQAVYHKELTEELKKEIRDIIIFYQRRLKSQKGLIGFCEFESRQIEVDIDGKKKIKTVGNRVISRSSPLFQEFKIWQILNNIEVTVVGKKR
KRRKLKENYSALFEELNDAEQLELNGSRRLCQEEKELLAQELFIRDKMTKSEVLKLLFDNPQELDLNFKTIDGNKTGYALFQAYSKMIEMS
GHEPVDFKKPVEKVVEYIKAVFDLLNWNTDILGFNSNEELDNQSYYKLWHLLYSFEGDNTPTGNGRLIQKMTELYGFEKEYATILANVSF
QDDYGSLSAKAIHKILPHLKEGNRYDVACVYAGYRHSESSLTREEIANKVLKDRLMLLPKNSLHNPVVEKILNQMVNVINVIIDIYGKPDEIR
VELARELKKNAKEREELTKSIAQTTKAHEEYKTLLQTEFGLTNVSRTDILRYKLYKELESCGYKTLYSNTYISREKLFSKEFDIEHIIPQARLF
DDSFSNKTLEARSVNIEKGNKTAYDFVKEKFGESGADNSLEHYLNNIEDLFKSGKISKTKYNKLKMAEQDIPDGFIERDLRNTQYIAKKALS
MLNEISHRVVATSGSVTDKLREDWQLIDVMKELNWEKYKALGLVEYFEDRDGRQIGRIKDWTKRNDHRHHAMDALTVAFTKDVFIQYFN
NKNASLDPNANEHAIKNKYFQNGRAIAPMPLREFRAEAKKHLENTLISIKAKNKVITGNINKTRKKGGVNKNMQQTPRGQLHLETIYGSGK
QYLTKEEKVNASFDMRKIGTVSKLAYRDALLKRLYENDNDPKKAFAGKNSLDKQPIWLDKEQMRKVPEKVKIVTLEAIYTIRKEISPDLKVD
KVIDVGVRKILIDRLNEYGNDAKKAFSNLDENPIWLNKEKGISIKRVTISGISNAQSLHVKKDKDGKPILDENGRNIPVDFVNTGNNHHVAVY
YRPVIDKRGQLVVDEAGNPKYELEEVVVSFFEAVTRANLGQPIIDKDYKTTEGWQFLFSMKQNEYFVFPNEKTGFNPKEIDLLDVENYGLI
SPNLFRVQKFSLKNYVFRHHLETTIKDTSSILRGITWIDFRSSKGLDTIVKVRVNHIGQIVSVGEY

>gi|256840409|ref|ZP_05545917.1|/1-1424
MKRILGLDLGTNSIGWALVNEAETDEEVSSIIRLGVRVNPLTVDETQNFEKGKSITTNAERTLKRSMRRNLQRYKLRRDALVEFLKELGFIT
DETILSEQGNRTTFETCRLRAKAAEEEISLEELSRVLLMINKKRGYKSSRKVKKEEDGVLIDGMEIAKKLYEEDLTPGQLCSQLLESGKKVL
PDFYRSDLQEELDRIWSVQKQFHPDVFCDAAKEEIKGKNRSQTWAILANYFVWKEEVVGWNDREAKNETIEKECKLVGLKRTTKGYELK
KENYRWRAQALTEQLGLEEIAVVLQEINGQINASSGYLGAISDRSKVLYFNHQTVGQYQMAELDKNPHASLRNMVFYRQDYLDEFDRIW
EKQAEFHKELTAELKKEIRDMIIFYQRRLKSQKGLISFCEFERLEIVVEKDGKKQTKVIGCKVIPRSHPLFQEFKVWQTLNDIKVLGREKKRN
ALSASESRALYPEEKEILARELAIKENMKKADVLKLLFENPQELDLNFKQIDGNRTGFSLFSAYSKMIEKYGYEPLDFKKPADEIIGQLRTIFT
DLGWNTNLFTIDLTKEGKELEMQPYFRLWHLLYSFEGDNTSTGNGKLIEKIMQLCGVEKEYAVELASLSFQDDYGSLSAKAIKRILPYLKE
GNLYDVACEYAGYRHSKSSLTKEEIENKVLKGKLEVLPKNSLRNPVVEKILSQMVNVINTIIDTYGKPDEIRVELARELKKSAKEREELTKAI
AKSTREHEEIRKLLQDEFGMMNVSRNDIIRYKLYEELKDNGYKTLYSNQYIPKEKIFSKEIDIEHIIPQSRLFDDSLSNKTLEYKAINIEKGNKT
AYDFVKEKYGEEGLQQYINRCESLFNDKKAKLRKLKMEQKDIPDGFIDRDLRNTQYIAKKALAMLNEICRRVVATTGSITDQLREDWQLVD
VMKELNLPKYEALGFVETYEDKDGRKIKRIKDWTKRNDHRHHAMDALTVAFTKDAFIQYFNNKNAAQDSGNRNHTNIIGIKTRYFDKGRA
LPPIPLEQFRIEAKQHLEKLLVSIKAKNKVVTVNVNRTKKRKGENTKIQQTPRGQLHLETVYGSHKQYVTKIEKVNASFDAAKIATVSKRAY
RNALLKRLETFGNDPKKAFTGKNALEKNPLYADKYQTIRVPEKVQTVEFETIYTIRKPVDPALNVDKVVDVKVRAILERRLKEYGGDPKKA
FVNFWENPIWLNKEKGISIKRVSIRGINNAQSIHVKKDKNGNPIWDENGKQIPVDFVNTGNNHHVAIYRKPVLDKKGQVTFDEDGNLMYEL
DEVVVPFFEAVTRANLGLPIIDKDYRKSEGWQFLFSMKQNEYFVFPNEKTGFNPKEVDLLNPDNYAMISPNLFRVQTMSKVMYGNNVVR
DYKFRHHLETTVKDMKELKDIAYKQYKTLSFGNSVVKIRINHIGQIVSVGEY

FIGURE 12 Q.

>gi|163754820|ref|ZP_02161941.1|/1-1391
MKKILGLDLGTNSIGWALVNEAEHEHEQSSIVKLGVRVNPLSVDEKTNFEKGRPLSINADRTSKRSARRNLQRFKLRRKNLINLLLKHHIIS
NETILTEVGKNTTHQTLKLRSKATQEKISLEEFARILLSINKKRGYKSSRKEKNEDEGFTIDGMAIAKELYDNQLTPGQYVYNILQDGKNYIP
DFYRSDLQAEFDDVWQFQKKFYSSILTDEFYLKLKGQGKENTRKLFLAIHQVYTAENKGKRDAVKLQYYKWRSEAVKKQLSIEEVAYVLV
EINNNINKSSGYLGAISDRSKALYFNNETVGEYLYGGQLVKNRHNSLRNQVFYRQDYLDEFEKIWETQAKFYDELTNELKQEIRDVVIFYQR
KLKSQKHLISDCLFEKHHKAIPKSSPLFQEFKIWQNLNYLEFVHTQSKEKIFFSELDDETRQFVFAELNIRGTLKIKEIKKLIGLETPVYWKSN
FDEIEGNRTNQVLYNVYQNIADDLGYGFDWSKKHAFEINAELQAIFKDVGIDTAILNFDASIEENDFDKQKSYQLWHLLYSAEDTDKATEE
DKLIYGNINTKLKKQLHQKYGFEAKYGNLLANISLQQDYGNLSAKAIKKILPYLESGNSFYEACTIVGYNPSNSLTKEQQQNRKLKDKLELL
PKNSLRNPVVEKILNQKVNLINQLIDEYGKPDEVRIELARELKKSAKQRADLSKHINDATLKNQNIRKTLQKDFGIQNPSKNDVIRYKLYDEL
KDNGYKTLFTNKYIPKEQLFSKDVDIEHIIPKAILFDDSFSNKTLAYRNVNLAKADATAFDFISSKHLDDLDNYKARVETLFQSKRISKGKYK
KLLMTRQNLPEDFIERDLRNSQYIAKKATAMLKEVFRSVVPTSGRITDKLREDWDLVNVMKELNLPKYRTLGLTEFIERKNGNKIEQIIDWT
KRNDHRHHAMDALTVAFTTHNHIQYLNNLNVTHGLNGVDTDHEKSRLYAIKQAITKVYIVGKTKKRKFIPPMENFRSHAKNHLEGILISFKN
KNKVVTKNKNRTKTKTKGVFNEKIQLTPRGQLHKETIYGHAKKVIVTKEKVSSKFTHEKILQVTKPAYREALLKRLASFKNDPKKAFSGKNA
LSKNPLYFDITSGKKVPDTVKLSYFENSYVIRKEVNSDNFKNEKSLQKVLDVGIRRILEKRLADYNGNAKEAFSNLEKNPIWKDKEKGVCIK
KVRIEALKNAEPLHVKKDHLGKIILDDENKSTPVDFISTGNNHHVAIYEDKDGKLHDKVVSFYEVVSRVNLGLPIIDKSFNSHLDWKFLFSM
KQNEMFVFPSEDFNPKEIDIGNPKNNELISKYLFRVQKFSKVTYGNSAVRDYVFRHHLETELIDKKELKGITYYQLKSIGMLEKLVKVRINHI
GKIVHIGEY

>gi|256819408|ref|YP_003140687.1|/1-1426
MKNILGLDLGTTSIGFAHIVEDENKEKSEIKELGVRIVSLTTDEQSDFEKGKSITTNANRTLKHGARLNLDRYQQRRKYLIDLLQKANLITPS
SILAENGKNTTHSTWQLRAKAVTERIEKEEFARILLAINKKRGYKSSRKAKTEDEGQAIDGMAIAKRLYDENLTPGQLSLQLLQQNKKLLPD
FYRSDLQKEFDLVWNFQKQFYPDILTDIFYKELQGKGKDATSKAFSKRYHFDTTENKGSKESVRLQAYQWRAEAISKQLSKEEVAYVLTE
INNNLNNASGYLGAISDRSKELYFNRQTVGQYLYAKLQENRHNSLKNKVFYRQDYLDEFERIWETQASFHKELTDELKKQIRDVVIFYQRK
PKSQKGLISFCEFESKEIEIEKDGKTITKNIGARVVPKSSPLFQEFKIWQILNNVICKRKGIRKKKISAKTTQLDLLNESSQTIFSLDMECKQLL
FDELNLKGDLKSDKVLKLLGYSPQEWEINYNQLEGNRTQKALYEAYLKIVEMEAHDVKDILQIKSAKDDWSLDESPLSASEIREKVKAIFQT
LGICTKILYFDPLLPVKEFEEQDSYQLWHLLYSYESDDSTSGNETLYRILEKKYAFKREHARILANVALQDDYGSLSTKAIRKIYPNIKENQY
STACEKAGYKHSKLSLTTEELEARELKNIIPLLKKNALRNPVVEKILNQMINVVNALIEKNSERDAEGKITKYFHFDEIRIELARELKKNAQKR
YEMTQNINKAKLEHQKISEILQKEFGIKNPTKSDIIRYRLYQELEHNGYKELYTNAPIARDMLFSKNIEIEHIVPKARVFDDSFSNKTLTFHRIN
SDKGEYTAFDYITSLNSEEELNQYLTRVENAYKTKSISPTKYKNLLKKASEIGDDFINRDLRDTQYIAKKAKEILFQVTKNVLSTSGSITDRLR
EDWGLVDVMKELNMPKYQSLGLTEVEERKDGNKVTVIKNWTKRNDHRHHAMDALTVAFTKPSYIQYLNHLNARKDENNKNYSVILAIEE
KETIKVPTNNGKNKRVFIEPIPNFRQVAKKHLEEIFISHKAKNKVVTKNTNKPAGTDKQQITLTPRGQLHKETIYGKYQYYINKEEKIGVKFD
ERTIAKVSNPVYREALLKRLQANDNDPKKAFAGKNALSKNPIYLDESKTKTLPEKVNLTYLEEDFSIRKDISPDNFKDLKSIEKVIDQGVKRI
LIKRLQAYDNDPKKAFVDLEKNPIWLNKEKGIAIKRVTISGVNNAQPLHIGKDHLGKTTLNKEGKEIPVDYVSTGNNHHVAIYRDKEGNLQE
QIVSFFDAVVRAQQGIPIIDKTYKQAEGWQFLFTMKQNEMFVFPNATTGFNPAEIDLLDPKNKKLISPNLFRVQKIATKDYFFRHHLETNVE
TDNILKNVTWKREGLSGLKDIVKVRINHLGDIVSIGEY

>gi|154250555|ref|YP_001411379.1|/1-1037
MERIFGFDIGTTSIGFSVIDYSSTQSAGNIQRLGVRIFPEARDPDGTPLNQQRRQKRMMRRQLRRRRIRRKALNETLHEAGFLPAYGSAD
WPVVMADEPYELRRRGLEEGLSAYEFGRAIYHLAQHRHFKGRELEESDTPDPDVDDEKEAANERAATLKALKNEQTTLGAWLARRPPS
DRKRGIHAHRNVVAEEFERLWEVQSKFHPALKSEEMRARISDTIFAQRPVFWRKNTLGECRFMPGEPLCPKGSWLSQQRRMLEKLNNL
AIAGGNARPLDAEERDAILSKLQQQASMSWPGVRSALKALYKQRGEPGAEKSLKFNLELGGESKLLGNALEAKLADMFGPDWPAHPRK
QEIRHAVHERLWAADYGETPDKKRVIILSEKDRKAHREAAANSFVADFGITGEQAAQLQALKLPTGWEPYSIPALNLFLAELEKGERFGAL
VNGPDWEGWRRTNFPHRNQPTGEILDKLPSPASKEERERISQLRNPTVVRTQNELRKVVNNLIGLYGKPDRIRIEVGRDVGKSKRERE EI
QSGIRRNEKQRKKATEDLIKNGIANPSRDDVEKWILWKEGQERCPYTGDQIGFNALFREGRYEVEHIWPRSRSFDNSPRNKTLCRKDVNI
EKGNRMPFEAFGHDEDRWSAIQIRLQGMVSAKGGTGMSPGKVKRFLAKTMPEDFAARQLNDTRYAAKQILAQLKRLWPDMGPEAPVK
VEAVTGQVTAQLRKLWTLNNILADDGEKTRADHRHHAIDALTVACTHPGMTNKLSRYWQLRDDPRAEKPALTPPWDTIRADAEKAVSEI
VVSHRVRKKVSGPLHKETTYGDTGTDIKTKSGTYRQFVTRKKIESLSKGELDEIRDPRIKEIVAAHVAGRGGDPKKAFPPYPCVSPGGPEI
RKVRLTSKQQLNLMAQTGNGYADLGSNHHIAIYRLPDGKADFEIVSLFDASRRLAQRNPIVQRTRADGASFVMSLAAGEAIMIPEGSKKGI
WIVQGVWASGQVVLERDTDADHSTTTRPMPNPILKDDAKKVSIDPIGRVRPSND

>gi|121608211|ref|YP_996018.1|/1-1068
MNKAVAYRLALDLGSTSLGWAIFRLNEAREPTAIIRAGVRIFSDGRNANSEPLAVHRRVARAMRRRRDRLLKRKKRMHDQLVQHGFFPA
EVAERKELERLNPYQLRAKGLHEALTPGEFARALFHINQRRGFKSNRKTDRKDNDSGALKQAISELRKRIQDSDCATAGEWFWKERMQ
QKPEGVRGQWVRARYRKTPSTTDEGKKRIGYDLYVDRAMVEQEFDALWAAQAALQPGLFTEAARAELKDTLLYQRDLRPVKPGRCTLL
PAEERAPLALPSTQRFRILQEVNNLRLLDEALREVPLNLAQRDAVVSALEGKKELSFAAIRKLLTLSGKFNLEDEKRAELKGNATSVILARK
DLFGDAWAGFDAAVQDEIVWRLVSQESEGALIAWLQQHTGVDGVCAEAIVNTRLPDGYGRLSRKALERIVPALQREVCTYDKAVQAAGF
AHHSDLGFDFDYAEDEVQQVGEHTIASTGEVQAQYAFKQLPYYGKALQRHVAFGSGDAKDHEEKCYGKIANPTVHIGLNQVRTVVNALI
RRYGHPTEVVVELARDLKQSREQKQQTQREQADNQKRNEDIRKRIAPILETSPERVRDRDIKKWILWEELNKKDIADRHCPYSGERISAT
MLLSEAVEIEHILPFSRTLDDSLNNRTVAMRRANRIKGDRTPWEARADFEAQGWRYEAILQRAEGMPPRKRYRFAEDGYQRWLGKDRN
FLARALNDTSYLSRLAANYLRLVCPQGVRVIPGQMTDKLRGKFGLNSVLGLDGKKNRNDHRHHAVDACVIGVTDQGLMQRFANASKQA
RENGLTRLVQDMLLPWWPSYYDHVERAVRHIRVSHRPDHGFEGAMMKGTAHGIREDGIREDGRIKQRPKAKGSADHKTITLIPIDEPRQ
LARHGVDAEGKPLPYKGYASGSNYCIEITKNGKGKWEGQVISTFDAYRIKAAADAAARAGQVISTVEADSIVRKSGWERLRGAQSQNGQ
PLVMRLVIGDSVRMEVDGRDEVMRVVKMSGKEMVFAPVREANVDKRNNMPDEQDPFTYTYKRADQLRKAKARQVTISPIGELRDPGFK
G

FIGURE 12 R.

>gi|148255343|ref|YP_001239928.1|/1-1064
MKRTSLRAYRLGVDLGANSLGWFVVWLDDHGQPEGLGPGGVRIFPDGRNPQSKQSNAAGRRLARSARRRRDRYLQRRGKLMGLLVK
HGLMPADEPARKRLECLDPYGLRAKALDEVLPLHHVGRALFHLNQRRGLFANRAIEQGDKDASAIKAAAGRLQTSMQACGARTLGEFLN
RRHQLRATVRARSPVGGDVQARYEFYPTRAMVDAEFEAIWAAQAPHHPTMTAEAHDTIREAIFSQRAMKRPSIGKCSLDPATSQDDVD
GFRCAWSHPLAQRFRIWQDVRNLAVVETGPTSSRLGKEDQDKVARALLQTDQLSFDEIRGLLGLPSDARFNLESDRRDHLKGDATGAIL
SARRHFGPAWHDRSLDRQIDIVALLESALDEAAIIASLGTTHSLDEAAAQRALSALLPDGYCRLGLRAIKRVLPLMEAGRTYAEAASAAGY
DHALLPGGKLSPTGYLPYYGQWLQNDVVGSDDERDTNERRWGRLPNPTVHIGIGQLRRVVNELIRWHGPPAEITVELTRDLKLSPRRLA
ELEREQAENQRKNDKRTSLLRKLGLPASTHNLLKLRLWDEQGDVASECPYTGEAIGLERLVSDDVDIDHLIPFSISWDDSAANKVVCMRY
ANREKGNRTPFEAFGHRQGRPYDWADIAERAARLPRGKRWRFGPGARAQFEELGDFQARLLNETSWLARVAKQYLAAVTHPHRIHVL
PGRLTALLRATWELNDLLPGSDDRAAKSRKDHRHHAIDALVAALTDQALLRRMANAHDDTRRKIEVLLPWPTFRIDLETRLKAMLVSHKP
DHGLQARLHEDTAYGTVEHPETEDGANLVYRKTFVDISEKEIDRIRDRRLRDLVRAHVAGERQQGKTLKAAVLSFAQRRDIAGHPNGIRH
VRLTKSIKPDYLVPIRDKAGRIYKSYNAGENAFVDILQAESGRWIARATTVFQANQANESHDAPAAQPIMRVFKGDMLRIDHAGAEKFVKI
VRLSPSNNLLYLVEHHQAGVFQTRHDDPEDSFRWLFASFDKLREWNAELVRIDTLGQPWRRKRGLETGSEDATRIGWTRPKKWP

>gi|122416577|sp|Q1QGC9|Q1QGC9_NI/1-1166
MHVEIDFPHFSRGDSHLAMNKNEILRGSSVLYRLGLDLGSNSLGWFVTHLEKRGDRHEPVALGPGGVRIFPDGRDPQSGTSNAVDRRM
ARGARKRRDRFVERRKELIAALIKYNLLPDDARERRALEVLDPYALRKTALTDTLPAHHVGRALFHLNQRRGFQSNRKTDSKQSEDGAIK
QAASRLATDKGNETLGVFFADMHLRKSYEDRQTAIRAELVRLGKDHLTGNARKKIWAKVRKRLFGDEVLPRADAPHGVRARATITGTKA
SYDYYPTRDMLRDEFNAIWAGQSAHHATITDEARTEIEHIIFYQRPLKPAIVGKCTLDPATRPFKEDPEGYRAPWSHPLAQRFRILSEARN
LEIRDTGKGSRRLTKEQSDLVVAALLANREVKFDKLRTLLKLPAEARFNLESDRRAALDGDQTAARLSDKKGFNKAWRGFPPERQIAIVA
RLEETEDENELIAWLEKECALDGAAAARVANTTLPDGHCRLGLRAIKKIVPIMQDGLDEDGVAGAGYHIAAKRAGYDHAKLPTGEQLGRL
PYYGQWLQDAVVGSGDARDQKEKQYGQFPNPTVHIGLGQLRRVVNDLIDKYGPPTEISIEFTRALKLSEQQKAERQREQRRNQDKNKA
RAEELAKFGRPANPRNLLKMRLWEELAHDPLDRKCVYTGEQISIERLLSDEVDIDHILPVAMTLDDSPANKIICMRYANRHKRKQTPSEAF
GSSPTLQGHRYNWDDIAARATGLPRNKRWRFDANAREEFDKRGGFLARQLNETGWLARLAKQYLGAVTDPNQIWVVPGRLTSMLRGK
WGLNGLLPSDNYAGVQDKAEEFLASTDDMEFSGVKNRADHRHHAIDGLVTALTDRSLLWKMANAYDEEHEKFVIEPPWPTMRDDLKAA
LEKMVVSHKPDHGIEGKLHEDSAYGFVKPLDATGLKEEEAGNLVYRKAIESLNENEVDRIRDIQLRTIVRDHVNVEKTKGVALADALRQLQ
APSDDYPQFKHGLRHVRILKKEKGDYLVPIANRASGVAYKAYSAGENFCVEVFETAGGKWDGEAVRRFDANKKNAGPKIAHAPQWRDA
NEGAKLVMRIHKGDLIRLDHEGRARIMVVHRLDAAAGRFKLADHNETGNLDKRHATNNDIDPFRWLMASYNTLKKLAAVPVRVDELGRV
WRVMPN

>gi|288957741|ref|YP_003448082.1|/1-1168
MARPAFRAPRREHVNGWTPDPHRISKPFFILVSWHLLSRVVIDSSSGCFPGTSRDHTDKFAEWECAVQPYRLSFDLGTNSIGWGLLNLD
RQGKPREIRALGSRIFSDGRDPQDKASLAVARRLARQMRRRRDRYLTRRTRLMGALVRFGLMPADPAARKRLEVAVDPYLARERATRE
RLEPFEIGRALFHLNQRRGYKPVRTATKPDEEAGKVKEAVERLEAAIAAAGAPTLGAWFAWRKTRGETLRARLAGKGKEAAYPFYPARR
MLEAEFDTLWAEQARHHPDLLTAEAREILRHRIFHQRPLKPPPVGRCTLYPDDGRAPRALPSAQRLRLFQELASLRVIHLDLSERPLTPAE
RDRIVAFVQGRPPKAGRKPGKVQKSVPFEKLRGLLELPPGTGFSLESDKRPELLGDETGARIAPAFGPGWTALPLEEQDALVELLLTEAE
PERAIAALTARWALDEATAAKLAGATLPDFHGRYGRRAVAELLPVLERETRGDPDGRVRPIRLDEAVKLLRGGKDHSDFSREGALLDALP
YYGAVLERHVAFGTGNPADPEEKRVGRVANPTVHIALNQLRHLVNAILARHGRPEEIVIELARDLKRSAEDRRREDKRQADNQKRNEER
KRLILSLGERPTPRNLLKLRLWEEQGPVENRRCPYSGETISMRMLLSEQVDIDHILPFSVSLDDSAANKVVCLREANRIKRNRSPWEAFG
HDSERWAGILARAEALPKNKRWRFAPDALEKLEGEGGLRARHLNDTRHLSRLAVEYLRCVCPKVRVSPGRLTALLRRRWGIDAILAEAD
GPPPEVPAETLDPSPAEKNRADHRHHALDAVVIGCIDRSMVQRVQLAAASAEREAAAREDNIRRVLEGFKEEPWDGFRAELERRARTIV
VSHRPEHGIGGALHKETAYGPVDPPEEGFNLVVRKPIDGLSKDEINSVRDPRLRRALIDRLAIRRRDANDPATALAKAAEDLAAQPASRGI
RRVRVLKKESNPIRVEHGGNPSGPRSGGPFHKLLLAGEVHHVDVALRADGRRWVGHWVTLFEAHGGRGADGAAAPPRLGDGERFLM
RLHKGDCLKLEHKGRVRVMQVVKLEPSSNSVVVVEPHQVKTDRSKHVKISCDQLRARGARRVTVDPLGRVRVHAPGARVGIGGDAGRT
AMEPAEDIS

>gi|159042956|ref|YP_001531750.1|/1-1079
MRLGLDIGTSSIGWWLYETDGAGSDARITGVVDGGVRIFSDGRDPKSGASLAVDRRAARAMRRRRDRYLRRRATLMKVLAETGLMPAD
PAEAKALEALDPFALRAAGLDEPLPLPHLGRALFHLNQRRGFKSNRKTDRGDNESGKIKDATARLDMEMMANGARTYGEFLHKRRQKA
TDPRHVPSVRTRLSIANRGGPDGKEEAGYDFYPDRRHLEEEFHKLWAAQGAHHPELTETLRDLLFEKIFFQRPLKEPEVGLCLFSGHHG
VPPKDPRLPKAHPLTQRRVLYETVNQLRVTADGREARPLTREERDQVIHALDNKKPTKSLSSMVLKLPALAKVLKLRDGERFTLETGVRD
AIACDPLRASPAHPDRFGPRWSILDADAQWEVISRIRRVQSDAEHAALVDWLTEAHGLDRAHAEATAHAPLPDGYGRLGLTATTRILYQL
TADVVTYADAVKACGWHHSDGRTGECFDRLPYYGEVLERHVIPGSYHPDDDDITRFGRITNPTVHIGLNQLRRLVNRIIETHGKPHQIVVE
LARDLKKSEEQKRADIKRIRDTTEAAKKRSEKLEELEIEDNGRNRMLLRLWEDLNPDDAMRRFCPYTGTRISAAMIFDGSCDVDHILPYSR
TLDDSFPNRTLCLREANRQKRNQTPWQAWGDTPHWHAIAANLKNLPENKRWRFAPDAMTRFEGENGFLDRALKDTQYLARISRSYLDT
LFTKGGHVWVVPGRFTEMLRRHWGLNSLLSDAGRGAVKAKNRTDHRHHAIDAAVIAATDPGLLNRISRAAGQGEAAGQSAELIARDTPP
PWEGFRDDLRVRLDRIIVSHRADHGRIDHAARKQGRDSTAGQLHQETAYSIVDDIHVASRTDLLSLKPAQLLDEPGRSGQVRDPQLRKAL
RVATGGKTGKDFENALRYFASKPGPYQAIRRVRIIKPLQAQARVPVPAQDPIKAYQGGSNHLFEIWRLPDGEIEAQVITSFEAHTLEGEKR
PHPAAKRLLRVHKGDMVALERDGRRVVGHVQKMDIANGLFIVPHNEANADTRNNDKSDPFKWIQIGARPAIASGIRRVSVDEIGRLRDGG
TRPI

FIGURE 12 S.

>gi|123763007|sp|Q13CC2|Q13CC2_RH/1-1064
MSECVTRRILGIDLGIASCGWGVIEVGEASGSIIASGVRCFDAPLIDKTGEPKSATRRTARGQRRIIRRRRQRMNAVRRLLAEFGVLTGRS
PDALHQALLRLSQSVAGSQVTPWTLRAAAHERKLTNDELAVVLGHIARHRGFRSNSKNDGGANAADETSKMKKAMETTREGLARYHSF
GAMIASDPKFADRKRNRDKDYSHTAKRSDLEDEVRTIFRSQTRFGSLVASEKLSQAFADAAFFQRPLQDSEDMVGSCPFEPGQKRTAR
RAPSFELFRFLSRLANLKLTVGRAPERRLTPDEIALAAKGFGETKKSITFKSLREALDLDPNARFSGVAKEKESTLDVAARTGGAAYGTKT
LKDALGDAPWRSLSRMPEKLDRIAEILSFREDMKAIRNGLEEVGLDGLVVDALMQATANGDFKDFTRAAHISALAARNIIPGLREGLVYSD
ACTRVGYDHAARPAVPLSQIGSPVTRKALSEALKQVRAVAREYGPIDYFHIELARSIGKSAEERKKLTDGIEARNVEKEKRRKEAAEHLGR
APSDDELLRYELAKEQNFKCIYSGDPIDPAGISANDTRYQVDHILPWSRFGDDSYVNKTLCTARSNQNKRGRTPFEWFDADKTEAEWME
YSARVEDLKEVKGRKKRNYSIKDAASVEDKFKARNLTDTQWATRLLADELKRMFPPRECERVVTVRADGGNDGLSIVEERRVFTRPGAI
TSKLRRAWGLEGLKKQDGKRVEDDRHHAVDALVLAATTESLLNRLTVEVQQREREGRQDDIFHCSQPWPGFRVDVQRTVYGSETMPGI
FVSRAERRRARGKAHDATVKQIRDIDGERIVFERKPIEKLTDKDLERIPVPEPYGKAADPKKLRDELVENLRAWIAAGKPKDKPPRSPKGD
IIRKVRIETKDKVAVEINGGTVDRGDMARVDVFRKKNKKGVWEFYVIPIYPHQIVASALPPNRAVIAYKAESEWTAIDGCFEFAWSLNPMS
YLELVKSNGELIEGYFRSMDRTTGAINLSPMSTNSETIRSIGVKTLSSFRKFTVDRLGRKFEIPREVRTWRGEACT

>gi|81833332|sp|Q7MRD3|Q7MRD3_WOL/1-1059
MIERILGVDLGISSLGWAIVEYDKDDEAANRIIDCGVRLFTAAETPKKKESPNKARREARGIRRVLNRRRVRMNMIKKLFLRAGLIQDVDLD
GEGGMFYSKANRADVWELRHDGLYRLLKGDELARVLIHIAKHRGYKFIGDDEADEESGKVKKAGVVLRQNFEAAGCRTVGEWLWRER
GANGKKRNKHGDYEISIHRDLLVEEVEAIFVAQQEMRSTIATDALKAAYREIAFFVRPMQRIEKMVGHCTYFPEERRAPKSAPTAEKFIAIS
KFFSTVIIDNEGWEQKIIERKTLEELLDFAVSREKVEFRHLRKFLDLSDNEIFKGLHYKGKPKTAKKREATLFDPNEPTELEFDKVEAEKKA
WISLRGAAKLREALGNEFYGRFVALGKHADEATKILTYYKDEGQKRRELTKLPLEAEMVERLVKIGFSDFLKLSLKAIRDILPAMESGARYD
EAVLMLGVPHKEKSAILPPLNKTDIDILNPTVIRAFAQFRKVANALVRKYGAFDRVHFELAREINTKGEIEDIKESQRKNEKERKEAADWIAE
TSFQVPLTRKNILKKRLYIQQDGRCAYTGDVIELERLFDEGYCEIDHILPRSRSADDSFANKVLCLARANQQKTDRTPYEWFGHDAARWN
AFETRTSAPSNRVRTGKGKIDRLLKKNFDENSEMAFKDRNLNDTRYMARAIKTYCEQYWVFKNSHTKAPVQVRSGKLTSVLRYQWGLE
SKDRESHTHHAVDAIIIAFSTQGMVQKLSEYYRFKETHREKERPKLAVPLANFRDAVEEATRIENTETVKEGVEVKRLLISRPPRARVTGQ
AHEQTAKPYPRIKQVKNKKKWRLAPIDEEKFESFKADRVASANQKNFYETSTIPRVDVYHKKGKFHLVPIYLHEMVLNELPNLSLGTNPEA
MDENFFKFSIFKDDLISIQTQGTPKKPAKIIMGYFKNMHGANMVLSSINNSPCEGFTCTPVSMDKKHKDKCKLCPEENRIAGRCLQGFLDY
WSQEGLRPPRKEFECDQGVKFALDVKKYQIDPLGYYYEVKQEKRLGTIPQMRSAKKLVKK

>gi|291276265|ref|YP_003516037.1|/1-1024
MIRTLGIDIGIASIGWAVIEGEYTDKGLENKEIVASGVRVFTKAENPKNKESLALPRTLARSARRRNARKKGRIQQVKHYLSKALGLDLECF
VQGEKLATLFQTSKDFLSPWELRERALYRVLDKEELARVILHIAKRRGYDDITYGVEDNDSGKIKKAIAENSKRIKEEQCKTIGEMMYKLYF
QKSLNVRNKKESYNRCVGRSELREELKTIFQIQQELKSPWVNEELIYKLLGNPDAQSKQEREGLIFYQRPLKGFGDKIGKCSHIKKGENSP
YRACKHAPSAEEFVALTKSINFLKNLTNRHGLCFSQEDMCVYLGKILQEAQKNEKGLTYSKLKLLLDLPSDFEFLGLDYSGKNPEKAVFLS
LPSTFKLNKITQDRKTQDKIANILGANKDWEAILKELESLQLSKEQIQTIKDAKLNFSKHINLSLEALYHLLPLMREGKRYDEGVEILQERGIF
SKPQPKNRQLLPPLSELAKEESYFDIPNPVLRRALSEFRKVVNALLEKYGGFHYFHIELTRDVCKAKSARMQLEKINKKNKSENDAASQLL
EVLGLPNTYNNRLKCKLWKQQEEYCLYSGEKITIDHLKDQRALQIDHAFPLSRSLDDSQSNKVLCLTSSNQEKSNKTPYEWLGSDEKKW
DMYVGRVYSSNFSPSKKRKLTQKNFKERNEEDFLARNLVDTGYIGRVTKEYIKHSLSFLPLPDGKKEHIRIISGSMTSTMRSFWGVQEKN
RDHHLHHAQDAIIIACIEPSMIQKYTTYLKDKETHRLKSHQKAQILREGDHKLSLRWPMSNFKDKIQESIQNIIPSHHVSHKVTGELHQETVR
TKEFYYQAFGGEEGVKKALKFGKIREINQGIVDNGAMVRVDIFKSKDKGKFYAVPIYTYDFAIGKLPNKAIVQGKKNGIIKDWLEMDENYEF
CFSLFKNDCIKIQTKEMQEAVLAIYKSTNSAKATIELEHLSKYALKNEDEEKMFTDTDKEKNKTMTRESCGIQGLKVFQKVKLSVLGEVLEH
KPRNRQNIALKTTPKHV

>gi|253828136|ref|ZP_04871021.1|/1-1007
MKVLGFDIGINSIGWAFVENNELKDCGVRLFTKAENPKTKESLALPRRNARSSRVRLRRRRSRLLSLKHIISKEFGLHYEDYIALDGNLPKA
YTGKIESIYKLRYEALSQAISKEDLARVILHIAKHRGYMNKNQKTSQDSEKGAILSAIKENSNKIKHYQTAGEYFYKELFENLRDGTKEFKNI
RNKAMKNEKTGNMESTYYNTILASDLEKELKLILEKQQAYHNYSDTFIKEVLEIAFYQRPLKDFSHLVGSCTFYENEKRACKNSYSAWEFI
ALGKIINTLESIKKDTGEIFSNEIITQILKEVLDKGSLSYTKLRKMINLDEKVKFKGLKYDKEGVEKTNLIEFKKLKEFKKALGEHNLERKTLDEI
ATHITLTKDENQLKKQLEQYALSQNQIDALAELDFNDHINLSFKALYEILPLMEKGKRYDEACNELGLKAKTSNQKSEFLPAFCDSIFAQEL
TNPIVNRAISEYRKVLNRLLKKYGKMHKIHIELARDAGLSKEARNKIEKEQKENKDINDKAEKTCKEFGLKPNAKNILKIKLWREQGEICAYS
GKKITIDDLKEDKALEVDHIYPYSRSYDDSYNNKVLVFIKENQLKLNQTPYEAFGSNQEKWSKIQVLAQKLPYKKKNRILDENFKNKDQIEFI
SRNLNDTRYATSLIAKYTKEYLEFLPLSDDEDISLKSGEKGSKTHVQTMNGMLTSVLRHAWGFSQKDRNNHLHHALDAIIIAYSTNSIIKAF
SDFKKNQEILKAKLYAKKLTSEEYKNQSKFTGFEKYRSKILEKLDSIFVSKPPNKNTKGALHEQTFYSYQDILEEYKTQEGLKRALECGKVR
KIGTKYVVNDRMVRLDIFKKDNKFYGIPIYTMDFALGVLPNKVVTIGTDKDKIKKEWKTIDETYEFCFSLYKGDLVLIQKKEMQESEFAYYN
GFDISSASISLEKHDNKFENLTENQKLLFPKAKEGDVEVRGIGIQGLKVFKKYMVTPLGEIQETRTEARQNIQPRKKHGLR

>gi|356486333|gb|EHI16317.1|/1-983
MARILAFDIGISSIGWAFSENDELKDCGVRIFTKAENPKTGESLALPRRLARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAK
AYKGSLISPYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNV
RNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLN
NLKNTEGILYTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQDDLNEIAKDITLIKDEIKLKK
ALAKYDLNQNQIDSLSKLEFKDHLNISFKALKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKEYRKV
LNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKML
EIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARL
VLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNS
AELYAKKISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEFYQSYGGKEGVLKALELGKIRKVNGKIVKNG
DMFRVDIFKHKKTNKFYAVPIYTMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSSTVSLIV
SKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK

FIGURE 12 T.

>gi|189485225|ref|YP_001956166.1|/1-1032
MKKRILGLDIGIASVGWALVEFDDEFQDNKKKGTIIKSGVRIFTRAETPKERESLAKPRREARSQRRRIRRRVERLNSIRELFVKNGVVEE
WSVSKKNPNNIYITLNQIKTPWQLRMESLDRKLTNEQFAVVLTHLAKHRGYKSLRKDDKAGKVLEAISQNKKRLNDSDYRTLGEMFCKNE
SYKEHKRNKGGEYINSVARELIEEEVEKIFVKQREKGNNFADVNTEREYLKIAFSQKPAGSISKMVGFCTFENGEKRAPRDSYSAELFKVL
CTINNTTFTDMKTGEIRSFSKEEIGKTLETVKNVKDVKYAKIRKDLGLEDSVKFSGVDYRENAKSKGEGEIFVSMKTYHEIKSKIEKLSKEK
WDNLKGNTELLDSIVNIIANLKDDESIKKEFQKLKIDEDIIDVLLNFEFKKFINLSFKALRKIIPFQFQGQKYITASTSAGYSLKQEGIKNKLVPP
AKENINVAVVARAFAQTRKVINAVIRKYGQFDQINIELATELKNSKFDRSKIEEGQKKFQSQKDKIFKEVQELRGGMTPSSSLLLKYRLWRE
QDERCVYSGRKISFEDLLDTGFLDIDHIIPYSKSMDDSFNNKVLCLAEENRNKKNDIPFNYYQRIDRDWNLLISIITSMKNMKIAKKSRLLKQ
ELSDIEGFIERNINDTRYATRYIKDFLENNLEFKENEFIKVKVQARSGGLTSVLSYNWGMSKDREESYFHHAKDAIIVACSTQGMVQYISKI
SQQYENDKDKMKKIKDKTENCIPKPWDSFRADVEESIEGIFVSFAPRHKVTGSAHKETIYSKKHLEKERYVTVKKSLDKIKLSDLENIPCNE
NCGIIRVLKERLEKFKNDSLKAFADPVHMPTKCPNKKGPVIRSVKIKESNKTGIEVRKGLAERGEMVRVDVFIKDKKYYLVPIYVSDFKNKD
LPNKAIAALKLENEWIMIDKTYSFKFSLFKDDLIKIKKKKEEIFGYFGGVHRSTGTIKIKSRHKPKEIDEGIGSRCLLDFRKFQVDVLGNYTEV
KHEKRMPAYITRKDKKH

>gi|229113166|ref|ZP_04242662.1|/1-1069
MNYDIGLRIGITSCGWSIINKDLERIEDLGVRVFEKAENPDGTASAAPRREARKSRRKHRRKKHRIERIKRLIVQHDLLSKKELDSLYLSPFN
IDVWNLQVEALERKLDNREFARVLIHLVQRRGFQTIRKSVEIQEEGKLLENVSENERIMKVNGYKTVGEMFSKHEKFTHHKRNKDGSYTS
VVTRSLLLKEIKTIFNSQRILGNVFSTQKLESDYLYIWGSQRPTLTYEQLMSMVGNCIYEKKEKRAPKTSWAFQYFSLLQKINKLKVSDDNT
LRKLSKEERDIVIELAFKKKNVSFVDIRKALKLNDNTKFNHLTYSHVVETKKVEKATFIELKGYHLIKKQLKDRNVELTQELQPKDYDAIAAA
STFFKNDIEFRDYLRNQYLDSNGKRKSNIANKVFEDKVIAAVFELNFSKVCHLSFKSLYKLIPHLEKGLDYKEAILCAGYNFRDENKRDKKL
FLPVIQQETNPVVHRALTQARKVINAIIREYGSPNTIRFVVANDLAKTYKQRKELESTLKKNRAENEKIKTKLLSMGIINPTGMDIVKYKLWY
EQSGICLYTGRQMELSSLFKPGYSIVNHIIPYNRSFDDTYHNRVLTLTETKYEKGNKIPFEYFGAEETCWNEYEKRVSTSKKIHDKKREKLL
IRNFDIEYENEMISNNVFDTHYIGSYFEKYIDEHLQFKKSTKKQKVFSIHGNMVAHLHSRWGFNKNSREIYLNHAVDAAIVAVLTDQLIRNIA
DYYRRDNILEQKGSLKFPEPWKSFSIELEARISPQVRDLIKKLNLDSYLDVDILHIKPVFISKKPRRKVNGEAHKSTIRSLIGINEIGKTVTCIK
TRLEDIPFDENGDFPMYGKDTDKYTYTAIKERYLEFNKDKEKAFSTPLYKLKRNGEPGNQIKSVKIIDTRNIVNKVNQGKGIAYNSNIIRVDV
FSRDQKYYVVPIYLRDFIKGSLPNKIITAAKGFMQWPETDNSYTFLFSIFPNDLIKVVPKEGKFIKAKNENKENINLTEILGYFKGLDSSTGAF
TLESHDGSLLARGIGTKNLLLIEKYQVDVLGRYNKVGNEERQDIKIQQN

>gi|222109285|ref|YP_002551549.1|/1-1131
MAQHVFGLDIGIASVGWAILGEQRIIDLGVRCFDKAETAKEGDPLNLTRRQARLLRRRLYRRAWRLTQLSRLLKRKGLIADAKLFAKAPSY
GDSAWELRRQGLDRLLTPLEWARVIYHQCKHRGFHWTSKAEEAKADSDAEGGRVKQGLAHTKALMQAKNYRSAAEMVLAEFPDAQRN
KRGQYDKALSRVLLGEELALLFATQRRLGNPHASDFFEKLILGDGDRKSGLFWQQKPALSGADLLKMLGKCTFEKGEYRAPKASFSVER
HVWLTRLNNLRIVVDGRSRPLNEAERQAALLLPYQTETSKYKTLKNAFIKAGLWGDGVRFGGLAYPSQAQIDAEKTKDPEDQFLVKLPA
WHELRKAFKAAGHEALWQQISTPALDGDPTLLDQIATVLSVYKDGAEVVQQLRQLALPEPAASIAVLEKISFDKFSSLSLKALRRIVPLMQS
GLRYDEAVAQIPEYGHHSQRIEPGAAKHLYLPPFYEAQRKYAGKGDHIGSMQFRDDADIPRNPVVLRALNQARKVVNALIREYGSPIAVNI
EMARDLSRPLDERNKVKRAQEEFRDRNDRARSEFERDFGYKPKAAAFEKWMLYREQLGQCAYSQQPLDIQRVLDDHNYAQVDHALPY
SRSYDDSKNNKVLVLTHENQNKGNRTAFEYLTSFPDGEDGERWRTFVAWVQGNKAYRMAKRNRLLRKNYGVDESKGFIDRNLNDTRYI
CKFFKNYVEEHLQLAARADGDTARRCVVVNGQLTAFLRARWGLTKVRGDSDRHHALDAAVVAACTHGMVKALADYSRRKEISFLQEGF
PDPETGEILNPAAFDRARQHFPEPWTHFAHELKARLFTDDLAALREDMQRLGSYTTEDLGRLRTLFVSRAPQRRSGGAVHKETIYAQPE
SLKQQGGVIEKILLTSLKLQDFDKLLNPESNDHFVEPHRNERLYAAIRQRLEQFGGRADKAFGPDNLFHKPDKNNQPTGPVVRSIKLVRG
KQTGIPIRGGLAKNDSMLRVDIFTKAGKFHLVPVYVHHRVTGLPNRAIVAFKDEDEWTLIDESFAFLFSVYPNDYVKVTLKKEQQSGYYSG
ADRSTGAMNLWAHDRAASVGKDGLIRGIGVKTALSVEKFNVDVLGRIYLAPPETRSGLA

>gi|182624245|ref|ZP_02952031.1|/1-1065
MNKDINYALGLDIGITSVGWAVINLDLNRIEDLGVRIFNAAENLKDGSSLALPRRLARGRRRLLRRKAYRVERVRKLILKNNILTKEELEGLF
KNKEVIDVWEARVKGLDYKLSREEWAKILINFCKRRGFKSNRKNEAKDKEAGQILSSISYNIEKMKETNSRTIGEYIYNEVKNSEDSYAPLR
NKFGEYKMCVSRDVIREEIHTLFEKQRELGNNFASDEIEDKYFEIFNSQRPFSNFEDLEKLVGFCTFERKKHKRAPKHCISSEEFTLYEGIN
KLSIIKDGEKRKLSEEERTLVVNEAFNKKEIKYTTLRKLLGLSEEELFSTLTYSIDKDISKTENTKFVSLKGYHEIKRAIEKGVSKKAWQDIEN
NRNLLNDIAYVLTLGKTDDEIEKQLRLRNVPEELFESLLDMSFNKFNNLSIYALEKILPFMKEGYQYNEACEKAGYNFKAIYEGVRTKKLPVI
EIDEIVNPVVNRALAQTRKVINSVIDKYGSPVRINIELARDLAKNFKDRKLIEKEQKENRANLDKIRENLRELMDKEPTVVEVLKYRLWEQQ
RGECAYTQSQIPLERLFSPGYCEIDHIIPFSRSFDDSLSNKILVLGSENQRKGNRTPYEYFGEDSERWNSFEIWVKGSYLNYKKKANLLKK
KFSEEEQRDWKARNLQDTKYICRYISNFVNNRLEFKESENKQKVITVNGRATAYLRSRWGLNKVREDGDKHHALDAAVVGVTTQGMVQ
KISKYSKAHELYTVRKNDEFVDVETGECVDVDEYKEIRKDILPRPWRGFSEELKLRLSDNPLEELKKSPIETYNEEFIKENVKPIFVSRVPF
RKISGKLFKETIYSERAFKDGYFVSKKSLVDLKRGDLDNFYNYECDKKLYDAISERMKEFNYNGKKAFESPFRKPTKSGKEGPVVRSVKIK
SAVPFKDGIPMNEGLVAKEGMVRIDIYEKEGKYFVVPVYRYQLAKGIVPKRAAIAKKDELKWPIMDQTYNFKFSIYKNDLIEVKYDKKGVYF
GYYDGFDRNSASLTIEKHDNSERYRSIGLKVGVLEINKYEVDVLGNYHKVKLGGN

FIGURE 12 U.

>gi|220930482|ref|YP_002507391.1|/1-1021
MKYTLGLDVGIASVGWAVIDKDNNKIIDLGVRCFDKAEESKTGESLATARRIARGMRRRISRRSQRLRLVKKLFVQYEIIKDSSEFNRIFDT
SRDGWKDPWELRYNALSRILKPYELVQVLTHITKRRGFKSNRKEDLSTTKEGVVITSIKNNSEMLRTKNYRTIGEMIFMETPENSNKRNKV
DEYIHTIAREDLLNEIKYIFSIQRKLGSPFVTEKLEHDFLNIWEFQRPFASGDSILSKVGKCTLLKEELRAPTSCYTSEYFGLLQSINNLVLVE
DNNTLTLNNDQRAKIIEYAHFKNEIKYSEIRKLLDIEPEILFKAHNLTHKNPSGNNESKKFYEMKSYHKLKSTLPTDIWGKLHSNKESLDNLF
YCLTVYKNDNEIKDYLQANNLDYLIEYIAKLPTFNKFKHLSLVAMKRIIPFMEKGYKYSDACNMAELDFTGSSKLEKCNKLTVEPIIENVTNP
VVIRALTQARKVINAIIQKYGLPYMVNIELAREAGMTRQDRDNLKKEHENNRKAREKISDLIRQNGRVASGLDILKWRLWEDQGGRCAYS
GKPIPVCDLLNDSLTQIDHIYPYSRSMDDSYMNKVLVLTDENQNKRSYTPYEVWGSTEKWEDFEARIYSMHLPQSKEKRLLNRNFITKDL
DSFISRNLNDTRYISRFLKNYIESYLQFSNDSPKSCVVCVNGQCTAQLRSRWGLNKNREESDLHHALDAAVIACADRKIIKEITNYYNERE
NHNYKVKYPLPWHSFRQDLMETLAGVFISRAPRRKITGPAHDETIRSPKHFNKGLTSVKIPLTTVTLEKLETMVKNTKGGISDKAVYNVLK
NRLIEHNNKPLKAFAEKIYKPLKNGTNGAIIRSIRVETPSYTGVFRNEGKGISDNSLMVRVDVFKKKDKYYLVPIYVAHMIKKELPSKAIVPLK
PESQWELIDSTHEFLFSLYQNDYLVIKTKKGITEGYYRSCHRGTGSLSLMPHFANNKNVKIDIGVRTAISIEKYNVDILGNKSIVKGEPRRG
MEKYNSFKSN

>gi|225377804|ref|ZP_03755025.1|/1-1152
MNAEHGKEGLLIMEENFQYRIGLDIGITSVGWAVLQNNSQDEPVRITDLGVRIFDVAENPKNGDALAAPRRDARTTRRRLRRRRHRLERI
KFLLQENGLIEMDSFMERYYKGNLPDVYQLRYEGLDRKLKDEELAQVLIHIAKHRGFRSTRKAETKEKEGGAVLKATTENQKIMQEKGYR
TVGEMLYLDEAFHTECLWNEKGYVLTPRNRPDDYKHTILRSMLVEEVHAIFAAQRAHGNQKATEGLEEAYVEIMTSQRSFDMGPGLQPD
GKPSPYAMEGFGDRVGKCTFEKDEYRAPKATYTAELFVALQKINHTKLIDEFGTGRFFSEEERKTIIGLLLSSKELKYGTIRKKLNIDPSLKF
NSLNYSAKKEGETEEERVLDTEKAKFASMFWTYEYSKCLKDRTEEMPVGEKADLFDRIGEILTAYKNDDSRSSRLKELGLSGEEIDGLLD
LSPAKYQRVSLKAMRKMQPYLEDGLIYDKACEAAGYDFRALNDGNKKHLLKGEEINAIVNDITNPVVKRSVSQTIKVINAIIQKYGSPQAVN
IELAREMSKNFQDRTNLEKEMKKRQQENERAKQQIIELGKQNPTGQDILKYRLWNDQGGYCLYSGKKIPLEELFDGGYDIDHILPYSITFD
DSYRNKVLVTAQENRQKGNRTPYEYFGADEKRWEDYEASVRLLVRDYKKQQKLLKKNFTEEERKEFKERNLNDTKYITRVVYNMIRQNL
ELEPFNHPEKKKQVWAVNGAVTSYLRKRWGLMQKDRSTDRHHAMDAVVIACCTDGMIHKISRYMQGRELAYSRNFKFPDEETGEILNR
DNFTREQWDEKFGVKVPLPWNSFRDELDIRLLNEDPKNFLLTHADVQRELDYPGWMYGEEESPIEEGRYINYIRPLFVSRMPNHKVTGS
AHDATIRSARDYETRGVVITKVPLTDLKLNKDNEIEGYYDKDSDRLLYQALVRQLLLHGNDGKKAFAEDFHKPKADGTEGPVVRKVKIEKK
QTSGVMVRGGTGIAANGEMVRIDVFRENGKYYFVPVYTADVVRKVLPNRAATHTKPYSEWRVMDDANFVFSLYSRDLIHVKSKKDIKTN
LVNGGLLLQKEIFAYYTGADIATASIAGFANDSNFKFRGLGIQSLEIFEKCQVDILGNISVVRHENRQEFH

>gi|162147907|ref|YP_001602368.1|/1-1050
MGENMIDESLTFGIDLGIGSCGWAVLRRPSAFGRKGVIEGMGSWCFDVPETSKERTPTNQIRRSNRLLRRVIRRRRNRMAAIRRLLHAA
GLLPSTDSDALKRPGHDPWELRARGLDKPLKPVEFAVVLGHIAKRRGFKSAAKRKATNISSDDKKMLTALEATRERLGRYRTVGEMFAR
DPDFASRRRNREGKYDRTTARDDLEHEVHALFAAQRRLGQGFASPELEEAFTASAFHQRPMQDSERLVGFCPFERTEKRAAKLTPSFE
RFRLLARLLNLRITTPDGERPLTVDEIALVTRDLGKTAKLSIKRVRTLIGLEDNQRFTTIRPEDEDRDIVARTGGAMTGTATLRKALGEALWT
DMQERPEQLDAIVQVLSFFEANETITEKLREIGLTLAVLDVLLTALDAGVFAKFKGAAHISTKAARNLLPHLEQGRRYDEACTMAGYDHAA
SRLSHHGQIVAKTQFNALVTEIGESIANPIARKALIEGLKQIWAMRNHWGLPGSIHVELARDVGNSIEKRREIEKHIEKNTALRARERREVH
DLLDLEDVNGDTLLRYRLWKEQGGKCLYTGKAIHIRQIAATDNSVQVDHILPWSRFGDDSFNNKTLCLASANQQKKRSTPYEWLSGQTG
DAWNAFVQRIETNKELRGFKKRNYLLKNAKEAEEKFRSRNLNDTRYAARLFAEAVKLLYAFGERQEKGGNRRVFTRPGALTAALRQAW
GVESLKKQDGKRINDDRHHALDALTVAAVDEAEIQRLTKSFHEWEQQGLGRPLRRVEPPWESFRADVEATYPEVFVARPERRRARGEG
HAATIRQVKERECTPIVFERKAVSSLKEADLERIKDGERNEAIVEAIRSWIATGRPADAPPRSPRGDIITKIRLATTIKAAVPVRGGTAGRGE
MVRADVFSKPNRRGKDEWYLVPVYPHQIMNRKAWPKPPMRSIVANKDEDEWTEVGPEHQFRFSLYPRSNIEIIRPSGEVIEGYFVGLHR
NTGALTISAHNDPKSIHSGIGTKTLLAISKYQVDRFGRKSPVRKEVRTWHGEACISPTPPG

>gi|291518097|emb|CBK73318.1|/1-765
MGITIGLDLGVASVGWAVVNDDYEILESCSNIFPSADASKNSERRGFRQGRRLTRRRKNRIHDFQKLWEDKGFVIPSQGTEDVLAIKIKGL
SEKLSVDEVYWVLLNSLKHRGISYLDDADSGDNSSDYAKSISRNEEELKEKLPCEIQWERLQKYGAYRGNISIVEDGEPITLRNVFTTSAY
KKEVEQFIDTQAKYNAQYSGDFKADYLEIFNRKREYYEGPGNELSRTDYGKYTTEINADGEYITVDNIFDKLVGKCSVNPDERRAAGASY
TAQEFNVLNDLNNLTISSESSFIEDGKLTEDAKRKIIETIKNAKTVNVKKIICDVIGDKKCQISGARIDKNEKEIFHSFEAYNKMRRALEEIGFDI
SSLSRENLDLIGDILTLNTDRESILNAFNRKGIELADEAKDILVKVRKTNGSLFNKWQSFGLSIMNELIPELYAQPKNQMELLTAMGVFKSR
GDRFLECKEIPGDLIVDDIYNPVVSKTVRITVRILNALIKKYGYPDRVVIEMPRDKNSDEEQQRLKKEQRDNENEIKDIKARVKTEYGREITE
EDFRQHSKLSLKLKLWNEQQGICPYSGKSIKIDDLLDNPNLFEVDHIIPLSISFDDSRNNKVLVYSSENQDKGNRTPLAYLASVNRQWDIH
SFMDYVLKTYAGAQKRKKRDNLLNEQDITKVEVLQGFVNRNINDTRYASKVVLNSLQEYFSSKECSTKVKVIRGSFTHQMRVNLKLKKTV
MNHMFIMQLMQCLLLFHRWDMTLIMNLQISI

>gi|257456748|ref|ZP_05621934.1|/1-992
MQKNSGEDENCGFEFLTEEEKELIKPLFFRISKPSFQFADIEKKLKGKNGFWKFNYRKDTNVSGCPVSAGLKNIFGDEASAESWKDKKVG
QYDMCDIWHVLFDFDDEEKLLEFAKRKLFLSDEAAKKFCAIRIQQGYANLSLKAIRKITPFLRKGYIYSTAVFLANIPFMIGRDIFLQNEKEIE
DSVKNIIGTLRDKNNIIVLANRCIESAFKDKDNDFRFEEWDKALVENSAWDLFGKKKWNEYDEEKRKTIISQVSEKVEDNLKIAVGKNPND
YKYPLLRTDDLIMDYLNQKGFVVKGELYHPSDTDYNFETPVPAEDGKIYLASPRSPSVKNPVVMRALHQLRKLVNYLIKTGKIDSTTKINVE
LANDVNDKNQRKAIEELAKTNEKNNADARKKIEELCNEAGFKVVPTESDIKKFRLWKEQNETCPYTGKHISFTDLFGPIPKFDFEHTIPRSL
SYDDSLENLTLCDSEFNRNIKKQRLPSELPDFEEINKRFLKFYEDKIDNCLRIIELNSKSGGSYEEPAVKDLRIVKKHKAQYELNYYKEKLRR
FSSTEITSGFKHSQLNDTRIITKFSLSYLKGVFDHVQPVKGSMTDTFKRQWGLMERNEIKDRSNHTHHTVDALTVACINRGKFNLLSEAIK
NSSDGKHLKFPKPWETFDTDVLNAVRYIIPKYFSDESSLRQSKKILRGRDGKPVLKNGKAVFIQGATARGSLHKDTFYGCIKTVPEKGGKS
EMIFVQRIPVSTLDEKVAEKIIDKRIRKTFEKNLSTGIQTLQEIQTDGILLPFKKEGRDVFVKRVRIKAHPTSPIILKKHHNVINKNPKDYKQNY
YVENEENYLLAIYRGKDAKGKDVSDHKLCNLLNAVKSRQNKTGFYPDFKEKKGINLQLYKVLKIGKIVILQNDIQEDVFALPKEKLWKRMY
RIAGLATSRNDIQIKLVHIIRETPWGYMKGEKDLNAGKECLLYGTANFKGLVEGQDFTVSPAGEIIQKARVC

FIGURE 12 V.

>gi|227512703|ref|ZP_03942752.1|/1-1239
MTEKRKFDIREVYLAIHHIVKYRGNFLNSATVDSFKTSNIDFTSQFDRLNELYRQVILEEPFQINMNQVDKMTNKLLDNDALKLDTQKQVAK
LLPVIYNDKPVDKQYTRLATEFSKAILGYKTKLDVILNLDTQNAKDWAIRLDDEDIDDKLPALVENLDESRQEIVTIIRDLYAQITLNAIVPKGK
SLSESMVDKYNDHKDHLDLLMGLIHELGDDSAKGIKLREAYTQYVGKSGDKTLNQDDFYAAIKKNLDEDSKRSPKIQRLIEQASFMPKQR
TSANGVIPHQLHQLELDRIIENQGRYYPFLKEPNSNTHKPNGGKYKLDELVAFRIPYYVGPLITKQDQEKTSGADFAWMIRKPGQGGEITP
WNFDQEVDRMASANTFIRRMTTKDSYLVAEDVLPDSSLIYEKFKVLNELNMLKVNDRRLSISQKQDLFNNLFKKQKTIRTKKLQTYIRTKW
ELPSVMISGLSDPDKFNSSLATYIDFRDVFGEKVDDPNRQADYEKIIEWSTVFEDRKIYQAKLNQLDWLTESQSKALLTKHYTGWGRLSK
KLLTGLKDQNGDSILDQLWKTNDNFMQIQSRDEFAKQIHDENAKQFHEANNVDDILDDAFTSPQNKKAIRQVDKVVQDVVKAVGYAPDKI
AIEFTRSPQDRPQRTFSRQRQLQSTYLNVAKALMKSNLNQELKSVIDSQQTLTDKLYLYFTQLGQDMYTGKEINFDELVNYQIDHILPQAFI
KDDSLDNRVLVSAPINNAKSDNVPVKKFGAKMGYFWKQLAENHLISKRKLNNLTTDPDKIGKFTAQGFIHRQLVETSQVIRLVANILGNEY
GHDGTTIIEVTAKMNHQMRKDFDLIKIRDVNDYHHAMDAYLTAYVGDYLYLRYPKLRSYFVYGDFKKLKDQSLKIRNFNFLHDLTNDEAGD
KIVDQETGEIIWDKHESVKQLKKVYHYKFMLVSQEVYTRQDALFNQTIYPASDADKRKLIPIKNNKPVDVYGGYSGNVDAYMAIVRIHGKK
EDKYKVVGVPMRAVADLKKAESKGRDNYLNAVHEVLKPQFTKKKKDRKTGEITEMMDDFDVLLGKVYYRQLIVDGNKKFMLGSSTYQY
NAKQLVLSDKAMQVLSKDDKLKNQDENQNLIDVYDEILEKVDQYFELYDINKFRQKLHDGRKKFIELPVDNDFNGKKLISYGKRATIISILNG
LHANATMSNLKYLGISSPFGMLQVPNGIILSPASRICYQSPTGLFERKVKLSDL

METHODS AND COMPOSITIONS FOR RNA-DIRECTED SITE-SPECIFIC DNA MODIFICATION

BACKGROUND

[0001] In recent years, engineered nuclease enzymes designed to target specific DNA sequences have attracted considerable attention as powerful tools for the genetic manipulation of cells and whole organisms, allowing targeted gene deletion, replacement and repair, as well as the insertion of exogenous sequences (transgenes) into the genome. Two major technologies for engineering site-specific DNA nucleases have emerged, both of which are based on the construction of chimeric endonuclease enzymes in which a sequence non-specific DNA endonuclease domain is fused to an engineered DNA binding domain. However, targeting each new genomic locus requires the design of a novel nuclease enzyme, making these approaches both time consuming and costly. In addition, both technologies suffer from limited precision, which can lead to unpredictable off-target effects.

[0002] There is need in the field for a technology that allows precise targeting of nuclease activity (or other protein activities) to distinct locations within a target DNA in a manner that does not require the design of a new protein for each new target sequence. The present disclosure addresses this need.

SUMMARY

[0003] The present disclosure provides a DNA-targeting RNA that comprises a targeting sequence and, together with a modifying polypeptide, provides for site-specific modification of a target DNA and/or a polypeptide associated with the target DNA. The present disclosure further provides modifying polypeptides. The present disclosure further provides methods of site-specific modification of a target DNA and/or a polypeptide associated with the target DNA; and kits and compositions for carrying out the methods.

BRIEF DESCRIPTION OF THE DRAWINGS

[0004] **Figures 1A and 1B** provides a schematic drawing of two exemplary subject DNA-targeting RNAs, each associated with a site-directed modifying polypeptide and with a target DNA. A DNA-targeting RNA comprises a single stranded "DNA-targeting segment" and a "protein-binding segment," which comprises a stretch of double stranded RNA. (**A**) A DNA-targeting RNA can comprise two separate RNA molecules (referred to as a "double-molecule" or "two-molecule" DNA-targeting RNA). A double-molecule DNA-targeting RNA comprises a "targeter-RNA" and an "activator-RNA." (**B**) A DNA-targeting RNA can comprise a single

RNA molecule (referred to as a "single-molecule" DNA-targeting RNA). A single-molecule DNA-targeting RNA comprises "linker nucleotides."

[0005] **Figure 2** depicts the amino acid sequence of a Cas9/Csn1 protein from *Streptococcus pyogenes*. Domains 1 and 2 are conserved to varying degrees among the Cas9/Csn1 proteins of many different species (also see Figures 10-12). The presence of domains 1 and 2 is a characteristic of Cas9/Csn1 proteins. The depicted domain 1 is 1 of 3 motifs present in Cas9/Csn1 that define a RuvC endonuclease domain. The depicted domain 2 is a combination of motifs 2 and 3 of the RuvC endonuclease domain as well as an HNH endonuclease domain.

[0006] **Figures 3A-C depict t**arget DNA cleavage by a site-directed modifying polypeptide (exemplified by the Cas9/Csn1 protein of *Streptococcus pyogenes*), which is directed by a DNA-targeting RNA. **(A)** Radiolabeled target DNAs were incubated in the presence of recombinant Cas9/Csn1 and various DNA-targeting RNA species (as indicated). Cleavage products were resolved using denaturing polyacrylamide gel electrophoresis and visualized by phosphorimaging. **(B)** Schematic diagrams of the DNA-targeting RNAs used in conjunction with the Cas9/Csn1 site-directed modifying polypeptide. Note that while one of the tested single-molecule DNA-targeting RNAs (RNA chimera A) supported efficient target DNA cleavage, the other tested single-molecule DNA-targeting RNA (RNA chimera B) did not. **(C)** Schematic representations of DNA-targeting RNA sequences and DNA targets.

[0007] **Figure 4** depicts target DNA editing through double-stranded DNA breaks introduced using a Cas9/Csn1 site-directed modifying polypeptide and a DNA-targeting RNA.

[0008] **Figures 5A and 5B depict target DNA cleavage. Figure 5 A.** Target DNA cleavage using a Cas9/Csn1 site-directed modifying polypeptide from a variety of different species (refer to sequences in Figure 12) and a DNA-targeting RNA. Target cleavage was carried out under the same conditions as in Figure 3. This experiment demonstrates that the various Cas9/Csn1 site-directed modifying polypeptides can utilize the same DNA-targeting RNA. **B.** A schematic representation of the DNA-targeting RNA used in Figure 5A.

[0009] **Figures 6A-C provide tracrRNA sequences. Figure 6 A depicts exemplary "tracrRNA" sequences ("activator-RNA" that can be included in a subject RNA.** . Figure 6**B provides a m**ultiple sequence alignment of selected tracrRNA orthologues (AlignX, VectorNTI package, Invitrogen) associated with CRISPR/Cas loci of similar architecture and highly similar Cas9/Csn1 sequences. Black boxes represent shared nucleotides. **C.** Multiple sequence alignment of selected tracrRNA orthologues (AlignX, VectorNTI package, Invitrogen) associated with CRISPR/Cas loci of different architecture and non-closely related Cas9/Csn1

sequences. Note the sequence similarity of N. meningitidis and P. multocida tracrRNA orthologues. Black boxes represent shared nucleotides.

[0010] **Figures 7A-C provide exemplary CRISPR ("targeter-RNA") sequences. Figure 7 A** depicts exemplary DNA comprising the sequences that encode the duplex-forming segment of the CRISPR repeat ("targeter-RNA Figures 7B and 7C**C provide m**ultiple sequence alignments of exemplary CRISPR repeat sequences (AlignX, VectorNTI package, Invitrogen) associated with the loci of different architecture and diverse Cas9 sequences. Black boxes represent shared nucleotides.

[0011] **Figures 8A-C** provide sequence alignments of exemplary duplex-forming segment of the CRISPR repeat ("targeter-RNA") with the duplex-forming segment of the corresponding tracrRNA orthologue ("activator-RNA"). Upper sequence, CRISPR repeat consensus sequence; lower sequence, tracrRNA othologue sequence annealing to the repeat. The CRISPR loci belong to the Type II (Nmeni/CASS4) CRISPR/Cas system. Nomenclature is according to the CRISPR database (CRISPR DB). Note that *S. thermophilus* LMD-9 and *W. succinogenes* contain each two Type II CRISPR loci. Two possible antirepeats were found within the F. *tularensis subsp. novicida*, *W. succinogenes* and *C. jejuni* CRISPR loci. Upper pair of sequences: antirepeat downstream of the CRISPR repeat spacer array; lower pair of sequences: antirepeat within the leader sequence of the CRISPR repeat-spacer array. An antirepeat sequence with low complementarity (11 mismatches) with the CRISPR repeat was found in the coding region of the gene upstream of Cas9/Csn1. Two possible antirepeat sequences were found within the *R. rubrum* CRISPR locus. Upper pair of sequences: antirepeat downstream of the repeat spacer array; lower pair: antirepeat upstream of cas1.

[0012] **Figures 9 A-G** depict tracrRNA (activator-RNA) and crRNA (targeter-RNA) sequences from various species. A degree of interchangeability exists; for example, the *S.pyogenes* Cas9/Csn1 protein is functional with tracrRNA and crRNA derived from *L.innocua*. (|) denotes a canonical Watson-Crick base pair while (•) denotes a G-U wobble base pair. "Variable 20nt" or "20nt" represents the DNA-targeting segment that is complementary to a target DNA. The schematics at the bottom show the design of single-molecule DNA-targeting RNA that incorporates features of the targeter-RNA and the activator-RNA. **A.** *Streptococcus pyogenes* (Cas9/Csn1 protein sequence from *Streptococcus pyogenes* is depicted in Figures 2 and 12A). ***B.*** *Listeria innocua* (Cas9/Csn1 protein sequence from *Listeria innocua* is depicted in Figure 12B).**C.** *Streptoccus mutans* (Cas9/Csn1 protein sequence from *Streptoccus mutans* is depicted in Figure 12C). **D.** *Streptoccus thermophilus* – locus 1 (or A) (Cas9/Csn1 protein sequence from *Streptoccus thermophilus* – locus 1 (or A) is depicted in Figure 12D). **E.** *Streptoccus thermophilus* – locus 2 (or B) (Cas9/Csn1 protein sequence from *Streptoccus thermophilus* – locus 2 (or B) is depicted

in Figure 12E). **F.** *Neisseria meningitides* (Cas9/Csn1 protein sequence from *Neisseria meningitides* is depicted in Figure 12F). **G.** *Pasteurella multocida* (Cas9/Csn1 protein sequence from *Pasteurella multocida* is depicted in Figure 12G).

[0013] **Figures 10A and 10B** depict the percent identity between the Cas9/Csn1 proteins from multiple species. **A.** Sequence identity relative to *Streptococcus pyogenes.* Domain 1 is amino acids 7-166 of and Domain 2 is amino acids 731-1003 Cas9/Csn1 from *Streptococcus pyogenes* as depicted in Figure 2. **B.** Sequence identity relative to *Neisseria meningitides*. Domain 1 is amino acids 13-139 of and Domain 2 is amino acids 475-750 Cas9/Csn1 from *Neisseria meningitides* as depicted in Figure 12F.

[0014] **Figures 11A-AM** depict a multiple sequence alignment of the Cas9/Csn1 proteins from various species (e.g., *Streptococcus pyogenes*, *Listeria innocua, Streptoccus mutans, Streptoccus thermophilus* – locus 1 (or A), *Streptoccus thermophilus* – locus 2 (or B), *Neisseria meningitides*, *Pasteurella multocida*, etc.).

[0015] **Figures 12 A-V** provide amino acid sequences of the Cas9/Csn1 proteins from various species (e.g., *Streptococcus pyogenes*, *Listeria innocua, Streptoccus mutans, Streptoccus thermophilus* – locus 1 (or A), *Streptoccus thermophilus* – locus 2 (or B), *Neisseria meningitides*, *Pasteurella multocida*, etc.).

DEFINITIONS

[0016] The terms "polynucleotide" and "nucleic acid," used interchangeably herein, refer to a polymeric form of nucleotides of any length, either ribonucleotides or deoxyribonucleotides. Thus, this term includes, but is not limited to, single-, double-, or multi-stranded DNA or RNA, genomic DNA, cDNA, DNA-RNA hybrids, or a polymer comprising purine and pyrimidine bases or other natural, chemically or biochemically modified, non-natural, or derivatized nucleotide bases. "Oligonucleotide" generally refers to polynucleotides of between about 5 and about 100 nucleotides of single- or double-stranded DNA. However, for the purposes of this disclosure, there is no upper limit to the length of an oligonucleotide. Oligonucleotides are also known as "oligomers" or "oligos" and may be isolated from genes, or chemically synthesized by methods known in the art. The terms "polynucleotide" and "nucleic acid" should be understood to include, as applicable to the embodiments being described, single-stranded (such as sense or antisense) and double-stranded polynucleotides.

[0017] A "stem-loop structure" refers to a nucleic acid having a secondary structure that includes a region of nucleotides which are known or predicted to form a double strand (step portion) that is linked on one side by a region of predominantly single-stranded nucleotides (loop portion). The

terms "hairpin" and "fold-back" structures are also used herein to refer to stem-loop structures. Such structures are well known in the art and these terms are used consistently with their known meanings in the art. As is known in the art, a stem-loop structure does not require exact base-pairing. Thus, the stem may include one or more base mismatches. Alternatively, the base-pairing may be exact, i.e. not include any mismatches.

[0018] A nucleic acid is "hybridizable" or "complementary" to another nucleic acid, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid can "base-pair" with or "anneal" to the other nucleic acid in a sequence specific manner under the appropriate *in vitro* or *in vivo* conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein; and Sambrook, J. and Russell, W., Molecular Cloning: A Laboratory Manual, Third Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (2001). The conditions of temperature and ionic strength determine the "stringency" of the hybridization.

[0019] Hybridization requires that the two nucleic acids contain complementary sequences, although mismatches between bases are possible. The conditions appropriate for hybridization between two nucleic acids depend on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of complementation between two nucleotide sequences, the greater the value of the melting temperature (Tm) for hybrids of nucleic acids having those sequences. For hybridizations between nucleic acids with short stretches of complementarity (e.g. complementarity over 35 or less, 30 or less, 25 or less, 22 or less, 20 or less, or 18 or less nucleotides) the position of mismatches becomes important (see Sambrook et al., supra, 11.7-11.8). Typically, the length for a hybridizable nucleic acid is at least about 10 nucleotides. Illustrative minimum lengths for a hybridizable nucleic acid are: at least about 15 nucleotides; at least about 20 nucleotides; at least about 22 nucleotides; at least about 25 nucleotides; and at least about 30 nucleotides). Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the region of complementation and the degree of complementation.

[0020] It is understood in the art that the sequence of polynucleotide need not be 100% complementary to that of its target nucleic acid to be specifically hybridizable or hybridizable. Moreover, a polynucleotide may hybridize over one or more segments such that intervening or adjacent segments are not involved in the hybridization event (e.g., a loop structure or hairpin structure). A polynucleotide can comprise at least 70%, at least 80%, at least 90%, at least 95%,

at least 99%, or 100% sequence complementarity to a target region within the target nucleic acid sequence to which they are targeted. For example, an antisense nucleic acid in which 18 of 20 nucleotides of the antisense compound are complementary to a target region, and would therefore specifically hybridize, would represent 90 percent complementarity. In this example, the remaining noncomplementary nucleotides may be clustered or interspersed with complementary nucleotides and need not be contiguous to each other or to complementary nucleotides. Percent complementarity between particular stretches of nucleic acid sequences within nucleic acids can be determined routinely using BLAST programs (basic local alignment search tools) and PowerBLAST programs known in the art (Altschul et al., J. Mol. Biol., 1990, 215, 403-410; Zhang and Madden, Genome Res., 1997, 7, 649-656) or by using the Gap program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, Madison Wis.), using default settings, which uses the algorithm of Smith and Waterman (Adv. Appl. Math., 1981, 2, 482-489).

[0021] The terms "peptide," "polypeptide," and "protein" are used interchangeably herein, and refer to a polymeric form of amino acids of any length, which can include coded and non-coded amino acids, chemically or biochemically modified or derivatized amino acids, and polypeptides having modified peptide backbones.

[0022] "Binding" as used herein (e.g. with reference to an RNA-binding domain of a polypeptide) refers to a non-covalent interaction between macromolecules (e.g., between a protein and a nucleic acid). While in a state of non-covalent interaction, the macromolecules are said to be "associated" or "interacting" or "binding" (e.g., when a molecule X is said to interact with a molecule Y, it is meant the molecule X binds to molecule Y in a non-covalent manner). Not all components of a binding interaction need be sequence-specific (e.g., contacts with phosphate residues in a DNA backbone), but some portions of a binding interaction may be sequence-specific. Binding interactions are generally characterized by a dissociation constant (K_d) of less than 10^{-6} M, less than 10^{-7} M, less than 10^{-8} M, less than 10^{-9} M, less than 10^{-10} M, less than 10^{-11} M, less than 10^{-12} M, less than 10^{-13} M, less than 10^{-14} M, or less than 10^{-15} M. "Affinity" refers to the strength of binding, increased binding affinity being correlated with a lower K_d.

[0023] By "binding domain" it is meant a protein domain that is able to bind non-covalently to another molecule. A binding domain can bind to, for example, a DNA molecule (a DNA-binding protein), an RNA molecule (an RNA-binding protein) and/or a protein molecule (a protein-binding protein). In the case of a protein domain-binding protein, it can bind to itself (to form homodimers, homotrimers, etc.) and/or it can bind to one or more molecules of a different protein or proteins.

[0024] The term "conservative amino acid substitution" refers to the interchangeability in proteins of amino acid residues having similar side chains. For example, a group of amino acids having aliphatic side chains consists of glycine, alanine, valine, leucine, and isoleucine; a group of amino acids having aliphatic-hydroxyl side chains consists of serine and threonine; a group of amino acids having amide containing side chains consisting of asparagine and glutamine; a group of amino acids having aromatic side chains consists of phenylalanine, tyrosine, and tryptophan; a group of amino acids having basic side chains consists of lysine, arginine, and histidine; a group of amino acids having acidic side chains consists of glutamate and aspartate; and a group of amino acids having sulfur containing side chains consists of cysteine and methionine. Exemplary conservative amino acid substitution groups are: valine-leucine-isoleucine, phenylalanine-tyrosine, lysine-arginine, alanine-valine, and asparagine-glutamine.

[0025] A polynucleotide or polypeptide has a certain percent "sequence identity" to another polynucleotide or polypeptide, meaning that, when aligned, that percentage of bases or amino acids are the same, and in the same relative position, when comparing the two sequences. Sequence identity can be determined in a number of different manners. To determine sequence identity, sequences can be aligned using various methods and computer programs (e.g., BLAST, T-COFFEE, MUSCLE, MAFFT, etc.), available over the world wide web at sites including ncbi.nlm.nili.gov/BLAST, ebi.ac.uk/Tools/msa/tcoffee/, ebi.ac.uk/Tools/msa/muscle/, mafft.cbrc.jp/alignment/software/. See, e.g., Altschul et al. (1990), J. Mol. Bioi. 215:403-10.

[0026] A DNA sequence that "encodes" a particular RNA is a DNA nucleic acid sequence that is transcribed into RNA. A DNA polynucleotide may encode an RNA (mRNA) that is translated into protein, or a DNA polynucleotide may encode an RNA that is not translated into protein (e.g. tRNA, rRNA, or a DNA-targeting RNA; also called “non-coding” RNA or “ncRNA”).

[0027] A "protein coding sequence" or a sequence that encodes a particular protein or polypeptide, is a nucleic acid sequence that is transcribed into mRNA (in the case of DNA) and is translated (in the case of mRNA) into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' terminus (N-terminus) and a translation stop nonsense codon at the 3' terminus (C-terminus). A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic nucleic acids. A transcription termination sequence will usually be located 3' to the coding sequence.

[0028] As used herein, a "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase and initiating transcription of a downstream (3' direction) coding or non-coding

sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site, as well as protein binding domains responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Various promoters, including inducible promoters, may be used to drive the various vectors of the present invention. The promoter may be a constitutively active promoter, i.e. a promoter is active in the absence externally applied agents, or it may be an inducible promoter (e.g., T7 RNA polymerase promoter, T3 RNA polymerase promoter, Isopropyl-beta-D-thiogalactopyranoside (IPTG)-regulated promoter, heat shock promoter, Tetracycline-regulated promoter, Steroid-regulated promoter, Metal-regulated promoter, etc). As used herein, an inducible promoter is a promoter whose activity is regulated upon the application of an agent to the cell, (e.g. doxycycline) or the induced presence of a particular RNA polymerase (e.g., T7 RNA polymerase).

[0029] The terms "DNA regulatory sequences," "control elements," and "regulatory elements," used interchangeably herein, refer to transcriptional and translational control sequences, such as promoters, enhancers, polyadenylation signals, terminators, protein degradation signals, and the like, that provide for and/or regulate transcription of a non-coding sequence (e.g., DNA-targeting RNA) or a coding sequence (e.g., site-directed modifying polypeptide, or Cas9/Csn1 polypeptide) and/or regulate translation of an encoded polypeptide.

[0030] The term "naturally-occurring" or "unmodified" as used herein as applied to a nucleic acid, a polypeptide, a cell, or an organism, refers to a nucleic acid, polypeptide, cell, or organism that is found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by a human in the laboratory is naturally occurring.

[0031] The term "chimeric" as used herein as applied to a nucleic acid or polypeptide refers to two components that are defined by structures derived from different sources. For example, where "chimeric" is used in the context of a chimeric polypeptide (e.g., a chimeric Cas9/Csn1 protein), the chimeric polypeptide includes amino acid sequences that are derived from different polypeptides. A chimeric polypeptide may comprise either modified or naturally-occurring polypeptide sequences (e.g., a first amino acid sequence from a modified or unmodified Cas9/Csn1 protein; and a second amino acid sequence other than the Cas9/Csn1 protein). Similarly, "chimeric" in the context of a polynucleotide encoding a chimeric polypeptide includes nucleotide sequences derived from different coding regions (e.g., a first nucleotide

sequence encoding a modified or unmodified Cas9/Csn1 protein; and a second nucleotide sequence encoding a polypeptide other than a Cas9/Csn1 protein).

[0032] "Heterologous," as used herein, means a nucleotide or polypeptide sequence that is not found in the native nucleic acid or protein, respectively. For example, in a chimeric Cas9/Csn1 protein, the RNA-binding domain of a naturally-occurring bacterial Cas9/Csn1 polypeptide may be fused to a heterologous polypeptide sequence (i.e. a polypeptide sequence from a protein other than Cas9/Csn1 or a polypeptide sequence from another organism). The heterologous polypeptide sequence may exhibit an activity (e.g., enzymatic activity) that will also be exhibited by the chimeric Cas9/Csn1 protein (e.g., methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, etc.). A heterologous nucleic acid sequence may be linked to a naturally-occurring nucleic acid sequence (e.g., by genetic engineering) to generate a chimeric nucleotide sequence encoding a chimeric polypeptide.

[0033] "Recombinant," as used herein, means that a particular nucleic acid (DNA or RNA) is the product of various combinations of cloning, restriction, polymerase chain reaction (PCR) and/or ligation steps resulting in a construct having a structural coding or non-coding sequence distinguishable from endogenous nucleic acids found in natural systems. DNA sequences encoding polypeptides can be assembled from cDNA fragments or from a series of synthetic oligonucleotides, to provide a synthetic nucleic acid which is capable of being expressed from a recombinant transcriptional unit contained in a cell or in a cell-free transcription and translation system. Genomic DNA comprising the relevant sequences can also be used in the formation of a recombinant gene or transcriptional unit. Sequences of non-translated DNA may be present 5' or 3' from the open reading frame, where such sequences do not interfere with manipulation or expression of the coding regions, and may indeed act to modulate production of a desired product by various mechanisms (see "DNA regulatory sequences", below). Alternatively, DNA sequences encoding RNA (e.g., DNA-targeting RNA) that is not translated may also be considered recombinant. Thus, e.g., the term "recombinant" polynucleotide or "recombinant" nucleic acid refers to one which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of sequence through human intervention. This artificial combination is often accomplished by either chemical synthesis means, or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques. Such is usually done to replace a codon with a codon encoding the same amino acid, a conservative amino acid, or a non-conservative amino acid. Alternatively, it is performed to join together nucleic acid segments of desired functions to generate a desired combination of functions. This artificial combination is often accomplished by either chemical synthesis means,

or by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

[0034] Similarly, the term "recombinant" polypeptide refers to a polypeptide which is not naturally occurring, e.g., is made by the artificial combination of two otherwise separated segments of amino sequence through human intervention. Thus, e.g., a polypeptide that comprises a chimeric amino acid sequence is a recombinant polypeptide.

[0035] A "vector" or "expression vector" is a replicon, such as plasmid, phage, virus, or cosmid, to which another DNA segment, i.e. an "insert", may be attached so as to bring about the replication of the attached segment in a cell.

[0036] An "expression cassette" comprises a DNA coding sequence operably linked to a promoter. "Operably linked" refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their intended manner. For instance, a promoter is operably linked to a coding sequence if the promoter affects its transcription or expression.

[0037] The terms "recombinant expression vector," or "DNA construct" are used interchangeably herein to refer to a DNA molecule comprising a vector and at least one insert. Recombinant expression vectors are usually generated for the purpose of expressing and/or propagating the insert(s), or for the construction of other recombinant nucleotide sequences. The insert(s) may or may not be operably linked to a promoter sequence and may or may not be operably linked to DNA regulatory sequences.

[0038] A cell has been "genetically modified" or "transformed" or "transfected" by exogenous DNA, e.g. a recombinant expression vector, when such DNA has been introduced inside the cell. The presence of the exogenous DNA results in permanent or transient genetic change. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones that comprise a population of daughter cells containing the transforming DNA. A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

[0039] Suitable methods of genetic modification (also referred to as "transformation") include viral infection, transfection, conjugation, protoplast fusion, electroporation, particle gun technology, calcium phosphate precipitation, direct micro injection, and the like. The choice of method is

generally dependent on the type of cell being transformed and the circumstances under which the transformation is taking place (e.g., *in vitro*, *ex vivo*, or *in vivo*). A general discussion of these methods can be found in Ausubel, et al., Short Protocols in Molecular Biology, 3rd ed., Wiley & Sons, 1995.

[0040] A "target DNA" as used herein is a DNA polynucleotide that comprises a "target site" or "target sequence." The terms "target site" or "target sequence" are used interchangeably herein to refer to a nucleic acid sequence present in a target DNA to which a DNA-binding molecule will bind, provided sufficient conditions for binding exist. For example, the target site (or target sequence) 5'-GAGCATATC-3' within a target DNA is targeted by (or is bound by, or hybridizes with, or is complementary to) the RNA sequence 5'-GAUAUGCUC-3'. Suitable DNA/RNA binding conditions include physiological conditions normally present in a cell. Other suitable DNA/RNA binding conditions (e.g., conditions in a cell-free system) are known in the art; see, e.g., Sambrook, *supra*.

[0041] By "site-directed modifying polypeptide" or "RNA-binding site-directed polypeptide" or "RNA-binding site-directed modifying polypeptide" it is meant a polypeptide that binds RNA and is targeted to a specific DNA sequence. A site-directed modifying polypeptide as described herein is targeted to a specific DNA sequence by the RNA molecule to which it is bound. The RNA molecule comprises a sequence that is complementary to a target sequence within the target DNA, thus targeting the bound polypeptide to a specific location within the target DNA (the target sequence).

[0042] By "cleavage" it is meant the breakage of the covalent backbone of a DNA molecule. Cleavage can be initiated by a variety of methods including, but not limited to, enzymatic or chemical hydrolysis of a phosphodiester bond. Both single-stranded cleavage and double-stranded cleavage are possible, and double-stranded cleavage can occur as a result of two distinct single-stranded cleavage events. DNA cleavage can result in the production of either blunt ends or staggered ends. In certain embodiments, a complex comprising a DNA-targeting RNA and a site-directed modifying polypeptide is used for targeted double-stranded DNA cleavage.

[0043] "Nuclease" and "endonuclease" are used interchangeably herein to mean an enzyme which possesses catalytic activity for DNA cleavage.

[0044] By "cleavage domain" or "active domain" or "nuclease domain" of a nuclease it is meant the polypeptide sequence or domain within the nuclease which possesses the catalytic activity for DNA cleavage. A cleavage domain can be contained in a single polypeptide chain or cleavage activity can result from the association of two (or more) polypeptides. A single nuclease domain may consist of more than one isolated stretch of amino acids within a given polypeptide.

[0045] The RNA molecule that binds to the site-directed modifying polypeptide and targets the polypeptide to a specific location within the target DNA is referred to herein as the "DNA-targeting RNA" or "DNA-targeting RNA polynucleotide." A subject DNA-targeting RNA comprises two segments, a "DNA-targeting segment" and a "protein-binding segment." By "segment" it is meant a segment/section/region of a molecule, e.g., a contiguous stretch of nucleotides in an RNA. A segment can also mean a region/section of a complex such that a segment may comprise regions of more than one molecule. For example, in some cases the protein-binding segment (described below) of a DNA-targeting RNA is one RNA molecule and the protein-binding segment therefore comprises a region of that RNA molecule. In other cases, the protein-binding segment (described below) of a DNA-targeting RNA comprises two separate molecules that are hybridized along a region of complementarity. As an illustrative, non-limiting example, a protein-binding segment of a DNA-targeting RNA that comprises two separate molecules can comprise (i) base pairs 40-75 of a first RNA molecule that is 100 base pairs in length; and (ii) base pairs 10-25 of a second RNA molecule that is 50 base pairs in length. The definition of "segment," unless otherwise specifically defined in a particular context, is not limited to a specific number of total base pairs, is not limited to any particular number of base pairs from a given RNA molecule, is not limited to a particular number of separate molecules within a complex, and may include regions of RNA molecules that are of any total length and may or may not include regions with complementarity to other molecules.

[0046] The DNA-targeting segment (or "DNA-targeting sequence") comprises a nucleotide sequence that is complementary to a specific sequence within a target DNA. The protein-binding segment (or "protein-binding sequence") interacts with a site-directed modifying polypeptide.

[0047] The protein-binding segment of a subject DNA-targeting RNA comprises two complementary stretches of nucleotides that hybridize to one another to form a double stranded RNA duplex (dsRNA duplex). In some embodiments, the subject DNA-targeting RNA comprises two separate RNA molecules (RNA polynucleotides) and is referred to herein as a "double-molecule DNA-targeting RNA" or a "two-molecule DNA-targeting RNA." In other embodiments, the subject DNA-targeting RNA is a single RNA molecule (single RNA polynucleotide) and is referred to herein as a "single-molecule DNA-targeting RNA." The term "DNA-targeting RNA" is inclusive, referring to double-molecule DNA-targeting RNAs and to single-molecule DNA-targeting RNAs.

[0048] A subject DNA-targeting RNA and a subject site-directed modifying polypeptide form a complex (i.e., bind via non-covalent interactions). The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA. The site-directed modifying polypeptide of the complex provides the

site-specific activity. In other words, the site-directed modifying polypeptide is guided to a target DNA sequence (e.g. a chromosomal sequence or an extrachromosomal sequence, e.g. an episomal sequence, a minicircle sequence, a mitochondrial sequence, a chloroplast sequence, a plasmid sequence, etc.) by virtue of its association with the protein-binding segment of the DNA-targeting RNA.

[0049] A "host cell," as used herein, denotes an *in vivo* or *in vitro* eukaryotic cell, a prokaryotic cell (e.g., bacterial or archaeal cell), or a cell from a multicellular organism (e.g., a cell line) cultured as a unicellular entity, which eukaryotic or prokaryotic cells can be, or have been, used as recipients for a nucleic acid, and include the progeny of the original cell which has been transformed by the nucleic acid. It is understood that the progeny of a single cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation. A "recombinant host cell" (also referred to as a "genetically modified host cell") is a host cell into which has been introduced a heterologous nucleic acid, e.g., an expression vector. For example, a subject bacterial host cell is a genetically modified bacterial host cell by virtue of introduction into a suitable bacterial host cell of an exogenous nucleic acid (e.g., a plasmid or recombinant expression vector) and a subject eukaryotic host cell is a genetically modified eukaryotic host cell (e.g., a mammalian germ cell), by virtue of introduction into a suitable eukaryotic host cell of an exogenous nucleic acid.

[0050] By “pluripotent stem cell” or “pluripotent cell” it is meant a cell that has the ability to differentiate into all types of cells in an organism. Pluripotent cells are capable of forming teratomas and of contributing to ectoderm, mesoderm, or endoderm tissues in a living organism. Examples of pluripotent stem cells are embryonic stem (ES) cells, embryonic germ stem (EG) cells, and induced pluripotent stem (iPS) cells.

[0051] By “embryonic stem cell” or “ES cell” it is meant a cell that a) can self-renew, b) can differentiate to produce all types of cells in an organism, and c) is derived from the inner cell mass of the blastula of a developing organism. ES cells can be cultured over a long period of time while maintaining the ability to differentiate into all types of cells in an organism. In culture, ES cells typically grow as flat colonies with large nucleo-cytoplasmic ratios, defined borders and prominent nuclei. In addition, ES cells express SSEA-3, SSEA-4, TRA-1-60, TRA-1-81, and Alkaline Phosphatase, but not SSEA-1. Examples of methods of generating and characterizing ES cells may be found in, for example, US Patent No. 7,029,913, US Patent No. 5,843,780, and US Patent No. 6,200,806, the disclosures of which are incorporated herein by reference.

[0052] By "embryonic germ stem cell" or "embryonic germ cell" or "EG cell" it is meant a cell that a) can self-renew, b) can differentiate to produce all types of cells in an organism, and c) is derived from germ cells and germ cell progenitors, e.g. primordial germ cells, i.e. those that would become sperm and eggs. Embryonic germ cells (EG cells) are thought to have properties similar to embryonic stem cells as described above. Examples of methods of generating and characterizing EG cells may be found in, for example, US Patent No. 7,153,684; Matsui, Y., et al., (1992) Cell 70:841; Shamblott, M., et al. (2001) Proc. Natl. Acad. Sci. USA 98: 113; Shamblott, M., et al. (1998) Proc. Natl. Acad. Sci. USA, 95:13726; and Koshimizu, U., et al. (1996) Development, 122:1235, the disclosures of which are incorporated herein by reference.

[0053] By "induced pluripotent stem cell" or "iPS cell" it is meant a cell that a) can self-renew, b) can differentiate to produce all types of cells in an organism, and c) is derived from a somatic cell. iPS cells have an ES cell-like morphology, growing as flat colonies with large nucleo-cytoplasmic ratios, defined borders and prominent nuclei. In addition, iPS cells express one or more key pluripotency markers known by one of ordinary skill in the art, including but not limited to Alkaline Phosphatase, SSEA3, SSEA4, Sox2, Oct3/4, Nanog, TRA160, TRA181, TDGF 1, Dnmt3b, FoxD3, GDF3, Cyp26a1, TERT, and zfp42. Examples of methods of generating and characterizing iPS cells may be found in, for example, U.S. Patent Publication Nos. US20090047263, US20090068742, US20090191159, US20090227032, US20090246875, and US20090304646, the disclosures of which are incorporated herein by reference.

[0054] By "somatic cell" it is meant any cell in an organism that, in the absence of experimental manipulation, does not ordinarily give rise to all types of cells in an organism. In other words, somatic cells are cells that have differentiated sufficiently that they will not naturally generate cells of all three germ layers of the body, i.e. ectoderm, mesoderm and endoderm. For example, somatic cells would include both neurons and neural progenitors, the latter of which may be able to naturally give rise to all or some cell types of the central nervous system but cannot give rise to cells of the mesoderm or endoderm lineages.

[0055] By "mitotic cell" it is meant a cell undergoing mitosis. Mitosis is the process by which a eukaryotic cell separates the chromosomes in its nucleus into two identical sets in two separate nuclei. It is generally followed immediately by cytokinesis, which divides the nuclei, cytoplasm, organelles and cell membrane into two cells containing roughly equal shares of these cellular components.

[0056] By "post-mitotic cell" it is meant a cell that has exited from mitosis, i.e., it is "quiescent", i.e. it is no longer undergoing divisions. This quiescent state may be temporary, i.e. reversible, or it may be permanent.

[0057] By "meiotic cell" it is meant a cell that is undergoing meiosis. Meiosis is the process by which a cell divides its nuclear material for the purpose of producing gametes or spores. Unlike mitosis, in meiosis, the chromosomes undergo a recombination step which shuffles genetic material between chromosomes. Additionally, the outcome of meiosis is four (genetically unique) haploid cells, as compared with the two (genetically identical) diploid cells produced from mitosis.

[0058] By "recombination" it is meant a process of exchange of genetic information between two polynucleotides. As used herein, "homology-directed repair (HDR)" refers to the specialized form DNA repair that takes place, for example, during repair of double-strand breaks in cells. This process requires nucleotide sequence homology, uses a "donor" molecule to template repair of a "target" molecule (i.e., the one that experienced the double-strand break), and leads to the transfer of genetic information from the donor to the target. Homology-directed repair may result in an alteration of the sequence of the target molecule (e.g., insertion, deletion, mutation), if the donor polynucleotide differs from the target molecule and part or all of the sequence of the donor polynucleotide is incorporated into the target DNA.

[0059] By "non-homologous end joining (NHEJ)" it is meant the repair of double-strand breaks in DNA by direct ligation of the break ends to one another without the need for a homologous template (in contrast to homology-directed repair, which requires a homologous sequence to guide repair). NHEJ often results in the loss (deletion) of nucleotide sequence near the site of the double-strand break.

[0060] The terms "treatment", "treating" and the like are used herein to generally mean obtaining a desired pharmacologic and/or physiologic effect. The effect may be prophylactic in terms of completely or partially preventing a disease or symptom thereof and/or may be therapeutic in terms of a partial or complete cure for a disease and/or adverse effect attributable to the disease. "Treatment" as used herein covers any treatment of a disease or symptom in a mammal, and includes: (a) preventing the disease or symptom from occurring in a subject which may be predisposed to acquiring the disease or symptom but has not yet been diagnosed as having it; (b) inhibiting the disease or symptom, i.e., arresting its development; or (c) relieving the disease, i.e., causing regression of the disease. The therapeutic agent may be administered before, during or after the onset of disease or injury. The treatment of ongoing disease, where the treatment stabilizes or reduces the undesirable clinical symptoms of the patient, is of particular interest. Such treatment is desirably performed prior to complete loss of function in the affected tissues. The subject therapy will desirably be administered during the symptomatic stage of the disease, and in some cases after the symptomatic stage of the disease.

[0061] The terms "individual," "subject," "host," and "patient," are used interchangeably herein and refer to any mammalian subject for whom diagnosis, treatment, or therapy is desired, particularly humans.

[0062] The amino acids described herein are preferred to be in the "L" isomeric form. The amino acid sequences are given in one-letter code (A: alanine; C: cysteine; D: aspartic acid; E: glutamic acid; F: phenylalanine; G: glycine; H: histidine; I: isoleucine; K: lysine; L: leucine; M: methionine; N: asparagine; P: proline; Q: glutamine; R: arginine; S: serine; T: threonine; V: valine; W: tryptophan; Y: tyrosine; X: any residue). In keeping with standard polypeptide nomenclature, NH_2 refers to the free amino group present at the amino terminus (the N terminus) of a polypeptide, while COOH refers to the free carboxyl group present at the carboxyl terminus (the C terminus) of a polypeptide.

[0063] General methods in molecular and cellular biochemistry can be found in such standard textbooks as Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook et al., HaRBor Laboratory Press 2001); Short Protocols in Molecular Biology, 4th Ed. (Ausubel et al. eds., John Wiley & Sons 1999); Protein Methods (Bollag et al., John Wiley & Sons 1996); Nonviral Vectors for Gene Therapy (Wagner et al. eds., Academic Press 1999); Viral Vectors (Kaplift & Loewy eds., Academic Press 1995); Immunology Methods Manual (I. Lefkovits ed., Academic Press 1997); and Cell and Tissue Culture: Laboratory Procedures in Biotechnology (Doyle & Griffiths, John Wiley & Sons 1998), the disclosures of which are incorporated herein by reference. Reagents, cloning vectors, and kits for genetic manipulation referred to in this disclosure are available from commercial vendors such as BioRad, Stratagene, Invitrogen, Sigma-Aldrich, and ClonTech.

[0064] Before the present invention is further described, it is to be understood that this invention is not limited to particular embodiments described, as such may, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting, since the scope of the present invention will be limited only by the appended claims.

[0065] Where a range of values is provided, it is understood that each intervening value, to the tenth of the unit of the lower limit unless the context clearly dictates otherwise, between the upper and lower limit of that range and any other stated or intervening value in that stated range, is encompassed within the invention. The upper and lower limits of these smaller ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one

or both of the limits, ranges excluding either or both of those included limits are also included in the invention.

[0066] Certain ranges are presented herein with numerical values being preceded by the term "about." The term "about" is used herein to provide literal support for the exact number that it precedes, as well as a number that is near to or approximately the number that the term precedes. In determining whether a number is near to or approximately a specifically recited number, the near or approximating unrecited number may be a number which, in the context in which it is presented, provides the substantial equivalent of the specifically recited number.

[0067] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can also be used in the practice or testing of the present invention, the preferred methods and materials are now described. All publications mentioned herein are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited.

[0068] All publications and patents cited in this specification are herein incorporated by reference as if each individual publication or patent were specifically and individually indicated to be incorporated by reference and are incorporated herein by reference to disclose and describe the methods and/or materials in connection with which the publications are cited. The citation of any publication is for its disclosure prior to the filing date and should not be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

[0069] It is noted that as used herein and in the appended claims, the singular forms “a,” “an,” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to “a polynucleotide” includes a plurality of such polynucleotides and reference to “the polypeptide” includes reference to one or more polypeptides and equivalents thereof known to those skilled in the art, and so forth. It is further noted that the claims may be drafted to exclude any optional element. As such, this statement is intended to serve as antecedent basis for use of such exclusive terminology as “solely,” “only” and the like in connection with the recitation of claim elements, or use of a “negative” limitation.

[0070] It is appreciated that certain features of the invention, which are, for clarity, described in the context of separate embodiments, may also be provided in combination in a single embodiment. Conversely, various features of the invention, which are, for brevity, described in the context of

a single embodiment, may also be provided separately or in any suitable sub-combination. All combinations of the embodiments pertaining to the invention are specifically embraced by the present invention and are disclosed herein just as if each and every combination was individually and explicitly disclosed. In addition, all sub-combinations of the various embodiments and elements thereof are also specifically embraced by the present invention and are disclosed herein just as if each and every such sub-combination was individually and explicitly disclosed herein.

[0071] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

DETAILED DESCRIPTION

[0072] The present disclosure provides a DNA-targeting RNA that comprises a targeting sequence and, together with a modifying polypeptide, provides for site-specific modification of a target DNA and/or a polypeptide associated with the target DNA. The present disclosure further provides modifying polypeptides. The present disclosure further provides methods of site-specific modification of a target DNA and/or a polypeptide associated with the target DNA; and kits and compositions for carrying out the methods.

NUCLEIC ACIDS

DNA-targeting RNA

[0073] The present disclosure provides a DNA-targeting RNA that directs the activities of an associated polypeptide (e.g., a site-directed modifying polypeptide) to a specific target sequence within a target DNA. A subject DNA-targeting RNA comprises: a first segment (also referred to herein as a "DNA-targeting segment" or a "DNA-targeting sequence") and a second segment (also referred to herein as a "protein-binding segment" or a "protein-binding sequence").

DNA-targeting segment of a DNA-targeting RNA

[0074] The DNA-targeting segment of a subject DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA. In other words, the DNA-targeting segment of a subject DNA-targeting RNA interacts with a target DNA in a sequence-specific manner via hybridization (i.e., base pairing). As such, the nucleotide sequence of the DNA-targeting segment may vary and determines the location within the target DNA that the DNA-targeting RNA and the target DNA will interact. The DNA-targeting segment of a subject DNA-targeting RNA can be modified (e.g., by genetic engineering) to hybridize to any desired sequence within a target DNA.

[0075] The DNA-targeting segment can have a length of from about 10 nucleotides to about 100 nucleotides. For example, the DNA-targeting segment can have a length of from about 15 nucleotides (nt) to about 80 nt, from about 15 nt to about 50nt, from about 15 nt to about 40 nt, from about 15 nt to about 30 nt or from about 15 nt to about 25 nt. For example, the DNA-targeting segment can have a length of from about 15 nt to about 20 nt, from about 20 nt to about 25 nt, from about 25 nt to about 30 nt, from about 30 nt to about 35 nt, from about 35 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. The nucleotide sequence of the DNA-targeting segment that is complementary to a nucleotide sequence of the target DNA can have a length greater than about 10 nt. For example, the nucleotide sequence of the DNA-targeting segment that is complementary to a nucleotide sequence of the target DNA can have a length greater than about 10 nt, greater than about 15 nt, greater than about 18 nt, greater than about 20 nt, greater than about 25 nt, greater than about 30 nt, greater than about 35 nt or greater than about 40 nt. For example, the nucleotide sequence of the DNA-targeting segment that is complementary to a nucleotide sequence of the target DNA can have a length of from about 10 nt to about 15 nt, from about 15 nt to about 18 nt, from about 18 nt to about 20 nt, from about 20 nt to about 25 nt, from about 25 nt to about 30 nt, from about 30 nt to about 35 nt, from about 35 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. The percent complementarity between a nucleotide sequence of the DNA-targeting segment and a nucleotide sequence of the target DNA can be greater than 65%. For example, the percent complementarity between a nucleotide sequence of the DNA-targeting segment and a nucleotide sequence of the target DNA can be greater than 65%, greater than 70%, greater than 75%, greater than 80%, greater than 85%, greater than 90%, greater than 95%, or greater than 99%. In some cases, the percent complementarity between a nucleotide sequence of the DNA-targeting segment and a nucleotide sequence of the target DNA is 100%.

Protein-binding segment of a DNA-targeting RNA

[0076] The protein-binding segment of a subject DNA-targeting RNA interacts with a site-directed modifying polypeptide. The subject DNA-targeting RNA guides the bound polypeptide to a specific nucleotide sequence within target DNA via the above mentioned DNA-targeting segment. The protein-binding segment of a subject DNA-targeting RNA comprises two stretches of nucleotides that are complementary to one another. The complementary nucleotides of the

protein-binding segment hybridize to form a double stranded RNA duplex (dsRNA) (see Figures 1A and 1B).

[0077] In some embodiments, a subject DNA-targeting RNA comprises two separate RNA molecules (RNA polynucleotides) and is referred to herein as a "double-molecule DNA-targeting RNA" or a "two-molecule DNA-targeting RNA" (Figure 1A). In other embodiments, a subject DNA-targeting RNA is a single RNA molecule (single RNA polynucleotide) and is referred to herein as a "single-molecule DNA-targeting RNA" (Figure 1B). If not otherwise specified, the term "DNA-targeting RNA" is inclusive, referring to both single-molecule DNA-targeting RNAs and double-molecule DNA-targeting RNAs.

[0078] A subject double-molecule DNA-targeting RNA comprises two separate RNA molecules. Each of the two RNA molecules of a subject double-molecule DNA-targeting RNA comprises a stretch of nucleotides that are complementary to one another such that the complementary nucleotides of the two RNA molecules hybridize to form the double stranded RNA duplex of the protein-binding segment (Figure 1A).

[0079] A subject single-molecule DNA-targeting RNA comprises two stretches of nucleotides (a targeter-RNA and an activator-RNA) covalently linked by intervening nucleotides ("linkers" or "linker nucleotides") that are complementary to one another and hybridize to form the double stranded RNA duplex (dsRNA duplex) of the protein-binding segment, thus resulting in a stem-loop structure (Figure 1B). The targeter-RNA and the activator-RNA can be covalently linked via the 3' end of the targeter-RNA and the 5' end of the activator-RNA. Alternatively, targeter-RNA and the activator-RNA can be covalently linked via the 5' end of the targeter-RNA and the 3' end of the activator-RNA.

[0080] The linker of a single-molecule DNA-targeting RNA can have a length of from about 3 nucleotides to about 100 nucleotides. For example, the linker can have a length of from about 3 nucleotides (nt) to about 90 nt, from about 3 nucleotides (nt) to about 80 nt, from about 3 nucleotides (nt) to about 70 nt, from about 3 nucleotides (nt) to about 60 nt, from about 3 nucleotides (nt) to about 50 nt, from about 3 nucleotides (nt) to about 40 nt, from about 3 nucleotides (nt) to about 30 nt, from about 3 nucleotides (nt) to about 20 nt or from about 3 nucleotides (nt) to about 10 nt. For example, the linker can have a length of from about 3 nt to about 5 nt, from about 5 nt to about 10 nt, from about 10 nt to about 15 nt, from about 15 nt to about 20 nt, from about 20 nt to about 25 nt, from about 25 nt to about 30 nt, from about 30 nt to about 35 nt, from about 35 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. In some embodiments, the linker of a single-molecule DNA-targeting RNA is 4 nt.

[0081] With regard to both a subject single-molecule DNA-targeting RNA and to a subject double-molecule DNA-targeting RNA, the protein-binding segment can have a length of from about 10 nucleotides to about 100 nucleotides. For example, the protein-binding segment can have a length of from about 15 nucleotides (nt) to about 80 nt, from about 15 nt to about 50 nt, from about 15 nt to about 40 nt, from about 15 nt to about 30 nt or from about 15 nt to about 25 nt.

[0082] Also with regard to both a subject single-molecule DNA-targeting RNA and to a subject double-molecule DNA-targeting RNA, the dsRNA duplex of the protein-binding segment can have a length from about 8 base pairs (bp) to about 50bp. For example, the dsRNA duplex of the protein-binding segment can have a length from about 8 bp to about 40 bp, from about 8 bp to about 30bp, from about 8 bp to about 25 bp, from about 8 bp to about 20 bp or from about 8 bp to about 15 bp. For example, the dsRNA duplex of the protein-binding segment can have a length from about from about 8 bp to about 10 bp, from about 10 bp to about 15 bp, from about 15 bp to about 18 bp, from about 18 bp to about 20 bp, from about 20 bp to about 25 bp, from about 25 bp to about 30 bp, from about 30 bp to about 35 bp, from about 35 bp to about 40 bp, or from about 40 bp to about 50 bp. In some embodiments, the dsRNA duplex of the protein-binding segment has a length of 36 base pairs. The percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment can be greater than about 60%. For example, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment can be greater than about 65%, greater than about 70%, greater than about 75%, greater than about 80%, greater than about 85%, greater than about 90%, greater than about 95%, greater than about 98%, or greater than about 99% . In some cases, the percent complementarity between the nucleotide sequences that hybridize to form the dsRNA duplex of the protein-binding segment is 100%.

[0083] An exemplary double-molecule DNA-targeting RNA comprises a crRNA-like (CRISPR RNA or "targeter-RNA" or crRNA repeat) molecule and a corresponding tracrRNA-like (trans-acting CRISPR RNA or "activator-RNA" or tracrRNA) molecule (see Figure 1A). A crRNA-like molecule (targeter-RNA) comprises both the DNA-targeting segment (single stranded) of the DNA-targeting RNA and a stretch ("duplex-forming segment") of nucleotides that forms one half of the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. A corresponding tracrRNA-like molecule (activator-RNA) comprises a stretch of nucleotides (duplex-forming segment) that forms the other half of the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA (see Figure 1A). In other words, a stretch of nucleotides of a crRNA-like molecule are complementary to and hybridize with a stretch of nucleotides of a tracrRNA-like molecule to form the dsRNA duplex of the protein-binding domain of the DNA-

targeting RNA. As such, each crRNA-like molecule can be said to have a corresponding tracrRNA-like molecule. The crRNA-like molecule additionally provides the single stranded DNA-targeting segment. Thus, a crRNA-like and a tracrRNA-like molecule (as a corresponding pair) hybridize to form a DNA-targeting RNA (see Figure 1A). The exact sequence of a given crRNA or tracrRNA molecule is characteristic of the species in which the RNA molecules are found. Various crRNAs and tracrRNAs are depicted individually and in corresponding complementary pairs in Figures 6-9. A subject double-molecule DNA-targeting RNA can comprise any corresponding crRNA and tracrRNA pair.

[0084] The term "activator-RNA" is used herein to mean a tracrRNA-like molecule of a double-molecule DNA-targeting RNA. The term "targeter-RNA" is used herein to mean a crRNA-like molecule of a double-molecule DNA-targeting RNA. The term "duplex-forming segment" is used herein to mean the stretch of nucleotides of an activator-RNA or a targeter-RNA that contributes to the formation of the dsRNA duplex by hybridizing to a stretch of nucleotides of a corresponding activator-RNA or targeter-RNA molecule. In other words, an activator-RNA comprises a duplex-forming segment that is complementary to the duplex-forming segment of the corresponding targeter-RNA. As such, an activator-RNA comprises a duplex-forming segment while a targeter-RNA comprises both a duplex-forming segment and the DNA-targeting segment of the DNA-targeting RNA. Therefore, a subject double-molecule DNA-targeting RNA can be comprised of any corresponding activator-RNA and targeter-RNA pair.

[0085] In some embodiments, the duplex-forming segment of the activator-RNA is greater than about 60% identical to one of the activator-RNA (tracrRNA) molecules depicted in Figures 6, 8, and 9 over a stretch of at least 25 contiguous nucleotides. For example, the duplex-forming segment of the activator-RNA (or the DNA encoding the duplex-forming segment of the activator-RNA) is greater than about 65% identical, greater than about 70% identical, greater than about 75% identical, greater than about 80% identical, greater than about 85% identical, greater than about 90% identical, greater than about 95% identical, greater than about 98% identical, greater than about 99% identical, or 100 % identical, to one of the tracrRNA sequences depicted in Figures 6, 8 and 9 over a stretch of at least 25 contiguous nucleotides.

[0086] In some embodiments, the duplex-forming segment of the targeter-RNA is greater than about 60% identical to one of the targeter-RNA (crRNA) molecules depicted in Figures 7-9 over a stretch of at least 12 contiguous nucleotides. For example, the duplex-forming segment of the targeter-RNA (or the DNA encoding the duplex-forming segment of the targeter-RNA) is greater than about 65% identical, greater than about 70% identical, greater than about 75% identical, greater than about 80% identical, greater than about 85% identical, greater than about 90% identical, greater than about 95% identical, greater than about 98% identical, greater than about

99% identical or 100 % identical to one of the crRNA sequences depicted in Figures 7-9 over a stretch of at least 12 contiguous nucleotides.

[0087] An exemplary single-molecule DNA-targeting RNA comprises two complementary stretches of nucleotides that hybridize to form a dsRNA duplex. In some embodiments, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is greater than about 60% identical to one of the activator-RNA (tracrRNA) molecules depicted in Figures 6, 8, and 9 over a stretch of at least 25 contiguous nucleotides. For example, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is greater than about 65% identical, greater than about 70% identical, greater than about 75% identical, greater than about 80% identical, greater than about 85% identical, greater than about 90% identical, greater than about 95% identical, greater than about 98% identical, greater than about 99% identical or 100 % identical to one of the tracrRNA sequences depicted in Figures 6, 8 and 9 over a stretch of at least 25 contiguous nucleotides.

[0088] In some embodiments, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is greater than about 60% identical to one of the targeter-RNA (crRNA) molecules depicted in Figures 7-9 over a stretch of at least 12 contiguous nucleotides. For example, one of the two complementary stretches of nucleotides of the single-molecule DNA-targeting RNA (or the DNA encoding the stretch) is greater than about 65% identical, greater than about 70% identical, greater than about 75% identical, greater than about 80% identical, greater than about 85% identical, greater than about 90% identical, greater than about 95% identical, greater than about 98% identical, greater than about 99% identical or 100 % identical to one of the crRNA sequences depicted in Figures 7-9 over a stretch of at least 12 contiguous nucleotides.

Site-directed modifying polypeptide

[0089] A subject DNA-targeting RNA and a subject site-directed modifying polypeptide form a complex. The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA (as noted above). The site-directed modifying polypeptide of the complex provides the site-specific activity. In other words, the site-directed modifying polypeptide is guided to a DNA sequence (e.g. a chromosomal sequence or an extrachromosomal sequence, e.g. an episomal sequence, a minicircle sequence, a mitochondrial sequence, a chloroplast sequence, etc.) by virtue of its association with at least the protein-binding segment of the DNA-targeting RNA (described above).

[0090] A subject site-directed modifying polypeptide modifies target DNA (e.g., cleavage or methylation of target DNA) and/or a polypeptide associated with target DNA (e.g., methylation or acetylation of a histone tail). A site-directed modifying polypeptide is also referred to herein as a "site-directed polypeptide" or an "RNA binding site-directed modifying polypeptide."

[0091] In some cases, the site-directed modifying polypeptide is a naturally-occurring modifying polypeptide. In other cases, the site-directed modifying polypeptide is not a naturally-occurring polypeptide (e.g., a chimeric polypeptide as discussed below or a naturally-occurring polypeptide that is modified, e.g., mutation, deletion, insertion).

[0092] Exemplary naturally-occurring site-directed modifying polypeptides are provided in Figure 12 as a non-limiting and non-exhaustive list of naturally occurring Cas9/Csn1 endonucleases. These naturally occurring polypeptides, as disclosed herein, bind a DNA-targeting RNA, are thereby directed to a specific sequence within a target DNA, and cleave the target DNA to generate a double strand break. A subject site-directed modifying polypeptide comprises two portions, an RNA-binding portion and an activity portion. In some embodiments, a subject site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that exhibits site-directed enzymatic activity (e.g., activity for DNA methylation, activity for DNA cleavage, activity for histone acetylation, activity for histone methylation, etc.), wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[0093] In other embodiments, a subject site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that modulates transcription within the target DNA (e.g., to increase or decrease transcription), wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[0094] In some cases, a subject site-directed modifying polypeptide has enzymatic activity that modifies target DNA (e.g., nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity).

[0095] In other cases, a subject site-directed modifying polypeptide has enzymatic activity that modifies a polypeptide (e.g., a histone) associated with target DNA (e.g., methyltransferase

activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity).

Exemplary site-directed modifying polypeptides

[0096] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 2, or to the corresponding domains in any of the amino acid sequences depicted in Figure 12.

Nucleic acid modifications

[0097] In some embodiments, a subject nucleic acid (e.g., a DNA-targeting RNA) comprises one or more modifications, e.g., a base modification, a backbone modification, etc. As is known in the art, a nucleoside is a base-sugar combination. The base portion of the nucleoside is normally a heterocyclic base. The two most common classes of such heterocyclic bases are the purines and the pyrimidines. Nucleotides are nucleosides that further include a phosphate group covalently linked to the sugar portion of the nucleoside. For those nucleosides that include a pentofuranosyl sugar, the phosphate group can be linked to the 2', the 3', or the 5' hydroxyl moiety of the sugar. In forming oligonucleotides, the phosphate groups covalently link adjacent nucleosides to one another to form a linear polymeric compound. In turn, the respective ends of this linear polymeric compound can be further joined to form a circular compound, however, linear compounds are generally suitable. In addition, linear compounds may have internal nucleotide base complementarity and may therefore fold in a manner as to produce a fully or partially double-stranded compound. Within oligonucleotides, the phosphate groups are commonly referred to as forming the internucleoside backbone of the oligonucleotide. The normal linkage or backbone of RNA and DNA is a 3' to 5' phosphodiester linkage.

Modified backbones and modified internucleoside linkages

[0098] Examples of suitable nucleic acids containing modifications include nucleic acids containing modified backbones or non-natural internucleoside linkages. Nucleic acids (having modified backbones include those that retain a phosphorus atom in the backbone and those that do not have a phosphorus atom in the backbone.

[0099] Suitable modified oligonucleotide backbones containing a phosphorus atom therein include, for example, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates, 5'-alkylene phosphonates and chiral phosphonates, phosphinates,

phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, phosphorodiamidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, selenophosphates and boranophosphates having normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein one or more internucleotide linkages is a 3' to 3', 5' to 5' or 2' to 2' linkage. Suitable oligonucleotides having inverted polarity comprise a single 3' to 3' linkage at the 3'-most internucleotide linkage i.e. a single inverted nucleoside residue which may be a basic (the nucleobase is missing or has a hydroxyl group in place thereof). Various salts (such as, for example, potassium or sodium), mixed salts and free acid forms are also included.

[00100] In some embodiments, a subject nucleic acid comprises one or more phosphorothioate and/or heteroatom internucleoside linkages, in particular $-CH_2-NH-O-CH_2-$, $-CH_2-N(CH_3)-O-CH_2-$ (known as a methylene (methylimino) or MMI backbone), $-CH_2-O-N(CH_3)-CH_2-$, $-CH_2-N(CH_3)-N(CH_3)-CH_2-$ and $-O-N(CH_3)-CH_2-CH_2-$ (wherein the native phosphodiester internucleotide linkage is represented as $-O-P(=O)(OH)-O-CH_2-$). MMI type internucleoside linkages are disclosed in the above referenced U.S. Pat. No. 5,489,677. Suitable amide internucleoside linkages are disclosed in t U.S. Pat. No. 5,602,240.

[00101] Also suitable are nucleic acids having morpholino backbone structures as described in, e.g., U.S. Pat. No. 5,034,506. For example, in some embodiments, a subject nucleic acid comprises a 6-membered morpholino ring in place of a ribose ring. In some of these embodiments, a phosphorodiamidate or other non-phosphodiester internucleoside linkage replaces a phosphodiester linkage.

[00102] Suitable modified polynucleotide backbones that do not include a phosphorus atom therein have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; riboacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts.

Mimetics

[00103] A subject nucleic acid can be a nucleic acid mimetic. The term "mimetic" as it is applied to polynucleotides is intended to include polynucleotides wherein only the furanose ring or both the furanose ring and the internucleotide linkage are replaced with non-furanose groups, replacement

of only the furanose ring is also referred to in the art as being a sugar surrogate. The heterocyclic base moiety or a modified heterocyclic base moiety is maintained for hybridization with an appropriate target nucleic acid. One such nucleic acid, a polynucleotide mimetic that has been shown to have excellent hybridization properties, is referred to as a peptide nucleic acid (PNA). In PNA, the sugar-backbone of a polynucleotide is replaced with an amide containing backbone, in particular an aminoethylglycine backbone. The nucleotides are retained and are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone.

[00104] One polynucleotide mimetic that has been reported to have excellent hybridization properties is a peptide nucleic acid (PNA). The backbone in PNA compounds is two or more linked aminoethylglycine units which gives PNA an amide containing backbone. The heterocyclic base moieties are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone. Representative U.S. patents that describe the preparation of PNA compounds include, but are not limited to: U.S. Pat. Nos. 5,539,082; 5,714,331; and 5,719,262.

[00105] Another class of polynucleotide mimetic that has been studied is based on linked morpholino units (morpholino nucleic acid) having heterocyclic bases attached to the morpholino ring. A number of linking groups have been reported that link the morpholino monomeric units in a morpholino nucleic acid. One class of linking groups has been selected to give a non-ionic oligomeric compound. The non-ionic morpholino-based oligomeric compounds are less likely to have undesired interactions with cellular proteins. Morpholino-based polynucleotides are non-ionic mimics of oligonucleotides which are less likely to form undesired interactions with cellular proteins (Dwaine A. Braasch and David R. Corey, Biochemistry, 2002, 41(14), 4503-4510). Morpholino-based polynucleotides are disclosed in U.S. Pat. No. 5,034,506. A variety of compounds within the morpholino class of polynucleotides have been prepared, having a variety of different linking groups joining the monomeric subunits.

[00106] A further class of polynucleotide mimetic is referred to as cyclohexenyl nucleic acids (CeNA). The furanose ring normally present in a DNA/RNA molecule is replaced with a cyclohexenyl ring. CeNA DMT protected phosphoramidite monomers have been prepared and used for oligomeric compound synthesis following classical phosphoramidite chemistry. Fully modified CeNA oligomeric compounds and oligonucleotides having specific positions modified with CeNA have been prepared and studied (see Wang et al., J. Am. Chem. Soc., 2000, 122, 8595-8602). In general the incorporation of CeNA monomers into a DNA chain increases its stability of a DNA/RNA hybrid. CeNA oligoadenylates formed complexes with RNA and DNA complements with similar stability to the native complexes. The study of incorporating CeNA structures into natural nucleic acid structures was shown by NMR and circular dichroism to proceed with easy conformational adaptation.

[00107] A further modification includes Locked Nucleic Acids (LNAs) in which the 2'-hydroxyl group is linked to the 4' carbon atom of the sugar ring thereby forming a 2'-C,4'-C-oxymethylene linkage thereby forming a bicyclic sugar moiety. The linkage can be a methylene (-CH_2-), group bridging the 2' oxygen atom and the 4' carbon atom wherein n is 1 or 2 (Singh et al., Chem. Commun., 1998, 4, 455-456). LNA and LNA analogs display very high duplex thermal stabilities with complementary DNA and RNA (Tm=+3 to +10° C), stability towards 3'-exonucleolytic degradation and good solubility properties. Potent and nontoxic antisense oligonucleotides containing LNAs have been described (Wahlestedt et al., Proc. Natl. Acad. Sci. U.S.A., 2000, 97, 5633-5638).

[00108] The synthesis and preparation of the LNA monomers adenine, cytosine, guanine, 5-methyl-cytosine, thymine and uracil, along with their oligomerization, and nucleic acid recognition properties have been described (Koshkin et al., Tetrahedron, 1998, 54, 3607-3630). LNAs and preparation thereof are also described in WO 98/39352 and WO 99/14226.

Modified sugar moieties

[00109] A subject nucleic acid can also include one or more substituted sugar moieties. Suitable polynucleotides comprise a sugar substituent group selected from: OH; F; O-, S-, or N-alkyl; O-, S-, or N-alkenyl; O-, S- or N-alkynyl; or O-alkyl-O-alkyl, wherein the alkyl, alkenyl and alkynyl may be substituted or unsubstituted C.sub.1 to C_{10} alkyl or C_2 to C_{10} alkenyl and alkynyl. Particularly suitable are $O((CH_2)_nO)_mCH_3$, $O(CH_2)_nOCH_3$, $O(CH_2)_nNH_2$, $O(CH_2)_nCH_3$, $O(CH_2)_nONH_2$, and $O(CH_2)_nON((CH_2)_nCH_3)_2$, where n and m are from 1 to about 10. Other suitable polynucleotides comprise a sugar substituent group selected from: C_1 to C_{10} lower alkyl, substituted lower alkyl, alkenyl, alkynyl, alkaryl, aralkyl, O-alkaryl or O-aralkyl, SH, SCH_3, OCN, Cl, Br, CN, CF_3, OCF_3, $SOCH_3$, SO_2CH_3, ONO_2, NO_2, N_3, NH_2, heterocycloalkyl, heterocycloalkaryl, aminoalkylamino, polyalkylamino, substituted silyl, an RNA cleaving group, a reporter group, an intercalator, a group for improving the pharmacokinetic properties of an oligonucleotide, or a group for improving the pharmacodynamic properties of an oligonucleotide, and other substituents having similar properties. A suitable modification includes 2'-methoxyethoxy (2'-O-CH_2 CH_2OCH_3, also known as 2'-O-(2-methoxyethyl) or 2'-MOE) (Martin et al., Helv. Chim. Acta, 1995, 78, 486-504) i.e., an alkoxyalkoxy group. A further suitable modification includes 2'-dimethylaminooxyethoxy, i.e., a $O(CH_2)_2ON(CH_3)_2$ group, also known as 2'-DMAOE, as described in examples hereinbelow, and 2'-dimethylaminoethoxyethoxy (also known in the art as 2'-O-dimethyl-amino-ethoxy-ethyl or 2'-DMAEOE), i.e., 2'-O-CH_2-O-CH_2-$N(CH_3)_2$.

[00110] Other suitable sugar substituent groups include methoxy (-O-CH_3), aminopropoxy (--O CH_2 CH_2 CH_2NH_2), allyl (-CH_2-CH=CH_2), -O-allyl (--O-- CH_2—CH=CH_2) and fluoro (F). 2'-sugar

substituent groups may be in the arabino (up) position or ribo (down) position. A suitable 2'-arabino modification is 2'-F. Similar modifications may also be made at other positions on the oligomeric compound, particularly the 3' position of the sugar on the 3' terminal nucleoside or in 2'-5' linked oligonucleotides and the 5' position of 5' terminal nucleotide. Oligomeric compounds may also have sugar mimetics such as cyclobutyl moieties in place of the pentofuranosyl sugar.

Base modifications and substitutions

[00111] A subject nucleic acid may also include nucleobase (often referred to in the art simply as "base") modifications or substitutions. As used herein, "unmodified" or "natural" nucleobases include the purine bases adenine (A) and guanine (G), and the pyrimidine bases thymine (T), cytosine (C) and uracil (U). Modified nucleobases include other synthetic and natural nucleobases such as 5-methylcytosine (5-me-C), 5-hydroxymethyl cytosine, xanthine, hypoxanthine, 2-aminoadenine, 6-methyl and other alkyl derivatives of adenine and guanine, 2-propyl and other alkyl derivatives of adenine and guanine, 2-thiouracil, 2-thiothymine and 2-thiocytosine, 5-halouracil and cytosine, 5-propynyl (-C=C-CH_3) uracil and cytosine and other alkynyl derivatives of pyrimidine bases, 6-azo uracil, cytosine and thymine, 5-uracil (pseudouracil), 4-thiouracil, 8-halo, 8-amino, 8-thiol, 8-thioalkyl, 8-hydroxyl and other 8-substituted adenines and guanines, 5-halo particularly 5-bromo, 5-trifluoromethyl and other 5-substituted uracils and cytosines, 7-methylguanine and 7-methyladenine, 2-F-adenine, 2-amino-adenine, 8-azaguanine and 8-azaadenine, 7-deazaguanine and 7-deazaadenine and 3-deazaguanine and 3-deazaadenine. Further modified nucleobases include tricyclic pyrimidines such as phenoxazine cytidine(1H-pyrimido(5,4-b)(1,4)benzoxazin-2(3H)-one), phenothiazine cytidine (1H-pyrimido(5,4-b)(1,4)benzothiazin-2(3H)-one), G-clamps such as a substituted phenoxazine cytidine (e.g. 9-(2-aminoethoxy)-H-pyrimido(5,4-(b) (1,4)benzoxazin-2(3H)-one), carbazole cytidine (2H-pyrimido(4,5-b)indol-2-one), pyridoindole cytidine (H-pyrido(3',2':4,5)pyrrolo(2,3-d)pyrimidin-2-one).

[00112] Heterocyclic base moieties may also include those in which the purine or pyrimidine base is replaced with other heterocycles, for example 7-deaza-adenine, 7-deazaguanosine, 2-aminopyridine and 2-pyridone. Further nucleobases include those disclosed in U.S. Pat. No. 3,687,808, those disclosed in The Concise Encyclopedia Of Polymer Science And Engineering, pages 858-859, Kroschwitz, J. I., ed. John Wiley & Sons, 1990, those disclosed by Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613, and those disclosed by Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B., ed., CRC Press, 1993. Certain of these nucleobases are useful for increasing the binding affinity of an oligomeric compound. These include 5-substituted pyrimidines, 6-azapyrimidines and N-2, N-6 and O-6 substituted purines, including 2-aminopropyladenine, 5-

propynyluracil and 5-propynylcytosine. 5-methylcytosine substitutions have been shown to increase nucleic acid duplex stability by 0.6-1.2° C. (Sanghvi et al., eds., Antisense Research and Applications, CRC Press, Boca Raton, 1993, pp. 276-278) and are suitable base substitutions, e.g., when combined with 2'-O-methoxyethyl sugar modifications.

Conjugates

[00113] Another possible modification of a subject nucleic acid involves chemically linking to the polynucleotide one or more moieties or conjugates which enhance the activity, cellular distribution or cellular uptake of the oligonucleotide. These moieties or conjugates can include conjugate groups covalently bound to functional groups such as primary or secondary hydroxyl groups. Conjugate groups include, but are not limited to, intercalators, reporter molecules, polyamines, polyamides, polyethylene glycols, polyethers, groups that enhance the pharmacodynamic properties of oligomers, and groups that enhance the pharmacokinetic properties of oligomers. Suitable conjugate groups include, but are not limited to, cholesterols, lipids, phospholipids, biotin, phenazine, folate, phenanthridine, anthraquinone, acridine, fluoresceins, rhodamines, coumarins, and dyes. Groups that enhance the pharmacodynamic properties include groups that improve uptake, enhance resistance to degradation, and/or strengthen sequence-specific hybridization with the target nucleic acid. Groups that enhance the pharmacokinetic properties include groups that improve uptake, distribution, metabolism or excretion of a subject nucleic acid.

[00114] Conjugate moieties include but are not limited to lipid moieties such as a cholesterol moiety (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989, 86, 6553-6556), cholic acid (Manoharan et al., Bioorg. Med. Chem. Let., 1994, 4, 1053-1060), a thioether, e.g., hexyl-S-tritylthiol (Manoharan et al., Ann. N.Y. Acad. Sci., 1992, 660, 306-309; Manoharan et al., Bioorg. Med. Chem. Let., 1993, 3, 2765-2770), a thiocholesterol (Oberhauser et al., Nucl. Acids Res., 1992, 20, 533-538), an aliphatic chain, e.g., dodecandiol or undecyl residues (Saison-Behmoaras et al., EMBO J., 1991, 10, 1111-1118; Kabanov et al., FEBS Lett., 1990, 259, 327-330; Svinarchuk et al., Biochimie, 1993, 75, 49-54), a phospholipid, e.g., di-hexadecyl-rac-glycerol or triethylammonium 1,2-di-O-hexadecyl-rac-glycero-3-H-phosphonate (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654; Shea et al., Nucl. Acids Res., 1990, 18, 3777-3783), a polyamine or a polyethylene glycol chain (Manoharan et al., Nucleosides & Nucleotides, 1995, 14, 969-973), or adamantane acetic acid (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654), a palmityl moiety (Mishra et al., Biochim. Biophys. Acta, 1995, 1264, 229-237), or an octadecylamine or hexylamino-carbonyl-oxycholesterol moiety (Crooke et al., J. Pharmacol. Exp. Ther., 1996, 277, 923-937.\

[00115] A conjugate may include a "Protein Transduction Domain" or PTD (also known as a CPP – cell penetrating peptide), which may refer to a polypeptide, polynucleotide, carbohydrate, or organic or inorganic compound that facilitates traversing a lipid bilayer, micelle, cell membrane, organelle membrane, or vesicle membrane. A PTD attached to another molecule, which can range from a small polar molecule to a large macromolecule and/or a nanoparticle, facilitates the molecule traversing a membrane, for example going from extracellular space to intracellular space, or cytosol to within an organelle. In some embodiments, a PTD is covalently linked to the amino terminus of an exogenous polypeptide (e.g., a site-directed modifying polypeptide). In some embodiments, a PTD is covalently linked to the carboxyl terminus of an exogenous polypeptide (e.g., a site-directed modifying polypeptide). In some embodiments, a PTD is covalently linked to a nucleic acid (e.g., a DNA-targeting RNA, a polynucleotide encoding a DNA-targeting RNA, a polynucleotide encoding a site-directed modifying polypeptide, etc.). Exemplary PTDs include but are not limited to a minimal undecapeptide protein transduction domain (corresponding to residues 47-57 of HIV-1 TAT comprising YGRKKRRQRRR; SEQ ID NO://); a polyarginine sequence comprising a number of arginines sufficient to direct entry into a cell (e.g., 3, 4, 5, 6, 7, 8, 9, 10, or 10-50 arginines); a VP22 domain (Zender et al. (2002) *Cancer Gene Ther.* 9(6):489-96); an Drosophila Antennapedia protein transduction domain (Noguchi et al. (2003) *Diabetes* 52(7):1732-1737); a truncated human calcitonin peptide (Trehin et al. (2004) *Pharm. Research* 21:1248-1256); polylysine (Wender et al. (2000) *Proc. Natl. Acad. Sci. USA* 97:13003-13008); RRQRRTSKLMKR (SEQ ID NO:/); Transportan GWTLNSAGYLLGKINLKALAALAKKIL (SEQ ID NO:/); KALAWEAKLAKALAKALAKHLAKALAKALKCEA (SEQ ID NO:/); and RQIKIWFQNRRMKWKK (SEQ ID NO:/). Exemplary PTDs include but are not limited to, YGRKKRRQRRR (SEQ ID NO:/), RKKRRQRRR (SEQ ID NO:/); an arginine homopolymer of from 3 arginine residues to 50 arginine residues; Exemplary PTD domain amino acid sequences include, but are not limited to, any of the following: YGRKKRRQRRR (SEQ ID NO://); RKKRRQRR (SEQ ID NO:2); YARAAARQARA (SEQ ID NO://); THRLPRRRRRR (SEQ ID NO://); and GGRRARRRRRR (SEQ ID NO://). In some embodiments, the PTD is an activatable CPP (ACPP) (Aguilera et al. (2009) *Integr Biol (Camb)* June; 1(5-6): 371-381). ACPPs comprise a polycationic CPP (typically Arg9 or R9) connected via a cleavable linker to a matching polyanion (typically Glu9 or E9), which reduces the net charge to nearly zero and thereby inhibits adhesion and uptake into cells. Upon cleavage of the linker, the polyanion is released, locally unmasking the polyarginine and its inherent adhesiveness, thus "activating" the ACPP to traverse the membrane.

Exemplary DNA-targeting RNAs

[00116] In some embodiments, a suitable DNA-targeting RNA comprises two separate RNA polynucleotide molecules. The first of the two separate RNA polynucleotide molecules (the activator-RNA) comprises a nucleotide sequence having at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C. The second of the two separate RNA polynucleotide molecules (the targeter-RNA) comprises a nucleotide sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence identity over a stretch of at least 12 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 7A-C.

[00117] In some embodiments, a suitable DNA-targeting RNA is a single RNA polynucleotide and comprises a first nucleotide sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C and a second nucleotide sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100% nucleotide sequence identity over a stretch of at least 12 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 7A-C.

[00118] In some embodiments, the DNA-targeting RNA is a double-molecule DNA-targeting RNA and the targeter-RNA comprises the sequence 5'GUUUUAGAGCUA-3'linked at its 5' end to a stretch of nucleotides that are complementary to a target DNA. In some embodiments, the DNA-targeting RNA is a double-molecule DNA-targeting RNA and the activator-RNA comprises the sequence 5' UAGCAAGUUAAAAUAAGGCUAGUCCG-3'.

[00119] In some embodiments, the DNA-targeting RNA is a single-molecule DNA-targeting RNA and comprises the sequence 5'-GUUUUAGAGCUA-linker-UAGCAAGUUAAAAUAAGGCUAGUCCG-3' linked at its 5' end to a stretch of nucleotides that are complementary to a target DNA (where "linker" denotes any a linker nucleotide sequence that can comprise any nucleotide sequence). The linker of a single-molecule DNA-targeting RNA can have a length of from about 3 nucleotides to about 100 nucleotides. For example, the linker can have a length of from about 3 nucleotides (nt) to about 90nt, from about 3 nucleotides (nt) to about 80nt, from about 3 nucleotides (nt) to about 70nt, from about 3 nucleotides (nt) to about 60nt, from about 3 nucleotides (nt) to about 50nt, from about 3

nucleotides (nt) to about 40nt, from about 3 nucleotides (nt) to about 30nt, from about 3 nucleotides (nt) to about 20nt or from about 3 nucleotides (nt) to about 10nt. In some embodiments, the linker has a length of from about 3 nt to about 5 nt, from about 5 nt to about 10 nt, from about 10 nt to about 15 nt, from about 15 nt to about 20 nt, from about 20 nt to about 25 nt, from about 25 nt to about 30 nt, from about 30 nt to about 35 nt, from about 35 nt to about 40 nt, from about 40 nt to about 50 nt, from about 50 nt to about 60 nt, from about 60 nt to about 70 nt, from about 70 nt to about 80 nt, from about 80 nt to about 90 nt, or from about 90 nt to about 100 nt. In some embodiments, the linker of a single-molecule DNA-targeting RNA is 4nt.

Nucleic acids encoding a subject DNA-targeting RNA and/or a subject site-directed modifying polypeptide

[00120] The present disclosure provides a nucleic acid comprising a nucleotide sequence encoding a subject DNA-targeting RNA and/or a subject site-directed modifying polypeptide. In some embodiments, a subject DNA-targeting RNA-encoding nucleic acid is an expression vector, e.g., a recombinant expression vector.

[00121] In some embodiments, a subject method involves contacting a target DNA or introducing into a cell (or a population of cells) one or more nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide. In some embodiments a cell comprising a target DNA is *in vitro.* In some embodiments a cell comprising a target DNA is *in vivo.* Suitable nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide include expression vectors, where an expression vector comprising a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is a "recombinant expression vector."

[00122] In some embodiments, the recombinant expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, a recombinant retroviral construct, etc.

[00123] Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson

et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00124] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

[00125] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00126] In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide in both prokaryotic and eukaryotic cells.

[00127] Non-limiting examples of suitable eukaryotic promoters (promoters functional in a eukaryotic cell) include those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector may also include appropriate sequences for amplifying expression. The expression vector may also include nucleotide sequences encoding protein tags (e.g., 6xHis tag, hemagglutinin tag, green fluorescent protein, etc.) that are fused to the site-directed modifying polypeptide, thus resulting in a chimeric polypeptide.

[00128] In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to an inducible promoter. In some

embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to a constitutive promoter.

[00129] Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a stem cell or progenitor cell. Suitable methods include, e.g., infection, lipofection, electroporation, calcium phosphate precipitation, polyethyleneimine (PEI)-mediated transfection, DEAE-dextran mediated transfection, liposome-mediated transfection, and the like.

CHIMERIC POLYPEPTIDES

[00130] The present disclosure provides a chimeric site-directed modifying polypeptide. A subject chimeric site-directed modifying polypeptide interacts with (e.g., binds to) a subject DNA-targeting RNA (described above). The DNA-targeting RNA guides the chimeric site-directed modifying polypeptide to a target sequence within target DNA (e.g. a chromosomal sequence or an extrachromosomal sequence, e.g. an episomal sequence, a minicircle sequence, a mitochondrial sequence, a chloroplast sequence, etc.). A subject chimeric site-directed modifying polypeptide modifies target DNA (e.g., cleavage or methylation of target DNA) and/or a polypeptide associated with target DNA (e.g., methylation or acetylation of a histone tail).

[00131] A subject chimeric site-directed modifying polypeptide modifies target DNA (e.g., cleavage or methylation of target DNA) and/or a polypeptide associated with target DNA (e.g., methylation or acetylation of a histone tail). A chimeric site-directed modifying polypeptide is also referred to herein as a "chimeric site-directed polypeptide" or a "chimeric RNA binding site-directed modifying polypeptide."

[00132] A subject chimeric site-directed modifying polypeptide comprises two portions, an RNA-binding portion and an activity portion. A subject chimeric site-directed modifying polypeptide comprises amino acid sequences that are derived from at least two different polypeptides. A subject chimeric site-directed modifying polypeptide can comprise modified and/or naturally-occurring polypeptide sequences (e.g., a first amino acid sequence from a modified or unmodified Cas9/Csn1 protein; and a second amino acid sequence other than the Cas9/Csn1 protein).

<u>RNA-binding portion</u>

[00133] In some cases, the RNA-binding portion of a subject chimeric site-directed modifying polypeptide is a naturally-occurring polypeptide. In other cases, the RNA-binding portion of a subject chimeric site-directed modifying polypeptide is not a naturally-occurring molecule (modified, e.g., mutation, deletion, insertion). Naturally-occurring RNA-binding portions of

interest are derived from site-directed modifying polypeptides known in the art. For example, Figure 12 provides a non-limiting and non-exhaustive list of naturally occurring Cas9/Csn1 endonucleases that can be used as site-directed modifying polypeptides. In some cases, the RNA-binding portion of a subject chimeric site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or 100%, amino acid sequence identity to the RNA-binding portion of a polypeptide having any of the amino acid sequences set forth in Figure 12 (Figures 12A-V).

[00134] In some cases, the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99%, or 100%, amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 2, or to the corresponding domains in any of the amino acid sequences depicted in Figure 12 (Figures 12A-V).

Activity portion

[00135] In addition to the RNA-binding portion, the chimeric site-directed modifying polypeptide comprises an "activity portion." In some embodiments, the activity portion of a subject chimeric site-directed modifying polypeptide comprises the naturally-occurring activity portion of a site-directed modifying polypeptide (e.g., Cas9/Csn1 endonuclease). In other embodiments, the activity portion of a subject chimeric site-directed modifying polypeptide comprises a modified amino acid sequence (e.g., substitution, deletion, insertion) of a naturally-occurring activity portion of a site-directed modifying polypeptide. Naturally-occurring activity portions of interest are derived from site-directed modifying polypeptides known in the art. For example, Figure 12 provides a non-limiting and non-exhaustive list of naturally occurring Cas9/Csn1 endonucleases that can be used as site-directed modifying polypeptides. The activity portion of a subject chimeric site-directed modifying polypeptide is variable and may comprise any heterologous polypeptide sequence that may be useful in the methods disclosed herein.

[00136] In some embodiments, a subject chimeric site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that exhibits site-directed enzymatic activity (e.g., activity for DNA methylation, activity for DNA cleavage, activity for histone acetylation, activity for histone methylation, etc.), wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00137] In other embodiments, a subject chimeric site-directed modifying polypeptide comprises: (i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting

RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) an activity portion that modulates transcription within the target DNA (e.g., to increase or decrease transcription), wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00138] In some cases, the activity portion of a subject chimeric site-directed modifying polypeptide has enzymatic activity that modifies target DNA (e.g., nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity).

[00139] In other cases, the activity portion of a subject chimeric site-directed modifying polypeptide has enzymatic activity (e.g., methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity) that modifies a polypeptide associated with target DNA (e.g., a histone).

[00140] In some cases, the activity portion of a subject chimeric site-directed modifying polypeptide exhibits enzymatic activity (described above). In other cases, the activity portion of a subject chimeric site-directed modifying polypeptide modulates transcription of the target DNA (described above). The activity portion of a subject chimeric site-directed modifying polypeptide is variable and may comprise any heterologous polypeptide sequence that may be useful in the methods disclosed herein.

Exemplary chimeric site-directed modifying polypeptides

[00141] In some embodiments, the activity portion of the chimeric site-directed modifying polypeptide comprises a modified form of the Cas9/Csn1 protein. In some instances, the modified form of the Cas9/Csn1 protein comprises an amino acid change (e.g., deletion, insertion, or substitution) that reduces the naturally-occurring nuclease activity of the Cas9/Csn1 protein. For example, in some instances, the modified form of the Cas9/Csn1 protein has less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nuclease activity of the corresponding wild-type Cas9/Csn1 polypeptide. In some cases, the modified form of the Cas9/Csn1 polypeptide has substantially no nuclease activity.

[00142] In yet other embodiments, the activity portion of the site-directed modifying polypeptide comprises a heterologous polypeptide that has DNA-modifying activity and/or transcription

factor activity and/or DNA-associated polypeptide-modifying activity. In some cases, a heterologous polypeptide replaces a portion of the Cas9/Csn1 polypeptide that provides nuclease activity. In other embodiments, a subject site-directed modifying polypeptide comprises both a portion of the Cas9/Csn1 polypeptide that provides nuclease activity and a heterologous polypeptide. In other words, in some cases, a subject chimeric site-directed modifying polypeptide is a fusion polypeptide comprising both the portion of the Cas9/Csn1 polypeptide that provides nuclease activity and the heterologous polypeptide. In other cases, a subject chimeric site-directed modifying polypeptide is a fusion polypeptide comprising a modified variant of the activity portion of the Cas9/Csn1 polypeptide (e.g., amino acid change, deletion, insertion) and a heterologous polypeptide. In yet other cases, a subject chimeric site-directed modifying polypeptide is a fusion polypeptide comprising a heterologous polypeptide and the RNA-binding portion of a naturally-occurring or a modified site-directed modifying polypeptide.

[00143] In some cases, the chimeric site-directed modifying polypeptide comprises an amino acid sequence having at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 99% or 100% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 2, or to the corresponding domains in any of the amino acid sequences depicted in Figure 12 (Figures 12A-V).

[00144] For example, in a chimeric Cas9/Csn1 protein, the RNA-binding domain of a naturally-occurring (or modified, e.g., mutation, deletion, insertion) bacterial Cas9/Csn1 polypeptide may be fused to a heterologous polypeptide sequence (i.e. a polypeptide sequence from a protein other than Cas9/Csn1 or a polypeptide sequence from another organism). The heterologous polypeptide sequence may exhibit an activity (e.g., enzymatic activity) that will also be exhibited by the chimeric Cas9/Csn1 protein (e.g., methyltransferase activity, acetyltransferase activity, kinase activity, ubiquitinating activity, etc.). A heterologous nucleic acid sequence may be linked to a naturally-occurring nucleic acid sequence (e.g., by genetic engineering) to generate a chimeric nucleotide sequence encoding a chimeric polypeptide.

Nucleic acid encoding a subject chimeric site-directed modifying polypeptide

[00145] The present disclosure provides a nucleic acid comprising a nucleotide sequence encoding a subject chimeric site-directed modifying polypeptide. In some embodiments, the nucleic acid comprising a nucleotide sequence encoding a subject chimeric site-directed modifying polypeptide is an expression vector, e.g., a recombinant expression vector.

[00146] In some embodiments, a subject method involves contacting a target DNA or introducing into a cell (or a population of cells) one or more nucleic acids comprising a chimeric site-directed modifying polypeptide. Suitable nucleic acids comprising nucleotide sequences

encoding a chimeric site-directed modifying polypeptide include expression vectors, where an expression vector comprising a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is a "recombinant expression vector."

[00147] In some embodiments, the recombinant expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, etc.

[00148] Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00149] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

[00150] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00151] In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in either a eukaryotic cell, e.g., a mammalian cell; or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is

operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a chimeric site-directed modifying polypeptide in both prokaryotic and eukaryotic cells.

[00152] Non-limiting examples of suitable eukaryotic promoters (promoters functional in a eukaryotic cell) include those from cytomegalovirus (CMV) immediate early, herpes simplex virus (HSV) thymidine kinase, early and late SV40, long terminal repeats (LTRs) from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art. The expression vector may also contain a ribosome binding site for translation initiation and a transcription terminator. The expression vector may also include appropriate sequences for amplifying expression. The expression vector may also include nucleotide sequences encoding protein tags (e.g., 6xHis tag, HA tag, green fluorescent protein, etc.) that are fused to the chimeric site-directed modifying polypeptide.

[00153] In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to an inducible promoter. In some embodiments, a nucleotide sequence encoding a chimeric site-directed modifying polypeptide is operably linked to a constitutive promoter.

[00154] Methods of introducing a nucleic acid into a host cell are known in the art, and any known method can be used to introduce a nucleic acid (e.g., an expression construct) into a stem cell or progenitor cell. Suitable methods include, e.g., infection, lipofection, electroporation, calcium phosphate precipitation, DEAE-dextran mediated transfection, liposome-mediated transfection, microinjection, and the like.

METHODS

[00155] The present disclosure provides methods for modifying a target DNA and/or a target DNA-associated polypeptide. Generally, a subject method involves contacting a target DNA with a complex (a "targeting complex"), which complex comprises a DNA-targeting RNA and a site-directed modifying polypeptide.

[00156] As discussed above, a subject DNA-targeting RNA and a subject site-directed modifying polypeptide form a complex. The DNA-targeting RNA provides target specificity to the complex by comprising a nucleotide sequence that is complementary to a sequence of a target DNA. The site-directed modifying polypeptide of the complex provides the site-specific activity. In some embodiments, a subject complex modifies a target DNA, leading to, for example, DNA cleavage, DNA methylation, DNA damage, DNA repair, etc. In other embodiments, a subject complex modifies a target polypeptide associated with target DNA (e.g., a histone), leading to, for example, histone methylation, histone acetylation, and the like. The target DNA may be, for

example, naked DNA *in vitro*, chromosomal DNA in cells *in vitro*, chromosomal DNA in cells *in vivo*, etc.

[00157] In some cases, the site-directed modifying polypeptide exhibits nuclease activity that cleaves target DNA at a target DNA sequence defined by the region of complementarity between the DNA-targeting RNA and the target DNA. The nuclease activity cleaves target DNA to produce double strand breaks. These breaks are then repaired by the cell in one of two ways: non-homologous end joining, and homology-directed repair (Figure 4). In non-homologous end joining (NHEJ), the double-strand breaks are repaired by direct ligation of the break ends to one another. As such, no new nucleic acid material is inserted into the site, although some nucleic acid material may be lost, resulting in a deletion. In homology-directed repair, a donor polynucleotide with homology to the cleaved target DNA sequence is used as a template for repair of the cleaved target DNA sequence, resulting in the transfer of genetic information from the donor polynucleotide to the target DNA. As such, new nucleic acid material may be inserted/copied into the site. In some cases, a target DNA is contacted with a subject donor polynucleotide. In some cases, a subject donor polynucleotide is introduced into a subject cell. The modifications of the target DNA due to NHEJ and/or homology-directed repair lead to, for example, gene correction, gene replacement, gene tagging, transgene insertion, nucleotide deletion, gene disruption, gene mutation, etc.

[00158] Accordingly, cleavage of DNA by a site-directed modifying polypeptide may be used to delete nucleic acid material from a target DNA sequence, for example, to disrupt a gene that makes cells susceptible to infection (e.g. the CCR5 or CXCR4 gene, which makes T cells susceptible to HIV infection), to remove disease-causing trinucleotide repeat sequences in neurons, to create gene knockouts and mutations as disease models in research, by cleaving the target DNA sequence and allowing the cell to repair the sequence in the absence of an exogenously provided donor polynucleotide.

[00159] Alternatively, if a DNA-targeting RNA and a site-directed modifying polypeptide are coadministered to cells with a donor polynucleotide sequence comprising homology to the target DNA sequence, the subject methods may be used to add, i.e. insert or replace, nucleic acid material to a target DNA sequence (e.g. to "knock in" a nucleic acid that encodes for a protein, an siRNA, an miRNA, etc.), to add a tag (e.g., 6xHis, a fluorescent protein (e.g., a green fluorescent protein; a yellow fluorescent protein, etc.), hemagglutinin (HA), FLAG, etc.), or to add a regulatory sequence to a gene (e.g. promoter, polyadenylation signal, internal ribosome entry sequence (IRES), 2A peptide, start codon, stop codon, splice signal, localization signal, etc.). As such, a complex comprising a DNA-targeting RNA and a site-directed modifying polypeptide is useful in any *in vitro* or *in vivo* application in which it is desirable to modify DNA

in a site-specific, i.e. "targeted", way, for example gene knock-out, gene knock-in, gene editing, gene tagging, etc., as used in, for example, gene therapy, e.g. to treat a disease or as an antiviral, antipathogenic, or anticancer therapeutic, the production of genetically modified organisms in agriculture, the large scale production of proteins by cells for therapeutic, diagnostic, or research purposes, the induction of iPS cells, biological research, the targeting of genes of pathogens for deletion or replacement, etc.

[00160] In some embodiments, a subject DNA-targeting RNA and a subject site-directed modifying polypeptide are used as an inducible system for shutting off gene expression in bacterial cells. In some cases, nucleic acids encoding an appropriate DNA-targeting RNA and an appropriate site-directed polypeptide are incorporated into the chromosome of a target cell and are under control of an inducible promoter. When the DNA-targeting RNA and the site-directed polypeptide are induced, the target DNA is cleaved at the location of interest (e.g., a target gene on a separate plasmid). As such, in some cases, bacterial expression strains are engineered to include nucleic acid sequences encoding an appropriate site-directed modifying polypeptide in the bacterial genome and an appropriate DNA-targeting RNA on a plasmid (e.g., under control of an inducible promoter), allowing experiments in which the expression of any targeted gene (expressed from a separate plasmid introduced into the strain) could be controlled by inducing expression of the DNA-targeting RNA and the site-directed polypeptide.

[00161] In some cases, the site-directed modifying polypeptide has enzymatic activity that modifies target DNA in ways other than introducing double strand breaks. Enzymatic activity of interest that may be used to modify target DNA (e.g., by fusing a heterologous polypeptide with enzymatic activity to a site-directed modifying polypeptide, thereby generating a chimeric site-directed modifying polypeptide) includes, but is not limited methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity). Methylation and demethylation is recognized in the art as an important mode of epigenetic gene regulation while DNA damage and repair activity is essential for cell survival and for proper genome maintenance in response to environmental stresses.

[00162] As such, the methods herein find use in the epigenetic modification of target DNA and may be employed to control epigenetic modification of target DNA at any location in a target DNA by genetically engineering the desired complementary nucleic acid sequence into the DNA-targeting segment of a DNA-targeting RNA. The methods herein also find use in the intentional and controlled damage of DNA at any desired location within the target DNA. The methods

herein also find use in the sequence-specific and controlled repair of DNA at any desired location within the target DNA. Methods to target DNA-modifying enzymatic activities to specific locations in target DNA find use in both research and clinical applications.

[00163] In some cases, the site-directed modifying polypeptide has activity that modulates the transcription of target DNA. In some cases, a chimeric site-directed modifying polypeptides comprising a heterologous polypeptide that exhibits the ability to increase or decrease transcription (e.g., transcriptional activator or transcription repressor polypeptides) is used to increase or decrease the transcription of target DNA at a specific location in a target DNA by altering the DNA-targeting segment of the DNA-targeting RNA. As such, in some cases, the subject method is used to control the expression of a targeted coding-RNA (protein-encoding gene) or a targeted non-coding RNA (e.g., tRNA, rRNA, snoRNA, siRNA, miRNA, long ncRNA, etc.).

[00164] In some cases, the site-directed modifying polypeptide has enzymatic activity that modifies a polypeptide associated with DNA (e.g. histone). In some embodiments, the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, demyristoylation activity glycosylation activity (e.g., from O-GlcNAc transferase) or deglycosylation activity. The enzymatic activities listed herein catalyze covalent modifications to proteins. Such modifications are known in the art to alter the stability or activity of the target protein (e.g., phosphorylation due to kinase activity can stimulate or silence protein activity depending on the target protein). Of particular interest as protein targets are histones. Histone proteins are known in the art to bind DNA and form complexes known as nucleosomes. Histones can be modified (e.g., by methylation, acetylation, ubuitination, phosphorylation) to elicit structural changes in the surrounding DNA, thus controlling the accessibility of potentially large portions of DNA to interacting factors such as transcription factors, polymerases and the like. A single histone can be modified in many different ways and in many different combinations (e.g., trimethylation of lysine 27 of histone 3, H3K27, is associated with DNA regions of repressed transcription while trimethylation of lysine 4 of histone 3, H3K4, is associated with DNA regions of active transcription). Thus, a site-directed modifying polypeptide with histone-modifying activity finds use in the site specific control of DNA structure and can be used to alter the histone modification pattern in a selected region of target DNA. Such methods find use in both research and clinical applications.

Target cells of interest

[00165] In some of the above applications, the subject methods may be employed to induce DNA cleavage and DNA modification in mitotic or post-mitotic cells *in vitro* and/or *ex vivo* and/or *in vitro* (e.g., to produce genetically modified cells that can be reintroduced into an individual). Because the DNA-targeting RNA provide specificity by hybridizing to target DNA, a mitotic and/or post-mitotic cell of interest in the disclosed methods may include a cell from any organism (e.g. a bacterial cell, an archaeal cell, a cell of a single-cell eukaryotic organism, a plant cell, an animal cell, a cell from an invertebrate animal (e.g. fruit fly, cnidarian, echinoderm, nematode, etc.), a cell from a vertebrate animal (e.g., fish, amphibian, reptile, bird, mammal), a cell from a mammal, a cell from a rodent, a cell from a human, etc.). Any type of cell may be of interest (e.g. a stem cell, e.g. an embryonic stem (ES) cell, an induced pluripotent stem (iPS) cell, a germ cell; a somatic cell, e.g. a fibroblast, a hematopoietic cell, a neuron, a muscle cell, a bone cell, a hepatocyte, a pancreatic cell etc.). Cells may be from established cell lines or they may be primary cells, where "primary cells", "primary cell lines", and "primary cultures" are used interchangeably herein to refer to cells and cells cultures that have been derived from a subject and allowed to grow in vitro for a limited number of passages, i.e. splittings, of the culture. For example, primary cultures are cultures that may have been passaged 0 times, 1 time, 2 times, 4 times, 5 times, 10 times, or 15 times, but not enough times go through the crisis stage. Typically, the primary cell lines of the present invention are maintained for fewer than 10 passages in vitro.

[00166] If the cells are primary cells, they may be harvest from an individual by any convenient method. For example, leukocytes may be conveniently harvested by apheresis, leukocytapheresis, density gradient separation, etc., while cells from tissues such as skin, muscle, bone marrow, spleen, liver, pancreas, lung, intestine, stomach, etc. are most conveniently harvested by biopsy. An appropriate solution may be used for dispersion or suspension of the harvested cells. Such solution will generally be a balanced salt solution, e.g. normal saline, phosphate-buffered saline (PBS), Hank's balanced salt solution, etc., conveniently supplemented with fetal calf serum or other naturally occurring factors, in conjunction with an acceptable buffer at low concentration, generally from 5-25 mM. Convenient buffers include HEPES, phosphate buffers, lactate buffers, etc. The cells may be used immediately, or they may be stored, frozen, for long periods of time, being thawed and capable of being reused. In such cases, the cells will usually be frozen in 10% DMSO, 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperatures, and thawed in a manner as commonly known in the art for thawing frozen cultured cells.

Nucleic acids encoding a subject DNA-targeting RNA and/or a subject site-directed modifying polypeptide

[00167] In some embodiments, a subject method involves contacting a target DNA or introducing into a cell (or a population of cells) one or more nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a donor polynucleotide. Suitable nucleic acids comprising nucleotide sequences encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide include expression vectors, where an expression vector comprising a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is a "recombinant expression vector."

[00168] In some embodiments, the recombinant expression vector is a viral construct, e.g., a recombinant adeno-associated virus construct (see, e.g., U.S. Patent No. 7,078,387), a recombinant adenoviral construct, a recombinant lentiviral construct, etc.

[00169] Suitable expression vectors include, but are not limited to, viral vectors (e.g. viral vectors based on vaccinia virus; poliovirus; adenovirus (see, e.g., Li et al., Invest Opthalmol Vis Sci 35:2543 2549, 1994; Borras et al., Gene Ther 6:515 524, 1999; Li and Davidson, PNAS 92:7700 7704, 1995; Sakamoto et al., H Gene Ther 5:1088 1097, 1999; WO 94/12649, WO 93/03769; WO 93/19191; WO 94/28938; WO 95/11984 and WO 95/00655); adeno-associated virus (see, e.g., Ali et al., Hum Gene Ther 9:81 86, 1998, Flannery et al., PNAS 94:6916 6921, 1997; Bennett et al., Invest Opthalmol Vis Sci 38:2857 2863, 1997; Jomary et al., Gene Ther 4:683 690, 1997, Rolling et al., Hum Gene Ther 10:641 648, 1999; Ali et al., Hum Mol Genet 5:591 594, 1996; Srivastava in WO 93/09239, Samulski et al., J. Vir. (1989) 63:3822-3828; Mendelson et al., Virol. (1988) 166:154-165; and Flotte et al., PNAS (1993) 90:10613-10617); SV40; herpes simplex virus; human immunodeficiency virus (see, e.g., Miyoshi et al., PNAS 94:10319 23, 1997; Takahashi et al., J Virol 73:7812 7816, 1999); a retroviral vector (e.g., Murine Leukemia Virus, spleen necrosis virus, and vectors derived from retroviruses such as Rous Sarcoma Virus, Harvey Sarcoma Virus, avian leukosis virus, a lentivirus, human immunodeficiency virus, myeloproliferative sarcoma virus, and mammary tumor virus); and the like.

[00170] Numerous suitable expression vectors are known to those of skill in the art, and many are commercially available. The following vectors are provided by way of example; for eukaryotic host cells: pXT1, pSG5 (Stratagene), pSVK3, pBPV, pMSG, and pSVLSV40 (Pharmacia). However, any other vector may be used so long as it is compatible with the host cell.

[00171] In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to a control element, e.g., a transcriptional control element, such as a promoter. The transcriptional control element may be functional in

either a eukaryotic cell, e.g., a mammalian cell, or a prokaryotic cell (e.g., bacterial or archaeal cell). In some embodiments, a nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide is operably linked to multiple control elements that allow expression of the nucleotide sequence encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide in both prokaryotic and eukaryotic cells.

[00172] Depending on the host/vector system utilized, any of a number of suitable transcription and translation control elements, including constitutive and inducible promoters, transcription enhancer elements, transcription terminators, etc. may be used in the expression vector (see e.g., Bitter et al. (1987) *Methods in Enzymology*, 153:516-544).

[00173] In some embodiments, a DNA-targeting RNA may be provided directly as RNA. In such cases, the DNA-targeting RNA may be produced by direct chemical synthesis or may be transcribed *in vitro* from a DNA encoding the DNA-targeting RNA. Methods of synthesizing RNA from a DNA template are well known in the art. In some cases, the DNA-targeting RNA will be synthesized *in vitro* using an RNA polymerase enzyme (e.g., T7 polymerase, T3 polymerase, SP6 polymerase, etc.). Once synthesized, the RNA may directly contact a target DNA or may be introduced into a cell by any of the well-known techniques for introducing nucleic acids into cells (e.g., microinjection, electroporation, transfection, etc).

[00174] Nucleotides encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a donor polynucleotide may be provided to the cells using well-developed transfection techniques; see, e.g. Angel and Yanik (2010) PLoS ONE 5(7): e11756, and the commercially available TransMessenger® reagents from Qiagen, Stemfect™ RNA Transfection Kit from Stemgent, and TransIT®-mRNA Transfection Kit from Mirus Bio LLC. See also Beumer et al. (2008) Efficient gene targeting in Drosophila by direct embryo injection with zinc-finger nucleases. PNAS 105(50):19821-19826. Alternatively, nucleic acids encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide may be provided on DNA vectors. Many vectors, e.g. plasmids, cosmids, minicircles, phage, viruses, etc., useful for transferring nucleic acids into target cells are available. The vectors comprising the nucleic acid(s) may be maintained episomally, e.g. as plasmids, minicircle DNAs, viruses such cytomegalovirus, adenovirus, etc., or they may be integrated into the target cell genome, through homologous recombination or random integration, e.g. retrovirus-derived vectors such as MMLV, HIV-1, ALV, etc.

[00175] Vectors may be provided directly to the subject cells. In other words, the cells are contacted with vectors comprising the nucleic acid encoding DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide such that the vectors are taken up by the cells. Methods for contacting cells with

nucleic acid vectors that are plasmids, such as electroporation, calcium chloride transfection, microinjection, and lipofection are well known in the art. For viral vector delivery, the cells are contacted with viral particles comprising the nucleic acid encoding a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide. Retroviruses, for example, lentiviruses, are particularly suitable to the method of the invention. Commonly used retroviral vectors are "defective", i.e. unable to produce viral proteins required for productive infection. Rather, replication of the vector requires growth in a packaging cell line. To generate viral particles comprising nucleic acids of interest, the retroviral nucleic acids comprising the nucleic acid are packaged into viral capsids by a packaging cell line. Different packaging cell lines provide a different envelope protein (ecotropic, amphotropic or xenotropic) to be incorporated into the capsid, this envelope protein determining the specificity of the viral particle for the cells (ecotropic for murine and rat; amphotropic for most mammalian cell types including human, dog and mouse; and xenotropic for most mammalian cell types except murine cells). The appropriate packaging cell line may be used to ensure that the cells are targeted by the packaged viral particles. Methods of introducing the retroviral vectors comprising the nucleic acid encoding the reprogramming factors into packaging cell lines and of collecting the viral particles that are generated by the packaging lines are well known in the art.

[00176] Vectors used for providing the nucleic acids encoding DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide to the subject cells will typically comprise suitable promoters for driving the expression, that is, transcriptional activation, of the nucleic acid of interest. In other words, the nucleic acid of interest will be operably linked to a promoter. This may include ubiquitously acting promoters, for example, the CMV-β-actin promoter, or inducible promoters, such as promoters that are active in particular cell populations or that respond to the presence of drugs such as tetracycline. By transcriptional activation, it is intended that transcription will be increased above basal levels in the target cell by at least about 10 fold, by at least about 100 fold, more usually by at least about 1000 fold. In addition, vectors used for providing a DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide to the subject cells may include nucleic acid sequences that encode for selectable markers in the target cells, so as to identify cells that have taken up the DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide and/or a donor polynucleotide.

[00177] A subject DNA-targeting RNA and/or a site-directed modifying polypeptide and/or a chimeric site-directed modifying polypeptide may instead be used to contact DNA or introduced

into cells as RNA. Methods of introducing RNA into cells are known in the art and may include, for example, direct injection, transfection, or any other method used for the introduction of DNA.

[00178] A subject site-directed modifying polypeptide may instead be provided to cells as a polypeptide. Such a polypeptide may optionally be fused to a polypeptide domain that increases solubility of the product. The domain may be linked to the polypeptide through a defined protease cleavage site, e.g. a TEV sequence, which is cleaved by TEV protease. The linker may also include one or more flexible sequences, e.g. from 1 to 10 glycine residues. In some embodiments, the cleavage of the fusion protein is performed in a buffer that maintains solubility of the product, e.g. in the presence of from 0.5 to 2 M urea, in the presence of polypeptides and/or polynucleotides that increase solubility, and the like. Domains of interest include endosomolytic domains, e.g. influenza HA domain; and other polypeptides that aid in production, e.g. IF2 domain, GST domain, GRPE domain, and the like. The polypeptide may be formulated for improved stability. For example, the peptides may be PEGylated, where the polyethyleneoxy group provides for enhanced lifetime in the blood stream.

[00179] Additionally or alternatively, the subject site-directed modifying polypeptide may be fused to a polypeptide permeant domain to promote uptake by the cell. A number of permeant domains are known in the art and may be used in the non-integrating polypeptides of the present invention, including peptides, peptidomimetics, and non-peptide carriers. For example, a permeant peptide may be derived from the third alpha helix of Drosophila melanogaster transcription factor Antennapaedia, referred to as penetratin, which comprises the amino acid sequence RQIKIWFQNRRMKWKK. As another example, the permeant peptide comprises the HIV-1 tat basic region amino acid sequence, which may include, for example, amino acids 49-57 of naturally-occurring tat protein. Other permeant domains include poly-arginine motifs, for example, the region of amino acids 34-56 of HIV-1 rev protein, nona-arginine, octa-arginine, and the like. (See, for example, Futaki et al. (2003) Curr Protein Pept Sci. 2003 Apr; 4(2): 87-96; and Wender et al. (2000) Proc. Natl. Acad. Sci. U.S.A 2000 Nov. 21; 97(24):13003-8; published U.S. Patent applications 20030220334; 20030083256; 20030032593; and 20030022831, herein specifically incorporated by reference for the teachings of translocation peptides and peptoids). The nona-arginine (R9) sequence is one of the more efficient PTDs that have been characterized (Wender et al. 2000; Uemura et al. 2002). The site at which the fusion is made may be selected in order to optimize the biological activity, secretion or binding characteristics of the polypeptide. The optimal site will be determined by routine experimentation.

[00180] A subject site-directed modifying polypeptide may be produced *in* vitro or by eukaryotic cells or by prokaryotic cells, and it may be further processed by unfolding, e.g. heat denaturation, DTT reduction, etc. and may be further refolded, using methods known in the art.

[00181] Modifications of interest that do not alter primary sequence include chemical derivatization of polypeptides, e.g., acylation, acetylation, carboxylation, amidation, etc. Also included are modifications of glycosylation, e.g. those made by modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps; e.g. by exposing the polypeptide to enzymes which affect glycosylation, such as mammalian glycosylating or deglycosylating enzymes. Also embraced are sequences that have phosphorylated amino acid residues, e.g. phosphotyrosine, phosphoserine, or phosphothreonine.

[00182] Also included in the subject invention are DNA-targeting RNAs and site-directed modifying polypeptides that have been modified using ordinary molecular biological techniques and synthetic chemistry so as to improve their resistance to proteolytic degradation, to change the target sequence specificity, to optimize solubility properties, to alter protein activity (e.g., transcription modulatory activity, enzymatic activity, etc) or to render them more suitable as a therapeutic agent. Analogs of such polypeptides include those containing residues other than naturally occurring L-amino acids, e.g. D-amino acids or non-naturally occurring synthetic amino acids. D-amino acids may be substituted for some or all of the amino acid residues.

[00183] The site-directed modifying polypeptides may be prepared by in vitro synthesis, using conventional methods as known in the art. Various commercial synthetic apparatuses are available, for example, automated synthesizers by Applied Biosystems, Inc., Beckman, etc. By using synthesizers, naturally occurring amino acids may be substituted with unnatural amino acids. The particular sequence and the manner of preparation will be determined by convenience, economics, purity required, and the like.

[00184] If desired, various groups may be introduced into the peptide during synthesis or during expression, which allow for linking to other molecules or to a surface. Thus cysteines can be used to make thioethers, histidines for linking to a metal ion complex, carboxyl groups for forming amides or esters, amino groups for forming amides, and the like.

[00185] The site-directed modifying polypeptides may also be isolated and purified in accordance with conventional methods of recombinant synthesis. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion chromatography, gel electrophoresis, affinity chromatography, or other purification technique. For the most part, the compositions which are used will comprise at least 20% by weight of the desired product, more usually at least about 75% by weight, preferably at least about 95% by weight, and for

therapeutic purposes, usually at least about 99.5% by weight, in relation to contaminants related to the method of preparation of the product and its purification. Usually, the percentages will be based upon total protein.

[00186] To induce DNA cleavage and recombination, or any desired modification to a target DNA, or any desired modification to a polypeptide associated with target DNA, the DNA-targeting RNA and/or the site-directed modifying polypeptide and/or the donor polynucleotide, whether they be introduced as nucleic acids or polypeptides, are provided to the cells for about 30 minutes to about 24 hours, e.g., 1 hour, 1.5 hours, 2 hours, 2.5 hours, 3 hours, 3.5 hours 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 12 hours, 16 hours, 18 hours, 20 hours, or any other period from about 30 minutes to about 24 hours, which may be repeated with a frequency of about every day to about every 4 days, e.g., every 1.5 days, every 2 days, every 3 days, or any other frequency from about every day to about every four days. The agent(s) may be provided to the subject cells one or more times, e.g. one time, twice, three times, or more than three times, and the cells allowed to incubate with the agent(s) for some amount of time following each contacting event e.g. 16-24 hours, after which time the media is replaced with fresh media and the cells are cultured further.

[00187] In cases in which two or more different targeting complexes are provided to the cell (e.g., two different DNA-targeting RNAs that are complementary to different sequences within the same or different target DNA), the complexes may be provided simultaneously (e.g. as two polypeptides and/or nucleic acids), or delivered simultaneously. Alternatively, they may be provided consecutively, e.g. the targeting complex being provided first, followed by the second targeting complex, etc. or vice versa.

[00188] Typically, an effective amount of the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide is provided to the target DNA or cells to induce cleavage. An effective amount of the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide is the amount to induce a 2--fold increase or more in the amount of target modification observed between two homologous sequences relative to a negative control, e.g. a cell contacted with an empty vector or irrelevant polypeptide. That is to say, an effective amount or dose of the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide will induce a 2-fold increase, a 3-fold increase, a 4-fold increase or more in the amount of target modification observed at a target DNA region, in some instances a 5-fold increase, a 6-fold increase or more, sometimes a 7-fold or 8-fold increase or more in the amount of recombination observed, e.g. an increase of 10-fold, 50-fold, or 100-fold or more, in some instances, an increase of 200-fold, 500-fold, 700-fold, or 1000-fold or more, e.g. a 5000-fold, or 10,000-fold increase in the amount of recombination observed. The amount

of target modification may be measured by any convenient method. For example, a silent reporter construct comprising complementary sequence to the targeting segment (targeting sequence) of the DNA-targeting RNA flanked by repeat sequences that, when recombined, will reconstitute a nucleic acid encoding an active reporter may be cotransfected into the cells, and the amount of reporter protein assessed after contact with the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide, e.g. 2 hours, 4 hours, 8 hours, 12 hours, 24 hours, 36 hours, 48 hours, 72 hours or more after contact with the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide. As another, more sensitivity assay, for example, the extent of recombination at a genomic DNA region of interest comprising target DNA sequences may be assessed by PCR or Southern hybridization of the region after contact with a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide, e.g. 2 hours, 4 hours, 8 hours, 12 hours, 24 hours, 36 hours, 48 hours, 72 hours or more after contact with the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide.

[00189] Contacting the cells with a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may occur in any culture media and under any culture conditions that promote the survival of the cells. For example, cells may be suspended in any appropriate nutrient medium that is convenient, such as Iscove's modified DMEM or RPMI 1640, supplemented with fetal calf serum or heat inactivated goat serum (about 5-10%), L-glutamine, a thiol, particularly 2-mercaptoethanol, and antibiotics, e.g. penicillin and streptomycin. The culture may contain growth factors to which the cells are responsive. Growth factors, as defined herein, are molecules capable of promoting survival, growth and/or differentiation of cells, either in culture or in the intact tissue, through specific effects on a transmembrane receptor. Growth factors include polypeptides and non-polypeptide factors. Conditions that promote the survival of cells are typically permissive of nonhomologous end joining and homology-directed repair.

[00190] In applications in which it is desirable to insert a polynucleotide sequence into a target DNA sequence, a polynucleotide comprising a donor sequence to be inserted is also provided to the cell. By a "donor sequence" or "donor polynucleotide" it is meant a nucleic acid sequence to be inserted at the cleavage site induced by a site-directed modifying polypeptide. The donor polynucleotide will contain sufficient homology to a genomic sequence at the cleavage site, e.g. 70%, 80%, 85%, 90%, 95%, or 100% homology with the nucleotide sequences flanking the cleavage site, e.g. within about 50 bases or less of the cleavage site, e.g. within about 30 bases, within about 15 bases, within about 10 bases, within about 5 bases, or immediately flanking the cleavage site, to support homology-directed repair between it and the genomic sequence to which it bears homology. Approximately 25, 50, 100, or 200 nucleotides, or more than 200

nucleotides, of sequence homology between a donor and a genomic sequence (or any integral value between 10 and 200 nucleotides, or more) will support homology-directed repair. Donor sequences can be of any length, e.g. 10 nucleotides or more, 50 nucleotides or more, 100 nucleotides or more, 250 nucleotides or more, 500 nucleotides or more, 1000 nucleotides or more, 5000 nucleotides or more, etc.

[00191] The donor sequence is typically not identical to the genomic sequence that it replaces. Rather, the donor sequence may contain at least one or more single base changes, insertions, deletions, inversions or rearrangements with respect to the genomic sequence, so long as sufficient homology is present to support homology-directed repair. In some embodiments, the donor sequence comprises a non-homologous sequence flanked by two regions of homology, such that homology-directed repair between the target DNA region and the two flanking sequences results in insertion of the non-homologous sequence at the target region. Donor sequences may also comprise a vector backbone containing sequences that are not homologous to the DNA region of interest and that are not intended for insertion into the DNA region of interest. Generally, the homologous region(s) of a donor sequence will have at least 50% sequence identity to a genomic sequence with which recombination is desired. In certain embodiments, 60%, 70%, 80%, 90%, 95%, 98%, 99%, or 99.9% sequence identity is present. Any value between 1% and 100% sequence identity can be present, depending upon the length of the donor polynucleotide.

[00192] The donor sequence may comprise certain sequence differences as compared to the genomic sequence, e.g. restriction sites, nucleotide polymorphisms, selectable markers (e.g., drug resistance genes, fluorescent proteins, enzymes etc.), etc., which may be used to assess for successful insertion of the donor sequence at the cleavage site or in some cases may be used for other purposes (e.g., to signify expression at the targeted genomic locus). In some cases, if located in a coding region, such nucleotide sequence differences will not change the amino acid sequence, or will make silent amino acid changes (i.e., changes which do not affect the structure or function of the protein). Alternatively, these sequences differences may include flanking recombination sequences such as FLPs, loxP sequences, or the like, that can be activated at a later time for removal of the marker sequence.

[00193] The donor sequence may be provided to the cell as single-stranded DNA, single-stranded RNA, double-stranded DNA, or double-stranded RNA. It may be introduced into a cell in linear or circular form. If introduced in linear form, the ends of the donor sequence may be protected (e.g., from exonucleolytic degradation) by methods known to those of skill in the art. For example, one or more dideoxynucleotide residues are added to the 3' terminus of a linear molecule and/or self-complementary oligonucleotides are ligated to one or both ends. See, for

example, Chang et al. (1987) Proc. Natl. Acad Sci USA 84:4959-4963; Nehls et al. (1996) Science 272:886-889. Additional methods for protecting exogenous polynucleotides from degradation include, but are not limited to, addition of terminal amino group(s) and the use of modified internucleotide linkages such as, for example, phosphorothioates, phosphoramidates, and O-methyl ribose or deoxyribose residues. As an alternative to protecting the termini of a linear donor sequence, additional lengths of sequence may be included outside of the regions of homology that can be degraded without impacting recombination. A donor sequence can be introduced into a cell as part of a vector molecule having additional sequences such as, for example, replication origins, promoters and genes encoding antibiotic resistance. Moreover, donor sequences can be introduced as naked nucleic acid, as nucleic acid complexed with an agent such as a liposome or poloxamer, or can be delivered by viruses (e.g., adenovirus, AAV), as described above for nucleic acids encoding a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide.

[00194] Following the methods described above, a DNA region of interest may be cleaved and modified, i.e. "genetically modified", ex vivo. In some embodiments, as when a selectable marker has been inserted into the DNA region of interest, the population of cells may be enriched for those comprising the genetic modification by separating the genetically modified cells from the remaining population. Separation may be by any convenient separation technique appropriate for the selectable marker used. For example, if a fluorescent marker has been inserted, cells may be separated by fluorescence activated cell sorting, whereas if a cell surface marker has been inserted, cells may be separated from the heterogeneous population by affinity separation techniques, e.g. magnetic separation, affinity chromatography, "panning" with an affinity reagent attached to a solid matrix, or other convenient technique. Techniques providing accurate separation include fluorescence activated cell sorters, which can have varying degrees of sophistication, such as multiple color channels, low angle and obtuse light scattering detecting channels, impedance channels, etc. The cells may be selected against dead cells by employing dyes associated with dead cells (e.g. propidium iodide). Any technique may be employed which is not unduly detrimental to the viability of the genetically modified cells.

[00195] Cell compositions that are highly enriched for cells comprising modified DNA are achieved in this manner. By "highly enriched", it is meant that the genetically modified cells will be 70% or more, 75% or more, 80% or more, 85% or more, 90% or more of the cell composition, for example, about 95% or more, or 98% or more of the cell composition. In other words, the composition may be a substantially pure composition of genetically modified cells.

[00196] Genetically modified cells produced by the methods described herein may be used immediately. Alternatively, the cells may be frozen at liquid nitrogen temperatures and stored

for long periods of time, being thawed and capable of being reused. In such cases, the cells will usually be frozen in 10% dimethylsulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperatures, and thawed in a manner as commonly known in the art for thawing frozen cultured cells.

[00197] The genetically modified cells may be cultured in vitro under various culture conditions. The cells may be expanded in culture, i.e. grown under conditions that promote their proliferation. Culture medium may be liquid or semi-solid, e.g. containing agar, methylcellulose, etc. The cell population may be suspended in an appropriate nutrient medium, such as Iscove's modified DMEM or RPMI 1640, normally supplemented with fetal calf serum (about 5-10%), L-glutamine, a thiol, particularly 2-mercaptoethanol, and antibiotics, e.g. penicillin and streptomycin. The culture may contain growth factors to which the regulatory T cells are responsive. Growth factors, as defined herein, are molecules capable of promoting survival, growth and/or differentiation of cells, either in culture or in the intact tissue, through specific effects on a transmembrane receptor. Growth factors include polypeptides and non-polypeptide factors.

[00198] Cells that have been genetically modified in this way may be transplanted to a subject for purposes such as gene therapy, e.g. to treat a disease or as an antiviral, antipathogenic, or anticancer therapeutic, for the production of genetically modified organisms in agriculture, or for biological research. The subject may be a neonate, a juvenile, or an adult. Of particular interest are mammalian subjects. Mammalian species that may be treated with the present methods include canines and felines; equines; bovines; ovines; etc. and primates, particularly humans. Animal models, particularly small mammals, e.g. murine, lagomorpha, etc. may be used for experimental investigations.

[00199] Cells may be provided to the subject alone or with a suitable substrate or matrix, e.g. to support their growth and/or organization in the tissue to which they are being transplanted. Usually, at least 1x103 cells will be administered, for example 5x103 cells, 1x104 cells, 5x104 cells, 1x105 cells, 1 x 106 cells or more. The cells may be introduced to the subject via any of the following routes: parenteral, subcutaneous, intravenous, intracranial, intraspinal, intraocular, or into spinal fluid. The cells may be introduced by injection, catheter, or the like. Examples of methods for local delivery, that is, delivery to the site of injury, include, e.g. through an Ommaya reservoir, e.g. for intrathecal delivery (see e.g. US Patent Nos. 5,222,982 and 5385582, incorporated herein by reference); by bolus injection, e.g. by a syringe, e.g. into a joint; by continuous infusion, e.g. by cannulation, e.g. with convection (see e.g. US Application No. 20070254842, incorporated here by reference); or by implanting a device upon which the cells

have been reversably affixed (see e.g. US Application Nos. 20080081064 and 20090196903, incorporated herein by reference).

[00200] The number of administrations of treatment to a subject may vary. Introducing the genetically modified cells into the subject may be a one-time event; but in certain situations, such treatment may elicit improvement for a limited period of time and require an on-going series of repeated treatments. In other situations, multiple administrations of the genetically modified cells may be required before an effect is observed. The exact protocols depend upon the disease or condition, the stage of the disease and parameters of the individual subject being treated.

[00201] In other aspects of the invention, the DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide are employed to modify cellular DNA in vivo, again for purposes such as gene therapy, e.g. to treat a disease or as an antiviral, antipathogenic, or anticancer therapeutic, for the production of genetically modified organisms in agriculture, or for biological research. In these in vivo embodiments, a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide are administered directly to the individual. A DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may be administered by any of a number of well-known methods in the art for the administration of peptides, small molecules and nucleic acids to a subject. A DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide can be incorporated into a variety of formulations. More particularly, a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide of the present invention can be formulated into pharmaceutical compositions by combination with appropriate pharmaceutically acceptable carriers or diluents.

[00202] Pharmaceutical preparations are compositions that include one or more a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide present in a pharmaceutically acceptable vehicle. "Pharmaceutically acceptable vehicles" may be vehicles approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in mammals, such as humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is formulated for administration to a mammal. Such pharmaceutical vehicles can be lipids, e.g. liposomes, e.g. liposome dendrimers; liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like, saline; gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. Pharmaceutical compositions may be formulated into preparations in solid, semi-

solid, liquid or gaseous forms, such as tablets, capsules, powders, granules, ointments, solutions, suppositories, injections, inhalants, gels, microspheres, and aerosols. As such, administration of the a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide can be achieved in various ways, including oral, buccal, rectal, parenteral, intraperitoneal, intradermal, transdermal, intratracheal, etc., administration. The active agent may be systemic after administration or may be localized by the use of regional administration, intramural administration, or use of an implant that acts to retain the active dose at the site of implantation. The active agent may be formulated for immediate activity or it may be formulated for sustained release.

[00203] For some conditions, particularly central nervous system conditions, it may be necessary to formulate agents to cross the blood-brain barrier (BBB). One strategy for drug delivery through the blood-brain barrier (BBB) entails disruption of the BBB, either by osmotic means such as mannitol or leukotrienes, or biochemically by the use of vasoactive substances such as bradykinin. The potential for using BBB opening to target specific agents to brain tumors is also an option. A BBB disrupting agent can be co-administered with the therapeutic compositions of the invention when the compositions are administered by intravascular injection. Other strategies to go through the BBB may entail the use of endogenous transport systems, including Caveolin-1 mediated transcytosis, carrier-mediated transporters such as glucose and amino acid carriers, receptor-mediated transcytosis for insulin or transferrin, and active efflux transporters such as p-glycoprotein. Active transport moieties may also be conjugated to the therapeutic compounds for use in the invention to facilitate transport across the endothelial wall of the blood vessel. Alternatively, drug delivery of therapeutics agents behind the BBB may be by local delivery, for example by intrathecal delivery, e.g. through an Ommaya reservoir (see e.g. US Patent Nos. 5,222,982 and 5385582, incorporated herein by reference); by bolus injection, e.g. by a syringe, e.g. intravitreally or intracranially; by continuous infusion, e.g. by cannulation, e.g. with convection (see e.g. US Application No. 20070254842, incorporated here by reference); or by implanting a device upon which the agent has been reversably affixed (see e.g. US Application Nos. 20080081064 and 20090196903, incorporated herein by reference).

[00204] Typically, an effective amount of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide are provided. As discussed above with regard to ex vivo methods, an effective amount or effective dose of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide in vivo is the amount to induce a 2 fold increase or more in the amount of recombination observed between two homologous sequences relative to a negative control, e.g. a cell contacted with an empty vector or irrelevant polypeptide. The amount of recombination may be measured by any convenient method, e.g. as

described above and known in the art. The calculation of the effective amount or effective dose of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide to be administered is within the skill of one of ordinary skill in the art, and will be routine to those persons skilled in the art. Thefinal amount to be administered will be dependent upon the route of administration and upon the nature of the disorder or condition that is to be treated.

[00205] The effective amount given to a particular patient will depend on a variety of factors, several of which will differ from patient to patient. A competent clinician will be able to determine an effective amount of a therapeutic agent to administer to a patient to halt or reverse the progression the disease condition as required. Utilizing LD50 animal data, and other information available for the agent, a clinician can determine the maximum safe dose for an individual, depending on the route of administration. For instance, an intravenously administered dose may be more than an intrathecally administered dose, given the greater body of fluid into which the therapeutic composition is being administered. Similarly, compositions which are rapidly cleared from the body may be administered at higher doses, or in repeated doses, in order to maintain a therapeutic concentration. Utilizing ordinary skill, the competent clinician will be able to optimize the dosage of a particular therapeutic in the course of routine clinical trials.

[00206] For inclusion in a medicament, a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may be obtained from a suitable commercial source. As a general proposition, the total pharmaceutically effective amount of the a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide administered parenterally per dose will be in a range that can be measured by a dose response curve.

[00207] Therapies based on a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotides, i.e. preparations of a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide to be used for therapeutic administration, must be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 μm membranes). Therapeutic compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle. The therapies based on a DNA-targeting RNA and/or site-directed modifying polypeptide and/or donor polynucleotide may be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-mL vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous solution of compound, and the resulting

mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized compound using bacteriostatic Water-for-Injection.

[00208] Pharmaceutical compositions can include, depending on the formulation desired, pharmaceutically-acceptable, non-toxic carriers of diluents, which are defined as vehicles commonly used to formulate pharmaceutical compositions for animal or human administration. The diluent is selected so as not to affect the biological activity of the combination. Examples of such diluents are distilled water, buffered water, physiological saline, PBS, Ringer's solution, dextrose solution, and Hank's solution. In addition, the pharmaceutical composition or formulation can include other carriers, adjuvants, or non-toxic, nontherapeutic, nonimmunogenic stabilizers, excipients and the like. The compositions can also include additional substances to approximate physiological conditions, such as pH adjusting and buffering agents, toxicity adjusting agents, wetting agents and detergents.

[00209] The composition can also include any of a variety of stabilizing agents, such as an antioxidant for example. When the pharmaceutical composition includes a polypeptide, the polypeptide can be complexed with various well-known compounds that enhance the in vivo stability of the polypeptide, or otherwise enhance its pharmacological properties (e.g., increase the half-life of the polypeptide, reduce its toxicity, enhance solubility or uptake). Examples of such modifications or complexing agents include sulfate, gluconate, citrate and phosphate. The nucleic acids or polypeptides of a composition can also be complexed with molecules that enhance their in vivo attributes. Such molecules include, for example, carbohydrates, polyamines, amino acids, other peptides, ions (e.g., sodium, potassium, calcium, magnesium, manganese), and lipids.

[00210] Further guidance regarding formulations that are suitable for various types of administration can be found in Remington's Pharmaceutical Sciences, Mace Publishing Company, Philadelphia, Pa., 17th ed. (1985). For a brief review of methods for drug delivery, see, Langer, Science 249:1527-1533 (1990).

[00211] The pharmaceutical compositions can be administered for prophylactic and/or therapeutic treatments. Toxicity and therapeutic efficacy of the active ingredient can be determined according to standard pharmaceutical procedures in cell cultures and/or experimental animals, including, for example, determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Therapies that exhibit large therapeutic indices are preferred.

[00212] The data obtained from cell culture and/or animal studies can be used in formulating a range of dosages for humans. The dosage of the active ingredient typically lines within a range of circulating concentrations that include the ED50 with low toxicity. The dosage can vary within this range depending upon the dosage form employed and the route of administration utilized.

[00213] The components used to formulate the pharmaceutical compositions are preferably of high purity and are substantially free of potentially harmful contaminants (e.g., at least National Food (NF) grade, generally at least analytical grade, and more typically at least pharmaceutical grade). Moreover, compositions intended for in vivo use are usually sterile. To the extent that a given compound must be synthesized prior to use, the resulting product is typically substantially free of any potentially toxic agents, particularly any endotoxins, which may be present during the synthesis or purification process. Compositions for parental administration are also sterile, substantially isotonic and made under GMP conditions.

[00214] The effective amount of a therapeutic composition to be given to a particular patient will depend on a variety of factors, several of which will differ from patient to patient. A competent clinician will be able to determine an effective amount of a therapeutic agent to administer to a patient to halt or reverse the progression the disease condition as required. Utilizing LD50 animal data, and other information available for the agent, a clinician can determine the maximum safe dose for an individual, depending on the route of administration. For instance, an intravenously administered dose may be more than an intrathecally administered dose, given the greater body of fluid into which the therapeutic composition is being administered. Similarly, compositions which are rapidly cleared from the body may be administered at higher doses, or in repeated doses, in order to maintain a therapeutic concentration. Utilizing ordinary skill, the competent clinician will be able to optimize the dosage of a particular therapeutic in the course of routine clinical trials.

GENETICALLY MODIFIED HOST CELLS

[00215] The present disclosure provides genetically modified host cells, including isolated genetically modified host cells, where a subject genetically modified host cell comprises (has been genetically modified with): 1) an exogenous DNA-targeting RNA; 2) an exogenous nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; 3) an exogenous site-directed modifying polypeptide; 4) an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide; or 5) any combination of the above. A subject genetically modified cell is generated by genetically modifying a host cell with, for example: 1) an exogenous DNA-targeting RNA; 2) an exogenous nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; 3) an exogenous site-directed modifying polypeptide;

4) an exogenous nucleic acid comprising a nucleotide sequence encoding a site-directed modifying polypeptide; or 5) any combination of the above.).

[00216] In some embodiments, a subject genetically modified host cell is *in vitro*. In some embodiments, a subject genetically modified host cell is *in vivo*. In some embodiments, a subject genetically modified host cell is a prokaryotic cell or is derived from a prokaryotic cell. In some embodiments, a subject genetically modified host cell is a bacterial cell or is derived from a bacterial cell. In some embodiments, a subject genetically modified host cell is an archaeal cell or is derived from an arachael cell. In some embodiments, a subject genetically modified host cell is a eukaryotic cell or is derived from a eukaryotic cell. In some embodiments, a subject genetically modified host cell is a plant cell or is derived from a plant cell. In some embodiments, a subject genetically modified host cell is an animal cell or is derived from an animal cell. In some embodiments, a subject genetically modified host cell is an invertebrate cell or is derived from an invertebrate cell. In some embodiments, a subject genetically modified host cell is a vertebrate cell or is derived from a vertebrate cell. In some embodiments, a subject genetically modified host cell is a mammalian cell or is derived from a mammalian cell. In some embodiments, a subject genetically modified host cell is a rodent cell or is derived from a rodent cell. In some embodiments, a subject genetically modified host cell is a human cell or is derived from a human cell.

[00217] The present disclosure further provides progeny of a subject genetically modified cell, where the progeny can comprise the same exogenous nucleic acid or polypeptide as the subject genetically modified cell from which it was derived. The present disclosure further provides a composition comprising a subject genetically modified host cell.

Genetically modified stem cells and genetically modified progenitor cells

[00218] In some embodiments, a subject genetically modified host cell is a genetically modified stem cell or progenitor cell. Suitable host cells include, e.g., stem cells (adult stem cells, embryonic stem cells, iPS cells, etc.) and progenitor cells (e.g., cardiac progenitor cells, neural progenitor cells, etc.). Suitable host cells include mammalian stem cells and progenitor cells, including, e.g., rodent stem cells, rodent progenitor cells, human stem cells, human progenitor cells, etc. Suitable host cells include *in vitro* host cells, e.g., isolated host cells.

[00219] In some embodiments, a subject genetically modified host cell comprises an exogenous DNA-targeting RNA nucleic acid. In some embodiments, a subject genetically modified host cell comprises an exogenous nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA. In some embodiments, a subject genetically modified host cell comprises an exogenous site-directed modifying polypeptide. In some embodiments, a subject genetically modified host cell comprises an exogenous nucleic acid comprising a nucleotide sequence

encoding a site-directed modifying polypeptide. In some embodiments, a subject genetically modified host cell comprises exogenous nucleic acid comprising a nucleotide sequence encoding 1) a DNA-targeting RNA and 2) a site-directed modifying polypeptide.

COMPOSITIONS

[00220] The present invention provides a composition comprising a subject DNA-targeting RNA and/or a site-directed modifying polypeptide. In some cases, the site-directed modifying polypeptide is a subject chimeric polypeptide. A subject composition is useful for carrying out a method of the present disclosure, e.g., a method for site-specific modification of a target DNA; a method for site-specific modification of a polypeptide associated with a target DNA; etc.

Compositions comprising a DNA-targeting RNA

[00221] The present invention provides a composition comprising a subject DNA-targeting RNA. The composition can comprise, in addition to the DNA-targeting RNA, one or more of: a salt, e.g., NaCl, $MgCl_2$, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) (HEPES), 2-(N-Morpholino)ethanesulfonic acid (MES), MES sodium salt, 3-(N-Morpholino)propanesulfonic acid (MOPS), N-tris[Hydroxymethyl]methy1-3-aminopropanesulfonic acid (TAPS), etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a nuclease inhibitor; nd the like. For example, in some cases, a subject composition comprises a subject DNA-targeting RNA and a buffer for stabilizing nucleic acids.

[00222] In some embodiments, a DNA-targeting RNA present in a subject composition is pure, e.g., at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or more than 99% pure, where "% purity" means that DNA-targeting RNA is the recited percent free from other macromolecules, or contaminants that may be present during the production of the DNA-targeting RNA.

Compositions comprising a subject chimeric polypeptide

[00223] The present invention provides a composition a subject chimeric polypeptide. The composition can comprise, in addition to the DNA-targeting RNA, one or more of: a salt, e.g., NaCl, $MgCl_2$, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, HEPES, MES, MES sodium salt, MOPS, TAPS, etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a protease inhibitor; a reducing agent (e.g., dithiothreitol); and the like.

[00224] In some embodiments, a subject chimeric polypeptide present in a subject composition is pure, e.g., at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or more than 99% pure, where "% purity" means that the site-directed modifying polypeptide is the recited percent free from other proteins,

other macromolecules, or contaminants that may be present during the production of the chimeric polypeptide.

Compositions comprising a DNA-targeting RNA and a site-directed modifying polypeptide

[00225] The present invention provides a composition comprising: (i) a DNA-targeting RNA or a DNA polynucleotide encoding the same; and ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same. In some cases, the site-directed modifying polypeptide is a subject chimeric site-directed modifying polypeptide. In other cases, the site-directed modifying polypeptide is a naturally-occurring site-directed modifying polypeptide. In some instances, the site-directed modifying polypeptide exhibits enzymatic activity that modifies a target DNA. In other cases, the site-directed modifying polypeptide exhibits enzymatic activity that modifies a polypeptide that is associated with a target DNA. In still other cases, the site-directed modifying polypeptide modulates transcription of the target DNA.

[00226] The present invention provides a composition comprising: (i) a DNA-targeting RNA, as described above, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00227] In some instances, a subject composition comprises: a composition comprising: (i) a subject DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00228] In other embodiments, a subject composition comprises: (i) a polynucleotide encoding a subject DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a polynucleotide encoding the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting

RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

[00229] In some embodiments, a subject composition includes both RNA molecules of a double-molecule DNA-targeting RNA. As such, in some embodiments, a subject composition includes an activator-RNA that comprises a duplex-forming segment that is complementary to the duplex-forming segment of a targeter-RNA (see Figure 1A). The duplex-forming segments of the activator-RNA and the targeter-RNA hybridize to form the dsRNA duplex of the protein-binding segment of the DNA-targeting RNA. The targeter-RNA further provides the DNA-targeting segment (single stranded) of the DNA-targeting RNA and therefore targets the DNA-targeting RNA to a specific sequence within the target DNA. As one non-limiting example, the duplex-forming segment of the activator-RNA comprises a nucleotide sequence that has greater than about 70%, greater than about 80%, greater than about 90%, greater than about 95%, greater than about 98%, or 100% identity with the sequence 5'UAGCAAGUUAAAAU-3' (see Figure 3). As another non-limiting example, the duplex-forming segment of the targeter-RNA comprises a nucleotide sequence that has greater than about 70%, greater than about 80%, greater than about 90%, greater than about 95%, greater than about 98%, or 100% identity with the sequence 5'GUUUUAGAGCUA-3' (see Figure 3).

[00230] The present disclosure provides a composition comprising: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00231] For example, in some cases, a subject composition comprises: (i) a DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00232] As another example, in some cases, a subject composition compries: (i) a DNA polynucleotide encoding a DNA-targeting RNA, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a polynucleotide encoding the site-directed modifying polypeptide, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00233] A subject composition can comprise, in addition to i) a subject DNA-targeting RNA, or a DNA polynucleotide encoding the same; and ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, one or more of: a salt, e.g., NaCl, $MgCl_2$, KCl, $MgSO_4$, etc.; a buffering agent, e.g., a Tris buffer, HEPES, MES, MES sodium salt, MOPS, TAPS, etc.; a solubilizing agent; a detergent, e.g., a non-ionic detergent such as Tween-20, etc.; a protease inhibitor; a reducing agent (e.g., dithiothreitol); and the like.

[00234] In some cases, the components of the composition are individually pure, e.g., each of the components is at least about 75%, at least about 80%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or greater than 99%, pure. In some cases, the individual components of a subject composition are pure before being added to the composition.

[00235] For example, in some embodiments, a site-directed modifying polypeptide present in a subject composition is pure, e.g., at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 98%, at least about 99%, or more than 99% pure, where "% purity" means that the site-directed modifying polypeptide is the recited percent free from other proteins (e.g., proteins other than the site-directed modifying polypeptide), other macromolecules, or contaminants that may be present during the production of the site-directed modifying polypeptide.

KITS

[00236] The present disclosure provides kits for carrying out a subject method. A subject kit can include one or more of: a site-directed modifying polypeptide; a nucleic acid comprising a nucleotide encoding a site-directed modifying polypeptide; a DNA-targeting RNA; a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; an activator-RNA; a nucleic acid comprising a nucleotide sequence encoding an activator-RNA; a targeter-RNA; and a nucleic acid comprising a nucleotide sequence encoding a targeter-RNA. A site-directed modifying polypeptide; a nucleic acid comprising a nucleotide encoding a site-directed modifying polypeptide; a DNA-targeting RNA; a nucleic acid comprising a nucleotide sequence

encoding a DNA-targeting RNA; an activator-RNA; a nucleic acid comprising a nucleotide sequence encoding an activator-RNA; a targeter-RNA; and a nucleic acid comprising a nucleotide sequence encoding a targeter-RNA, are described in detail above. A kit may comprise a complex that comprises two or more of: a site-directed modifying polypeptide; a nucleic acid comprising a nucleotide encoding a site-directed modifying polypeptide; a DNA-targeting RNA; a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; an activator-RNA; a nucleic acid comprising a nucleotide sequence encoding an activator-RNA; a targeter-RNA; and a nucleic acid comprising a nucleotide sequence encoding a targeter-RNA.

[00237] In some embodiments, a subject kit comprises: a site-directed modifying polypeptide; and a reagent for reconstituting and/or diluting the site-directed modifying polypeptide. In other embodiments, a subject kit comprises: a nucleic acid comprising a nucleotide encoding a site-directed modifying polypeptide; and a reagent for reconstituting and/or diluting the site-directed modifying polypeptide.

[00238] A subject kit comprising a site-directed modifying polypeptide, or a nucleic acid comprising a nucleotide encoding a site-directed modifying polypeptide, can further include one or more additional reagents, where such additional reagents can be selected from: a buffer for introducing the site-directed modifying polypeptide into a cell; a wash buffer; a control reagent; a control expression vector or RNA polynucleotide; a reagent for *in vitro* production of the site-directed modifying polypeptide from DNA, and the like. In some cases, the site-directed modifying polypeptide included in a subject kit is a chimeric site-directed modifying polypeptide, as described above.

[00239] In some embodiments, a subject kit comprises: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In some cases, the kit comprises: (i) a DNA-targeting RNA; and a site-directed modifying polypeptide. In other cases, the kit comprises: (i) a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; and (ii) a nucleic acid comprising a nucleotide sequence encoding site-directed modifying polypeptide.

[00240] As another example, a subject kit can include: (i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, comprising: (a) a first segment comprising a nucleotide

sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, comprising: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA In some cases, the kit comprises: (i) a DNA-targeting RNA; and a site-directed modifying polypeptide. In other cases, the kit comprises: (i) a nucleic acid comprising a nucleotide sequence encoding a DNA-targeting RNA; and (ii) a nucleic acid comprising a nucleotide sequence encoding site-directed modifying polypeptide.

[00241] The present disclosure provides a kit comprising: (1) a recombinant expression vector comprising (i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a nucleotide sequence encoding the site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.; and (2) a reagent for reconstitution and/or dilution of the expression vector.

[00242] The present disclosure provides a kit comprising: (1) a recombinant expression vector comprising: (i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises: (a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (b) a second segment that interacts with a site-directed modifying polypeptide; and (ii) a nucleotide sequence encoding the site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA; and (2) a reagent for reconstitution and/or dilution of the recombinant expression vector.

[00243] The present disclosure provides a kit comprising: (1) a recombinant expression vector comprising a nucleic acid comprising a nucleotide sequence that encodes a DNA targeting RNA comprising: (i) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and (ii) a second segment that interacts with a site-directed modifying polypeptide; and (2) a reagent for reconstitution and/or dilution of the recombinant expression vector. In some embodiments of this kit, the kit comprises: a recombinant expression vector

comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA. In other embodiments of this kit, the kit comprises: a recombinant expression vector comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises: (a) an RNA-binding portion that interacts with the DNA-targeting RNA; and (b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

[00244] In some embodiments of any of the above kits, the kit comprises an activator-RNA or a targeter-RNA. In some embodiments of any of the above kits, the kit further comprises a donor polynucleotide to effect the desired genetic modification. Components of a subject kit can be in separate containers; or can be combined in a single container.

[00245] Any of the above-described kits can further include one or more additional reagents, where such additional reagents can be selected from: a dilution buffer; a reconstitution solution; a wash buffer; a control reagent; a control expression vector or RNA polynucleotide; a reagent for *in vitro* production of the site-directed modifying polypeptide from DNA, and the like.

[00246] In addition to above-mentioned components, a subject kit can further include instructions for using the components of the kit to practice the subject methods. The instructions for practicing the subject methods are generally recorded on a suitable recording medium. For example, the instructions may be printed on a substrate, such as paper or plastic, etc. As such, the instructions may be present in the kits as a package insert, in the labeling of the container of the kit or components thereof (i.e., associated with the packaging or subpackaging) etc. In other embodiments, the instructions are present as an electronic storage data file present on a suitable computer readable storage medium, e.g. CD-ROM, diskette, flash drive, etc. In yet other embodiments, the actual instructions are not present in the kit, but means for obtaining the instructions from a remote source, e.g. via the internet, are provided. An example of this embodiment is a kit that includes a web address where the instructions can be viewed and/or from which the instructions can be downloaded. As with the instructions, this means for obtaining the instructions is recorded on a suitable substrate.

EXAMPLES

[00247] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the present invention, and are not

intended to limit the scope of what the inventors regard as their invention nor are they intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g. amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Celsius, and pressure is at or near atmospheric. Standard abbreviations may be used, e.g., bp, base pair(s); kb, kilobase(s); pl, picoliter(s); s or sec, second(s); min, minute(s); h or hr, hour(s); aa, amino acid(s); kb, kilobase(s); bp, base pair(s); nt, nucleotide(s); i.m., intramuscular(ly); i.p., intraperitoneal(ly); s.c., subcutaneous(ly); and the like.

Example 1:

[00248] A recombinant DNA-targeting polypeptide based on the sequence of *Streptococcus pyogenes* Cas9/Csn1 endonuclease was heterologously expressed in *Escherichia coli* and purified by a combination of affinity, ion exchange and gel filtration chromatographic steps according to standard procedures in the art. The protein was eluted and stored in 20 mM HEPES 7.5 150 mM potassium chloride and 1 mM TCEP. Targeting RNAs were obtained either by chemical synthesis (for the targeter and activator segments of the double molecule DNA-targeting RNA) or by in vitro transcription using T7 RNA polymerase (single-molecule DNA-targeting RNAs) according to standard protocols. Target DNAs were obtained by chemical synthesis of the individual single strands. One of the strands (the strand complementary to the targeting sequence in the DNA-targeting RNA) was subsequently [^{32}P]-labeled at its 5' end using T4 polynucleotide kinase and [gamma-^{32}P]-ATP as per standard procedures. The two strands were subsequently annealed by mixing equimolar amounts of the two DNA strands, heating to 95°C for 3 min and slow-cooling to room temperature.

[00249] DNA cleavage was performed by incubating the above DNA-targeting polypeptide (500 nM final concentration) together with the DNA-targeting RNA (500 nM concentration) and the target DNA (10 nM) in a total volume of 10 microliter, in a cleavage buffer containing 20 mM HEPES pH 7.5, 100 mM potassium chloride, 5 mM magnesium chloride, 1mM dithiothreitol and 5% (v/v) glycerol. For reactions guided by double-molecule DNA-targeting RNA, the targeter and activator RNAs were mixed in equimolar amounts and annealed by heating to 95oC for 1 min and slow-cooling to room temperature, prior to addition to the DNA-targeting polypeptide. The DNA-targeting RNA/polypeptide complexes were assembled by incubation in the cleavage buffer for 15 min at room temperature. Subsequently, the assembled complex was added to target DNA and incubated for 1 hr at 37 °C. Reactions were quenched by the addition of 20

microliter of formamide quench buffer (5% glycerol, 1 mM EDTA, 0.025% sodium dodecyl sulfate in formamide) and subsequently resolved on a 12% polyacrylamide, 7M urea denaturing gel. Clevage products were visualized by phosphorimaging according to standard procedures.

[00250] The results are depicted in Figures 3 and 5.

[00251] Figures 3A-Cdepict target DNA cleavage by a site-directed modifying polypeptide (exemplified by the Cas9/Csn1 protein of *Streptococcus pyogenes*), which is directed by a DNA-targeting RNA. (**A**) Radiolabeled target DNAs were incubated in the presence of Cas9/Csn1 and various DNA-targeting RNA species (as indicated). Cleavage products were resolved using denaturing polyacrylamide gel electrophoresis and visualized by phosphorimaging. (**B**) Schematic diagrams of the DNA-targeting RNAs used in conjunction with the Cas9/Csn1 site-directed modifying polypeptide. Note that while one of the tested single-molecule DNA-targeting RNAs (RNA chimera A) supported efficient target DNA cleavage, the other tested single-molecule DNA-targeting RNA (RNA chimera B) did not. (**C**) Schematic representations of DNA-targeting RNA sequences and DNA targets.

[00252] **Figures 5A and 5B depict target DNA cleavage. Figure 5 A.** Target DNA cleavage using a Cas9/Csn1 site-directed modifying polypeptide from a variety of different species (refer to sequences in Figure 12) and a DNA-targeting RNA. This experiment demonstrates that the various Cas9/Csn1 site-directed modifying polypeptides can utilize the same DNA-targeting RNA. **B.** A schematic representation of the DNA-targeting RNA used in Figure 5A.

[00253] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

CLAIMS

What is claimed is:

1. A DNA-targeting RNA comprising:

(i) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and

(ii) a second segment that interacts with a site-directed modifying polypeptide.

2. The DNA-targeting RNA of Claim 1, wherein the first segment comprises 20 nucleotides that have at least 65% complementarity to a sequence in the target DNA.

3. The DNA-targeting RNA of Claim 1, wherein the second segment comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 12 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 7A-C.

4. The DNA-targeting RNA of Claim 1, wherein the second segment comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C.

5. The DNA-targeting RNA of Claim 1, wherein the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 2, or to the corresponding domains in any of the amino acid sequences depicted in Figure 12.

6. A DNA polynucleotide comprising a nucleotide sequence that encodes the DNA-targeting RNA of Claim 1.

7. A recombinant expression vector comprising the DNA polynucleotide of Claim 6.

8. The recombinant expression vector of claim 7, wherein the nucleotide sequence encoding the DNA-targeting RNA is operably linked to a promoter.

9. The recombinant expression vector of claim 8, wherein the promoter is an inducible promoter.

10. The recombinant expression vector of claim 7, wherein the nucleotide sequence encoding the DNA-targeting RNA of Claim 1 further comprises a multiple cloning site.

11. An *in vitro* genetically modified host cell comprising the DNA polynucleotide of Claim 1.

12. A recombinant expression vector comprising:

(i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises:

(a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and

(b) a second segment that interacts with a site-directed modifying polypeptide; and

(ii) a nucleotide sequence encoding the site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

13. A recombinant expression vector comprising:

(i) a nucleotide sequence encoding a DNA-targeting RNA, wherein the DNA-targeting RNA comprises:

(a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and

(b) a second segment that interacts with a site-directed modifying polypeptide; and

(ii) a nucleotide sequence encoding the site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

14. A chimeric site-directed modifying polypeptide comprising:

(i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and

(ii) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

15. The chimeric site-directed modifying polypeptide of Claim 14 comprising an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 2, or to the corresponding domains in any of the amino acid sequences depicted in Figure 12.

16. The chimeric site-directed modifying polypeptide of Claim 14, wherein the DNA-targeting RNA further comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 12 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 7A-C.

17. The chimeric site-directed modifying polypeptide of Claim 14, wherein the DNA-targeting RNA further comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C.

18. The chimeric site-directed modifying polypeptide of Claim 14, wherein the enzymatic activity modifies the target DNA.

19. The chimeric site-directed modifying polypeptide of Claim 18, wherein the enzymatic activity is nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity.

20. The chimeric site-directed modifying polypeptide of Claim 19, wherein the enzymatic activity is nuclease activity.

21. The chimeric site-directed modifying polypeptide of Claim 20, wherein the nuclease activity introduces a double strand break in the target DNA.

22. The chimeric site-directed modifying polypeptide of Claim 14, wherein the enzymatic activity modifies a target polypeptide associated with the target DNA.

23. The chimeric site-directed modifying polypeptide of Claim 22, wherein the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase

activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity.

24. The chimeric site-directed modifying polypeptide of Claim 22, wherein the target polypeptide is a histone and the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity or deubiquitinating activity.

25. An RNA polynucleotide comprising a nucleotide sequence encoding the chimeric site-directed modifying polypeptide of Claim 14.

26. A DNA polynucleotide comprising a nucleotide sequence encoding the chimeric site-directed modifying polypeptide of Claim 14.

27. A recombinant expression vector comprising the polynucleotide of Claim 26.

28. The recombinant expression vector of Claim 27, wherein the polynucleotide is operably linked to a promoter.

29. The recombinant expression vector of Claim 28, wherein the promoter is an inducible promoter.

30. An *in vitro* genetically modified host cell comprising the polynucleotide of Claim 26.

31. A chimeric site-directed modifying polypeptide comprising:

(i) an RNA-binding portion that interacts with a DNA-targeting RNA, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in a target DNA; and

(ii) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

32. The chimeric site-directed modifying polypeptide of Claim 31, wherein the activity portion increases transcription within the target DNA.

33. The chimeric site-directed modifying polypeptide of Claim 31, wherein the activity portion decreases transcription within the target DNA.

34. A composition comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising:

(a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and

(b) a second segment that interacts with a site-directed modifying polypeptide; and

(ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

35. The composition of Claim 34, wherein the first segment of the DNA-targeting RNA comprises 20 nucleotides that have at least 90% complementarity to a sequence in the target DNA.

36. The composition of Claim 34, wherein the second segment of the DNA-targeting RNA comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 12 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 7A-C.

37. The composition of Claim 34, wherein the second segment of the DNA-targeting RNA comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C.

38. The composition of Claim 34, wherein the site-directed modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 2, or to the corresponding domains in any of the amino acid sequences depicted in Figure 12.

39. The composition of Claim 34, wherein the enzymatic activity modifies the target DNA.

40. The composition of Claim 39, wherein the enzymatic activity is nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity,

deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity.

41. The composition of Claim 40, wherein the enzymatic activity is nuclease activity.

42. The composition of Claim 41, wherein the nuclease activity introduces a double strand break in the target DNA.

43. The composition of Claim 34, wherein the enzymatic activity modifies a target polypeptide associated with the target DNA.

44. The composition of Claim 43, wherein the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity.

45. The composition of Claim 43, wherein the target polypeptide is a histone and the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity or deubiquitinating activity.

46. The composition of Claim 34, wherein the DNA-targeting RNA is a double-molecule DNA-targeting RNA and the composition comprises both a targeter-RNA and an activator-RNA, the duplex-forming segments of which are complementary and hybridize to form the second segment of the DNA-targeting RNA.

47. The composition of Claim 46, wherein the duplex-forming segment of the activator-RNA comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C.

48. A composition comprising:
(i) the DNA-targeting RNA of Claim 34, or a DNA polynucleotide encoding the same; and
(ii) a buffer for stabilizing nucleic acids.

49. A composition comprising:

(i) the site-directed modifying polypeptide of Claim 34, or a polynucleotide encoding the same; and

(ii) a buffer for stabilizing nucleic acids and/or proteins.

50. A composition comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising:

(a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and

(b) a second segment that interacts with a site-directed modifying polypeptide; and

(ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

51. The composition of Claim 50, wherein the activity portion increases transcription within the target DNA.

52. The composition of Claim 50, wherein the activity portion decreases transcription within the target DNA.

53. A composition comprising:

(i) the site-directed modifying polypeptide of Claim 50, or a polynucleotide encoding the same; and

(ii) a buffer for stabilizing nucleic acids and/or proteins.

54. A method of site-specific modification at target DNA, the method comprising:

contacting the target DNA with a complex comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in the target DNA; and

(ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide exhibits site-directed enzymatic activity.

55. The method of Claim 54, wherein the target DNA is extrachromosomal.

56. The method of Claim 54, wherein the target DNA is part of a chromosome *in vitro.*

57. The method of Claim 54, wherein the target DNA is part of a chromosome *in vivo.*

58. The method of Claim 54, wherein the target DNA is part of a chromosome in a cell.

59. The method of Claim 58, wherein the cell is an archaeal cell.

60. The method of Claim 58, wherein the cell is a bacterial cell.

61. The method of Claim 58, wherein the cell is a eukaryotic single-cell organism.

62. The method of Claim 58, wherein the cell is a somatic cell.

63. The method of Claim 58, wherein the cell is a germ cell.

64. The method of Claim 58, wherein the cell is a stem cell.

65. The method of Claim 58, wherein the cell is a plant cell.

66. The method of Claim 58, wherein the cell is an animal cell.

67. The method of Claim 66, wherein the cell is a vertebrate cell.

68. The method of Claim 67, wherein the cell is a mammalian cell.

69. The method of Claim 68, wherein the cell is a human cell.

70. The method of Claim 54, wherein the DNA-targeting RNA comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 12 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 7A-C.

71. The method of Claim 54, wherein the DNA-targeting RNA comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C.

72. The method of Claim 54, wherein the DNA-modifying polypeptide comprises an amino acid sequence having at least about 75% amino acid sequence identity to amino acids 7-166 or 731-1003 of the Cas9/Csn1 amino acid sequence depicted in Figure 2, or to the corresponding domains in any of the amino acid sequences depicted in Figure 12.

73. The method of Claim 54, wherein the enzymatic activity modifies the target DNA.

74. The method of Claim 73, wherein the enzymatic activity is nuclease activity, methyltransferase activity, demethylase activity, DNA repair activity, DNA damage activity, deamination activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity.

75. The method of Claim 74, wherein the DNA-modifying enzymatic activity is nuclease activity.

76. The method of Claim 75, wherein the nuclease activity introduces a double strand break in the target DNA.

77. The method of Claim 76, wherein the contacting occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair.

78. The method of Claim 77, the method further comprising contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide or copies of the donor polynucleotide integrate into the target DNA.

79. The method of Claim 77, wherein the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that nucleotides within the target DNA are deleted.

80. The method of Claim 54, wherein the enzymatic activity modifies a target polypeptide associated with the target DNA.

81. The method of Claim 80, wherein the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity or demyristoylation activity.

82. The method of Claim 80, wherein the target polypeptide is a histone and the enzymatic activity is methyltransferase activity, demethylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity or deubiquitinating activity.

83. The method of Claim 54, wherein the complex further comprises an activator-RNA.

84. The method of Claim 83, wherein the activator-RNA comprises a nucleotide sequence with greater than 75% identity over a stretch of at least 25 contiguous nucleotides to any one of the nucleotide sequences set forth in Figures 6A-C.

85. A method of site-specific modification at target DNA, the method comprising:

contacting the target DNA with a complex comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in the target DNA; and

(ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide modulates transcription within the target DNA.

86. The method of Claim 85, wherein the site-directed modifying polypeptide increases transcription within the target DNA.

87. The method of Claim 85, wherein the site-directed modifying polypeptide decreases transcription within the target DNA.

88. A method of promoting site-specific cleavage and modification of a target DNA in a cell, the method comprising:

introducing into the cell a complex comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in the target DNA; and

(ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein:

(a) the site-directed modifying polypeptide exhibits nuclease activity that creates a double strand break in the target DNA;

(b) the site of the double strand break is determined by the DNA-targeting RNA;

(c) the contacting occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair; and

(d) the target DNA is cleaved and rejoined to produce a modified DNA sequence.

89. The method of Claim 88, the method further comprising contacting the target DNA with a donor polynucleotide, wherein the donor polynucleotide or copies of the donor polynucleotide integrate into the target DNA.

90. The method of Claim 88, wherein the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that nucleotides within the target DNA are deleted.

91. The method of Claim 88, wherein the cell is an archaeal cell.

92. The method of Claim 88, wherein the cell is a bacterial cell.

93. The method of Claim 88, wherein the cell is a eukaryotic single-cell organism.

94. The method of Claim 88, wherein the cell is a somatic cell.

95. The method of Claim 88, wherein the cell is a germ cell.

96. The method of Claim 88, wherein the cell is a stem cell.

97. The method of Claim 88, wherein the cell is *in vitro*.

98. The method of Claim 88, wherein the cell is *in vivo*.

99. A method of producing a genetically modified cell in a subject, the cell comprising a site-specific modification in a chromosomal target DNA, the method comprising:

(I) introducing into the cell a complex comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in the target DNA; and

(ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein:

(a) the site-directed modifying polypeptide exhibits nuclease activity that creates a double strand break in the target DNA;

(b) the site of the double strand break is determined by the DNA-targeting RNA;

(c) the contacting occurs under conditions that are permissive for nonhomologous end joining or homology-directed repair; and

(d) the target DNA is cleaved and rejoined to produce a modified DNA sequence; and

(II) transplanting the genetically modified cell into the subject.

100. The method of Claim 99, the method further comprising contacting the cell with a donor polynucleotide, wherein the donor polynucleotide or copies of the donor polynucleotide integrate into the target DNA.

101. The method of Claim 99, wherein the method does not comprise contacting the cell with a donor polynucleotide, wherein the target DNA is modified such that nucleotides within the target DNA are deleted.

102. The method of Claim 99, wherein the cell is a somatic cell.

103. The method of Claim 99, wherein the cell is a germ cell.

104. The method of Claim 99, wherein the cell is a stem cell.

105. The method of Claim 99, wherein the subject is a plant.

106. The method of Claim 99, wherein the subject is an animal.

107. The method of Claim 106, wherein the subject is a vertebrate.

108. The method of Claim 107, wherein the subject is mammalian.

109. The method of Claim 108, wherein the subject is human.

110. A method of modulating site-specific transcription within a target DNA, the method comprising:

contacting the target DNA with a complex comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, wherein the DNA-targeting RNA comprises a nucleotide sequence that is complementary to a sequence in the target DNA; and

(ii) a site-directed modifying polypeptide, or a polynucleotide encoding the same, wherein the site-directed modifying polypeptide modulates transcription within the target DNA,

wherein said contacting results in modulating transcription within the target DNA.

111. The method of Claim 110, wherein transcription within the target DNA is increased.

112. The method of Claim 110, wherein transcription within the target DNA is decreased.

113. A kit comprising:

(i) the DNA-targeting RNA of Claim 1, or a DNA polynucleotide encoding the same; and

(ii)a reagent for reconstitution and/or dilution

114. The kit of method 113, the kit further comprising a reagent selected from the group consisting of: a buffer for introducing into cells the DNA-targeting RNA, a wash buffer, a control reagent, a control expression vector or RNA polynucleotide, a reagent for transcribing the DNA-targeting RNA from DNA and combinations thereof.

115. A kit comprising:

(i) the site-directed modifying polypeptide of Claim 34, or a polynucleotide encoding the same; and

(ii)a reagent for reconstitution and/or dilution

116. The kit of Claim 115, the kit further comprising a reagent selected from the group consisting of: a buffer for introducing into cells the site-directed modifying polypeptide, a wash buffer, a control reagent, a control expression vector or RNA polynucleotide, a reagent for *in vitro* production of the site-directed modifying polypeptide from DNA, and combinations thereof.

117. A kit comprising:

(i) the site-directed modifying polypeptide of Claim 50, or a polynucleotide encoding the same; and

(ii)a reagent for reconstitution and/or dilution

118. A kit comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, the DNA-targeting RNA comprising:

(a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and

(b) a second segment that interacts with a site-directed modifying polypeptide; and

(ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, the site-directed modifying polypeptide comprising:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

119. A kit comprising:

(i) a DNA-targeting RNA, or a DNA polynucleotide encoding the same, comprising:

(a) a first segment comprising a nucleotide sequence that is complementary to a sequence in a target DNA; and

(b) a second segment that interacts with a site-directed modifying polypeptide; and

(ii) the site-directed modifying polypeptide, or a polynucleotide encoding the same, comprising:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

120. A kit comprising:

(i) the recombinant expression vector of Claim 12; and

(ii) a reagent for reconstitution and/or dilution

121. A kit comprising:

(i) the recombinant expression vector of Claim 13; and

(ii) a reagent for reconstitution and/or dilution

122. A kit comprising:

(i) the recombinant expression vector of Claim 7; and

(ii) a reagent for reconstitution and/or dilution

123. A kit comprising:

(i) the recombinant expression vector of Claim 7; and

(ii) a recombinant expression vector comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that exhibits site-directed enzymatic activity, wherein the site of enzymatic activity is determined by the DNA-targeting RNA.

124. A kit comprising:

(i) the recombinant expression vector of Claim 7; and

(ii) a recombinant expression vector comprising a nucleotide sequence that encodes a site-directed modifying polypeptide, wherein the site-directed modifying polypeptide comprises:

(a) an RNA-binding portion that interacts with the DNA-targeting RNA; and

(b) an activity portion that modulates transcription within the target DNA, wherein the site of modulated transcription within the target DNA is determined by the DNA-targeting RNA.

METHODS AND COMPOSITIONS FOR RNA-DIRECTED SITE-SPECIFIC DNA MODIFICATION

ABSTRACT OF THE DISCLOSURE

The present disclosure provides a DNA-targeting RNA that comprises a targeting sequence and, together with a modifying polypeptide, provides for site-specific modification of a target DNA and/or a polypeptide associated with the target DNA. The present disclosure further provides modifying polypeptides. The present disclosure further provides methods of site-specific modification of a target DNA and/or a polypeptide associated with the target DNA; and kits and compositions for carrying out the methods.

Doc Code: **TR.PROV**
Document Description: Provisional Cover Sheet (SB16)

PTO/SB/16 (11-08)
Approved for use through 09/30/2010 OMB 0651-0032
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

Provisional Application for Patent Cover Sheet

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c)

Inventor(s)

Inventor 1 — Remove

Given Name	Middle Name	Family Name	City	State	Country
Martin		Jinek	Berkeley	CA	US

Inventor 2 — Remove

Given Name	Middle Name	Family Name	City	State	Country
Jennifer	A.	Doudna	Berkeley	CA	US

Inventor 3 — Remove

Given Name	Middle Name	Family Name	City	State	Country
Emmanuelle		Charpentier	Berkeley	CA	US

Inventor 4 — Remove

Given Name	Middle Name	Family Name	City	State	Country
Krzysztof		Chylinski	Berkeley	CA	US

All Inventors Must Be Listed – Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. — Add

Title of Invention	METHODS AND COMPOSITIONS FOR RNA-DIRECTED SITE-SPECIFIC DNA MODIFICATION
Attorney Docket Number (if applicable)	BERK-187PRV

Correspondence Address

Direct all correspondence to (select one):

◉ The address corresponding to Customer Number	○ Firm or Individual Name
Customer Number	84220

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

◉ No.

○ Yes, the name of the U.S. Government agency and the Government contract number are:

Doc Code: **TR.PROV**
Document Description: Provisional Cover Sheet (SB16)

PTO/SB/16 (11-08)
Approved for use through 09/30/2010 OMB 0651-0032
U.S. Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number

Entity Status

Applicant claims small entity status under 37 CFR 1.27

- (●) Yes, applicant qualifies for small entity status under 37 CFR 1.27
- ( ) No

Warning

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

Signature

Please see 37 CFR 1.4(d) for the form of the signature.

Signature	/Paula A. Borden, Ph.D., Reg. No. 42,344/			Date (YYYY-MM-DD)	2012-05-25
First Name	Paula	Last Name	Borden	Registration Number (If appropriate)	42344

This collection of information is required by 37 CFR 1.51. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.11 and 1.14. This collection is estimated to take 8 hours to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **This form can only be used when in conjunction with EFS-Web. If this form is mailed to the USPTO, it may cause delays in handling the provisional application.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or paten. Accordingly, pursuant to the requirements of the Act, please be advised that : (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1. The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether disclosure of these records is required by the Freedom of Information Act.
2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.
3. A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.
4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).
5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent Cooperation Treaty.
6. A record in this system of records may be disclosed, as a routine use, t o a n other federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).
7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspection or an issued patent.
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.