Supplementary Material for:

Targeted metagenomics: finding rare tryptophan dimer natural products in the environment.

Fang-Yuan Chang, Melinda A. Ternei, Paula Y. Calle, and Sean F. Brady*
Laboratory of Genetically Encoded Small Molecules, The Rockefeller University, Howard Hughes Medical Institute, 1230 York Avenue, New York, NY 10065.

RECEIVED DATE (automatically inserted by publisher); sbrady@rockefeller.edu

Supplementary Material Table of Contents

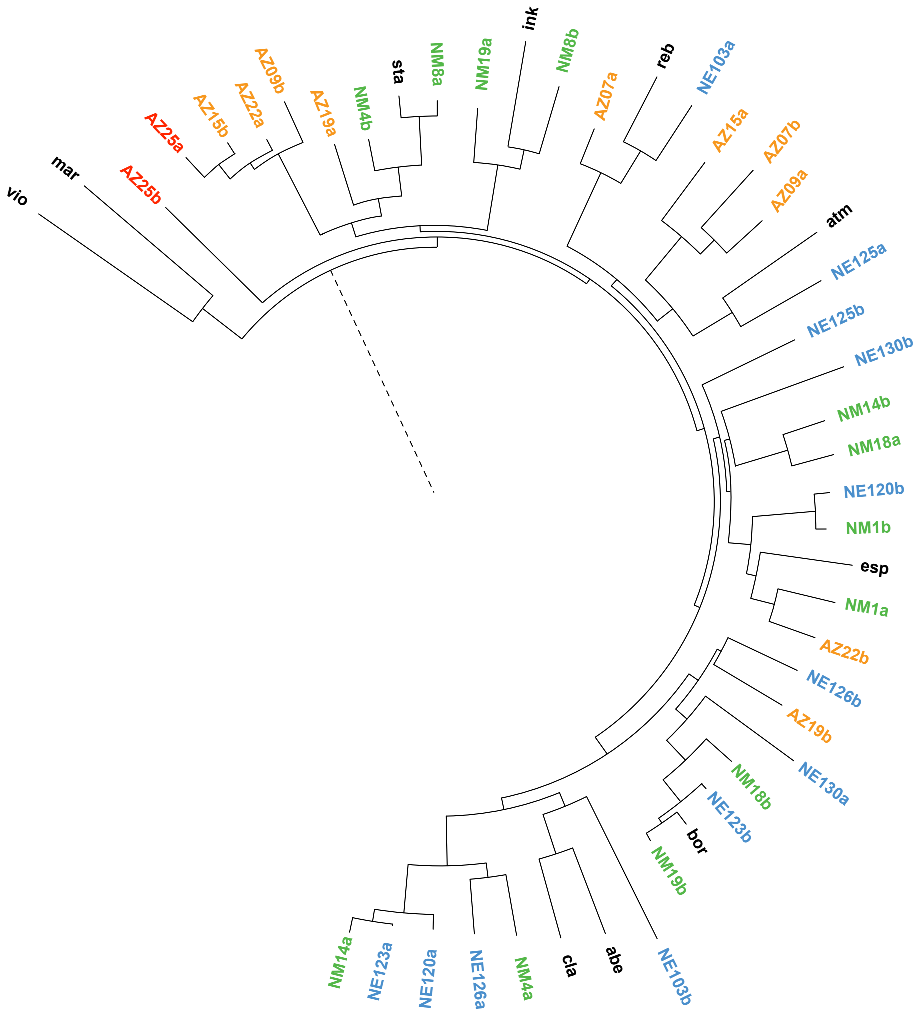
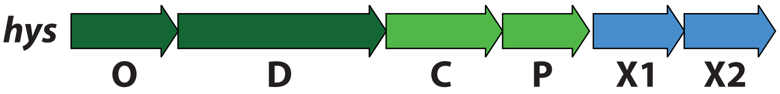
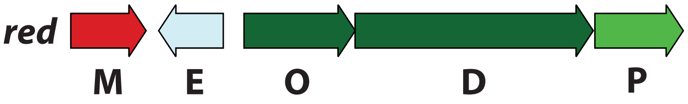


Figure S1. CLUSTALW-based phylogenetic tree of TD environmental sequence tags. CPAS gene sequence tags from Arizona (AZ; orange), New Mexico (NM; green), and New England (NE; blue) soil samples are aligned with the knowns (black). The two tags (AZ25a, AZ25b) investigated in this study are highlighted in red.



N	Gene	Size (aa)	Homolog	ID/SM (%)	Origin	Accession Number
1	orf-25	645	sensor kinase	48/62	Rhodococcus wratislaviensis NBRC 100605	GAF48609.1
2	orf-24	225	LuxR family transcriptional regulator	77/88	Rhodococcus opacus	WP_025433290.1
3	orf-23	322	dehydrogenase	78/85	Pseudonocardia acaciae	WP_028921473.1
4	orf-22	134	transcriptional regulator	71/84	Pseudonocardia dioxanivorans	WP_013677698.1
5	orf-21	286	RNA ligase	78/84	Streptomyces viridochromogenes	WP_003995917.1
6	orf-20	140	archease	59/66	Verrucosispora maris	WP_013733957.1
7	orf-19	173	transporter	35/52	Nitrolancea hollandica	WP_008477262.1
8	orf-18	372	MFS transporter	65/79	Myxococcus xanthus	WP_020478826.1
9	orf-17	166	endonuclease	71/80	Pseudonocardia asaccharolytica	WP_028931995.1
10	orf-16	191	alcohol dehydrogenase	32/41	Mycobacterium cosmeticum	CDO08914.1
11	orf-15	198	reductase	74/81	Actinokineospora enzanensis	WP_018687140.1
12	orf-14	402	FAD-dependent oxidoreductase	72/81	Amycolatopsis mediterranei	WP_013226545.1
13	orf-13	119	hypothetical protein	60/75	Nocardia brasiliensis NBRC 14402	GAJ84167.1
14	orf-12	215	hypothetical protein	54/63	Micromonospora sp. ATCC 39149	WP_007071715.1
15	orf-11	123	NADH-quinone oxidoreductase	77/87	Pseudonocardia dioxanivorans	WP_013678831.1
16	orf-10	331	hypothetical protein	40/51	Saccharomonospora xinjiangensis	WP_006240734.1
17	orf-9	249	NADH dehydrogenase	79/86	Pseudonocardia dioxanivorans	WP_013678829.1
18	orf-8	188	NADH-ubiquinone oxidoreductase	71/78	Streptomyces afghaniensis	WP_020275586.1
19	orf-7	100	NADH-ubiquinone oxidoreductase	74/90	Actinomadura madurae	WP_021596654.1
20	orf-6	610	NADH dehydrogenase	64/72	Streptomyces sp. NRRL WC-3641	WP_031120050.1
21	orf-5	502	oxidoreductase	73/81	Streptomyces sp. NRRL F-2305	WP_030835849.1
22	orf-4	476	NADH-quinone oxidoreductase	60/70	Streptomyces sp. NRRL S-37	WP_030871967.1
23	orf-3	223	none			
24	orf-2	287	RNA polymerase sigma factor	56/74	Kutzneria albida	WP_025359630.1
25	orf-1	260	methyltransferase	37/49	Microlunatus phosphovorus	WP_013864415.1
26	O	490	indole-3-pyruvic acid imine synthase StaO	51/64	Streptomyces sp. NRRL WC-3719	WP_031128187.1
27	D	954	chromopyrrolic acid synthase StaD	52/62	uncultured bacterium	AHE14639.1
28	C	533	FAD-binding monooxygenase StaC	67/77	Streptomyces sp. TP-A0274	BAF47693.1
29	P	397	cytochrome P450 StaP	57/68	Actinomadura melliaura	ABC02792.1
30	X1	417	cytochrome P450 hydroxylase	46/61	uncultured bacterium	AHE14835.1
31	X2	417	cytochrome P450 hydroxylase	47/64	uncultured bacterium	AHE14835.1
32	orf+1	140	cyclase	50/63	Frankia sp. QA3	WP_009739180.1
33	orf+2	175	cyclase	58/68	Streptosporangium amethystogenes	WP_030923309.1
34	orf+3	379	alcohol dehydrogenase	76/86	Streptosporangium amethystogenes	WP_030923305.1
35	orf+4	243	cytochrome C oxidase	77/91	Actinoalloteichus cyanogriseus	WP_030104074.1
36	orf+5	364	membrane protein	30/50	Xenococcus sp. PCC 7305	WP_006512167.1
37	orf+6	440	dehydratase	42/55	Oceanicola granulosus	WP_007253650.1
38	orf+7	266	dehydratase	34/55	Mesorhizobium ciceri	WP_024505168.1

Figure S2. Annotation of the *hys* gene cluster. Tryptophan dimer gene cluster found in clone P204 (GenBank accession: KP274855). The annotation is based on BLASTP (NCBI) homology search of translated genes predicted by FGENESB.



N	Gene	Size (aa)	Homolog	ID/SM (%)	Origin	Accession Number
1	orf-2	110	flavodoxin	75/85	Streptomyces filamentosus	WP_006129488.1
2	orf-1	107	multidrug transporter	71/84	Amycolatopsis nigrescens	WP_020667160.1
3	M	344	methyltransferase	64/75	Streptomyces mobaraensis	WP_004943262.1
4	E	295	imine reductase (6-phosphogluconate dehydrogenase)	66/74	Streptomyces mobaraensis	WP_004937742.1
5	O	509	indole-3-pyruvic acid imine synthase StaO	66/73	Streptomyces mobaraensis	WP_004942121.1
6	D	1091	chromopyrrolic acid synthase StaD	57/70	uncultured bacterium	AHE14826.1
7	P	388	cytochrome P450 StaP	65/73	Streptomyces mobaraensis	WP_004942127.1
8	T	399	ABC transporter	57/67	Saccharopolyspora erythraea	WP_011874711.1
9	orf+1	209	tryptophan 2,3-dioxygenase	42/51	Amycolatopsis nigrescens	WP_020673272.1
10	orf+2	221	putative membrane protein	78/84	Streptomyces aurantiacus	WP_016638623.1
11	orf+3	139	ketosteroid isomerase	87/92	Streptomyces afghaniensis	WP_020273417.1
12	orf+4	448	putative glucarate dehydratase	93/96	Streptomyces ambofaciens ATCC 23877	CAJ89466.1
13	orf+5	156	gluconate permease	58/63	Streptomyces ambofaciens ATCC 23877	CAJ89467.1
14	orf+6	92	hypothetical protein	77/86	Streptomyces chartreusis	WP_010040441.1
15	orf+7	587	peptidase S8	84/88	Streptomyces venezuelae	WP_015038177.1
16	orf+8	439	protease	71/80	Streptomyces niveus NCIMB 11891	EST31045.1
17	orf+9	692	ABC transporter	90/95	Streptomyces sp. CNB091	WP_018956698.1
18	orf+10	396	transcriptional regulator	57/71	Streptomyces niveus NCIMB 11891	EST33937.1
19	orf+11	67	hypothetical protein	63/67	Streptomyces niveus	WP_023536416.1
20	orf+12	582	asparagine synthase	55/64	Streptomyces niveus	WP_031228349.1
21	orf+13	214	hypothetical protein	49/60	Streptomyces niveus	WP_031228351.1
22	orf+14	43	none			
23	orf+15	274	phosphoesterase	71/77	Streptomyces sp. Mg1	WP_008737016.1
24	orf+16	651	diacylglycerol kinase	47/59	Streptomyces griseus	WP_003967271.1
25	orf+17	312	ribonuclease	53/67	Streptomyces sp. FxanaC1	WP_018090524.1
26	orf+18	210	hypothetical protein	40/59	Cellulomonas massiliensis	WP_019136883.1
27	orf+19	288	G-D-S-L family lipolytic protein	70/80	Streptomyces pratensis	WP_014152362.1

Figure S3. Annotation of the *red* gene cluster. Tryptophan dimer gene cluster found in clone P121 (GenBank accession: KP274854). The annotation is based on BLASTP (NCBI) homology search of translated genes predicted by FGENESB.

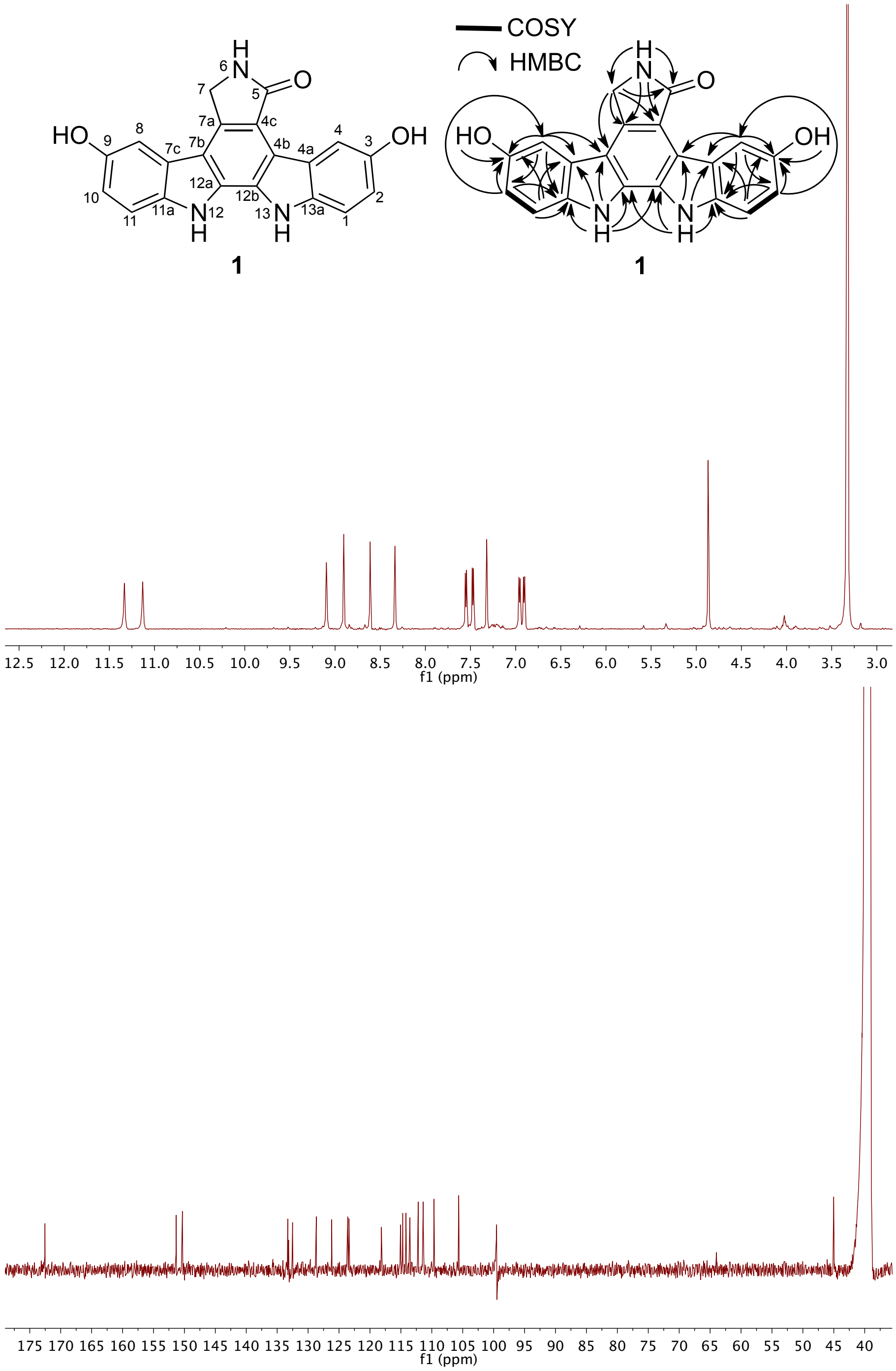


Figure S4. NMR spectra of compound 1. Numbering scheme (**top left**), the observed 2-D NMR correlations (**top right**), ^{1}H NMR (**top,** 600 MHz) and ^{13}C NMR (**bottom,** 150 MHz) in DMSO-d_6.

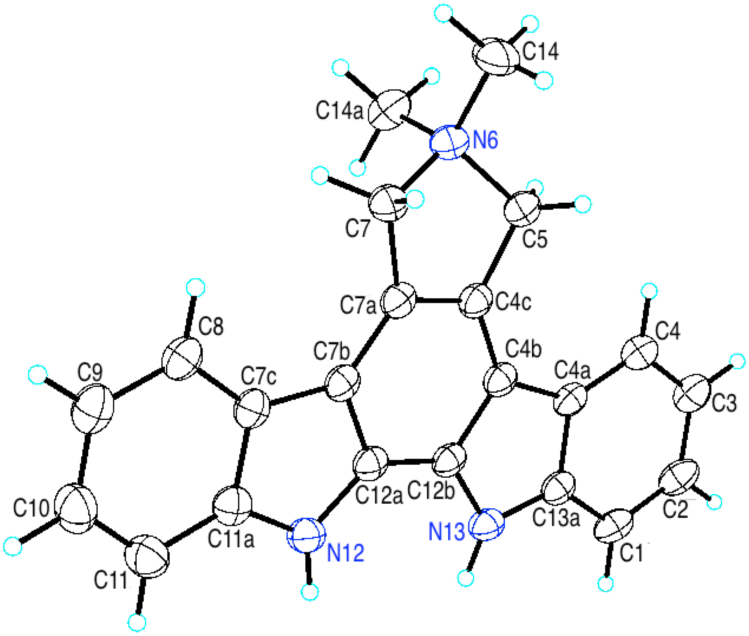


Figure S5. ORTEP diagram of compound 2. CDCC deposition number: CCDC 1043291. Formula: $C_{22}H_{20}N_3^+$, $C_2F_3O_2^-$ (trifluoroacetate salt). Space group: C 2/c. Cell: a=28.3226(19), b=10.5984(7), c=16.3567(8); alpha=90, beta=113.507(3), gamma=90. Temperature: 203 K. Z: 8. R (reflections): 0.0486(3644).

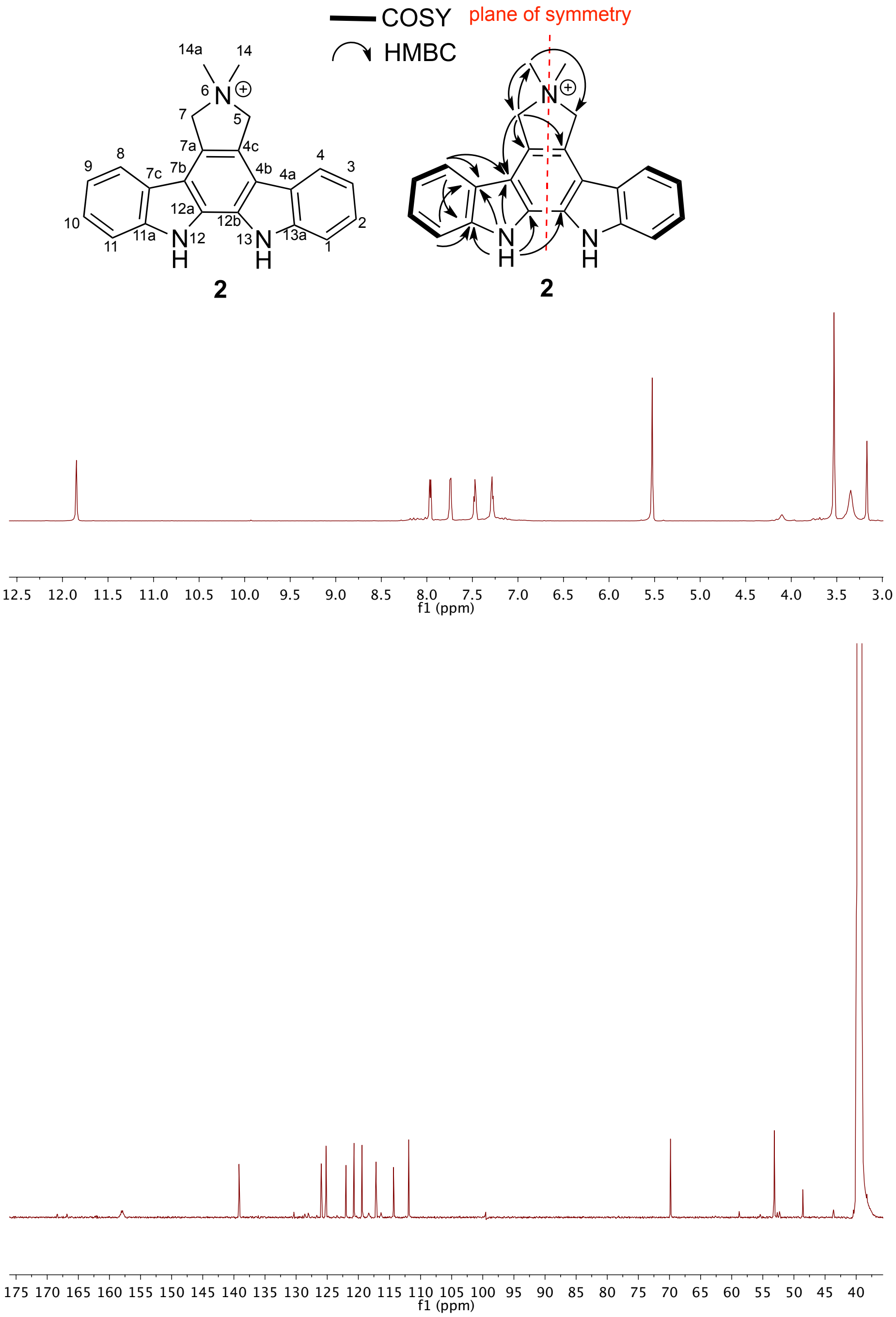


Figure S6. NMR spectra of compound 2. Numbering scheme (**top left**), the observed 2-D NMR correlations (**top right**), ^{1}H NMR (**top,** 600 MHz) and ^{13}C NMR (**bottom,** 150 MHz) in DMSO-d_6.

Table S1. PCR primer list.

Name	Sequence
StaDVF	GTSATGMTSCAGTACCTSTACGC
StaDVR	YTCVAGCTGRTAGYCSGGRTG
AZ25aF	ACTGGGTCTGTGGAACGAG
AZ25aR	TCCCTTCTCCACGACGAAG
AZ25bF	ACGCTGCATGACGGGATG
AZ25bR	CCTTCTCCACCAGAAACAGG
EspOF	GAGACCATGGCAATGAGGTGGGATGAGGCG
EspOR	GAGAGCGGCCGCCACTCATCACGACACCTCC
EspDF	GAGACATATGAGTGTTTTTGATCTGCCCC
EspDR	GAGACAATTGCGTCATTGTTCGTCCTCGG
EspPF	GAGATCATGATCATGACGCAGCGCGGTACA
EspPR	GAGAAAGCTTCGTTCACAGCGGATCGAC
RedOF	GAGACATATGGCCGAGATCCCACACG
RedOR	GAGACAATTGCGAATACGCTCATCACGGGG
RedDF	GAGAACATGTCTATGAGCGTATTCGACCTGCCG
RedDR	GAGAAAGCTTGCATCACGTACCTTCCGCG
RedPF	GAGACCATGGCAATGGACCCGTACTCCTTTTACC
RedPR	GAGAAAGCTTGATCTCGAATGTCCGTCACAG
RedMF	GCGCCCATGGCAATGAGTGACACGTCTCCCG
RedMR	GCGCAAGCTTGTCACTTGGTGGCGATCAC
RedEF	GCGCCATATGGGTGCCAAGGTGACAGT
RedER	GCGCCAATTGGGTACTCAGATGGGGCTG

(a) Underlined sequences indicate the restriction sites added for ligation.

Table S2. Tryptophan dimer sequence tag list.

AZ07a:
GTCATGATGCAGTACCTCTACGCGGCCTACTCGATTCCCACCTACGGCGCGGGCGAAGAGCTGGTGCGGC
GCGGTGACTGGACGCCAGAGCAGCTGACTCTGATGTGCGGCGACGGTGGCGAGACCCGCGAGGGCGGCC
TGCGGGCGAGCCTGATGGACGTCGCACGCGAGGAGATGATCCACTTCCTGGTCATCAACAACGTGCTGA
TGGCCACTGGTGAACCGTTCTACGTGCCGGACATCGACTTCGGCACGCTGAACGAGCAGCTGCCGGTGCC
GCTGGACTTCTGCCTCGAAGCGTTCAACCTCGGCAGCGTGCAGCGGTTCATCCACATCGAGCAGCCGGAG
GGCCTCGTGGGCGCGGTCCGGCTGGGCGATCTGCCGTCCTCATCGGACAGTCACGACTACGCGTCGCTGA
GCGAGCTGTACGGGGACATCAGGGAAGGCCTGCAACGGGTTCCGGACCTGTTCCTGGTAGACCAGGGCC
GTGGCGGTGGCGAGCACCGGCTGTTCATGCGCGAGTCGGTCAACGCCCGGCATCCCGACTACCAGCTCG
AG

AZ07b:
GTGATGATGCAGTACCTGTACGCAGCGTTCTCAATCCCGACCTACGGTGCCGGCCTGGAATATGTGCGAC
GCGGCGAGTGGACAGTCGAACAGTTGCGGCTGGTATGCGGCAACGGGGGAGAAACACGCGACAAGGGG
ATACGCAGCAGCCTGCTCACCGTAGCACGGGAGGAGATGATCCACTTCCTGATCGTCAACAACATCATC
ATGGCCATTGGCGAACCATTTCACATCCCGGCCATCGACTTCGGCACTATCAACAACCAGCTGCTGGTAC
CGCTTGACTTTTCACTCGAGCGCCTGAACATTGGCAGCGTGGCACGGTTTATCGCCATCGAGCAGCCAGA
GGGCCTCCTCGGTGAGGTCAGGCGTGAGGACATCACAAGAAGTTCGGACTCTCCCGCGGAGGACCACCT
CTACAGCTCACTTAGCGAGCTTTACGGAAACATCCGCGAAGGGCTACAACGCGTCCCGAATATCTTCATG
GTAGACAAGGGCAGAGGCGGCGGAGAACATCACCTCTTCCTCCGAGAATCGATCAACGCAGTCCATCCC
GACTACCAGCTGGAG

AZ09a:
GTGATGCTGCAGTACCTCTACGCGGCGTTCTCGATCCCGACTTACGGCGCAGGCGTGGAGTACGTGCGAC
GCAACGAGTGGACCGCCGACCAGTTGCGGCTGCTATGCGGCGACGGGGGAGAGACGCGCGACAACGGC
ATGCGGAGCAGCCTGCTCACCGTGGCCCGCGAGGAAATGATCCACTTCCTCCTCGTCAACAACATCATCA
TGGCCATTGGCGAACCATTCCACATCCCGTCCATCGACTTCGGCACTATCAACAGCCAGCTGCTGGTCCC
GCTTGACTTCTCGCTCGAACCCCTCGGTGTCGGCAGCGTCGCGCGGTTCATCGCCATCGAGCAGCCAAAC
GATCTCATCGGCGAGGTCAGGCGCGGGGACATCGCGAGCTCGGGACCGCGGGAGGAGAACCACCTCTAC
AGCTCACTGAGCGAGCTCTACGCGAACATCCGCGAAGGGCTACAACGCGTTCCAGATCTCTTCATGGTGG
ACAAGGGAAGAGGCGGTGGAGAGCACCACCTCTTCCTCCGAGAATCGATCAACGCGGTCCATCCCGGCT
ACCAGCTCGAG

AZ09b:
CGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGTCATGATGCAGTACCTGTACGCGGCCTACTCGGT
GCCCACGCACGGCGCTGGCTTGGAGTACGTGCGCCTGGGTCTGTGGAACTCCAGGCAGCTCGCTGTGGCC
TGCGGTGACGGCGGTGAGTCGCTGGACACCGGAATCCGCACCACGTTGCTGTCCATAGCGCGCGAGGAG
ATGATGCATTTCCTGATGGTCAACAACGTCCTGATGGCCATCGGAGAACCGTTCACGGTGCCGCGGATCG
ACTTCGGCACGATCAACGACGAGTTGCCGGTGGCGCTGGATTTCTCCCTCGAGGGACTGGGGCTTGGTAG
CGTCCAGCGGTTCATCGCGATCGAGCAGCCCGAGAACGCGGTCGGCGATGTCCAGCTCGGCGACGTGAT
GGCGACGAGAACCCAGCCGCTTGCGAATTATGCCGGGCTGAGCGAGCTGTACGGCGATATCCGCGACGG
GCTCGAGCGGATTCCGGGCCTCTTCGTGGTCGAGAAGGGACGCGGCGGCGGCGAGCACCATCTGTTCCT
GCGTCGCTCCATCGACCTGGACCACCCGGGCTACCAGCTTGAAAAGGGCGAATTCCAGCACACTGGCGG
CC

AZ15a:
GTCATGATGCAGTACCTGTACGCGGCGTTTTCGATCCCGACCTATGGTGCAGGTGGGGAGTATGTGCGCC
GTGGCGAGTGGACGCCGCAGCAGTTACAGCTGGTCTGCGGCGACGGTGGAGAAACGCGCGACAACGGC
ATGCGAGGCAGCTTGCTCACTGTGGCGCGCGAGGAAATGATCCACTTTCTGGTCGTCAACAACATCATCA
TGGCGATTGGTGAACCGTTCTACGTTCCAGCCATCGACTTCGGCACCATCAACCACCAGCTACCGGTTCC
GCTGGATTTCTCGCTTGAGCCCCTTGGTCTCGGCAGTGTGCAGCGCTTCATCGCCATCGAGCAGCCTGAG
GGTCTCGTCGGTGAGGTCAAGCGTGGCGGCATCGTGAGTACGGGGTCGGGCGAGGAGAACTACCCCTAC
AGCTCACTGAGCGAGCTCTACGGGAGCATCCGCGAGGGTCTGCAACGCGTCCCGGATCTCTTCATGGTGC
ACAAGGGCAGAGGCGGAGGAGAGCATCACCTCTTCCTGCCAGAGTCGATCAACACGGTTCATCCCGGCT
ACCAGCTGGAG

AZ15b:
GATCCAGTACCTGTACGCCGCCTACTCGGTGCCCACGCACGGCGCCGGGCTGGAGTACGTGCGACTGGG
TCTCTGGAACGCCCAGCAGCTGGCCGTGGCCTGCGGTGACGGCGGTGAGTCGCTGGACACCGGCATCCG
CACCACGTTGCTGTCCATAGCCCGCGAGGAGATGATGCACTTCTTGATGGTCAACAACGTCCTGATGGCC
ATCGGGGAACCTTTCACGGTGCCGCGGATCGACTTCGGCACGATCAACGACGAGCTGCCGGTTGCGCTG
GACTTCGCACTCGAGGGGCTGGGGCTCGGCAGCGTCCAGAGGTTCATCGAGATCGAGAAGCCCGAGAGT
GCGGTCGGCGATGTACATCTCGGTGACGTCAGGGCGACCGGACCCCGGCCGCCGTTCAGTTATGCCTCGC
TGAGCGAGCTGTACGGCGATATCCGCGACGAGCTCCGGAGAGTTCCGGGCCTCTTCGTCGTGGAGAAGG
GACGCGGCGGCGGCGAGCACCACTTGTTCCTGCGCCGCTCCATCGACATGGATCACCCCGGCTACCAGCT
TGAG

AZ19a:
GCGGCCGCCAGTGTGATGGATATCTGCAGAATTCGCCCTTGCAGTACCTGTACGCGGCCTACTCGGTGCC
CACCTACGGCGCCGGCGCCGAGTACGTGCGGCTCGGCCTGTGGACCCCCGAGCAGCTCCAGCTCGCCTG
CGGTGACGGCGGGGAGACCCTCGACGAGGGCATCCGCAGCATGCTGCTCGGCATCGCGCGGGAGGAGAT
GGTCCACTTCCTGCTCGTCAACAACATCATCATGGCGCTGGGCGAGCCGTTCCACGTGCCGCGGATCGAC
TTCGCCACCGTCAACCACGAGCTGCCGGTACCGCTCGACTTCGCCCTGGAGCGGCTGAACCTCGGCAGCG
TGGAGCGCTTCACCCAGATCGAGAAGCCGGAGCACCAGGTCAGCCACCTGCGCTCCCGTGACGGCGGCA
GCGCGGAGCCGCACGCCGGGCACGGGTACGCCTCGCTCAGCGAGCTGTACGCCGACATCCGCGAGGCCC
TCGAACGGATTCCCGGCCTCTTCCTCGTGGAGAAGGGCAGGGGCGGCGGTGAGCACCACCTGTTCCTCCG
CGAGTCGATCAACCACGAGCATCCCGGCTACCAGCTGGAGAAGGGCGAATTCCAGCACACTGGCGGCCG

AZ19b:
GCAGTACCTCTACGCGGCATGGTCCATCCCCACGTCCGGCGCCGGCGCCGAGTACGTGCGCCGCGGTGA
GTGGACCCCGCGGCAGCTGCGCCTGGTGTGCGGGGACGGCGGCTCCACTCTAGACTACGGCATCCGCGG
GATGCTGCTCGACGTGGCCCGCGAGGAGATGATCCATTTCCTGGTCATCAACAACATCATCATGGCCATC
GGCGAGCCGTTCCACCTGCCGGCCATCGACTTCGGCACGATCAACGAGCAGATACCGGTCCCGCTCGACT
TCTGCCTGGAGGGCTTCGGCATCGGCGCCGTCGCGCGGTTCATCGCCATCGAACGGCCGGATCACCTCAC
CGTCGAGCTGAGCGGGGACGGGTCCGCACCGGCCGGGAGCGGTGAGGCGCCTTATCCGTACGGCTCGCT
CAGCCAGCTCTACGCGGCGATTCGCGAGGCAGTGCAACGGATCCCGGACCTCTTCCTGGTCCGCAAGGG
CCGTGGCGGTGGCGAGCACCACCTGTTCATGCGGGAGTCGGTCAACGCGGTGCACCCCGGCTACCAGCT
GGAAAAGGGCGAAATTCCAGCACACTGGCGGCCG

AZ22a:
GTGATGCTGCAGTACCTGTACGCCGCCTACTCGGTGCCCACGCACGGCGCCGGGCTGGAGTACGTGCGCC
TGGGTCTCTGGAACGCCCAGCAGCTGGCCGTGGCCTGCGGTGACGGCGGTGAGTCGCTGGACACCGGGA
TCCGCACCACGTTGCTGTCCATAGCCCGCGAGGAGATGATGCATTTCCTGATGGTCAACAACGTCCTGAT
GGCCATCGGGGAACCCTTCACGGTGCCGCGGATCGACTTCGGCACGATCAACGACGAGCTGCCGGTTGC
GCTGGATTTCTCACTCGAGGGGTTGGGGCTCGGCAGCGTCCAGCGCTTCATCGAGATCGAGAAGCCCGA
GAGTGCGGTCGGCGACGTCCAGCTCGGTGACCTCATGGCGACGGGACCCCGACCGTCGCTCAATTACGC
CTCGTTGAGCGAGCTGTACGGCGATATCCGCGACGGGCTCCAGCGAGTTCCAGGACTTTTGTCGTGGAG
AAGGGACGGGGCGGCGGTGAGCACCACTTGTTCCTGCGCCGCTCCATCGACCTGGATCACCCGGACTAC
CAGCT

AZ22b:
GTGATGCTCCAGTACCTGTACGCCGCCCACTCGATCCCGACGTACGGTGCCGGGCAGGAGCATGTGCGTC
GCGGGCTGTGGACACCGGAGCAGCTCACGCTGACCTGCGGCGACGGCGGCCAGACCACGGCCGGCGGA
ATCCGTGGCACCCTGCTGTCGGTCGCCCGTGAGGAAATGATCCACTTCTTGCTGATCAACAACATCATCA
TGGCGATGGGTGAGCCGTTCTTCGTGCCGGCCGTCGACTTCGGCACGGTCAACAACACCCTGCCACTGCC
GCTGGACCTCGCCCTGGAGCAGTTCGGCATAGCCAGCGTCCAGCGGTTCATCGCGATCGAACAGCCGCA
CACCCAGATCGGTGAGATCAGCGTCGGCACGGATCCGGGACTGTCGGAGCCGCTGTCGGAGCCGACACA
TACATACGGTTCGCTCAGCGGCCTGTACGCCGACATCCGGGAAGGACTGCAGCGGGTGCCGGACCTGTT
CATGGTGAAGAAGGGCCGTGGCGGAGGCGAGCACCACCTGTTCCTCCGTCAGGGCGTCAACGCGGTCCA
CCCGGACTACCAGC

AZ25a:
GTGATGCTGCAGTACCTGTACGCCGCCTACTCGGTGCCCACGCATGGCGCCGGGCTGGAGTACGTGCGAC

TGGGTCTGTGGAACGAGCAGCAACTGGCCGTGGCCTGCGGTGACGGCGGTGAGTCGCTGGACACCGGCA
TCCGCACCACGCTGCTGTCCATAGCCCGCGAGGAGATGATGCATTTCCTGATGGTCAACAACGTCCTGAT
GGCCATCGGAGAACCGTTCACGGTGCCGCGGATCGACTTCGGCACGATCAACGACGAGTTGCCGATCGC
GCTGGAGTTCTCACTCGAGGGGCTGGGGCTCGGCAGCGTCCAAAGGTTCATCGAGATCGAGAAGCCCGA
GAGTGCGGTCGGCGATGTACATCTCGGCGACGTCAGGGCGACCGGACTCCGGCCGCCGTTCAGTTATGC
CTCGCTGAGCGAGCTGTACGGCGATATCCGCGACGGGCTCCGGAGAGTTCCGGGCCTCTTCGTCGTGGAG
AAGGGACGCGGCGGCGGCGAGCACCACTTGTTCCTGCGCCGCTCCATCGACATGGATCACCCGGACTAC
CAG

AZ25b:
CTCCAGTACCTGTACGCGGCCTTCTCCGTACCGACCCACGGGGCAGGGCTCACGTACGTGCGCCGGGGCC
AGTGGACGGCGGCGCAGCTGGAACTGGCCTGCGGCGGGGGAGGGGCGACGCTGCATGACGGGATGCGT
GGCAGCCTGCTGACCGTGGCCCGCGAGGAGATGATCCACTTCTTGCTGGTGAACAACATCATCATGGCG
ATGGGGGAGCCGTTCCACGTCCCGGTGATCGACTTCGCCACGCTCAACGAGAAGCTGCCGGTCCCGCTG
GACTTCGCCCTGGAGCCGCTGAGTGTCGGCAGCGTGCAGCGTTACGTCGCCCTTGAGCAGCCGGCGGCA
CTCACCCCGGAGATCCGGCGGGGCCCCCGTACCACCCCCCAGGAAAGGGCAGCGGGCTTCACCTGGGGA
TCGGTCAGCGAACTGTACGCGGGGGTGCGCGAAGGGCTGCAGCGCGTCCCGGACCTGTTTCTGGTGGAG
AAGGGGCGCGGCGGCGGTGAACACCATCTTTTCCTGCGGGAGTCGGTCAACAGGAACCATCCGGACTAT
CAGCTGGAG

NM1a:
GCCTACTCGATCCCCACGTACGGCGCCGGCCAGGAGCACGTGCGCCGTGGGCTGTGGACGCCCGAACAG
CTCGCGCTGGTCTGCGGCGACGGCGGGGAGACCACCGCCGGCGGGATCCGTGGCACGTTGCTGTCCGTG
GCGCGGGAAGAGATGATCCACTTTTTGTTGATCAATAACGTGATCATGGCGATGGGTGAGCCGTTCTTCG
TGCCGACCGTCGACTTCGGCACCATCAACAACACCCTGCCACTCCCCCTCGACCTGGCGTTGGAGCAGTT
CGGCATCGGCAGCGTCCAGCGGTTCATCGCCATCGAACGACCCCACGCCCAGGACGGCGAGATCCAGAT
CGGCACCGACCAGGGCAGCGGCGGGCCGCTCGCCGAACCGACCCACACCTACAGCTCGCTCAGTGGCCT
CTACGCCGACATCCGGGAAGGACTGCAGCGGG

NM1b:
GCGTACTCGGTGCCGGCCTACGGGGCGGGGGAGGAGTACGTCCGGCGCGGGCTGTGGACCCCCGAACAG
CTGCGGCTCGCCTGTGGGGACGGCGGCCGGACCCGCGACGGGGGCATCCGCGGCACGCTGCTCGGCATC
GCCCGCGAGGAAATGATCCACTTCCTGATCGTCAACAACATCATCATGGCGATGGGCGAACCGTTCTACG
TCCCCGACGTCGACTTCGGCACGATCAACAACACCCTGCCCGTGCCGCTGGACTTCGCCCTGGAGCCCCT
CGGCGTGGGCAGCGTGCAGCGGTTCATCGCGATCGAACGGCCGGAGGACCAGGTCGGCGAGCTCCACCC
CCCGGTCCCGGTTCGGTTCCCCGCCGGCCGCCGAGCACCCTTACGCCTCGCTCAGCGAGCTGTACGGCGA
CATCCGCGAGGCCTGCAACGGGTCCCCGACGTC

NM4a:
CTGTGGTCCATCCCCACCTACTCGGCGGGCGCCCAGTACGTCAGGCGGGGGAGTGGACCCCCGAACAGC
TCCGGCTGATGTGCGGCGCGGGGCCGCACCGCATCGACGGAGGCGTCCGGCAGCGCCTGTTCGAGGTGG
CCAGGGAGGAGATGATCCATTTCCTGCTGATCAACAACATCGTCATGGCGACCGGGCAGCCGTTCCACCT
TCCGGCGATCGACTTCGGCACGGTGAACAACGAGCTGCCGGTGCCGCTCGACTTCTGCCTGGAGCCCTTC
GGACGCGGCGCCCTGGAGCGGTTCATCGCGCTGGAGCGGCCCTACGACCTGGTCAGGGACATCGCCGGG
AACGACGCGCCGGCCGGCGGCGTGCCCGAGGGGCGCGCCCCGTACGGTACGGCTCGCTGAGCGAGCTGT
ACTCGGCCATCCGCGAAGCCGTCAGGCCATCC

NM4b:
GCCTATTCGGTGCCCACCCACGGTGCCGGCCTGGAGTACGTGCGGCGGGGCACGTGGACACCCGAGCAG
CTCCGGCTCGCCTGCGGCGACGGCGGGGAGACCCTAGACGAGGGCATCCGCAGCATGCTGCTGAACATC
GCCCGCGAGGAGATGATCCACTTTCTCCTCGTCAACAACATCCTCACCGCGGTGGGCGAGCCGTTCCACG
TGCCCCGGCTCGACTTCGGCACCGTCAACCACGAGCTGCCGGTACCGCTGGACTTCTCCCTGGAGAGGCT
GGGGCCCGGCAGCGTCGAACGGTTCGTGCAGATCGAACGGCCGGAAGACCTCGTCCACGACGTACGGCG
GGGCGACACACCGGCGGCCCCCGCGCCCCACGACGACCGGCACCCCGTACGCCTCGCTGAGCGAGCTGT
ACACGGACATCCGCGAGGGCCTGCACCGCGTCCC

NM8a:
GCCTACTCGGTACCCACCCACGGCGCCGGCGCGGAGTACGTACGCCGGGGCCTGTGGACGCCCGACCAA
CTGCGGCTCGCGTGCGGTGACGGCGGGGAGACCCTCGACGAGGGCATCCGCAGCATGCTGCTGACCATC
GCCCGCGAGGAGATGATCCACTTCCTCCTCGTCAACAACATCCTCATGGCGGTGGGCGAACCCTTCCACG
CGCCCCGGATCGACTTCGGCACCGTCAACCGACAGCTGGCCGTCCCGCTGGACTTCGCCCTGGAGCGCCT
GGGGCCCGGCAGCGTGGAGCGGTTCGTACAGATCGAACGCCCCGAGGACCTCGTCGACGAGGTACGGCG
CGGCGACGCTCCGGCGCCCCCGCCGCGTACGACGAGCGGCACCCGTACGCCTCGCTGAGCGAGCTGTAC
GCGGACATCCGGGAAGGGCTGGAGAGCATCCCG

NM8b:
GCGTACTCGGTCCCGGCGCACGGCGCCGGCCAGGAGTACGTCCGCACCGGCGACTGGACCGAGGAACAG
CTCCGGACCGCCTGCGGTGACGGGGGAGAGACGCTCGACGGCGGCATCCGGGGCGTCCTGGTCCGGGTC
GCGCGCGAGGAGATGATCCATTTCCTGGCGGTCAACAACATCCTCACGGCCATCGGTGAGCCGTTCCACG
TCCCGGACGTCGACTTCGGCACCCTGAACGCGACACTGCCCGTCCCACTGGACCTCTGCCTGGAACGCCT
GAACCTCGCCAGCGTCGAGCGGTTCGTCGAGATCGAGCGGCCCGAGGGCCTGACCGGCGACGTGCGGCT
GGGGGAGCGGGCGGGGCGCGGTGAGCTCGACCGGCACACCTACTCGTCGATCAGCGAGCTGTACGGCGA
CATCAGGCAGGGCCTGCAGGACATCCCGG

NM14a:
CTGTGGTCCATCCCCACCCACTCGGCCGGGATCGAGTACGTGCGGCGACGCGAGTGGACGCCCGGGCAG
CTCCGGCTCATGTGCGGCGAGGGCCCGCACAGCCTCGACGGCGGCGTCCGGCAGGACCTGTTCGCCGTC
GCCCGCGAGGAAATGATCCACTTCCTGCTGATAAACAACATCATCATGGCGACGGGTCAGCCGTTCCACC
TGCCCCGGATCGACTTCGGCACAGTGAACGGCGAGCTCCCGTCCCATCGACCTGTGCCTGGAGCCCTTCG
GGCGGGCAGCCTGCAACGGTTCGCCGCGCTGGAACGGCCCTACGATCTGGTCCGCGACCTCGCCGCCGA
CCGGCCACCGGTGACCGCGTGCGACCCGTATCCCTACGGTTCGCTCAGCGAGCTGTACAAGGCCATCCGC
CAGGCGATCCAGGGACATCCCGGACGT

NM14b:
GCGTACTCCATCCCGACCCACGGCGCCGGGGTCGAGCACGTCCGCCGCGGCCTGTGGACCCCGCAGCAG
CTGGAGCTGGCCTGCGGCGACGGCGGCGCCACCACCGCCGGCGGGCTGCGCGGCATGCTGCTCGGTGTG
GCCCGCGAGGAGATGATCCACTTCCTGCTGGTCAACAACATCATCATGGCCATGGGGGAGCCGTTCCAC
GTGCCGGTGGTCGACTTCGGCACCGCCAACACCACCCTGCCGGTGCCGCTGGACGTCAGCCTGGAGGCG
CTCAACCTCGGCAGCGTCCAGCGGTTCATCGCCATCGAGCGGCCGGACTGCGAGGTCGGCGAGCTGCGC
CGGGCGCCGGGCACCGGAGACCCGCTCGCCGCCGCCGGCCGGTCGCCGCGCTGCTACGGCACCGTCAGC
GAGCTGTAACGCCGAGACGTCCGGGGAGGGCC

NM18a:
GCGTACTCGGTGCCGACCCACGGCGCCGGGGCCGAGCTCGTCCGGCGCGGCCGGTGGACGCCCGAGCAG
CACCGGCTGGCCTGCGGCGACGGCGGTGAGACCACCGACGCCGGGGTGCGCGGCATGCTGCTGTCGGTG
GCCCGGGAGGAGATGATCCACTTCCTGCTGATCAACAACATCATCATGGCCATGGGTGAGCCGTTCCACG
TGCCGGTGGTCGACTTCGGCACCATCAACACCACCCTGCCGGTGCCGCTGGACGTCAGCCTGGAGGCGCT
GAACGTCGGCAGTGTGCAGCGGTTCATCGCGATCGAACGCCCCGACACCAGATCGGCGAGCTGCACCCG
GCGGCCGGTACCGGCTACCCACCGGTCGTCGACCCGACCCGACCGCCGCGCAGCTACGGCACCCTCCAG
CGAGCTGTACGCCGAGACTCCGCGACGGGCT

NM18b:
GCCTGGTCGATCCCGACGGCCGGCGCCGCCGCCGAGTTCGTCCGCCGCGGCGAATGGACGCCGGAGCAG
CTGCGGCTGGCGTGCGGCGACGGCGGACCCACCCTCGACTACGGGATGCGCGGCACGCTGCTCAACGTG
GCCCGCGAGGAAATGATCCACTTCCTGGTCATCAACAACATCATCACCGCGACCGGCGACGCGTTCCAC
GTGCCGGCGATCGACTTCGGCACCCTCAACGAGCAGCTGCCGGTGCCGCTGGACTTCAGCCTGGAGGCG
TTCGGGCTGGGGCCGCTGCAACGCTTCATCGCCATCGAACGGCCGGACGACCAGACCGTCGAGTTCGCC
GGGACCGACACGCTGCTCGACCGGGGCGACGCGCTGTACCCGTACGGCTCGCTCAGCGAGCTCTACGCG
GCCATCCCGCGAGGGGCCATTCCCAGCCGGGGT

NM19a:
GCGTACTCGGTGCCGACGTACGCGGCGGCCGAGGCCTTCGTCCACCGCGGTCTGTGGACCCCCGAGCAG
CTGCGGGTCGCCCTGGGTGACGGCGGAGAGACCCTCGACGGCGGCGTCCGGGGAAAGCTGGTGGAGATC

GCGCGCGAGGAGATGATCCATTTCCTCCTCGTCAACAACATCCTCATGGCTCTCGGTGAGCCCTTCTGCG
TCCCCGCGCTGGACTTCGGCGCGCTCGGCACCGACCTGCCGGTGCCCCTGGACCTGTGCCTCGAAGGGCT
CGACATCGCCACCGTGGCGCGGTTCATCGCGATCGAGCAGCCCGCGGTGGGCACGCCGGAGGTACGGCG
ACCCGACCTGCCCACCACCACCGGCGCGGCGGGTGCGGGCCGGTACGAGACGTTGAGCGAGATGTACGC
GCGGATCCGTCAGGGGCTGCAGGACGTCCCG

NM19b:
GCCTGGTCCATCCCGACGGCCGGTGCCGGGGCGGAACTCGTCCGCCGTGGCGAGTGGACCCCGGAGCAG
TTGCAGCTGGCCTGCGGCGACGGCGGTCCGACGCTCGACTACGGGATGCGCGGCACGCTGCTCAACGTG
GCCCGGGAGGAGATGATCCACTTCCTGGTCATCAACAACATCATCACCGCCACCGGGGACAGCTTCCAC
CTGCCGGCGATCGACTTCGGCACCCTCAACGAGCAGCTGCCGGTGCCCCTGGACTTCAGCCTGGAGGGTT
CGGGATCGGGCCGTTGCAGCGGTTCATCGCCATCGAGCAGCCGGACGACCTGACCGTCGAGTTCGCCGG
GACCGACACCCGTGCGTCGACCGGGGCGACGCCATCGTACCCGTACGGCTCGCTCAGCGAGCTGTACGC
GGCGAGTCCGGGAGGCCGTCCGAGCCGGGGGC

NE103a:
GCGTTCTCGATCCCCACCTACGGCGCGGGTCAGCAGCTGGTGCGGCGCGGTGACTGGACACCGGAACAG
CTGCGGCTCATGTGCGGTGACGGGGGTGAGACGGCGGACGGCGGCGTGCGGGGCGCGCTGCTCGGCGTC
GCCCGCGAGGAGATGATCCACTTCCTGGTGGTCAACAACGTCCTCGTGGCCATCGGCGAGCCGTTCCACC
TGCCCGACATCGACTTCGGCACGATCAACGACAACCTCATGGTGCCGCTCGACTTCTCGCTCGAGGCGCT
CGGGCTCGGCAGCGTACAGCGCTTCATCCAGATCGAGCAGCCGGACGGGCTCGTCGGCGCCGTGCGGCT
CGGCGACCTCCCGCCCCCGTTGAGTCCGAACAGGACTACGACTACGCCTCGCTCAGCGAACGTGTACGGT
CGGACGACTACACGGAAGGGGTTAGGTCGACC

NE103b:
TGTGGTCGCTGCCGACTCACCGTATGGGGCCGCGCTGGTCCACCAGGGCGAGTGGACCGAGGACCAGTA
CACGCTGGTCTGCGGGCGCGGGCCGGCGACGGCTGACGGGGGGATCCGGGGCGCTCTGTTCGGGGTGGC
CCGAGAGGAGATGATTCATTTCCTCGTCATCAACAACATCATCATGGCCACGGGTCAGCCCTTCCACGTG
CCCGACATCGACTTCTCCTCGCTCAACGCGCAGATCGACCTGCCGATGGACTTCTGCCTGGAGCGGTTCG
GCCTGTCCTCGCTGAGCAGGTTCGTCGAGTTTGAAAAGCCGTTTTCGCTCACCGTCGAGCCAGCGCCGGC
ACCGGGAGGCCCGCAGCGAAAGGGTGCCCGTTACGGGTCGCTGAGCGAGCTCTACGCGTCGATCCGGGA
CGGCCTCGCCCGGTTCCCGAGGCCTTCCT

NE120a:
CTGTGGTCCATCCCCACCCACTCGGCCGGGACCGAGTACGTGCGGCGGGGCGAGTGGACGCCCGGGCAG
CTCCGGCTCATGTGCGGCGAGGGCCCGCACAGCCTCGACGGCGGTATCCGGCAGGGCCTGTTCGCCGTC
GCCCGCGAGGAAATGATCCACTTCCTGCTGATCAACAACATCATCATGGCGACCGGCCAGCCGTTCCACC
TGCCCCGGATCGACTTCGGCACGGTGAACGGCGAACTCCCCGTCCCCTCGACCTGTGCCTGGAGCCCTTC
GGGCGGGGCAGCCTGCAACGGTTCGCCGCGCTGGAACGACCCCACGACCTGACCCGCGACCTCGCCGAG
GACCGGCCGCCGGTGACCACGCGCGACCCGTACCCCTACGGTTCGCTCAGCGAGCTGTACGCCGCCATCC
GCCAGGCCGTCCAGGACATCCCGGACGTCTTCC

NE120b:
GCGTACTCGGTCCCGGCCTACGGGGCGGGGAGGAGTACGTCCGGCGGGGGCTGTGGACCCCCGAGCAGC
TGCGGCTCGCCTGCGGGGACGGCGGCCGGAGCCGCGACGGGGGCATCCGCGGCACGCTGCTCGGCATCG
CCCGCGAGGAGATGATCCACTTCCTGATCGTCAACAACATCATCATGGCGATGGGCGAACCCTTCCACGT
CCCCGACGTCGACTTCGGCACGATCAACAACACCCTGCCGGTGCCGCTGGACTTCGCCCTGGAGCCCCTC
GGCGTGGGCAGCGTCCAGCGGTTCATCGCGATCGAACGGCCGGAGGACCAGGTCGGCGAGCTCCACCCC
CCGGTCCGGTACGGCTCCACGCCGGACGCCGGGCACCCGTACGCCTCGCTCAGCGAGCTGTACGGCGAC
ATCCGCGAGGGCCTGCAACGGGTCCCCGAGTC

NE123a:
CTGTGGTCCATCCCCACCCACTCGGCCGGGACCGAGTACGTGCGGCGAGGCGAGTGGACGCCCGGGCAG
CTCCGGCTCATGTGCGGCGAGGGCCCGCACAGCCTCGACGGCGGCGTCCGGCAGGACCTGTTCGCCGTC
GCCCGCGAGGAAATGATCCACTTCCTGCTGATCAACAACATCATCATGGCGACGGGTCAGCCGTTCCACC
TGCCCCGGATCGACTTCGGCACGGTGAACGGCGAGCTCCCCGTCCCGTCGACCTGTGCCTGGAGCCCTTC
GGGCGGGGCAGCCTGCAACGGTTCGCCGCTCTGGAACGGCCCTACGACCTGGTCCGCGACCTGGCCGCC

GACCGGCCACCGGTGACCGCGCGACCCGTACCCCTACGGTTCGCTCAGCGAGCTGTACGGGGCCATC
CGCCAGGCGATCCAGGACATCCCGGACGTCTTTC

NE123b:
GCCTGGTCCATCCCGACGGCCGGTGCCGGGGCGGAACTCGTCCGCCGTGGCGAGTGGACCCCGGAGCAG
TTGCAGCTGGCCTGCGGCGACGGCGGTCCGACGCTCGACTACGGGATGCGCGGCACGCTGCTCAACGTG
GCCCGGGAGGAGATGATCCACTTCCTGGTCATCAACAACATCATCACCGCCACCGGGGACAGCTTCCAC
CTGCCGGCGATCGACTTCGGCACCCTCAACGAGCAGCTGCCGGTGCCCCTGGACTTCAGCCTGGAGGGTT
CGGGATCGGGCCGTTGCAGCGGTTCATCGCCATCGAGCAGCCGGACGACCTGACCGTCGAGTTCTCCGG
GACCGACACCCTGCTCGACCGGGGGCGACGCCATGTACCCGTACGGCTCGCTCAGCGAGCTGTACGCGC
GATCCGGGAGGCATCCGAGCGGGTGCCCGACC

NE125a:
GCCTTCTCCATCCCGGCCTACGGGGCCGGAACCGAGTACGTGCGTCGTGGTGTGTGGACACCCGGGCAGT
TGCGGCTCGCTTGTGGCAGTGGCGGCGAGAACCGGGACGACGGCATGCGTGGCGCCCTGCTCACCGTGG
CGCGTGAGGAGATGATCCACTTCCTGTTGGTCAACAACATCATCATGGCCACCGGGGAGACGTTCTACGT
CCCGTCCATCGATTTCGGGACGATCAACAACGAGCTGCCGGTACCGCTGGACTTCTCCCTCGAACCGTTC
GGGATCGGCAGTGTCCAGCGGTTCATCGCGATCGAGCAACCCGCGGGCCTGGCCGGGGAGATCCGGCGC
GGTGACGTCCGGGGTGACGACGCGCGAGGTCCCCGGCACCCGTACGGCTCACTCAGCGAGCTCTACGCG
AGCATCCGGGAAGGTCTGTGCCGGGTGCCCGGT

NE125b:
GCCTACTCCATCCCGACGTACAGCGCGGGCCGGCGGCACGTGCGGCACGGGCTGTGGACGCCGGAGGAG
CTGTGGCTCGTGTGCGGCGACGGCGGCGAGACGACGACGGGTGGCATGCGCGGCAGGCTGCTGACCGTC
GCCCGCGAGGAGATGATCCACTTCCTGGTCATCAACAACATCATCATGGCGATCGGCGAGCCGTTCTACG
TGCCCGACGTCGACTTCGGCACGGTCGACGCGGTCCTGCCGGTGCCGCTCGACCTCGCGCTCGAGGCCTT
CGGCCTGGGCAGCGTGCAGCGGTTCATCGCCCTCGAGCGGCCCGACGCGCAGGTCGGCGAGGTGCGCCG
CGCCGGCGCCGCCCGACCGCCGTGTCCGACGGAGCCGGCGCGCACGTACTCCCCGCCAAGCGCGCGTGT
ACGCGACCGATCCGGTACGGGGCGGTGAAGGCGC

NE126a:
CTGTGGTCCGTCCCCACCTACTCGGCGGGGGTCCGCCACGTCGAACGCGGCGAGTGGACGCCGGAGCAG
CTGCGCCTGATGTGCGGTGCGGGGCCGCACGGCCTCGACGGCGGGCTCCGGGGGAGCCTGCTGGGCGTG
GCCCGGGAGGAGATGATCCACTTCCTGGTGATCAACAACCTGATCATGGCGACCGGCCAGCCCTTCCACC
TGCCCGACATCGACTTCGGCACGGTGAACAGCGTGCTGCCCGTCCCGGTCGACTTCTGCCTGGAGCCGTT
CGGCATCGCCGCCCTGCAGCGGTTCATCGCCCTGGAGCGGCCGTACGGCCTGGTCCGGGACTTCGCGGG
GAGCGGCGGGAACGGCGGCGGCGGGTCGGCGACGGCGCGCCGCGAGGGGGAACCGCACCGGTACGGCT
CGCTGAGCGAGCTGTACGCGGCCATCCGCCGGGCG

NE126b:
GCCTGGTCCGTCCCGACCCACGGCGCGGCCGGCGAGTACGTGCGCCGGGGCGCCTGGACCCCCCGGCAG
CTCCGGCTCGCCTGCGGGGACGGGGGCCCGACCGGTGACTACGGCGCGCGCGGCCTGCTGCTCGACGTG
GCCCGGGAAGAGATGATCCACTTTCTGGTGGTCAACAACATCCTGACCGCCATGGGGGAGCCGTTCCAC
CTGCCCGCGATCGACTTCGGCACCCTCAACGCCGAGCTGCCGATCCCGCTGGACCTCTGCCTGGAGGGGT
TCGGGATCGGGGCGCTGCAACGCTTCATCGCCGTCGAACAGCCCGACGGGCTGACGCCGGGGCTCGCCG
GGGACCGGCCGGCGGCGGCCGGTGGGCACCCGTACGGGTCGCTCAGCGAGCTCTACGCCGCGATCCGCG
AGGCGGTCACCCGGATCCCCGACCTGTTCGTGGTC

NE130a:
GCCTGGTCGATCCCGCTGTACGGGGCGGGCCGGGAACTGGTACGGCGCGGGGAGTGGAGCGCCCGGCAG
CTCCGCCTGGCGTGCGGCGACGGCGGGGAGACCCTCAGCAACGGCATGCGCGGCAGCCTGATGAACGTG
GCCCGGGAGGAGATGACGCACTTCCTGGTCATCAACAACGTCATCACCGCGGCGGGACTTCCCTTCCACC
TGCCGCGGATCGACTTCGGGACCGTCAACAACCGGCTGCGGGTGCCGCTCGATGTGAGCCTGGAGGGTC
GGGCTGGGCAGCGTGCAGCGGTTCATCGCCCTGGAGCAGCCGCACGGACTGAGCCCCGACTCGCCGAGG
AGACCGAGAACGCCGCCCCCCGGGACGACGGGCAGCCGTACCGTACCAGTCGCTCAGCGACCTGTACGC
CGCGATCCGGGAGGGCATCGCCACCATC

NE130b:
GCCTACTCGATCCCGACCCACGGTGCGGCCCGGGAGTTCGTCCGCCGCGGTCTGTGGACGCCCGACCAG
GCCCGGCTGGCGTGCGGGGACGGCGGCCGCAGCCGCGACGCCGGGGTGCGCGGCCGGCTGCTGGAGGT
GGCCCGGGAAGAGATGATCCACTTCCTCGCCGTCAACAACATCGTCATGGCGATGGGGGAGGCGTTCCA
CGTCCCCGACATCGCGTTCGGCCGGCTCAACGCCACCCTGCCGGTGGCGATGGACTTCTGCCTGGAACCC
TTCGGGCTGGGCACCCTCCAGCGGTTCATCGCGATCGAGCGGCCCCTGGACGCGGCACCGCCCCCGCTCA
GCGCGGTGTCCGAGGACCGGCCGGCGATGCCGGGACCCCGCCGCTACCAGGGCATCCCGGAGTTGTACG
AGGAGATCCGGCAGGGCCTGCACCGCGTCCCCGAC