FIG. 1

>CasY1
MRKKLFKGYILHNKRLVYTGKAAIRSIKYPLVAPNKTALNNLSEKIIYDYEHLFGPLNVA
SYARNSNRYSLVDFWIDSLRAGVIWQSKSTSLIDLISKLEGSKSPSEKIFEQIDFELKNK
LDKEQFKDIILLNTGIRSSSNVRSLRGRFLKCFKEEFRDTEEVIACVDKWSKDLIVEGKS
ILVSKQFLYWEEEFGIKIFPHFKDNHDLPKLTFFVEPSLEFSPHLPLANCLERLKKFDIS
RESLLGLDNNFSAFSNYFNELFNLLSRGEIKKIVTAVLAVSKSWENEPELEKRLHFLSEK
AKLLGYPKLTSSWADYRMIIGGKIKSWHSNYTEQLIKVREDLKKHQIALDKLQEDLKKVV
DSSLREQIEAQREALLPLLDTMLKEKDFSDDLELYRFILSDFKSLLNGSYQRYIQTEEER
KEDRDVTKKYKDLYSNLRNIPRFFGESKKEQFNKFINKSLPTIDVGLKILEDIRNALETV
SVRKPPSITEEYVTKQLEKLSRKYKINAFNSNRFKQITEQVLRKYNNGELPKISEVFYRY
PRESHVAIRILPVKISNPRKDISYLLDKYQISPDWKNSNPGEVVDLIEIYKLTLGWLLSC
NKDFSMDFSSYDLKLFPEAASLIKNFGSCLSGYYLSKMIFNCITSEIKGMITLYTRDKFV
VRYVTQMIGSNQKFPLLCLVGEKQTKNFSRNWGVLIEEKGDLGEEKNQEKCLIFKDKTDF
AKAKEVEIFKNNIWRIRTSKYQIQFLNRLFKKTKEWDLMNLVLSEPSLVLEEEWGVSWDK
DKLLPLLKKEKSCEERLYYSLPLNLVPATDYKEQSAEIEQRNTYLGLDVGEFGVAYAVVR
IVRDRIELLSWGFLKDPALRKIRERVQDMKKKQVMAVFSSSSTAVARVREMAIHSLRNQI
HSIALAYKAKIIYEISISNFETGGNRMAKIYRSIKVSDVYRESGADTLVSEMIWGKKNKQ
MGNHISSYATSYTCCNCARTPFELVIDNDKEYEKGGDEFIFNVGDEKKVRGFLQKSLLGK
TIKGKEVLKSIKEYARPPIREVLLEGEDVEQLLKRRGNSYIYRCPFCGYKTDADIQAALN
IACRGYISDNAKDAVKEGERKLDYILEVRKLWEKNGAVLRSAKFL (SEQ ID NO: 1)

>CasY2
MQKVRKTLSEVHKNPYGTKVRNAKTGYSLQIERLSYTGKEGMRSFKIPLENKNKEVFDEF
VKKIRNDYISQVGLLNLSDWYEHYQEKQEHYSLADFWLDSLRAGVIFAHKETEIKNLISK
IRGDKSIVDKFNASIKKKHADLYALVDIKALYDFLTSDARRGLKTEEEFFNSKRNTLFPK
FRKKDNKAVDLWVKKFIGLDNKDKLNFTKKFIGFDPNPQIKYDHTFFFHQDINFDLERIT
TPKELISTYKKFLGKNKDLYGSDETTEDQLKMVLGFHNNHGAFSKYFNASLEAFRGRDNS
LVEQIINNSPYWNSHRKELEKRIIFLQVQSKKIKETELGKPHEYLASFGGKFESWVSNYL
RQEEEVKRQLFGYEENKKGQKKFIVGNKQELDKIIRGTDEYEIKAISKETIGLTQKCLKL
LEQLKDSVDDYTLSLYRQLIVELRIRLNVEFQETYPELIGKSEKDKEKDAKNKRADKRYP
QIFKDIKLIPNFLGETKQMVYKKFIRSADILYEGINFIDQIDKQITQNLLPCFKNDKERI
EFTEKQFETLRRKYYLMNSSRFHHVIEGIINNRKLIEMKKRENSELKTFSDSKFVLSKLF
LKKGKKYENEVYYTFYINPKARDQRRIKIVLDINGNNSVGILQDLVQKLKPKWDDIIKKN
DMGELIDAIEIEKVRLGILIALYCEHKFKIKKELLSLDLFASAYQYLELEDDPEELSGTN
LGRFLQSLVCSEIKGAINKISRTEYIERYTVQPMNTEKNYPLLINKEGKATWHIAAKDDL
SKKKGGGTVAMNQKIGKNFFGKQDYKTVFMLQDKRFDLLTSKYHLQFLSKTLDTGGGSWW
KNKNIDLNLSSYSFIFEQKVKVEWDLTNLDHPIKIKPSENSDDRRLFVSIPFVIKPKQTK
RKDLQTRVNYMGIDIGEYGLAWTIINIDLKNKKINKISKQGFIYEPLTHKVRDYVATIKD
NQVRGTFGMPDTKLARLRENAITSLRNQVHDIAMRYDAKPVYEFEISNFETGSNKVKVIY
DSVKRADIGRGQNNTEADNTEVNLVWGKTSKQFGSQIGAYATSYICSFCGYSPYYEFENS
KSGDEEGARDNLYQMKKLSRPSLEDFLQGNPVYKTFRDFDKYKNDQRLQKTGDKDGEWKT
HRGNTAIYACQKCRHISDADIQASYWIALKQVVRDFYKDKEMDGDLIQGDNKDKRKVNEL
NRLIGVHKDVPIINKNLITSLDINLL (SEQ ID NO: 2)

FIG. 1 (Cont.)

>CasY3
MKAKKSFYNQKRKFGKRGYRLHDERIAYSGGIGSMRSIKYELKDSYGIAGLRNR
IADATISDNKWLYGNINLNDYLEWRSSKTDKQIEDGDRESSLLGFWLEALRLGFVFSKQS
HAPNDFNETALQDLFETLDDDLKHVLDRKKWCDFIKIGTPKTNDQGRLKKQIKNLLKGNK
REEIEKTLNESDDELKEKINRIADVFAKNKSDKYTIFKLDKPNTEKYPRINDVQVAFFCH
PDFEEITERDRTKTLDLIINRFNKRYEITENKKDDKTSNRMALYSLNQGYIPRVLNDLFL
FVKDNEDDFSQFLSDLENFFSFSNEQIKIIKERLKKLKKYAEPIPGKPQLADKWDDYASD
FGGKLESWYSNRIEKLKKIPESVSDLRNNLEKIRNVLKKQNNASKILELSQKIIEYIRDY
GVSFEKPEIIKFSWINKTKDGQKKVFYVAKMADREFIEKLDLWMADLRSQLNEYNQDNKV
SFKKKGKKIEELGVLDFALNKAKKNKSTKNENGWQQKLSESIQSAPLFFGEGNRVRNEEV
YNLKDLLFSEIKNVENILMSSEAEDLKNIKIEYKEDGAKKGNYVLNVLARFYARFNEDGY
GGWNKVKTVLENIAREAGTDFSKYGNNNNRNAGRFYLNGRERQVFTLIKFEKSITVEKIL
ELVKLPSLLDEAYRDLVNENKNHKLRDVIQLSKTIMALVLSHSDKEKQIGGNYIHSKLSG
YNALISKRDFISRYSVQTTNGTQCKLAIGKGKSKKGNEIDRYFYAFQFFKNDDSKINLKV
IKNNSHKNIDFNDNENKINALQVYSSNYQIQFLDWFFEKHQGKKTSLEVGGSFTIAEKSL
TIDWSGSNPRVGFKRSDTEEKRVFVSQPFTLIPDDEDKERRKERMIKTKNRFIGIDIGEY
GLAWSLIEVDNGDKNNRGIRQLESGFITDNQQQVLKKNVKSWRQNQIRQTFTSPDTKIAR
LRESLIGSYKNQLESLMVAKKANLSFEYEVSGFEVGGKRVAKIYDSIKRGSVRKKDNNSQ
NDQSWGKKGINEWSFETTAAGTSQFCTHCKRWSSLAIVDIEEYELKDYNDNLFKVKINDG
EVRLLGKKGWRSGEKIKGKELFGPVKDAMRPNVDGLGMKIVKRKYLKLDLRDWVSRYGNM
AIFICPYVDCHHISHADKQAAFNIAV (SEQ ID NO: 3)

>CasY4
MSKRHPRISGVKGYRLHAQRLEYTGKSGAMRTIKYPLYSSPSGGRTVPREIVSAINDDYV
GLYGLSNFDDLYNAEKRNEEKVYSVLDFWYDCVQYGAVFSYTAPGLLKNVAEVRGGSYEL
TKTLKGSHLYDELQIDKVIKFLNKKEISRANGSLDKLKKDIIDCFKAEYRERHKDQCNKL
ADDIKNAKKDAGASLGERQKKLFRDFFGISEQSENDKPSFTNPLNLTCCLLPFDTVNNNR
NRGEVLFNKLKEYAQKLDKNEGSLEMWEYIGIGNSGTAFSNFLGEGFLGRLRENKITELK
KAMMDITDAWRGQEQEEELEKRLRILAALTIKLREPKFDNHWGGYRSDINGKLSSWLQNY
INQTVKIKEDLKGHKKDLKKAKEMINRFGESDTKEEAVVSSLLESIEKIVPDDSADDEKP
DIPAIAIYRRFLSDGRLTLNRFVQREDVQEALIKERLEAEKKKKPKKRKKKSDAEDEKET
IDFKELFPHLAKPLKLVPNFYGDSKRELYKKYKNAAIYTDALWKAVEKIYKSAFSSSLKN
SFFDTDFDKDFFIKRLQKIFSVYRRFNTDKWKPIVKNSFAPYCDIVSLAENEVLYKPKQS
RSRKSAAIDKNRVRLPSTENIAKAGIALARELSVAGFDWKDLLKKEEHEEYIDLIELHKT
ALALLLAVTETQLDISALDFVENGTVKDFMKTRDGNLVLEGRFLEMFSQSIVFSELRGLA
GLMSRKEFITRSAIQTMNGKQAELLYIPHEFQSAKITTPKEMSRAFLDLAPAEFATSLEP
ESLSEKSLLKLKQMRYYPHYFGYELTRTGQGIDGGVAENALRLEKSPVKKREIKCKQYKT
LGRGQNKIVLYVRSSYYQTQFLEWFLHRPKNVQTDVAVSGSFLIDEKKVKTRWNYDALTV
ALEPVSGSERVFVSQPFTIFPEKSAEEEGQRYLGIDIGEYGIAYTALEITGDSAKILDQN
FISDPQLKTLREEVKGLKLDQRRGTFAMPSTKIARIRESLVHSLRNRIHHLALKHKAKIV
YELEVSRFEEGKQKIKKVYATLKKADVYSEIDADKNLQTTVWGKLAVASEISASYTSQFC
GACKKLWRAEMQVDETITTQELIGTVRVIKGGTLIDAIKDFMRPPIFDENDTPFPKYRDF
CDKHHISKKMRGNSCLFICPFCRANADADIQASQTIALLRYVKEEKKVEDYFERFRKLKN
IKVLGQMKKI (SEQ ID NO: 4)

FIG. 1 (Cont.)

>CasY5
MAESKQMQCRKCGASMKYEVIGLGKKSCRYMCPDCGNHTSARKIQNKKKRD
KKYGSASKAQSQRIAVAGALYPDKKVQTIKTYKYPADLNGEVHDSGVAEKIAQAIQEDEI
GLLGPSSEYACWIASQKQSEPYSVVDFWFDAVCAGGVFAYSGARLLSTVLQLSGEESVLR
AALASSPFVDDINLAQAEKFLAVSRRTGQDKLGKRIGECFAEGRLEALGIKDRMREFVQA
IDVAQTAGQRFAAKLKIFGISQMPEAKQWNNDSGLTVCILPDYYVPEENRADQLVVLLRR
LREIAYCMGIEDEAGFEHLGIDPGALSNFSNGNPKRGFLGRLLNNDIIALANNMSAMTPY
WEGRKGELIERLAWLKHRAEGLYLKEPHFGNSWADHRSRIFSRIAGWLSGCAGKLKIAKD
QISGVRTDLFLLKRLLDAVPQSAPSPDFIASISALDRFLEAAAESSQDPAEQVRALYAFHL
NAPAVRSIANKAVQRSDSQEWLIKELDAVDHLEFNKAFPFFSDTGKKKKKGANSNGAPSE
EEYTETESIQQPEDAEQEVNGQEGNGASKNQKKFQRIPRFFGEGSRSEYRILTEAPQYFD
MFCNNMRAIFMQLESQPRKAPRDFKCFLQNRLQKLYKQTFLNARSNKCRALLESVLISWG
EFYTYGANEKKFRLRHEASERSSDPDYVVQQALEIARRLFLFGFEWRDCSAGERVDLVEI
HKKAISFLLAITQAEVSVGSYNWLGNSTVSRYLSVAGTDTLYGTQLEEFLNATVLSQMRG
LAIRLSSQELKDGFDVQLESSCQDNLQHLLVYRASRDLAACKRATCPAELDPKILVLPVG
AFIASVMKMIERGDEPLAGAYLRHRPHSFGWQIRVRGVAEVGMDQGTALAFQKPTESEPF
KIKPFSAQYGPVLWLNSSSYSQSQYLDGFLSQPKNWSMRVLPQAGSVRVEQRVALIWNLQ
AGKMRLERSGARAFFMPVPFSFRPSGSGDEAVLAPNRYLGLFPHSGGIEYAVVDVLDSAG
FKILERGTIAVNGFSQKRGERQEEAHREKQRRGISDIGRKKPVQAEVDAANELHRKYTDV
ATRLGCRIVVQWAPQPKPGTAPTAQTVYARAVRTEAPRSGNQEDHARMKSSWGYTWGTYW
EKRKPEDILGISTQVYWTGGIGESCPAVAVALLGHIRATSTQTEWEKEEVVFGRLKKFFP
S (SEQ ID NO: 6)

>CasY6
MAESKQMQCRKCGASMKYEVIGLGKKSCRYMCPDCGNHTSARKIQNKKKRD
KKYGSASKAQSQRIAVAGALYPDKKVQTIKTYKYPADLNGEVHDRGVAEKIEQAIQEDEI
GLLGPSSEYACWIASQKQSEPYSVVDFWFDAVCAGGVFAYSGARLLSTVLQLSGEESVLR
AALASSPFVDDINLAQAEKFLAVSRRTGQDKLGKRIGECFAEGRLEALGIKDRMREFVQA
IDVAQTAGQRFAAKLKIFGISQMPEAKQWNNDSGLTVCILPDYYVPEENRADQLVVLLRR
LREIAYCMGIEDEAGFEHLGIDPGALSNFSNGNPKRGFLGRLLNNDIIALANNMSAMTPY
WEGRKGELIERLAWLKHRAEGLYLKEPHFGNSWADHRSRIFSRIAGWLSGCAGKLKIAKD
QISGVRTDLFLLKRLLDAVPQSAPSPDFIASISALDRFLEAAAESSQDPAEQVRALYAFHL
NAPAVRSIANKAVQRSDSQEWLIKELDAVDHLEFNKAFPFFSDTGKKKKKGANSNGAPSE
EEYTETESIQQPEDAEQEVNGQEGNGASKNQKKFQRIPRFFGEGSRSEYRILTEAPQYFD
MFCNNMRAIFMQLESQPRKAPRDFKCFLQNRLQKLYKQTFLNARSNKCRALLESVLISWG
EFYTYGANEKKFRLRHEASERSSDPDYVVQQALEIARRLFLFGFEWRDCSAGERVDLVEI
HKKAISFLLAITQAEVSVGSYNWLGNSTVSRYLSVAGTDTLYGTQLEEFLNATVLSQMRG
LAIRLSSQELKDGFDVQLESSCQDNLQHLLVYRASRDLAACKRATCPAELDPKILVLPAG
AFIASVMKMIERGDEPLAGAYLRHRPHSFGWQIRVRGVAEVGMDQGTALAFQKPTESEPF
KIKPFSAQYGPVLWLNSSSYSQSQYLDGFLSQPKNWSMRVLPQAGSVRVEQRVALIWNLQ
AGKMRLERSGARAFFMPVPFSFRPSGSGDEAVLAPNRYLGLFPHSGGIEYAVVDVLDSAG
FKILERGTIAVNGFSQKRGERQEEAHREKQRRGISDIGRKKPVQAEVDAANELHRKYTDV
ATRLGCRIVVQWAPQPKPGTAPTAQTVYARAVRTEAPRSGNQEDHARMKSSWGYTWSTYW
EKRKPEDILGISTQVYWTGGIGESCPAVAVALLGHIRATSTQTEWEKEEVVFGRLKKFFP
S (SEQ ID NO: 7)

FIG. 1 (Cont.)

>CasY7
MKRILNSLKVAALRLLFRGKGSELVKTVKYPLVSPVQGAVEELAEAIRHDNLHLFGQKEI
VDLMEKDEGTQVYSVVDFWLDTLRLGMFFSPSANALKITLGKFNSDQVSPFRKVLEQSPF
FLAGRLKVEPAERILSVEIRKIGKRENRVENYAADVETCFIGQLSSDEKQSIQKLANDIW
DSKDHEEQRMLKADFFAIPLIKDPKAVTEEDPENETAGKQKPLELCVCLVPELYTRGFGS
IADFLVQRLTLLRDKMSTDTAEDCLEYVGIEEEKGNGMNSLLGTFLKNLQGDGFEQIFQF
MLGSYVGWQGKEDVLRERLDLLAEKVKRLPKPKFAGEWSGHRMFLHGQLKSWSSNFFRLF
NETRELLESIKSDIQHATMLISYVEEKGGYHPQLLSQYRKLMEQLPALRTKVLDPEIEMT
HMSEAVRSYIMIHKSVAGFLPDLLESLDRDKDREFLLSIFPRIPKIDKKTKEIVAWELPG
EPEEGYLFTANNLFRNFLENPKHVPRFMAERIPEDWTRLRSAPVWFDGMVKQWQKVVNQL
VESPGALYQFNESFLRQRLQAMLTVYKRDLQTEKFLKLLADVCRPLVDFFGLGGNDIIFK
SCQDPRKQWQTVIPLSVPADVYTACEGLAIRLRETLGFEWKNLKGHEREDFLRLHQLLGN
LLFWIRDAKLVVKLEDWMNNPCVQEYVEARKAIDLPLEIFGFEVPIFLNGYLFSELRQLE
LLLRRKSVMTSYSVKTTGSPNRLFQLVYLPLNPSDPEKKNSNNFQERLDTPTGLSRRFLD
LTLDAFAGKLLTDPVTQELKTMAGFYDHLFGFKLPCKLAAMSNHPGSSSKMVVLAKPKKG
VASNIGFEPIPDPAHPVFRVRSSWPELKYLEGLLYLPEDTPLTIELAETSVSCQSVSSVA
FDLKNLTTILGRVGEFRVTADQPFKLTPIIPEKEESFIGKTYLGLDAGERSGVGFAIVTV
DGDGYEVQRLGVHEDTQLMALQQVASKSLKEPVFQPLRKGTFRQQERIRKSLRGCYWNFY
HALMIKYRAKVVHEESVGSSGLVGQWLRAFQKDLKKADVLPKKGGKNGVDKKKRESSAQD
TLWGGAFSKKEEQQIAFEVQAAGSSQFCLKCGWWFQLGMREVNRVQESGVVLDWNRSIVT
FLIESSGEKVYGFSPQQLEKGFRPDIETFKKMVRDFMRPPMFDRKGRPAAAYERFVLGRR
HRRYRFDKVFEERFGRSALFICPRVGCGNFDHSSEQSAVVLALIGYIADKEGMSGKKLVY
VRLAELMAEWKLKKLERSRVEEQSSAQ (SEQ ID NO: 5)

>CasY18
MKRIAKFRHDKPVKREAWSKGYRVHKNRIINKVTRSIKYPLVVKDEWKKRLIDDAAHDYRWLVG
PINYSDWCRDPNQYSILEFWIDFLCVGGVFQSSHSNICRLAIQLSGGSVFEQEWKDLSPFVRAN
LIQGIKPAEFIGFLTAEFRSSSNPKNFISKFFEGSNEDLESLTNEFASIVDFIKAKDISLLRKS
LPSCKKIAPNLWEKAVGSHSTNELLKLLTKYTRVMLVAEPSHSDRVFSQTVLQSNDQDDPELTG
PLPSHKVGKASYLFIPEFIREVNLDKISKLDLSAKSKLAVEQVKKLSELTSDFKQIENQSEAYF
GLSTSFNELSNFLGILIRTLRNAPEAILKDQIALCAPLDKDILKITLDWLCDRAQALPENPRFE
TNWAEYRSYLGGKIKSWFSNYENFFEIPQAASSQQNNNREKKLGNRSAIRALNLKKEAFEKARE
TFKGDKGTLEKIDLAYRLLGSISPEVLQCDEGLKLYQQFNDELLVLNETINQKFQDAKRDIKAK
KEKESFEKLQRNLSSPLPRIPEFFGERAKKGYQKARVSPKLARHLLECLNDWLARFAKVEESAF
SEKEFQRILDWLRTSDFLPVFIRKSKDPPSWLRYIARVATGKYYFWVSEYSRKRVQIIDKPIAQ
NPLKELISWFLLNKDAFSRDNELFKGLSSKMVTLARIMAGILRDRGEGLKELQAMTSKLDNIGL
LHPSFSVPVTDSLKDAAFYRAFFSELEGLLNIGRSRLIIERITLQSQQSKNKKTRRPLMPEPFI
NEDKEVFLAFPKFETKNKVKGTRVVYNSPDEVNWLLSPIRSSKGQLSFMFRCLSEDAKIMTTSG
GCSYIVEFKKLLEAQEEVLSIHDCDIIPRAFVSIPFTLERESEETKPDWKPNRFMGVDIGEYAV
AYCVIEKGTDSIEILDCGIVRNGAHRVLKEKVDRLKRRQRSMTFGAMDTSIAAARESLVGNYRN
RLHAIALKHGAKLVYEYEVSAFESGGNRIKKVYETLKKSDCTGETEADKNARKHIWGETNAVGD
QIGAGWTSQTCAKCGRSFGADLKAGNFGVAVPVPEKVEDSKGHYAYHEFPFEDGLKVRGFLKPN
KIISDQKELAKAVHAYMRPPLVALGKRKLPKNARYRRGNSSLFRCPFSDCGFTADADIQAAYNI
AVKQLYKPKKGYPKERKWQDFVILKPKEPSKLFDKQFYRPN (SEQ ID NO: 8)

FIG. 2

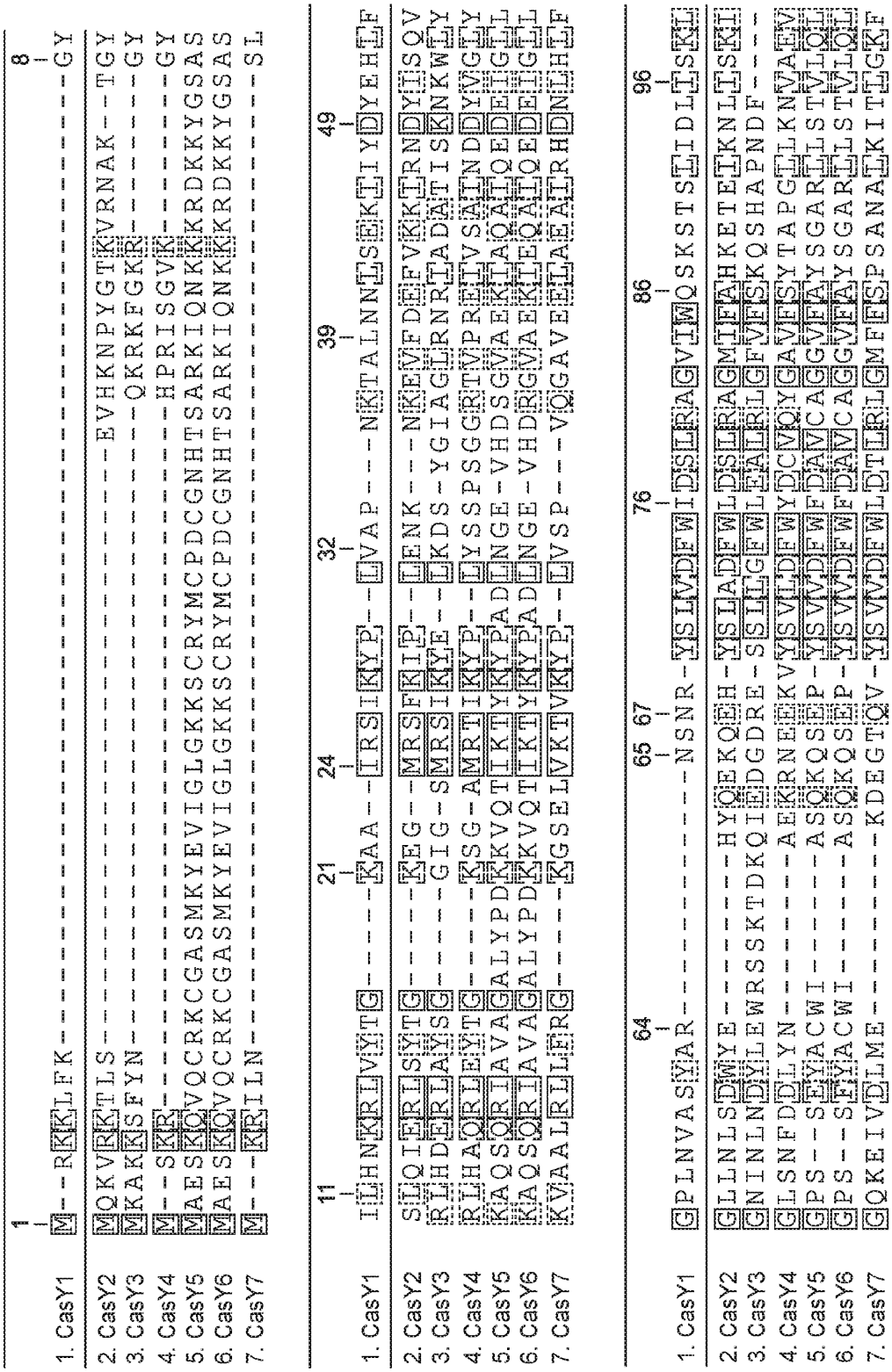


FIG. 2 (Cont.)

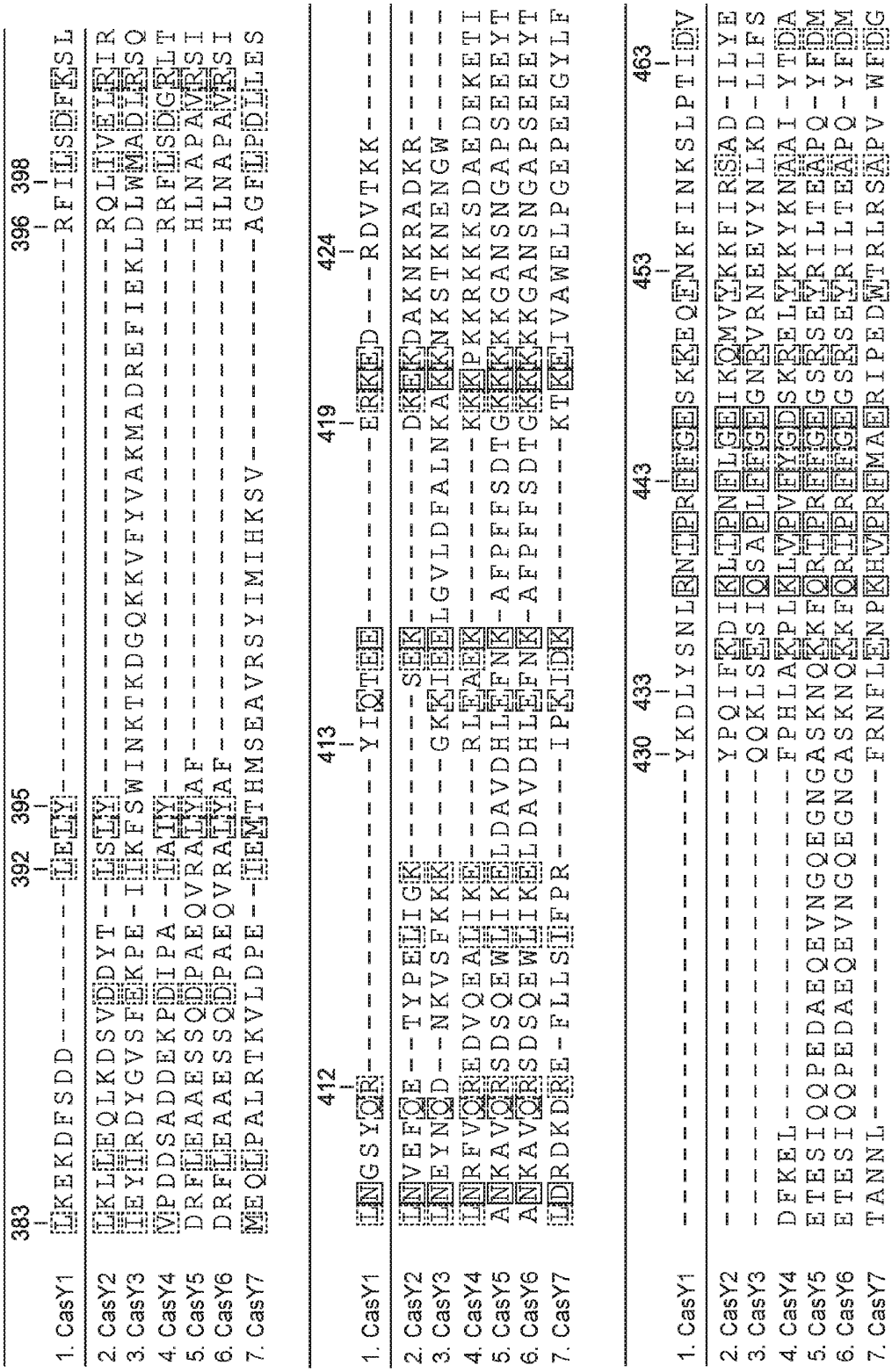
A (Cont.)

FIG. 2 (Cont.)

A (Cont.)

FIG. 2 (Cont.)

A (Cont.)



A (Cont.)

FIG. 2 (Cont.)

A (Cont.)

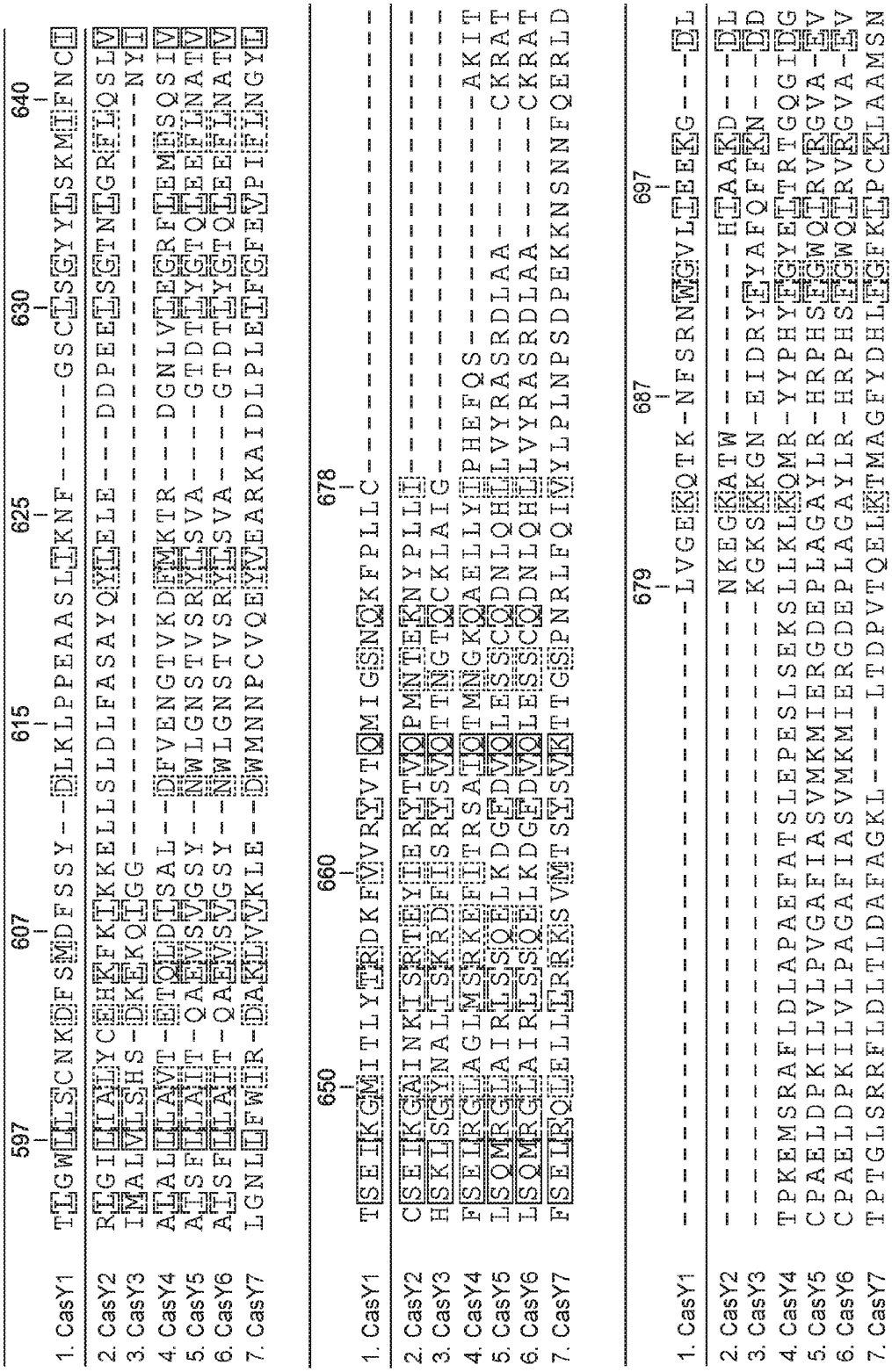


FIG. 2 (Cont.)

FIG. 2 (Cont.)

A (Cont.)

FIG. 2 (Cont.)

A (Cont.)

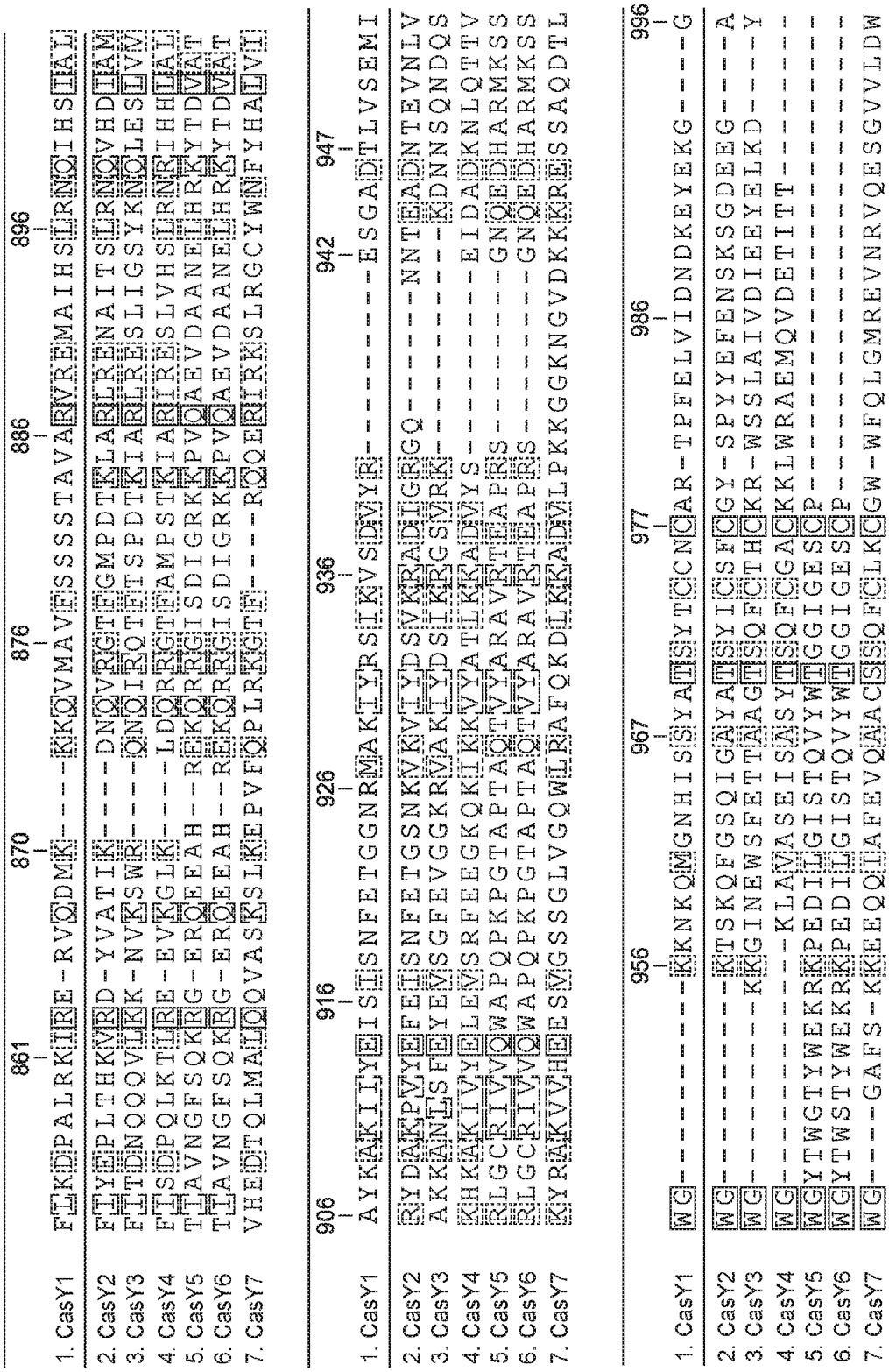


FIG. 2 (Cont.)

A (Cont.)

FIG. 2 (Cont.)

B

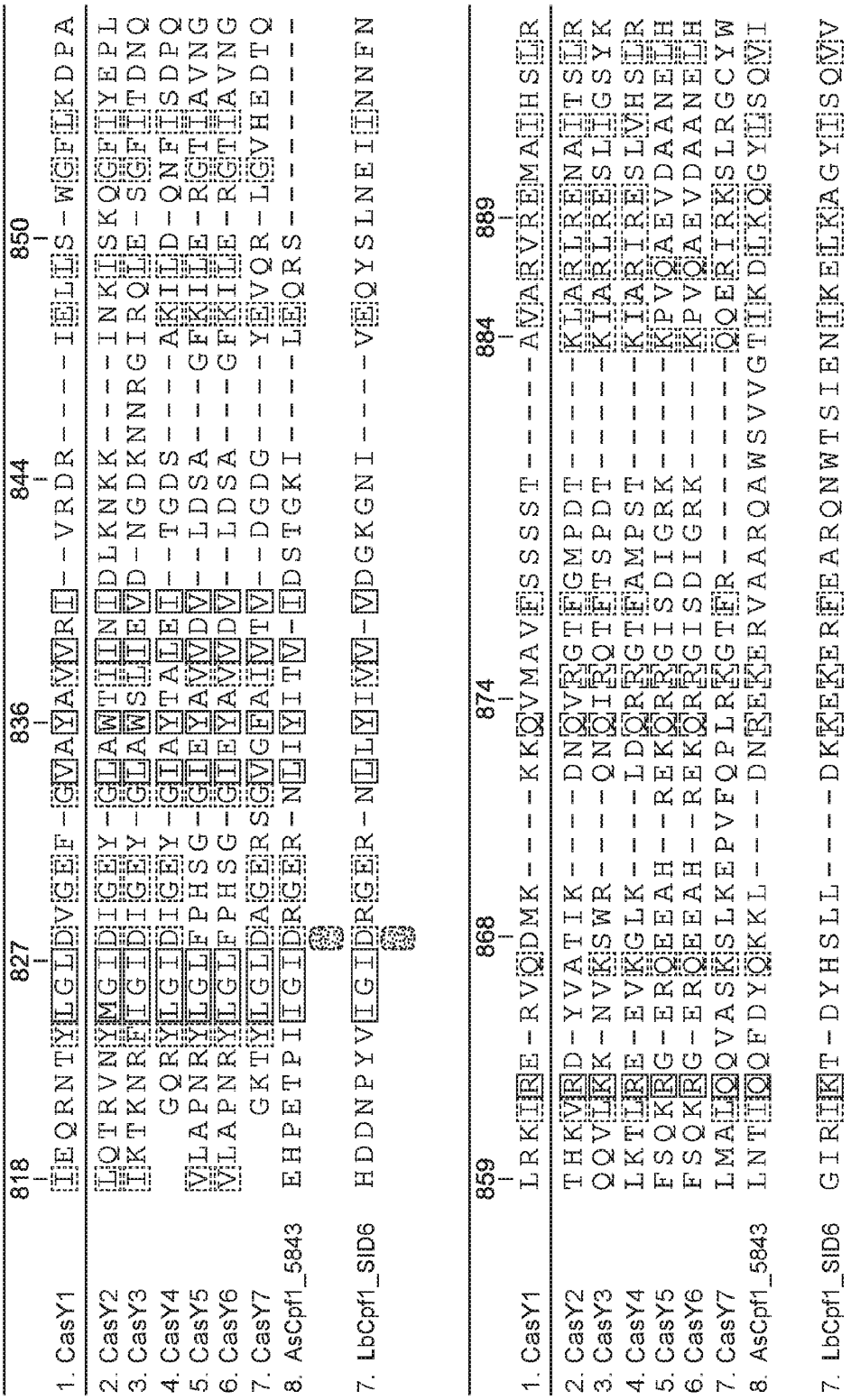

818
827
836
844
850
1. CasY1
2. CasY2
3. CasY3
4. CasY4
5. CasY5
6. CasY6
7. CasY7
8. AsCpf1_5843
7. LbCpf1_SID6
859
868
874
884
889
1. CasY1
2. CasY2
3. CasY3
4. CasY4
5. CasY5
6. CasY6
7. CasY7
8. AsCpf1_5843
7. LbCpf1_SID6

FIG. 2 (Cont.)

B (Cont.)

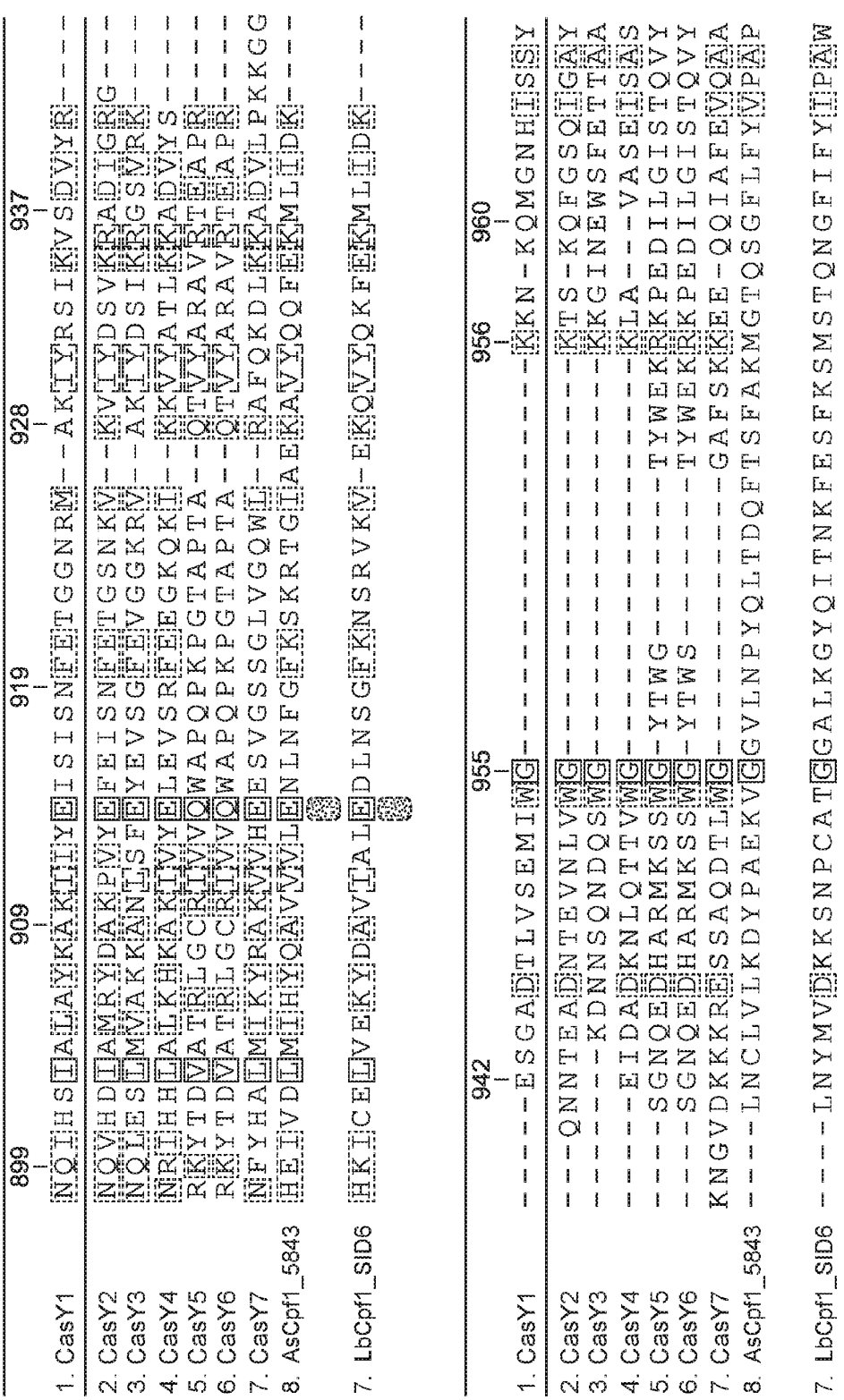


FIG. 2 (Cont.)

B (Cont.)

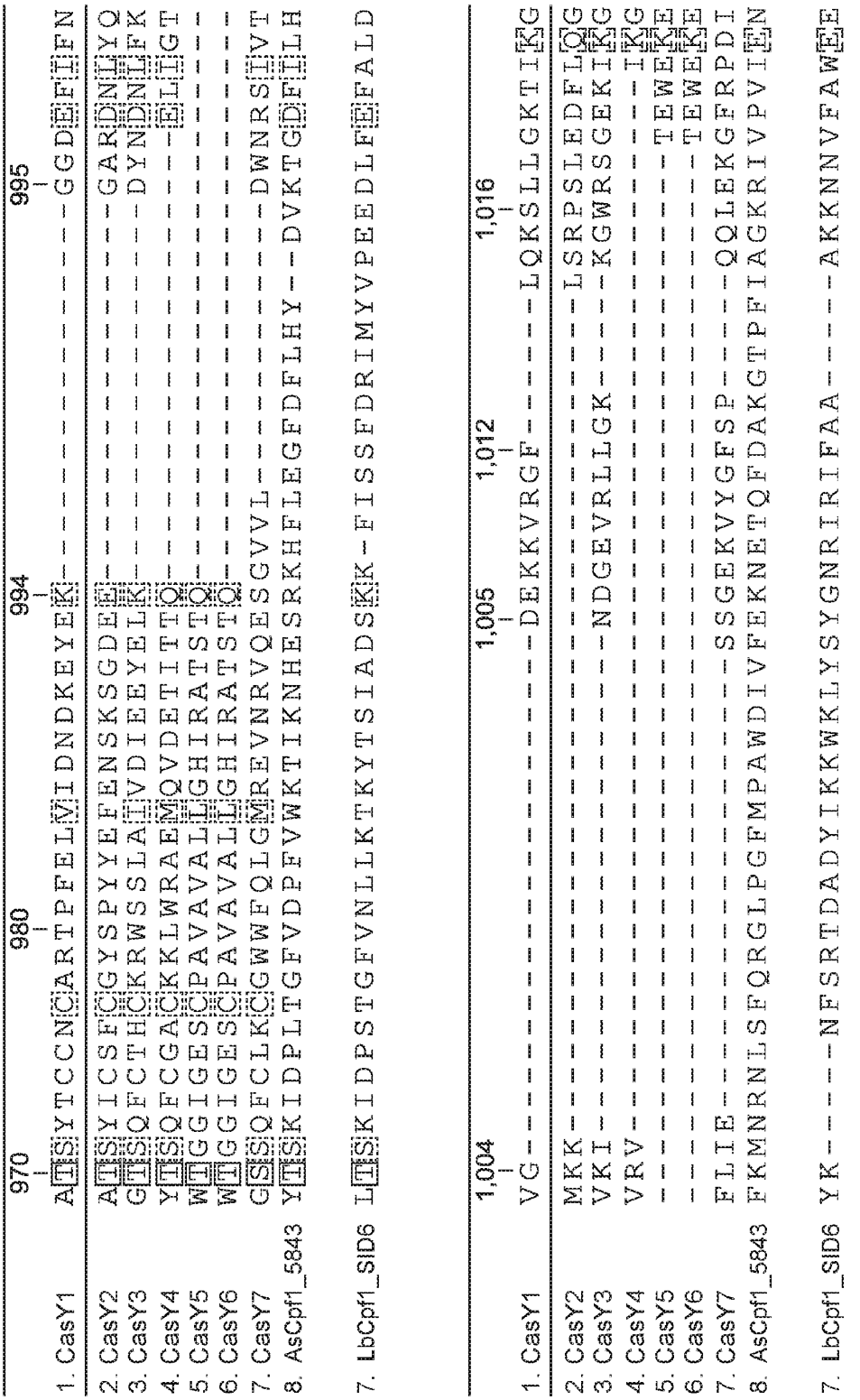


FIG. 2 (Cont.)

B (Cont.)

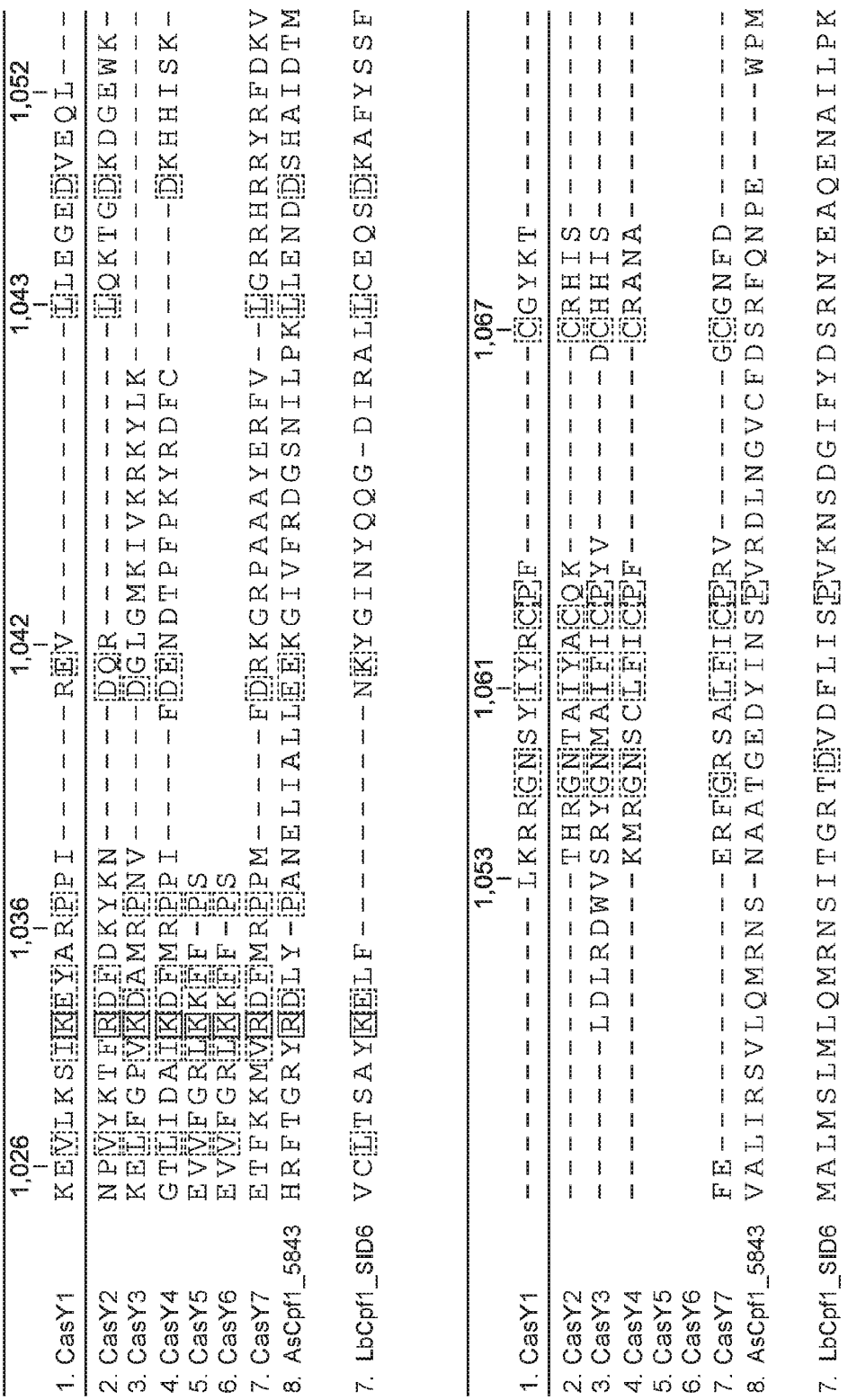


FIG. 2 (Cont.)

B (Cont.)

FIG. 3

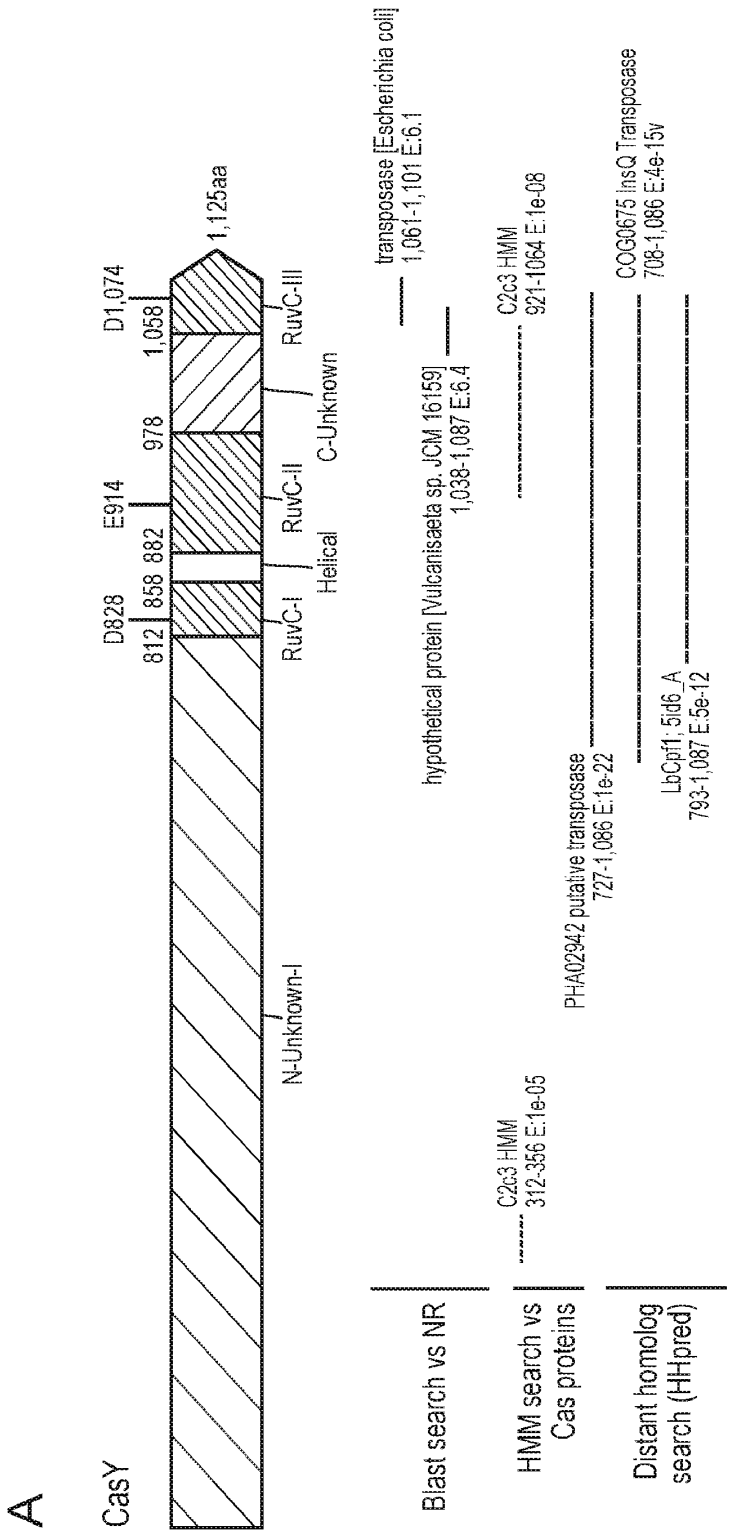

A
CasY
1,125aa
D1,074
1,058
978
E914
882
858
D828
812
N-Unknown-I
RuvC-I
Helical
RuvC-II
C-Unknown
RuvC-III
Blast search vs NR
transposase [Escherichia coli]
1,061-1,101 E:6.1
hypothetical protein [Vulcanisaeta sp. JCM 16159]
1,038-1,087 E:6.4
HMM search vs Cas proteins
C2c3 HMM
312-356 E:1e-05
C2c3 HMM
921-1064 E:1e-08
Distant homolog search (HHpred)
PHA02942 putative transposase
727-1,086 E:1e-22
COG0675 InsQ Transposase
708-1,086 E:4e-15v
LbCpf1, 5id6_A
793-1,087 E:5e-12

FIG. 3 (Cont.)

B

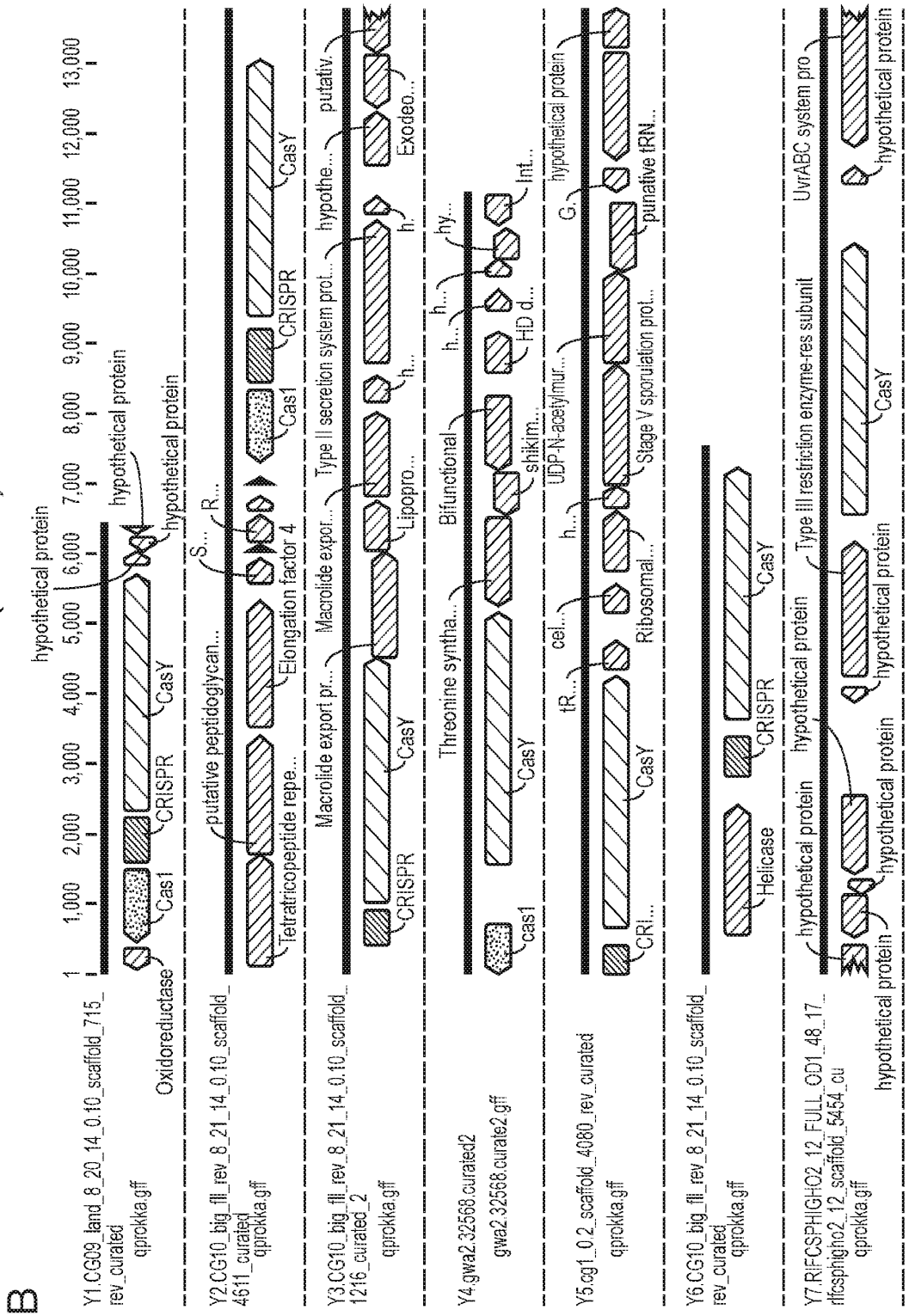

1
1,000
2,000
3,000
4,000
5,000
6,000
7,000
8,000
9,000
10,000
11,000
12,000
13,000
hypothetical protein
Y1.CG09_land_8_20_14_0.10_scaffold_715_rev_curated qprokka.gff
Oxidoreductase
Cas1
CRISPR
CasY
hypothetical protein
hypothetical protein
Y2.CG10_big_fll_rev_8_21_14_0.10_scaffold_4611_curated qprokka.gff
putative peptidoglycan...
S...
R...
Tetratricopeptide repe...
Elongation factor 4
Cas1
CRISPR
CasY
Y3.CG10_big_fll_rev_8_21_14_0.10_scaffold_1216_curated_2 qprokka.gff
Macrolide export pr...
Macrolide expor...
Type II secretion system prot...
hypothe...
putativ...
CRISPR
CasY
Lipopro...
h...
h.
Exodeo...
Y4.gwa2.32568.curated2 gwa2.32568.curate2.gff
Threonine syntha...
Bifunctional
h...
h...
hy...
cas1
CasY
shikim...
HD d...
Int...
Y5.cg1_0.2_scaffold_4080_rev_curated qprokka.gff
tR...
cel...
h...
UDP-N-acetylmur...
G...
hypothetical protein
CRI...
CasY
Ribosomal...
Stage V sporulation prot...
putative tRN...
Y6.CG10_big_fll_rev_8_21_14_0.10_scaffold_rev_curated qprokka.gff
Helicase
CRISPR
CasY
Y7.RIFCSPHIGHO2_12_FULL_OD1_48_17_rifcsphigho2_12_scaffold_5454_cu qprokka.gff
hypothetical protein
hypothetical protein
Type III restriction enzyme-res subunit
UvrABC system pro
hypothetical protein
hypothetical protein
hypothetical protein
CasY
hypothetical protein

FIG. 4

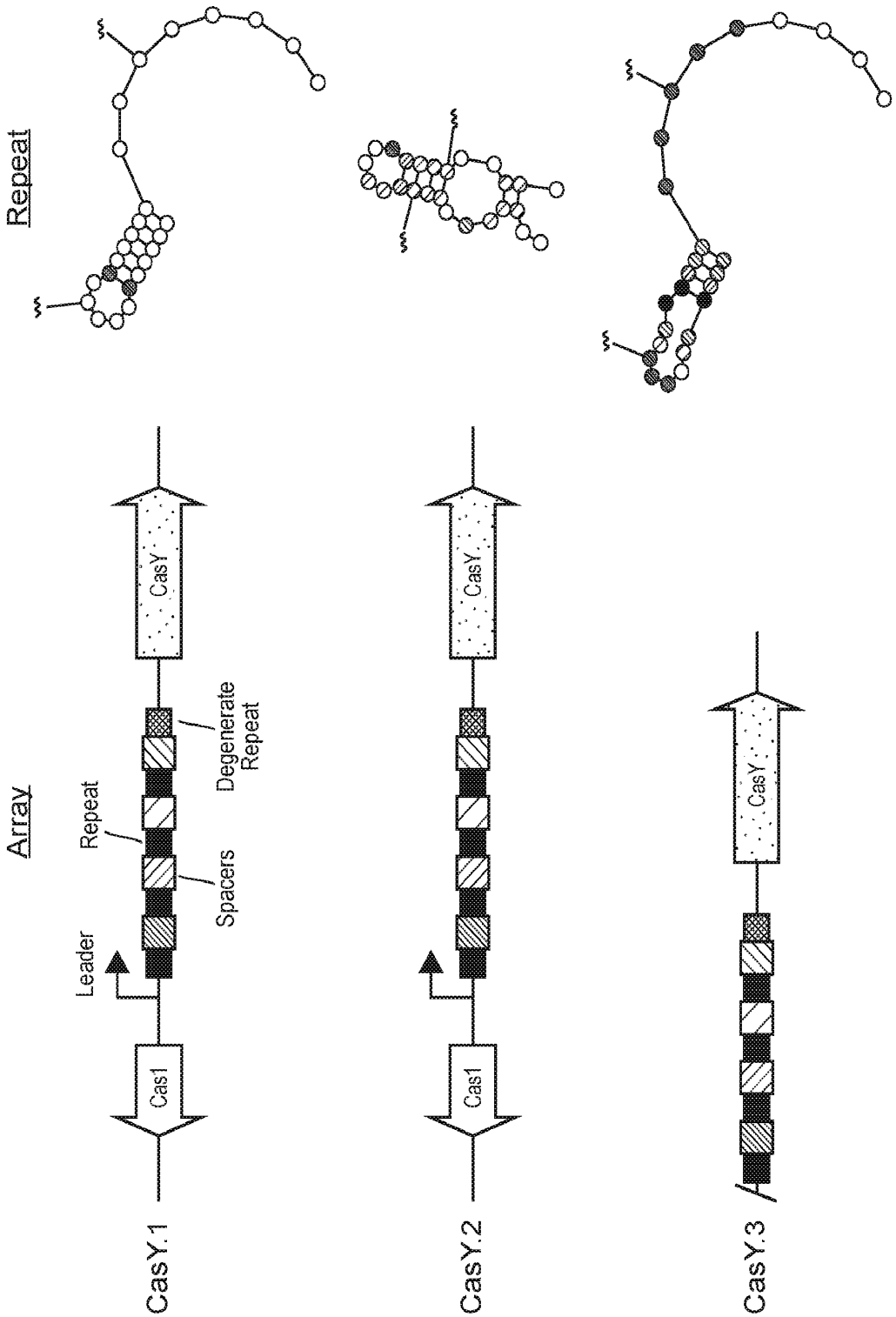

Repeat
Array
CasY
Degenerate Repeat
Repeat
Spacers
Leader
Cas1
CasY.1
CasY
Cas1
CasY.2
CasY
CasY.3

FIG. 4 (Cont.)

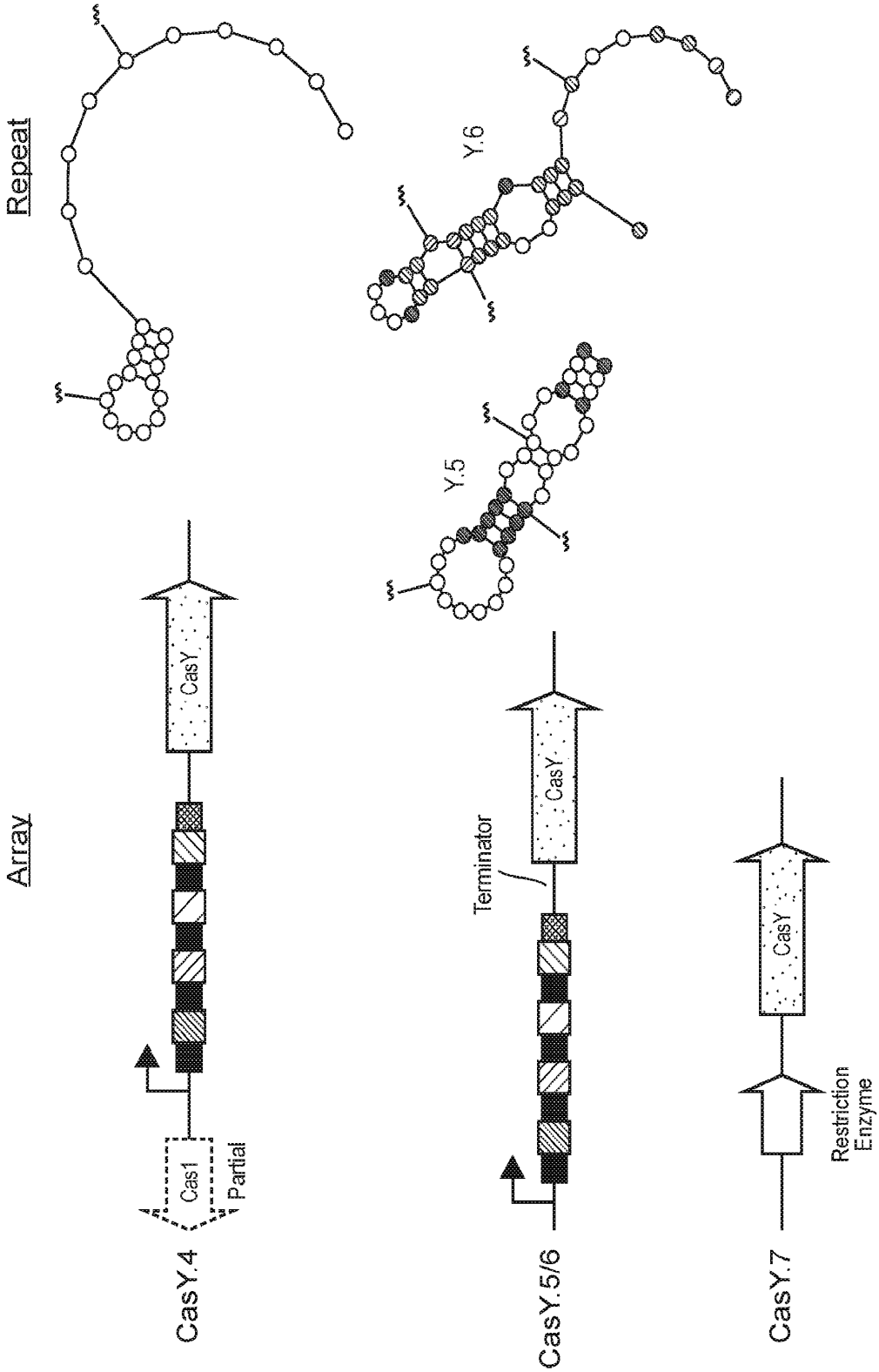

Repeat
Y.6
Y.5
Array
CasY
Cas1
Partial
CasY.4
Terminator
CasY
CasY.5/6
CasY
Restriction Enzyme
CasY.7

FIG. 4 (Cont.)

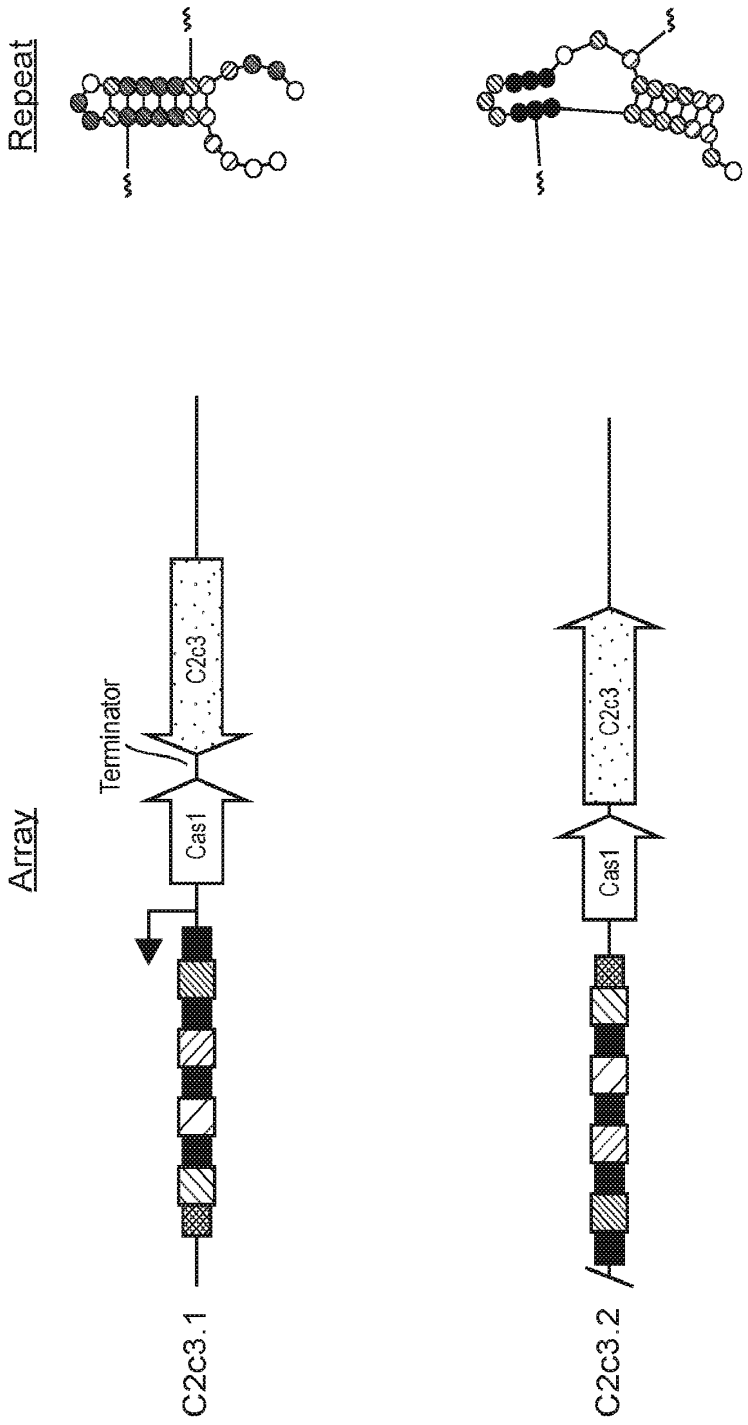


FIG. 5

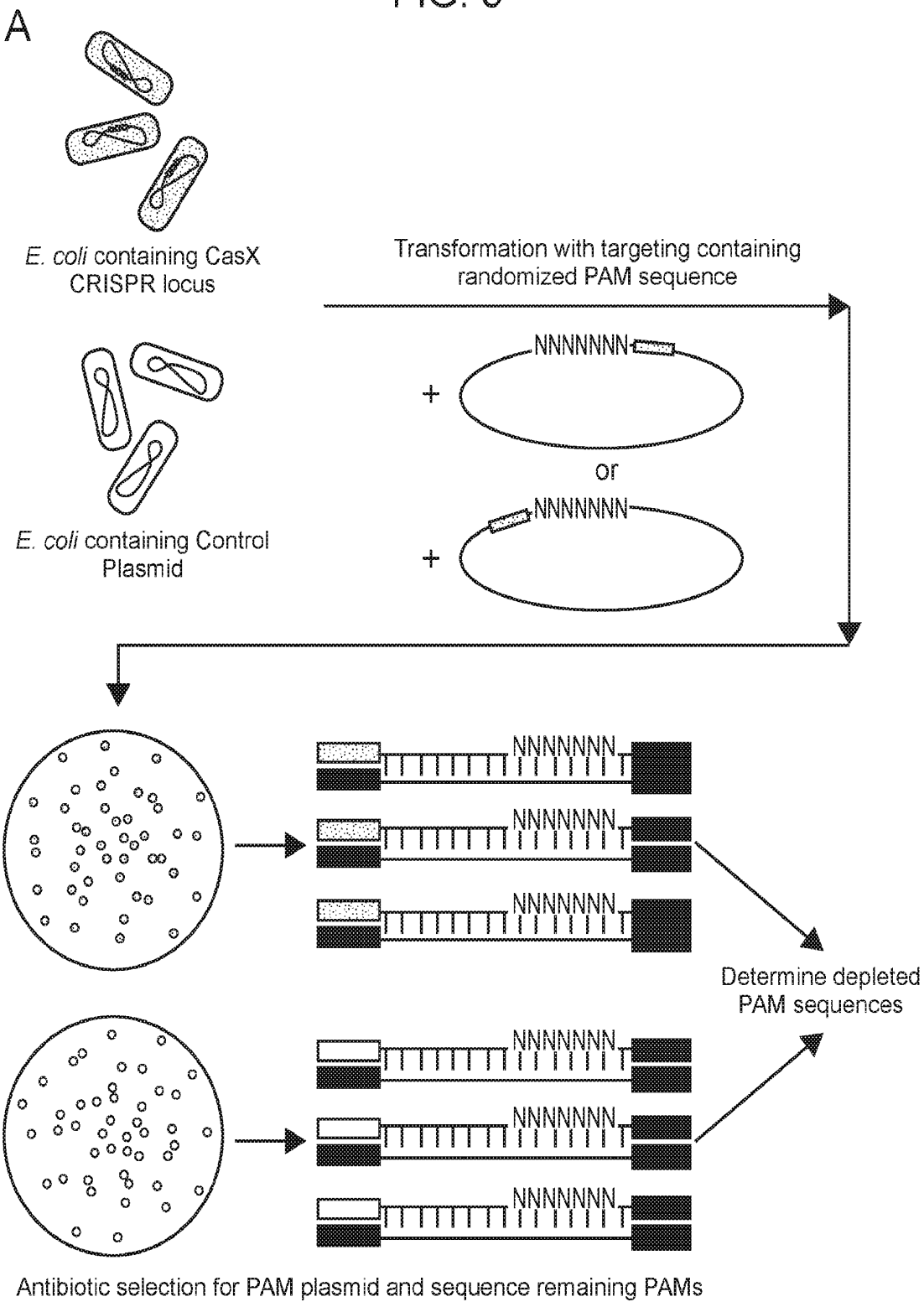

A
E. coli containing CasX CRISPR locus
E. coli containing Control Plasmid
Transformation with targeting containing randomized PAM sequence
NNNNNNN
+
or
NNNNNNN
+
NNNNNNN
NNNNNNN
NNNNNNN
Determine depleted PAM sequences
NNNNNNN
NNNNNNN
NNNNNNN
Antibiotic selection for PAM plasmid and sequence remaining PAMs

FIG. 5 (Cont.)

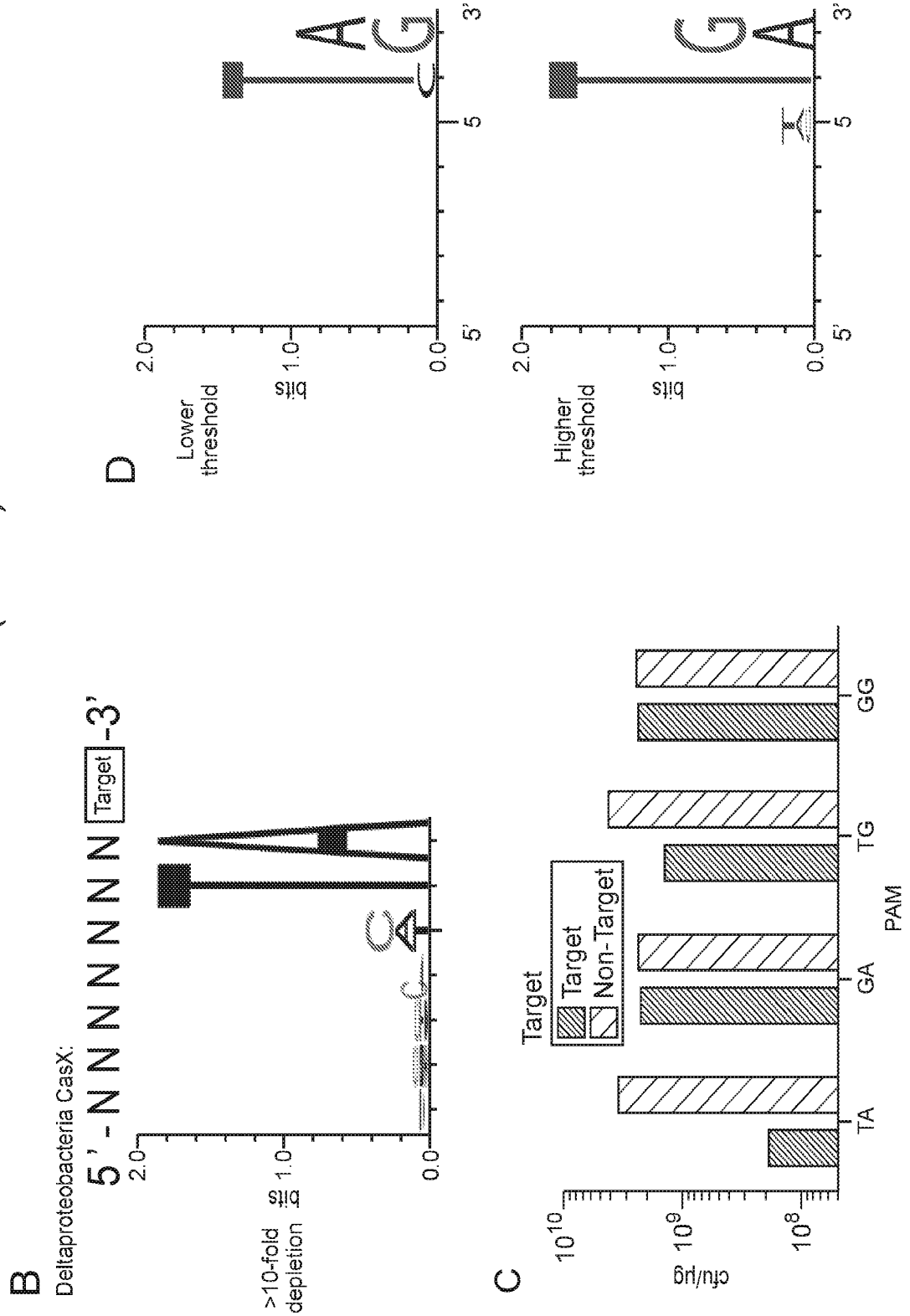

B
Deltaproteobacteria CasX:
5'-NNNNNNN Target -3'
>10-fold depletion
bits
2.0
1.0
0.0
C
Target
Target
Non-Target
cfu/µg
10^{10}
10^{9}
10^{8}
TA
GA
TG
GG
PAM
D
Lower threshold
Higher threshold
bits
2.0
1.0
0.0
5'
3'

FIG. 6

A

	Repeat
CasY1	CTCCGAAAGTATCGGGGATAAAGGC
CasY2	CACCGAAATTTGGAGAGGATAAGGC
CasY3	CTCCGAATTATCGGGAGGATAAGGC
CasY4	CCCCGAATATAGGGGACAAAAAGGC
CasY5 (and Y6)	GTCTAGACATACAGGTGGAAAGGTGAGAGTAAAGAC

B

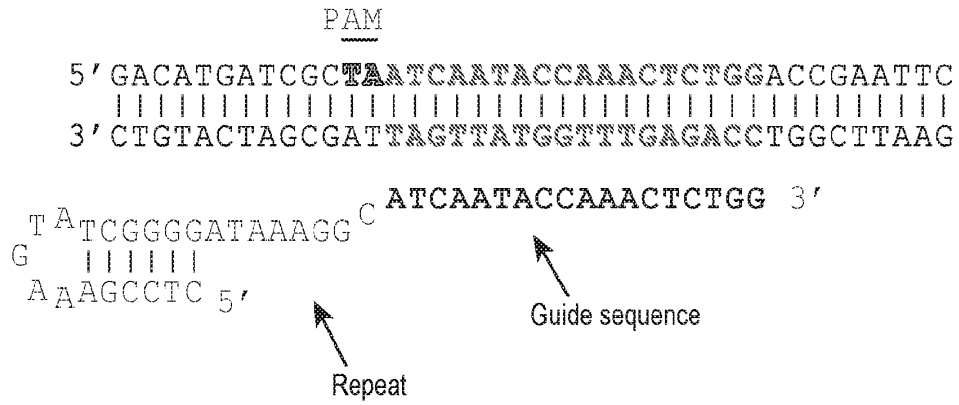


FIG. 7

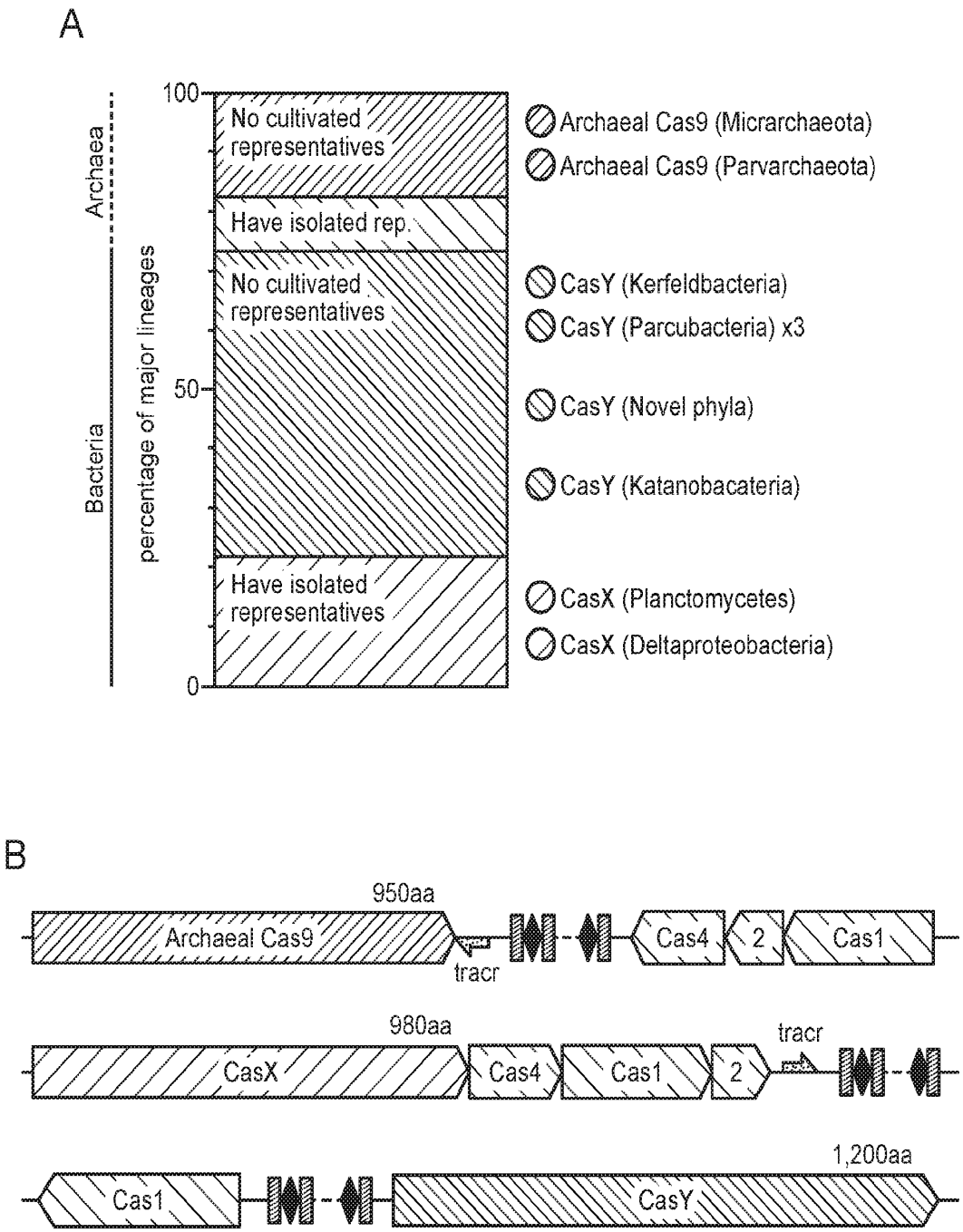

A
Archaea
Bacteria
percentage of major lineages
100
50
0
No cultivated representatives
Have isolated rep.
No cultivated representatives
Have isolated representatives
Archaeal Cas9 (Micrarchaeota)
Archaeal Cas9 (Parvarchaeota)
CasY (Kerfeldbacteria)
CasY (Parcubacteria) x3
CasY (Novel phyla)
CasY (Katanobacateria)
CasX (Planctomycetes)
CasX (Deltaproteobacteria)
B
950aa
Archaeal Cas9
tracr
Cas4
2
Cas1
980aa
CasX
Cas4
Cas1
2
tracr
1,200aa
Cas1
CasY

FIG. 8

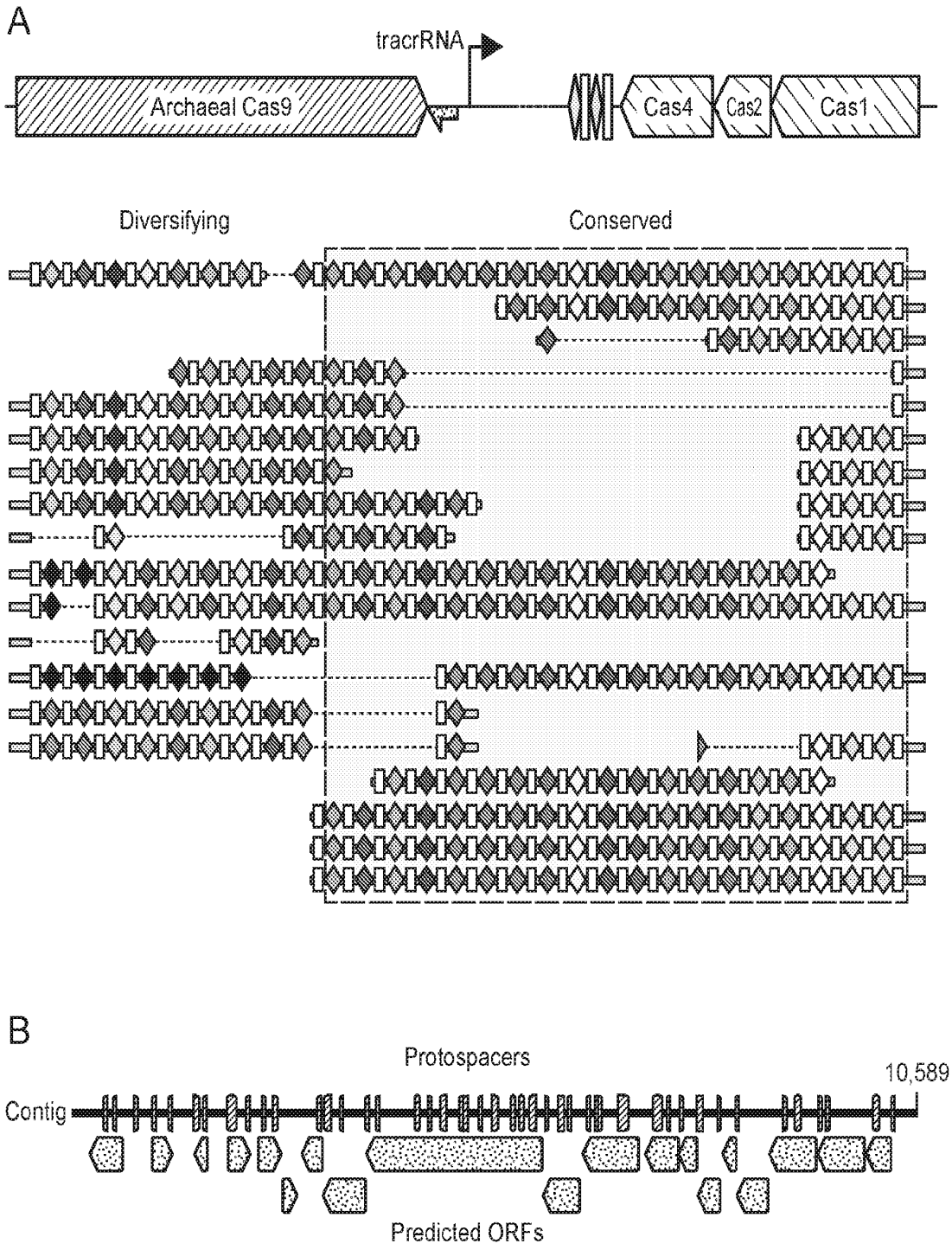

A
tracrRNA
Archaeal Cas9
Cas4
Cas2
Cas1
Diversifying
Conserved
B
Protospacers
10,589
Contig
Predicted ORFs

FIG. 8 (Cont.)

C

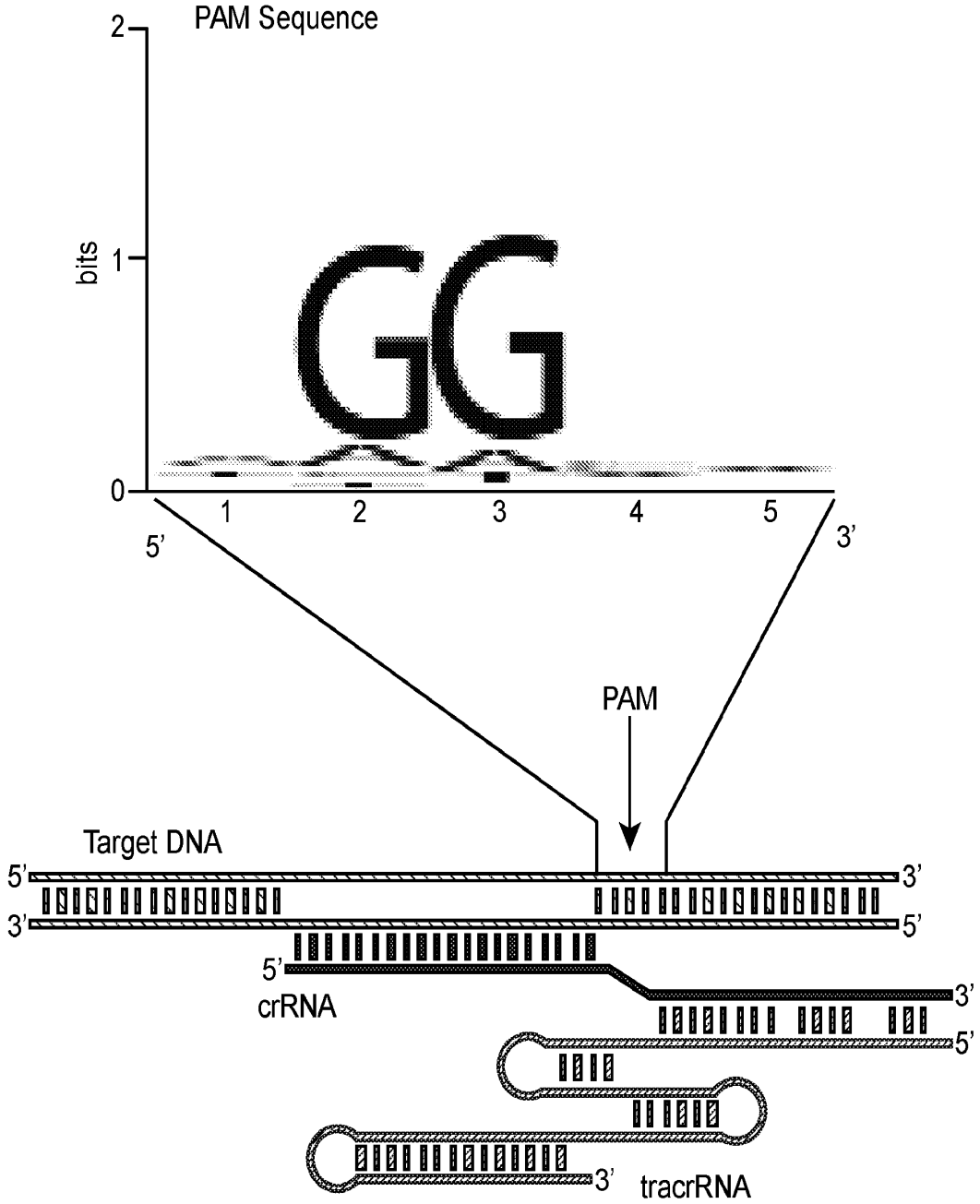

PAM Sequence
bits
2
1
0
GG
1
2
3
4
5
5'
3'
PAM
Target DNA
5'
3'
3'
5'
5'
crRNA
3'
5'
3'
tracrRNA

FIG. 9

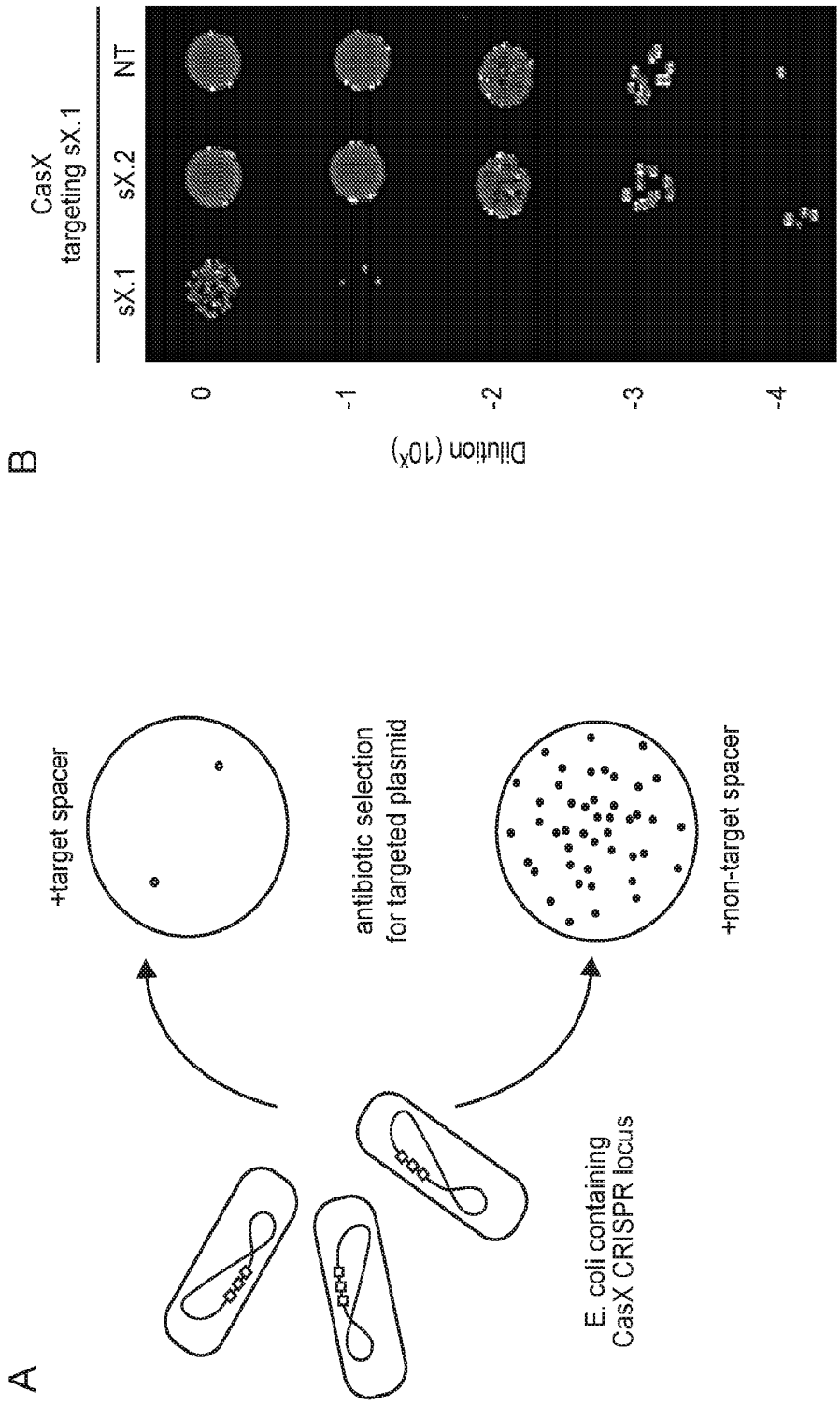
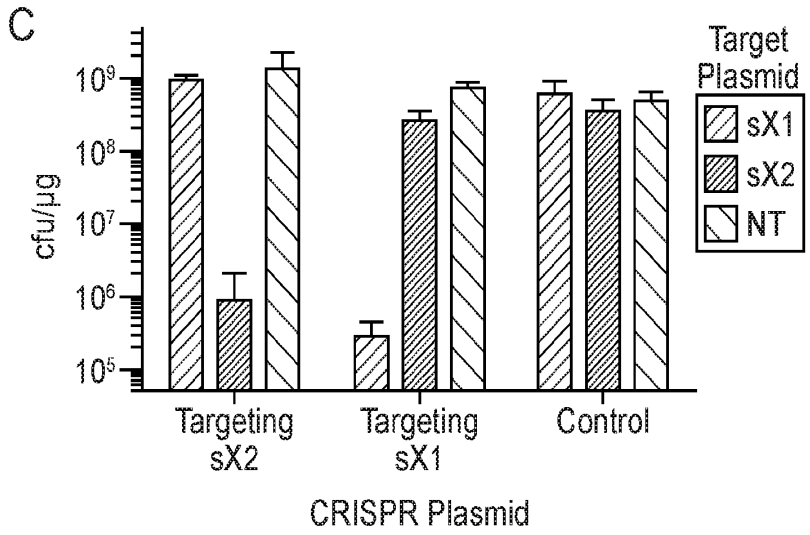

A
E. coli containing
CasX CRISPR locus
antibiotic selection
for targeted plasmid
+target spacer
+non-target spacer
B
CasX
targeting sX.1
sX.1
sX.2
NT
0
-1
-2
-3
-4
Dilution (10x)

FIG. 9 (Cont.)


C
cfu/µg
10^9
10^8
10^7
10^6
10^5
Target Plasmid
sX1
sX2
NT
Targeting sX2
Targeting sX1
Control
CRISPR Plasmid

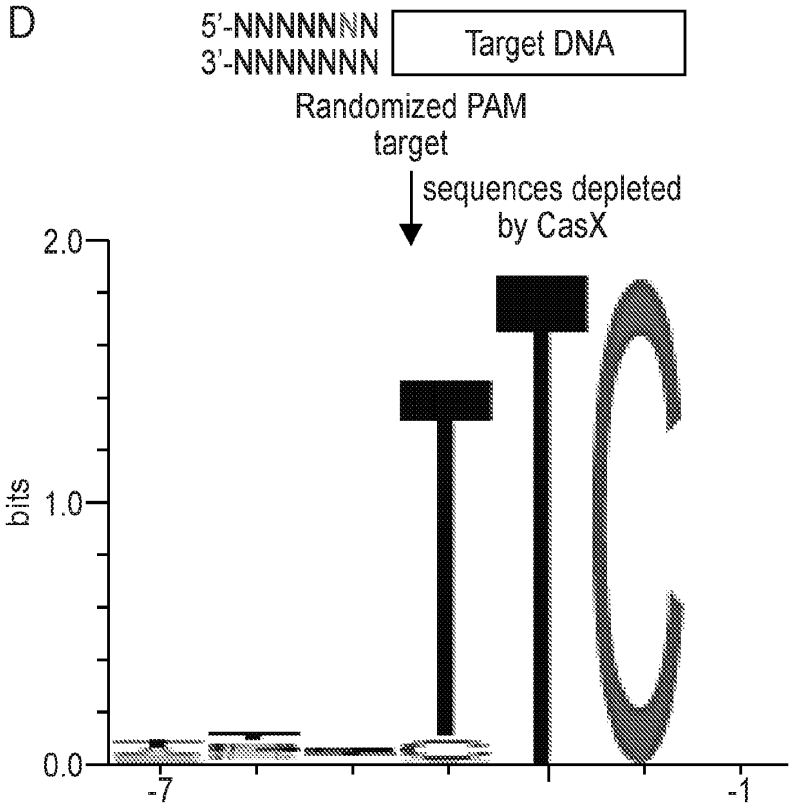

D
5'-NNNNNNN
3'-NNNNNNN
Target DNA
Randomized PAM target
sequences depleted by CasX
bits
2.0
1.0
0.0
-7
-1

FIG. 10

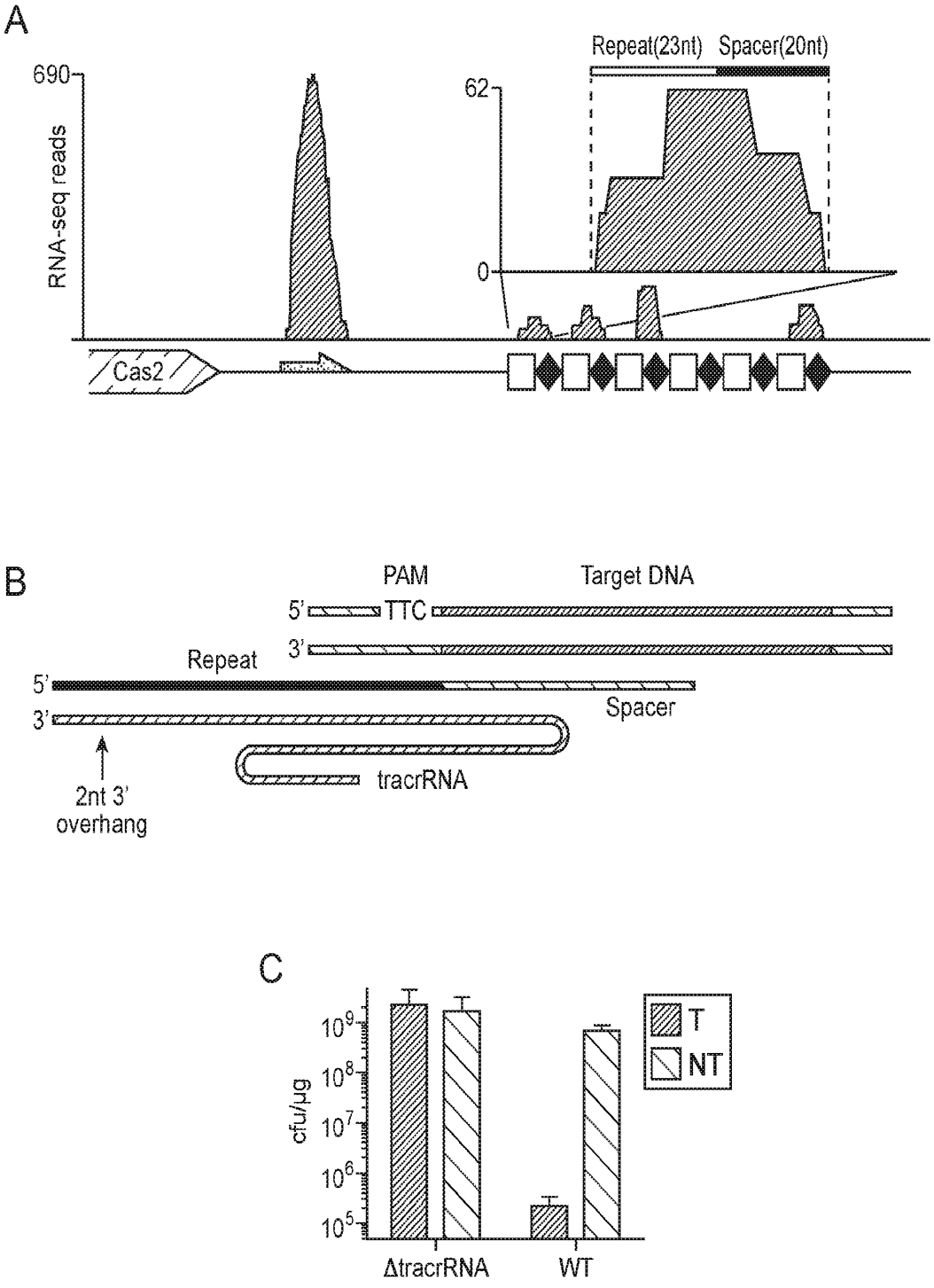

A
690
RNA-seq reads
62
0
Repeat(23nt)
Spacer(20nt)
Cas2
B
PAM
Target DNA
5'
TTC
3'
Repeat
5'
3'
Spacer
tracrRNA
2nt 3'
overhang
C
cfu/μg
10^9
10^8
10^7
10^6
10^5
T
NT
ΔtracrRNA
WT

FIG. 11

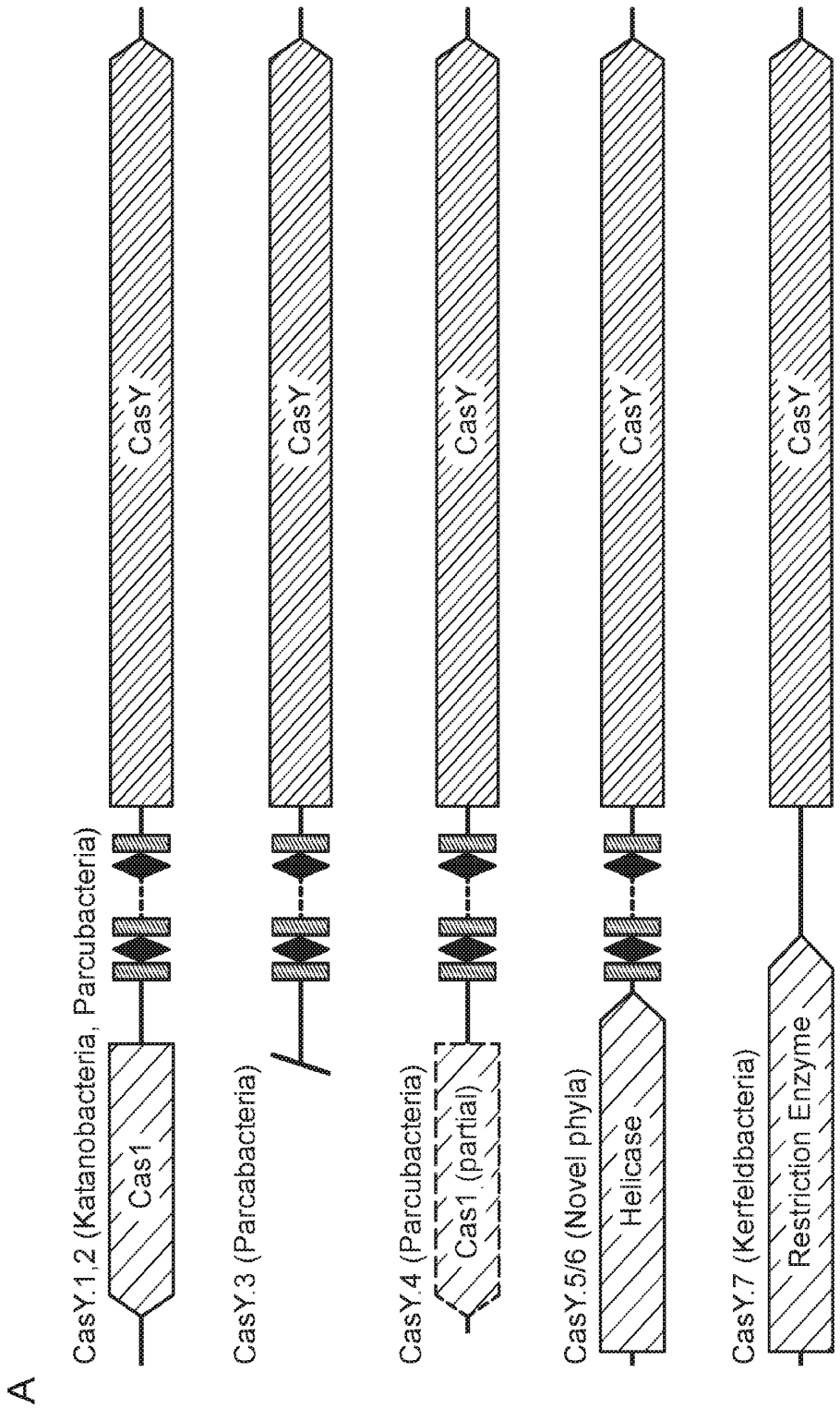

A
CasY.1,2 (Katanobacteria, Parcubacteria)
CasY
Cas1
CasY.3 (Parcabacteria)
CasY
CasY.4 (Parcubacteria)
CasY
Cas1 (partial)
CasY.5/6 (Novel phyla)
CasY
Helicase
CasY.7 (Kerfeldbacteria)
CasY
Restriction Enzyme

FIG. 11 (Cont.)

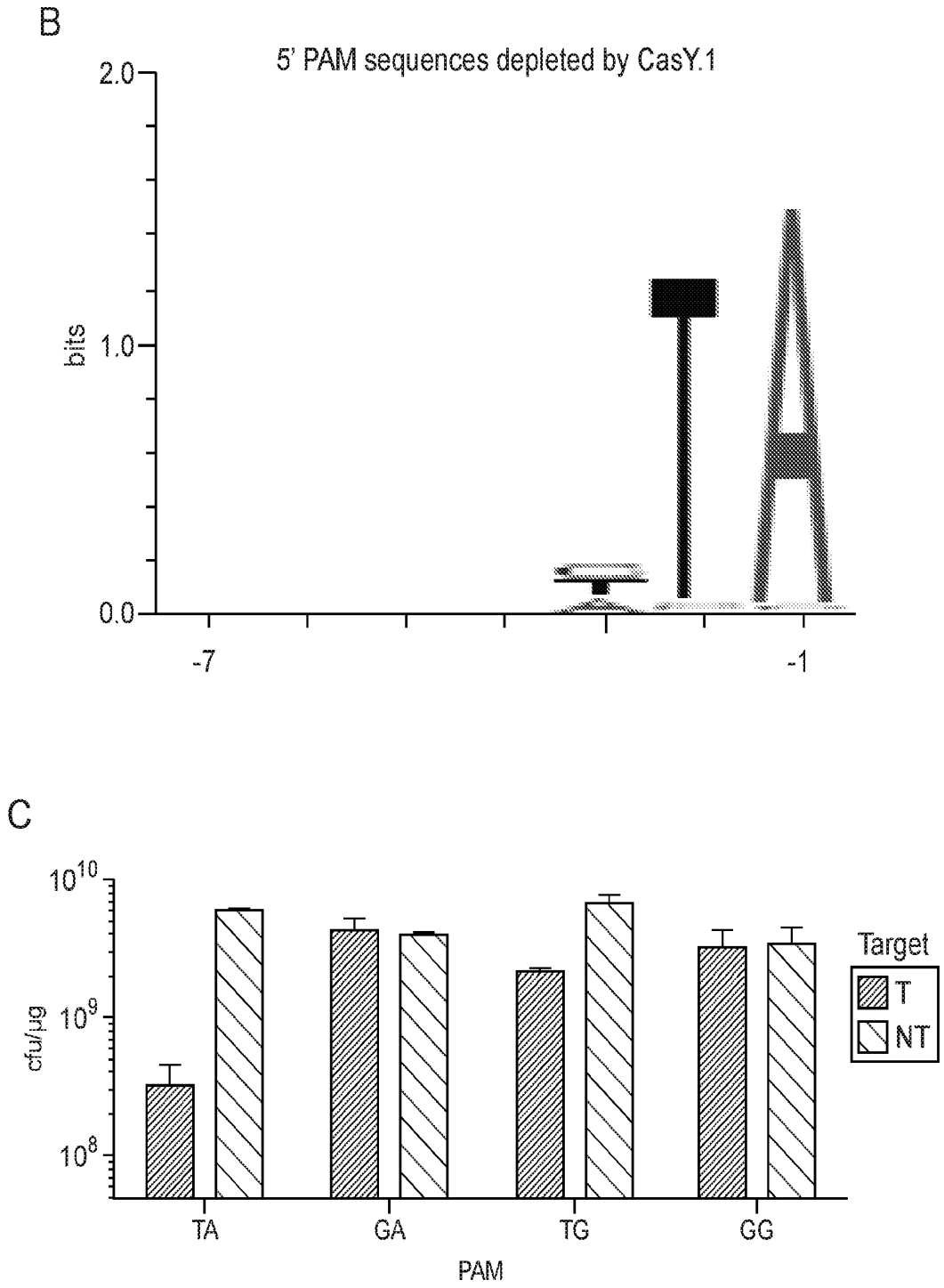
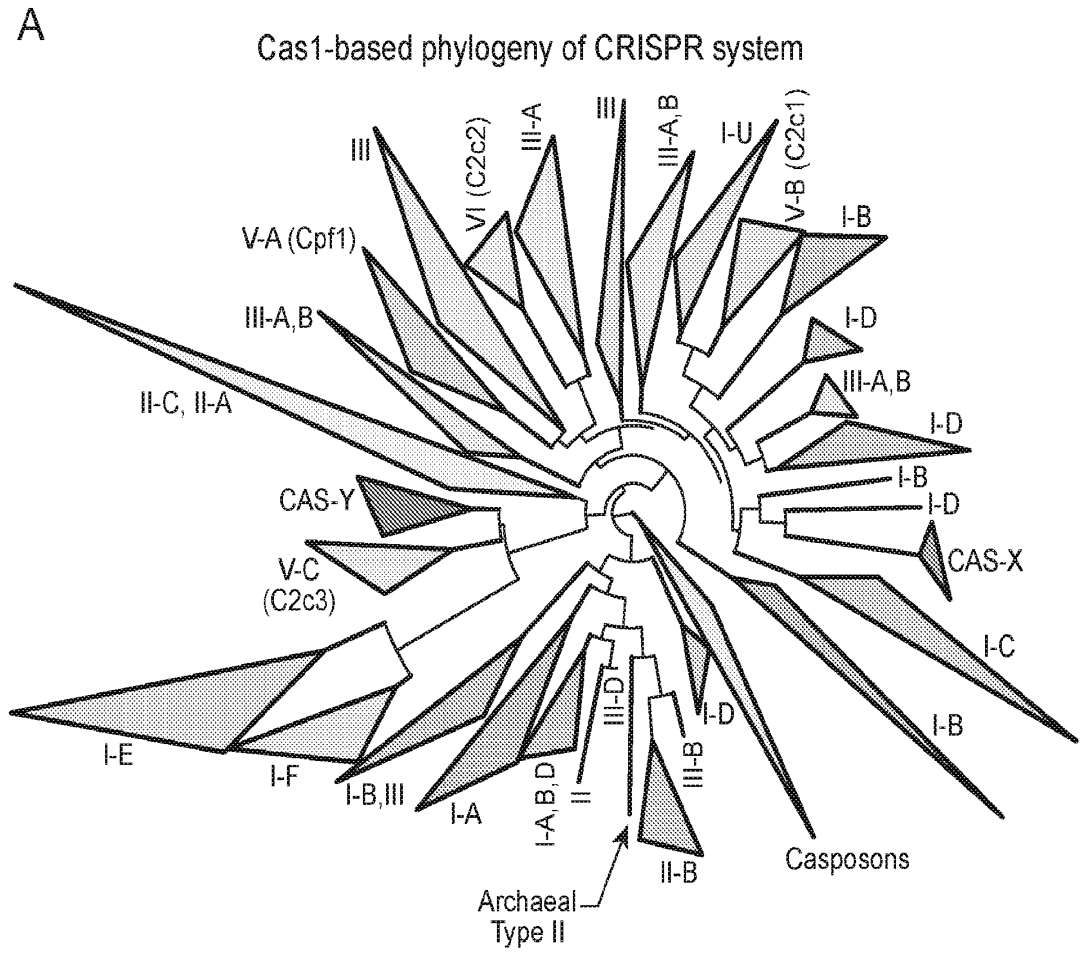

B
5' PAM sequences depleted by CasY.1
2.0
1.0
0.0
bits
-7
-1
C
10^{10}
10^{9}
10^{8}
cfu/µg
TA
GA
TG
GG
PAM
Target
T
NT

FIG. 12


A
Cas1-based phylogeny of CRISPR system
III
VI (C2c2)
III-A
III
III-A,B
I-U
V-B (C2c1)
I-B
V-A (Cpf1)
III-A,B
I-D
II-C, II-A
III-A,B
I-D
I-B
CAS-Y
I-D
V-C
(C2c3)
CAS-X
I-C
I-E
I-F
I-B,III
I-A
I-A,B,D
II
III-D
III-B
I-D
I-B
II-B
Casposons
Archaeal
Type II

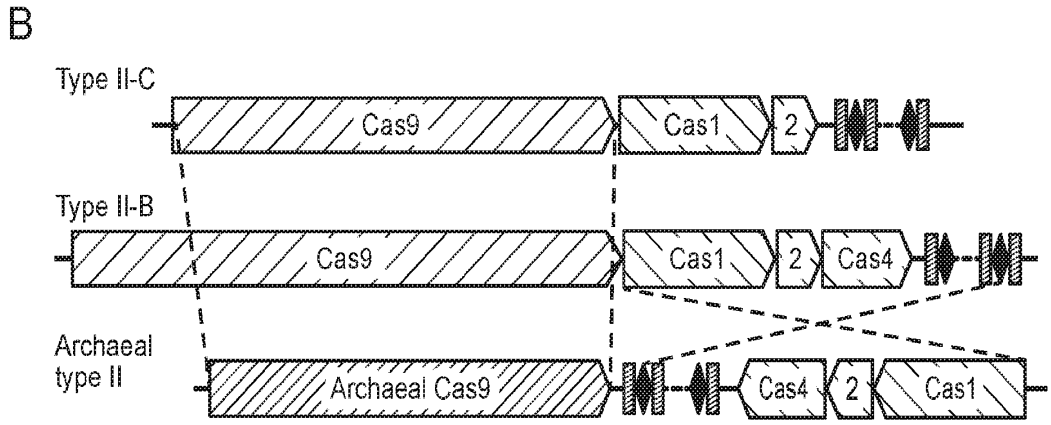
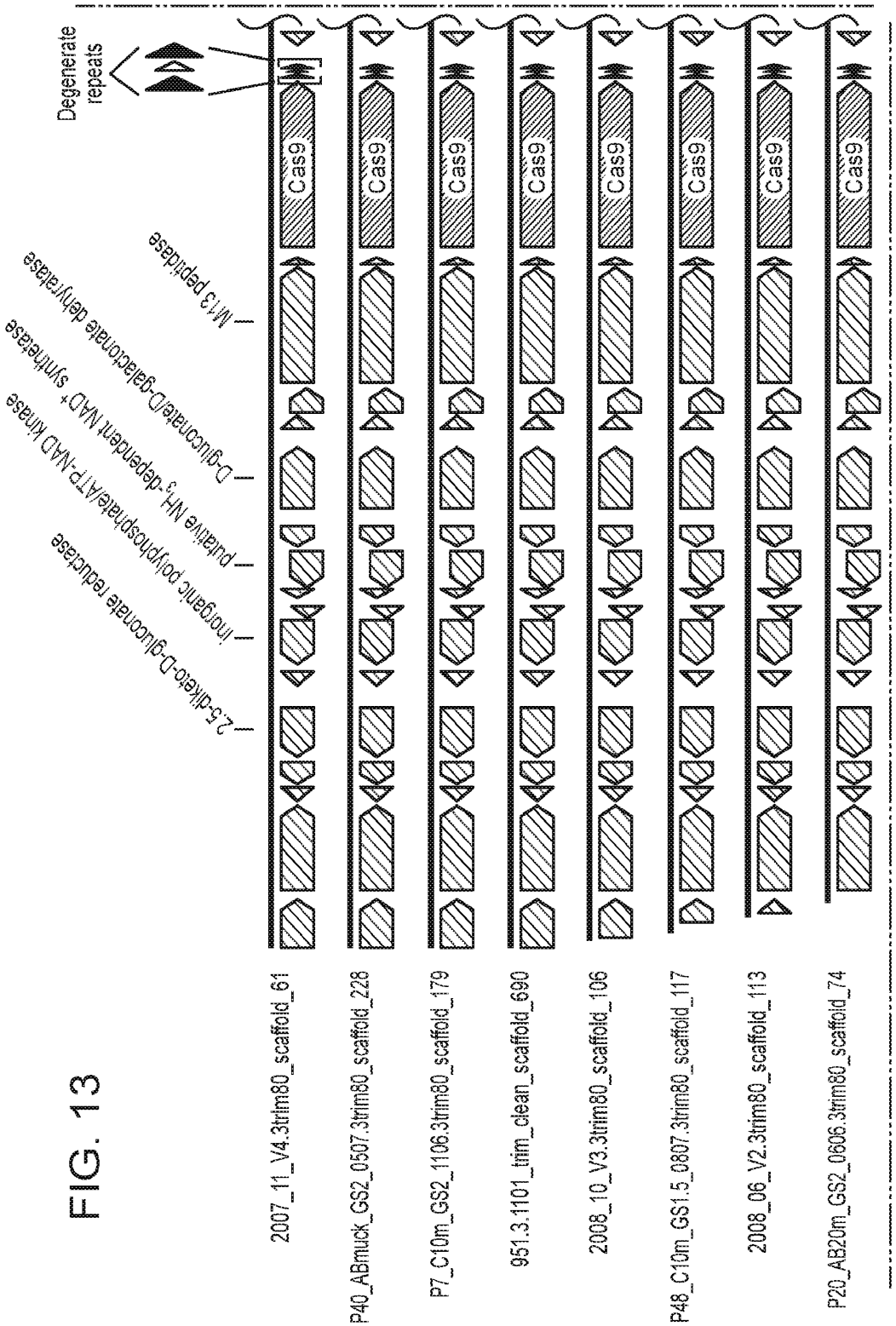

B
Type II-C
Cas9
Cas1
2
Type II-B
Cas9
Cas1
2
Cas4
Archaeal
type II
Archaeal Cas9
Cas4
2
Cas1

FIG. 13


2007_11_V4.3trim80_scaffold_61
P40_ABmuck_GS2_0507.3trim80_scaffold_228
P7_C10m_GS2_1106.3trim80_scaffold_179
951.3.1101_trim_clean_scaffold_690
2008_10_V3.3trim80_scaffold_106
P48_C10m_GS1.5_0807.3trim80_scaffold_117
2008_06_V2.3trim80_scaffold_113
P20_AB20m_GS2_0606.3trim80_scaffold_74
Cas9
Degenerate repeats
M13 peptidase
D-gluconate/D-galactonate dehydratase
putative NH3-dependent NAD+ synthetase
inorganic polyphosphate/ATP-NAD kinase
2,5-diketo-D-gluconate reductase

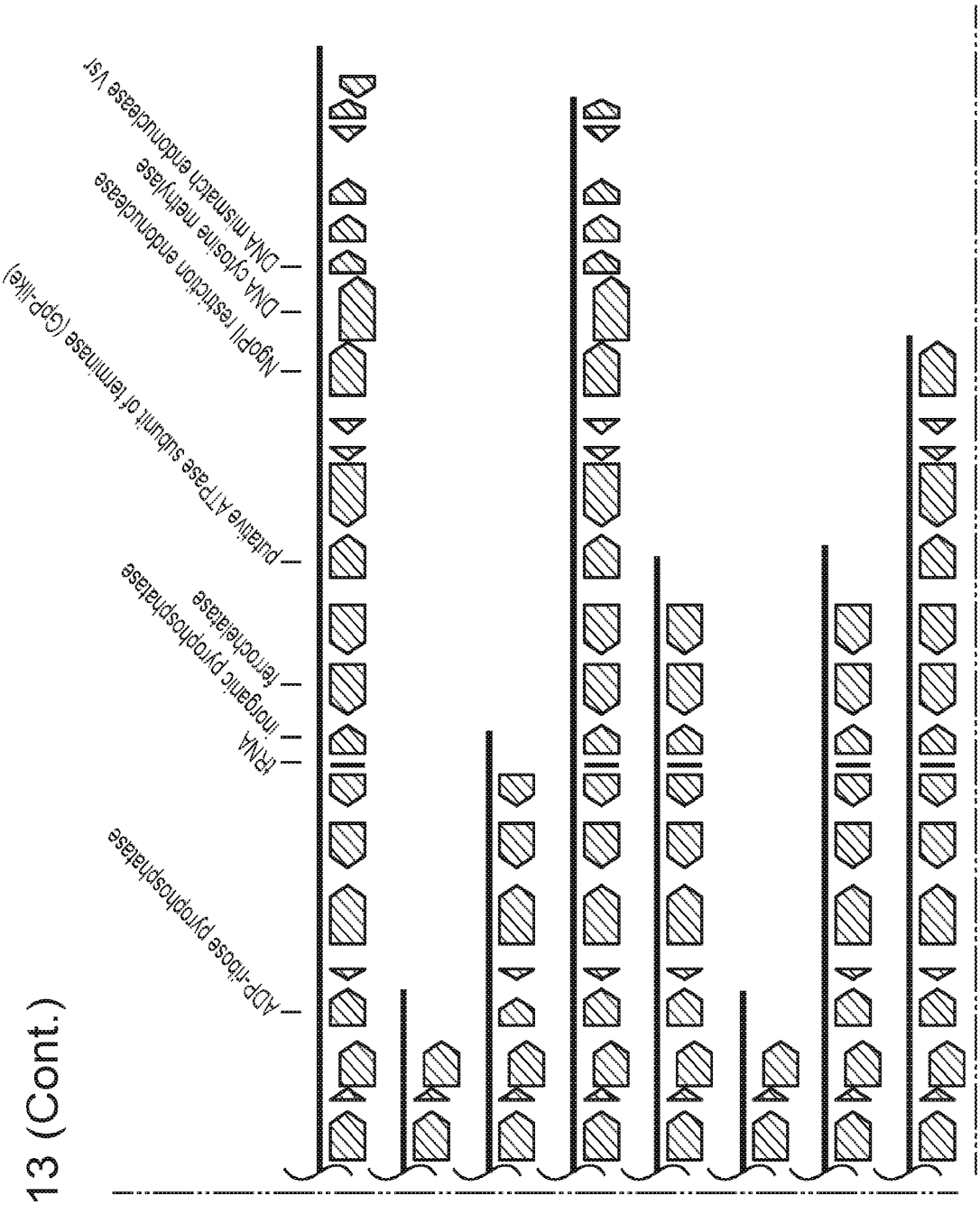

ADP-ribose pyrophosphatase
tRNA
inorganic pyrophosphatase
ferrochelatase
putative ATPase subunit of terminase (GpP-like)
NgoPII restriction endonuclease
DNA cytosine methylase
DNA mismatch endonuclease Vsr

FIG. 13 (Cont.)

FIG. 13 (Cont.)

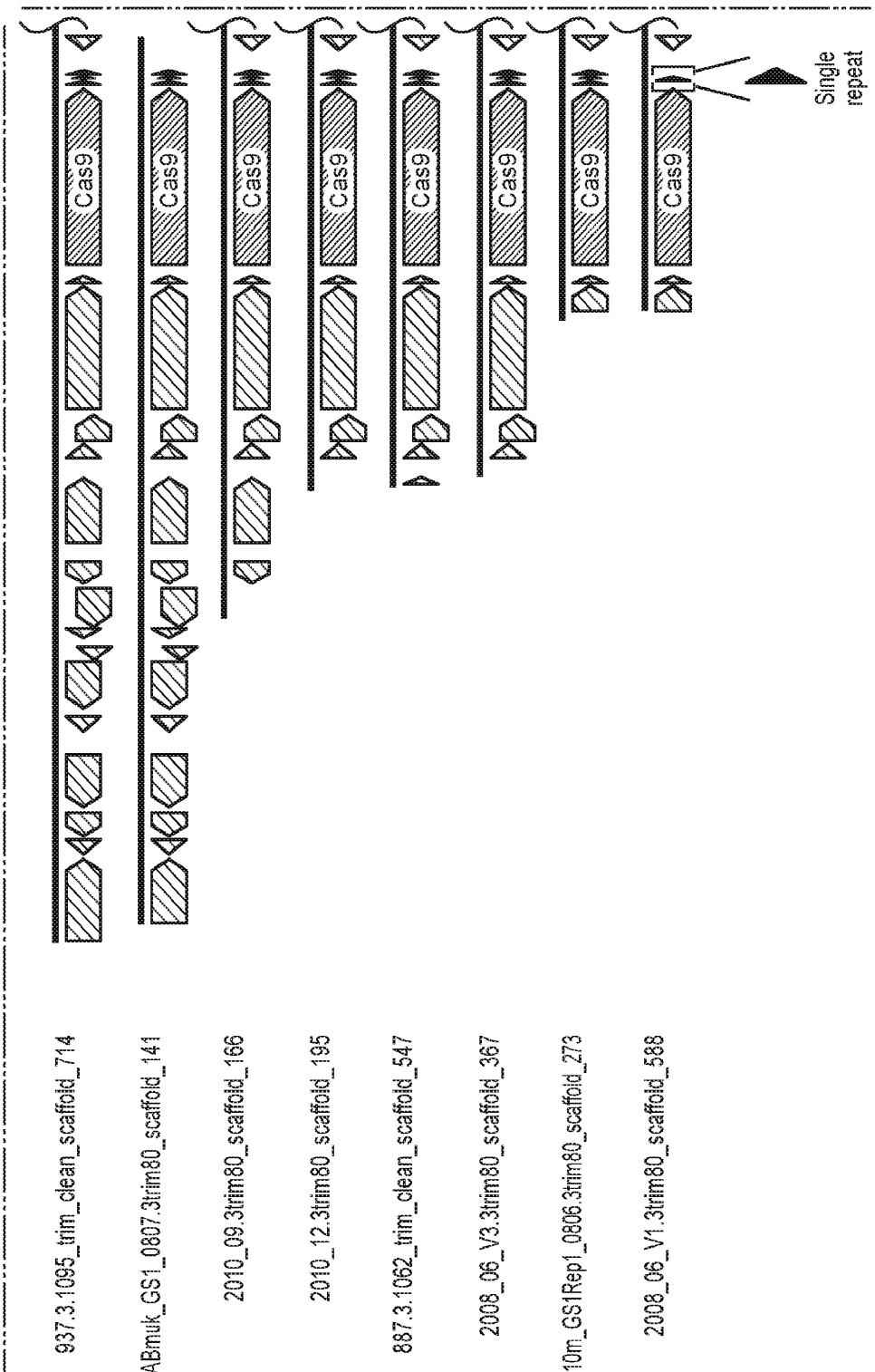

937.3.1095_trim_clean_scaffold_714
P41_ABmuk_GS1_0807.3trim80_scaffold_141
2010_09.3trim80_scaffold_166
2010_12.3trim80_scaffold_195
887.3.1062_trim_clean_scaffold_547
2008_06_V3.3trim80_scaffold_367
P23_C10m_GS1Rep1_0806.3trim80_scaffold_273
2008_06_V1.3trim80_scaffold_588
Cas9
Single repeat

FIG. 13 (Cont.)

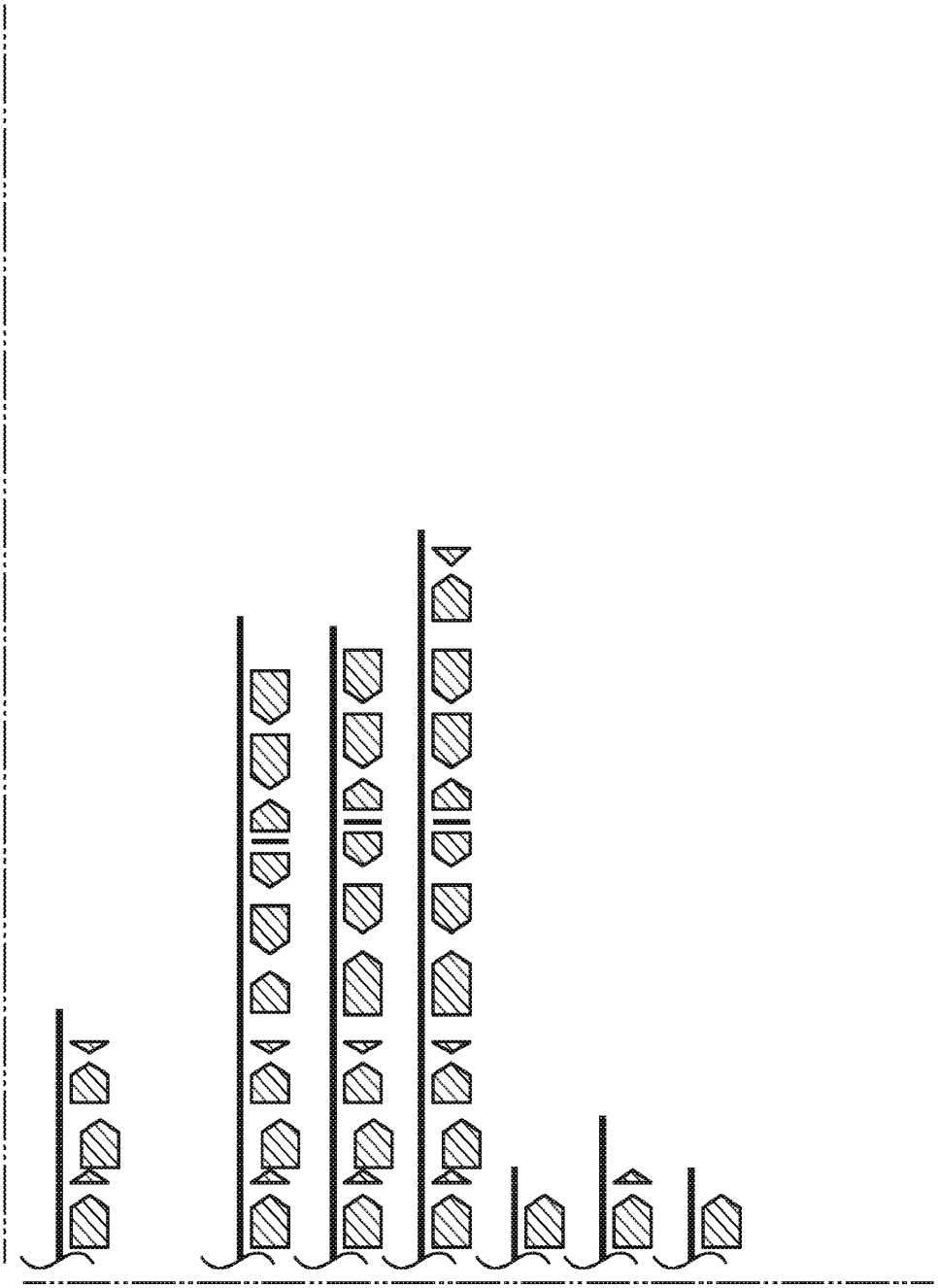


FIG. 14

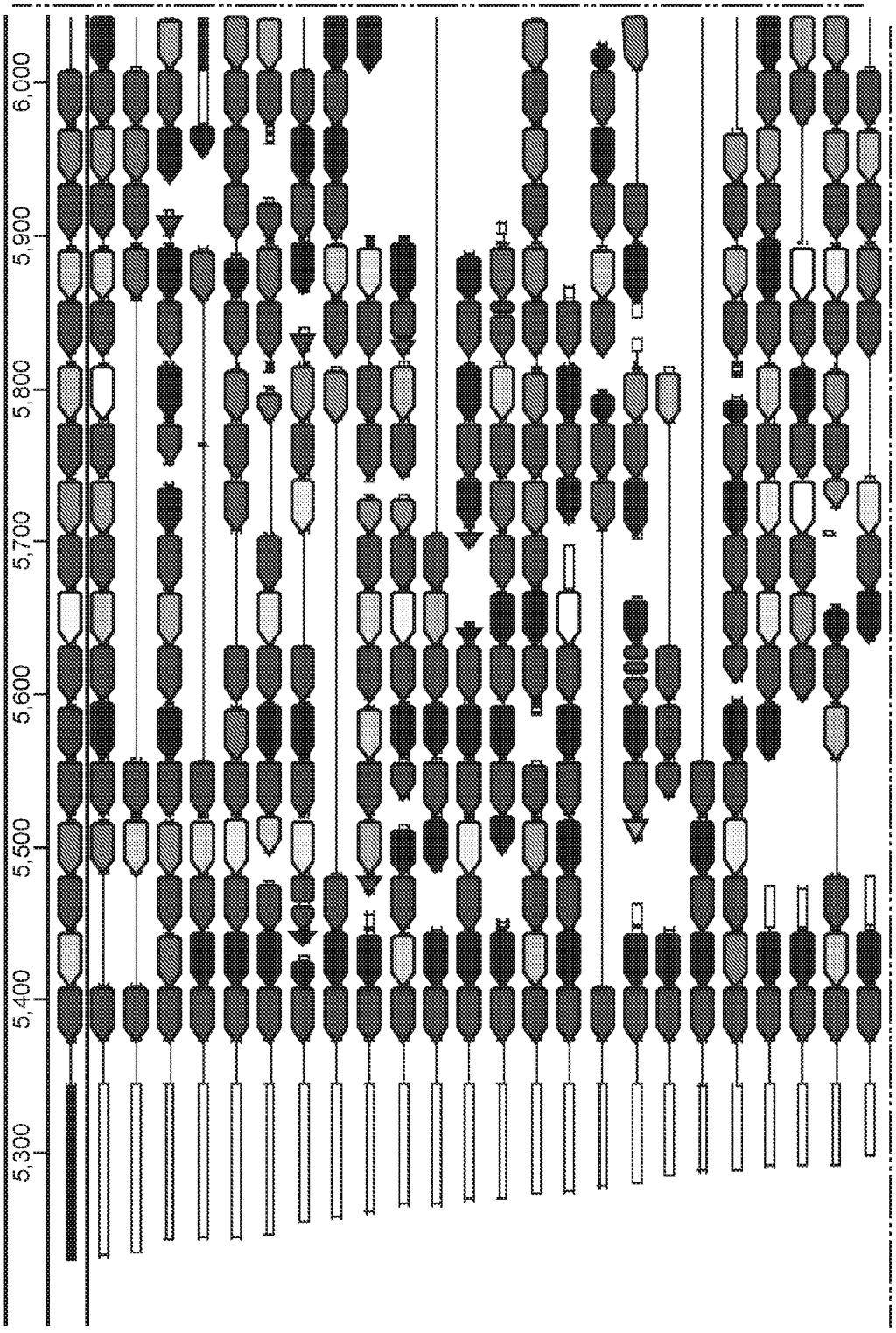


FIG. 14 (Cont.)

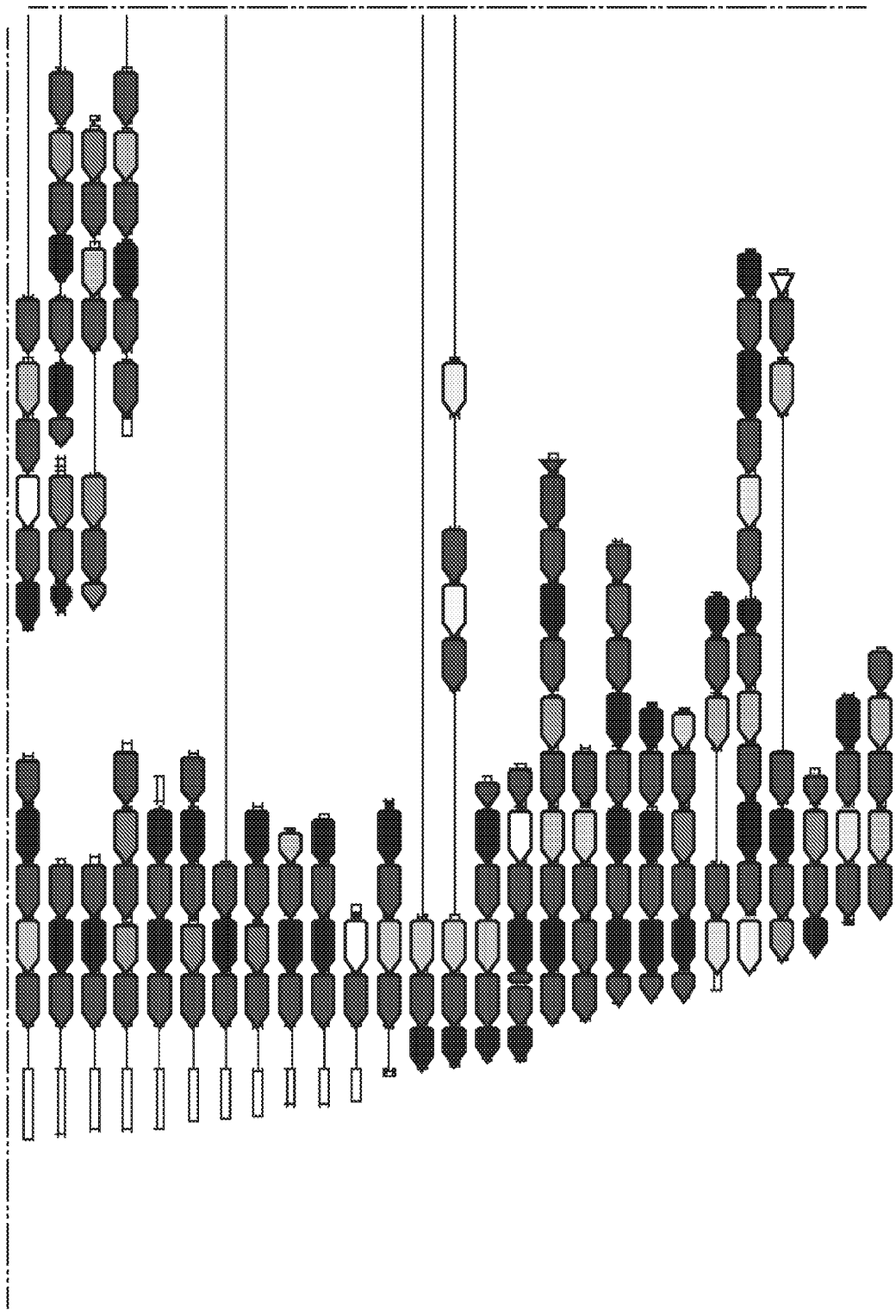


FIG. 14 (Cont.)

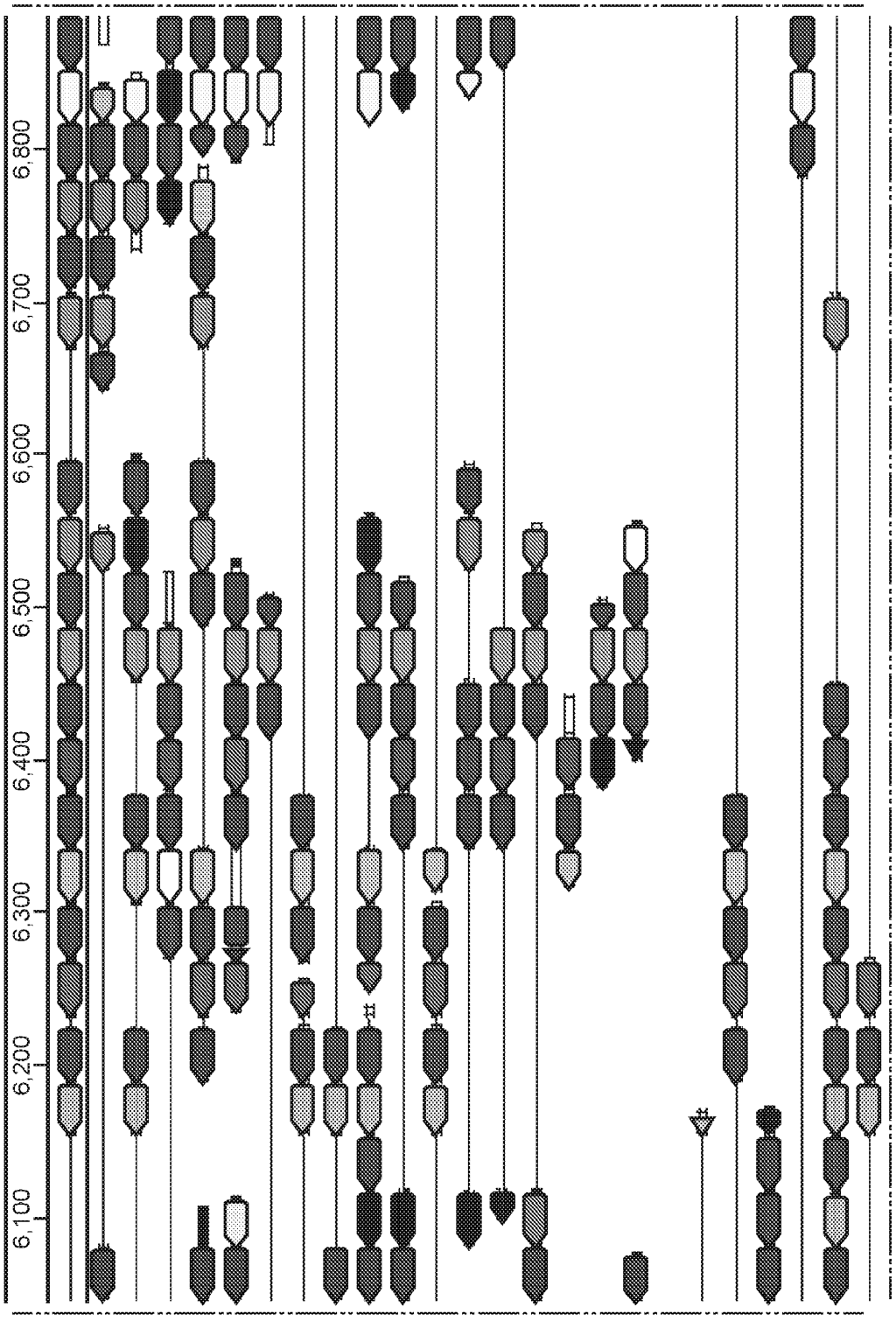


FIG. 14 (Cont.)

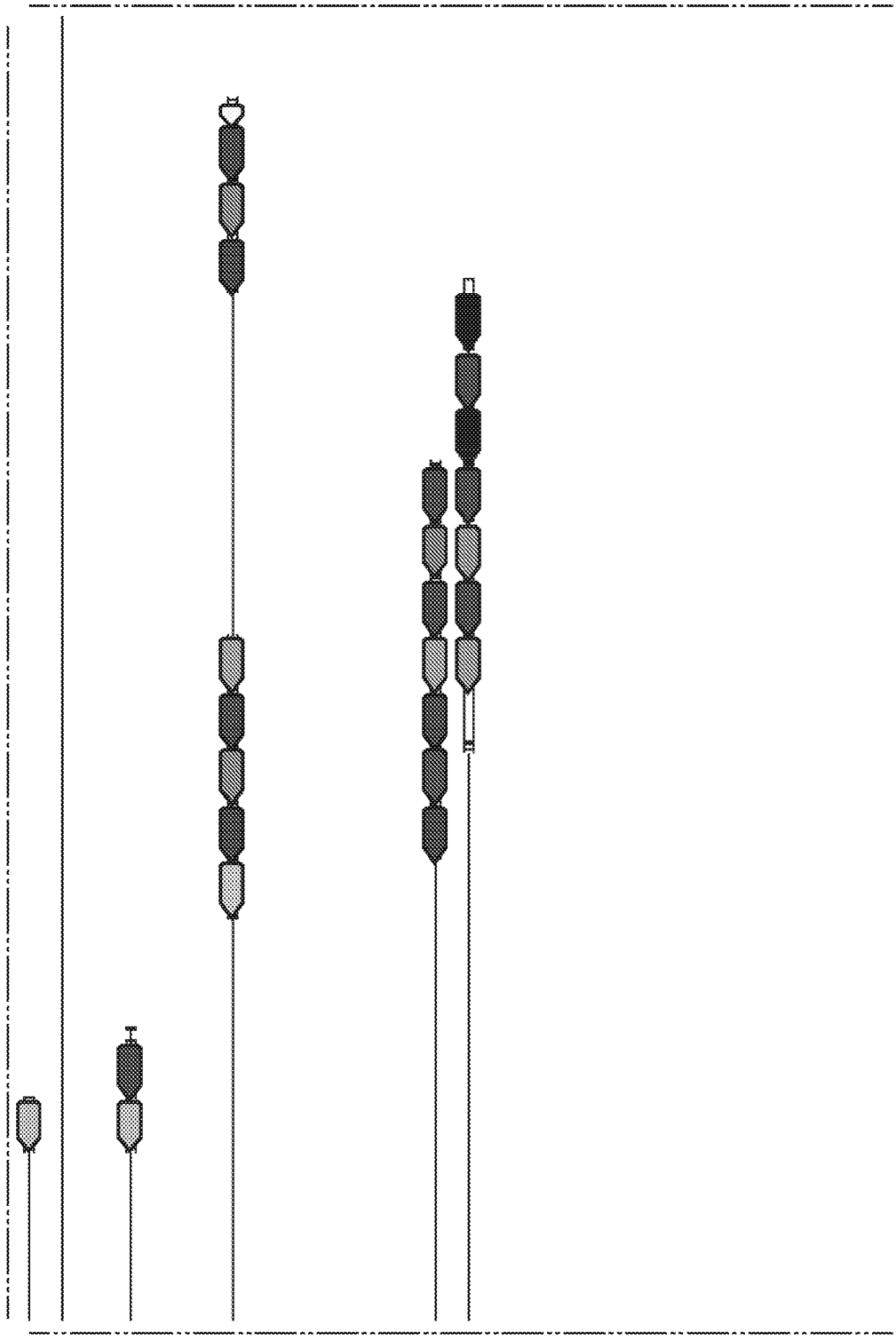


FIG. 14 (Cont.)

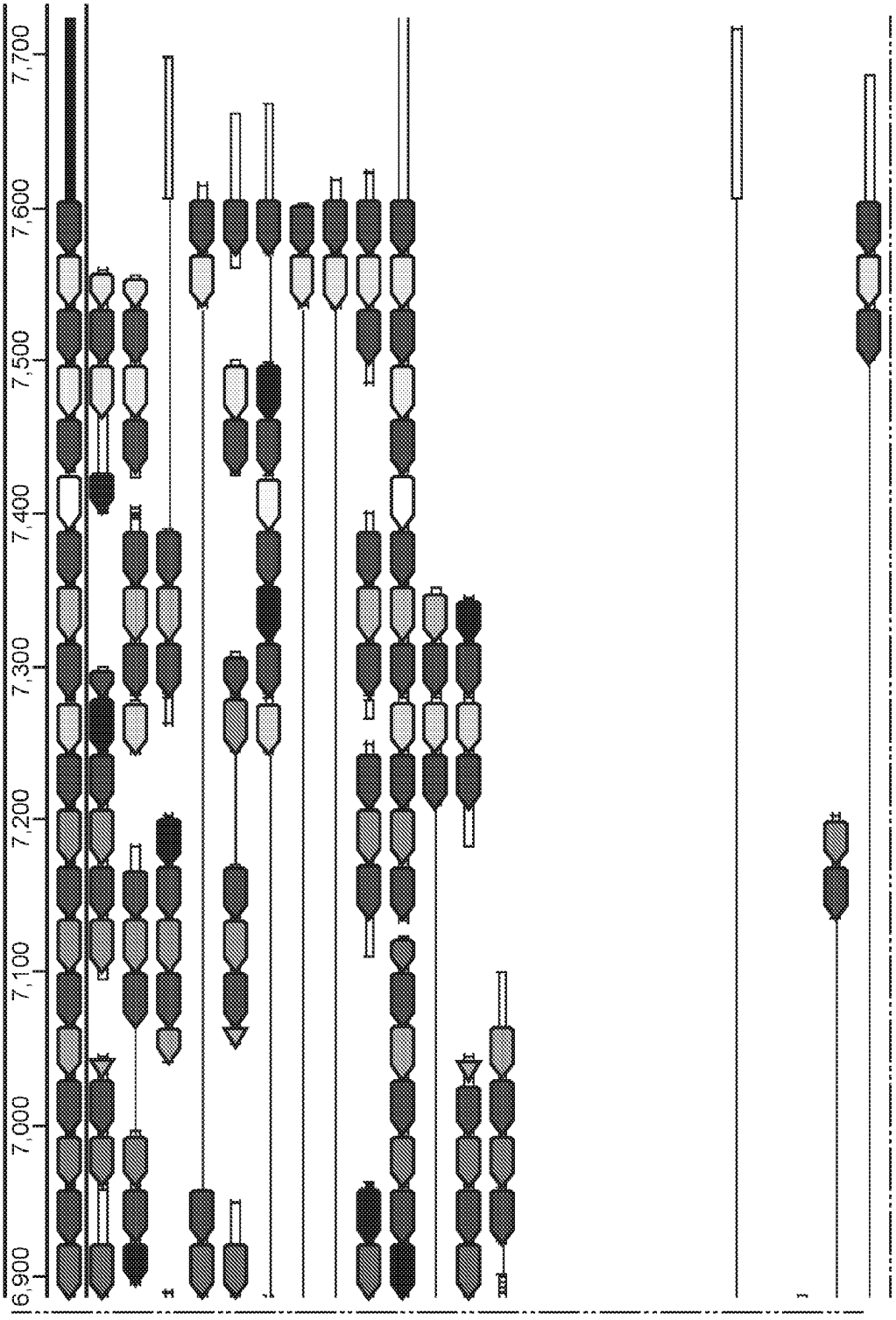


FIG. 14 (Cont.)

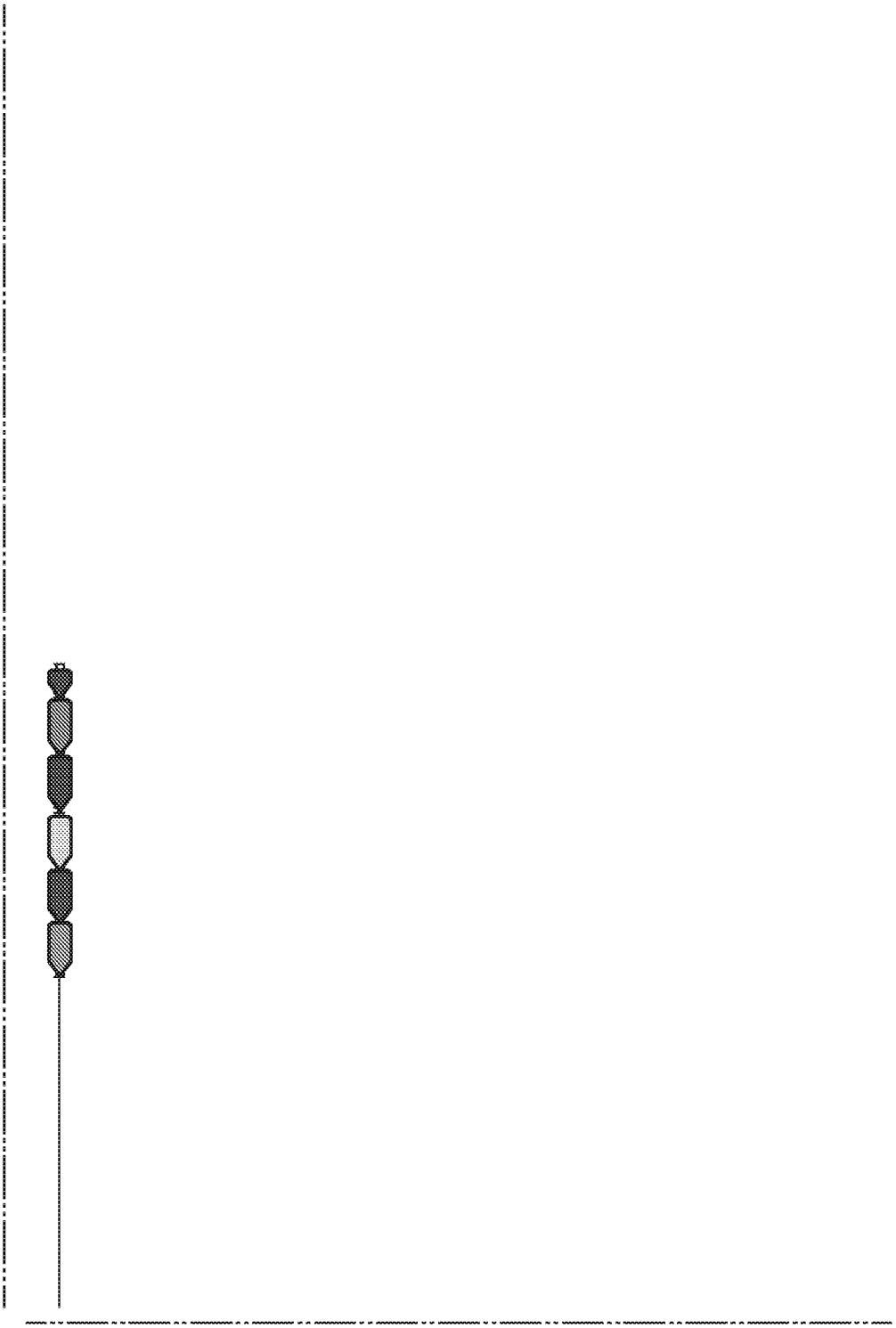


FIG. 15

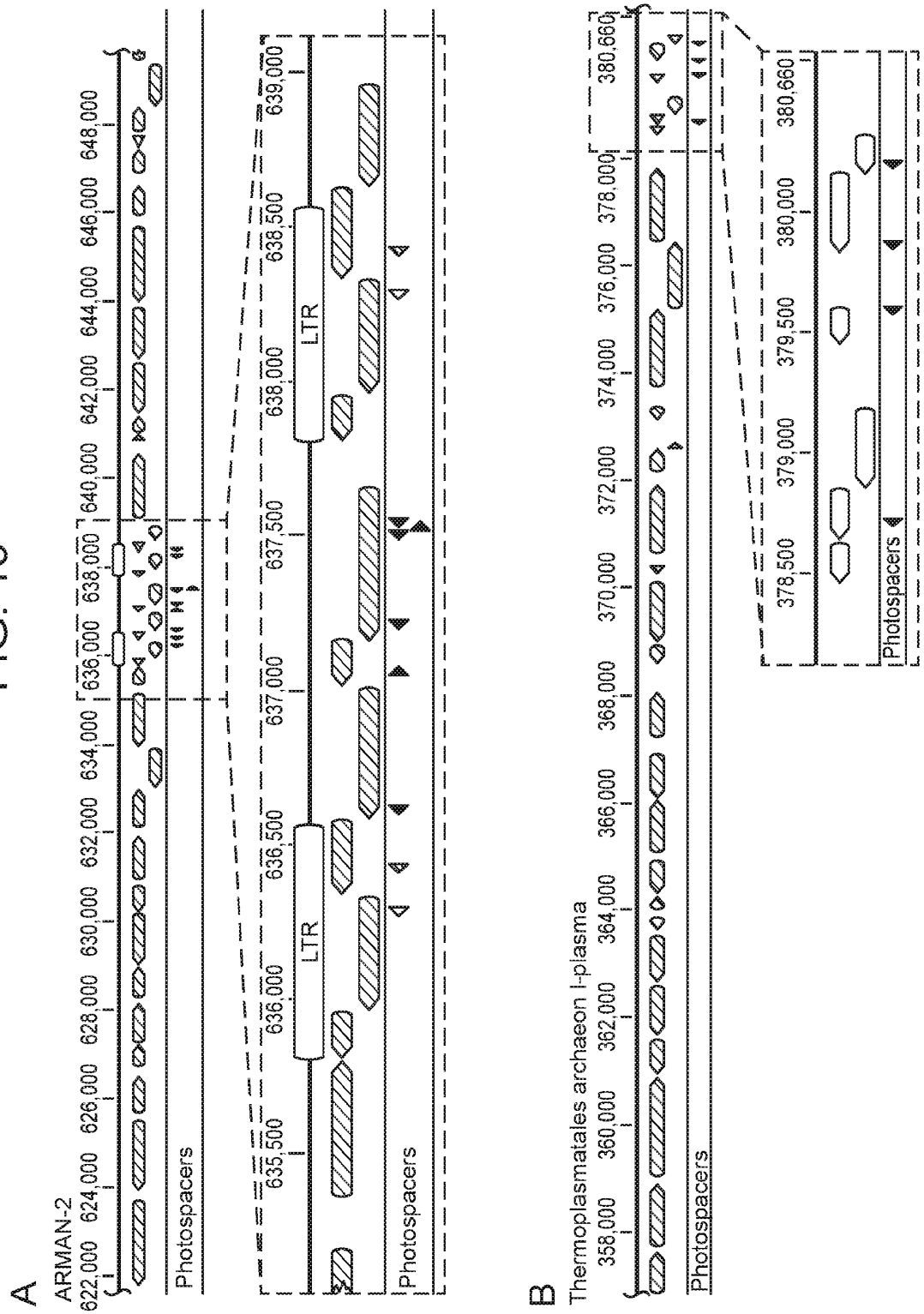

A
ARMAN-2
622,000
624,000
626,000
628,000
630,000
632,000
634,000
636,000
638,000
640,000
642,000
644,000
646,000
648,000
Photospacers
635,500
636,000
636,500
637,000
637,500
638,000
638,500
639,000
LTR
LTR
Photospacers
B
Thermoplasmatales archaeon I-plasma
358,000
360,000
362,000
364,000
366,000
368,000
370,000
372,000
374,000
376,000
378,000
380,660
Photospacers
378,500
379,000
379,500
380,000
380,660
Photospacers

FIG. 16

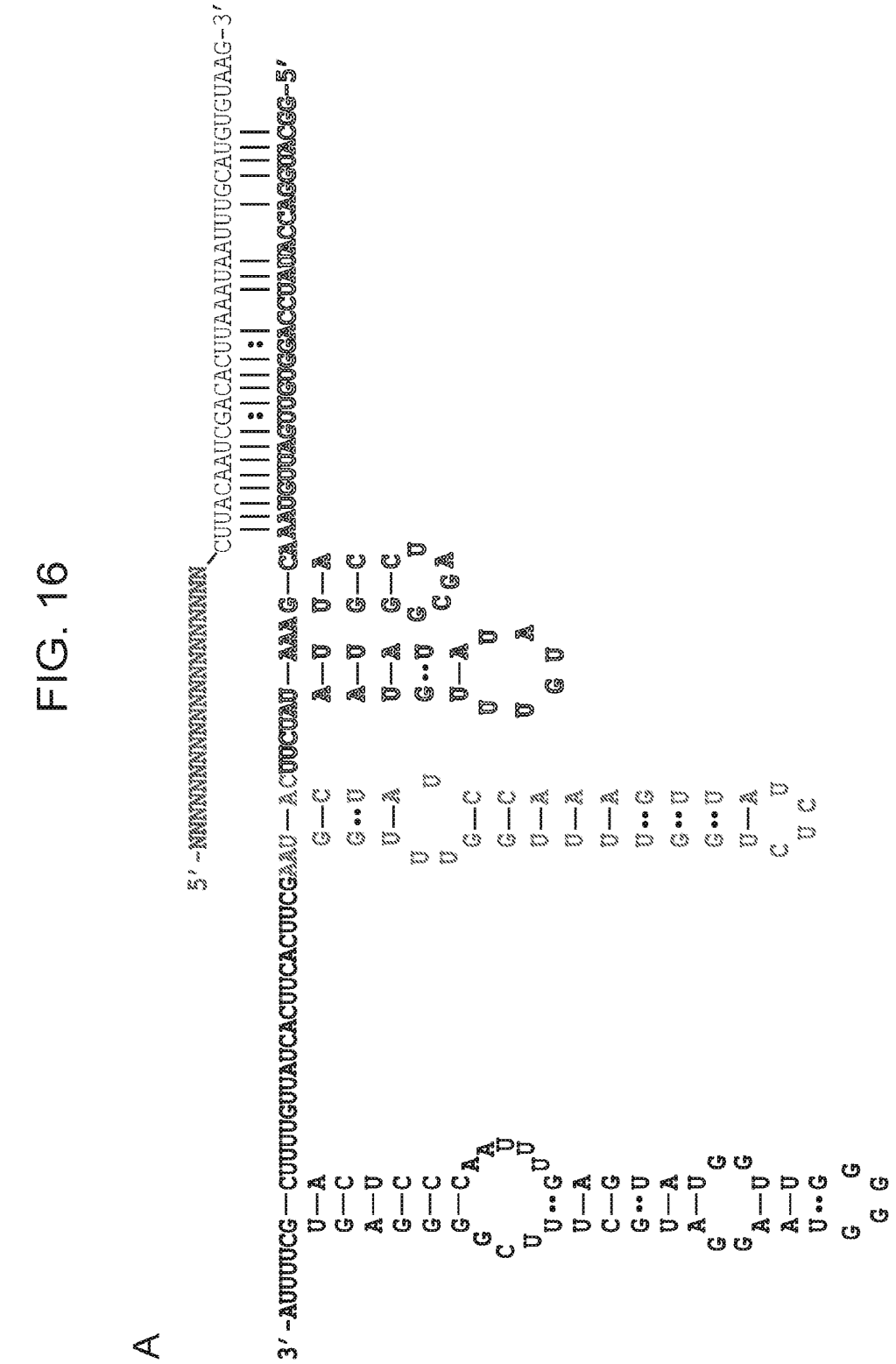

A
5′-NNNNNNNNNNNNNNNNNNNN
CUUACAAUCGACACUUAAAUAAUUUGCAUGUGUAAG-3′
3′-AUUUUCG—CUUUUGUUAUCACUUCACUUCGAAU—ACUUCUAU—AAAG—CAAAUGUUAGUUGCUCGACCUAUUACCAGUUACGG-5′

FIG. 16 (Cont.)

B

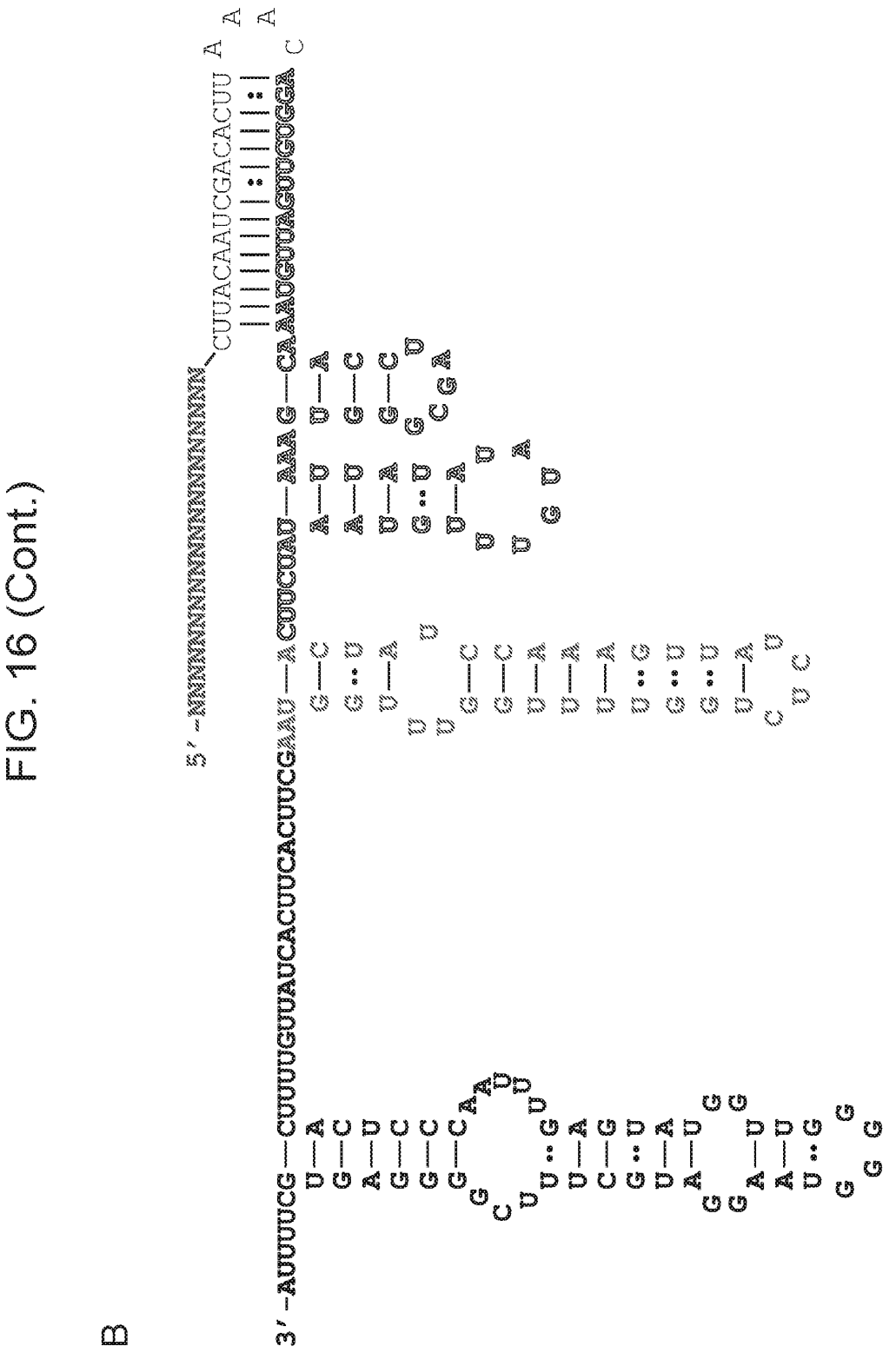
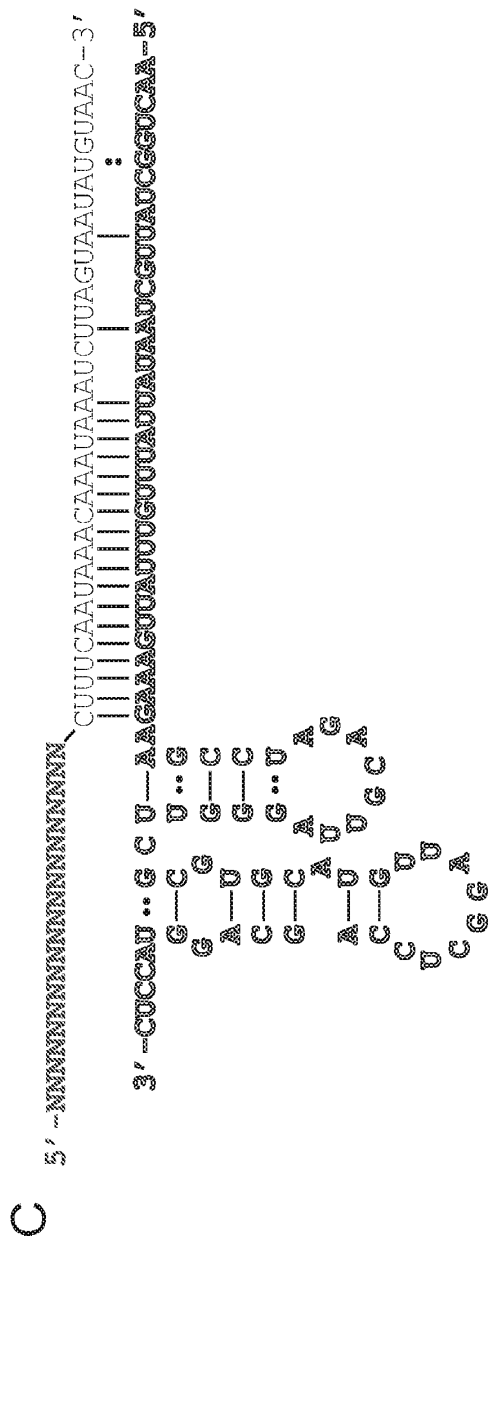

5′ –NNNNNNNNNNNNNNNNNNNN
CUUACAAUCGACACUU A
A
C
3′ –AUUUUCG—CUUUUGUUAUCACUUCACUUCGAAU—A CUUCUAU—AAA G—CA AUGUUAGUUGUGGA
U—A C—C A—U U—A
G—C G·U A—U G—C
A—U U—A U—A G—C
G—C D D G·U G—C D
G—C D D D—A C G A
C A A D D—A
G D D
D·G D—A C—G
D—A C—G G·U
C—G G·U D—A
G·U D—A A—U G
A—U G G
G A—U
A—U
D·G
G G
G

FIG. 16 (Cont.)



Protein	Guide
AR1-Cas9 AR1-dCas9	crRNA sgRNA-69 sgRNA-104 sgRNA-179
AR4-Cas9 AR4-dCas9	crRNA sgRNA-75 sgRNA-122

FIG. 17

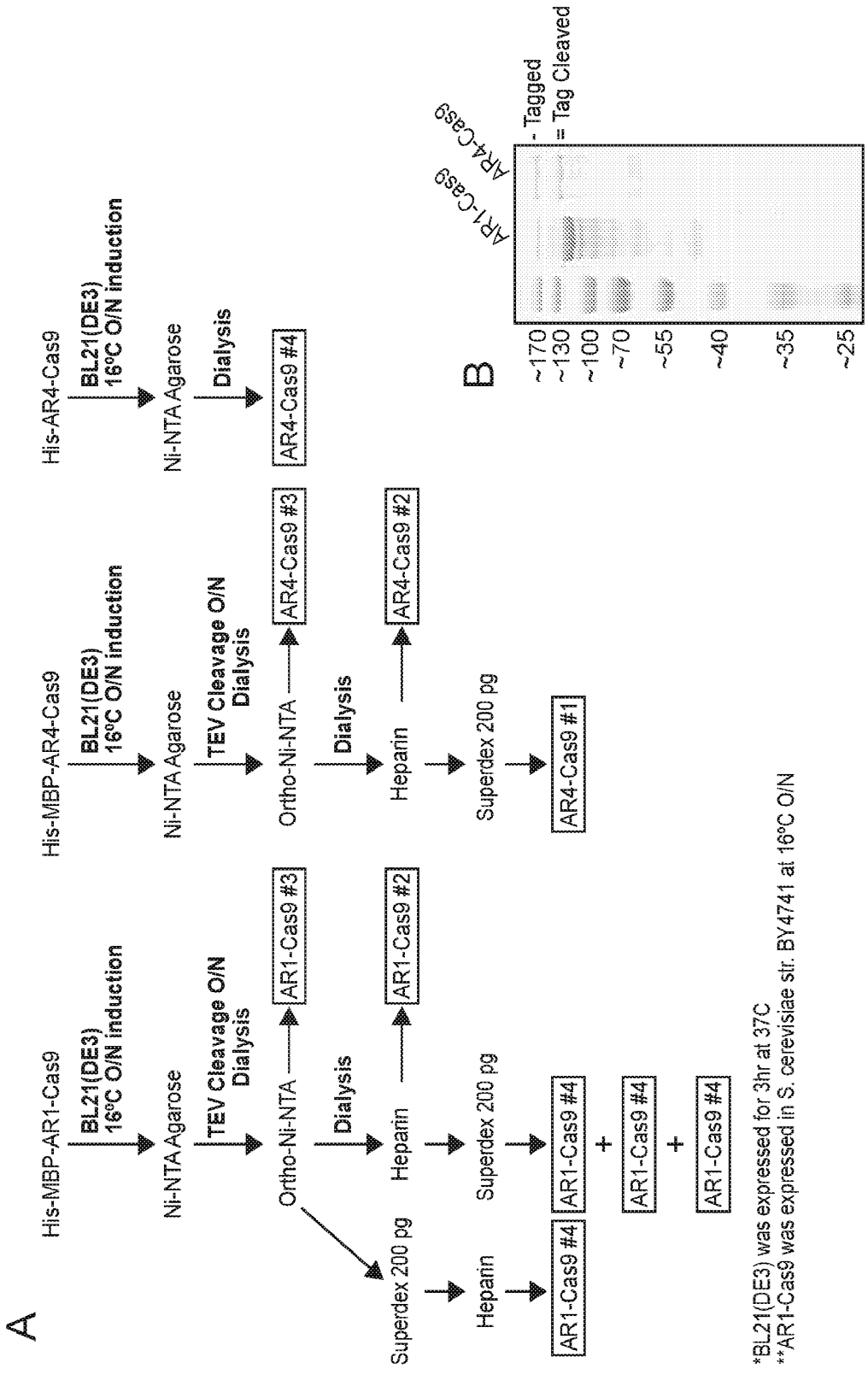

A
His-MBP-AR1-Cas9
BL21(DE3)
16°C O/N induction
Ni-NTA Agarose
TEV Cleavage O/N
Dialysis
Ortho-Ni-NTA
AR1-Cas9 #3
Dialysis
Heparin
AR1-Cas9 #2
Superdex 200 pg
AR1-Cas9 #4
His-MBP-AR4-Cas9
BL21(DE3)
16°C O/N induction
Ni-NTA Agarose
TEV Cleavage O/N
Dialysis
Ortho-Ni-NTA
AR4-Cas9 #3
Dialysis
Heparin
AR4-Cas9 #2
Superdex 200 pg
AR4-Cas9 #1
His-AR4-Cas9
BL21(DE3)
16°C O/N induction
Ni-NTA Agarose
Dialysis
AR4-Cas9 #4
Superdex 200 pg
Heparin
AR1-Cas9 #4
AR1-Cas9 #4
+
AR1-Cas9 #4
+
AR1-Cas9 #4
*BL21(DE3) was expressed for 3hr at 37C
**AR1-Cas9 was expressed in S. cerevisiae str. BY4741 at 16°C O/N
B
AR1-Cas9
AR4-Cas9
- Tagged
= Tag Cleaved
~170
~130
~100
~70
~55
~40
~35
~25

FIG. 18

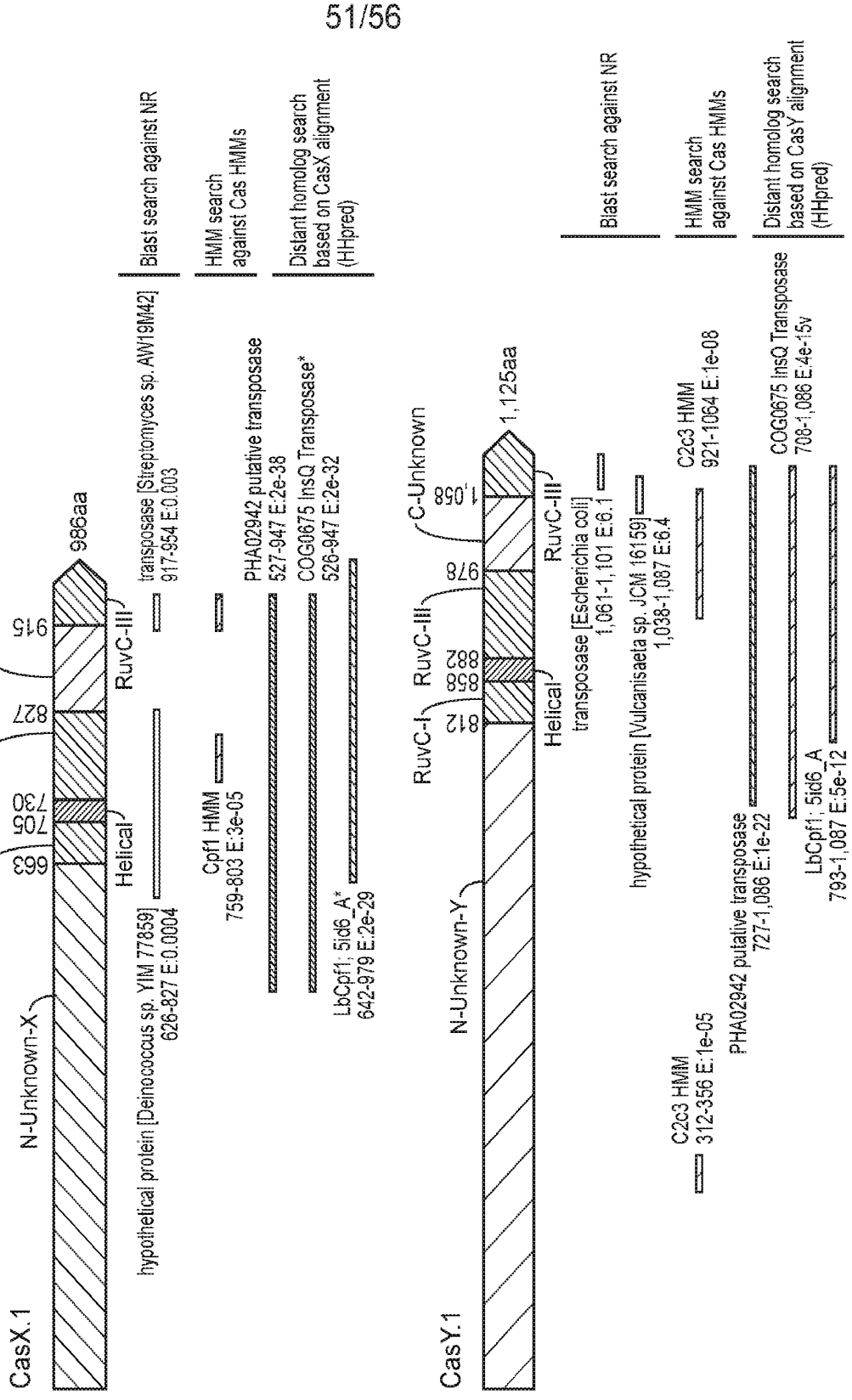

CasX.1
N-Unknown-X
RuvC-I
RuvC-II
RuvC-III
Helical
C-Unknown
663
705
730
827
915
986aa
hypothetical protein [Deinococcus sp. YIM 77859]
626-827 E:0.0004
transposase [Streptomyces sp. AW19M42]
917-954 E:0.003
Cpf1 HMM
759-803 E:3e-05
PHA02942 putative transposase
527-947 E:2e-38
COG0675 InsQ Transposase*
526-947 E:2e-32
LbCpf1; 5id6_A*
642-979 E:2e-29
Blast search against NR
HMM search against Cas HMMs
Distant homolog search based on CasX alignment (HHpred)
CasY.1
N-Unknown-Y
RuvC-I
RuvC-III
Helical
RuvC-III
C-Unknown
812
838
882
978
1,058
1,125aa
transposase [Escherichia coli]
1,061-1,101 E:6.1
hypothetical protein [Vulcanisaeta sp. JCM 16159]
1,038-1,087 E:6.4
C2c3 HMM
312-356 E:1e-05
PHA02942 putative transposase
727-1,086 E:1e-22
C2c3 HMM
921-1064 E:1e-08
COG0675 InsQ Transposase
708-1,086 E:4e-15v
LbCpf1; 5id6_A
793-1,087 E:5e-12
Blast search against NR
HMM search against Cas HMMs
Distant homolog search based on CasY alignment (HHpred)

FIG. 19

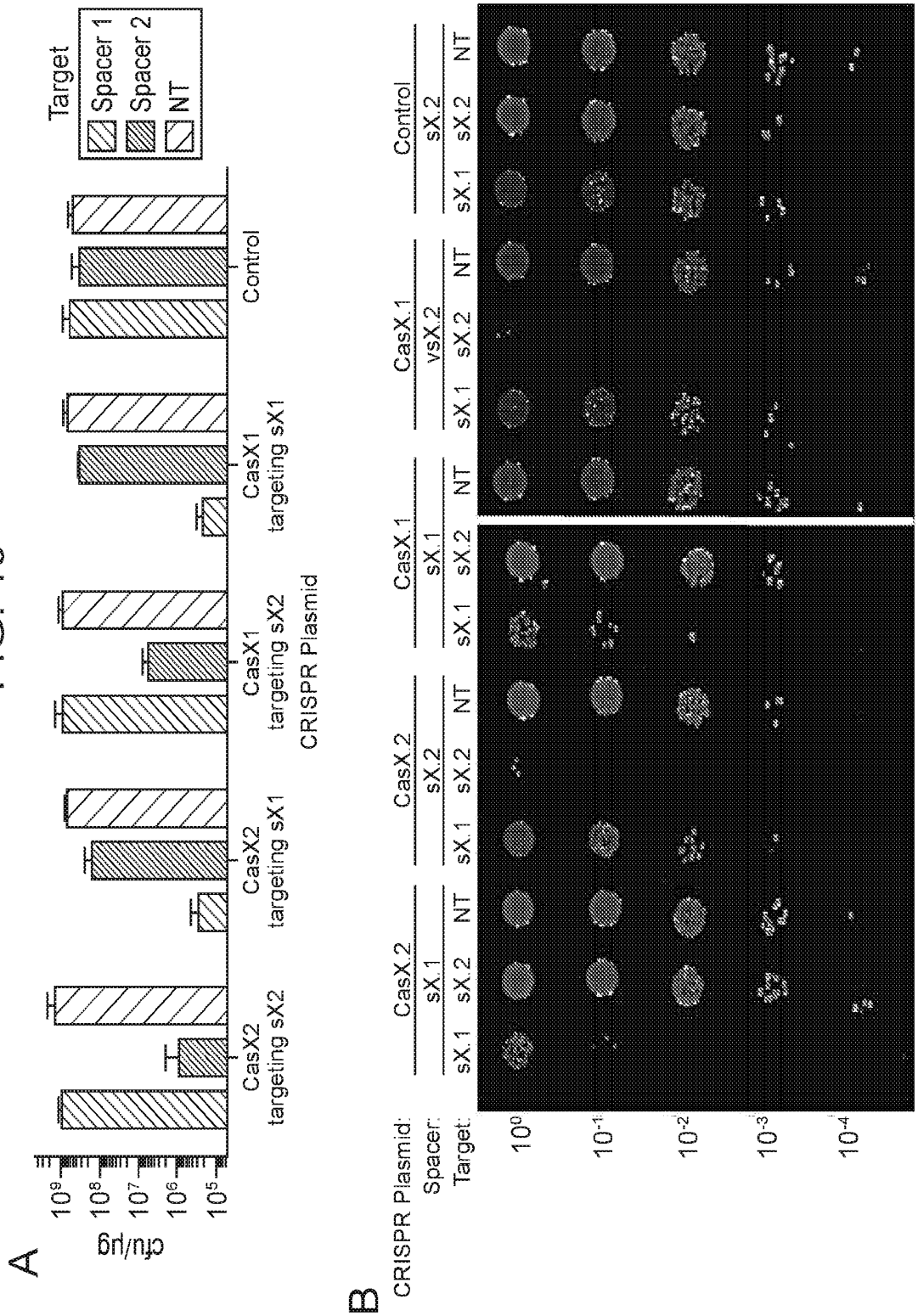

A
cfu/µg
10^9
10^8
10^7
10^6
10^5
Target
Spacer 1
Spacer 2
NT
CasX2 targeting sX2
CasX2 targeting sX1
CasX1 targeting sX2
CasX1 targeting sX1
Control
CRISPR Plasmid
B
CRISPR Plasmid:
Spacer:
Target:
CasX.2
sX.1
sX.2
NT
sX.1
sX.2
NT
CasX.1
sX.1
sX.2
NT
CasX.1
vsX.2
sX.1
sX.2
NT
Control
sX.2
sX.1
sX.2
NT
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}

FIG. 19 (Cont.)

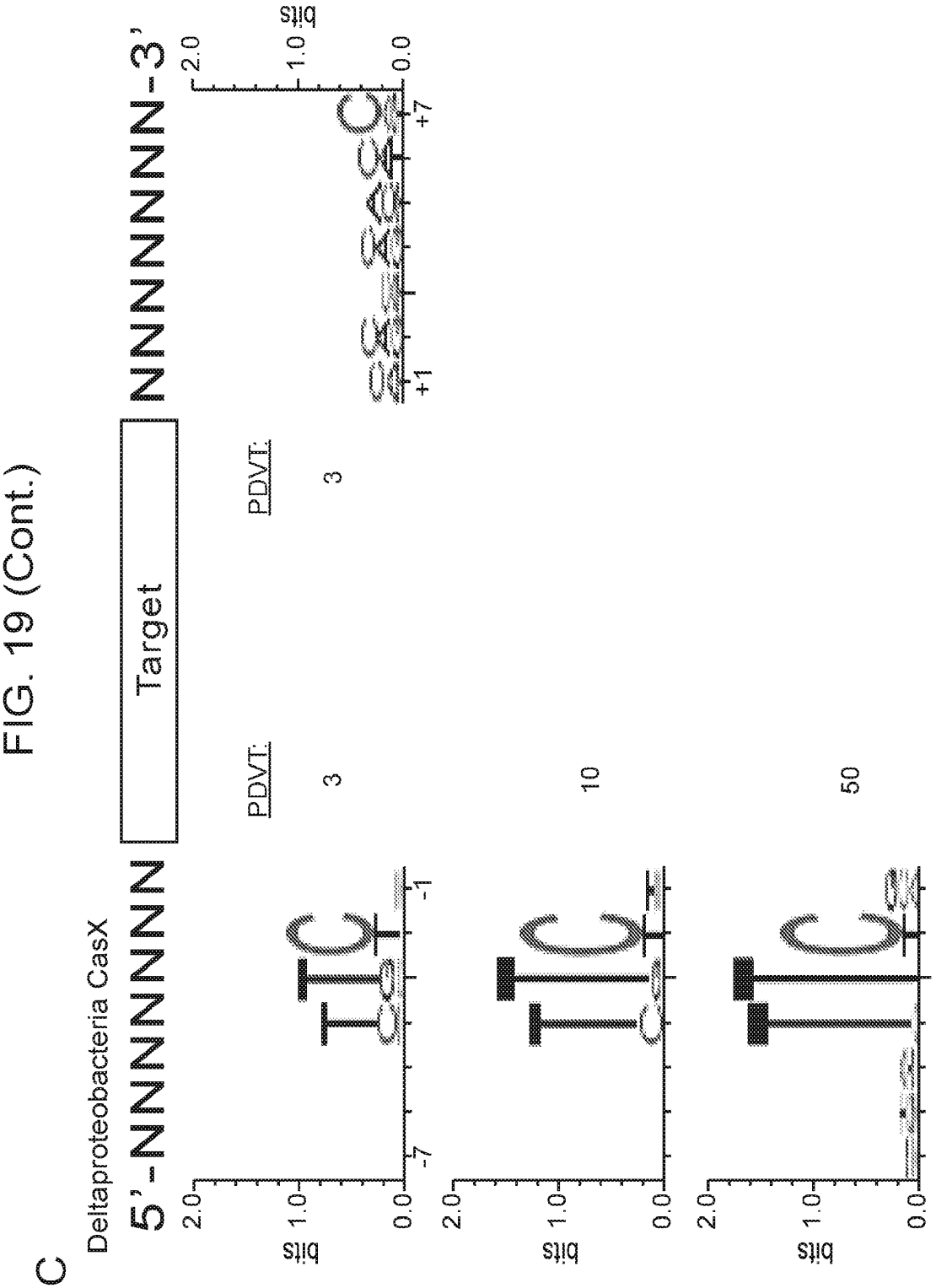

C
Deltaproteobacteria CasX
5'-NNNNNNN
Target
NNNNNNN-3'
PDVT:
3
10
50
bits
2.0
1.0
0.0
-7
-1
+1
+7

FIG. 19 (Cont.)

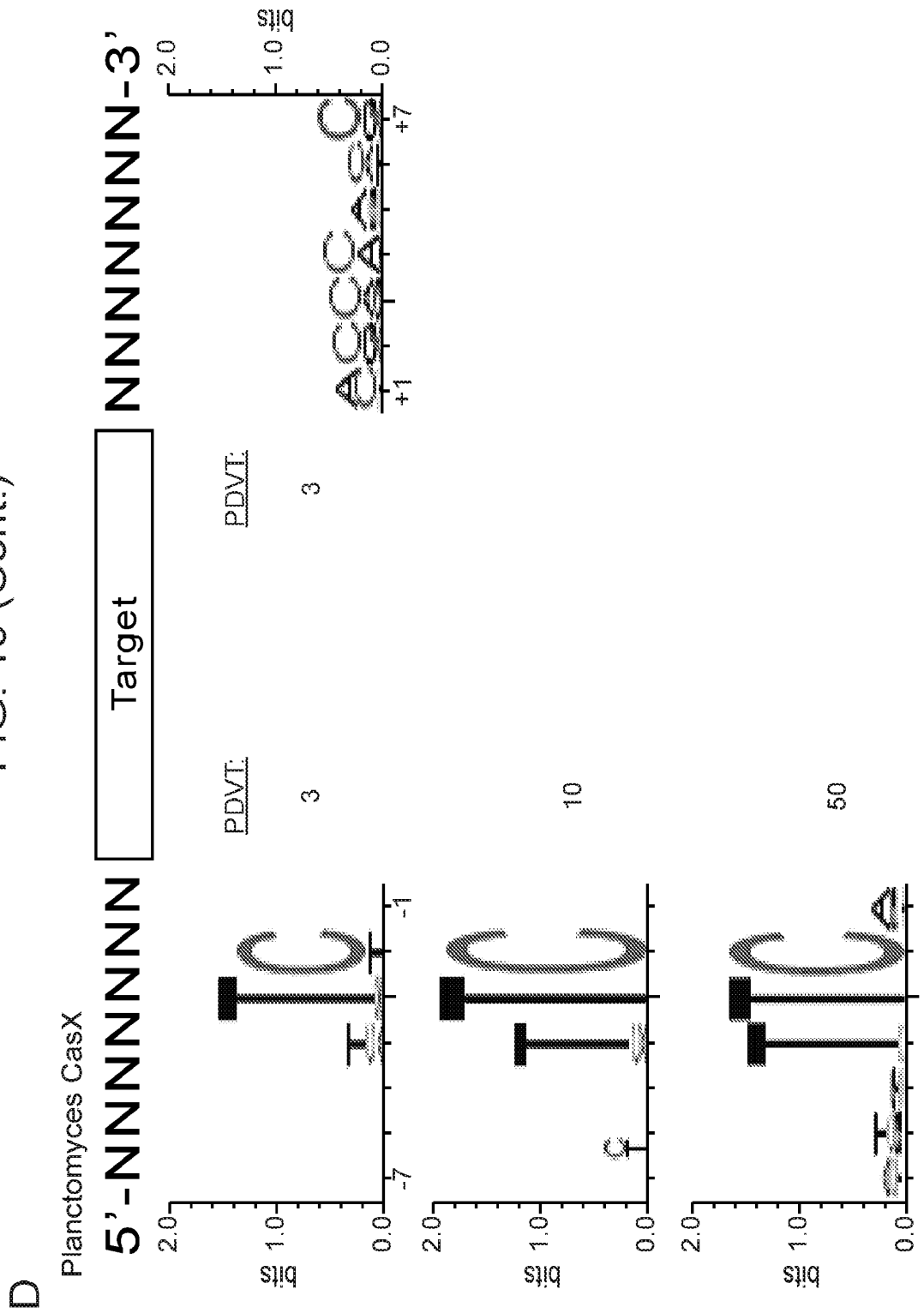

D
Planctomyces CasX
5'-NNNNNN
Target
NNNNNN-3'
PDVT
3
10
50
bits
2.0
1.0
0.0
-7
-1
+1
+7

FIG. 20

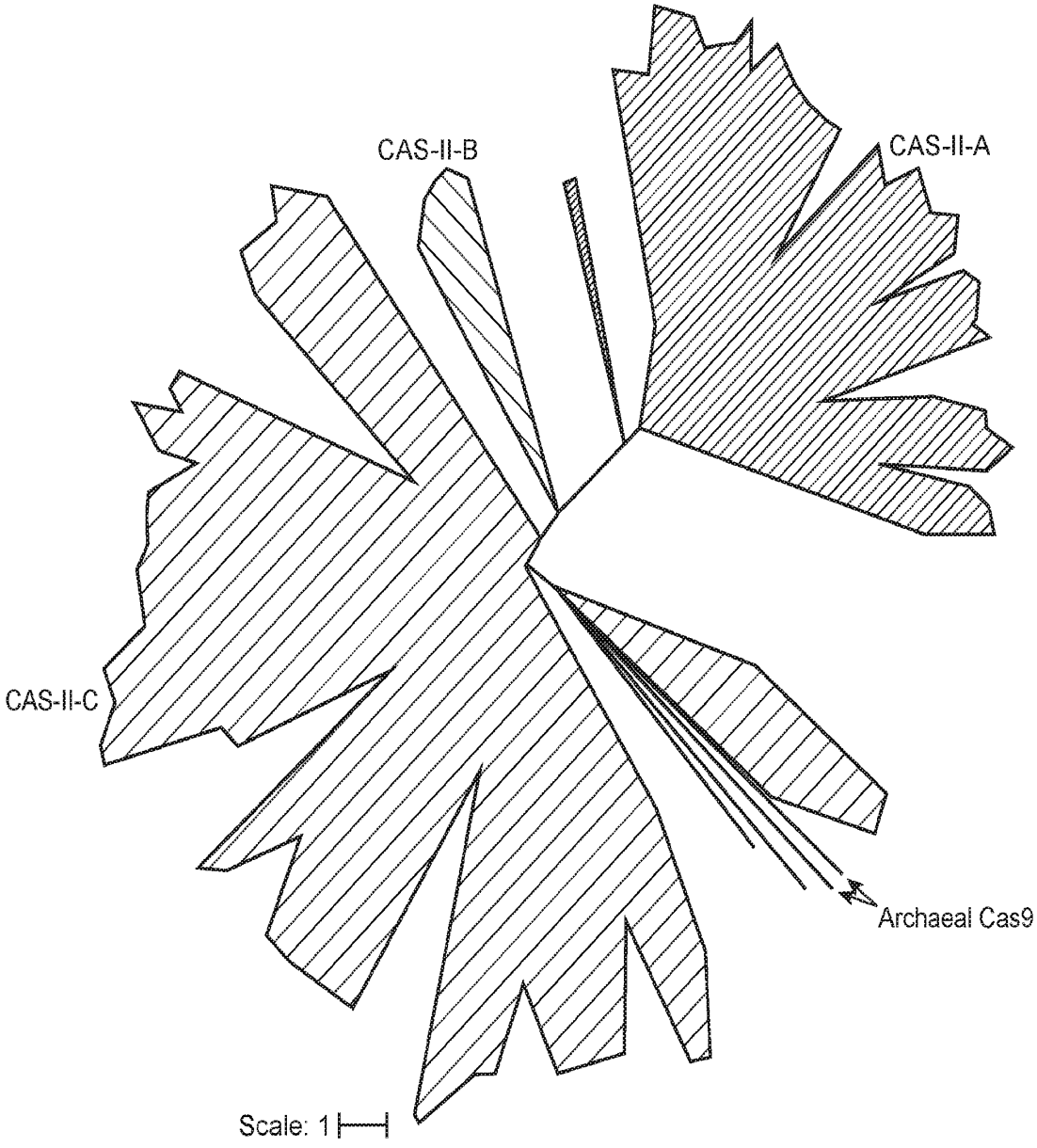


FIG. 21

In vitro cleavage conditions assayed for Cas9 from ARMAN-1

Protein Purification	Buffer	Salt (mM)	Metal	Guide	Target	Temperature
AR1-Cas9 #1	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	crRNA cr:69 cr:69 cr:69	dsDNA ssDNA DNA Bubble ssRNA dsDNA	37
AR1-Cas9 #1	Tris ph 7.5	100-500	Mg^{2+}	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #1	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	cr:69 cr:104 cr:179	dsDNA	30-48
AR1-Cas9 #1	MOPS: pH 6 pH 6.5 pH 7.0 pH 7.5	300	Mg^{2+}	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #1	Citrate: pH 5 pH 5.5 pH 6	300	Mg^{2+}	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #1	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	cr:69 cr:104 cr:179	plasmid	37-50
AR1-Cas9 #2	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #3	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #4	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #5	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	cr:69 cr:104 cr:179	dsDNA	37
AR1-Cas9 #6	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	cr:69 cr:104 cr:179	ssDNA dsDNA	37
AR4-Cas9 #1	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	sgRNA-122	dsDNA	37
AR4-Cas9 #2	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	sgRNA-122	dsDNA	37
AR4-Cas9 #3	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	sgRNA-122	dsDNA	37
AR4-Cas9 #4	Tris ph 7.5	300	Mg^{2+} Mn^{2+} Zn^{2+}	sgRNA-122	dsDNA	37