NEW R2 RETROTRANSPOSONS FOR GENE WRITING

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit under 35 U.S.C. § 119(e) of the United States Provisional Application Serial Nos. 63/569,037 and 63/569,042 filed March 22, 2024, the content of which is hereby incorporated by reference in its entirety.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0002] The content of the electronic sequence listing (385419.xml; Size: 560,415 bytes; and Date of Creation: March 21, 2025) is herein incorporated by reference in its entirety.

BACKGROUND

[0003] Unlike DNA transposons, which move via a "cut and paste" mechanism, retrotransposons move in a "copy and paste" manner, using RNA as an intermediate. Based on whether they are flanked by long terminal repeat (LTR), retrotransposons are classified as LTR type or non-LTR type. Non-LTR retrotransposons are sub-classified into two major groups, namely long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). LINEs are autonomous elements that encode proteins to mediate their own mobility, whereas SINEs are nonautonomous elements that do not encode protein and consequently require LINEs for their propagation.

[0004] Non-LTR retrotransposons are mobilized by a mechanism in which proteins translated from their open reading frame proteins combine with their own mRNA to form a ribonucleoprotein (RNP) complex. The RNP complex is inserted into the target site via a mechanism termed target-primed reverse transcription (TPRT). The TPRT process is initiated by an encoded EN domain that nicks one strand of DNA at a target site and creates a 3'-hydroxyl end, which is used as a primer for reverse transcription of the LINE mRNA onto the DNA target via RT.

[0005] R2 is a long interspersed element (LINE) found in a specific sequence of the 28S rDNA among a wide variety of animals. Known R2 retrotransposons include those from medaka fish

Oryzias latipes (R2Ol) and from Bombyx mori (R2Bm). Efforts have been made to use these R2 retrotransposons for targeted gene knock-in for practical application such as gene therapy.

SUMMARY

[0006] Through bioinformatic approaches with confirmative testing in human cells, the instant inventors discovered highly efficient new R2 retrotransposons from different species. These new retrotransposons, as well as engineered versions thereof, can be suitably used for targeted gene delivery.

[0007] Accordingly, one embodiment of the present disclosure provides a method for inserting an exogenous sequence to a genomic DNA of a cell, comprising introducing to the cell (a) a polypeptide or a first polynucleotide encoding the polypeptide, wherein the polypeptide comprises (i) a DNA binding domain or a CRISPR Cas protein, (ii) a reverse-transcriptase comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 6, 2, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118, 122, 126, 130, 134, 138, 142, 146, 150, 154, 158, 162, 166, 170, 174, 178, 182, 186, 190, 194, 198, 202, 206, 210, 214, 218, 222, 226, 230, 234, 238, 242, 246, 250, 254, 258, 262, 266, 270, 274, 278, 282, 286, 290, 294, 298, 302, 306, 310, 314, 318, 322, 326, 330, 334, 338, 342, 346, 350, 354, 358, 362, 366, 370, 374, 378, 382, 386, 390, 394, 398, 402, 406, 410, 414, 418 and 422, or an amino acid sequence having at least 85% sequence identity to any of the amino acid selected from the group, and (iii) an endonuclease or a nickase, and (b) an RNA comprising (i) a template for the exogenous sequence and (ii) a 3' UTR, or a second polynucleotide encoding the RNA.

[0008] In some embodiments, the method further comprises introducing the first polynucleotide and the second polynucleotide to the cell. In some embodiments, the first polynucleotide and the second polynucleotide are on a same vector or on separate vectors.

[0009] In some embodiments, the 3'UTR is an R2 retrotransposon 3' UTR. In some embodiments, the 3'UTR comprise a sequence selected from the group consisting of SEQ ID NO: 8, 4, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100, 104, 108, 112, 116, 120, 124, 128, 132, 136, 140, 144, 148, 152, 156, 160, 164, 168, 172, 176,

180, 184, 188, 192, 196, 200, 204, 208, 212, 216, 220, 224, 228, 232, 236, 240, 244, 248, 252, 256, 260, 264, 268, 272, 276, 280, 284, 288, 292, 296, 300, 304, 308, 312, 316, 320, 324, 328, 332, 336, 340, 344, 348, 352, 356, 360, 364, 368, 372, 376, 380, 384, 388, 392, 396, 400, 404, 408, 412, 416, 420 and 424.

[0010] In some embodiments, the reverse-transcriptase comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 6, 2, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118, 122, 126, 130, 134, 138, 142, 146, 150, 154, 158, 162, 166, 170, 174, 178, 182, 186, 190, 194, 198, 202, 206, 210, 214, 218, 222, 226, 230, 234, 238, 242, 246, 250, 254, 258, 262, 266, 270, 274, 278, 282, 286, 290, 294, 298, 302, 306, 310, 314, 318, 322, 326, 330, 334, 338, 342, 346, 350, 354, 358, 362, 366, 370, 374, 378, 382, 386, 390, 394, 398, 402, 406, 410, 414, 418 and 422.

[0011] In some embodiments, the polypeptide comprises the endonuclease. In some embodiments, the polypeptide comprises the nickase, which optionally is Cas nickase. In some embodiments, the CRISPR Cas protein is selected from the group consisting of Cas9, Cas12 and Cas13.

[0012] In some embodiments, the DNA binding domain is selected from the group consisting of a transcription activator-like effector (TALE) DNA binding domain, a homeodomain protein, a zinc finger, a helix-turn-helix, a leucine zipper, and a DNA-binding domain of a homing endonuclease, a transposon or retro-transposon.

[0013] In some embodiments, the DNA binding domain is a TALE. In some embodiments, the homing endonuclease is selected from the group consisting of I-SceI, I-CreI and I-DmoI. In some embodiments, the retro-transposon is a long interspersed nuclear element (LINE), optionally selected from the group consisting of R1, R2, R4, R5, R6, R7, R8, and R9.

[0014] In some embodiments, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 5, 1, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69, 73, 77, 81, 85, 89, 93, 97, 101, 105, 109, 113, 117, 121, 125, 129, 133, 137, 141, 145, 149, 153, 157, 161, 165, 169, 173, 177, 181, 185, 189, 193, 197, 201, 205, 209, 213, 217, 221, 225, 229, 233, 237, 241, 245, 249, 253, 257, 261, 265, 269, 273, 277, 281, 285, 289, 293, 297, 301,

305, 309, 313, 317, 321, 325, 329, 333, 337, 341, 345, 349, 353, 357, 361, 365, 369, 373, 377, 381, 385, 389, 393, 397, 401, 405, 409, 413, 417 and 421, or an amino acid sequence having at least 85% sequence identity to any amino acid sequence selected from the group.

[0015] In some embodiments, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 5, 1, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69, 73, 77, 81, 85, 89, 93, 97, 101, 105, 109, 113, 117, 121, 125, 129, 133, 137, 141, 145, 149, 153, 157, 161, 165, 169, 173, 177, 181, 185, 189, 193, 197, 201, 205, 209, 213, 217, 221, 225, 229, 233, 237, 241, 245, 249, 253, 257, 261, 265, 269, 273, 277, 281, 285, 289, 293, 297, 301, 305, 309, 313, 317, 321, 325, 329, 333, 337, 341, 345, 349, 353, 357, 361, 365, 369, 373, 377, 381, 385, 389, 393, 397, 401, 405, 409, 413, 417 and 421.

[0016] In some embodiments, the RNA further comprises a 5'UTR. In some embodiments, the 5'UTR comprises a sequence selected from the group consisting of SEQ ID NO: 7, 3, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63, 67, 71, 75, 79, 83, 87, 91, 95, 99, 103, 107, 111, 115, 119, 123, 127, 131, 135, 139, 143, 147, 151, 155, 159, 163, 167, 171, 175, 179, 183, 187, 191, 195, 199, 203, 207, 211, 215, 219, 223, 227, 231, 235, 239, 243, 247, 251, 255, 259, 263, 267, 271, 275, 279, 283, 287, 291, 295, 299, 303, 307, 311, 315, 319, 323, 327, 331, 335, 339, 343, 347, 351, 355, 359, 363, 367, 371, 375, 379, 383, 387, 391, 395, 399, 403, 407, 411, 415, 419 and 423.

[0017] In some embodiments, the cell is a mammalian cell, preferably a human cell. In some embodiments, the exogenous sequence encodes a protein, preferably a therapeutic protein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0018] **FIG. 1** shows the integration activities of selected new R2 elements with digital PCR.

[0019] **FIG. 2** illustrates the *cis* and *trans* configurations for R2-based reverse transcription.

[0020] **FIG. 3** shows the gene delivery rates of new R2 elements indicated by the GFP levels.

[0021] **FIG. 4** shows representative GFP images and charts indicating the cargo delivery rates of selected R2 elements in a *cis* configuration.

[0022] **FIG. 5** shows representative GFP images and charts indicating the cargo delivery rates of selected R2 elements in a *trans* configuration.

[0023] **FIG. 6** shows the GFP cargo delivery rates of selected R2 elements in a *cis* configuration.

[0024] **FIG. 7** shows the GFP cargo delivery rates of selected R2 elements in a *trans* configuration, where the RNA templates included pseudouridine-modified nucleotides.

DETAILED DESCRIPTION

[0025] The following description sets forth exemplary embodiments of the present technology. It should be recognized, however, that such description is not intended as a limitation on the scope of the present disclosure but is instead provided as a description of exemplary embodiments.

Newly Discovered R2 Retrotransposons

[0026] Through bioinformatic approaches with confirmative testing in human cells, the instant inventors discovered highly efficient new R2 retrotransposons. Their retrotransposition activities were confirmed in human cells *in vitro*.

[0027] Each of the R2 mRNA includes a 5' untranslated region (UTR), a single open reading frame (ORF) encoding the endonuclease (EN) and reverse-transcriptase (RT) and DNA-binding domains, and a 3' UTR. The R2 mRNA can anneal with the target site (*i.e.*, the 28S rDNA in the human genome) and initiates the reverse transcription mechanism.

[0028] Generally, each of the R2 retrotransposons includes one or more N-terminal zinc finger domains, a SANT DNA-binding domain, a 3' UTR binding domain, a template switching domain, a reverse-transcriptase (RT), a C-terminal zinc finger domain (C-ZnF), and an endonuclease (EN). Among them, the RT domain carries the retrotransposition activity, while the other domains are contemplated to be switchable to similar domains. For instance, the N-terminal zinc fingers can be substituted with other DNA binding domains, as discussed below, and so does the EN domain.

[0029] These new R2 retrotransposons, the species with which they are associated, and the corresponding ORF, RT and 5'- and 3'-UTR sequences, are summarized in **Tables A and B**. The corresponding sequences are provided in **Tables 1 and 2**.

Table A. Newly Discovered R2 Retrotransposons

No.	R2 Name	Species	SEQ ID NO: for			
			ORF	RT	5'UTR	3'UTR
1	R2ApuApu	*Apus Apus*	1	2	3	4
2	R2AspTig	*Aspidoscelis tigris*	5	6	7	8
3	R2AspTig'	*Aspidoscelis tigris*	9	10	11	12
4	R2ChaBot	*Charina bottae*	13	14	15	16
5	R2ColStr	*Colius striatus*	17	18	19	20
6	R2CroOre	*Crotalus oregenus helleri*	21	22	23	24
7	R2HemCom	*Hemiprocne comata*	25	26	27	28
8	R2MelGeo	*Melospiza georgiana*	29	30	31	32
9	R2PeCa	*Pelusios castaneus*	33	34	35	36
10	R2CarCar	*Caretta caretta*	37	38	39	40
11	R2PhrBla	*Phrynosoma blainvillii*	41	42	43	44
12	R2PoEx	*Podocnemis expansa*	45	46	47	48
13	R2PufMau	*Puffinus mauretanicus*	49	50	51	52
14	R2Hal	*Haliaeetus albicilla*	53	54	55	56
15	R2EmyOrb	*Emys orbicularis*	57	58	59	60
16	R2OpiFel	*Opisthorchis felineus*	61	62	63	64
17	R2CapEur	*Caprimulgus europaeus*	65	66	67	68
18	R2DeCl	*Denticeps clupeoides*	69	70	71	72
19	R2Onki	*Oncorhynchus kisutch*	73	74	75	76
20	R2AhaPra	*Ahaetulla prasina*	77	78	79	80
21	R2AnnSte	*Anniella stebbinsi*	81	82	83	84
22	R2AriEle	*Arizona Elegans*	85	86	87	88
23	R2AspMar	*Aspidoscelis marmoratus*	89	90	91	92
24	R2AzeSeq	*Azemiops Seq 1*	93	94	95	96
25	R2AzeSeq2	*Azemiops Seq 2*	97	98	99	100
26	R2AzeSeq3	*Azemiops Seq 3*	101	102	103	104
27	R2ChaBot2	*Charina bottae (Second R2)*	105	106	107	108
28	R2CroAda	*Crotalus adamanteus*	109	110	111	112
29	R2CucCan	*Cuculus canorus*	113	114	115	116
30	R2HalLeu	*Haliaeetus leucocephalus*	117	118	119	120
31	R2AqCh	*Aquila chrysaetos chrysaetos*	121	122	123	124
32	R2HarHar	*Harpia harpyja*	125	126	127	128
33	R2CorAlt	*Corapipo Altera*	129	130	131	132
34	R2GopFla	*Gopherus flavomarginatus*	133	134	135	136
35	R2MelCri	*Melozone crissalis*	137	138	139	140

36	R2MyCa	*Myiozetetes cayennensis*	141	142	143	144
37	R2PoAt	*Poecile atricapillus*	145	146	147	148
38	R2ChPi	*Chrysemys Picta*	149	150	151	152
39	R2ATLH	*Aspidoscelis tigris*	153	154	155	156
40	R2ATLH'	*Aspidoscelis tigris*	157	158	159	160
41	R2ATLH"	*Aspidoscelis tigris*	161	162	163	164
42	R2Mac	*Manacus candei*	165	166	167	168
43	R2Pif	*Pipra filicauda*	169	170	171	172
44	R2TyAl	*Tyto alba*	173	174	175	176
45	R2ViLa	*Viper Latastei*	177	178	179	180
46	R2Tral	*Trimeresurus albolabris*	181	182	183	184
47	R2Boja	*Bothrops jararaca*	185	186	187	188
48	R2Bofu	*Boaedon fuliginosus*	189	190	191	192
49	R2CroTig	*Crotalus tigris*	193	194	195	196
50	R2GraSmi	*Grayia smithii*	197	198	199	200
51	R2SibCol	*Sibynophis collaris*	201	202	203	204
52	R2LacBil	*Lacerta bilineata*	205	206	207	208
53	R2LiaOli	*Liasis olivaceus*	209	210	211	212
54	R2MasLat	*Masticophis lateralis*	213	214	215	216
55	R2MasLat2	*Masticophis lateralis Second R2*	217	218	219	220
56	R2OxyLep	*Oxyrhabdium leporinum*	221	222	223	224
57	R2PanAll	*Pantherophis alleghaniensis*	225	226	227	228
58	R2PanGut	*Pantherophis guttatus*	229	230	231	232
59	R2VipUrs	*Vipera ursinii*	233	234	235	236
60	R2FasGig1	*Fasciola gigantica*	237	238	239	240
61	R2FasGig2	*Fasciola gigantica*	241	242	243	244
62	R2FasGig3	*Fasciola gigantica*	245	246	247	248
63	R2FasGig4	*Fasciola gigantica*	249	250	251	252
64	R2FasHep	*Fasciola hepatica*	253	254	255	256
65	R2OpiViv1	*Opisthorchis viverrini*	257	258	259	260
66	R2OpiViv2	*Opisthorchis viverrini*	261	262	263	264
67	R2OpiViv3	*Opisthorchis viverrini*	265	266	267	268
68	R2CalDau1	*Calicophoron daubenyi*	269	270	271	272
69	R2CalDau2	*Calicophoron daubenyi*	273	274	275	276
70	R2SchJap	*Schistosoma japonicum*	277	278	279	280
71	R2SchSol	*Schistocephalus solidus*	281	282	283	284
72	R2SpiEri	*Spirometra erinaceiruropaei*	285	286	287	288
73	R2DipCan	*Dipylidium caninum*	289	290	291	292
74	R2MicRuf	*Micrastur ruficolis*	293	294	295	296
75	R2ChaBot	*Charina bottae*	297	298	299	300
76	R2LiaOli	*Liasis olivaceus*	301	302	303	304
77	R2BotJar	*Bothrops jararaca*	305	306	307	308
78	R2BoaFul1	*Boaedon fuliginosus*	309	310	311	312
79	R2BoaFul2	*Boaedon fuliginosus*	313	314	315	316
80	R2TriAlb1	*Trimeresurus albolabris*	317	318	319	320
81	R2TriAlb2	*Trimeresurus albolabris*	321	322	323	324

82	R2CroTig	*Crotalus tigris*	325	326	327	328
83	R2CroHel	*Crotalus helleri*	329	330	331	332
84	R2OxyLep1	*Oxyrhabdium leporinum*	333	334	335	336
85	R2OxyLep2	*Oxyrhabdium leporinum*	337	338	339	340
86	R2AzeFea1	*Azemiops feae*	341	342	343	344
87	R2AzeFea2	*Azemiops feae*	345	346	347	348
88	R2SisCat	*Sistrurus catenatus*	349	350	351	352
89	R2VipLat	*Vipera latastei*	353	354	355	356
90	R2VipUrs	*Vipera ursinii*	357	358	359	360
91	R2AhaPra	*Ahaetulla prasine*	361	362	363	364
92	R2PanGut	*Pantherophis guttauts*	365	366	367	368
93	R2MasLat1	*Masticophis lateralis*	369	370	371	372
94	R2MasLat2	*Masticophis lateralis*	373	374	375	376
95	R2AriEle	*Arizona elegans*	377	378	379	380
96	R2SibCol	*Sibynophis collaris*	381	382	383	384
97	R2GraSmi	*Grayia smithii*	385	386	387	388
98	R2RhaNuc	*Rhabdophis nuchalis*	389	390	391	392
99	R2MyxAsi1	*Myxocyprinus asiaticus*	393	394	395	396
100	R2MyxAsi2	*Myxocyprinus asiaticus*	397	398	399	400
101	R2MyxAsi3	*Myxocyprinus asiaticus*	401	402	403	404
102	R2AcrFas	*Acrossocheilus fasciatus*	405	406	407	408
103	R2RhiKla	*Rhinichthys klamathensis*	409	410	411	412
104	R2Mobbir	*Mobula birostris*	413	414	415	416
105	R2PetZak	*Petrocephalus zakoni*	417	418	419	420
106	R2Ptesp	*Pterobdella sp.*	421	422	423	424

Engineered R2 Retrotransposons

[0030] The retrotransposition of R2 retrotransposons does not require homology-directed repair (HDR) or donor DNA. Compared to HDR, which is a passive process in cells, reverse transcription of mRNA into cDNA via TPRT using the RT domain shows relatively higher processivity and fidelity, which is advantageous for synthesizing long cDNAs at the target site rather than recombination at the target site as donor DNA.

[0031] It has been demonstrated herein that the newly discovered R2 retrotransposons specifically target the 28S gene in the human genome. Based on sequence analysis, it is contemplated that the specific 28S binding is attributed to the N-terminal zinc finger (ZnF) domains, likely assisted by the SANT DNA-binding domain. Replacement of these N-terminal regions of the R2 proteins, therefore, may alternate the target specificity of the R2 retrotransposons.

[0032] Accordingly, in some embodiments, the present disclosure provides an engineered R2, in the form of a fusion protein, that includes a truncated version of an R2 ORF and one or more other heterologous elements. The truncated R2 ORF, in some embodiments, includes the reverse-transcriptase (or a biological variant thereof) of one of the disclosed R2, but does not include portion of the N-terminal region of the R2, such as one or more of the N-Znf1, N-Znf2, N-Znf3, and SANT DNA-Binding Domain.

[0033] In some embodiments, the heterologous elements may be a DNA binding domain or a CRISPR Cas protein and/or an endonuclease or a nickase. It is appreciated that it is not necessary that both elements are heterologous. For instance, the fusion protein can include one or more heterologous DNA binding domains but the native endonuclease. In another example, the fusion protein can include the native DNA binding domains and a heterologous endonuclease.

[0034] Likewise, another embodiment of the present disclosure provides an engineered R2, in the form of a fusion protein, that includes a truncated version of one of the new R2 ORF and one or more other heterologous elements. The truncated R2 ORF, in some embodiments, includes the reverse-transcriptase (or a biological variant thereof) of R2, but does not include portion of the N-terminal region of R2, such as one or more of the N-Znf1, N-Znf2, N-Znf3, and SANT DNA-Binding Domain.

[0035] As used herein, a "biological variant" of a reference protein refers to a protein or fragment that has at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to the reference protein. In some embodiments, the biological variant retains the desired activity (*e.g.*, reverse-transcriptase) of the reference protein. Likewise, a "biological variant" of a reference nucleic acid sequence refers to a nucleic acid sequence or fragment that has at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to the reference nucleic acid sequence.

A. Truncated R2 ORF

[0036] As provided, an engineered R2 protein of the present disclosure can include a truncated version of the R2 ORF fused to at least one heterologous element.

[0037] In one embodiment, the truncated R2 ORF includes a reverse-transcriptase comprising the amino acid sequence of any one of SEQ ID NO: 6, 2, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118, 122, 126, 130, 134, 138, 142, 146, 150, 154, 158, 162, 166, 170, 174, 178, 182, 186, 190, 194, 198, 202, 206, 210, 214, 218, 222, 226, 230, 234, 238, 242, 246, 250, 254, 258, 262, 266, 270, 274, 278, 282, 286, 290, 294, 298, 302, 306, 310, 314, 318, 322, 326, 330, 334, 338, 342, 346, 350, 354, 358, 362, 366, 370, 374, 378, 382, 386, 390, 394, 398, 402, 406, 410, 414, 418 and 422 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 6, 2, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118, 122, 126, 130, 134, 138, 142, 146, 150, 154, 158, 162, 166, 170, 174, 178, 182, 186, 190, 194, 198, 202, 206, 210, 214, 218, 222, 226, 230, 234, 238, 242, 246, 250, 254, 258, 262, 266, 270, 274, 278, 282, 286, 290, 294, 298, 302, 306, 310, 314, 318, 322, 326, 330, 334, 338, 342, 346, 350, 354, 358, 362, 366, 370, 374, 378, 382, 386, 390, 394, 398, 402, 406, 410, 414, 418 and 422.

[0038] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 2 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 2. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 6 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 6. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 10 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 10.

[0039] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 14 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 14. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 18 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 18. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 22 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 22.

[0040] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 26 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 26. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 30 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 30. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 34 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 34.

[0041] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 38 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 38. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 42 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 42. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 46 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 46.

[0042] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 50 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 50. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 54 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 54. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 58 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 58.

[0043] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 62 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 62. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 66 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to

SEQ ID NO: 66. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 70 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 70. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 74 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 74.

[0044] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 78 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 78. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 82 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 82. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 86 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 86.

[0045] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 90 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 90. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 94 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 94. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 98 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 98.

[0046] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 102 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 102. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 106 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 106. In one embodiment, the reverse-transcriptase includes the amino acid

sequence of SEQ ID NO: 110 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 110.

[0047] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 114 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 114. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 118 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 118. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 122 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 122.

[0048] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 126 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 126. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 130 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 130. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 134 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 134.

[0049] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 138 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 138. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 142 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 142. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 146 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 146.

[0050] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 150 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 150. In one embodiment, the reverse-

transcriptase includes the amino acid sequence of SEQ ID NO: 154 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 154. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 158 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 158.

[0051] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 162 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 162. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 166 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 166. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 170 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 170.

[0052] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 174 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 174. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 178 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 178. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 182 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 182.

[0053] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 186 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 186. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 190 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 190. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 194 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 194.

[0054] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 198 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 198. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 202 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 202. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 206 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 206.

[0055] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 210 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 210. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 214 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 214. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 218 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 218.

[0056] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 222 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 222. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 226 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 226. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 230 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 230.

[0057] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 234 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 234. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 238 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to

SEQ ID NO: 238. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 242 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 242.

[0058] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 246 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 246. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 250 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 250. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 254 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 254.

[0059] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 258 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 258. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 262 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 262. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 266 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 266.

[0060] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 270 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 270. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 274 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 274. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 278 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 278.

[0061] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 282 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%,

95%, 98% or 99% sequence identity to SEQ ID NO: 282. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 286 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 286. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 290 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 290.

[0062] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 294 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 294. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 298 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 298. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 302 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 302.

[0063] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 306 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 306. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 310 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 310. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 314 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 314.

[0064] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 318 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 318. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 322 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 322. In one embodiment, the reverse-transcriptase includes the amino acid

sequence of SEQ ID NO: 326 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 326.

[0065] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 330 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 330. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 334 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 334. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 338 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 338.

[0066] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 342 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 342. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 346 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 346. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 350 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 350.

[0067] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 354 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 354. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 358 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 358. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 362 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 362.

[0068] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 366 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 366. In one embodiment, the reverse-

transcriptase includes the amino acid sequence of SEQ ID NO: 370 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 370. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 374 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 374.

[0069] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 378 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 378. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 382 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 382. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 386 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 386.

[0070] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 390 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 390. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 394 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 394. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 398 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 398.

[0071] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 402 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 402. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 406 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 406. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 410 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 410.

[0072] In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 414 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 414. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 418 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 418. In one embodiment, the reverse-transcriptase includes the amino acid sequence of SEQ ID NO: 422 or an amino acid sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to SEQ ID NO: 422.

[0073] In some embodiments, the truncated R2 ORF further includes the 3' UTR binding domain of the R2 ORF. In some embodiments, the truncated R2 ORF further includes the template switching domain of the R2 ORF. In some embodiments, the truncated R2 ORF further includes the C-terminal ZnF. In some embodiments, the truncated R2 ORF further includes the endonuclease domain. In some embodiments, the truncated R2 ORF includes all elements of the wild-type R2 ORF, except those that are specifically substituted with a corresponding heterologous element (*e.g.*, a heterologous endonuclease).

B. Heterologous Targeting Domain

[0074] In some embodiments, one or more of the N-terminal ZnF domains is replaced with a heterologous targeting domain (*e.g.*, a DNA-binding domain or a CRISPR Cas protein).

[0075] A "DNA-binding domain" refers to a protein domain capable of binding a nucleic acid sequence of interest, preferably a double strand nucleic acid molecule. The DNA binding domain recognizes and binds nucleic acid at specific polynucleotide sequences, further referred to as "nucleic acid target sequence." The DNA-binding domain comprises any of the DNA-binding domain known in the art, such as transcription activator-like effector (TALE), homeodomain protein, a zinc finger, a helix-turn-helix, a leucine zipper, and a DNA-binding domain of a homing endonuclease, a transposon or retro-transposon. In certain embodiments, the DNA-binding domain recognizes a target DNA sequence in the genome.

[0076] A "homeodomain protein" consists of three linked alpha helices (helices 1, 2 and 3). Helices 2 and 3 are arranged in a conspicuous helix turn helix motif. A 60 amino acid long

region (homeodomain) within helix 3 binds specifically to DNA segments that contain the sequence 5'ATTA3'. In animals, there are 16 major classes of homeodomain protein, ANTP, PRD, PRD-LIKE, POU, HNF, CUT (with four subclasses: ONECUT, CUX, SATB, and CMP), LIM, ZF, CERS, PROS, SIX/SO, plus the TALE superclass with the classes IRO, MKX, TGIF, PBC, and MEIS. Further homeodomain proteins include but not limited to HEX motif, EH1 motif, Octapeptide/Hep/EH1/TN/GEH motif, WRPW motif, TUP1, OAR motif, CUT domain, such as ONECUT, CUX, SATB, CMP domains, COMPASS domain, HNF1A (LFB1) transcription factor, POU, OCT-1, OCT-2, Pit 1, LIM domain, PROSPERO (PROS), SIX/SO and CERS (LASS).

[0077] The "zinc finger" (ZF) includes at least nine types, C2H2, C2HC, C3H, C4, C6, C3HC4, C2HC5, C4HC3, and C8, in which C and H represent cysteine and histidine, respectively. C2H2 zinc finger is a loop of 12 amino acids with two cysteines and two histidines at the base of the loop (Cys2His2 zinc finger motif consisting of an α helix and an antiparallel β sheet) that tetrahedrally coordinate a zinc ion. C2H2 zinc finger is represented by TFIIIA and further comprises Adnp, Tshz, Zeb, Zfhx and Zhx. C2HC-typezinc finger domains, also referred to as retroviral-type (RT) zinc finger sequences, are required for viral genome packaging and RNA or single-stranded DNA binding in eukaryotes. The C3H family proteins are divided into 18 groups based on the different amino acid spacing numbers between C and H in zinc finger motif. C4 finger has one alpha helix and contains a zinc atom bound to four cysteine amino acids. The C4 zinc finger contains a 70 amino acid long region near the zinc atom that binds specifically to DNA segments. The C4 family includes seven types: GATA, FYVE, Tim10/DDP, LSD1, A20, TFIIB, and Zn-finger in Ran-binding protein. C3HC4-type finger, also termed RING finger, can be categorized into seven types with different conserved motifs, such as RING-H2, RING-HC, RING-v, RING-D, RING-S/T, RING-G and RING-C2. The C2HC5 motif, also referred to as LIM. More information on the zinc finger can be found in the prior art publication, such as Li et al., International Journal of Molecular Sciences, (2020) 21(4):1361. Further examples of the zinc finger proteins include Sp1, Glucocorticoid receptor, estrogen receptor, progesterone receptor, thyroid hormone receptor (erbA), retinoid acid receptor, and the vitamin D3 receptor.

[0078] A "leucine zipper" consists of an alpha helix that contains a region in which every seventh amino acid is leucine, which has the effect of lining up all the leucine residues on one side of the alpha helix. The leucine residues allow for the dimerization of the two lecine zipper proteins and formation of Y shaped dimer. Dimerization may occur between two of the same proteins (homodimers, e.g., Jun-Jun) or two different proteins (heterodimers, e.g., Fos-Jun). A leucine zipper contains a 20 amino acid long region that binds specifically to DNA segments. Examples of the Leucine Zipper proteins include but not limited to CCAAT/ enhancer binding protein (C/EBP), Cyclic AMP response element binding protein (CREB), Finkel osteogeneic sarcoma virus (Fos) protein, Jun, GCN4, and HSF.

[0079] A "helix loop helix" (HLH) or "helix-turn-helix" consists of a short alpha helix connected by a loop to a loner alpha helix. The loop allows for dimerization of two HLH proteins and formation of Y shaped dimer. The dimerization may occur between two of the same proteins (homo dimers) or two different proteins (heterodimers). Example of HLH domains include but not limited to MyoD, Myc, Pho4, SREBP-1a, and Max-Mad.

[0080] A "Transcription Activator like Effector" (TALE) are proteins that are encoded by phytopathogenic bacteria of the genus Xanthomonas and Ralstonia to influence the gene expression of host plant cells during bacterial infection. These proteins are helix-loop-helix-containing transcriptional factors that comprise a DNA binding region and an N-terminal domain that appears to interact with the bacterial transport machinery for introducing the protein into the plant cell. The C-terminal domain of the TALE protein seems to interact with the plant host's transcriptional machinery to induce expression of sets of plant genes that are beneficial to the invading bacteria. The DNA binding portion of the proteins is found in the middle section of the protein and is made of an array of repeat units, each approximately 33-35 amino acids in length, which have been shown to be responsible for interacting with the target DNA. In a preferred embodiment, said DNA-binding domain is derived from a TALE engineered to bind a specific nucleic acid target sequence. Unlimited examples of the TALE include KNOX and BEL, such as PBC, MEIS, PREP, IRO, MKX and TGIF.

[0081] A "homing endonuclease (HEs)" or meganuleases (MNs), are sequence-specific endonucleases with large cleavage sites (14–25 bp) that can create double-stranded breaks at

specific locations. The homing endonucleases are encoded by mobile genetic elements that induce recombination in a process called homing. The homing endonucleases genes (HEGs) have invaded many genomic niches including group I and group II introns. Some HEGs can move from an ORF-containing intron to an "ORF-less" intron. There are four major families of homing endonucleases (HEs), naming is based on conserved amino acid motifs: the H-N-H, HIS-CYS, LAGLIDADG, and GIY-YIG families of HEs. The LAGLIDADG family of HEs is the most frequently encountered among group I introns. However GIY-YIG endonucleases have been identified within numerous group I introns, and LAGLIDADG HEs and the H-N-H domain are present within the ORFs of some group II introns. Additional HE-like proteins have been described, the PD-(D/E)XK HEs are found in bacterial tRNA group I introns, the very-short patch repair (Vsr) endonucleases (a predicted family of phage HEs based on metagenomic), and the Holliday junction resolvase-like HEs found in some phage introns. In certain embodiments of the present disclosure, the homing endonuclease comprises PI-SceI, PI-PfuI, I-CreI, I-CeuI, I-DmoI, F-TevI, F-TevII, I-TevI, I-TevII, I-PpoI, I-DirI, I-NjaI, I-NanI, I-NitI, I-SceV, I-SceVI and I-LlaI, I-HmuI, I-HmuII, I-TevIII and I-CmoeI. In a particular embodiment, the homing endonuclease is I-PpoI (Intron-Encoded Endonuclease).

[0082] Retrotransposon is a class of eucaryotic genes capable of replicating to new locations within their own genome through an RNA intermediate. Retrotransposons are widespread in metazoan genomes. The proportion of the genome that contains retrotransposons often exceeds that containing DNA transposons, particularly in higher vertebrates such as humans. Most retrotransposons integrate into random sites of the host genome, but some have a sequence preference. In particular, a few non-LTR subclades integrate into the genome in a highly sequence-specific manner. By leaving their 5' and/or 3' ends of genomic copies (usually with a poly (A) stretch and target-site duplication (TSD)), the site-specific non-LTR retrotransposons can be searched throughout the DNA database.

[0083] There are five clades of restriction enzyme-like endonuclease (RLE)-encoding elements based on their RT sequence similarity. Most of three clades (NeSL, R2, and R4) are target specific and two clades (HERO and CRE) have some site-specific elements. R2 (R2 clade) was found at a specific sequence in the 28S rDNA of many invertebrate and vertebrate species. R4 in

Ascaris lumbricoides and Dong (R4 clade) were found at another site of the 28S rDNA and microsatellite TAA repeats, respectively.

[0084] In contrast to RLE-encoding elements, most of the apurinic/apyrimidinic endonuclease (APE)-encoding non-LTR elements do not insert themselves in a sequence-specific manner, but do have weak target-site specificity, *e.g.*, human L1 for TAAA repeats. However, two clades of APE-encoding non-LTR elements, Tx1 and R1, are known to be sequence-specific.

[0085] Among the site-specific non-LTR retrotransposons, R1 clade (R1/R6/R7/RT), R2 clade (R2/R8/R9), R4 clade (R4) and NeSl clade (R5) target the rRNA genes, whereas R7 and R8 target the 18S rDNA, and R1, R2, R4, R5, R6, R9, and RT target the 28S rDNA. All R-element target sites within 28S and 18S rDNA are highly conserved among organisms (see Fujiwara et al., Microbiology Spectrum, (2015) Vol. 3, Issue 2).

[0086] In a preferred embodiment, the retrotransposon is rDNA specific and comprises R1, R2, R4, R5, R6, R7, R8, R9 and RT. In certain embodiments, the heterologous DNA-binding domain provided herein comprises the DNA-binding domain of any one of the R1, R2, R4, R5, R6, R7, R8, R9 and RT.

[0087] In certain embodiments, the one or more N-terminal ZnF domains is replaced with a CRISPR Cas protein. In certain embodiments, the CRISPR Cas protein comprises Cas9, Cas12a/Cpf1 and Cas13. The clustered regularly interspaced short palindromic repeat (CRISPR)/Cas mediated genome editing technology has seen wild use since its invention due to its simplicity and efficiency, leading to a new era of genetic engineering and gene therapy. The targeting and cleavage are achieved by a single CRISPR RNA ("crRNA")-bound Cas protein in Class 2 CRISPR-Cas systems. Class 2 systems can be subdivided into Type II Cas9 and Type V Cas12 that target DNA, as well as Type VI Cas13 that target RNA. Cas9 and Cas12a/Cpf1 have been well studied in the past and broadly harnessed for gene editing in various cell types and organisms in prokaryotes and eukaryotes. Both Cas9 and Cas12 systems utilize guide RNA to recognize the target site and protospacer adjacent motifs (PAMs) to determine the cleavage site and generate double strand breaks. However, the binding and cleavage of DNA by Cas9 and Cas12 are quite different. Cas9 recognizes a 3'-G-rich PAM and produces blunt ends cleaved by the RuvC and HNH domains, whereas Cas12 recognizes a 5' T-rich PAM and produces

staggered ends cleaved solely by the RuvC domain. Pre-assembled Cas 13 and crRNA recognizes target RNAs. Upon RNA-binding, Cas13 will undergo a conformational change and induce the catalytic activity of its nuclease domains, resulting in the cleavage of target transcripts. In certain embodiments, the gRNA or the crRNA can be designed to target a DNA sequence in the genome. In certain embodiments, the Cas protein comprises an inactive nuclease domain. In certain embodiments, the Cas protein does not comprise an active nuclease domain.

[0088] In certain embodiments, the guide RNA ("gRNA") comprises two short, non-coding RNA species referred to as crRNA and trans-acting RNA ("tracrRNA"). In an exemplary system, the gRNA forms a complex with a Cas protein of the present disclosure. The gRNA:Cas protein complex binds a target polynucleotide sequence in the genome. The gRNA can be supplied separately that works with the Cas protein in guiding the Cas protein (or the fusion) to a target location on a genomic sequence.

[0089] In one aspect of the present disclosure, a fusion protein provided herein that comprises a first fragment that is a truncated R2 protein and a second fragment that can be a DNA-binding domain or a CRISPR Cas protein. In certain embodiments, the second fragment is N-terminal to the first fragment. In certain embodiments, the second fragment can be alternatively C-terminal to the first fragment. The first and second fragments can be fused directly or via a linker sequence.

C. Heterologous Endonuclease or Nickase

[0090] In some embodiments, the native endonuclease of the R2 ORF can be replaced with a biological variant (*e.g.*, a sequence having at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% sequence identity to the native endonuclease) or a heterologous endonuclease or a nickase.

[0091] An endonuclease is an enzyme that cleaves internal phosphodiester bonds of polynucleotides. These enzymes are either specific or non-specific to the sequences being cleaved. The endonucleases that are specific to a particular sequence are termed restriction endonucleases, which cleave large DNA molecules at specific sequences of four to six nucleotides.

[0092] In some embodiments the endonuclease is a eukaryotic endonuclease. Non-limiting examples include Neurospora endonuclease, S1 nuclease, P1-nuclease, Mung bean nuclease I, Ustilago nuclease, DNase I, AP endonuclease and Endo R. Other examples include Fok1 nuclease, type-II restriction 1-like endonucleases (RLE-type nuclease), and RLE-type endonucleases (REL).

[0093] Many known transposons and retrotransposons also include related endonucleases. In some embodiments, the heterologous endonuclease is one from R1, R2, R4, R5, R6, R7, R8, R9 or RT.

[0094] In some embodiments, the R2 endonuclease is substituted with a nickase. A nikcase (or nicking endonuclease) is an enzyme that cuts one strand of a double-stranded DNA at a specific recognition nucleotide sequence, to produce DNA molecules that are nicked. Nickases can be prepared from Cas enzyme with site mutations (*e.g.*, D10A Cas9).

[0095] The heterologous DNA-binding domain/Cas and/or the heterologous endonuclease can be used to the truncated R2 ORF through a linker sequence. The term "linker sequence" refers to an innocuous length of nucleic acid or protein that joins two other sections of nucleic acid or protein.

[0096] Linkers within the scope of the present disclosure are characterized in terms of amino acid content, length, rigidity and secondary structure. Linkers within the scope of the present disclosure separate the first fragment and the second fragment and allow proper folding and functioning of each domain. In this manner, a linker can be tailored to the particular first and second fragments. According to one aspect, functional independence of the structural and fused (heterologous) domains is maximized by a suitable linker to limit steric interference between domains during the export and assembly processes of the bacterial cell. According to an additional aspect, cell stress is minimized by limiting the overall length of the fusion protein. Longer linker sequences and higher induction levels stress the biosynthetic machinery of the cells, inhibiting cell growth and leading to cell lysis in extreme cases.

[0097] Linkers within the scope of the present disclosure facilitate functioning of the first and second fragments. Linkers within the scope of the present disclosure allow efficient protein

processing and export through the bacterial curli secretion machinery as well as provide the proper spatial and physicochemical separation of the polypeptides to retain their respective functions.

[0098] Linkers within the scope of the present disclosure include amino acid residues. The amino acid residues may be any of the naturally occurring amino acid residues. Amino acid residues may also be synthetic amino acids known to those of skill in the art. Representative amino acids which may be used in linkers include Glycine, Alanine, Valine, Leucine, Isoleucine, Serine, Cysteine, Selenocysteine, Threonine, Methionine, Proline, Phenylalanine, Tyrosine, Tryptophan, Histidine, Lysine, Arginine, Aspartate, Glutamate, Asparagine, and Glutamine.

[0099] According to one aspect, a linker sequence is a polypeptide sequence of at least 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 24, 48 or more amino acids. In some embodiments, the linker sequence comprises from about 3 amino acids to about 50 amino acids. In some embodiments, the linker sequence comprises from about 3 amino acids to about 40 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 30 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 25 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 24 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 23 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 22 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 21 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 20 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 19 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 18 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 17 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 16 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 15 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 14 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 13 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 12 amino acids. In some embodiments, the linker sequence comprises

from about 5 amino acids to about 11 amino acids. In some embodiments, the linker sequence comprises from about 5 amino acids to about 10 amino acids.

[0100] In some embodiments, the linker sequence comprises a flexible polypeptide, *e.g.*, a polypeptide not having a rigid secondary and/or tertiary structure. In some embodiments, the linker sequence comprises glycine and serine residues. In some embodiments at least 50% of the amino acids comprised by the linker sequence are glycine or serine residues, *e.g.* at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or more are glycine or serine residues. In some embodiments, the linker sequence consists of glycine and serine residues. In some embodiments, the linker sequence comprises a rigid polypeptide.

[0101] In some embodiments, the truncated R2 ORF and the one or more heterologous elements are not delivered and expressed together. Instead, they can be delivered and expressed separately, or are conjugated or bound to one another.

[0102] Therefore, in one aspect, the present disclosure provides a complex comprising a first polypeptide comprising the truncated R2 ORF and a second polypeptide comprising the targeting domain (*e.g.*, a DNA-binding domain or a CRISPR Cas protein) and/or the endonuclease/nickase, wherein the first polypeptide is conjugated or bound to the second polypeptide. The term "conjugate" or "bind" as used herein refers to conjugation of a first conjugation moiety and a second conjugation moiety which form a complex and stabilized by chemical reactions with few of several available chemical groups (chemical conjugation) or via one or more covalent bonds (biological conjugation).

[0103] In certain embodiments, the first polypeptide includes a first conjugation moiety and the second polypeptide comprises a second conjugation moiety, wherein the first polypeptide and the second polypeptide are brought together by the conjugation moieties and form a stable complex.

[0104] In certain embodiments, the first polypeptide and the second polypeptide may respectfully but regardless of order, further include (a) a receptor and a corresponding ligand, (b) a peptide and a corresponding antibody, or (c) an enzyme and a corresponding substrate.

Protein-Coding Sequence for Delivery

[0105] In one embodiment, the R2 ORF, or an engineered version thereof, can be provided as a mRNA. The mRNA can further include the corresponding 5'UTR and/or 3'UTR of the native R2.

[0106] In some embodiments, to insert a different sequence (*e.g.*, a second protein-coding sequence) to a target genome, the mRNA can further include that second protein-coding sequence. In some embodiments, the second protein-coding sequence is 3' to the R2 ORF. In some embodiments, the second protein-coding sequence is included in the mRNA in a reverse-complement orientation such that it cannot be directly translated by the mRNA. Rather it is translated after integration to the target genome. In some embodiments, the second protein-coding sequence is inserted at 3' to the R2 ORF via a self-cleaving peptide sequence, such as 2A self-cleaving peptide.

[0107] Alternatively, in some embodiments, the second protein-coding sequence is provided on a separate mRNA that flanks the second protein-coding sequence with the 5'UTR and/or 3'UTR of the native R2.

[0108] In some embodiments, the second protein is a therapeutic protein that can be helpful in treating a disease or condition in a patient. Examples include:

- Insulin - used to treat diabetes
- Phenylalanine hydroxylase - used to treat PKU (Phenylketonuria) disease
- Erythropoietin - used to treat anemia
- Filgrastim (G-CSF) - used to treat neutropenia
- Interferons - used to treat cancer, multiple sclerosis, and viral infections
- Tumor necrosis factor (TNF) inhibitors - used to treat autoimmune diseases, such as rheumatoid arthritis and Crohn's disease
- Monoclonal antibodies - used to treat cancer, autoimmune diseases, and some viral infections
- Follitropin alfa - used to treat infertility
- Coagulation factor - used to treat bleeding disorders, such as Hemophilia

- Enzyme replacement therapy enzymes - used to treat lysosomal storage disorders, such as Gaucher's disease
- Growth hormone - used to treat growth hormone deficiency.

[0109] In some embodiments, the polynucleotide (*e.g.*, mRNA) further includes a downstream flanking sequence that is 3' to the protein-coding sequence. Addition of a downstream flanking sequence can improve the retrotransposition efficiency. In certain embodiments, the protein-coding sequence may comprise, at the 3' end of the gene encoding sequence, one or more consecutive nucleotides (rDNA sequences) of the "downstream sequence" starting from the EN cleavage site. The protein-coding sequence may comprise, at the 3' end of the gene encoding sequence, different lengths of the downstream sequence depending on the type of polynucleotide or vector used for delivery. In certain embodiments, a length of the downstream sequence of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 200, 250 nt can be included to the 3' mRNA of the fusion protein-encoding sequence. In some embodiments, the length can be 1-250, 1-200, 4-180, 10-150, 15-150, 40-90, 50-90, 70-90 or 71-90 nt, without limitation.

[0110] In some embodiments, the polynucleotide (*e.g.*, mRNA) further includes an upstream flanking sequence that is 5' to the protein-coding sequence. Addition of an upstream flanking sequence may also improve the retrotransposition efficiency. In certain embodiments, the protein-coding sequence may comprise, at the 5' end of the gene encoding sequence, one or more consecutive nucleotides (rDNA sequences) of the "upstream sequence" starting from the EN cleavage site. The protein-coding sequence may comprise, at the 5' end of the gene encoding sequence, different lengths of the downstream sequence depending on the type of polynucleotide or vector used for delivery. In certain embodiments, a length of the downstream sequence of at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 200, 250 nt can be included to the 5' mRNA of the fusion protein-encoding sequence. In some embodiments, the length can be 1-250, 1-200, 4-180, 10-150, 15-150, 40-90, 50-90, 70-90 or 71-90 nt, without limitation.

[0111] In some embodiments, the R2 3'UTR has the nucleic acid sequence of any one of SEQ ID NO: 8, 4, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96,

100, 104, 108, 112, 116, 120, 124, 128, 132, 136, 140, 144, 148, 152, 156, 160, 164, 168, 172, 176, 180, 184, 188, 192, 196, 200, 204, 208, 212, 216, 220, 224, 228, 232, 236, 240, 244, 248, 252, 256, 260, 264, 268, 272, 276, 280, 284, 288, 292, 296, 300, 304, 308, 312, 316, 320, 324, 328, 332, 336, 340, 344, 348, 352, 356, 360, 364, 368, 372, 376, 380, 384, 388, 392, 396, 400, 404, 408, 412, 416, 420 and 424. In some embodiments, the 3'UTR has at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% sequence identity to any one of SEQ ID NO: 8, 4, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 92, 96, 100, 104, 108, 112, 116, 120, 124, 128, 132, 136, 140, 144, 148, 152, 156, 160, 164, 168, 172, 176, 180, 184, 188, 192, 196, 200, 204, 208, 212, 216, 220, 224, 228, 232, 236, 240, 244, 248, 252, 256, 260, 264, 268, 272, 276, 280, 284, 288, 292, 296, 300, 304, 308, 312, 316, 320, 324, 328, 332, 336, 340, 344, 348, 352, 356, 360, 364, 368, 372, 376, 380, 384, 388, 392, 396, 400, 404, 408, 412, 416, 420 and 424. In some embodiments, the 3'UTR is directly connected to the downstream flanking sequence.

[0112] In some embodiments, the polynucleotide (*e.g.*, mRNA) further includes a 5' untranslated region (UTR) 5' to the protein-coding sequence. When an upstream flanking sequence is also included, preferably the upstream flanking sequence is 5' to the 5'UTR.

[0113] In some embodiments, the R2 5'UTR has the nucleic acid sequence of any one of SEQ ID NO: 7, 3, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63, 67, 71, 75, 79, 83, 87, 91, 95, 99, 103, 107, 111, 115, 119, 123, 127, 131, 135, 139, 143, 147, 151, 155, 159, 163, 167, 171, 175, 179, 183, 187, 191, 195, 199, 203, 207, 211, 215, 219, 223, 227, 231, 235, 239, 243, 247, 251, 255, 259, 263, 267, 271, 275, 279, 283, 287, 291, 295, 299, 303, 307, 311, 315, 319, 323, 327, 331, 335, 339, 343, 347, 351, 355, 359, 363, 367, 371, 375, 379, 383, 387, 391, 395, 399, 403, 407, 411, 415, 419 and 423. In some embodiments, the 5'UTR has at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% sequence identity to any one of SEQ ID NO: 7, 3, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63, 67, 71, 75, 79, 83, 87, 91, 95, 99, 103, 107, 111, 115, 119, 123, 127, 131, 135, 139, 143, 147, 151, 155, 159, 163, 167, 171, 175, 179, 183, 187, 191, 195, 199, 203, 207, 211, 215, 219, 223, 227, 231, 235, 239, 243, 247, 251, 255, 259, 263, 267, 271, 275, 279, 283, 287, 291, 295, 299, 303, 307, 311, 315, 319, 323, 327, 331, 335, 339, 343, 347, 351, 355, 359, 363, 367, 371, 375, 379, 383, 387, 391, 395, 399, 403, 407, 411, 415, 419 and 423.

[0114] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:2 or a biological variant, or the R2 ORF includes SEQ ID NO: 1 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 4 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 3 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:6 or a biological variant, or the R2 ORF includes SEQ ID NO: 5 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 8 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 7 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:10 or a biological variant, or the R2 ORF includes SEQ ID NO: 9 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 12 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 11 or a biological variant.

[0115] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:14 or a biological variant, or the R2 ORF includes SEQ ID NO: 13 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 16 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 15 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:18 or a biological variant, or the R2 ORF includes SEQ ID NO: 17 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 20 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 19 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:22 or a biological variant, or the R2 ORF includes SEQ ID NO: 21 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 24 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 23 or a biological variant.

[0116] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:26 or a biological variant, or the R2 ORF includes SEQ ID NO: 25 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 28 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 27 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID

NO:30 or a biological variant, or the R2 ORF includes SEQ ID NO: 29 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 32 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 31 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:34 or a biological variant, or the R2 ORF includes SEQ ID NO: 33 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 36 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 35 or a biological variant.

[0117] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:38 or a biological variant, or the R2 ORF includes SEQ ID NO: 37 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 40 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 39 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:42 or a biological variant, or the R2 ORF includes SEQ ID NO: 41 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 44 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 43 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:46 or a biological variant, or the R2 ORF includes SEQ ID NO: 45 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 48 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 47 or a biological variant.

[0118] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:50 or a biological variant, or the R2 ORF includes SEQ ID NO: 49 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 52 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 51 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:54 or a biological variant, or the R2 ORF includes SEQ ID NO: 53 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 56 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 55 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID

NO:58 or a biological variant, or the R2 ORF includes SEQ ID NO: 57 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 60 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 59 or a biological variant.

[0119] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:62 or a biological variant, or the R2 ORF includes SEQ ID NO: 61 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 64 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 63 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:66 or a biological variant, or the R2 ORF includes SEQ ID NO: 65 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 68 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 67 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:70 or a biological variant, or the R2 ORF includes SEQ ID NO: 69 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 72 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 71 or a biological variant.

[0120] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:74 or a biological variant, or the R2 ORF includes SEQ ID NO: 73 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 76 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 75 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:78 or a biological variant, or the R2 ORF includes SEQ ID NO: 77 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 80 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 79 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:82 or a biological variant, or the R2 ORF includes SEQ ID NO: 81 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 84 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 83 or a biological variant.

[0121] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:86 or a biological variant, or the R2 ORF includes SEQ ID NO: 85 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 88 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 87 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:90 or a biological variant, or the R2 ORF includes SEQ ID NO: 89 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 92 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 91 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:94 or a biological variant, or the R2 ORF includes SEQ ID NO: 93 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 96 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 95 or a biological variant.

[0122] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:98 or a biological variant, or the R2 ORF includes SEQ ID NO: 97 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 100 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 99 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:102 or a biological variant, or the R2 ORF includes SEQ ID NO: 101 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 104 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 103 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:106 or a biological variant, or the R2 ORF includes SEQ ID NO: 105 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 108 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 107 or a biological variant.

[0123] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:110 or a biological variant, or the R2 ORF includes SEQ ID NO: 109 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 112 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 111 or a

biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:114 or a biological variant, or the R2 ORF includes SEQ ID NO: 113 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 116 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 115 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:118 or a biological variant, or the R2 ORF includes SEQ ID NO: 117 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 120 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 119 or a biological variant.

[0124] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:122 or a biological variant, or the R2 ORF includes SEQ ID NO: 121 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 124 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 123 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:126 or a biological variant, or the R2 ORF includes SEQ ID NO: 125 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 128 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 127 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:130 or a biological variant, or the R2 ORF includes SEQ ID NO: 129 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 132 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 131 or a biological variant.

[0125] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:134 or a biological variant, or the R2 ORF includes SEQ ID NO: 133 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 136 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 135 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:138 or a biological variant, or the R2 ORF includes SEQ ID NO: 137 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 140 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 139 or a

biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:142 or a biological variant, or the R2 ORF includes SEQ ID NO: 141 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 144 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 143 or a biological variant.

[0126] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:146 or a biological variant, or the R2 ORF includes SEQ ID NO: 145 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 148 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 147 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:150 or a biological variant, or the R2 ORF includes SEQ ID NO: 149 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 152 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 151 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:154 or a biological variant, or the R2 ORF includes SEQ ID NO: 153 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 156 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 155 or a biological variant.

[0127] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:158 or a biological variant, or the R2 ORF includes SEQ ID NO: 157 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 160 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 159 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:162 or a biological variant, or the R2 ORF includes SEQ ID NO: 161 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 164 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 163 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:166 or a biological variant, or the R2 ORF includes SEQ ID NO: 165 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 168 or a biological variant. In some

embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 167 or a biological variant.

[0128] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:170 or a biological variant, or the R2 ORF includes SEQ ID NO: 169 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 172 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 171 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:174 or a biological variant, or the R2 ORF includes SEQ ID NO: 173 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 176 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 175 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:178 or a biological variant, or the R2 ORF includes SEQ ID NO: 177 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 180 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 179 or a biological variant.

[0129] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:182 or a biological variant, or the R2 ORF includes SEQ ID NO: 181 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 184 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 183 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:186 or a biological variant, or the R2 ORF includes SEQ ID NO: 185 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 188 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 187 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:190 or a biological variant, or the R2 ORF includes SEQ ID NO: 189 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 192 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 191 or a biological variant.

[0130] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:194 or a biological variant, or the R2 ORF includes SEQ ID NO: 193 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 196 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 195 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:198 or a biological variant, or the R2 ORF includes SEQ ID NO: 197 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 200 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 199 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:202 or a biological variant, or the R2 ORF includes SEQ ID NO: 201 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 204 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 203 or a biological variant.

[0131] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:206 or a biological variant, or the R2 ORF includes SEQ ID NO: 205 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 208 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 207 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:210 or a biological variant, or the R2 ORF includes SEQ ID NO: 209 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 212 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 211 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:214 or a biological variant, or the R2 ORF includes SEQ ID NO: 213 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 216 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 215 or a biological variant.

[0132] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:218 or a biological variant, or the R2 ORF includes SEQ ID NO: 217 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 220 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 219 or a

biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:222 or a biological variant, or the R2 ORF includes SEQ ID NO: 221 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 224 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 223 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:226 or a biological variant, or the R2 ORF includes SEQ ID NO: 225 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 228 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 227 or a biological variant.

[0133] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:230 or a biological variant, or the R2 ORF includes SEQ ID NO: 229 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 232 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 231 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:234 or a biological variant, or the R2 ORF includes SEQ ID NO: 233 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 236 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 235 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:238 or a biological variant, or the R2 ORF includes SEQ ID NO: 237 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 240 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 239 or a biological variant.

[0134] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:242 or a biological variant, or the R2 ORF includes SEQ ID NO: 241 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 244 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 243 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:246 or a biological variant, or the R2 ORF includes SEQ ID NO: 245 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 248 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 247 or a

biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:250 or a biological variant, or the R2 ORF includes SEQ ID NO: 249 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 252 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 251 or a biological variant.

[0135] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:254 or a biological variant, or the R2 ORF includes SEQ ID NO: 253 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 256 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 255 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:258 or a biological variant, or the R2 ORF includes SEQ ID NO: 257 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 260 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 259 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:262 or a biological variant, or the R2 ORF includes SEQ ID NO: 261 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 264 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 263 or a biological variant.

[0136] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:266 or a biological variant, or the R2 ORF includes SEQ ID NO: 265 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 268 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 267 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:270 or a biological variant, or the R2 ORF includes SEQ ID NO: 269 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 272 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 271 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:274 or a biological variant, or the R2 ORF includes SEQ ID NO: 273 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 276 or a biological variant. In some

embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 275 or a biological variant.

[0137] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:278 or a biological variant, or the R2 ORF includes SEQ ID NO: 277 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 280 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 279 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:282 or a biological variant, or the R2 ORF includes SEQ ID NO: 281 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 284 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 283 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:286 or a biological variant, or the R2 ORF includes SEQ ID NO: 285 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 288 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 287 or a biological variant.

[0138] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:290 or a biological variant, or the R2 ORF includes SEQ ID NO: 289 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 292 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 291 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:294 or a biological variant, or the R2 ORF includes SEQ ID NO: 293 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 296 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 295 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:298 or a biological variant, or the R2 ORF includes SEQ ID NO: 297 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 300 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 299 or a biological variant.

[0139] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:302 or a biological variant, or the R2 ORF includes SEQ ID NO: 301 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 304 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 303 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:306 or a biological variant, or the R2 ORF includes SEQ ID NO: 305 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 308 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 307 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:310 or a biological variant, or the R2 ORF includes SEQ ID NO: 309 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 312 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 311 or a biological variant.

[0140] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:314 or a biological variant, or the R2 ORF includes SEQ ID NO: 313 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 316 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 315 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:318 or a biological variant, or the R2 ORF includes SEQ ID NO: 317 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 320 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 319 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:322 or a biological variant, or the R2 ORF includes SEQ ID NO: 321 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 324 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 323 or a biological variant.

[0141] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:326 or a biological variant, or the R2 ORF includes SEQ ID NO: 325 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 328 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 327 or a

biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:330 or a biological variant, or the R2 ORF includes SEQ ID NO: 329 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 332 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 331 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:334 or a biological variant, or the R2 ORF includes SEQ ID NO: 333 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 336 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 335 or a biological variant.

[0142] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:338 or a biological variant, or the R2 ORF includes SEQ ID NO: 337 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 340 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 339 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:342 or a biological variant, or the R2 ORF includes SEQ ID NO: 341 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 344 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 343 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:346 or a biological variant, or the R2 ORF includes SEQ ID NO: 345 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 348 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 347 or a biological variant.

[0143] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:350 or a biological variant, or the R2 ORF includes SEQ ID NO: 349 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 352 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 351 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:354 or a biological variant, or the R2 ORF includes SEQ ID NO: 353 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 356 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 355 or a

biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:358 or a biological variant, or the R2 ORF includes SEQ ID NO: 357 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 360 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 359 or a biological variant.

[0144] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:362 or a biological variant, or the R2 ORF includes SEQ ID NO: 361 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 364 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 363 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:366 or a biological variant, or the R2 ORF includes SEQ ID NO: 365 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 368 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 367 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:370 or a biological variant, or the R2 ORF includes SEQ ID NO: 369 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 372 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 371 or a biological variant.

[0145] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:374 or a biological variant, or the R2 ORF includes SEQ ID NO: 373 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 376 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 375 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:378 or a biological variant, or the R2 ORF includes SEQ ID NO: 377 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 380 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 379 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:382 or a biological variant, or the R2 ORF includes SEQ ID NO: 381 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 384 or a biological variant. In some

embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 383 or a biological variant.

[0146] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:386 or a biological variant, or the R2 ORF includes SEQ ID NO: 385 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 388 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 387 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:390 or a biological variant, or the R2 ORF includes SEQ ID NO: 389 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 392 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 391 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:394 or a biological variant, or the R2 ORF includes SEQ ID NO: 393 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 396 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 395 or a biological variant.

[0147] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:398 or a biological variant, or the R2 ORF includes SEQ ID NO: 397 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 400 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 399 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:402 or a biological variant, or the R2 ORF includes SEQ ID NO: 401 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 404 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 403 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:406 or a biological variant, or the R2 ORF includes SEQ ID NO: 405 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 408 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 407 or a biological variant.

[0148] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:410 or a biological variant, or the R2 ORF includes SEQ ID NO: 409 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 412 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 411 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:414 or a biological variant, or the R2 ORF includes SEQ ID NO: 413 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 416 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 415 or a biological variant. In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:418 or a biological variant, or the R2 ORF includes SEQ ID NO: 417 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 420 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 419 or a biological variant.

[0149] In some embodiments, the R2 ORF includes an RT that includes SEQ ID NO:422 or a biological variant, or the R2 ORF includes SEQ ID NO: 421 or a biological variant, and the corresponding R2 3'UTR includes SEQ ID NO: 424 or a biological variant. In some embodiments, a corresponding R2 5'UTR is used, which includes SEQ ID NO: 423 or a biological variant.

Gene Delivery and Treatments

[0150] The R2 retrotransposition systems disclosed here are suitable for delivering exogenous genes to a target cell, such as a mammalian cell.

[0151] Accordingly, in some embodiments, the present disclosure provides a method for integrating an exogenous polynucleotide to the genome of a target cell. In one embodiment, the method entails introducing to the cell the polynucleotide comprising an R2 retrotransposon of the instant disclosure. In some embodiments, the retrotransposon is delivered to the target cell as an mRNA molecule directly.

[0152] In some embodiments, when the polynucleotide comprises a CRISPR Cas protein encoding sequence, the method further comprises introducing to the cell a guide RNA (gRNA) that binds to the CRISPR Cas protein and recognizes a target DNA sequence in the genome.

[0153] mRNAs may be synthesized according to any of a variety of known methods. For example, the mRNAs may be synthesized via *in vitro* transcription (IVT). Briefly, IVT is typically performed with a linear or circular DNA template containing a promoter, a pool of ribonucleotide triphosphates, a buffer system that may include DTT and magnesium ions, and an appropriate RNA polymerase (*e.g.*, T3, T7 or SP6 RNA polymerase), DNase I, pyrophosphatase, and/or RNase inhibitor. The exact conditions will vary according to the specific application.

[0154] The mRNA may be synthesized as unmodified or modified mRNA. Typically, mRNAs are modified to enhance stability. Modifications of mRNA can include, for example, modifications of the nucleotides of the RNA. A modified mRNA can thus include, for example, backbone modifications, sugar modifications or base modifications. In some embodiments, antibody encoding mRNAs (*e.g.*, heavy chain and light chain encoding mRNAs) may be synthesized from naturally occurring nucleotides and/or nucleotide analogues (modified nucleotides) including, but not limited to, purines (adenine (A), guanine (G)) or pyrimidines (thymine (T), cytosine (C), uracil (U)), and as modified nucleotides analogues or derivatives of purines and pyrimidines, such as *e.g.* 1-methyl-adenine, 2-methyl-adenine, 2-methylthio-N-6-isopentenyl-adenine, N6-methyl-adenine, N6-isopentenyl-adenine, 2-thio-cytosine, 3-methyl-cytosine, 4-acetyl-cytosine, 5-methyl-cytosine, 2,6-diaminopurine, 1-methyl-guanine, 2-methyl-guanine, 2,2-dimethyl-guanine, 7-methyl-guanine, inosine, 1-methyl-inosine, pseudouracil (5-uracil), dihydro-uracil, 2-thio-uracil, 4-thio-uracil, 5-carboxymethylaminomethyl-2-thio-uracil, 5-(carboxyhydroxymethyl)-uracil, 5-fluoro-uracil, 5-bromo-uracil, 5-carboxymethylaminomethyl-uracil, 5-methyl-2-thio-uracil, 5-methyl-uracil, N-uracil-5-oxyacetic acid methyl ester, 5-methylaminomethyl-uracil, 5-methoxyaminomethyl-2-thio-uracil, 5'-methoxycarbonylmethyl-uracil, 5-methoxy-uracil, uracil-5-oxyacetic acid methyl ester, uracil-5-oxyacetic acid (v), 1-methyl-pseudouracil, queosine, 13-D-mannosyl-queosine, wybutoxosine, and phosphoramidates, phosphorothioates, peptide nucleotides, methylphosphonates, 7-deazaguanosine, 5-methylcytosine and inosine. The preparation of such analogues is known to a person skilled in the art *e.g.* from the U.S. Pat. Nos. 4,373,071, 4,401,796, 4,415,732, 4,458,066,

4,500,707, 4,668,777, 4,973,679, 5,047,524, 5,132,418, 5,153,319, 5,262,530 and 5,700,642, the disclosure of which is included here in its full scope by reference.

[0155] In some embodiments, the mRNAs may contain RNA backbone modifications. Typically, a backbone modification is a modification in which the phosphates of the backbone of the nucleotides contained in the RNA are modified chemically. Exemplary backbone modifications typically include, but are not limited to, modifications from the group consisting of methylphosphonates, methylphosphoramidates, phosphoramidates, phosphorothioates (*e.g.* cytidine 5'-O-(1-thiophosphate)), boranophosphates, positively charged guanidinium groups etc., which means by replacing the phosphodiester linkage by other anionic, cationic or neutral groups.

[0156] In some embodiments, the mRNAs may contain sugar modifications. A typical sugar modification is a chemical modification of the sugar of the nucleotides it contains including, but not limited to, sugar modifications chosen from the group consisting of 2'-deoxy-2'-fluoro-oligoribonucleotide (2'-fluoro-2'-deoxycytidine 5'-triphosphate, 2'-fluoro-2'-deoxyuridine 5'-triphosphate), 2'-deoxy-2'-deamine-oligoribonucleotide (2'-amino-2'-deoxycytidine 5'-triphosphate, 2'-amino-2'-deoxyuridine 5'-triphosphate), 2'-O-alkyloligoribonucleotide, 2'-deoxy-2'-C-alkyloligoribonucleotide (2'-O-methylcytidine 5'-triphosphate, 2'-methyluridine 5'-triphosphate), 2'-C-alkyloligoribonucleotide, and isomers thereof (2'-aracytidine 5'-triphosphate, 2'-arauridine 5'-triphosphate), or azidotriphosphates (2'-azido-2'-deoxycytidine 5'-triphosphate, 2'-azido-2'-deoxyuridine 5'-triphosphate).

[0157] In some embodiments, the mRNAs may contain modifications of the bases of the nucleotides (base modifications). A modified nucleotide which contains a base modification is also called a base-modified nucleotide. Examples of such base-modified nucleotides include, but are not limited to, 2-amino-6-chloropurine riboside 5'-triphosphate, 2-aminoadenosine 5'-triphosphate, 2-thiocytidine 5'-triphosphate, 2-thiouridine 5'-triphosphate, 4-thiouridine 5'-triphosphate, 5-aminoallylcytidine 5'-triphosphate, 5-aminoallyluridine 5'-triphosphate, 5-bromocytidine 5'-triphosphate, 5-bromouridine 5'-triphosphate, 5-iodocytidine 5'-triphosphate, 5-iodouridine 5'-triphosphate, 5-methylcytidine 5'-triphosphate, 5-methyluridine 5'-triphosphate, 6-azacytidine 5'-triphosphate, 6-azauridine 5'-triphosphate, 6-chloropurine riboside

5'-triphosphate, 7-deazaadenosine 5'-triphosphate, 7-deazaguanosine 5'-triphosphate, 8-azaadenosine 5'-triphosphate, 8-azidoadenosine 5'-triphosphate, benzimidazole riboside 5'-triphosphate, N1-methyladenosine 5'-triphosphate, N1-methylguanosine 5'-triphosphate, N6-methyladenosine 5'-triphosphate, O6-methylguanosine 5'-triphosphate, pseudouridine 5'-triphosphate, puromycin 5'-triphosphate or xanthosine 5'-triphosphate. In some embodiments, the RNA contains pseudouridine-modified nucleotides.

[0158] In some embodiments, the mRNAs include a 5' cap structure. A 5' cap is typically added as follows: first, an RNA terminal phosphatase removes one of the terminal phosphate groups from the 5' nucleotide, leaving two terminal phosphates; guanosine triphosphate (GTP) is then added to the terminal phosphates via a guanylyl transferase, producing a 5'5'5 triphosphate linkage; and the 7-nitrogen of guanine is then methylated by a methyltransferase. Examples of cap structures include, but are not limited to, m7G(5')ppp (5'(A,G(5')ppp(5)A and G(5)ppp(5')G.

[0159] In some embodiments, the retrotransposon of the present disclosure is delivered by a vector such as a plasmid vector and a viral vector (*e.g.*, adenoviral vector).

[0160] In some embodiments, the target mammalian cell is a human cell. In some embodiments, the human cell is an induced pluripotent stem cell (iPSC). When an mRNA is directly introduced to the target stem cell, in a preferred embodiment, the downstream flanking sequence of the retrotransposon has a length of at least 5 nt, such as 5 to 50 nt, 7 to 30 nt, or 8 to 20 nt.

[0161] In some embodiments, the R2 ORF and the exogenous gene are included in the same retrotransposon polynucleotide. In some embodiments, only the exogenous gene is included in the retrotransposon polynucleotide, while the R2 ORF is delivered separately on a different polynucleotide.

[0162] In some embodiments, the delivery is *in vitro*. In some embodiments, the delivery is *in vivo* or ex vivo. In some embodiments, the retrotransposon polynucleotide is delivered by parenteral means, or locally.

[0163] The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intra-articular injection and infusion.

[0164] Delivery can be systemic or local. In addition, it may be desirable to introduce the retrotransposon polynucleotide of the disclosure into the central nervous system by any suitable route, including intraventricular and intrathecal injection; intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir. Pulmonary administration can also be employed, *e.g.*, by use of an inhaler or nebulizer, and formulation with an aerosolizing agent.

EXAMPLES

[0165] The following examples are included to demonstrate specific embodiments of the disclosure. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques to function well in the practice of the disclosure, and thus can be considered to constitute specific modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the disclosure.

Example 1: Discovery and Testing of New R2 Retrotransposons

[0166] This example describes the discovery and testing of new R2 retrotransposons from various species.

[0167] The amino acid sequence of the open reading frame (ORF) of known R2 sequences, in particular those of the main functional domains, was used to search for analogs in the GenBank database. The genomic and amino acid sequences of the hits were examined with respect to domains/motif common to R2 retrotransposons, including zinc finger domains, SANT/Myb domains, reverse-transcriptase domain, and endonuclease domains.

[0168] R2 retrotransposons selected based on such sequence analysis (**Table 1**) were subjected to *in vitro* retrotransposition activity assessment. For each R2 retrotransposon, the entire

sequence of the open reading frame (ORF), its reverse-transcriptase (RT) domain (underlined in the ORF sequence and also separated listed), and the corresponding 5'UTR and 3'UTR sequences are provided.

[0169] The retrotransposition activities of these newly discovered R2 retrotransposons were tested with cultured human cell line, 293T. Cells were grown in Dulbecco's minimal essential medium (DMEM) supplemented with 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin (Invitrogen) in 5% CO_2 at 37 °C. Cells were grown as adherent cell cultures and passaged every 72 h at a 1:5 dilution (volume of cells: final volume of medium) until confluency was reached, in order to maintain log-phase growth.

[0170] Cells were seeded at 60% confluence in 24-well cluster plates one day before transfection. For transfection, OptiMEM medium (Invitrogen) was pre-warmed to room temperature. Transfection of mRNA into cells was performed with Lipofectamine® MessengerMAX™ (Invitrogen). mRNA *Lipofectamine*® MessengerMAX™ Reagent (1.5 µL of each) were mixed with 25 µl OptiMEM medium and incubated for 10 min at temperature. 0.5 µg of mRNA was mixed with OptiMEM medium in total volume 25 µL. Then, mixed 25 µL of lipofectamine reagent mixture with 25 µL mRNA mixture and incubated for 5 min at temperature. Mixtures were applied to each well. Helper virus infected the plasmid transfected cells at the same time and continue to incubate for 72 h. Integration of the donor sequence into the human genome was measured with quantitative PCR (qPCR) or or digital PCR (dPCR) in isolated genomic DNA.

Table 1. R2 Sequences

Name	Sequence	SEQ ID NO:
(1) R2ApuApu ORF	MWQVAQAFFVSVATQTAWADWRTTSVTTNICDQTDKVGLDVEVDVTKRLEFLNSKYGECEYE RPPSDSEQLASLPEVYEVLTSASLTGNDTSPQLEVAEAGLGVIDPVSPDDSMYSFISPTPSP AKEPSEGGTPLIKMPDKNPPCPCCGTRMGKMTELIDHLKRKHGIRKLLFQCARCGKTNEKHH SIACHIPKCTGSSEPATVGGFECEECHKTFETNIGLGQHKRLAHPALRNLERILASRPKETT ARGKHKQCWSEEELELLQKLEKQLEGQKNINKLIGEQIPTKTAKQISDKRRAIREKKCPIRP EGHKRWVEAQDVRSSPVRNKSGRLGSLYRKVISGWLSSNKLSILNKIFEKMLLAKENWEHLN DTAMTEFYNILGASTKKQRPGAKQPEARKTRPSGQQKEWMKKRSRKRGNFLRHQRLFHLDRG KLANIILDDVESLTCDIPPDEVHSVFKSRWESTGAYNGLGSFKAFERADNSAFNIMISAKEI SKNIKEISKTAAPGPDKITLRELTKIDPHCNHLRELFNLWLLTGMIPDSVK**NCRTVLIAKSA** **VPERLGDINNWRPITIGSIILRLFSRIITARLAKACPINPRQRGFIRSAGCGENLKLLQLII** **RHAKSQHQPLGVVFVDLAKAFDTVSHAHIISALEQRGVDQHIIALIKNMYHNVTTNIPLKDG** **HSDSIKIQIGVKQGDPMSPLLFNLAVDPLLCKLQECGEGFQIHSQNITALAFADDLVLLSNS** **WEGMMKNIEIVETFCNLTGLKTQGEKCHGFYIQPTKDSYKINDCPAWPINGTPLNMIDPGSS** **EKYLGMNI**DPWVGISKPELLSKVKNWLHRIDRAPLKPFQKVEILRTYAMPRLIYSADQANVK ITHLEMLDQEIRSAVKQWLHLPPSTCDAILYASNRDGGFGLLRFSKLIPSVQARRLHRLAQS	1

	PDEVITTIVRLEGIEKEYEKLWVKAGGRKDKIPSIWDPRTVFRLPNEVGGEEITSEWEAPAP KNIYPKPCNWRKDEFNNWTKLASQGKGISNFENDKISNDWLTHYRGIPHRKLLTALQLRANV YCNREFLARGRHDSQDASCRHCQADVESCAHIIGYCPKVQEARIKRHNNICNLLIEEAKNKD WVVFQEPNLRDKNNLFKPDMIFVKDKQALVVDVTIRYESNPSSLSEAAAGKVQKYKHLTQQV QELTNATHVKFVGFPIGARGKWYSENHDLLTEIGLSKSHQQKLARKMVNWALFSSVEIVHMF ASQARV	
(1) R2ApuApu RT	NCRTVLIAKSAVPERLGDINNWRPITIGSIILRLFSRIITARLAKACPINPRQRGFIRSAGC GENLKLLQLIIRHAKSQHQPLGVVFVDLAKAFDTVSHAHIISALEQRGVDQHIIALIKNMYH NVTTNIPLKDGHSDSIKIQIGVKQGDPMSPLLFNLAVDPLLCKLQECGEGFQIHSQNITALA FADDLVLLSNSWEGMMKNIEIVETFCNLTGLKTQGEKCHGFYIQPTKDSYKINDCPAWPING TPLNMIDPGSSEKYLGMNI	2
(1) R2ApuApu 5'UTR	AATCAGGGGGCGTAATTGTGGAATTTTATCTGATCGGGCTTTAGCTTCCCACAACTCCTGTT GAGGATAAGTGGGCGGGAGTAACTATGACTCTCTTAAGACGGGATCATAGTTACTAGCTGGG CTTTGCTGCAGAGGTCCCACCTCCTCGTGGTCCCGCTGGATGCCCTACTGGGTAACCAAGAG GACCTCTGCCCCAGTGCTGGAGTAGTAGTAGCTGCTACCGCCCTACCGAGTGGACTCACCAC CCCTCGGTAGAATTCCCTAAGGTGAAACACAATTGTTGACAGTGTCTGTGCTGTCCAAGGCC TTGCCCACGGGCTGCTGTTGGGCAAAACCGACTCCCTTGGAGGGGGGATCTCGGTAAAGACT AATCCTCAACCGATTTGGACAAA	3
(1) R2ApuApu 3'UTR	TGTTAGCGTCGGATTCGACTGTAAGCGTGCGGGTTTCACTGTTTGTTTTATTAAATGTTGCA GGATAGGGGTGGGACAAAAAACTGGGATCTCGTAATGATTACCTCAAACATGACGCATCACA TTGGACTTGACGAGGGTGGACGAGCCACCTTTACTTAACTCGGAAAAGGAACATAGCCCGCT ATGTTCTCTAAATAGTCACCCTTTTCTTCCAATTAATTCATCCGCACTTCTGCCCCTTGCTC ACACCGAGCCGTACGCAATTCCCACAGAGTGTA	4
(2) R2AspTig ORF	MHRFHRHAEELNEERSARLPTKKLTWSKEEDSSLVRHASRLWAPNLLKKHLLDLLAPLFPGR STESIKKRLQILKWLPPTTPTEASSTSSQDATSSSTQNATSTSNQEATIQDSTPTPSWRDDL LDFAVKNLQEPKLGSDDLRIIASELQSGTISCADAAQLLEDHARHWFPHTWKPSKEKHHSLD PSASKKQVRRANYAAIQRLYHLRRKDAASCVLSGKWRRAHRGPTAKVDGLDEYWSEIFELPS APDHCVTGPSKDKDWSLLGVISAAEVTSALRDMGNSAAGLDRLTPSTLLTWDQPSIAALFNL ILATESVPRHLA**KSRVIFLPKSEDPASPSDFRPISIASVLLRALHKILARRITSTLKFSQLQ** **HAFLKRDGCLDASSLLHAVFRKVHDNRVPVSMVFLDISKAFDSVSVHTLLRVAKSSGLPPPL** **TNYLDSLYDRPCVQLESRTSICGRGVRQGDPLSPLLFIMTMEVLLQSSIPEIGFDLDGNHIP** **ALAYADDLVLFAEKPERLQEKINGLFGSLSSAGMTINAKKSMGLTVAKDGKHKTMVLLPHAY** **MSGSTPIRPMGPSDTTRYLGLEF**SWKGKVTPKHTSKLDQMLQEISKAPLKPYQRLDLLKNFA IPKLTHELVLGNAHRNTLKKLDVLVRASVRSWLRLPKDTPLGYLYAHVRDGGLGLPCLHTSI PVLQKNRFAKLLSNPGSIHKAITSQKSFRVVQRRIDLPVTIGGHLVQSKKDARAAWLNDLKR SVDGKDLFGDPIDPVSHRWVSHPQRVFPRLHLRGIQLRGGLLATRARNARGRNLDGTPKTCK GLCGQTETINHILQVCDRTHDARCARHNRIVQKLQRFLRKKKIPTWVEPMIPSGHSYIKPDI IATINNELLVIDVSIVSGPRLSESWQLKSSKYGSSTNTTAIQAWHGSPITVRQAPLIISNRG LLFAQSGQALRSIGLTSRDMTDLCLLAIIGSLKCYDSYTRSTQDYKLQRGGQHAGNS	5
(2) R2AspTig RT	KSRVIFLPKSEDPASPSDFRPISIASVLLRALHKILARRITSTLKFSQLQHAFLKRDGCLDA SSLLHAVFRKVHDNRVPVSMVFLDISKAFDSVSVHTLLRVAKSSGLPPPLTNYLDSLYDRPC VQLESRTSICGRGVRQGDPLSPLLFIMTMEVLLQSSIPEIGFDLDGNHIPALAYADDLVLFA EKPERLQEKINGLFGSLSSAGMTINAKKSMGLTVAKDGKHKTMVLLPHAYMSGSTPIRPMGP SDTTRYLGLEF	6
(2) R2AspTig 5'UTR	GGGTCGTAGTTACATGGCCTGGGCGCCTCCTCGTGGCGACCGCTGGAGAAGATGGGGACGAA TCTCGTCTCCATCCGAGCTAAGATCCAGGCCACCCCCCTCCTCTGGGCTAACCATCCATTGG AGGGGTCTCTCTGCTGGTTACGATAAACAAAAAGAAAAATTCGCCTGGTCCCTCCGGACTTC AGGCTAGCAGCAGAAACAATGAAGTTGCAGGTGCTCCCGGCGCTCCTGCCCCCCCCTCTGAC AGTTTTGACTGCTCCTGTGGTCGAGTGTTTCCATCAAAGAGAAGTCTAGCC	7
(2) R2AspTig 3'UTR	TGATCTGCTTTCCCAAGTGGTGACTACCCTGCAAGTTTATTGTCAAAGCGTTAGCCGCCTCG GAGCTCATTACGTTACTATGGCCTGCATTGGGCTCGCCAAAATGTAACGACTGACTCTGATA ATATATGACACC	8
(3) R2AspTig' ORF	MLVTINKKKNSPGPSGLQASSRNNEVAGAPGAPAPPSDSFDCSCGRVFPSKRSLAMHRFHRH AEELNEERSARLPTKKLTWSKEEDSSLVRHASRLWAPNLLKKHLLDLLAPLFPGRSTESIKK RLQILKWLPPTTPTEASSTSSQDATSSSTQKATSTSNQEATIQDSTPTPSWRDDLLDFAVKN LQEPKLGSDDLRIIASELQSGTISCADAAQLLEDHARHWFPHTWKPSKEKHHSLDPSASKKQ VRRANYAAIQRLYHLRRKDAASCVLSGKWRRAHRGPTAKVDGLDEYWSEIFELPSAPDHCVT GPSKDKDWSLLGVISAAEVTSALRDMGNSAAGLDRLTPSTLLTWDQPSIAALFNLILATESV PRHLA**KSRVIFLPKSEDPASPSDFRPISIASVLLRALHKILARRITSTLKFSQLQHAFLKRD** **GCLDASSLLHAVFRKVHDNRVPVSMVFLDISKAFDSVSVHTLLRVAKSSGLPPPLTNYLDSL** **YDRPCVQLESRTSICGRGVRQGDPLSPLLFIMTMEVLLQSSIPEIGFDLDGNHIPALAYADD** **LVLFAEKPERLQEKINGLFGSLSSAGMTINAKKSMGLTVAKDGKHKTMVLLPHAYMSGSTPI** **RPMGPSDTTRYLGLEF**SWKGKVTPKHTSKLDQMLQEISKAPLKPYQRLDLLKNFAIPKLTHE	9

	LVLGNAHRNTLKKLDVLVRASVRSWLRLPKDTPLGYLYAHVRDGGLGLPCLHTSIPVLQKNR FAKLLSNPGSIHKAITSQKSFRVVQRRIDLPVTIGGHLVQSKKDARAAWLNDLKRSVDGKDL FGDPIDPVSHRWVSHPQRVFPRLHLRGIQLRGGLLATRARNARGRNLDGTPKTCKGLCGQTE TINHILQVCDRTHDARCARHNRIVQKLQRFLRKKKIPTWVEPMIPSGHSYIKPDIIATINNE LLVIDVSIVSGPRLSESWQLKSSKYGSSTNTTAIQAWHGSPITVRQAPLIISNRGLLFAQSG QALRSIGLTSRDMTDLCLLAIIGSLKCYDSYTRSTQDYKLQRGGQHAGNS	
(3) R2AspTig' RT	KSRVIFLPKSEDPASPSDFRPISIASVLLRALHKILARRITSTLKFSQLQHAFLKRDGCLDA SSLLHAVFRKVHDNRVPVSMVFLDISKAFDSVSVHTLLRVAKSSGLPPPLTNYLDSLYDRPC VQLESRTSICGRGVRQGDPLSPLLFIMTMEVLLQSSIPEIGFDLDGNHIPALAYADDLVLFA EKPERLQEKINGLFGSLSSAGMTINAKKSMGLTVAKDGKHKTMVLLPHAYMSGSTPIRPMGP SDTTRYLGLEF	10
(3) R2AspTig' 5'UTR	GTCGTAGTTACATGGCCTGGGCGCCTCCTCGTGGCGACCGCTGGAGAAGATGGGGACGAATC TCGTCTCCATCCGAGCTAAGATCCAGGCCACCCCCCTCCTCTGGGCTAACCATCCATTGGAG GGGTCTCT	11
(3) R2AspTig' 3'UTR	TGATCTGCTTTCCCAAGTGGTGACTACCCTGCAAGTTTATTGTCAAAGCGTTAGCCGCCTCG GAGCTCATTACGTTACTATGGCCTGCATTGGGCTCGCCAAAATGTAACGACTGACTCTGATA ATATATGACACC	12
(4) R2ChaBot ORF	MASGGANEFRCDCGTLWPSARSLAMHNFHRHTEKVMAERREKQPAKRLRWSAEEEAALHRQA DRLWSEGMLSKDLCLLLTDFFPGRSAEAIKKRLQSTRWIPPGQRPPVSPEGNEESSLAEQFD FAMWSKDLLEEAVGNLQEPRMGAELLSRIARGLLTGEFSKEEGVKRLEDHACQWFPYRWKSS VPRKSGNPQMHSNKKIRRMNYAAIQRMYHRRLKDAATCVLSGQWRTAYQPRPEVEGLNEYWT KIFETRSVEDTGCLRTPTRTLYSLIGRIIPSEVTAALKTMRSSAPGLDRITASELLAWHQPS IASYYNLLLVLEVCPRYLS**CSRVAFLPKVPVPSAPDDLRPISITSVLLRAFHKILARRLNDE LHLTSLQFAFLQRDGCLDATTLLHSILRNIHHTRSPGAMLFLDISKAFDTVSHSTIARIITS AGFPIPLIHYICNLYSNSTVQLGEKMIKCGRGVRQGDPLSPLLFILVIEEILLASLPDVGYN FGGQRLGAIAYADDLVLLAEHPQRLQEKFKIFSEALHTTGMVLNPKKSIGLTISRDGKRKAI VWQPHCYNYLGTPIRPLKIDESVKYLGLQF**NWKGRINPRHVGKLDEMLGEISAAPLKPYQRI RILTSFLIPKLTHSLVLGLAYRKTLKRADRLVYAAVRSWLRLPKDTALGYFHASVADGGLGI PCLSTTIPLLQRDQFAKLISQTDPLSVIIREQEAFKVVGNIINRPCRINSRVVTSKEEARAE WKTSLSQSVDGSELTIPDVDLASHGWLKYPVRVFPRLFLRGIQLRGGVLPTKARSARGNNRV VVDRSCKGLCGTSETLNHVLQSCEITHNARCSRHNKIVQQLEYLLNKIGIKIWVEPIIPSGS SFIKPDLVVMKDNQLWVLDVSIVSGMRLRESWNIKIVKYGSTERVNDIVCWFGSPIAVKQLP IIISNRGLCSTFSGSGLRALGLSRRDITDLCLLTVKGSLMCYDLYNRSTRSSGR	13
(4) R2ChaBot RT	CSRVAFLPKVPVPSAPDDLRPISITSVLLRAFHKILARRLNDELHLTSLQFAFLQRDGCLDA TTLLHSILRNIHHTRSPGAMLFLDISKAFDTVSHSTIARIITSAGFPIPLIHYICNLYSNST VQLGEKMIKCGRGVRQGDPLSPLLFILVIEEILLASLPDVGYNFGGQRLGAIAYADDLVLLA EHPQRLQEKFKIFSEALHTTGMVLNPKKSIGLTISRDGKRKAIVWQPHCYNYLGTPIRPLKI DESVKYLGLQF	14
(4) R2ChaBot 5'UTR	GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGCGGGAGTAACTATGACTCTCTTAACTTTCTT GTCATAGTTACAGGCCTGGACACCTCCGCGTGGTGCCCGCTGGATAGGGTATGGGATCTTAT ATCCCATGCCTGAACTAAGTTCCAGGCCACCCTCCTCTCGGGAAGACTTAATTACCCGTTAG GAGGTTGCTGCTGTTTTTTTGTTAAATTGATTAATTTATGTCCGGAGCCCTCTGGTTGACCG GTGATACAGCAGGTAAGTGAGGTAAGACAGTTGTCAAGAATCCACA	15
(4) R2ChaBot 3'UTR	TTTTCGATCTGGATTCGATGAACGACGATTGTCTCCTGTTATGATTTGAATTATTTTCTTGG AGATGTTGTATTTATAAATTTGCTAAATTATAATTGTATCTTTTTGAACATACTCCAATAAT AAATGTAAA	16
(5) R2ColStr ORF	MRQPLRVDFISVATQTSWAELWAARDNVFGFKSVSDKNEFHLNDPRFLKSMCNKGGGHLNDP GFLKDVSDKNGFIFNDFECLNSQFNFNTLEFLNSKFGLSDLTLCLDKLSVLPEPYEEPAPPP LSDNINRPGNQTSRPHPLQRDRELGNRPPNKAFEAFETLEAPAGESTITLIPSGEPLTPPAN PRGESTSPLYSRGCEPAAALTPIIKIPDRNLACPCCGVQFNNAQGILDHLKRAHGNKKVLFQ CAKCGRTGEKHHSIACHYPKCKGLPTASPPKGWTCEECQASFGTKIGLGQHKRLAHPVTRNN ERIAASRPKEKNNRGKHKQCWTEEEIELLKILDKKFQGSKNINKLIAEHITTKTAKQISDKR RDLNNSVTAKQLEFITETVTETVPIREQPEDLGGQEKNYIKGFYRKILSKSLTSKKKDFFHD LAKKVIKGKDHLSLSKEAIDEYYNLLHRRTAELHANQSRKIREAKITTSSKHSKEWMKKRAR RRGAFLRFQRLFHLNRGKLAGIILDEKESPQCDIPLEEIHRVFKSRWESPGTFKGLADFRSS GLADNTAFESLITAKEIAKNIKEMNTESSPGPDGVKVKDLAKIDPDFTRIAELFNLWLTTGE VPDKVK**ECRSVLLPKAADPGRLIDINNWRPITIGSAILRLFSRILTARLAKACPINPRQRGF IQSAGCAENLKLLQLIMQNAKKEHHALGVVFVDLAKAFDSVNHSHILSVLKQKGVDEHIIAL ITELYRNVTTSIQVKKARTESIGIHIGVKQGDPMSPLLFNLALDPLLCKLENSGYGYQHCDK NITAMAFADDLVLLSSSWDGMNRNIKILETFCDLTGLKIQGEKCHGFFIKPTKDSYTINDCE	17

	AWAINNTPLKLIAPGDSEKYLGLQIDPWSGLAKPELLSKMKKWLDLIDKAPLKPCQKVDILK GYAIPRVIYLADQSNAKVTNLEVLDLTIRNSVKLWLHLPPSTCDALLYARTRDGGLGITRLT ALIPSIQARRLHRIAQSSDEIIRSVARNENIEKEYSKLWVTAGGDKDKIPCIWEPQSVLVVS EEGSDLDGGTEWERPRLNVYYPKPCNWRKQEFQKWIKLQSQGRGISNFEGDKISNNWIGNNR SIPQRKIITALQLRANVYPTREFLARGRQDNFNKTCRHCNSEIESTAHIIGFCPSVQEARIK RHNYVCEILIEEARKCDWLVFQEPLIRDECNELFKPDLVFVKDQTARVVDVTIRFENKLATL KEAAEEKVNKYKHLEHQIKELTNASIVKFSGFPLGSRGKWFEGNFELLTELGLSASRREKVA RRLSNKALFTSVDIIHMFASSAKRPRDPNE	
(5) R2ColStr RT	ECRSVLLPKAADPGRLIDINNWRPITIGSAILRLFSRILTARLAKACPINPRQRGFIQSAGC AENLKLLQLIMQNAKKEHHALGVVFVDLAKAFDSVNHSHILSVLKQKGVDEHIIALITELYR NVTTSIQVKKARTESIGIHIGVKQGDPMSPLLFNLALDPLLCKLENSGYGYQHCDKNITAMA FADDLVLLSSSWDGMNRNIKILETFCDLTGLKIQGEKCHGFFIKPTKDSYTINDCEAWAINN TPLKLIAPGDSEKYLGLQI	18
(5) R2ColStr 5'UTR	AAGGTAAGTGGGGAGACCTGGCAGCCTAACTCTCCTAGACGCTCTCTGGGGGGGGGTGGACC TGGCAGTCCTGCTTGCTTTCTTTTTGGGGGCTCATAGTTACTAACTGGGCTCAGCCGCAGAG ACCACACCTCCACGTGGTCCCGCTGGATACCCTACCGGGTAACTAAGCTGGTCTCTGCCCCA GTGCCGGGAGTTAAAATCTCCCTTTTTTGCTGTGCGAACTCACCGAACGTACAGCAGTAAAA ACTCGGTGTTACAAAGTTTCTCTGACTGCAAATAAACTTTGCAGTCTCAAGCCCGAGCCCGT GGGAAAACTGGGCGCAGCTACCAGCACCTGAGGGGTGTGCTGATTAGCTATTGTCTTACCCT CAATACCGGACTGACTAGCTATTGTCTCCGCCCGGCCAGACCGGTCCGACAGGAATTCCTGG A	19
(5) R2ColStr 3'UTR	ACTGTTAGTATTAAGGTTTTTCTTTTACGTTATTTTGCCCAGTATACTGTATAGCTTTACTG TACTACCCTGTTTTTAAATACCCCCGAATGTTTTGCAGTGTCCCCTTATTTTAATTATGTAT TGTAGACCCTGTAAACCCCCTGTTGTTTATTATAAATTGTTTTATTCCCCAATCAATAAATC CACCCGACTGTGACGGCCAAGCACCACCCGACGGCTGTGAGAGTAGCGATACGGAATCTCGT TCCGGATCCACATGACGTGTCATTTAAAGTTCTGATTGGACAAGTAATAACCGGACTTGCCA CAGGTGGACGGGCCACCTTTACCTAACCCGGAAAAGAGACAGTCTCGTCTGTCTTCGTAA	20
(6) R2CroOre ORF	MLTDTVVLRPPGLPDPSCPGSCPRVASPKSDAECPKHSCPCGSEFGTKRGLAMHRFHCHPAE VNAERLTTLPTKKLVWTQEEMDGLLRHANGAWKEGMLKTQIFTSLMPHFPGRSAEAIKKRLQ SLKWVPARDNPERNSQPLHATSGNVAPATLPTPDPSRVDDQGANIEDPAASWRRALLAAAAD HFKEPKLGAGTLKQITQDLLNGKISSGECRVLMGEHAQHWFPHTWKPSSARPEPKARQWKAR EIRRANYAAVQKLYKLRRRDAAQAVFSGSWRNAYRGRAPQPDKMDQYWAEVFETASATDVTP PGSPSEIYWSVVSKITPSEVADALKGMRNSAPGLDRITAEELLTWDHPSLAAYYNLMLAAGG PPEHLA**CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQR DGCLEATALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKC LYSRSTVRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYAD DLVLLAEKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEP IKPMGTDDSVRYLGLHF**NWKGRIVPKHTGKLDSLLKELTKAPLKPYQRLQLLKFHAVPKFTH ELVLGHAHRNTVKKLDCLTRAAVRKWLRLPQDTPLGYLHANVKDGGLGIPCFSTSIPLLQKK RFEKIATNPATIFQIMQRQDSFRTQGRRLDKPCRLNSTVVTSKTEVREEWGNMLSNSVDGRE LRHPEVDKASHEWLLRPDRVFPRLHMRGIQLRGGLLPTKARSGRGNNRQVTSKTCRGACQNV ETLNHILQVCDVTHDARCARHNRVLHRLERLLHKKGLRTTLEPVIPAGSSFIKPDLIIETDT QTYVLDVSVVAGYRMTESWDIKVEKYGRPDRVEAIRKWAAIPATKRVEHLPVIVSNRGLCYQ PSGDGLRALGLTSRDISDLCLLAISGSLRCYDLYMRGTRAV	21
(6) R2CroOre RT	CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQRDGCLEA TALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKCLYSRST VRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYADDLVLLA EKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEPIKPMGT DDSVRYLGLHF	22
(6) R2CroOre 5'UTR	CGACCATTTGGTGTTGATATTGATACTGGCGACTGGCTAGTTGGTGTTGGTTGAAAGCGCTG GCGAGGTTGAGCTGATAATTGGTTGCACGGGTTGGGCACCTCCTAGTGGTGACCGCTGGATA GGGCTTCGGCCTGAGCCAAGATCCAACCCACCCTCCGTTGCTGGCAGACCCCAGTAACTGAG ATTAAACAGCTGGT	23
(6) R2CroOre 3'UTR	GACTACCGAGCGCCACCGGAGGGGACCCAGAGGACCGTCTCGGTGATTAGGCGGTGACCGCC CTGGCCTCTACCTGAAATAGCGATCAAATCGCCTTAGAGGGCCTGATTTGGGGGGTTGGGCA GACACTCTGTGTCATTACTGCCCTTTACCTTTCCACTTCTTGCCTGTCTAATAATAAATG	24
(7) R2HemCom ORF	MWRLAQACFVSVATQTTWADWHTNLVTTNISDQLKPVELEVDVRGLEFVNSKIGEIDGYHLP TESEQLASLPEVYEVLTLTSLTGNDTSPQSEVLEAGAEVTDPVSPDDSILSFISPTPSPAKD LTDESTPIVKIPDKNPSCPCCGTRMGKITSLIDHLKRTHGERKILFQCARCGKTNEKHHSIA CHFPKCTDKVASATVGGFECEECHKTFETKIGLGQHKRLVHPAIRNIERTLASRPKETSARG KHRQCWSEEEVELLQKLENQFEGHKNINKLIAEQIPTKTAKQISDKRRALREKKDSKQSESP EWGEAQEIRPSPVRPRGGRLGSHYRKVISDWLTSNKLPILRKIFEKVMLAKGDLRTLVDLAI KEHYKHLVGSSVERRKGHGSRQHEARKIHPSGQPKEWMKRRCRKRGNFLKYQRLFHLDRGKL	25

	ADIILDDVESLTCDIPPDEVHEVFKSRWGSTGTFTGLGPFQAFERAGNSAFDIMISAKEISK NIKEMNKNAAPGPDRITLKHLIKLDPHCYQLKELFNLWLLTGMIPDVVK**DCRTILIAKSAIP** **ERLKDINNWRPITIGSIVLRLFSRIRTSRLTKACPINPRQRGFIRSPGCAENLKLLQLMIRH** **AKSHHLPLGVVFVDLAKAFDTVSHDHIISALRQRGVDDHIIALIENMYQKVTTKIHLKDGQS** **DSIKIQIGVKQGDPMSPLLFNLAVDPLLCKLQESGNGFLIHSKTITAMAFADDLLLLSDSWE** **GMMKNIKIVETFCNLTGLKTQGEKCHGFYIQPTKDSYKINDCPAWTINGTPLNMIDPGSSEK** **YLGMCI**DPWVGISKPELLSKVKGWLQRIGKAPLKPFQKVEILKVYAMPRLIYSADQADVKIS CLEMLDLEIRSTVKQWLHLPPSTCDAILYSSTRDGGFGLTRFSKLIPSVQARRLHRLAQSPD EVITTIVRSEGIESEYEKLWVRAGGRKEKIPSIWDPRTVIKLTSVAGGEELTSEWEVSAPKN IYPKPCNWRKEEFNNWTKLVXQGKGISNFENDKISNDWLIHYRGIPHRKLLTALQLRANVYC TREFLARGRHGSQNTSCRHCQAEIESCAHXIGYCPKVQDARIKRHNGICDLLIEEAKTKDWI VYQEPNLRADNNLFKPDLILVKDKQALVIDVTVRYESNPSSLSEAAAEKAQKYQHLTQQIQD LTNATNVKYVGFPVGARGKWYLGNYDVLSEIGLSKSHQTKIARRIVNRALFTSVEIVHMFAS KARV	
(7) R2HemCom RT	DCRTILIAKSAIPERLKDINNWRPITIGSIVLRLFSRIRTSRLTKACPINPRQRGFIRSPGC AENLKLLQLMIRHAKSHHLPLGVVFVDLAKAFDTVSHDHIISALRQRGVDDHIIALIENMYQ KVTTKIHLKDGQSDSIKIQIGVKQGDPMSPLLFNLAVDPLLCKLQESGNGFLIHSKTITAMA FADDLLLLSDSWEGMMKNIKIVETFCNLTGLKTQGEKCHGFYIQPTKDSYKINDCPAWTING TPLNMIDPGSSEKYLGMCI	26
(7) R2HemCom 5'UTR	ACGGGATCATAGTTACTGACTGGGCTCAGCTGCAGAGGTCCCACCTCCTCGTGGTCCCGCCG GATGCCCTCGTGGGTAACCGAGAGGACCTCTGCCCCAGTGTTGGAGTGAGCAGTAGCTGCTA CCGCCCTATCGAGTGGACTCACCACCCGCTCGGTAGTACTTTTTGAAAGGTGAAACAAATTC CCGACAGCATATCTGCTGTCCAAGGCCTGGCCCACGGGCTACGTTGGGCAAAACCGATTGCC TTGAGGGGGGCTTTCGGTGTAGACTTATCCTCAACCTGAATTGACAAA	27
(7) R2HemCom 3'UTR	ATTGCGTCGGGTTAAACTGTAAGCGTGCGGGTTCACTGAATAAATGTTATATTGGATAAGGG TGGGACAAAAATTGGGATTCTCGTAATGATTACCTAAACCATGACGCGTCACACCGGACTGG ACTAATCAGGTGGACGAGCCACCTTATTTAACTCGGAAAAGGAACATATCTTACGTTCGATA AA	28
(8) R2MelGeo ORF	MAKIPDPFLGGGVWGPPVRLSRVSVGTQTSPSDWVDLVSRALPRTKCISRGVDLVSQNYLTS KTTNLAENELDSLVNFYFEMYRSDLHEHVHEGVRFSVDWEVLDGLPEVYEQLAPQPCMGENL DASPPRDSEPFVLGEGSGKEVREGTPEALPPVPGEHGGQSPDKIVKVTVPDKNPPCPCCGTR VNSVLSLIEHLKGSHGKKKVCFRCAKCGKENINYHSVICHFPKCRGPVIEKAPAGDWICEVC GRDFTTKIGLGQHKRLAHPLIRNKERIVASQPKETSNRGAHKRCWTKEEEELLIKLVAQFEG NKNINKLIAEHITTKTAKQISDKRRLLPKKTTDEADKELGMRLRSRRVVGTEPGSSQQAQAE AENNGPGKNHLPRRPAAGGETMEKIRPTDKDNGHQETSGQQREGQLQAHYHEKIKESLSAGA INNFPRAFKQLMDGQEMQTMINQAAQDCFGCLESISQIRTATRGKKEKVTHEKQPKKPFQKW MKTRAIKKGNYLRFQRLFHLDRGKLAKIILDDIECLSCDIPPSEIYSVFKARWETPGNFAGL GDFEIKGKADNSAFRDLITAKEIERNVQEMSKGSAPGPDGITLGDIRKMDPGYSRTAELFNL WLTSGEIPDMVR**GCRTVLIPKSTKPERLKDINNWRPITIGSILLRLFSRIVTARMSKACPLN** **PRQRGFIRAAGCSENLKLLQTIIRTAKSEHRPLGVVFVDIAKAFDTVSHQHIINALQQRGVD** **PHIIGLVNNMYKDISTYVTTKRNTHTDKIQIRVGVKQGDPLSPLLFNLAMDPLLCKLEEQGK** **GFHRGQRSITAMAFADDLVLLSDSWENMKTNIKILETFCDLTGLKTQGEKCYGFYIKPTKDS** **YTVNNCAAWTINGTPLNMINPGESEKYLGLQF**DPWTGIAKTNLTTKLDFWLERIDQAPLKPL QKLDILKTYTIPRLTYLADHTEMKAGALEALDQKIRTAVKDWLHLPLSTCDAILYSSTKDGG LGVTKLAGLIPSVQARRLHRIAQSSDETMKEFLEKHQMEQLYKKLWIQAGGESEKMPSIWEA LPATVPSNSADTTSEWEAPTMKTKFPKPCDWRRNEFEKWTKLVSQGRGIKSFEGDKISNDWL QYYRRIPHRKLLTALQLRANVYPTREFLSRGREENYIKACRHCKADNETCAHIIGCCPVTQD ARIKRHNYICDVLTEEAKKKDWVVFKEPHIRDTIKELYKPDLIFVKDGRAFVVDVTVRYEAS KTSLEEAAAEKVNKYKHLGSEVRRLTNAKDVVFLGFPLGARGKWYEGNFELLDTLGLSKSRQ VRVAKTLSTTALLSSVDIVHMFASRARRIHDGG	29
(8) R2MelGeo RT	GCRTVLIPKSTKPERLKDINNWRPITIGSILLRLFSRIVTARMSKACPLNPRQRGFIRAAGC SENLKLLQTIIRTAKSEHRPLGVVFVDIAKAFDTVSHQHIINALQQRGVDPHIIGLVNNMYK DISTYVTTKRNTHTDKIQIRVGVKQGDPLSPLLFNLAMDPLLCKLEEQGKGFHRGQRSITAM AFADDLVLLSDSWENMKTNIKILETFCDLTGLKTQGEKCYGFYIKPTKDSYTVNNCAAWTIN GTPLNMINPGESEKYLGLQF	30
(8) R2MelGeo 5'UTR	CGCCGCGCGAGCGGCGGCGGGTGTTACAGCCCCCGGGCAGCAGGTCTCGCCGAATCCCGGGG CCGAGGGAGAGGACCGCCGCCGCGCCCTCCTCCCCCCCCGTCGGCGGCGCCGTGGCGGGGGG CGTCGCTGCGGCGGCGCCCCCGCAGCGCGGCGTTTCGCGCGGGGGTGCGGGGGCCGGGCCCG CCGGCCCCCGGCGCCGCTGTCAACCGGGGCGGACTGCGCTCAGTGCGCCCCGACCGCGCGGC GCCGCCGGGCCGGGCTCCCGAGCTGGGTCTTTCTCTGCGCTGCGTTCGCGCGGGGTTGCGGG GGCCGGTCCCGCCGGCTCCCGGCGCCGCGGTCGGCCGAGGCTTGCTGCGCTCGGTTGGCCCC GGCGGAGCGGCGTTGCCGGTCGGGGATCGCGGGCTGGGTCGATCTCTGCGCGGCGTTTTGCG CGGGGGTTGCGGGGGCCGCTCCTGCCGGCCCCCGGCGCCGCTGTCGGCCGGGGCCGGCTTCG	31

	CTCGGTTCGCCCCGGCGGCGCGGCGTTGCCGGTCCGGGATCGCGGGCTAAATCGATCTTAGC CGCAGGTAGGGAGCCGTTGGGATTCACCAGCCCGACGAACTCCTATAAAATTCGGTGAAACA AATTCCTCGGTAAAAGCCGCGTGGCTTATTGCCTTAAACCTGGCCCCTTCGGTTTCAGACAG GGGCAAAAAGAGTTCAGAATCACTCCGTTTG	
(8) R2MelGeo 3'UTR	AGAATTATTTACTTTTTTTTTTTTTCTGTTTTAAGTTGTATTAATTATTTTTATTTTGTAT TTGTTTATTTAACCTTTTTGTGTTTTTGTAGTTTGTTGCTTATTGCTTTTGTATGTTTTGTT TCATTTCCCCTTTATATTTTAATAAATTAGACGGTAGCTAGGTTCGTATGGCAGCCACAAGC CAGTTAGGTAGCTCATAGTGAGTAGGGACAGAAAACTTTTTTCATAACGCGTCATTTAACAT CTGATTAGGACCAAGTTACCGGATTTGGACCAATTTACCGGATTTGTCCAAGGTGGACGGGC CACCTTTACTTAACCCGGAAAGGGAACATATATTATTATATGTGTTCGAAATA	32
(9) R2PeCa ORF	MEALSCALLSPLRRALTLLDTKIIPWGRCIFRPLEPLARYRAKVSTVATQTECPNTTTSAVQ SVIGTSQTLSTLNKTNHNCAVETHAVKPFPILTLEIVNSAFKLTGKSSIPLQPNIPIVSSTT TEMEGDGGSSEMLLENNERPPMTSTTTEENCGTENQEKFLVMLPNDDPECPVCGDHVGKPTA LSIHMKRVHGNKVVEYQCLLCGKQESRVHSIICHIPKCKGPETQEPEGDWPCPACEKQFRTK SGLSQHKRSAHPAIRNAERIEACQPKPNSQRGKHKSCWTDEEVTLLKDLETLYLGHRSINKM IADDMKTKTAKQVGEKRRLLGTSKKNNTAIVTPEPGTQKNTLEAQPLKPKAKTGLKNAYKRE IEKYLQTLGHLNGNMAQKNYPDFANEIFRAWIAEDPELNDQIEQTSSSVLSTFLSSTPRLTK RGKKKTESTKSGKEKAWMRKRAVERGTYLRHQQLFDRDRCKLASIILDGTERLQCKIPPDEV LRVYKDKWETCTPFEGLGQFTSYGEADNSAFDLLISRKEVVKNMKEMTTNSSPGPDGITMSD LLGEDPEGEKLAGLFNMWLVTGFIPTGVK**ECKSILIPKTMEPEALKELGNWRPITIGSIVLR** **LFSRILTTRLAKACPMNERQRGFIAAPGCSENLKVLHLLTRDAKRSKKPLGVVFVDIAKAFD** **SVSHDHIMWVLRERNLDQHMVNIIKDSYKNVKTRMELDKEFTEQIAINVGVKQGDPMSPLLF** **NLALDPLISKLENAGMGYLHGDKKITALAFADDLVMLSDSWDGMKTNIEILEAFCKLSGLKV** **QARKCHGFFLSPTHDSFTINNCEAWKIATDNLNMIAPGESEKYLGLKV**DPWIGFSKPALQNK LDTWLERITRAPLKPPQKLEILNTFTIPRLIYLADHTDYKKTMLASLDHSIRRAVKEWLHLP ADTCDNFFYSRARDGGLGVARLASLIPSIQARRLHRIAHSQDEIIRYVALTNGINEDFQKLW VAAGGDEKKIPVITDPVYIDYKLPQGILEQLTEWEKPRPKAMYPMPCSWRENEFSRWKTLHC QGSGIEHFKNDPISNHWIRNNRGFPARHYILALKMRSNVYPTREYQGRGKNNAPVSCRHCPA KYESLSHILGQCPAVQGARIRRHNKLCELLKREAKKQKWTVYQEPQLRTEDNELRKPDLVFV KDKKALVVDVTVRFEYKEKVFTDAAAEKTQHYESLTPQIRKLTGADEVEYYGFPLGARGKWP EISEQVLTALGLTDSEQRRAARFFSRRTLLYSVDVLNTFTSIGKNKPDFVNKDLAKGGDRWE SCND	33
(9) R2PeCa RT	ECKSILIPKTMEPEALKELGNWRPITIGSIVLRLFSRILTTRLAKACPMNERQRGFIAAPGC SENLKVLHLLTRDAKRSKKPLGVVFVDIAKAFDSVSHDHIMWVLRERNLDQHMVNIIKDSYK NVKTRMELDKEFTEQIAINVGVKQGDPMSPLLFNLALDPLISKLENAGMGYLHGDKKITALA FADDLVMLSDSWDGMKTNIEILEAFCKLSGLKVQARKCHGFFLSPTHDSFTINNCEAWKIAT DNLNMIAPGESEKYLGLKV	34
(9) R2PeCa 5'UTR	CGGCTGTGTCAAGAGTCTAGTTGCTAGGGCGCAGCTCTGAGACCTCAACCTCCTCGTGGTCC CGCTGGACTACTCCCTGGTGCCTACTGATGCCTAGGAGGGAGATGACTAACTGAGGCTCAAC TGCAGAGTAATACTCCCGACCCTACTGTCCTGGTTTGCCGGCCAGGTAGTAGTATCAATACG GTGAAAAGCAAAGATA	35
(9) R2PeCa 3'UTR	ATTATCATGACCGCGATCACCTGATATGACATGTGGGTTGGACAGGCCACCCGGGGTACCCT GTACTCCGAATAACCTGAAAAAGAAACCTGCAAGGGTTTGGTAAA	36
(10) R2CarCar ORF	MMAALASAQISPRRGATSTGLPSTGLNMEIVPWGRCIFRPFEPLARYRVQVSSVATQTESPM STDGERHPSLGSSQTITTAGEVYPISGPSQRITTAGAISSVTVAQTEAEDMLPALPLEFLNS VFGETVRPDTSPGPCVPTTPTSEKERDAEGSESLVDDRNDTIGVSPTTSTVAVTDGESETRV QGKVLVRLPNDNPTCPFCGDHVGKPTALQVHLKRHHGGKDVEFQCSLCDKIDPKAHSILCHI PKCKGLTSQEPAGDWECEVCAKRFGTKSGLSQHKRRAHPVVRNAERIQASQPKPNSQRGKHK SCWSAEEELLLIDFDKMYSGEKCINKLIAEAMKTKTAKQISDKRRLLNLSSKTKSECADPGP GANTHCLEVRSDEPPTITGLESMYMREVKKRLRDQSKGPQHLGLAPFESEILSAWVRGDPNV RSLVETTSIDVLSSFLAPEHETEGKGKKTAQKPKSGKNKSWMKKRAAKKSTYRRYQHLFAND RCKLASIILDGTERLQCQIPLEEVFRAYKDKWETRTPFEGLGQFKSHATADNSAFDLFISAK EVLKNIKEMNMKTAPGPDQVSLRDLVRVDPEGDTLAGLFNAWLVTGVIPDGIK**ECRSLLIPK** **TMDPEALKDLGNWRPLTIGSIVLRLFSRILTNRLAKACPLNARQRGFIAAPGCSENLKVLHL** **ITKQAKKEKKPLGVVFIDIAKAFDSVSHDHIMWVLRERGLDQHMINIISDSYRNVHTRMEVG** **KELTPPIAIDVGVKQGDPMSPLLFNLALDPLISTLEKTGKGFPVDAQRVTALAFADDLVMLS** **DSWEGMNANIQILESFCKLSGLKVQARKCHGFLLSPTHDSFTVNNCEAWKIGCDSLNMIAPG** **ESEKYLGLKV**DPWIGFSKPALSDKLDTWLERINRAPLKPSQKLEMLNTFTVPRVIYLADHTD CKKALLSSLDDKIRRAVKEWLHLPSDTCNNFMYARSRDGGLGVTKLASLIPSIQARRLHRVA NSKDWAIRYIALANDIEDEYKKLWLAAGGDQKAMPAITEPVVVDHKLPEHVLEILNEWERPA PKAVYPNPCNWRDAEFANWENLPCQGSGVSHFGNDPISNDWLKHHPGFSERQFLTALKLRAN VYPTREYLGRGKNNHNVQCRHCHASYESLSHILGQCPAVQSARILRHNKLCGILAREARKLK WVTYQEPHLHTAKNELRKPDLVFVKDGTALVVDVTVRFEYKEKTFSDAAAEKVRHYASLTTE	37

	VRKLTGATKVKFFGFPLGARGKWPKINEQVLSALGLSESAQGRAAKLFSRRALLYSTDILNT FAGIGKGLLKPADAKPEGTPRRHVGIS	
(10) R2CarCar RT	ECRSLLIPKTMDPEALKDLGNWRPLTIGSIVLRLFSRILTNRLAKACPLNARQRGFIAAPGC SENLKVLHLITKQAKKEKKPLGVVFIDIAKAFDSVSHDHIMWVLRERGLDQHMINIISDSYR NVHTRMEVGKELTPPIAIDVGVKQGDPMSPLLFNLALDPLISTLEKTGKGFPVDAQRVTALA FADDLVMLSDSWEGMNANIQILESFCKLSGLKVQARKCHGFLLSPTHDSFTVNNCEAWKIGC DSLNMIAPGESEKYLGLKV	38
(10) R2CarCar 5'UTR	GGCTGTGTCAAGAGCGTGGTTGCTGGGCCGCAGTATTGAGACCTCAACCTCCTCGTGGTCCC GCTGGACGACTCTCATGGGAACCTCAATATCGGGAGTTGACTGACTGAGGCTCAGCTGCTTT CCTCCCTGCCCTACTGTCCTGGTACACCGGCTAGGCAGTAGAACGCTACGGTGAAAAACACA A	39
(10) R2CarCar 3'UTR	TGATTACATCAAAAACTAATTCATGACCGCGATCACTGGACAGGACAAACTGGGTTGGACGG GCCACCCGGGGTTCGGAACACCCCAAATAACCCGAAAAAGAAACCCGCGTGGGTTTG	40
(11) R2PhrBla ORF	MGCYSLCSEECPYFLARKESGVPTSLQWEIKGYGCDKILIRLNGRSTAGCQTSPPGTPKRVT GQGREKPERPAGGGAPILVFEGLRADKEEKRSERQKQTENLGTEVGPTRPEPGADKVQEERA DKVQKEQRADKVTQEGEGKQKRERKLDKDEPTIWLPDNKPECPYCKEVVGKPTALQKHLTGV HGNQAVTFGCMKCQRKDPRVHAILVHIPKCKGPNIPDARQAQNKHRCEDCGNQYDTQSGLSQ HTRHKHPETRNKQRIEEKEKEKGGKRGTHKSCWTQEETEQLALLWEKYEGHANINKLIQQEL HTKTAKQIGEKRRNILEKRNNKPEIGEEIPKTKTQSINRQTEKEGLANKFKQTAITLIQEHR IKNESIARALHTWLQENNNTRQTVEKATEEILNNFPRTAGIHNIPKRKKTKKPRGPYRGKWR QKRRIQKLTYKQLQELYDKDRAAAAAAYILDGEKKTNCDLPIQEVYQTYKDKWETKTEFKGLQ NFHPWGGTNNDILSNLITGEEVKEHLQAMCNKTAAGPDNISVKDLKQVLHVEQKLAELFNLW LITGQIPNTVK**KSRSILIPKTSDETEKKKVGNWRPLTISSVVLRLFSKIMTSRITKACPLNN RQRGFIAASGCSENLWLLHNIIKRAKDKKKELGVVLVDIAKAFDTVSHDHIRWVLRERGMDK HIIQLIMSAYDNATTSIKLKEGNTPDIHIRSGVKQGDPLSPLLFNLAMDPLITELETEGHGV EDENWTITTLAFADDIAVMSDSCKEMKQNLQIIEAFCNLTGLKVQTTKSYGFALQPTKDSYI VNNIKPWTIKDQPIQMVQPAHTTRYLGIQV**GPWKGMEKPNMIADTKKWLQNITKAPLKPTQK LEILRTYTIPRLIYHNEQTGTGKTKLKELDNLIRAGIKSWLHLAHDTCTNLIYTKTKDGGLG ITRLETTLPHQRCNTLLKIINSPDLTTRTIARALGIEEQFSKWWKKAGGSEDNKPKIYPQPQ NTTQKPTEKEEIPVTTKIQPTHKKKEAEEWAKLPCQGGAVENFINDSISNDWIRYKKDMTEK EFITAIKLRTNTYPTREFLARGKENYERKCRKCPAKYESLSHILGQCPGLQNARIWRHNAIG KLLANQAKKKKWTVHQEPHLRNKDNQLRKPDYIFTKERTAMVVDITVRWEGKMGGLEEAAQE KIKYYEELRDTIQKEWGVKDISFHGFVCGARGKWPDQNFGILTKLGIESHRLDGLAKLFSRR TVQYSCRMLHTFERDKKNQTGHVDC	41
(11) R2PhrBla RT	KSRSILIPKTSDETEKKKVGNWRPLTISSVVLRLFSKIMTSRITKACPLNNRQRGFIAASGC SENLWLLHNIIKRAKDKKKELGVVLVDIAKAFDTVSHDHIRWVLRERGMDKHIIQLIMSAYD NATTSIKLKEGNTPDIHIRSGVKQGDPLSPLLFNLAMDPLITELETEGHGVEDENWTITTLA FADDIAVMSDSCKEMKQNLQIIEAFCNLTGLKVQTTKSYGFALQPTKDSYIVNNIKPWTIKD QPIQMVQPAHTTRYLGIQV	42
(11) R2PhrBla 5'UTR	GTTGCTGGTTGGTGCCGATAGAGATCCCTCCGCCTCCTCGTGGCCCCGCCGTAAACGAGTGT ATCACTCGGCCAAGAGAGGGATCGCGCCATACCACCCAAACTATTACC	43
(11) R2PhrBla 3'UTR	GTGTTAGTTTGTACACCTTGTTAGTTTGTACACTTTGTTGTTGTTGTTTGTTTGATGTTATT TAACTATTTATTCATTAAAAACCGGACATGGGATCTCGTAATGATTACCGCCCTGCCCTCCA ACCTAAACACCTGAGGTGGAGGCTGGGATCATGCAACTCACCCTGACTTGACCCTGACCTTA AAAGATGGATGGGCCATCGCGGGGGTTGGTCTCCCCGCAACAACTCAAAAACGAACTGGCGC GAGTCAGTTTC	44
(12) R2PoEx ORF	MEALTVALLSPPRRALNLLDKKIVPWGRCIFRPSEPLARYRAKVCTVATQTECPDTTISEVQ CTRTSQTTNIKDKYGYSYVAEKRNMKPFVVQTNEILNSIYNIKDVLLTPRRREDRLSRLWLE ESMGFDEDEGVYESSPSQETSISEGTPTLHEPAKYVVSLPNDDPECPVCGDHIGKPTALSIH LKRTHGDKEVEYLCLLCGKSEQRVHSILCHIPKCKGETTTEPEGDWPCPECEKQFQSKSGLS QHKRSAHPALRNAERIQASQPKSDSQRGNHKRCWTDEEVALLLELEKTYLGHKHVNKMIATE IKTKTAKQVGEKRRLLAQSKKKVTVENTNPGPGAERHSNEKGNFKLKETPLSTLEVRPVQPA SKLGLNGAYKRKLKVYLENQKKMKQDNGFPDFAKDIFSAWVKGDHNIQEMVEQTSSDILSNF RPTLPKPKRTERQNNEKGNFKLKEAKSWMRKRAVKRGTFLRYQHLFDRDRCKLASIILDDTE RLKCEVSPDEILKVYKDKWETRTPFNGLGHFTSHGKIDNSAFELLISAKEVIKNVKEMNKKS APGPDGLTIEDLLGEDPEGERLAGLFNIWLVTGYIPNKIK**DCRSILIPKTMDPVALRELGNW RPITIGSIVLRLYSRILTTRLAKACPLNARQRGFIATPGCSENLKTLHLLTRDAKRSKKPLG VVFVDIAKAFDSVSHDHIMWVLRERGLDQHMAKIIADSYTNVHTRIEIDNKSTEPISLNVGV KQGDPMSPLLFNLALDPLISQLETTGCGFSHGNKTITALAFADDLVMLSESWDGMTKNIEIL EAFCSLSGLKVQAKKCHGFFLSPTHDSYTVNNCKPWQISDDALNMITPGESEKYLGLRV**DPW IGFSKPALSDKLDNWLQRITRAPLKPPQKLEILNTFTIPRLIYLADHTDIKKTMLSSLDKNI RRAVKAWLHLPADACDNFLYSRTRDGGLGVARLSSLIPSIQARRLHRIAHSQDETIRSIALA	45

	NGITDDYHKLWITAGGDENKIPVITAPVYIDYKLPQGVLDELTEWEKPAPKAIYPIPCSWRE AELSEWKTKQCQGSGIQHFKNDPISNTWLRNQRGFPARHYITALKLRSNVYPSREYQGRGNK NMNVNCRHCPAKYESLSHILGQCPAVQGARIQRHNKLCNILKQRARKLKWTIYQEPQLRTAE NELRKPDFIFVKDQKALVVDVTVRFEYKDKVFQDAADEKIQHYCSLTPQIKDLTGAEEVEYF GFPLGARGKWPEINYKVLTALGLSDTEQGRTAKFFSRRTLLYSIDVINTFASLGKNKPSTVN LYPVNKDHNQEGAAGKSCND	
(12) R2PoEx RT	DCRSILIPKTMDPVALRELGNWRPITIGSIVLRLYSRILTTRLAKACPLNARQRGFIATPGC SENLKTLHLLTRDAKRSKKPLGVVFVDIAKAFDSVSHDHIMWVLRERGLDQHMAKIIADSYT NVHTRIEIDNKSTEPISLNVGVKQGDPMSPLLFNLALDPLISQLETTGCGFSHGNKTITALA FADDLVMLSESWDGMTKNIEILEAFCSLSGLKVQAKKCHGFFLSPTHDSYTVNNCKPWQISD DALNMITPGESEKYLGLRV	46
(12) R2PoEx 5'UTR	TAGGCTGTGTCAAGAGTCGAGTTACTGGGGCGCAGCTCTGAGACCTCAACCTCCTCGTGGTC CCGCTGGACTACTCCCCTGGACCTTCCCATGTCCAGGAGGGAGCTGACCAACTGAGGCTCAA CTGCACGTGTGTTTCCCGCCCTACTACCTTGGTATGCCGGCCAGGTGGTAGTATCAATACGG TGAAAAACAAACA	47
(12) R2PoEx 3'UTR	ATTATCATGACCGCGATCACCAGATTGACGTAGGTTGGACGGGCCACCTGGGGTACCAGTTA CTCCGAATAACCCGTAAAAGAAACCCGCAAGGGTTTGTTAAA	48
(13) R2PufMau ORF	MIDKMELLMRFAFVSVGTQTSLSDLQDNVVKETITDHGKLVEVEMNWKGLEFINSEIDELER YHPSTWTDRLAILPEVYEIPTSPNLEGTTNSPCLREVESEEEEEPSSPGNMAQFSPLPLTPT PGSEVDPPQVDTPVVKVPDNNPSCPCCGNRMGKVSGLIEHLKRVHGRKKILFQCARCGRTNV KHHSIACHFPKCKGAIEIAPTEGWICGECQRVFESKIGLGQHKRLAHPVIRNMERIIASQPK ENSARGAHRQCWTEEEVELLRKLVKQYEGSKNINKLIAEHIPSKTSKQISDKRRGLHKSPVK NKGKDLEWRQHLEEEEHLNDIGSHEEEGLKSHYQRMINKWITSEKLPILRGSFTKVLEGKEK PSTIIDEAIEDCYSCLFKKSGELEVSQDPNHQVTRKTPSSKPRKKWMKKRAKKKGQFLRFQR LFYRDRCKLAGIILDDVESLKCGIPLDEVHKVFKSRWESPGTFRGLGAFRTLGRADNSAFEA MITAKEISKNIKEMSKNAAPGPDGVNLRDLINIDPNCDQLMELFNLWLVTGTIPDVVR**ECRT** **VLISKSAAPERRNDINNWRPITIGSTVLRLFSRILTARLTKACPINPRQRGFIKSAGCAENL** **KLLQLLIRNAKKEHRPLGVVFVDLAKAFDTVSHSHIIAALKQKGVDEHIIALISNLYYNINT** **YIDLKKEKSDPIGIQIGVKQGDPMSPILFNLSIDPLLCKLEEDGSGFQHCAKKITAMAFADD** **LVLLSGSWEGMQKNIEILEVFCELTGLKTQGEKCHGFYIQPTKDSYTINDCPPWTINGTPLN** **LIEPGSSEKYLGLQI**DPWVGISKPELMKKIEDWLQRIGKAALKPFQKVDILKGYTIPRLIYL ADQADVKITYLETLDLAIRSAVKEWLHLPPSTCDAILYSSTRDGGLGITRLSGLIPSVQARR LHRIAQSSDELTKAVAREEGIEKEFEKLWIAAGGDKEKIPSIWDPRTVIIISELNPEEQVSE WEVPSPKTIFPRPCDWRKQEFSNWTKLLSQGRGISNFERDKISNNWIEHYRGIPHRKLLTAL QLRANVYPTREFLARGRQERQIKSCRHCQAEIESCSHIIGYCPAVQDARIKRHNQLCELLIT EAKKKDWIVFQEPLLKDNQNELFKPDLVFVKDNQALVVDVTVRYESKPTSLADAATEKVKKY QHLKDQVQELTNATFIKFMGFPIGARGKWYPGNYELLAELGLSSSRQDKVARRMSNRALFSS VDIVHMFASRSRTILKPLHANDSPLTSQNS	49
(13) R2PufMau RT	ECRTVLISKSAAPERRNDINNWRPITIGSTVLRLFSRILTARLTKACPINPRQRGFIKSAGC AENLKLLQLLIRNAKKEHRPLGVVFVDLAKAFDTVSHSHIIAALKQKGVDEHIIALISNLYY NINTYIDLKKEKSDPIGIQIGVKQGDPMSPILFNLSIDPLLCKLEEDGSGFQHCAKKITAMA FADDLVLLSGSWEGMQKNIEILEVFCELTGLKTQGEKCHGFYIQPTKDSYTINDCPPWTING TPLNLIEPGSSEKYLGLQI	50
(13) R2PufMau 5'UTR	GGGGTCATAGTTACTAACTGGGCTTAGCTGCAGAGGTCACACCTCCTCGTGGTCCCGCTGGA TACCCTTCTGGGTAACTAAGTTGACCTCTGCCCCAGATTGGGAGGGTTACCTCCCCTATCGA TTGGCTTCACCAACCATTTGATAGTAATTAGCAGGTGATACAAATTTCTGGCACCGAAATTA TTTCGCTGCCCAAAGCCTATCCCCACGGGCTAAAGTTGGGGATATCTCGACAGTTCAGAGGG TGGTTCTGTCGGATTAGTCTAAACCCTCAAAC	51
(13) R2PufMau 3'UTR	CCACTCTCACGACGGTCTTACTCGTTCAATCCAAATAACATGACGTGTCACCTTTGCCAAAT TTGACACCAATTACTGGCCATTCACAGGTGGACGGGCCACCTTTACTTAACCCGGAAAAGGA ACATATATAATTTATATGTGTTCGATACC	52
(14) R2Hal ORF	MGRPMRISFVSVATQTSPLDLRDNAVSRSIRDHKPGGLLEGRFEIRGLEFVNSDLEGLKSQQ SPTLSDLLAVLPEVYEVPNSFSLRDKNTCPTPVEDENEEEESSPLGAVAGTSPVPSTPVSRS RDDLPAESTPVVKIPDKNTSCPCCGNVTGKLSGLIEHLKRAHGKKKILFQCARCGRTDVKHH SIACHFPKCKGPVEASPERSWKCEVCEKTFETKIGLGQHKRLVHPVTRNLERITAARPKTGT NMRGAHRRCWTEDEVELLRKLDKQYEGQKNINKLIAEHIPTKTTKQISDKRRDLHKSPVKKG KDQGQKKHLELEGEHLAEGEHLVCVGSETEGGLKSRYRKTVSEWVTSGKLSILHGMFEQVIE GEEDPAIVADKAMKDCFACLASKSLELKASQGRPRQVAQTARPSEPPKGWMKKRAVKRGQFL RFQRLFYLDRGKLAGIILDDVESLKCGIPQDEIHNVFKSRWESPGTFKGLGAFRSKGLADNS AFEAMITAKEISKNIQEMNKNSAPGPDRLSLRDIIKIDPHFLQLMEVFNLWLVTGIVPDVVK **ECRTVLISKSSLPERQMDINNWRPITIGSIILRLFSRILTARLAKACPINPRQKGFIRSAGC** **AENLKLLQLLIKDAKKEHRPLGVVFIDLAKAFDTVSHHHLIAVLKQKGVDHHIITLITNMYQ**	53

	NTRTRITMKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGSGYHHDSKGVTSMA **FADDLVLLSGSWEGMQSNIKILEAFCELTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQING** **TPLNMIEPGSSEKYLGLRI**DPWTGISKPELLVKIRDWLQRIGEAPLKPFQKVDILKGYTIPR LIYLADQADVKATYLEELDLTIRTTVKEWLHLPPSTCDAILYSSTRDGGLGITRLSGLIPSI QARRLHRIAQSSDEVIRSIVKQEGIEKDFEKLWLTAGGDRNKIPSIWDPRAIRVTSEERGED EFASEWEIPTPKTIFPKPCDWRKQEFINWTKLQSQGRGIANFEKDKISNNWLERYRGIPHRK LLTALQLRANVYPTREFLARGRQESHIKSCRHCQAENESSAHIIGYCPAVQDARIKRHNQLC DLLITEVKKKDWIVFQEPILKDEQNEIYKPDLVFVKGDQAMVIDVTIRYESKLSSLADAAAE KVKKYCHLKSQIQELTNAKNIKFKGFPVGARGKWYSGNYEVLKALGISTSRQDKIARCMSRK ALFTSVDIVHIFVSHSRSIVRT	
(14) R2Hal RT	ECRTVLISKSSLPERQMDINNWRPITIGSIILRLFSRILTARLAKACPINPRQKGFIRSAGC AENLKLLQLLIKDAKKEHRPLGVVFIDLAKAFDTVSHHHLIAVLKQKGVDHHIITLITNMYQ NTRTRITMKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGSGYHHDSKGVTSMA FADDLVLLSGSWEGMQSNIKILEAFCELTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQING TPLNMIEPGSSEKYLGLRI	54
(14) R2Hal 5'UTR	ACGGGATCATAGTTACTAACTGGGCAGAGCTGCAGAGGTTACACCTCCTCGTGGTCCCGCTG GATGCCCTTCGGGGTAACCAAGTTAACCTCTGCCCCAGTGTGGGAATTATAGATACTTCCCC CCTATCGTCTGGCCTCACCGACCATTCGATAGACTCAGCTTGGTGATATAAATTCTTGGCAA CGATTTTTTAATCGCTGCCCAAAGCCTTGCCTACGGGCACAGTTGGGCTCAGTCACACAGCA CTTGAGGGGGTGTTGTTTGACTTAGACTACTCCTCTATCTGGACTGGCAAA	55
(14) R2Hal 3'UTR	TTATTGCTACTGTTCTGGTTATATTGTTTTTTGCTGTATTACTGTTGGTTGTTTTTATTGCA TTTATTGTTCTGTCTTATATTGTTCTTGCTATATTATCATTCATTGCCTTGCTGTATTACTA TTAATTCTTTAATAAAACTACCCCTGGTACTCCTGCTCTGTTCCCTTGGCACTGGAAAGGTT GGGAGAGTAAGTCGGAACTCGTATTCATGACGTGTCACATCAACCTAAATTTTGACAAAATA GTAATTGGAATTGACAAATTTTGCCAAATTGGACTACTCACTGGATTGGACTCAGGTGGACG GGCCACCTTTACTTAACTCGGAAAAGAAACATATATCCTTTTATGTGTTTGTTAAA	56
(15) R2EmyOrb ORF	MEKSNLYFYTGLALDRVNNAPAVGFDAPADSVKYIRGVIRVAGPMDKMVWTIMKSHEHESEN QVLPVEHGEEVEETVCDDHGEEILSILPMVFTRDTFNDLTMESFNPDFKYLSTFDDPYVSES GYLNAHLVNGSSLLNSFRPSTDFNRDRNKDLNSTGSNINILESSSWKDGVFYLEYPVDSFNC PICPQILPTFGKWAKHINVVHKSKAIRVVCSKCGKSSQYHNIACHYPKCIPKKADTPMKSHT CSVCGLGFHTRSGLSQHCRHRHPAVRNEQRKEDMAAKSKIKEGSTRSRIWTKDEVAQLIQLE RKYIGKRNINKCIESEMRTKTNKQISDKRRELAKKKLVVTGDREEVTEVEDIRVNILEIPPP RERIFPLKGHPEVDLVEKICKRGKQSREGREIINSLGEIEGLLKEDLTKALGWAMLEKLCYS LVEGKDPRNESQIELRNKGKGKMRKEGRRKQFNAIRRYATKKAKYKFYQQLFDKDRRRLTRI IMGEPDKAKCAIPMKEIEEYYVNILGTIGHGNMIGWPSSTENADNDILMSPISVDEIRIALR EIRKDSAPGPDKVKVSNLWDIFNLDSLILTKIFNIWFLNRQVPDEFK**KNRTILIPKTQDEEE** **MKKLTSWRPLTIGSVWIRIYTKILAKRLVETVKICSRQKGFISAPGCEENIAILDNLIKGAK** **RNGNEIAIVFVDLAKAFDSVGHGHIIAGLRRFGIDEHFIDIIRDLYDNCTTRVWVGDEVTES** **ILIRRGVKQGDPLSPILFNIAMDPLLTRLEVEGSGFYVKGNGVTSLAFADDVAILSSSYKGM** **IKNLGILEEFCKNTNLSVNVKKTKGFHILTKHKTYVVNPKDNWQIGSEEIEYVQPGDFEKYL** **GARI**DPWIGVGGPVLKEKLKDWSENLIKAPLKPAQKILILQTYILPKLFYNLLLIEPPFNLL NDLDVLVRKIVKEILHLPQCTTDGIFYSRGRDGGLALTKFSVQIPNMLIRKWRRHIVSRDDW MDLSYLYAGENLVDKILSFSAVTSHPKKWREEEFLRWSELRCQGIGIKYFKNDLISNSIFRN SSKVKSSAYIAALQLRANIYPTRETLARGRGGVNALCRWCGKVRETVPHISGQCYKTKNARI KRHNRVVSAVVDKVGKLGWKAMIEPHLRNSKGELRKPDIIFIKGDTAVVVDVTVRWEDNAGS LEQAHSEKVAYYLELEEEIKNLTGGKNIHFFGFVLGARGKWSEGNNDLLQLLGMDRSKNFKE SICRLVLYSTIGMMAIFSDR	57
(15) R2EmyOrb RT	KNRTILIPKTQDEEEMKKLTSWRPLTIGSVWIRIYTKILAKRLVETVKICSRQKGFISAPGC EENIAILDNLIKGAKRNGNEIAIVFVDLAKAFDSVGHGHIIAGLRRFGIDEHFIDIIRDLYD NCTTRVWVGDEVTESILIRRGVKQGDPLSPILFNIAMDPLLTRLEVEGSGFYVKGNGVTSLA FADDVAILSSSYKGMIKNLGILEEFCKNTNLSVNVKKTKGFHILTKHKTYVVNPKDNWQIGS EEIEYVQPGDFEKYLGARI	58
(15) R2EmyOrb 5'UTR	TGGCTCTCCATGCGTCTCTGGACCCTCTCAGCTGTCGGGCTTGGACGGAAACATTGCCTGAG GGTAGTGAGAGAATGGGTTATCGGTGAGGTGGGGGTGGGCGTGAGCCCGGTCAATAGGGGCA ACCCCTACCAAGATCTGAGCCTTTTCTCACACGCCCGTTAGAGCTCGACATTATGCTGTTTG GGACCCCTCTTTGAGGAAAGCCTCTGGATTGCAAGTCCAACCGCTACCTCCTCGTGGTGAGA GTCGGTAACTGGCTTGGATAGCCAGGCGGATTGCGGAGGACGAACCCACATCCCACATTCAG TGGGGTGTGGGACGGGGTTGGGCAGCCCCATATGGTGCCACATGGATGCCCCCCTGGTAAGG GTAGCCTCCCCCAGGTACGGGTTGACTCCCCCTGGTAAGGGTTAGATTCCCCCTGGTAAGGG TAGTCTCCCCCAGGCACGGGTTAGTTTCCCCCTGGAAAGGGTAAGTTTCCCCCAGGTACGGG TTAGTTTCCCCCTGGAAAGGGTTAGTTTCCCCCTGAGAAGGGTAAATCTTTTAGCT	59
(15) R2EmyOrb	AGATCCCGGAGTTTCCATTCCGTGTATCTTGCCAAAACTAAATTTTAGCTATTTTGGTTTCC ATGCTCTAAATTTTCAAATTTTTAATATTGTTTTGAACATTAATTTATTTAGTGTTATTTAA	60

3'UTR	TATCTGTCCTTCAATAGAATGGTTTTAATTATATAATAAATACTTATTAATGGATGAAATTC CAATTTAAATTATTAGTTTTTTTTTTTTTTTTAATTACTGGATGTGGATGCGGACTGGCCG CGTATAATAACCAGGAAAGGGCAGGTAGTCACGCCTGTACATAAAT	
(16) R2OpiFel ORF	MVAQLACADCSRLFPTKIGLSQHRRHVHPTQHNADKLGRVKYSGARWSQQESQSLLRLANNL YPSCGTQTALFARLEQYFPGRSAISIKTRLRVLNWQAQQDESSSGGPDQTIGQITAYSSEAD DYSVWFKQTVDCAVSLLESHADSSLASVDLLAFARGLQSGIMTPEQVLSLLDLHASRTFPHI WKTVSRRRRQLAHRIPVNRKQIRRANYAAIQTLYHQRRKDAASAVLDGSWKDLYKGNCGLPL NAEQYWKQVLSAPKHVDSRPSRVVVPSDWSLIEPITGEEVGRTIRLMGNSSPGLDKLTPRML RRFNANALAGYFNLLLLSGGCPPHLC**RARITLVPKVPNPTSPDQLRPISVSSILVRCFHKVL** **ADRWSRRLQLPSLQFAFLHRDGCLEATSLLHALLRHSSATASNLSLAFVDISKAFDSVSHDT** **IVRSAKAFGAPSPLVRYIAQSYENAVAVFPNSEVHCHRGVRQGDPLSPLLFIMAMDEVLGLS** **MPQLGYQFHDTLVDGFAFADDWVVCAESQARLKEKLEAAAVELGRAGMKINARKTKAMVICG** **DRKHRATAVSVEPFCFAEELITPLGPTDTVTYLGIPF**TFKGKGVFNHRQHLLKLLDEVTRAP LKPHQRMEITRNYLIPRLTYSLVLGQVHRNTLKRLDNYIRQSIRGWLRLPKDTPISYIHAGK QHGGLGIPSLSATIPMQRRARMEKLFSTQCRVLRNVVNDSAFGKIVRDLSLPIRVHGACVNT KEELVAAWGDSLHNSVDGRGLRELVASPLSNRWLVFPERVFPRIFIRGIQLRCNLLRTRVRS ARHGHGGQTILCRGNCGQPESLVHILQSCWITHDARCARHNRVARELAKRLRRLGYTVFEEL RAPTSTSFIKPDLIAVRERRATVIDVSIVSDGRGVTVWNEKKQKYGADEYSLAIISALRAIG CDVDFLVHQPMIISYRGICFPQSAKAVIGLGLPKFTVSDLCLLAIVGSLRTYDTFMRGTWR	61
(16) R2OpiFel RT	RARITLVPKVPNPTSPDQLRPISVSSILVRCFHKVLADRWSRRLQLPSLQFAFLHRDGCLEA TSLLHALLRHSSATASNLSLAFVDISKAFDSVSHDTIVRSAKAFGAPSPLVRYIAQSYENAV AVFPNSEVHCHRGVRQGDPLSPLLFIMAMDEVLGLSMPQLGYQFHDTLVDGFAFADDWVVCA ESQARLKEKLEAAAVELGRAGMKINARKTKAMVICGDRKHRATAVSVEPFCFAEELITPLGP TDTVTYLGIPF	62
(16) R2OpiFel 5'UTR	GTCGGCGAGCTGAACCACCTCCTCGTGGTGCCGACTGGGGAGCCTGGACGCCAGTCTGGGTG AGCTAAGAGTTCAGCAACTCCAGACAGGGCTAACCACCCTGTTTGGTTTCAATGTGCTGGTT ACTTTCGACAATGTACACTACGAATCATCTGGCGCCTCCGCACCAGTTCCAACCAGTACACA GGAGAGACTCGTGGGGCTCACGTGTGAGGAATGTGGCAAGTGCTGCAAATCGAAAGCTGGCT TGGTAGCCCACCATCGAGTTCACAACAATGAAAGTGTTGGTACGAAC	63
(16) R2OpiFel 3'UTR	GTGCACATTCTTGCCATTGAATCTCACACCAGACCCAGTATGACGGACACTTTGTGCCTGAT GTGCGAGTCTTGACTGTGCTAGCGCTTACCGCGCCTTGAAGAGCATTCAGCATTGTTTGTCC TTTCTTCGGCTGTTGATCTTTACACGCGTGTTTCCATACCAAAATTCTTACATACGTTGGGA TTCATCTTATCTGACTGGAACTGTTGGTGGCATGACTTCGAAAGAGACGTTTCTCTCTTTTA TCTCATATTCTCCTGTAACCCTTTCACATTCATCGCTTGCATTTACTTTTTATGTCTGTGAA CATGCTCTTCAAAAATAAACGATAA	64
(17) R2CapEur ORF	MCWSTRVNFLSVGTQTDPKDLQEITTTKTTEDCEEKSEEASEGLGLIYLEFVNSGSMLFSEG GWDEWLGGLPEPYEEPGPPVLGAAVGPPELSGSVPGSHAGPGSGAACRGSAAGHGSIVACRG SAADPGSGGDCRGSAVDQGSGGDCRGSVADLETGGVCSGSAAGPGSRGDCRGSAAGRVSGGD CRGSAAAPGSKRDCRGSVAGRVSGGVCPGSATGPGSRGDCRGSAAGRVSGGDCRGSAAAPGS KRDCRGSAAGRVSGGVCPGSATGPGSGGECRGSAAGRVSGGDCRGSAAGPGLKRDCRGSAAG RVSGGVCPGTATGPGSGGDCRGSAAGRVSGGVCPGTATGPGSGGECRGSAAGRVSGGDCRGS AAAPGLKRDCRGSAAGRVSGGVCPGSAAGRVSGGVCPGSAAGPGSGGECRGSAAGRVSGGDC RGSAAAPGLKRDCRGSAAGRVSGGVCPGSAAGRVSGGVCPGSAAGPGSGGERRGSAAGRVSG GDCRGSAAAPGLKRDCRGSAAGRVSGGVCPGSATVPGSGGECRGSAAGRVSGGGGSGSVAEQ GSGRDCRGSADNQGSGLDYQGSGPNSLEEIKDPSPAGNPAAVNTPIVKIPDSQPSCPCCGNK VSKISGLVEHLKRNHGIKKILFQCARCGKTNEKQHSIACHFPKCKGPEEPPSIKGLICEECQ KIFPTKIGLGQHKRHAHPVTRNNERINASRPKAISARGAHKRCWTDEEVEQLQRLIKQYEGN KNINKLIAEHIPTKTAKQISDKRRLLHNSPTEGETITQDEMGHLEQETEGEQLVPSESHTEG GLKNHYLQLISQWMTDGKLSILQGSFDKIREMNGEATELIPEVIKDSYTRLFKKNTETKASQ GPKRRECKKPSSGPPKQWMKKRAIRRGQFLRFQQLFDLDKSKLASIILDDSESSKCNIPSDQ IHEVFKSRWESPGTFQSLGAFRANGHADNSAFSTLITAKEIDKNIKEMKKNSAPGPDGVSLK DLIKMDPRYTHLVELYNLWLVTGTIPDEMR**ECRTTLLPKSSKPERLKDINNWRPITIGSTIL** **RLYSRILTARLTKACPINPRQKGFIRAPGCAENLKLLQLILKTAKKEHRPLGVVFVDIAKAF** **DTVDHQHIITALRQKRVDEHIIALITDLYHKISTRITLKKEQSDPIGIQVGVKQGDPMSPIL** **FNLSMDPLLCKLEEEGEGFQHCSKKITTMAFADDLVLLSGSWEGMQKNIKIVEVFCELTGLK** **TQGEKCHGFYIKPTKDSYTINDCPPWTIKGTPLNMIEPGNSEKYLGLHI**DPWVGISKPELTE KVKDWLGRIGRAKLKPWQKVDILKGYAIPRLIYLADQADVKTTHLESLDLEIRNTVKEWLHL PPSTCDAILYCSTRDGGLGITRLTGLIPSIQARRLHRIAQSSDEITRTFAREAGIEKDFAKL WVAAGGDKDKIPSIWDPKTVSVLPEELSGEDPLSEWEVAAPRIIYPKPCNWRKQEFTHWTKL ASQGRGISNFEKDTISNKWLEQHRGIPQRKLITALQLRANVFPTREFLARGREGTQIKACRH CQAESETCAHIIGYCPAVQDARIKRHNQICELLMETAKKEEWVVFQEPQLRDTSHELYKPDI IFVKDKKAIVVDVTIRYESKPTSLSEAAAEKVRKYSHLKDQIQELTNATDIKFMGFPLGARG KWHEDNYELLEDLGLSGSKREKLARRLSNRALFTSVDIIHMFTSKTRTVIRV	65

(17) R2CapEur RT	ECRTTLLPKSSKPERLKDINNWRPITIGSTILRLYSRILTARLTKACPINPRQKGFIRAPGC AENLKLLQLILKTAKKEHRPLGVVFVDIAKAFDTVDHQHIITALRQKRVDEHIIALITDLYH KISTRITLKKEQSDPIGIQVGVKQGDPMSPILFNLSMDPLLCKLEEEGEGFQHCSKKITTMA FADDLVLLSGSWEGMQKNIKIVEVFCELTGLKTQGEKCHGFYIKPTKDSYTINDCPPWTIKG TPLNMIEPGNSEKYLGLHI	66
(17) R2CapEur 5'UTR	CAGGAGCTCTCTCCGCTCTCTCCCGCTCGAAATTAAACTAAACTAACGAAATTTCAACGAAA TTAAACGAAAACGCGGCAAATTCGGCTAATTCGGCAAATCCGGGAACAGCAGAACCTCGGCA AAACCAGCAATTTCGACAAAAACGGCATTTTCGGCAAAAACGGCAAAAACGGCAAAATCGGC AAAATCGGCAAATTCGGCAAGAGCAGCTAATCAAGCGAAATTAACGAAATTTCAACGAAAGT AAACGAAAAAGCAGCAAATTCGGCTAATTCGGCAAAACCGGGACGAGCAGAACCTCGGCAGA AACACCAATTTCGACAAAAACGGCGTTTTCGGCAAAAACGGCAAAATCGGCACAAGCGGCAA ATTCGGCAAAACCGGCAAACCAAACGAAATTAAACGAAAAAGCGGCAAATTCGGCTAATTCG GCAAAACCGGGAACAACAGAACCTCGGCAGAAACAGCAATTTCGACAAAAAAAACGGCATTT TAGGCAAATTCGGCAAAGGCGGCAAACCCGGCAAACACGGCAAAAAGAAAAATAAGGTAAAA CGGCTCAATTGGCAATACAGAAAAAGAAAAGACCGCAGCAAAACAGGAAAATTCGCTAAAAG CGGCAAATTCAGGAATTTTTACAAACGCGGTCAGAACTTTGTTGAGGTAAGACATTTTCTAA TTTCTAATCTTACCTCCATGCGGTCCCACGAGACCCACTTCGGGGTAACCGAGATGGCTCTG CCCCAGCCCGGGAACGAACACTAGCTGTTCCCCGCTACCGACTGGCTTCACCAGCCTTTCGG TAGTAAACCTAAGGTGAAACACAGTTTCGGCAACATTTGTTGCCCAAAGCCTTGCCCACAGG CTGCGCTGGGAATTTTTTGCCTTTTTCTTTCACCGGACTTACCTCCATCTCGGACTGACAAA	67
(17) R2CapEur 3'UTR	TAAACCACACACACACACCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA AAAAAGAAAAACAAAAACAAACAAACAAAAAAAAAACCCCAAAAAACCCCTACACTTTCCAC ACAATATGCATGTTACACATAGATTCCTTCACTGTTATTACCTGTTTATGACATATGTATTA CCAATTGTAAAACTAACCTGCTCACGACATCTAAAGCCACAAGCCAGTTAGAAGGGTAAGAT CTGGGAAGGACTCCCCACTTGAACACCGATGATAGTATGTGGGGACGAAAAGTGGACTCGAA GGCATGACGTGTCACCTTTACCGATTTGGCTAACCACCACCAAATTTGACATAGGTGGACGG GCCACCTCTACTTAACCCGGAAAGAAAGCATAGAAAATAATTTCTATGTATATTAA	68
(18) R2DeCl ORF	MKVSDILSYRMARGPGYQRAMGGSQEAGSQPIITRPPRKLDMARVMFVSVGTQTNYEDAKPV ASDDTQTWTGNENELIPFSLTIWNSSFQEPRPSESEIETDGADIGWLQTGESVEDLFDDNGE EDEAHLSELEASGYALEHSLLVEPVLGSENTSEGETVTITLPNSMIHCQLCGEVPGGPMKAQ EHFVKFHRDVPVVFSCGKCGKGDQNPRSILSHHAKCKSGAPQTSPARGHECTLCGQSFGTIS GLSQHKRHVHPTERNEERIVAASQPSKKKGRKLQLWSQQEIERLRQLEVTYKGARTINKLIA TELGNKSHIQVAEKRRQLRESATPESTQVVGADPLAFLMTAPTMISVPSEIADSLRRETARL YGPAQEDGEDEVTTALRRWAVGNEPIEKTVEKVTAELLSEAVSSCDKKTHYLREKRDKVRTY TSYEWMKRRARSRALFGRYQCQYFKNRARLAALILDGNITSKCQIPASEVHDVYKKKWESST QYGGLGSFTLGGCADNSVFEGLITAEEVGENLAAMNRTSATGPDKMNWKDVKDLCDGSTVVD LYNAWWATGKIPVAVK**QCRSILLPKTADQEQLMEIGNWRPITIGSMLLRLFSRILTKRLARA** **CPLNPRQRGFVQSPGCSENLMVLKSIMEQAKRERKHLAVVFIDVAKAFDSVSHNHIVDVLRR** **RGLDSHVIGVIQNSYEDCHTSIEVGGTKTPEIAIRTGVKQGDPMSPLLFNLAIDPLLATLET** **HGVGFQTEGGKIAALAFADDLVLLSDGWKGMSHNLSILEKFCNLTGLTVQAKKCHGFMITPT** **HDSYTVNNCQPWSIERAELHMVGPEESEKYLGVRI**GPWTGISKPDTESQLQTWIQKIDEAPL KPSQKVVILNDYALPRLIYRADHAEETSVRLARLDGLVKRAVKAWLRLPHSTCDGLLYSWRK DGGMGLLKLADSVPSIQVRRIQRVAHSEDPTIRQLMNSSRAQTRFDKAWKRAGGTPDTAPAL TDPITQSVNLETVSDGSILPQKVNYPVPRDWRKDEFEYWCRLPVQGMGTVCFRGDTISNHWL RNHRGFRERHYIAALQLRANVYPTRESLNRGRKGALLHCRKCSSKYESCSHILGQCPAIQAA RIARHNKICDILGDEAKKLGWETHKEPRLYTTMGELRKPDMIFVREGVAIVVDVTVRWEYDS LSLAKAASEKVEYYQNLEQQVRDYTSTHTVHFFGFPVGARGKWHSKNDLVLQLLGVTKGRMK ALGRLVARRALLYSLDMLVMFGGGTYSEHI	69
(18) R2DeCl RT	QCRSILLPKTADQEQLMEIGNWRPITIGSMLLRLFSRILTKRLARACPLNPRQRGFVQSPGC SENLMVLKSIMEQAKRERKHLAVVFIDVAKAFDSVSHNHIVDVLRRRGLDSHVIGVIQNSYE DCHTSIEVGGTKTPEIAIRTGVKQGDPMSPLLFNLAIDPLLATLETHGVGFQTEGGKIAALA FADDLVLLSDGWKGMSHNLSILEKFCNLTGLTVQAKKCHGFMITPTHDSYTVNNCQPWSIER AELHMVGPEESEKYLGVRI	70
(18) R2DeCl 5'UTR	AGCGTGCTTCTCCTCTGGGCAAGGAGACGGAGTCATAGTGAAGGCCCTACCTCCCCGTGGTT CCCGCTGGAAATGACAGGTAAACACTCCTGTCAGCTAAGAGGGTCTGCACGTGAAAACATCG CCCCCGGGTGGTGATGACCTGACACCACTCCGGTAAAAAAGGTCAACTCAAAC	71
(18) R2DeCl 3'UTR	TGATCGGATCAGACATTTCAACCAGACATGGACCTCCCAGCTGGTAGGCCCTGGGAAACTAG GCCTCCCCGTGTGATGCCACCCTACACGGTTGACGAAACCCACCGGAATTGACACCCTTTCC CCCACCCCCATCCCCTTTCCCTTTCCCCTACGGGAATCTCGTTCATCCATGATGATTATTTC ACTGGATGAACGAGGGGCTGATGGGCAGCTACTAATAACCCAGAAAGGATGCAACACACCAT CGTTGCAGCATCGAAAC	72
(19) R2Onki ORF	MSQGWHNGRKRPRHPPPPSPSNGRKMESIHLSSQWDQTGIPKLAPVVGAFGHLQVERSGSTT ALPLGSPGSKSYERPTQNKSPKGGRSEGPPTVRAGAPVAPSEAGGGTGERVRVKPPKRKKPS	73

	EPPKGAAAAGESGKAIATDSMCIEIPRTEIECGTCGATAATVKRLVDHMAAKHNEVKLTFKC AKCGKRSVNRHSIACHAPKCKGVQGTPEGKSKKFRCMACQGGFGTASGLSQHMRHRHIAAWN QERIQRERIAEGRLEGKKSWSPDEVRALRELSRTHSNGESPERKIAEGLGTNTSEEQVRWKR RKLALDNVPEGTEGPSWGATIMGLTKPAMQPPEVKGVAAALREQKQVAKEEEEGETAMGRIE REAAEIIQLLGKTDRGKGGTKPAKAKTQPALQRGWVRRRREKRDEYKQHQTLYKQDKRKLAE LILDGTAKRECRIPIEDLTKVYSAQWGDTGPFEGLGQFRSTGKADNTHFNTPITASEVLDNL TAGKKNSAAGPDGISRKELLEWDAKGEKLATMFASWQTVGKIPKAFK**QGRTTLLPKSMREEE** **LRETSGWRPITIGSTVLRLFSRILTGRLERACPINPRQRGFIATPGCAENLMILNGIIYQSK** **AERSELAVVFIDFAKAFDTVSHGHLIHVLRERGLDSHVVGLLKDSYEGVYTTVRTREGETQP** **IEMKKGVKQGDPMSPLLFNLALDPLINKLENVGEGYRFEGKKVTTLAFADDLVLMSDSWAGM** **QHNVNIVEAFCKLAGLRVQPKKCHGFMVRRGDAAFTINNCREWAIGGENMHLIDPEKTEKYL** **GLKV**NPWLGIAKKDRGAQMAEWIERIHRSALKPSQKISILNIHALTRVMYTIDHGDVGRTQL NALDGLVKTAVKRWLRLPPSTCDGLLYARNRDGGLGITKLASVIPAMQARRLFCLSNSPDQL TRAVTTYMTPQAEYEERWAQAGGDREQTPTLGKRQEDTPADRDDTERTGQEGAGKGEERTYK TPCDWRSIEFKGWAALPVQGDGIGEFQQDHTSNSWLADPEAAGFRERHYMAALQLRANVYPT REALARGRSKEGAACRRCDHKLETASHILGKCPGVKMARLRRHDLICELLVEEEAEKLGWAVT KEPRVRTAEGRLRIPDLVFVKGRLGVVADVTVRYERNEGTLERGRKEKVDKYLPIAPIIAKQ FKLSRVLVYGFPVGGRGKWPTSNYELLQTLGMGGTRSASFAKLISRRTLLYSLDILKAFGLE CKAGGPTGRKGGREDDDDE	
(19) R2Onki RT	QGRTTLLPKSMREEELRETSGWRPITIGSTVLRLFSRILTGRLERACPINPRQRGFIATPGC AENLMILNGIIYQSKAERSELAVVFIDFAKAFDTVSHGHLIHVLRERGLDSHVVGLLKDSYE GVYTTVRTREGETQPIEMKKGVKQGDPMSPLLFNLALDPLINKLENVGEGYRFEGKKVTTLA FADDLVLMSDSWAGMQHNVNIVEAFCKLAGLRVQPKKCHGFMVRRGDAAFTINNCREWAIGG ENMHLIDPEKTEKYLGLKV	74
(19) R2Onki 5'UTR	AGAGGGGAAAGAGCACCCAGGGGGCCCGACGGTCCTGACGACGCAGGCTAGGTAAGGCGCCC TGG	75
(19) R2Onki 3'UTR	TGAAGGGGGCAGGAAGGGCCAGCAAAGGCGCGAACGAGCCAGATCCGTAACCTATGGGATAA GGATTGGCACGAAAAATAGGCCGAATGAAGCGTGACCTCGAGCGGACGCCATCGAGAGAAGA GGTCGCAGAGTGAGCTAAGTGAGGCAGGTAGCACCCCCTGCGACCCGAAAAAAGGTGCCGCC AGTGGGTTGGTAGCCACTGGGGGGAACCCCCGATCTACCCCGTCCTCCGGCACCGGATCGGA GCGTCCCGAAGGGCCAGTTAGTAGGGAAGTCCGTTGATTACGCCGAAGAACCGCTCGTTCAA CCTAGGCTGGGTTAAAGCGCTAGACGTGAGGAGTTTTTGGGAATTGTCGCCACGAGTGTGTT GCATCGTGGTTATGTCGCCTGTAAGGCCTCCGGTCGGAAGGTAGGCGGCTGACGCGCAGCCC AAAAACGCGGAATTGTCAGAGCTCGTCGGGCCTGACGCGACCGGATGTCAAAGTGAATGATA TTCAGACAGCGGGGCTCGTGTAATACCCTCGTGCTCCTCATGCATAGCATGTAAAAATCGGG GTG	76
(20) R2AhaPra ORF	MPYRPSRGIATGYLTIVYSPSPSGHMDSSDKDGNAIGAGTQNRADNPRYPCFCGSTFRSKRG LGLHRAHKHPAVLNAERLAVLPTKRLTWTQEDDEALKRLSDKHWERGMLKKDLITLLTSYFP DRSFEAVKKRLQLTKWAPPEENPERLERPHRAETPERVDAHLADPTETLQSHGPDPTTATSW RRALLVAAADRFKETKLGAEELNQMVRDIQTGELSRKQCRARLRQHAKQWFPHKWKPSKPKP ANASCLKNKKAIRRANYAMVQKLYRLRRRDAASTVLSGRWRNAYRGPAPLPPEMEEYWASVF QTPSANDDRSPAEPAAVLWSLVAAITPSEVTSALNGMKDSAPGLDKIPAEELLTWDQPSLAA YYNLLLAAGGPPKHLS**CSRVVFLPKTDEPSAPGDYRPISIASTILRAFHKILAWRLRDTLQL** **SPLQHAFLQRDGCLEASTLLHSIFRRIHTSRVPCAMLFLDVAKAFDTVSHKSLFRVAYEAGL** **PPPLVSYLQHLYKRSSVKLAETTKCGRGVRQGDPLSPLLFILVIDDIVKRSLPDVGFDFDGQ** **RVGSLAYADDLVLFAEKSTRLQEKLNLLSEALQQAGMALNAKKSHGLTIARDGKRKCMALVP** **TTYECKGDTIQPLGTDDQVRYLGLQF**NWKGRIVPKHTGKLDSLLNELTKAPLKPHQRIKLLT TFLVPKFTHELVLGHAHRNTLKRLDCLTRAAVRRWLRLPTDTPTGYFHANVLDGGLGLPCFS TSIPLLQQKRLEKIALNPATLFQIVQRQDSFRSLAQRMDRPCRLGGAVATTKAEVREGWQIM LSASVDGKELSIPQVDKASHEWLRRPARVFPRLYIRGIQLRGGLLPTKARSARGNNRQVTDR TCRGACRAPETQNHILQVCELTHDARCARHNRVLRTLEKQLRRTEARTRMEPVIPSGSSFIK PDLIVEAANQMWVLDVSIVAGYRMEETWGIKARKYGREDRVADIRRWADVPEHTPVFQMPVV ISNRGLLYGPSGEGLRAMGLSRRDLTDLCLLAISGSLMCFDLYKRGTKDWKD	77
(20) R2AhaPra RT	CSRVVFLPKTDEPSAPGDYRPISIASTILRAFHKILAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHSIFRRIHTSRVPCAMLFLDVAKAFDTVSHKSLFRVAYEAGLPPPLVSYLQHLYKRSS VKLAETTKCGRGVRQGDPLSPLLFILVIDDIVKRSLPDVGFDFDGQRVGSLAYADDLVLFAE KSTRLQEKLNLLSEALQQAGMALNAKKSHGLTIARDGKRKCMALVPTTYECKGDTIQPLGTD DQVRYLGLQF	78
(20) R2AhaPra 5'UTR	TTAAGTGATATAGGACAATAGTCATAGTTACAGGCTTGGACGCCTCCGCGTGGCGACCGCCG GGGAGGGTACTTGGTACCCAAGCCAAGATCCAAGCCACCCCTCTGC	79
(20) R2AhaPra 3'UTR	ACAATCTGTTTTGTGTGTTGCTTGTATCCGTTGACATAACTGTCTTCTAGCGATTGTATCGC CTTAGAAAGCCTGGCGGGCGGGAAGGGGATAGAGAGATGTATTATGCTTTATATGTATTGTA	80

	TACTGATGTATTTTTGGTGTCCGAATACTGATGTATCTTACTGTAAATGTATCATGCACTGC TATAACCTGTACTGTTTTGCCTTGCCTTTCTAAAAATAAACAGTTAAA	
(21) R2AnnSte ORF	MPYRPSRGIATGYLTIVYSPSPSGHMDSSDKDGNAIGAGTQNRADNPRYPCFCGSTFRSKRG LGLHRAHKHPAVLNAERLAVLPTKRLTWTQEDDEALKRLSDKHWERGMLKKDLITLLTSYFP DRSFEAVKKRLQLTKWAPPEENPERLERPHRAETPERVDAHLADPTETLQSHGPDPTTATSW RRALLVAAADRFKETKLGAEELNQMVRDIQTGELSRKQCRARLRQHAKQWFPHKWKPSKPKP ANASCLKNKKAIRRANYAMVQKLYRLRRRDAASTVLSGRWRNAYRGPAPLPPEMEEYWASVF QTPSANDDRSPAEPAAVLWSLVAAITPSEVTSALNGMKDSAPGLDKIPAEELLTWDQPSLAA YYNLLLAAGGPPKHLS**CSRVVFLPKTDEPSAPGDYRPISIASTILRAFHKILAWRLRDTLQL** **SPLQHAFLQRDGCLEASTLLHSIFRRIHTSRVPCAMLFLDVAKAFDTVSHKSLFRVAYEAGL** **PPPLVSYLQHLYKRSSVKLAETTKCGRGVRQGDPLSPLLFILVIDDIVKRSLPDVGFDFDGQ** **RVGSLAYADDLVLFAEKSTRLQEKLNLLSEALQQAGMALNAKKSHGLTIARDGKRKCMALVP** **TTYECKGDTIQPLGTDDQVRYLGLQF**NWKGRIVPKHTGKLDSLLNELTKAPLKPHQRIKLLT TFLVPKFTHELVLGHAHRNTLKRLDCLTRAAVRRWLRLPTDTPTGYFHANVLDGGLGLPCFS TSIPLLQQKRLEKIALNPATLFQIVQRQDSFRSLAQRMDRPCRLGGAVATTKAEVREGWQIM LSASVDGKELSIPQVDKASHEWLRRPARVFPRLYIRGIQLRGGLLPTKARSARGNNRQVTDR TCRGACRAPETQNHILQVCELTHDARCARHNRVLRTLEKQLRRTEARTRMEPVIPSGSSFIK PDLIVEAANQMWVLDVSIVAGYRMEETWGIKARKYGREDRVADIRRWADVPEHTPVFQMPVV ISNRGLLYGPSGEGLRAMGLSRRDLTDLCLLAISGSLMCFDLYKRGTKDWKD	81
(21) R2AnnSte RT	CSRVVFLPKTDEPSAPGDYRPISIASTILRAFHKILAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHSIFRRIHTSRVPCAMLFLDVAKAFDTVSHKSLFRVAYEAGLPPPLVSYLQHLYKRSS VKLAETTKCGRGVRQGDPLSPLLFILVIDDIVKRSLPDVGFDFDGQRVGSLAYADDLVLFAE KSTRLQEKLNLLSEALQQAGMALNAKKSHGLTIARDGKRKCMALVPTTYECKGDTIQPLGTD DQVRYLGLQF	82
(21) R2AnnSte 5'UTR	GTTACTGGGTTGATGACGATAGTATAATCCTCCGCCTCCTCGTGGCCCCGCCGGACAACGAG TCTCTGACTCGGCCAAGTGAGGATTACATCAACCCCTATTTTAACCCTCGGTCGTCATGGTG TGCAACAGTTTGTGCACAGAAAACTGCCCGTATTTTCTCGCGAGGGCGAAG	83
(21) R2AnnSte 3'UTR	CTCAAGCTCAATTTGGACGACCCCTAAATTTTAATCCTAATTTTAAAAGCTTTTTAATCTTC CTTTTAAAAATTAAAATTTGTTGTATTTTTACAAAGAAATTTTGTTTCTAATTGGCTTATTT TTAAATTTTAGTTTAGGTAGGTTGACATGTCGGGAGCTTGTGATGATTATACCAAACCCTGG TCAAAGGCTAACACCAACATGGTCTGAGGCCAGGTGTCATGCTGCGTTAATTGGATTGACTG ACCTTAAATCGATGGACGGGCCATCAAGGGGCTGGTCACCCCTTACTAACGCGAAATAAGAG GCTTCCCCCTCAAACTTCAA	84
(22) R2AriEle ORF	MDPSNKDGNASSVGTQDTNTRPSYRCFCGSQFNTQRGLGRHQAHRHPAQVNASRVAALPTKR LAWTQEYDEALKRHSDRLWQSGMTKTKLYPLLIPYFPERSLEAIKKHLLLTHWVPPIEASEH NPPQEHPERPENPDRPPTIQQPPTSPVAPPQNGPGCRNATSWRGAILAFAADRLKETKMGGE ELAQIARDIQTGVISRKQCRARLNQHAKSWFPHKWKPSVTRPAGANCQWSKNAIRRANYAEV QKLYKLWRRDAASTVLSGRWXNAYKGPAPVPPEMEEYWAAVLKTPSAIDRVPAEPPAVLWSL ITPIVLSEVTAALNGMKSSAPGLDRLSATELLSWDQPSLAAYYNLLLAAGGPPKHLS**CSRVV** **FLPKTEDPKTPGDYRPISIASTILWAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEASTLLH** **TIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKSTVRLAE** **TTKCGWGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAERSTRL** **QEKLDLLSEALQRAGMALNAKKSHGFTIARDGKRKYMALVPTPYECKGDAIQPLGTDDSVRY** **LGLQF**NWKGRIIPKHTGMLDSLLNELSKAPLKPHQQLKLLSVHLVPKFTHELVLSHAHRNTL KRLDCLTRAAVRHWLRLPTDTPVGYFHAHVLDGGLGLPCFSTMIPLLQQKRLEKVASNPATL FQIVQRQTSFQTLESRMDRPCRLGGTVVTTKAEVREGWQIMLSNSVDGSELSISEVDKASHE WLRHPARVYPRLYIRGIQLRGGLLPTKARSTRGNNRQVTDRTCRGACRAPETQNHILQVCEL THDARCACHNRVLRCLEKLLRRKGLCTRLEPVIPSGASFIKPDLIVEVTNQLWILDVSIVAG YRVDETWDIKTTKYGRPDRVADIRRWAAVPATTPVVQHPVVISNRGLCYRPSGDGLRALGLS RRDTMDLCLLAIIGSLKCYDLYQRGTKDWNV	85
(22) R2AriEle RT	CSRVVFLPKTEDPKTPGDYRPISIASTILWAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLAETTKCGWGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAE RSTRLQEKLDLLSEALQRAGMALNAKKSHGFTIARDGKRKYMALVPTPYECKGDAIQPLGTD DSVRYLGLQF	86
(22) R2AriEle 5'UTR	GAAGAGCGTAGTTGCAGGCTTGGGCACCTCCTCGTGGTGACCGCTGGATATAGTACTTCATG TACTAGAACTAAGATCCAAGCCACCCCTCTACACGACTATTGCTTTAGTAGAGGGTTAGTTA CTGGTTTACTTATTTATATATCGTACTAGTCCAAGTCCCTCTGGAACT	87
(22) R2AriEle 3'UTR	TATGTCTTGTTATGCTTATGCTTGTATGTTGTTACCTGTCTATTAGCGAGCAATCGCACTTG GAGGGCCTGTCGGGCGGGAAAGGTCTAGTCGGTTGTATTTGCTATATTAACTTGTATACCAT GTATATTGATATCATGCATATGTGATGTATCTTGTTGTATCGTACTATGCACTGCTTTACCA TGTACTGTCTTGCCTTGCTTAATCCAAAAATAAACTGTCAA	88

(23) R2AspMar ORF	MFPSKRSLAMHRHHVHPDEVNRERAMRVPTKRLAWSEEDDSALVRHATLLWAPNMLRKHHLE LLVPFFPGRSAEALKKRLQILKWLPPSHPTLIDPPTPTTRVEDPATSTSSRTEDIPLLSTRI PSTPATFISQDPLLPPVTPTPTVTPQVTSTSTPSWKEVLLDFAANNLHDPRLGADDLLAIVS KIQSGSISSAEAARLLEDHAERWFPHTWAPSKRKHRPLGPTPTSRQLRRANYAAVQKLFHLR KKDAASCVLSGQWRQAHSGPKPTVPGMDDYWAQIFEHPSAPDRREPTPQDDVAWSLLDAITA AEVTSALRDMGNSAAGLDRLSPSTLSTWDPSSIAAYFNLILATESVPKQIA**KARVIFLPKTE** **EPSSPSDFRPISISSVLLRAFHKILARRMTSTLRFSQLQHAFLKRDGCLEAGSLLHAILRKI** **HDDRVRASLIFLDISKAFDSVSFHTLLRVAKSSGLPPPLTNYLVSLYDHPTVQLESRTVTCG** **RGVRQGDPLSPLLFIMAFEDILQSSLPEVGFKLDGHKIPALAYADDLVLFARKPERLQEKLD** **GLVVSLAAAGMSVNAKKSSGLTITKIGKTKTMVLLPHIYTAGSTAIKPMGPSDTIRYLGVDF** NWKGRVAPKHTSKLDQMLKDLSMAPLKPYQRLDLLKNFAIPRLTHELVLGNAHRNTLKKMDV LVRSAVRRWLRLPKDTPLGYMYAHIRDGGVGLPCLNTSIPLIQKKRFAKLLSNPSPIQRTVV NQKSFKVIQRATDLPVLIGGQAIQSKKEARTAWRDDLRRSYDGRDLFGDPVDPTSHHWVSHP QRVFPRLHLRGIQLRGGLLPTRARNARGRDPTGTPTTCRGACGQVETLNHILQICDKTHDAR CARHNRIVQRLCQFLRKKGIKFWTEPLVPTDTSYIKPDIIVLTNSELLVMDVSIVSGPRMSY SWTLKAEKYGTAKCSSALQHWHGSPIPIRHSPIILSNRGLLYAQSGQALKSIGLTTRDLSDL CLLTIMGSLRCFDHYTRSTMDSSDRGCT	89
(23) R2AspMar RT	KARVIFLPKTEEPSSPSDFRPISISSVLLRAFHKILARRMTSTLRFSQLQHAFLKRDGCLEA GSLLHAILRKIHDDRVRASLIFLDISKAFDSVSFHTLLRVAKSSGLPPPLTNYLVSLYDHPT VQLESRTVTCGRGVRQGDPLSPLLFIMAFEDILQSSLPEVGFKLDGHKIPALAYADDLVLFA RKPERLQEKLDGLVVSLAAAGMSVNAKKSSGLTITKIGKTKTMVLLPHIYTAGSTAIKPMGP SDTIRYLGVDF	90
(23) R2AspMar 5'UTR	GGGTCGTAGTTACATGGCCTGGGCGCCTCCTCGTGGCGACCGCTGGAGATGATGGGGTACGC GAAAGGGGGCAAGTCCCTCGATCCGTCCCCGTCTGAGCTAAGACCCAGGCCACCCCTCTTCT GGGCCAACCACCCAGTTGAGGGTTTCTCTGCTGGTTACAGAAAGGAAAATTGTGAATGCCCC TGGTCCCTCCGGACATCAGGGTAACAGCAGAGACGTCGACCCTGAAGGTGCTCCCAGCCTTC CCTCGTGTCCGCCCGGCGATGTCACCTGCCACTGCGGCCGA	91
(23) R2AspMar 3'UTR	GAGGATGGAAGCAGCTCATGACCCGCCACAAGCGGCGATCCCCCTTCATAGTCGTCGTCGAT AGCGGTTATCCGCCTCAAAGTTTACTGTATGCTGTATTTCTTTGTTATTTACTGCTGTACGG TCTACTTTGATAATAAAATGATAA	92
(24) R2AzeSeq ORF	MLTYTEDPKPPGFTDHSSTGRCAGVESLERTPEYSRHLCTCGSEFQSKQGLVMHQFHRHPAE VNAKRLASLPVKKLVWTQEEETALLRHASQAWKKGMMKVELCACLLPQFPGRSGEAIKKRLQ RLKWVPPGENIEREVPERDSLPQLTPAIKGHVAPATLPPHKDLVVGDQGQNNNIKAPWRRAL LTNMADHLSETKLGVEKLHQIVDGLNSNQMSTDKCRHLMNEHAQQQFPHKWKPSSAWPETKA GLYSKREIRRANYAIVQKLYKLRKRDVAHTVLSGGWKNAYRGHAPLPKNMEEYWSAIFETPS ETDVTTPEVPTEIHWSVMEKITPSEVTNSLKGMRNSAPGLDRISAEELLTWHQPSLAAYYNL MLATGGPPEHLA**CSRVAFIPKIESPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQ** **HGFLQRDGCLEATALLHAILRKVHQNRKPCAMLFLDVAKAFDTVSHHTLFRVASAMGLPPPL** **VDYLEHLYNRSTVKIANKATKCGRGVRQGDPLSPLLFIMVMEDIVRKALPEIGFNLDEQRVD** **SLAYADDLVLLAEKSTRLQEKLHLLTEALGKAGMTLNANKSHGLTLNKISKRKQMVLTPTAY** **DCKGMMIKPMGTDDSVRYLGLHF**NWKGKIVPKHTCKLDSMLKEMTKAPLKPYQRLQLLKYHL VPKFTHELVLGHAHRNTLKKLDCLTRAAVRHWLRLPKDTPLGYLHANVKDGGLGIPCFSTSI PLLQQRRFEKIAANPATIFQVIQRQDSFRTQGRRLDLPCRLNGTVVNSRAEVRKEWGNVLAD SIDGRELRYPKVDKASHGWLQHPDRVFPRLHMRGIQLRGGLLPTKARSARGKDRQVAEKTCR GACHCTETLNHILQVCEITHDARCARHNRVLHRLEQLLRKKGLRTTLEPVIPSGPSFIKPDL IVETGTQTLVIDVSVVASYRMKESWDIKVEKYGSSERVNDIRKWAAIPHQVRLHQLPVIVSN RGLCYQPSGEGLRALGVTCRDITDLCLLSINGSLRCYDLYMRGTQGAKVHRLPKNTEGPRRV PPGLHWQ	93
(24) R2AzeSeq RT	CSRVAFIPKIESPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQHGFLQRDGCLEA TALLHAILRKVHQNRKPCAMLFLDVAKAFDTVSHHTLFRVASAMGLPPPLVDYLEHLYNRST VKIANKATKCGRGVRQGDPLSPLLFIMVMEDIVRKALPEIGFNLDEQRVDSLAYADDLVLLA EKSTRLQEKLHLLTEALGKAGMTLNANKSHGLTLNKISKRKQMVLTPTAYDCKGMMIKPMGT DDSVRYLGLHF	94
(24) R2AzeSeq 5'UTR	CCTATGACGCATTGATTGTTGCAAGGCGAATCGTTGTAGAGTTAAGAGTGGAAAACAGTTAT AAGGGTTGGGCACCTCCGTGTGGTGACCGCTGGATAGGGCTATTACTAGCCTGAACTAAGAT CCAACCCAAAACCCCTGTTTTGCACCAGAACAGTGGGTAGGCTGCTGAT	95
(24) R2AzeSeq 3'UTR	CCCCTCCAGGCTCTATCCTGTAATAGCGATAGGATTATCGCCTTAGAAGGCTTAAGGGGAAG GACGAGTAGGCCAATTCGTGTCCGAACCACTCCATACCTCTCCTCTATTCGCCTATCTAATA ATTTATGTTAACA	96
(25) R2AzeSeq2 ORF	MLTYTEDPKPPGFTDHSSTGRCAGVESLERTPEYSRHLCTCGSEFQSKQGLVMHQFHRHPAE VNAKRLASLPVKKLVWTQEEETALLRHASQAWKKGMMKVELCACLLPQFPGRSGEAIKKRLQ RLKWVPPGENIEREVPERDSLPQLTPAIKGHVAPATLPPHKDLVVGDQGQNNNIKAPWRRAL LTNMADHLSETKLGVEKLHQIVDGLNSNQMSTDKCRHLMNEHAQQQFPHKWKPSSAWPETKA	97

	GLYSKREIRRANYAIVQKLYKLRKRDVAHTVLSGGWKNAYRGHAPLPKNMEEYWSAIFETPS ETDVTTPEVPTEIHWSVMEKITPSEVTNSLKGMRNSAPGLDRISAEELLTWHQPSLAAYYNL MLATGGPPEHLA**CSRVAFIPKIESPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQ** **HGFLQRDGCLEATALLHAILRKVHQNRKPCAMLFLDVAKAFDTVSHHTLFRVASAMGLPPPL** **VDYLEHLYSRSTVKIANKATKCGRGVRQGDPLSPLLFIMVMEDIVRKALPEIGFNLDEQRVD** **SLAYADDLVLLAEKSTRLQEKLHLLTEALGKAGMTLNANKSHGLTLNKISKRKQIVLTPTAY** **DCKGMMIKPMGTDDSVRYLGLHF**NWKGKIVPKHTCKLDSMLKEMTKAPLKPYQRLQLLKYHL VPKFTHELVLGHAHRNTLKKLDCLTRAAVRHWLRLPKDTPLGYLHANVKDGGLGIPCFSTSI PLLQQRRFEKIAANPATIFQVIQRQDSFRTQGRRLDLPCRLNGTVVNSRAEVRKEWGNVLAD SIDGRELRYPKVDKASHGWLQHPDRVFPRLHMRGIQLRGGLLPTKARSARGKDRQVAEKTCR GACHCTETLNHILQVCEITHDARCARHNRVLHRLEQLLRKKGLRTTLEPVIPSGPSFIKPDL IVETGTQTLVIDVSVVASYRMKESWDIKVEKYGSSERVNDIRKWAAIPHQVRLHQLPVIVSN RGLCYQPSGEGLRALGVTCRDITDLCLLSINGSLRCYDLYMRGTQGAKVHRLPKNTEGPRRV PPGLHWQ	
(25) R2AzeSeq2 RT	CSRVAFIPKIESPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQHGFLQRDGCLEA TALLHAILRKVHQNRKPCAMLFLDVAKAFDTVSHHTLFRVASAMGLPPPLVDYLEHLYSRST VKIANKATKCGRGVRQGDPLSPLLFIMVMEDIVRKALPEIGFNLDEQRVDSLAYADDLVLLA EKSTRLQEKLHLLTEALGKAGMTLNANKSHGLTLNKISKRKQIVLTPTAYDCKGMMIKPMGT DDSVRYLGLHF	98
(25) R2AzeSeq2 5'UTR	CCTATGACGCATTGATTGTTGCAAGGCGAATCGTTGTAGAGTTAAGAGTGGAAAACAGTTAT AAGGGTTGGGCACCTCCGTGTGGTGACCGCTGGATAGGGCTATTACTAGCCTGAACTAAGAT CCAACCCAAAACCCCTGTTTTGCACCAGAACAGTGGGTAGGCTGCTGAT	99
(25) R2AzeSeq2 3'UTR	CCCCTCCAGGCTCTATCCTGTAATAGCGATAGGATTATCGCCTTAGAAGGCTTAAGGGGAAG GACGAGTAGGCCAATTCGTGTCCGAACCACTCCATACCTCTCCTCTATTCGCCTATCTAATA ATTTATGTTAACA	100
(26) R2AzeSeq3 ORF	MLTRSKDPRPPGSPDLSSTGSCDGVASLESSPEFSRHPCPCGSEFRTKQGLAMHRYHRHPAE VNAERLVSLPTKKLVWTQEEEAALLRHASQAWKKGMMKVELFTCLLPHFQGRSEEAIKKRLN RLQWVPPVDNPERDVPERNSLPLPIPAVSGHVAPATHPLPEAKMVEDQGRNTGSVESWRRDL LATVADHLSETKLGIEALSQILGGLKSGQISTDKCRALMNDHARQQFPHTWKPSSARPGAEA RKWSQREMRRANYATVQKLYKLRRRDAAHTVLSGRWRTAYRGHAPLPSNMEEYWSAIFETPS ETDAAIPEVPTEIFWSVMGRIAPSEVTSALKGMRNSAPGLDRISAEELLTWHQPSLAAYYNL MLAAGGPPEHLA**CSRVAFIPKNDNPQTPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQ** **HGFLQRDGCLEATALLHTILRKVHHNRKPCAMLFLDVAKAFDTVSHQTLFRVAAAMGLPPPL** **VDYLKNLYSRSTVRIADKATRCGRGVRQGDPLSPLLFIMVMEDIVRRVLPEVGFNLDGQRVD** **SLAYADDLVLLAEKSSRLQEKLHLLTEALQKAGMALNASKSHGLTLDKVSKRKQMVLTPTTY** **ECKGGTIKPMGTDDSVRYLGLHF**NWKGKIVPKHTGKLDSLLKELTKAPLKPHQRLQLLKGHL VPKFIHELVLGHAHRNTLKKLDCLTRAAVRHWLRLPQDTPIGYLHAHVKDGGLGIPCFSTSI PLLQQKRFEKIATNPATIFQIIQRQDSFRTQGRRLDRPCRLNGAVMTSKAEVREEWGNVLSN SVDGKELRCSEVDKASHGWLQHPGRVFPRLHMRGIQLRGGLLSTKARSARGKDRQVTDKTCR GSCHCPETLNHILQVCEITHDARCARHNRVLHRLEQMLRKKGLRTTLEPVIPSGSSYIKPDL IIETGTRRYVVDVSVVAGYRMKESWDIKVEKYGRPDRVNNIRKWAAIPEQDLVIQLPVIVSN RGLCYRPSGEGLRALGLTCRDISDLCLMSINGSLRCYDLYMRGTRSDKVRGRPRNTEGG	101
(26) R2AzeSeq3 RT	CSRVAFIPKNDNPQTPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQRDGCLEA TALLHTILRKVHHNRKPCAMLFLDVAKAFDTVSHQTLFRVAAAMGLPPPLVDYLKNLYSRST VRIADKATRCGRGVRQGDPLSPLLFIMVMEDIVRRVLPEVGFNLDGQRVDSLAYADDLVLLA EKSSRLQEKLHLLTEALQKAGMALNASKSHGLTLDKVSKRKQMVLTPTTYECKGGTIKPMGT DDSVRYLGLHF	102
(26) R2AzeSeq3 5'UTR	TCGCCCCACAACTTGACCTTGGTTTTGATGTTGAATGTGAATTGTTGCAAAGCGAATTGTTG TGGACGCGAGATTGACGAACAGTTGCACGGGTTGGACACCTCCCAGTGGTGAACGCCGGATA GGGCCTTGTGCCTGAGCTAAGAGCCAACCCAAACCCCTGTTACGTACCAGTACAGTGGGCAA GCTGCTGAAT	103
(26) R2AzeSeq3 3'UTR	GGGACTCCCACAGGACTGCATAGGCAGTGTCCCCTCCACGCACTATCCTGAAGTAGCGATGA GTATCGCCTTAGAAGGCCTGATGGGGGAGGCTGAGTGGACACAATGTGTCACAAACGCTCTT TACCTCTCCTCGTCTTGCCTGTCTAATAATATATGTTAACA	104
(27) R2ChaBot2 ORF	MASGGANEFRCDCGTLWPSARSLAMHNFHRHTEKVMAERREKQPAKRLRWSAEEEAALHRQA DRLWSEGMLSKDLCLLLTDFFPGRSAEAIKKRLQSTRWIPPGQRPPVSPEGNEESSLAEQFD FAMWSKDLLEEAVGNLQEPRMGAELLSRIARGLLTGEFSKEEGVKRLEDHACQWFPYRWKSS VPRKSGNPQMHSNKKIRRMNYAAIQRMYHRRLKDAATCVLSGQWRTAYQPRPEVEGLNEYWT KIFETRSVEDTGCLRTPTRTLYSLIGRIIPSEVTAALKTMRSSAPGLDRITASELLAWHQPS IASYYNLLLVLEVCPRYLS**CSRVAFLPKVPVPSAPDDLRPISITSVLLRAFHKILARRLNDE** **LHLTSLQFAFLQRDGCLDATTLLHSILRNIHHTRSPGAMLFLDISKAFDTVSHSTIARIITS** **AGFPIPLIHYICNLYSNSTVQLGEKMIKCGRGVRQGDPLSPLLFILVIEEILLASLPDVGYN** **FGGQRLGAIAYADDLVLLAEHPQRLQEKFKIFSEALHTTGMVLNPKKSIGLTISRDGKRKAI**	105

	VWQPHCYNYLGTPIRPLKIDESVKYLGLQFNWKGRINPRHVGKLDEMLGEISAAPLKPYQRI RILTSFLIPKLTHSLVLGLAYRKTLKRADRLVYAAVRSWLRLPKDTALGYFHASVADGGLGI PCLSTTIPLLQRDQFAKLISQTDPLSVIIREQEAFKVVGNIINRPCRINSRVVTSKEEARAE WKTSLSQSVDGSELTIPDVDLASHGWLKYPVRVFPRLFLRGIQLRGGVLPTKARSARGNNRV VVDRSCKGLCGTSETLNHVLQSCEITHNARCSRHNKIVQQLEYLLNKIGIKIWVEPIIPSGS SFIKPDLVVMKDNQLWVLDVSIVSGMRLRESWNIKIVKYGSTERVNDIVCWFGSPIAVKQLP IIISNRGLCSTFSGSGLRALGLSRRDITDLCLLTVKGSLMCYDLYNRSTRSSGR	
(27) R2ChaBot2 RT	CSRVAFLPKVPVPSAPDDLRPISITSVLLRAFHKILARRLNDELHLTSLQFAFLQRDGCLDA TTLLHSILRNIHHTRSPGAMLFLDISKAFDTVSHSTIARIITSAGFPIPLIHYICNLYSNST VQLGEKMIKCGRGVRQGDPLSPLLFILVIEEILLASLPDVGYNFGGQRLGAIAYADDLVLLA EHPQRLQEKFKIFSEALHTTGMVLNPKKSIGLTISRDGKRKAIVWQPHCYNYLGTPIRPLKI DESVKYLGLQF	106
(27) R2ChaBot2 5'UTR	CTTTCTTGTCATAGTTACAGGCCTGGACACCTCCGCGTGGTGCCCGCTGGATAGGGTATGGG ATCTTATATCCCATGCCTGAACTAAGTTCCAGGCCACCCTCCTCTCGGGAAGACTTAATTAC CCGTTAGGAGGTTGCTGCTGTTTTTTTGTTAAATTGATTAATTTATGTCCGGAGCCCTCTGG TTGACCGGTGATACAGCAGGTAAGTGAGGTAAGACAGTTGTCAAGAATCCACA	107
(27) R2ChaBot2 3'UTR	TTTTCGATCTGGATTCGATGAACGACGATTGTCTCCTGTTATGATTTGAATTATTTTCTTGG AGATGTTGTATTTATAAATTTGCTAAATTATAATTGTATCTTTTTGAACATACTCCAATAAT AAATGTAAA	108
(28) R2CroAda ORF	MLTDTVVLRPPGLPDPSCPGSCPRVASPKSDAECPKHSCPCGSEFGTKRGLAMHRFHCHPAE VNAERLATLPTKKLVWTQEEMDGLLRHANGAWKEGMLKTQIFTSLMPHFPGRSAEAIKKRLQ SLKWVPARDNPERNSQPLHATSGNVAPATLPTPDPSRVDDQGANIEDPAASWRRALLAAAAD HLKEPKLGAGTLKQITQDLLNGKISSGECRVLMGEHAQHWFPHTWKPSSARPEPKARQWKAR EIRRANYAAVQKLYKLRRRDAAQAVFSGSWRNAYRGRAPQPDKMDQYWAEVFETASATDVTP PGSPSEIYWSVVSKITPSEVADALKGMRNSAPGLDRITAEELLTWDHPSLAAYYNLMLAAGG PPEHLA**CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDTLHLSPFQHGFLQR DGCLEATALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLRC LYSRSTVRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYAD DLVLLAEKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVTTPTTYECEGEP IKPMGTDDSVRYLGLHF**NWKGRIVPKHTGKLDSLLKELTKAPLKPYQRLQLLKFHAVPKFTH ELVLGHAHRNTVKKLDCLTRAAVRKWLRLPQDTPLGYLHANVKDGGLGIPCFSTSIPLLQKK RFEKIATNPATIFQIMQRQDSFRTQGRRLDKPCRLNSTVVTSKTEVREEWGNMLSNSIDGRE LRHPEVDKASHEWLLRPDRVFPRLHMRGIQLRGGLLPTKARSGRGNNRQVTSKTCRGACQNV ETLNHILQVCDVTHDARCARHNRVLHRLERLLHKKGLRTTLEPVIPAGSSFIKPDLIIETDT HTYVLDVSVVAGYRMTESWDIKVEKYGRPDRVEAIRKWAAIPATKRVEHLPVIVSNRGLCYQ PSGDGLRALGLTSRDISDLCLLAISGSLRCYDLYMRGTRAV	109
(28) R2CroAda RT	CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDTLHLSPFQHGFLQRDGCLEA TALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLRCLYSRST VRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYADDLVLLA EKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVTTPTTYECEGEPIKPMGT DDSVRYLGLHF	110
(28) R2CroAda 5'UTR	TCGACCATTTGGTGTTGATATTGATACTGGCGACTGGCTAGTTGGTGTTGGTTGAAAGCGCT GGCGAGGTTGAGCTGATAATTGGTTGCACGGGTTGGGCACCTCCTAGTGGTGACCGCTGGAT AGGGCTTCGGCCTGAGCCAAGATCCAACCCACCCTCCGTTGCTGGCAGACCCCAGTAACTGA GATTAAACAGCTGGT	111
(28) R2CroAda 3'UTR	GACTACCGAGCGCCACCGGAGGGGACCCAGAGGACCGTCTCGGTGATTAGGCGGTGACCGCC CTGGCCTCTACCTGAAATAGCGATCAAATCGCCTTAGAGGGCTTGATTTGGGGGGTTGGGCA GACACTCTGTGTCATTACTGCCCTTTACCTTTCCACTTCTTGCCTGTCTAATAATAAATGTT AACA	112
(29) R2CucCan ORF	MGRQTRMHFMSVATQTYYGDWRKVKVNVIESDHNPPIDVEWEISGLEFVNSGLKEEGYQPFI GVSDLAAFPDVYEALKTPDLVGNVNSPVLTVVEEREEEREDERQSSHGEVAELNPAPSTSMS GSCTELSADGIPVVKVPDKQPTCPSCGALFVKVSGLIEHLKRAHGKDKIWFECVKCGRQNVK YHSIVCHYPKCPGRIEVVESDDFVCGECKRSFSSKIGLSQHKRLAHPALRNEERIENSLPKE STRRGVHRSCWTDEEVELLLKWERQFEGEKNINKLIAEKIPSKSAKQISDKRRCLSKQKEKL KGKDLEEVSHDIEGEHPAKSEGGLNLGLRDIYKRILTEWVSAGKITTFNQTYSDLLEGREHH MSSITEATMTYLASLGKVSGELEASMREGTRKATKQLSSRKHSKGWMKRRAAKKGKFYRFQR LFYRDRVKLAEIILDGRESLKCEIPLDEVHDVFKSRWESQGTHRGLGVFKSQGMADNSVFLT MITEREISKNIKEMKKNAAPGPDGLRRRDIESCDPQFKDLMEIYNLWLVSGKVPDVVK**ECRT ILIPKSTNAERLNDINNWRPITIGSTILRLFTKILTARLTKACPLNPRQRGFIRSAGCAENL KLLQLIIKNAKREHHALGVVFIDLEKAFDSVSHEHLIEALENRGVEEHIITLIKDLYRNVTT SVEIKSGKSESIKINIGVKQGDPMSPLLFNLSIDPLLCRLEEKGHGFKHGSRCITSLAFADD LVLLSESWEGMVENIKILEAFCSLTGLKTQGIKCQGFYISPTKDSYTINNCPVWEINGTPLR MIDPEKSGKYLGLEI**NPWVGISKPELEKDLSEWLDRIGKASLKPLQKVDILKGYAIPRMIYK	113

	ADQAEVKASCLEALDLMIRSNVKEWLHLPPCTCDAILYSRSRDGGLGINRLSAIIPSIQARR LHRLAQSSDEAIREFVKKVSIEKEYEKLWITAGGDKNKIPSIWDPRSVMQISEITSGENPGS EWEVSAPKTIFPRPSDWRKREFNNWTNLGSQGIGIRNFVNDKISNDWIINYKGIPHRKLLTA LQLRANVYPNREFLSRGRREAHNKTCRHCPGSTESCAHIIGYCPRVQEARIKRHNIVCKILA DEAKKKEWKIYMEPLIRNDNNELFKPDLIIVKDGAAMVIDVTIRYESTLSTLKEAATEKVKK YEHLKDQIKDLTGASIVRFFGFPLGARGKWYDGNSDVLSDIGLSQSRINKTAIRLSNLALFS SVDIVHIFTSSAATGVT	
(29) R2CucCan RT	ECRTILIPKSTNAERLNDINNWRPITIGSTILRLFTKILTARLTKACPLNPRQRGFIRSAGC AENLKLLQLIIKNAKREHHALGVVFIDLEKAFDSVSHEHLIEALENRGVEEHIITLIKDLYR NVTTSVEIKSGKSESIKINIGVKQGDPMSPLLFNLSIDPLLCRLEEKGHGFKHGSRCITSLA FADDLVLLSESWEGMVENIKILEAFCSLTGLKTQGIKCQGFYISPTKDSYTINNCPVWEING TPLRMIDPEKSGKYLGLEI	114
(29) R2CucCan 5'UTR	ACGGGACAATAGTTACTACCTGGGCTCAGCTGCAGAGGTCACACCTCCTCGTGGTCCCGCTG GATGCCTGTAAAGGTAACCAAGTTGGCCTCTGCCCCAGGGAAAAGTTTTGGAGTTAAAGGTT GCTCCGCCCTATCGAGTAAACTCACCACTTACTCGGTAGTAGTTACAAGAGGTGAAACAAAA ATATTTGTGGCTTGAATTCGTTTAAGCTATATCCCTTGCCCACGGGAAGACATAGTTGGGCG GGCAGCTGACCACCGAGAGGGGGTGGTTACGTTGAAAGTCATCCCTCAAA	115
(29) R2CucCan 3'UTR	ATTGTTCTGGCTTTAAAGTTGAATTGTTCTGGTTCTAAAGTACAGTTGTTAATTTTATAGTT GTTAATTTTATGTATTTTAACAGTTATTTATTTATTATGTTTTCCAGATTTTAATAAACAAT ATTGACAGTCTGGTATACAGTACGGGATCTCGTAGTGATTACAGAATCATGACGCTGTCACT TTGGACATGACGATAGGTGGACGGGCCACCCTCACTTAACCCGGAAAAGGGACGCATCACCA TATGCGTTCGACAC	116
(30) R2HalLeu ORF	MGRPMRISFVSVATQTSPLDLRDNAVSRSIRDHKPGGLLEGRFEIRGLEFVNSDLEGLKSQQ SPTLSDLLAVLPEVYEVPNSFSLRDKNTCPTPVEDENEEEESSPLGAVAGTSPVPSTPVSRS RDDLPAESTPVVKIPDKNTSCPCCGNVTGKLSGLVEHLKRAHGKKKILFQCARCGRTDVKHH SIACHFPKCKGPVEASPERSWRCEVCEKTFETKIGLGQHKRLVHPVTRNLERITVARPKTGT NMRGAHRRCWTEDEVELLRKLDKQYEGQKNINKLIAEHIPTKTTKQISDKRRDLHKSPVKKG KDQGRKKHLELEGEHLAEGEHLVCVGSEMEGGLRSHYRKTISEWVTSGKLSILHGMFQRVIE GEEDLTTVADKAMEDCFACLAFKSSELKAGQGRPRQVAQTARPSEPPKGWMKKRAVKRGQFL RFQRLFYLDRGKLAGIILDDVESLKCGIPQDEIHNVFKSRWESPGTFKGLGAFRSKGLADNS AFEAMITAKEISKNIQEMNKNSAPGPDRLSLRDIIKIDPHFLQLMEVFNLWLVTGIVPDVVK **ECRTVLISKSSLPERQMDINNWRPITIGSIILRLFSRILTARLAKACPINPQQKGFIRSVGC AENLKLLQLLIKNAKKEHRPLGVVFIDLAKAFDTVSHYHLIAVLKRKGVNHQIITLITNMYQ NTRTRIAMKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGRRYHHGSKCVTSMA FADDLVLLSGSWEGMQSNIKILEAFCNLTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQING APLNMIEPGSSEKYLGLRI**DPWTGISKPKLLMKVRDWLQRIGEAPLKPFQKVDILKGYTIPR LIYLADQADVKATYLEELVLTIRTTVKEWLHLPPSTCDAILYSSTQDGGLGITRLSCLIPSI QARRLHRIAQSSDNVIRSVVKEEGMEKEFEKLWLTAGGDKNKIPSIWDPSAIRVTSEALGDD EFVSEWEVPAPKTIFPKPCDWRKQEFINWTKLQSQGRGIANFEKDKISNNWLKRYRGIPHRK LLTALQLSANVYPTREFLARGRQESQIKSCQHCQAENESSAHIIGYCPAVQDARIKRHNQPC DLLITEVKKKDCIVFQEPILKDERNEIYKPDLVFVKGDQAMVIDVTIRYESKLSSLVDAVAE KVKKYGYLKSQIQELTNAKNIKFKGFPVGARGKWYSGNYKVLKTLGLSNSRQEKVACYLLNK ALFTSVDIMYIFMSQSRTIVRT	117
(30) R2HalLeu RT	ECRTVLISKSSLPERQMDINNWRPITIGSIILRLFSRILTARLAKACPINPQQKGFIRSVGC AENLKLLQLLIKNAKKEHRPLGVVFIDLAKAFDTVSHYHLIAVLKRKGVNHQIITLITNMYQ NTRTRIAMKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGRRYHHGSKCVTSMA FADDLVLLSGSWEGMQSNIKILEAFCNLTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQING APLNMIEPGSSEKYLGLRI	118
(30) R2HalLeu 5'UTR	ACGGGATCATAGTTACTAACTGGGCAGAGCTGCAGAGGTTACACCTCCTCGTGGTCCCGCTG GATGCCCTTCGGGGTAACCAAGTTAACCTCTGCCCCAGTGTGGGAATTATAGCTACTTCCCC CCTATCGTCTGGCCTCACCAACCATTCGATAGATTCGACTTGGTGATCTAAATTCCTGGCAC CGATTCCTTGTCGCTGCCCAAAGCCTCGCCTACGGGCACAGTTGGGCTCAGTCACACAGCAC TTGAGGGGGTGTTGTTTGACTTAGACTACTCCTCTATCTGGACTGGCAAA	119
(30) R2HalLeu 3'UTR	TCATCTGTTACTGTTCTGTTTTATATCGTTTTTTGCTGTATTACTACTTACTGCTTTTGTTG TATTTACTGTTCTGCCTTATATTGTCTTTGCTGCATTACCACTTATTGTTTTGCTGTATTAC TACTTAATTCTTTAATAAAACTACCCCTGGTACTCCTGCTCTCTTACCCTGGCACTGGAAAG GTCGGGATAGTAAATTGGAACTCGTGTTCATGACGTGTCACGCTAACCTAAATTTGGACAAA ATATTAATTGGAATTGTCAAAATTTTTACCAATTTGGACCACTTTACCGGATTTGACACAGG TGGACGGGCCACCTTTACTTAACTCGGAAAAGAAACATATATATTTTTATGTGTTTGTTAAA	120
(31) R2AqCh ORF	MGRPMRISFVSVATQTSPLDLRDNVVSCPTKDHGGLLEIRFEIRGLEFVNSDLVGLEGQQSS TLLDCLALLPEVYEVSNSLSLRDKNICPTPVEEENEEEEESSPLGAAARISPLPSTPVARSR DDLPAEGTPVVKIPDKNTACPCCGIVAGKLSGLIEHLKRAHGKKKISFQCARCGRTDVRHHS IACHFPKCKGLVEASPERGWKCEVCEKTFETKIGLGQHKRLVHPVTRNLERISAARPNAGTM	121

	RGAHRRCWTEDEVELLRKLDKQYEGQKNINKLIAEHIPTKTTKQISDKRRDLHKSPVKKGKD QAREKHLELKGEYPEEGEHLVCAGSKTEGGLRSCYRKTISEWVTSGKLSILHAMFEQVIEGK EDLTTVTEKATEDCFACLAFKSSELKAGQGRSHQVAQIARPSEPPKEWMKKRAVKRGQFLRF QRLFYLDRGKLAGIILDDVESLKCGIPQDEIHDVFKSRWESPGTFKGLGAFRSRGQADNSAF EAMITAKEISKNIKEMNKNSAPGPDRLSLRDIIKIDPHFVQLMEVFNLWLVTGVVPDMVK**EC RTVLISKSSLPERLMDINNWRPITIGSIILRLFSRILTARLAKACPINPRQKGFIRSAGCAE NLKLLQLLIKNAKKEHRPLGVVFIDLAKAFDTVSHYHLIAVLKQKGVDHHIITLITNMYQNI KTRIAVKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGNGYHHGSKRVTSMAFA DDLVLLSGSWEGMQSNIKILEAFCDLTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQVNGTP LNMIEPGSSEKYLGLRI**DPWTGISKPELLLKIRDWLQRIGEAPLKPFQKVDILKGYTIPRLI YLADQADVKPTYLEELDLTIRTTVKEWLHLPPCTCDAILYSSTRDGGLGITRLSGLIPSIQA RRLHRIAQSSDEVIRSIVKQEGIEKEFEKLWLTAGGDRNKIPSIWDPSAIRVTSEELGDDEV VSEWEVPTPKTIFPKPCDWRKQEFINWTKLQSQGRGITNFEKDKISNNWLERYRGIPHRKLL TALQLRANVYPTREFLARGRQESQIKSCRHCQAANKSSANIIGYCPAVQDASIKRHNQLCDL IITEVKKKDWIVFQEPILKDERNEIYKPDLVFVNGDQAMVIDITIRYESKLSSLADAAAEKV KKYGHLKSRINELMNAKNIKFKGFPVGARGKWYSGNYKVLKALGISNSRQEKIARCISRKAL FTSVDIVRMFASHSRSIVRT	
(31) R2AqCh RT	ECRTVLISKSSLPERLMDINNWRPITIGSIILRLFSRILTARLAKACPINPRQKGFIRSAGC AENLKLLQLLIKNAKKEHRPLGVVFIDLAKAFDTVSHYHLIAVLKQKGVDHHIITLITNMYQ NIKTRIAVKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGNGYHHGSKRVTSMA FADDLVLLSGSWEGMQSNIKILEAFCDLTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQVNG TPLNMIEPGSSEKYLGLRI	122
(31) R2AqCh 5'UTR	ACGGGATCATAGTTACTAACTGGGCTGAGCTGCAGAGGTTACACCTCCTCGTGGTCCCGCTG GATGCCCTTCGGGGTAACCAAGTTAACCTCTGCCCCAGTGTGGGAATTATAGCTACTTCCCC CCTATCGTCTGGCCTCACCAACCATTCGATAGATTTGATTTGGTGATTCAAATTCCTGGCAC CGATTTCTCGTCGCTGCCCAAAGCCTCGCCTACGGGCATAGTTGGGCTCAGTCATACGACAG TTGAGGGGGTGTTGTTTGACTTATATTACTCCTCTAACTGGACTGGCAAA	123
(31) R2AqCh 3'UTR	TCATCTGTTACTGTTCTGTTTTATATTGGTTTTTGCTGTATTACTACTTACTGCTTTTGTTG TATTTACTGTTCTGCCTTATATTGTCTTTGCTGCAGTACCACTTATTGCTTTGCTGTATTAC TACTTAATTCTTTAATAAAACTACCCCTGGTACTCCTGCTCTCTTACCCTGGCACTGGAAAG GTCGGGATAGTAAATTGGAACTCATATTCATGACGTGTCACACTAACCTAAATTTGGACAAA ATATTAATTGGAAATGTCAAAATTTTTACCAATTTGGACCACTTTACCGGATTTGACACAGG TGGACGGGCCACCTTTACTTAACTCGGAAAAGAAACATATATAATTTTTATGTGTTTGTTAA A	124
(32) R2HarHar ORF	MGRPMRISFVSVATQTSPLDLRDNVVSCSTKDHGELLEVRLEIRGLEFVNSDWVGPEGQQSS TLIDRLALFPEVYEVSNSLSLRDKNICPTPGESENEEEEESSPLGATARISPLPSTPVSRPR DDLPAEGTPVVKVPDKNTSCPCCGNVTGKLSGLIEHLKRAHGKKKILFQCARCGRTDVRHHS IACHFPKCKGLVEASPERSWKCEVCEKTFETKIGLGQHKRLVHPVTRNLERITAARPKAGTM RGAHRRCWTEDEIELLRKLDKQYEGQKNINKLIAEHIPTKTAKQISDKRRDLHKSPVKKKGK ELGRKEHLELEGEHLAEGENLVCVGSEMEGGLRSRYRKTISEWVTSGKLSILHEMFEQVIEG EEDLTTVADKAMKDCFACLAFKSSELKAGQDRPRQVAQTARPSEPPKGWMKKRAVKRGQFLR FQRLFYLDRGKLAGIILDDVESLKCGIPQDEIHGVFKSRWESPGTFKGLGAFRSRGQADNSA FEAMITAKEISKNIKEMNKNSAPGPDRLSLRDIIKIDPHFAQLMEVFNLWLVTGVVPDVVK**E CRTVLISKSSLPERQMDINNWRPITIGSIILRLFSRILTARLAKACPIHPRQKGFIRSAGCA ENLKLLQLLIKNAKKDHRPLGVVFIDLAKAFDTVSHHHLIAVLKQKGVDHHIITLITNMYQN IKTRIAMKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGNGYHHGSKCVTSMAF ADDLVLLSGSWEGMQSNIKILEAFCGLTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQINGT PLNMIEPGSSEKYLGLRI**DPWTGISKPELLMKIRDWLQRIGEAPLKPFQKVDILKGYTIPRL IYLADQADVKATYLEELDLTIRTTVKDWLHLPPSTCDAILYSSTRDGGLGITRLSGLIPSIQ ARRLHRIAQSSDEVIRSIMKQEGIEKEFEKLWVTAGGDKNKIPSIWDPSAIRVTSEELGDDE FVSEWEVPAPKTIFPKPCDWRKQEFINWTKLQSQGRGIANFEKDKISNNWLDRYRGIPHRKL LTALQLRANVYPTREFLARGRQESQIKSCRHCQAENESSAHIIGYCPAVQDARIKRHNQLCD LLITEVKRKDWVVFQEPILKDEQNEIYKPDLVFVKGDQALVIDVTIRYESKLSSLADAAAEK VKKYGHLKSQIQELTNAKNIKFKGFPVGARGKWYSGNYEVLKALGISNSRQDKIARCMSRKA LFTSVDIVHMFASQSRSIVRT	125
(32) R2HarHar RT	ECRTVLISKSSLPERQMDINNWRPITIGSIILRLFSRILTARLAKACPIHPRQKGFIRSAGC AENLKLLQLLIKNAKKDHRPLGVVFIDLAKAFDTVSHHHLIAVLKQKGVDHHIITLITNMYQ NIKTRIAMKNEQSDPIGIQIGVKQGDPMSPLLFNLSVDPLLCKLEEEGNGYHHGSKCVTSMA FADDLVLLSGSWEGMQSNIKILEAFCGLTGLRTQGEKCHGFYIQPTKDSYTINDCPPWQING TPLNMIEPGSSEKYLGLRI	126
(32) R2HarHar 5'UTR	ACGGGATCATAGTTACTAACTGGGCAGAGCTGCAGAGGTTACACCTCCTCGTGGTCCCGCTG GATGCCCTTCGGGGTAACCAAGTTAACCTCTGCCCCAGTGTGGGAATTATAGCTACTTCCCC CCTATCGTCTGGCCTCACCAACCATTCGATAGACTCTATTGGGTGATCTAAATTCCTGGCAC	127

	CGATTCTTGATCGCTGCCCAAAGCCTCGCCCACGGGCACAGTTGGGCTCAGTCATACAGCAC TTGAGGGGGTGTTGTTTGACTTATACTACTCCTCTATCTGGACTGGCAAA	
(32) R2HarHar 3'UTR	TCACCTGTTACTGTTCTGTTTTTTATATCGGGTTTGCTATACTATTACGTGTTGTTCTTATT GTATTTACTGTTCTGTCTTATGTTGTCTTTGCTGTATTATCAGTTATTGCTTTGCTGTATTA CTACTCAATTCTTTAATAAAACTACCCCTGGTACTCCTGCTCTGTTACCTTGGCACTGGAAA GGTCGGGATAGTAAATTGGAACTCGTTTTCATGACGTGTCACACCAACCTAAATTTGGACAA AATACTAATTGGAAATGTCAAAATTTTACCAATTTGGACCATTTTACCGGATTTGACACAGG TGGACGGGCCACCTTTACTTAACTCGGAAAAGAAACATATATAATTTTTATGTGTTTGTTAA A	128
(33) R2CorAlt ORF	MGRPLKLLYVSTATQTDSKDWKPVETEINKTSTPIITSQEIDLQQLEFASSKEIENWQATQS LKQGDNLMTFPEPYESLTPDSVRGNEINATDNDNESNRGEDDSFILATESILDSQKDQTPEE IPLVKLPDKNPSCPCCKTKISGVQRLVEHLLRLHGKRKIIFQCNKCGLQSDNYRSTVCHYPL CKGPKEMVTPGEWICEECERSFQSKIGLGQHKRLSHPEVRNQERIAATRPKVRTARASARKC WSKEEEELLESLLKRYEGERYINKLIAEHIPSKTMKQISDKRRYMIPVQKERKDQEAHVQVE RRKELVTSVETSTQDQLRAHYVTKIEKGIRSRKIKNYPRAFQKIVEGTTDVHAIINDITKEW YGHLEARCQMDKISNKKPRRGQNQDTQSKRKATKWMKKRAIKRGQYLRFQKLFYKDRGKLAR IILDNIESVKCEIPLPEVYSVFKSRWESTGGFKGLGKFRSKGMADNSAFKAMITEKEIIKNI NEINKKAAPGPDGITIKDLTKMDPNYSQTTEIFNLWLTSGKIPDMIR**GCRTILLPKSTKPDR LKDINNWRPITIGSVILRLFSRILTARLAKACPINPRQRGFIKSPGCSENLKLLQLLIKNAK RERKELGVVFVDIAKAFDTVSHEHIITGLKQKGVDAHIIDLITNMYENSYTQIGTKNEQSDP IQIKVGVKQGDPMSPLLFNLAMDPLLCELEQYGKGFHHRDNSITAMAFADDLVLLSGSWEGM KENIKILETFCDLTGLKTQGEKCHGFFIKPTRDAYTINDCPPWTIGNTPLNMVCPGSSEKYL GLQI**DPWIQIAKPDVTSRIESWLDRITQAPLKPFQRVDILKTYAIPRVIYLADHAEIKAGFL EALDLKIRATVKKWLHLPANTCDAILYSSIKDGGLGITKLQALIPSVQARRLHRLAQCSDNT LRTVLQEENITKTYEKLWITAGGDKDKIPNIWTTKPMVEISTLGEEETSEWEVSTPRNIYPR PCNWKKKEFQKWTQLQSQGIGISNFEGDTISNSWIKNYKGIPHRKLLTALQLRANVYPTREY LARGQKEAYIRSCRHCQAGNESCAHIIGQCPVVKDARIKRHNYICGMLVKEAKKKDWVVYQE PNIRDRDGELFKPDLIFVKDKQALVVDVTVRYEADDTTLEKAEKEKVKKYQHLEKEVQNLTN TDKVTFMGFPLGARGKWYTKNFDLLQTLGLSASKWERTARALASRALLTSVDIIHMFASKSS HSSVS	129
(33) R2CorAlt RT	GCRTILLPKSTKPDRLKDINNWRPITIGSVILRLFSRILTARLAKACPINPRQRGFIKSPGC SENLKLLQLLIKNAKRERKELGVVFVDIAKAFDTVSHEHIITGLKQKGVDAHIIDLITNMYE NSYTQIGTKNEQSDPIQIKVGVKQGDPMSPLLFNLAMDPLLCELEQYGKGFHHRDNSITAMA FADDLVLLSGSWEGMKENIKILETFCDLTGLKTQGEKCHGFFIKPTRDAYTINDCPPWTIGN TPLNMVCPGSSEKYLGLQI	130
(33) R2CorAlt 5'UTR	AAGAGTATAGTTACTAACTGGGCTTAGCTGCAGAGATCGCACCTCCCCGTGGTCCCGCTGGA TGCCCTTTAAGGGTAACTGAGCCGATCTCTGCCCCAGTCCGGAGAAGAGGTTACACACCTCT TTCTCCCTATCGTCTGGATTCACTACCCAGCGATAGTAATTTAAAATAGTGATACACAATCA TTTTGGTATTAACTTATAAGAGTTAACTACCCAGTGCCTTGCCCACAGGTTAAAAGTTGGGC ATAGGTTGACAGATATTGAGGGGTACTGTTAACCTTATTCTCCCTCACTTGGACTGACAAAT TACCGGATTTGACACA	131
(33) R2CorAlt 3'UTR	TTCATTTTGTTGTTATCTGTTAGAACTCACTGTTATTTACTCATGTATACTGTCTCTTTATC ATATTTTATTGTCTTATTGATGTTATTAGATTTTCACCTGTAGTTTACTAGACTTTTACTGT TATTAAAGCTGTGGTACCTATTATGTCCCTATACCCCGAATGTGGTGAATCTCGTAATGATT ACTTCCCGCGCTCACTTGACATGACGATGGTGGACGGGCCACCCCTAACCCGGAAAAGGTAC ATACACTTTGTGTATGTTCGTTTAA	132
(34) R2GopFla ORF	MMAALASAQLRPSRRANSTILDTEIVPWGRCISRPFEPPMRYHVKVSTVATQTDSPGTTDGV ALRRRESTQTLLTAGEMYPIPGTSQTSFTDGVTSGNTFAQTEVEEPLPPLPMEILNSVFGET VRPGTSPRPCVPTTPASEKEGDTEGSATLAENIDNSCVPTDASTVTTTEGDGETRAQGKVLV PLPHDNPSCPFCGDHVGKPTALNVHLKRHHGGKDVAFQCALCGKIDPNPHSILCHIPKCKGL TLQVPAGDWACEVCAKQFSTKSGLSQHKRRAHPVVRNTERILASQPKQNSQRGKHKSCWSAD EELLLLDFDKMYSGDRCINKRIAEAIGTKTAKQVSEKRRLLNLYPRATSVTADQGPGASPHY LEARSDAPSPIKGLHSIYMRKVQQRLSKEQQGQQHKGQEKLESEILSAWVAGDPNVRSLVEA ASIDVLSSFLAPESGTTKKRARAKVSKPKSGKGKSWMKKRAAKRSTYMRFQHLFENDRCKLA SIILDGTERLQCQIPLAEVFQAYKQKWETQTAFQGLGQFKSQTRADNSAFDLLISAKEVIKN IKEMNKNSAPGPDKVSVKDLVRVDPEGDILAGLFNTWLVTGIIPDGIK**TCRSLLIPKTMDPE ALKDLGSWRPLTIGSIVLRLFSRILTNRLAKACPLNERQRGFIAAPGCSENLKVLHTIMKQA KRDKKPLGVVFIDIAKAFDSVSHKHIMWVLQQRGLDQHVINIICDSYRNVHTRMEIGKELTE PIAIEVGVKQGDPMSPLLFNLALDPLISTLEMAGKGFSVGTHKVTALAFADDLVMLSDSWEG MNANIRILESFCTLSGLKVQAKKCHGFQINPTHDSFTINNCEAWKIGATPLNMIIPGESEKY LGVKV**DPWIGFTKPALSEKLDTWLERINRAPLKPSQKLAMLNTFTVPRVIYLADHTDCKQAL LSSLDDKIRRAVKEWLHLPADTCNGLLYARSRDGGLGVTRLASLIPSIQARRLHRIAHSEDA TIRSIALANDIESEYRKLWLVAGGTQDTMPAITDAVIVDHKLPQHVLDKLTEWEKPAPRVIY	133

	PIPCNWRNAEFSQWKDLPCQGSGVSHFENDPTSNDWLTHHKGFSERQFLTALKLRANVYPTR EYLGRGKTNSIVRCRHCSAPYESLSHILGQCPAVQGARIRRHNKLCEILMREARKLRWEIYQ EPHLRTAQNELRKPDLIFAREGTALVVDVTVRFEYKDKTFSDAAAEKVRHYQSLTEEIRKLT GANKVSYFGLPLGARGKWPVINDRVLSALGLSVSAKGRAARLLSRRALLYTTDVINTFESIG KGFSTPAETKPDGTPKRQKGIS	
(34) R2GopFla RT	TCRSLLIPKTMDPEALKDLGSWRPLTIGSIVLRLFSRILTNRLAKACPLNERQRGFIAAPGC SENLKVLHTIMKQAKRDKKPLGVVFIDIAKAFDSVSHKHIMWVLQQRGLDQHVINIICDSYR NVHTRMEIGKELTEPIAIEVGVKQGDPMSPLLFNLALDPLISTLEMAGKGFSVGTHKVTALA FADDLVMLSDSWEGMNANIRILESFCTLSGLKVQAKKCHGFQINPTHDSFTINNCEAWKIGA TPLNMIIPGESEKYLGVKV	134
(34) R2GopFla 5'UTR	CGCTGTGTCAAGAGTCTGGTTGCTGGGCCGCAGTCCTGAGATCTCAACCTCCCCGTGGTCCC GCTGGACAACTCTCTGGGAGCCTCGATTTTGGGAGTTGACTGACTGAGACTCGGCTGCTTTC GATCCTCCTGCCCTACTGTTCTGGTGCACCGGCTAGGCAGTAGAACCCTACGGTGAAAAATT CA	135
(34) R2GopFla 3'UTR	TGATTACATTCATACTAATTCATGATCGCGATCACTGGACAAGGCAACCAGGGTTGGACGGG CCACCCGGGGTTTGGAACACCCCAAATAACTCGAAACAGAAACCCGCGTGGGTTTGTTAAA	136
(35) R2MelCri ORF	MAKIPDPFLGGGVWGPPVRLSRVSVGTQTSPADWVDLVSRALPGTKSISQGVDLVSLFPRHW VDLLSKTDQVDLVSQNCLTSKTTNLAENELDLLVNFCLEVYRSDLHEHVHEGVHFSVDWEVL GGLPEVYEQLAPQPCVGDNLDPSPPQDSEPFVLGEGSGKEVSEGTPEALPPVPGEHGGQSPD NIVKVTVPDKNPPCPCCGTRVNSVLSLIEHLKGSHGKKRVCFRCAKCGKENINYHSVICHFP KCRGPEIGKVPAGEWICEVCGRDFATKIGLGQHKRHMHAVERNQERIVASQPKETSNRGAHK RCWTKEEEELLIKLVEQFEGNKNINKLIAEHITTKTAKQISDKRRLLPKKLTTDGADKEPGM SLRPRRAAGGVGTEPGSSQQAQAGAEDNGPGKNHLPGRPAAGGKTMEKIRRPADKDNGRQKT SGQQREGQLQAHYHKTIKESLSAGAINNFPRAFKQLMDGQEMQTMVNQAAQDCFGCLESISQ IRTATRGKKDNLGTHEKQPKKPFQRWMKTRAIEKGNYLRFQRLFHLDRGKLAKIILDDIECL SCDIPPSEIYSVFKARWETPGNFAGLGDFEIKGKADNSAFRDLITAKEIERNVQEMSKGSAP GPDGITLGDIRKMDPGYSRTAELFNLWLTSGEIPDMVR**GCRTVLIPKSTQPERLKDINNWRP** **ITIGSILLRLFSRIVTARMSKACPLNPRQRGFIRAAGCSENLKLLQTIIRTAKSEHRPLGVV** **FVDIAKAFDTVSHQHIIHALQQRGVDPHIIGLVNNMYKDISTYVTTKRNTQTDKIQIRVGVK** **QGDPLSPLLFNLAMDPLLCKLEERGKGFHRGQRSITAMAFADDLVLLSDSWENMKTNIKILE** **TFCDLTGLKTQGQKCHGFYIKPTKDSYTINNCAAWTINGTPLNMINPGESEKYLGLQF**DPWT GIAKTNLTTKLEFWLERIDQAPLKPLQKLDILKTYTIPRLTYLADHTEMKAGALEALDQKIR TAVKDWLHLPSCTCDAILYSSTKDGGLGVTKLAGLIPSVQARRLHRIAQSSDETMKEFLEKH QMEQLYKKLWIQAGGESEKMPSIWEALPATVPSNSADTTSEWEAPTLKTKFPKPCDWRRKEF EKWTKLVSQGRGIKNFEGDKISNDWLQYYRRIPHRKLLTALQLRANVYPTREFLSRGREENY IKACRHCNADNETCAHIIGYCPVTQDARIKRHNHICDVLIEEAKKKDWVVFKEPHIRDTIKE LYKPDLIFLKDGRAFVVDVTVRYEASKTSLEEAAAEKVNKYKHLGSEVRNLTNAKDVVFMGF PLGARGKWYEGNFELLETLGLSKSRQVRVAKTLSTAALLSSVDIVHMFASRARRLHDAG	137
(35) R2MelCri RT	GCRTVLIPKSTQPERLKDINNWRPITIGSILLRLFSRIVTARMSKACPLNPRQRGFIRAAGC SENLKLLQTIIRTAKSEHRPLGVVFVDIAKAFDTVSHQHIIHALQQRGVDPHIIGLVNNMYK DISTYVTTKRNTQTDKIQIRVGVKQGDPLSPLLFNLAMDPLLCKLEERGKGFHRGQRSITAM AFADDLVLLSDSWENMKTNIKILETFCDLTGLKTQGQKCHGFYIKPTKDSYTINNCAAWTIN GTPLNMINPGESEKYLGLQF	138
(35) R2MelCri 5'UTR	CGGCTTTGGAGAGTGTAGTTACAAACTGGGCATCGCTGCAGAGATCGCACCTCCTCGTGGTC CCGCTGGTAGCCCTGCGTAGGGTGACTAAGTCGATCTCTGCCCCAGGTACGGAGCCGTTGGG GCTCACCAGCCCAACGAACTCCTGCCAAATTCGGTGAAACAAATTCCTCGGTAAAAGCCGCG TGGCTTATTGCCTGAAACCTGGCCCCCCGGTTTCAGTCAGGGGCAAAGAGTTCGGAAGTGGA CTGACCACCCACTCCGTTCTCACCGTCGAATGTGGTCCCAATTCG	139
(35) R2MelCri 3'UTR	TGAATTTTTTAACATTTTATTTTTATTTTTACTTTTATTTTTATTTGTTTATCTTTTATGTT TTAATTTTTATTTTTATTTTTTCATTCTTTGTAGTTTTTCAGTATTGTTTTTGTATGTTTTA ATAAATTAGACGGTAGCTAGGTTCGTAAGGCAGCCACAAGCCAGTTAGGTAGCTCATAGTGA GTAGGGACAGAAAACTTTAATTCATAACGCGTCAATTACCATCTGATTCGGACCAAATAACC GGATTTGGACCAATTTACCGGATTTGTCCAAGGTGGACGGGCCACCTTTACTTAACCCGGAA AAGGAGCATATATTATTTATATGTGTTCGAAATA	140
(36) R2MyCa ORF	MGLLMRLMHVSTATQTNINDLKPVTTKLSETITPVTLPSDDIELCQLEFSNSNEIEDWQTTK FVDQREKLALYPEVYESLPSDSVRENNIILSGNDIDADGEGEDSFALATEFVLDSQKDNLPG ETPLVKLPDKTPSCPCCKIKIMGIEKLVDHLLRSHGKRKIIFQCVTCGLQNENYRSIVCHYP LCKGASEVIPSGDCVCEECGRSFQTKRGLGQHKRHAHPEVRNQERIDAARLKVKTSREPARK CWSKEEEELLEKLVQQYEGERCINKLIAEHIPTKTLKQISDKRRCMKPVRKEKEDQEICVQV AKRKNQPAVTETSNPDQLRAHYVETIKKGIKLRTINIYPRAFRKIVEGSKDQLTLVNDTTED WYGHLKARCQAERKIKVGSKKKKAQETQSKRKATNWMKKRAIKKGQYLRYQKLFYKDRGKLA NIILDEIESVKCEIPLSEVFSVFKSKWETTGKFKGLGHFQSKGMADNSAFEVMITEKEVAKN	141

	IKEANKNAAPGPDGIIMKDLIKMDSKYTEITELFNLWLTSGKIPDSLR**RCRTILIPKSTKPD** **RLNDINNWRPITIGSAVLRLFSKILTARLEKACPINLRQRGFIKSPGCSENLKLLELIIRNA** **KKQHRELGVVFVDIAKAFDTVSHQHIIMGLKQKGVDSHIINLIANMYENIYTEIGTKEEKSE** **PIQIKVGVKQGDPMSPILFNLAMDPLLCELEQYGKEYHYDTKSITAMAFADDLVLLSDSWEG** **MKENIEILEVFCDLTGLKTQGEKCHGFFIKPTKDSYTINDCPPWTVGGTPLNMICPGSSEKY** **LGLHI**DPWMQIAKPDLSSKIDTWLDRISESPLKPLQKVDILKIYMIPRIIYLADHTEIKTGF LEALDLKLRARVKKWLHLPNDTCDAIFYSSMKDGGLGIIKLQSLIPSVQARRIHRLAQSSDE TLRDILQQENISQKYEKLWITAGGDKDQIPDIWAPKPTVQTSYLGEEDTSEWEVTTSKNTYP RPCNWRKTEFRKWTKLQSQGIGIVNFEGDTISNSWIKNYRGIPHRKLLTALKLRANVYPTRE YLARGKKDGWTTSCRHCKAENESCAHVIGQCPAVKDARIKRHNYICEMLLNEARRKDWVVYQ EPNIRDSDGELFKPDLVFVKEKQALVVDVTVRYEGDVNTLEKAAKEKVKKYQHLEKEIQELT NTEKVIFVGFPLGARGKWFNKNFDLLQTLGLSSSRRERVARAMASRALLTSVDIMHIFTSKV RTAINK	
(36) R2MyCa RT	RCRTILIPKSTKPDRLNDINNWRPITIGSAVLRLFSKILTARLEKACPINLRQRGFIKSPGC SENLKLLELIIRNAKKQHRELGVVFVDIAKAFDTVSHQHIIMGLKQKGVDSHIINLIANMYE NIYTEIGTKEEKSEPIQIKVGVKQGDPMSPILFNLAMDPLLCELEQYGKEYHYDTKSITAMA FADDLVLLSDSWEGMKENIEILEVFCDLTGLKTQGEKCHGFFIKPTKDSYTINDCPPWTVGG TPLNMICPGSSEKYLGLHI	142
(36) R2MyCa 5'UTR	TGAGAGCATAGTTACAAACTGGGCTCAGCTGCAGAGGTCGCACCTCCACGTGGTCCCGCTGG ATGCCCCTTGTGGGTAACTAAGTCGACCTACTGCCCCAGTTCGGAGGAGGGGTTTTGGCCTC TCTCTCCCTATCGTCTGGAGTCACTGCCCAGCGATAGTATTCAAAAACAGTGATACAAATTT TTTTTTGGTAATAGCTTACGAAGCAAATTACCCAAGCACCTTGCTCACGGGTTCTAGTTGAG CACAGGTTGACAGATATTGAGGGATACTGTTGACCTTATACCCCCTCACTTGAATTGGCAAT TATATATACCGCAAG	143
(36) R2MyCa 3'UTR	GTAATCTTTTATTGTTTTATTGTTATTATAGTATTTTAGTTTTTATAGTTTTAATGTTTAGA TATGTTATGTTTTTATTTCTTTTATTGTGTGGTTATGTTATAGTTTTAAAGTTCTTATTAAA GGCTGTGCGTTTTAACCCCCTTCCCCCCAGTGGAGAATCTCGTGATGATTACACTCTAAGCG CGATCACCGGACATGACGATGGTGGACGGGCCACCGTTACTTAACCCGTAAAAGGGACACAC TCTGTGTGTGTTCGTATAA	144
(37) R2PoAt ORF	MVKALKLLRVSTATQTNFKDWHIKSPQKLNFKPSEPGREVWKSLKLEFYNSLHCSEECTAHC LGREGLEGYPEVYVEETMACLSLNDNGTGESQEKITTPITAQEEVTALSEPEETTIQPQNNG ESEDCPVISVPDKNPPCPRCGMIMNKMHGLIEHLLRAHGIRECKFKCTKCDLENTNFHRIAC HYPKCKGKAKEPPPGKWPCDVCQMSFQTKIGLGQHKRIAHPAVRNQERIEATKPKETSARGA HKKCWSEEEENLLIQLIKQYEGNENINKLIAEHIPSKTAKQINDKRRCLVSKQPAPFGLQLR GKTQEQPQEKTRKPPKRKETTINQLKKHYEKMIPSKISDGTFIVYKEAFQEILKKKNKNPAH TISEITKDCIKVFKGMHSQEKNTRKPKQGHSQPKRRQAAKWMSKRAIKKGQYLRFQKLFYTD RGKLARIVLDDVEGLKCQIPANEISSVFRERWETPGNFKGFEDFEAEGTTDNTKFLPLITAK EIDKNIKEMAKTSAPGPDDITLGDLCKADPKFSLLTEIFNLWLTSGTVPDQVK**NCRTVLLPK** **SDRPDRLKDINNWRPITIGSTVLRLFSRVLTARLSKACPINPRQKGFTRAPGCSENLQLLQL** **LLKTAKREHNEIAVVFVDIAKAFDTVSHQHIIESLRQRKVDEHIINLIQSMYVNTSTCIKTQ** **DETSAPINILKGVKQGDPMSPLLFNLALDPLLCKLEKEGQGYKCEDAKITAMAFADDLVLLS** **DSWEGMMHNIKILETFCDITGLKTQGEKCHGFYIKPTKESYTINDCPPWTVNNTPLNMIPPG** **ESEKYLGLQV**DPWVGIAQAELLAKTNTWLERIGRTKLKPLQKVDLLKMYTIPRLYYIADHTD TPFVLLKNIDQAVRTKVKEWLHLPACTCDALLYSSTRDGGLGISRLEGLIPSIQARRLHRLA QSKDEILTNLLKGEGADLRHKKLWVKAGGDKTKIPSIWTTRPTVAVQAPTSENLSEWEVPTA KLKYPKPCNLRKKEFEKWKQLVSQGRGITHFENDSISNAWLTSYRAIHHRKLITALQLRANV YPTREFLARGRGGQYIKSCRHCNAENETCSHIIGKCPVTQDARIQRHNKICDLLTNIAQRKD WTVFHEPHIRDSNGQLFKPDLIFVKDNQAQVVDVTVRYESEESTLENAAEEKVSKYQHLEKE IQELTNAQQVTFIGFPLGSRGKWYPGNSMLLETLGLSKREQDRTSKTLASRALESSVDIIHI FASKTRTPG	145
(37) R2PoAt RT	NCRTVLLPKSDRPDRLKDINNWRPITIGSTVLRLFSRVLTARLSKACPINPRQKGFTRAPGC SENLQLLQLLLKTAKREHNEIAVVFVDIAKAFDTVSHQHIIESLRQRKVDEHIINLIQSMYV NTSTCIKTQDETSAPINILKGVKQGDPMSPLLFNLALDPLLCKLEKEGQGYKCEDAKITAMA FADDLVLLSDSWEGMMHNIKILETFCDITGLKTQGEKCHGFYIKPTKESYTINDCPPWTVNN TPLNMIPPGESEKYLGLQV	146
(37) R2PoAt 5'UTR	CAGAGGTAAGTCAGAGTTACTAACTGGGCTTTGCCGCGGAAATCAGACCTCCTCGTGGTCCC GCTGGATACCCGCAAGGGTAACTAACTGATTTCTGCCCCAGTGTGAGCGAGGCTGTGTTTTA CACTGCTTCCCTGCCTTTGGACTCGCCACCCAGAGTGCAGTTTCGCGATTTTGCGTTGAAAA GTTATCGGCAGAAGCTCTTAAAAAGCTCCCTGCCTGAAACCGTGCCCACCGGTAAAGTTGGG CAAAGCTAACAGCGCTTGAGGGGGTACTGTTAGTTAAATCACCCTCAATTGGACTGACTTTA AACCTAAAACGGCACC	147
(37) R2PoAt	ATAGTAAAACTCTGTATTAGTTAGATTACAGTTATTTTTTTTAGGTTAAGCATATTTTAGTT AGTTTCTTTCTGTATTAGTTAGTATTGTTGCTGATTAGTTCTTATATGTTTCTTTATATTTT	148

3'UTR	TGAATGCTTTTATCACTGTATATACGTTATATTTTTCAATAAAGGTGTGGGAGTAGCTGGAT AGATAGGGAACCCCCACCCCATCCCCCGCCCCCCATTTTCTGTTCCATCCTGAGTGATGGAT GTATCATCCGACATGCGACGTTACACCTAATTTGGCACTAAAATCTCACGTTTGTGACATAG GTGGTCGGACCACCTTTATAAAAACCTTGAATTTGACCTATACCTACCGATTTTGACACAGG TGGACGGGCCACCTATACTTAACCCGGAAAAGGTGCATATACGTTAGTATGTGTTCGTTTAA	
(38) R2ChPi ORF	MMEALASAQLGPGRRANSTHLDTDVVPWGRCTFRTFEPLARYRVKVSTVATQTDSHQLTDNE AYPTLKSSKTTVTEGEMYPIPGTSQTHTAAGAISCETVAQTEAEEILPALPLEILNSAFGET VGPDTSPEPCVPTTPVSEKERDAGGSETLADGTDGTICVPSSAPTVVVTEGECENRVQGKVL VTLPHDNPACPFCGDHVGKPKALSVHMKHHHGGKDVEFQCSLCTRADPNVHSILCHIPKCKG LTPQEPAGDWACEVCARRFGTKSGLSQHKRRAHPVVRNAERIQASQPKQNSQRGKHKSCWSA EEELLLTDFDTMYSGEKYINKLIAEAIKTKTAKQVSDKRRLLNLYRKKSSGTADPGPGAQKH NLEVRSETPISTTGLKNSYMRTITNRLRASNRGPLRQDLEKFEAEVLSAWVAGDPNIRSLVE TTSIDVLSSFLAPEHEHKHKQKGNTKGRKSKSGKGQSWRQKRATKRGTYIRFQHLFENDRCK LASIILDGTERLQCEIPLEKVFQTYKDKWESRTPFQGLGQFKSHALADNSAFDILISAKEVL KNISEMNKNSAPGPDKVSLKDLVRVDPEGDTLAGLFNTWLVTGIIPDGIK**ECRSLLIPKTMD** **PEALQDLGNWRPLTIGSIVLRLFSRILTNRLAKACPINARQRGFIAAPGCSENLKVLHTITK** **QAKKEKKPLGVVFIDIAKAFDSVSHDHIMWVLQERGLDQHIINIIGDSYRNVHTRMEVGKEL** **TSPIAIDVGVKQGDPMSPLLFNLALDPLISTLETAGKGFSFGAQKVTALAFADDLVMLSDSW** **EGMHTNIQILESFCHLSGLKVQAKKCHGFLLSPTHDSFTVNNCEAWKIGADGLNMIAPGESE** **KYLGLKV**DPWIGFSKPALSDKLSTWLERINRAPLKPSQKLEMLNTFTIPRVIYLADHTDCKK TLLSSLDDNIRRAVKEWLHLPADTCNSFLYTRSRDGGLGVTRLASLIPSIQARRLHRIAHSE DMTIRFIALANDIESEYQSLWLAAGGTQDTMPAITDPVTIDHRLPQHALELLNEWERPAPRV IYPIPCNWRDAEFSQWKDLPCQGSGVSHFGNDPTSNSWIKHLTGFSQRQFLTALKLRANVYP TREYLGRGKNEYMVKCRHCDAPYESLSHILGQCPAVQGARIRRHNKICEILTREARKLEWTV YQEPHLRTSTNELRKPDLVFVKDGTALVVDVTVRFEYKEQTFSDAAAEKVRHYQSLTSEISR LTGASKVRYFGFPLGARGKWPKINDRVLSALGLSESAQGRAARLLSRRALLYSIDVLNTFSG IGKGLPNSADTKSVGTQREQLGIS	149
(38) R2ChPi RT	ECRSLLIPKTMDPEALQDLGNWRPLTIGSIVLRLFSRILTNRLAKACPINARQRGFIAAPGC SENLKVLHTITKQAKKEKKPLGVVFIDIAKAFDSVSHDHIMWVLQERGLDQHIINIIGDSYR NVHTRMEVGKELTSPIAIDVGVKQGDPMSPLLFNLALDPLISTLETAGKGFSFGAQKVTALA FADDLVMLSDSWEGMHTNIQILESFCHLSGLKVQAKKCHGFLLSPTHDSFTVNNCEAWKIGA DGLNMIAPGESEKYLGLKV	150
(38) R2ChPi 5'UTR	CTGTGTCAAGAGTTTGGTTGCTGGGCCGCAGTATTGAGATCTCAACCTCCCCGTGGTCCCGC TGGACGACTCTCTGGGATGCCCGTTAGGGAGTTGACTGACTGAGACTCAGCTGCTTTCGTTC CCCTGCCCTACTGCTCTGGTACACCGGCTAGACAGTAGAACGCTACGGTGAAAAACACTA	151
(38) R2ChPi 3'UTR	ATGATTAAAATCTCAATAATTCATGACCGCGATCACTGGATGGGACTTTCTGGGTTGGATGA GCCACCCGGGGTTTGGAACACCCCAAATAACTCAAAACAAAAACCCGCGTGGGTTTGTAAAA A	152
(39) R2ATLH ORF	METTLVSIRAKIQATPLLWANHPLEGFLLVMKRKEMTSPGPSGPQGNSRDNDAVGVPSIPAC PPGDFTCPCGRTFPSKRSLAMHRHHSHPAEVNEERTTRLPLKRLAWSQEDDSALLRHATRLW APSLLKKKHLELLAPLFPGRSVEALKKRLQILKWLPPPPPETVSSPGTTPTMVSPPAFIRSP ETPTIVFPDTPSMVFPDTPTMVFPDTPTVVFPDTPTLDVQETQPPTSSWKVDLLDFAARNLR EPKMGADDLLNITSKLQSGTVSQSEAAHLLNEHAKRWFPHSWAPSKRTHRSLGPAPSNRQLR RANYAAVQKLYHLRKKDAASCVLSGQWKSAHRGSKDTIDGLEDYWSQTFEHPSATDPRDPRT PEDTAWTLLDTITAAEVTSALRDMGNSAAGLDRLTPATLSSWDQPSLAAYFNLILATESVPK HIS**KSRVIFLPKTDEPTAPSDFRPISISSVILRAFHKILARRITSTLQFSKLQHAFLKRDGC** **LEASSLLHAILRKIHEDRTPASLIFLDISKAFDSVSIHTLLRVAKSSGLPPPLTNYLVSLYD** **HPRVKLEAREVTCGRGVRQGDPLSPLLFIMAFEDILQSALPDIGFDLDGHNIPALAYADDLV** **LFARRPERLQEKLDILVDSLAAAGMSINAKKSIGLTIAKDGKRKTMVLLPHIYRSGSTEIKP** **LGPSDVTRYLGLDF**TWKGKVSPKHTSKLDQMLLELTRAPLKPYQRLDLLKNFAIPRLTHELV LGNAHRNTLKKLDVLIRSAVRSWLRLPKDTPLGFLHAHIRDGGIGLPCLNTSIPILQRNRFA KLLSNPEPIQKVVANQKSFRVIQRITNRPVQIGNHIIQSKKEARTAWKNELRRSVDGRDLFG DPVDPVSHRWVSHPQRVFPRLHLRGIQLRGGLLPTRTRNARGRDPPVTSTSCRGLCGQAETL NHILQVCDKTHDARCARHNRIVHRLSQLLRRKNVPCWIEPIIPAGATYIKPDIIAKVNNELL VMDVSVVSGSRLNETWSIKTEKYGSSKSSTALQSWHGSRIPLRHYPIIISNRGLLFAQSGQA LRSIGLSARDLTDICLLTIIGSLKCFDHYTRSTLDL	153
(39) R2ATLH RT	**KSRVIFLPKTDEPTAPSDFRPISISSVILRAFHKILARRITSTLQFSKLQHAFLKRDGCLEA** **SSLLHAILRKIHEDRTPASLIFLDISKAFDSVSIHTLLRVAKSSGLPPPLTNYLVSLYDHPR** **VKLEAREVTCGRGVRQGDPLSPLLFIMAFEDILQSALPDIGFDLDGHNIPALAYADDLVLFA** **RRPERLQEKLDILVDSLAAAGMSINAKKSIGLTIAKDGKRKTMVLLPHIYRSGSTEIKPLGP** **SDVTRYLGLDF**	154
(39) R2ATLH	GAGTTGTAGTTACACGGCCTGGGCGCCTCCTCGTGGCGACCGCTGGAGAAG	155

5'UTR		
(39) R2ATLH 3'UTR	AGCCGGTGTCGCATGTAGGTTACTGGGAACAGCTCATGACCCGCCACAAGCGGCGATCCCCC TGCAACGTCGTTATCATTAGCGGTTATCCGCCTCAAGGCTACCGTTTGCGGTAGTGTAATTG TTACATTGCTGCTGTCGGTTGACCTTGATAATCCATGACACC	156
(40) R2ATLH' ORF	MKRKEMTSPGPSGPQGNSRDNDAVGVPSIPACPPGDFTCPCGRTFPSKRSLAMHRHHSHPAE VNEERTTRLPLKRLAWSQEDDSALLRHATRLWAPSLLKKKHLELLAPLFPGRSVEALKKRLQ ILKWLPPPPPETVSSPGTTPTMVSPPAFIRSPETPTIVFPDTPSMVFPDTPTMVFPDTPTVV FPDTPTLDVQETQPPTSSWKVDLLDFAARNLREPKMGADDLLNITSKLQSGTVSQSEAAHLL NEHAKRWFPHSWAPSKRTHRSLGPAPSNRQLRRANYAAVQKLYHLRKKDAASCVLSGQWKSA HRGSKDTIDGLEDYWSQTFEHPSATDPRDPRTPEDTAWTLLDTITAAEVTSALRDMGNSAAG LDRLTPATLSSWDQPSLAAYFNLILATESVPKHIS**KSRVIFLPKTDEPTAPSDFRPISISSV ILRAFHKILARRITSTLQFSKLQHAFLKRDGCLEASSLLHAILRKIHEDRTPASLIFLDISK AFDSVSIHTLLRVAKSSGLPPPLTNYLVSLYDHPRVKLEAREVTCGRGVRQGDPLSPLLFIM AFEDILQSALPDIGFDLDGHNIPALAYADDLVLFARRPERLQEKLDILVDSLAAAGMSINAK KSIGLTIAKDGKRKTMVLLPHIYRSGSTEIKPLGPSDVTRYLGLDF**TWKGKVSPKHTSKLDQ MLLELTRAPLKPYQRLDLLKNFAIPRLTHELVLGNAHRNTLKKLDVLIRSAVRSWLRLPKDT PLGFLHAHIRDGGIGLPCLNTSIPILQRNRFAKLLSNPEPIQKVVANQKSFRVIQRITNRPV QIGNHIIQSKKEARTAWKNELRRSVDGRDLFGDPVDPVSHRWVSHPQRVFPRLHLRGIQLRG GLLPTRTRNARGRDPPVTSTSCRGLCGQAETLNHILQVCDKTHDARCARHNRIVHRLSQLLR RKNVPCWIEPIIPAGATYIKPDIIAKVNNELLVMDVSVVSGSRLNETWSIKTEKYGSSKSST ALQSWHGSRIPLRHYPIIISNRGLLFAQSGQALRSIGLSARDLTDICLLTIIGSLKCFDHYT RSTLDL	157
(40) R2ATLH' RT	KSRVIFLPKTDEPTAPSDFRPISISSVILRAFHKILARRITSTLQFSKLQHAFLKRDGCLEA SSLLHAILRKIHEDRTPASLIFLDISKAFDSVSIHTLLRVAKSSGLPPPLTNYLVSLYDHPR VKLEAREVTCGRGVRQGDPLSPLLFIMAFEDILQSALPDIGFDLDGHNIPALAYADDLVLFA RRPERLQEKLDILVDSLAAAGMSINAKKSIGLTIAKDGKRKTMVLLPHIYRSGSTEIKPLGP SDVTRYLGLDF	158
(40) R2ATLH' 5'UTR	GAGTTGTAGTTACACGGCCTGGGCGCCTCCTCGTGGCGACCGCTGGAGAAGATGGAGACGAC TCTCGTCTCCATCCGAGCTAAGATCCAGGCCACCCCCCTCCTCTGGGCTAACCACCCTCTGG AGGGTTTTCTGCTGGTT	159
(40) R2ATLH' 3'UTR	AGCCGGTGTCGCATGTAGGTTACTGGGAACAGCTCATGACCCGCCACAAGCGGCGATCCCCC TGCAACGTCGTTATCATTAGCGGTTATCCGCCTCAAGGCTACCGTTTGCGGTAGTGTAATTG TTACATTGCTGCTGTCGGTTGACCTTGATAATCCATGACACC	160
(41) R2ATLH" ORF	MTSPGPSGPQGNSRDNDAVGVPSIPACPPGDFTCPCGRTFPSKRSLAMHRHHSHPAEVNEER TTRLPLKRLAWSQEDDSALLRHATRLWAPSLLKKKHLELLAPLFPGRSVEALKKRLQILKWL PPPPPETVSSPGTTPTMVSPPAFIRSPETPTIVFPDTPSMVFPDTPTMVFPDTPTVVFPDTP TLDVQETQPPTSSWKVDLLDFAARNLREPKMGADDLLNITSKLQSGTVSQSEAAHLLNEHAK RWFPHSWAPSKRTHRSLGPAPSNRQLRRANYAAVQKLYHLRKKDAASCVLSGQWKSAHRGSK DTIDGLEDYWSQTFEHPSATDPRDPRTPEDTAWTLLDTITAAEVTSALRDMGNSAAGLDRLT PATLSSWDQPSLAAYFNLILATESVPKHIS**KSRVIFLPKTDEPTAPSDFRPISISSVILRAF HKILARRITSTLQFSKLQHAFLKRDGCLEASSLLHAILRKIHEDRTPASLIFLDISKAFDSV SIHTLLRVAKSSGLPPPLTNYLVSLYDHPRVKLEAREVTCGRGVRQGDPLSPLLFIMAFEDI LQSALPDIGFDLDGHNIPALAYADDLVLFARRPERLQEKLDILVDSLAAAGMSINAKKSIGL TIAKDGKRKTMVLLPHIYRSGSTEIKPLGPSDVTRYLGLDF**TWKGKVSPKHTSKLDQMLLEL TRAPLKPYQRLDLLKNFAIPRLTHELVLGNAHRNTLKKLDVLIRSAVRSWLRLPKDTPLGFL HAHIRDGGIGLPCLNTSIPILQRNRFAKLLSNPEPIQKVVANQKSFRVIQRITNRPVQIGNH IIQSKKEARTAWKNELRRSVDGRDLFGDPVDPVSHRWVSHPQRVFPRLHLRGIQLRGGLLPT RTRNARGRDPPVTSTSCRGLCGQAETLNHILQVCDKTHDARCARHNRIVHRLSQLLRRKNVP CWIEPIIPAGATYIKPDIIAKVNNELLVMDVSVVSGSRLNETWSIKTEKYGSSKSSTALQSW HGSRIPLRHYPIIISNRGLLFAQSGQALRSIGLSARDLTDICLLTIIGSLKCFDHYTRSTLD L	161
(41) R2ATLH" RT	KSRVIFLPKTDEPTAPSDFRPISISSVILRAFHKILARRITSTLQFSKLQHAFLKRDGCLEA SSLLHAILRKIHEDRTPASLIFLDISKAFDSVSIHTLLRVAKSSGLPPPLTNYLVSLYDHPR VKLEAREVTCGRGVRQGDPLSPLLFIMAFEDILQSALPDIGFDLDGHNIPALAYADDLVLFA RRPERLQEKLDILVDSLAAAGMSINAKKSIGLTIAKDGKRKTMVLLPHIYRSGSTEIKPLGP SDVTRYLGLDF	162
(41) R2ATLH" 5'UTR	GAGTTGTAGTTACACGGCCTGGGCGCCTCCTCGTGGCGACCGCTGGAGAAGATGGAGACGAC TCTCGTCTCCATCCGAGCTAAGATCCAGGCCACCCCCCTCCTCTGGGCTAACCACCCTCTGG AGGGTTTTCTGCTGGTTATGAAAAGAAAAGAA	163
(41) R2ATLH" 3'UTR	AGCCGGTGTCGCATGTAGGTTACTGGGAACAGCTCATGACCCGCCACAAGCGGCGATCCCCC TGCAACGTCGTTATCATTAGCGGTTATCCGCCTCAAGGCTACCGTTTGCGGTAGTGTAATTG TTACATTGCTGCTGTCGGTTGACCTTGATAATCCATGACACC	164

(42) R2Mac ORF	MDLLFYFPGIGKMVRSMRLLRLSAATQTSPSDWEDIPTRNLFNKGLVGEEIELKRLEFYNSC MINNLLGYHPPIRWEGLARYPEVYEVPFPEPDSGDNTSIREAVEENKKNRDDSPTLTSAPIR EKDSGQTPKTPVVKVPDKNPPCPCCGNRISGLLGLVDHLKRDHGRRKIRFQCAKCGRENLNC HSIACHFPHCKGPADRAPAGEWTCEVCGRNFQSKIGLGQHKRLTHPEERNRERIAASCPKTK AKPKKTGAHRKCWTSEEEELLIKLMEQYKGKKHINKLIAEHIPSKTTKQISDKRRLLEPGKK NGEDQDLDLLIEDEWEQLAVADSSREGYLQAHYLRKVGDRLSAQTLHTFQGAFKKIIDGQQD PLSVVNETSRDCYECLGVSCQLGKTSQNKQKQQNPSQNPRPGKKAKNWMRKRAIRKGQYLRF QRLFHLDKGKIARIILDDLESMRCEIPLQDIHSVFQSRWESPGTFKSLGGFQSKGLADNSAF EAMITAQEVSKNISEMSKNSAPGPDGITLRDLNKMDPKYSLTAEIFNLWLTSGMIPDTVR**GC** **RTVLIPKSSRPERLKDINNWRPITIGSIILRLFSRILTARLSKACPINPRQRGFIRAAGCSE** **NLKLLQLLIQNAKREHREMGVVFVDIAKAFDTVSHQHILMGLKQKGVDPHITNLIKNMYENI** **YTQIDTKKGTTDQIKMRVGVKQGDPMSPLLFNLAIDPLLCKLEECGNGYQQGKHTITAMAFA** **DDLVLLSGSWEGMQNNIKILETFCELTGLRTQGEKCHGFYIKPTAKSTYSINDCPAWTVNGT** **SLNMITPGSSEKYLGLRV**DPWVGIAKPGLSQQVEHWLQRLEGAPLKPLQKIDLLKTFAIPKI IYLADHTDTKAGLLESLDLQIRRSVKEWLHLPPCTCDAFLYSSTKDGGLGITRLAGLIPSVQ ARRLHRLAQSSDETIRAVLKEEQIEKDYEKLWVKAGGDKDRMPSIWEPRLVVEPLEHREEGA PSEWEVSAPKIKYPRPCNWRKREFNEWSKLASQGRGIDNFEGDKISNFWIERYRGIPHRKLL TALQLRANVYPSKEFLARGRRDKYSTSCRHCPAENESCSHIIGNCPITQEARIKRHNYICGM LSEQAKKADWAVFQEPNIRDDNNELYKPDLIFVKDDKAYVVDVTVRYEHSKKNLEIAAKEKA KKYQHLEKQIQDLTNATKVSFFGFPVGARGKWYDKNFELLSALGLSKSRQEKTARALARRAL FASVDIVHMFASKARTVLVS	165
(42) R2Mac RT	GCRTVLIPKSSRPERLKDINNWRPITIGSIILRLFSRILTARLSKACPINPRQRGFIRAAGC SENLKLLQLLIQNAKREHREMGVVFVDIAKAFDTVSHQHILMGLKQKGVDPHITNLIKNMYE NIYTQIDTKKGTTDQIKMRVGVKQGDPMSPLLFNLAIDPLLCKLEECGNGYQQGKHTITAMA FADDLVLLSGSWEGMQNNIKILETFCELTGLRTQGEKCHGFYIKPTAKSTYSINDCPAWTVN GTSLNMITPGSSEKYLGLRV	166
(42) R2Mac 5'UTR	AATGGGAGCATAGTTACTAACTGGGCTTAGCTGCAGAGGTCGCACCTCCTCGTGGTCCCGCT GGATACCCCTAGGGGTAACCGAGTCGACCTCTGCCCCAGTTCGGAGAATTGGCTAGAGCCTC TCTCTCCCTATCGCTTGGACTCGCCAACCAGCGATAGTATTGCTTTGGGTGAAACAAACTTA TGGTAACAGCCCTTACTAGGTCTGTTACCCAGAGCCTTGCCCACTGGCTTTAGCTGGGCACA GTTTGACAGATCCTGAGGGGGTACTGTCGACTTATACAACCCTCAC	167
(42) R2Mac 3'UTR	TTCAATAAAGTTGCGACCGCTAGGGACGTAGGATCGCCACAAGCCAGGTAGAAAATTCAGGT AGGATAGTTTGGAGGGACTAGGATCTGGGACTTAACTCATGACGTGTCTTTTACCATCGACT TTGGGACAAATCTTTACCGATTTTGACCGAGGTGGACGGGCCACCTTTACTTAACCCGGAAA AGGGACATATATTAATTTATATGTGTTCGAAAAA	168
(43) R2Pif ORF	MGWLLRLMWVSTGTQTDSRDWKPVVPKLNEVSSSLVTTSDEIELQHLEFASSKEIGNWQVTQ ILSQEEKLALFPDVYESSTSDSVRDNENNISDNNNEINGDEDDSFMLATEPALDSQKNRSPG KIPLVRLPDKNPSCPCCKTKIAGVERLIDHLLRSHGIRKTVFQCIKCDLQSDNYRSIVCHYP KCKGAIDMGPSGDWICEVCGRSFQTNRGLGQHKRLSHPEVRNQERIAATRPKERTARESSRK CWSKEEEELLEKLVKQYEGEKYINKLIAEHIPSKTIKQISDKRRHMFPVQKEREDRDGYVQV DKRKRHLEIETPSPDQLRVHYLAMITKGLKSRKMNNYSRVFQKIVEGKPEVHAIINDITIEW YGHLEARCRRRETSDVKMERGQIQGPQSKRKAANWMKKRAIKKGQYLRFQKLFYKDRGKLAR IILDDIESVKCEIPLPEVHSVFKTRWETTGRFKNLGNFRSIGMADNNVFKAMITEKEIIKNI NEMNKNTAPGPDGITVRELIKMDPKFTQITEMFNLWLTSGRIPDIIR**ECRTVLLPKSTKPDR** **LKDINNWRPITIGSVILRLFSRILTARLIKACPINPRQRGFIKSPGCSENLKLLQLLMKNAK** **REHNELGVVFVDIAKAFDTINHQHIITGLKQKGVDAHIINLIANMYENIYTRIGINNKQLDP** **IQINVGVKQGDPMSPLLFNLAMDPLLCELEQYGEGFKFRNKSITAMAFADDLVLLSGSWEGM** **KRNIEILETFCDLTGLKTQGEKCHGFLIKPTREAYTINDCPPWTIDGTPINMVNPGSSEKYL** **GVQI**DPWIQIAKSDLSSRIETWLCRVSEAPLKPFQKVDMLKTYVIPRIIYLADHTEIKSGSL EALDLKIRMTVKKWLHLPASTCDAILYSSIKDGGLGIIKLQALIPSVQARRLHRLAQGSDDI LRAVLVEENITQMYEKLWISAGGDKEKVPNIWAPKPTVQTATLGEEEISEGENNILKNNYPR PCNWKKKEFQKWTQLQSQGIGITNFEGDTISNNWIKNYRGIPHRKLLTALQLRANVYPTREY LARGKKDGYIRSCRHCQAENESCAHIIGQCPIVKDARIKRHNYICGMLAEEAKMKEWMVYQE PNIRDTDGELFKPDLIFVKDKKALVVDVTIRYEADDTTLDKAAKEKVRKYRHLEKEIKELTN TEKVTFIGFPLGARGKWYNKNFDLLQTLGLSASKRERTARALASRALFTSVDIVHIFASKSS TQLCK	169
(43) R2Pif RT	ECRTVLLPKSTKPDRLKDINNWRPITIGSVILRLFSRILTARLIKACPINPRQRGFIKSPGC SENLKLLQLLMKNAKREHNELGVVFVDIAKAFDTINHQHIITGLKQKGVDAHIINLIANMYE NIYTRIGINNKQLDPIQINVGVKQGDPMSPLLFNLAMDPLLCELEQYGEGFKFRNKSITAMA FADDLVLLSGSWEGMKRNIEILETFCDLTGLKTQGEKCHGFLIKPTREAYTINDCPPWTIDG TPINMVNPGSSEKYLGVQI	170
(43) R2Pif 5'UTR	CGCGACAGCATAGTTACTAACTGGGCTAAGCTGCAGAGATCGCACCTCCCCGTGGTCCCGCT GGATGCCCGTAAAGGGTAACTAAGTCGATCTCTGCCCCAGTCCGGAGGAGAAGGCCATAACC	171

	TCTCTCTCCCTATCGTCTGGACTCACTACCCAGCGATAGTACTCAAAAATAGTGAAACAAAT TTTTGTGGTATTAACTTACCGAGGTTAATTGCCCGAAACCTTGCCCACGGGCTAAAGTTGGG CTCTAGGTTGACAGGTATTGAGGGGTACTGTTAACCTTGACTTCCCTCACTTGTTGGATCTA TTGGTATTGACAA	
(43) R2Pif 3'UTR	TCTCTTCATGTTGTTACTAAGTTTTAGATTTTCTTTGATTTTATCATTTTTTATTGTCTTTT ATTGTCTTTTTAGCCTCTTTTAGTTTTTTAGTAATTATCTGTATTAGGTTTCATCTTACTTT TAATTGTATTAGTTTTATTTGTATTTTATTAGTTGCTATTAAAGCTGTGTTCTTTTATTGTG ACCCTTTCCCCAAAGTGGTGAATCTCGTAATGATTACATCTAGCGCTCACTGGACATGACGA TGGTGGACGGGCCACCTCTAACCCGGAAAAGGTACATACACTCTGTGTGTGTTCGTTTAA	172
(44) R2TyAl ORF	MCLAARIAFVSVGTQTSLLDLRDDVVPRTIRDHAKLVDVGINFRRLEFVNSDVEGLEGGDLS TVTESLAMLPEVYEEPSPSALEGKDISPLLRESESEGEGELSPPGAVAQLFPLPSSPILASG GEPSPEGTPVVKVPDKNTSCPCCGTRIGKVSGLVEHLRRVHGRRKILFQCALCGKTNEKHHS IACHVPKCKGAIEATPKEGLACEECGRTFETKIGLGQHKRLAHPVTRNLERIMAARPKETSA RGVHRKCWTEEEVEILRKLEKQYEGNRNINKLIAEELPFKTAKQISDKRRILHKSPVKKREE DQEWKKHLREGEHLVSIGPRKRGGGLKSHYRKTISKWITAGKLRILHATFEKVIEGGEDLPT TVGGAVEDCYARLFKRSSELEASQGRKLQTARTARSSEPPKEWMKKRARKRGEFLRFQRLFH LDRGKLAGIILDDVESLKCGLPLDEVHCVFKSRWESPGTFKGLGGFKTLGRADNSAFEAMIT SKEISKNIKEMSKNSAPGPDGLSLRGLMAIDPHYTMLAELYNLWLLTGTIPDMVK**ECRTVLI** **SKSALPERQKDINNWRPITIGSTVLRLFSRIITARLTKACPISPRQRGFIRSAGCAENLKLL** **QLLIRNAKREKHPLGVVFVDLAKAFDTVSHNHIIEVLKQKGVDNHIISLISNMYSNTHTYID** **MKSEQSDPIGIRIGVKQGDPMSPILFNLSVDPLLCVLEKEGSGYQHYSTSITSMAFADDLVL** **LSGSWEGMRRNIEIVEAFCDLTGLKTQGEKCHGFYIQPTKDSYTINDCPAWTIKGTPLNMIE** **PGSSEKYLGLKI**DPWVGISKPELLAKIKDWLHRIDRAPLKPMQKVDILKVYTIPRMIYLADQ SDVKVSYLESLDLTIRTAVKEWVHLPPSTCDAILYSSTRDGGLGLTRLSGLIPSVQARRLHR IAQSSDEVIREIVSKEGIEKEYEKLWLAAGGNKEKIPSIWDPSSAMVITEVLGGIDQYSEWE RPAPKTIFPKPSDWRKQEFIKWTKLASQGRGITNFEKDKISNSWLERYRGIPQRKIVTALQL RANVYPTREFLARGRQETHVRSCRHCGAEIESCSHIIGYCPAVQDARIKRHNQICELLIEEA KKKEWIAFQEPLIRDEKKELFKPDLVFVREEIAIVVDVTIRYESKETSLEDAAKEKIKKYQH LEDQIQELTNATTIKFMGFPLGARGKWYQGNYTLLKELGLSSSRQDKVARRLSNRALFSSVD IIHMFSSKVRTVENV	173
(44) R2TyAl RT	ECRTVLISKSALPERQKDINNWRPITIGSTVLRLFSRIITARLTKACPISPRQRGFIRSAGC AENLKLLQLLIRNAKREKHPLGVVFVDLAKAFDTVSHNHIIEVLKQKGVDNHIISLISNMYS NTHTYIDMKSEQSDPIGIRIGVKQGDPMSPILFNLSVDPLLCVLEKEGSGYQHYSTSITSMA FADDLVLLSGSWEGMRRNIEIVEAFCDLTGLKTQGEKCHGFYIQPTKDSYTINDCPAWTIKG TPLNMIEPGSSEKYLGLKI	174
(44) R2TyAl 5'UTR	ACGGGATCATAGTTACTAACTGGGCTCAGCTGCAGAGGTCACGCCTCCTCGTGGTCCCGCCG GATACCCTCTTGGGTAACTGAAGTGGCCTCTGCCCCAGTGTGATGAGATATGGTTACTTCCT CCCTATCAAGTGACTTCACCGACCACTTGATAGTATCTAATTGGTGATACAAAATTTCGGAG TCTTAATTGACTACCCAGAGCCTGGCCCACGGGCCTATGTTGGGCGCAGTCTGACAGCTCGT AAGGGAGTGCTGTCTGGCTTAAATTTGACCCTTACCTGGATTTGACAAA	175
(44) R2TyAl 3'UTR	TTGTATCATCACATTGTAGTTGTAATATCTGTGCACTGTATTACTTTTAATTAATAAAATAA AGTGGTATGAAGAGCCGACATGGAAGGATCGTGGGGGGGTTAGATCAGGAGAGAAAACGATC AGAGTATTGGGATTTCAACATAGTAGGTGGGAATCTTGTTCTTTTTATGACGCGTCTCAACC GAAATTGACATTAACAATTGGACTTGACAAAATTTTTCACCAATTTGGACCCAGGTGGACGG GCCACCTTTACTTAACTCGGAAAAGAAACATATATAATTTATATGTGTTTGATAAA	176
(45) R2ViLa ORF	MFPCIDDPRPPGPPDRSSPGRRSRVACRESDAKDPGHVCPCGSEFQSKQGLAMHRYHRHMAE ENAKRLASLPSKRLAWSQEEDRALLRLATAQWTGGMAKVTLFGLLQPQFPGRSVEAIKKRLQ GLKWTAPADNPERDPLPVPPSTAQGDVEPATSPAPTADPRVFDQNTDDPTVSWRRALLTAVE GKLSEPKMGADTLCRIASQLKNGEINGEACRALMNEHAREWFPHTWKPSSVKDQTDTRRWSK RAIRRANYATVQRLFKLRRRDAAHTVLSGRWRNAYRGQAPLPKGMDQYWSKVFEAGSAKDEK AQVVPDKTLWETLAAVTPSEVTSALKGMKASAPGLDRISAEDLLAWNQPSLAAYFNLMLAAG GPPEQLAC**SRVTFIPKVDEPKTPGDYRPISIASVILRAFHKILAWRLRDTVQLSPFQHGFLQ** **RDGCLEASALLHTILRKSHRDRKPCALLFLDVAKAFDTVSHQTLFRVATAMGLPPPLVDYLK** **HLYGRSTVQIAGQNTKCGRGVRQGDPLSPLLFIMVMEDIVRDALPEVGFDLDGQRVDSLAYA** **DDLVLLAEKSSRLQEKLTLLTAALQNAGMELNARKSHGLTIEKASKRKCMVLRDTKYQCKGE** **GIQPMGTDDSVRYLGLHF**NWKGRLVPKHTGKLDSLLKELSAAPLKPFQRLQLLKLHVVPKFV HELVLGHAHRNTLQKLDCLIRAAVRQWLRLPTDTPLGYLHANIKDGGLGIPCFSTSIPLLQQ KRFEKIAQNPATIFQIIQRQDSFRTLGRRLDRPCRLQGAVLTSKAEAREEWGNVLSNSVDGR ELRYPDVDKASHEWLLRPSRVFPRLHMRGIQLRGGLLPTKARSARGNNRQVQDKTCRGACRQ PETLNHILQVCEVTHDGRCARHNRVLHRLAQLLRKKGVRTTLEPVIPSGSSFIKPDLIAETD KQTYVLDVSVVAGGRMKTSWDIKDEKYGRPDRVNDIRKWGAIPEASRVCQLPIIISNRGLCY QATGEGLRALGLTTREISDLCLLAISGSLKCYDLYMRGTRGVRGAPEHRYGAPKGPH	177

(45) R2ViLa RT	CSRVTFIPKVDEPKTPGDYRPISIASVILRAFHKILAWRLRDTVQLSPFQHGFLQRDGCLEA SALLHTILRKSHRDRKPCALLFLDVAKAFDTVSHQTLFRVATAMGLPPPLVDYLKHLYGRST VQIAGQNTKCGRGVRQGDPLSPLLFIMVMEDIVRDALPEVGFDLDGQRVDSLAYADDLVLLA EKSSRLQEKLTLLTAALQNAGMELNARKSHGLTIEKASKRKCMVLRDTKYQCKGEGIQPMGT DDSVRYLGLHF	178
(45) R2ViLa 5'UTR	CTCCCGTCGCTTGACATTGGCGTTTTTTCGCAGTGCGTATATCAATTGCGAGAGCTAGTCTT TGTCCTTGGCGTGAGTTGATTGCATAGTTGCACGGGTTAGACACCTCCTAGTGGTGACCGCT GGATAAGGCCTCCGGGCCCGAACCAAGATCTAACCCACCCTCCGTTGCAGAAACAGCGGAGG CTAAACTGCTGCTA	179
(45) R2ViLa 3'UTR	GTGGTGACCCCTCTGGTCGACACCTGGACTAGCGACCAAGTCGCCTTAGAAGACCTGATGGG GGGGGTCTCCGTACCCTTTCTCGTCTTGTCTCTCTAATAATACATGTGAAAA	180
(46) R2Tral ORF	MLTLHMDPRPPGSPDPSWPGSCPREVCQESAAEPPKHSCPCGSEFRTKRGLAMHRSHRHQAE VNAERLASLPTRKLVWTQEEEAGLLRHATQAWKKGMAKTQIFASLLPHFPGRSVEAIKKRLQ SLKWVPPRDTPERNSQPLNATSGNVAPATLPAPDSSVTIDDQGRVEDPAASWRRALLTAAAV RLSEPKLGAETLRQITQDLQDGKISSKKCRVLMNEHAQHWFPHKWKPTSVRPEPKTRQWSTR EIRRANYASVQKLYKLRKRDAAQTVFSGRWRNAYRGQAPLPAEMDEYWAKIFETPSAIDSVT PEVPSKVLWSVMGKIVPSEVTDALKGMRNSAPGLDRISAEELLTWHQPSLAAYYNLMLAAGG PPEHLA**CSRVTFIPKLENPQAPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQR DGCLEASALLHAILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVASAMGLPPPLVHYLKH LYSRSTVRLANKVTTCGRGVRQGDPLSPLLFIMVMEDIVREALPEVGFNLDGQRVDSLAYAD DLVLLAEKSPRLQEKLHRLAEALRKAGMSLNARKSHGLTVTKVSKRKQMVLSPTSYECEGET IKPMGTDDSVRYLGLHF**NWKGRIVPKHTGLWDSLLKELKKAPLKPYQRLQLLRFHAVPKFTH QLVLGHAHRNTLKKLDCLTRAAVRQWLRLPPDTPLGYLHANVKDGGLGIPCFSTSIPLLQQK RFEKIATNPATIFQVIERQDSFRTQRRRLDRPCRLNKTVVTSKAEVREGWGNMLSNSVDGKE LRYPEVDKASHEWLLRPSRVFPRLHLRGIQLRGGLLPTKARSGRGKNRQVTNKTCRGACQNV ETLNHILQVCEVTHDARCARHNRVLHLLEKLIRKKGARATLEPVIPAGSSFIKPDLIIETDT QMFVLDVSVVAGYRMAESWDIKVEKYGRPDRVEAIRRWAAIPEKIRVEQLPVIVSNRGLCYR PSGVGLRALGLTSRDISDLCLLAISGSLRCYDLYMRGTRASKEN	181
(46) R2Tral RT	CSRVTFIPKLENPQAPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQRDGCLEA SALLHAILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVASAMGLPPPLVHYLKHLYSRST VRLANKVTTCGRGVRQGDPLSPLLFIMVMEDIVREALPEVGFNLDGQRVDSLAYADDLVLLA EKSPRLQEKLHRLAEALRKAGMSLNARKSHGLTVTKVSKRKQMVLSPTSYECEGETIKPMGT DDSVRYLGLHF	182
(46) R2Tral 5'UTR	TACCCATTGGTCTTGATTTTGATACTGGTGTTTTGCGAATTGGTGTTGATTGCAAAGTGTTG TTGCGGTGAATTGACGATTGGTTGCTAGGGTTGGGCACCTCCAAGTGGTACCCGCTGGATAG GGCCTTCGTGCCTGAACTAAGCCCCAACCCACCTCTCCGTTGCTGGCAGACCACAGTAACTG AGATTAAACAGCTGGT	183
(46) R2Tral 3'UTR	GCAGTGCCGAACGGGACCCAGAGGACCGCCTCGCCGATCAGGCGTGGGCGGTGACCACCCTG GCCTCTACCTGAAATAGCGATCTAATCGCCTTAGAAGGCTTGATGCTGGAGGGTCGAGCAGA CACCAAGTGTCACGAATGCTCCTTACCCTACCTCTTCTTGCTTCTCTAATAATCAATGTTAA CA	184
(47) R2Boja ORF	MLTRTDDPRPPGHPDPSCPGSCSRDCRESVAESPRHACPCGCEFPTKRGLAMHRFHRHMAEV NAERLATLPTKKLAWTQEEEAGLLRHASQAWKEGMTKTQIFAAVLPHFPGRSMEAIKKRLQS LKWVPPQNTPERNLQPLHATSGNVAPAALPSPDSSVKVDGQGQNTEDPAASWRRALLAAAVD HLEEPKLGAETLRQITQDLENGKISSKKCRVLMNEHAKHWFPHTWKPSSARPEPKARQWSNR EIRRGNYGAIQKLYKLRKRDAAQTVLSGRWRTAYRGRAPMPAKMEEYWSEVFQAPSATDAIP PGSPSDTLWSVMGKITSSEVTAALKGMRNSAPGLDRITAEELLTWHQPSLASYYNLMLVAGG PPKHLA**CSRVTFIQKVELPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQHGFLQR DGCLEATALLHAILRKVHNSRKPCAMLFLDVAKAFDTVSHQTLFRVASEMGLPPPLVNYLKY LYSRSTIKLADKVIKCGRGVRQGDPLSPLLFIMVMEVIVREALPEVGFNLDGQRVDSLAYAD DLVLLAEKSTRLQEKLERLSEALRKAGMALNARKSHGLTVTKASKRKQMVLTPTTYECEGDT IRPMGTDDSVRYLGLHF**NWKGRIVPKHTGKLDTLLKELTRAPLKPYQRLELLRFHTVPKLTH ELVLGHAHRNTLKKLDCLTRAAVRQWLRLPQDTPLGYLHAHVKDGGLGIPCFSTSIPLLQQR RFEKIATNPATIFQVLQRQDSFRTQWRRLDKPCRLNSTVVTSKVEVRKEWENMLSNSVDGRE LRYPEVDKASHEWLLRPSRVFPRLHMRGIQLRGGLLPTKARSGRGKNRQVTSKTCRGACQTV ETLNHILQVCEVTHDARCARHNRVLHHLEKLLRKKGIRTTLEPVIPAGSSFIKPDLIIETDT KLFVLDVSVVAGYRMAESWSIKVEKYGRPDRVEAIRRWAAIPEEIRVEHLPVIVSNRGLCYR PSGDGLRALGLTSRDISDLCLLAINGSLRCYDLYMRGTRATRETERHLP	185
(47) R2Boja RT	CSRVTFIQKVELPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQHGFLQRDGCLEA TALLHAILRKVHNSRKPCAMLFLDVAKAFDTVSHQTLFRVASEMGLPPPLVNYLKYLYSRST IKLADKVIKCGRGVRQGDPLSPLLFIMVMEVIVREALPEVGFNLDGQRVDSLAYADDLVLLA EKSTRLQEKLERLSEALRKAGMALNARKSHGLTVTKASKRKQMVLTPTTYECEGDTIRPMGT	186

	DDSVRYLGLHF	
(47) R2Boja 5'UTR	CCACTTGGTGTTGATTTGGATTTTGGTACTGGCGACTTGCTAGTTGGTATTGATTGCAAAGT GTTGTTGCAGGTTGAGCGGACGATTGGTTGCGCGGGTTGGACACCTCCACGTGGTGACCGCT GGATAAGGCTACGGGCCTTAACTAAGATCCAACCCACCCCTCCGTTACTGACAGACCACAGT AACTGAGATTAAACAGCTGGC	187
(47) R2Boja 3'UTR	CAAAGAGGACCCCTCTCTGATTAGGCGGTGACCACCCTGGCCTCTACCGAAATAGCGATCAA ATCGCCTTAGAGGGCTTGATTTGGGAGGCTGAGCTGACACTGTGTGTCGCGAACGCTCTTTG CCTTTCCTCTTCTTGCCTATCTAATAACCCGTGTTAATA	188
(48) R2Bofu ORF	MDRNTNVKGPHASSPSNAFTCSCGRQFSTKRGLGMHRRHCQPVEVNIERMAVLPTKRLIWSL EEDEALRRCADHQWKTGMLKKDLYTLLLPHFPGRSQEAIKKRLQTIKWVPPTDPALVSHPPP GVDTGSASDRGRDSSAASDSDPDPAATWVSSLLTATTSNLQEAKMGAEELYVIARDLAMGVI SQTQSRALLQDHAQKWFPHLWKPSKPRPLSRTHQWSKRTVRRANYAVIQRLYKLRRRDAAST IFSGQWKNAYRGSPPLPPDMESYWSSVFQTPSAIDDRVPATPSEVYWTLLCAMTPSEVTAAL KSMASSAPGLDRITAATLLTWHQPSIAAYFNLLLAAGGPPEHLA**CSRVVFLPKTEAPSTPGE FRPISIASTILRAFHKILSWRLRDTLQLSPFQHGFLQQDGCLEATTLLHTILRRVHLKCEPC AMLFLDIAKAFDTVSHHTLFRVATAAGLPPPLVCYLRHLYERSTIQLAGTTTVCGRGVRQGD PLSPLLFILVIEDIVTSSLPSIGYDLDGHRIHSIAYADDLVLFAEKSPRLQEKLHLLSEALS KAGMTLNAKKSIGLTIAKDGKRKRTALLPTTYRCGIGNINPLGPGETLRYLGIQF**NWKGRVT PKHTGKLETLLQELTQAPLKPQQRLLLLKFFLVPKFTHELVLGHAHRNTLRKMDWLIRAAVR RWLRLPKDTPVGYLHAHIKDGGLGVPCLSTTIPLLQASRLEKIAAHPTTLFQQIQRQDALGP LRRRMNQPCRINTTTVTTKAEAREEWKSTLVASFDGRELGVPKVDPVSHDWLLRPSRVFPRL FLRGIQLRGGVLPTKSRSARGANRKVVDKICRGACRATETINHTLQVCEVTHDARCARHNRI VRRLDQLLRRQGIQPSVEPIIPSGTSFIKPDLIAEINNRLWVMDVSVVAGHRMQETWKIKVD KYGKPEQVASILRWAAVDSSVPVHHLPIVITNRGLIFEPSAAGLRTLGLSRRDLVDLCLLTI MGSLKCYDVYMRGTF	189
(48) R2Bofu RT	CSRVVFLPKTEAPSTPGEFRPISIASTILRAFHKILSWRLRDTLQLSPFQHGFLQQDGCLEA TTLLHTILRRVHLKCEPCAMLFLDIAKAFDTVSHHTLFRVATAAGLPPPLVCYLRHLYERST IQLAGTTTVCGRGVRQGDPLSPLLFILVIEDIVTSSLPSIGYDLDGHRIHSIAYADDLVLFA EKSPRLQEKLHLLSEALSKAGMTLNAKKSIGLTIAKDGKRKRTALLPTTYRCGIGNINPLGP GETLRYLGIQF	190
(48) R2Bofu 5'UTR	TGGTTACGCGGCCTGGACGCCTCCTAGTGGCGACCGCTGGATAAGGGGTAGATAGTTTAACT TCTGCTCCCGAGCTAAGAGCCAGGCCACCCCTCGGCGAGGTTAGTACCTTGTTGAGGTTTCA CTGCTGGTTTTCTAATTGAAAGGAATTGTCCCGGTCCCCCTGGACTTAAGGACTCCAGCAGC	191
(48) R2Bofu 3'UTR	GGTTGAACTGGCTTCTATTGGTCCCCCCCCCAACACATTAGTGTACTACTCAATTAGCGACT TGTCGCTCTCAGGGAGGCTCTCCTCCCTGATAATATATGTCACG	192
(49) R2CroTig ORF	MLTDTVVLRPPGLPDPSCPGSCPRVASPKSDAECPKHSCPCGSEFRTKRGLAMHRFHCHPAE VNAERLTTLPTKKLVWTQEEMDGLLRHANGAWKEGMLKTQIFTSLMPHFPGRSAEAIKKRLQ SLKWVPARDNPERNSQPLHATSGNVAPATLPTPDPSRVDDQGANIEDPAASWRRALLTAAAD HFKEPKLGAGTLKQITQDLLNGKISSRECRVLMGEHAQHWFPHTWKPSSARPEPKARQWKAR EIRRANYAAVQKLYKLRRRDAAQAVFSGSWRNAYRGRAPQPDKMDQYWAEVFETASATDVTP PGSPSEIYWSVVSKITPSEVADALKGMRNSAPGLDRITAEELLTWDHPSLAAYYNLMLAAGG PPEHLA**CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQR DGCLEATALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKC LYSRSTVRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYAD DLVLLAEKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEP IKPMGTDDSVRYLGLHF**NWKGRIVPKHTGKLDSLLKELTKAPLKPYQRLQLLKFHAVPKFTH ELVLGHAHRNTVKKLDCLTRAAVRKWLRLPQDTPLGYLHANVKDGGLGIPCFSTSIPLLQKK RFEKIATNPATIFQIMQRQDSFRTQGRRLDKPCRLNSTVVTSKTEVREEWGNMLSNSVDGRE LRHPEVDKASHEWLLRPDRVFPRLHMRGIQLRGGLLPTKARSGRGNNRQVTSKTCRGACQNV ETLNHILQVCDVTHDARCARHNRVLHRLERLLHKKGLRTTLEPVIPAGSSFIKPDLIIETDT QTYVLDVSVVAGYRMTESWDIKVEKYGRPDRVEAIRKWAAIPATKRVEHLPVIVSNRGLCYQ PSGDGLRALGLTSRDISDLCLLAISGSLRCYDLYMRGTRAV	193
(49) R2CroTig RT	CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQRDGCLEA TALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKCLYSRST VRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYADDLVLLA EKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEPIKPMGT DDSVRYLGLHF	194
(49) R2CroTig 5'UTR	TCGACCATTTGGTGTTGATATTGATACTGGCGACTGGCTAGTTGGTGTTGGTTGAAAGTGCT GGCGAGGTTGAGCTGATAATTGGTTGCACGGGTTGGGCACCTCCTAGTGGTGACCGCTGGAT AGGGCTTCGGCCTGAGCCAAGATCCAACCCACCCTCCGTTGCTGGCAGACCCCAGTAACTGA GATTAAACAGCTGGT	195

(49) R2CroTig 3'UTR	GACTACCGAGCGCCACCGGAGGGGACCCAGAGGACCGTCTCGGTGATTAGGCGGTGACCGCC CTGGCCTCTACCTGAAATAGCGATCAAATCGCCTTAGAGGGCCTGATTTGGGGGGTTGGGCA GACACTCTGTGTCATTACTGCCCTTTACCTTTCCACTTCTTGCCTGTCTAATAATAAATGTT AACA	196
(50) R2GraSmi ORF	MTIITNPSPPGPSDPSDTDRYPGHTDTRDSPRERQEDRKETQQYICHCGSRFKSIRGLGRHR AHRHMAEVNAERVAALPTKRLKWSQEDDEALCRLASRHWVLGMTKKTLIAQITSFFPGRSHE ALKKRLNSVNWGPPNVTPERTNIPQPEHPERLANLIGPVAHTASEPGHPDRPTKVATPDGPV AHSHPEPDQDAATSWRCALLGAAADRLKEVKLGAGELAQIARDLQNGCTRKQCRARLNHHAK KWFPHTWKPSVPRAQDANRQWSKTAIRRANYAQVQKLYKLRQRDAASTVLSGSWRTAYKGPA PVPPKMEEYWAAVFNTPSVVDTSVPAEPPAVLWSLMAAITPSEVTAALTGMKASAPGLDRLS ALELLSWDQPSLAAYYNLLLAAGGPPKHLT**CSRVVFLPKTENPSAPGDYRPISIASTILRAF HKIMAWRLRDTLQFSPLQHAFLQRDGCLEASTLLHNILRRIHTDRVQGAMLFLDVAKAFDTV SHSSLFRVAYEAGLPPPLVSYLRHLYRHSTVRLAETTTKCGRGVRQGDPLSPLLFIMVIDDV VKNSLPDVGFDFDGHRVGSLAYADDLILFAEKHTRLQEKLHLLSEALQRAGMALNTKKSHGL TIAKDGKRKCMALVPTPYECKGATILPLGTDDTVRYLGLQF**NWKGRVIPKHTGRLDTLLDEL TRAPLKPHQRLKLLRVYTCPKLTHELVLGHAHRNTLKKLDCLTRAAVRRWLRLPTDTPTGYF HAHVSDGGLGLPCFSTTIPLLQQKRLEKVACNPATLFQILQRQSSFRSLERRLDQPCRLNGA VVTSKAEVREGWQIMLSDSIDGRELSIPGVDKASYEWLRHPARVFPRLYIRGIQLRCGLLPT KARSARGANRQVIARTCKGACCASETQNHILQVCELTHDVRCARHNRILRYLNKQLHRLGLQ TWVEPVIPSGSSFIKPDLIVETEKQLWVLDVSIVAGYRLEESWNIKAAKYGRSDRVVDIRRW ATVPDSTFVCQLPVIISNRGLLYAPSGVGLRALGLSRRDIMDLCLLAISGSLKCYDLYMRGN	197
(50) R2GraSmi RT	CSRVVFLPKTENPSAPGDYRPISIASTILRAFHKIMAWRLRDTLQFSPLQHAFLQRDGCLEA STLLHNILRRIHTDRVQGAMLFLDVAKAFDTVSHSSLFRVAYEAGLPPPLVSYLRHLYRHST VRLAETTTKCGRGVRQGDPLSPLLFIMVIDDVVKNSLPDVGFDFDGHRVGSLAYADDLILFA EKHTRLQEKLHLLSEALQRAGMALNTKKSHGLTIAKDGKRKCMALVPTPYECKGATILPLGT DDTVRYLGLQF	198
(50) R2GraSmi 5'UTR	TTAAGAAATCGTTCATAGTTACAGGCTTGGGCGCCTCCGTGTGGCGACCGCTGGATAGGGCA CATTGTGCCTGAGCTAAGACCCAAGCCACCCCTCTCCTGGTCTAATTGCTTTGGAGAGGGAC TGTCACTGGTTCAATACATATT	199
(50) R2GraSmi 3'UTR	GCTTGGACTGGATTGTTTTCTTATCTGTATATATGTATATATTATTATGATGTCTGTATTGT GACGTACCATGTGGTCTCTATCTCATAGCGAGTGATTCGCCCTAGAAAGCTTGGTGGGCGGG ACGGGAATTGTCAGGTTGCTTTTTATATCGTACCATTGGGTTATCGTGCACTGCTTTCCTAT GCCTGTACTGCCTTGCTATTCTAAAATAAACAGTTAAC	200
(51) R2SibCol ORF	MTIVFSPSPPGHPDPSDIDGNANSTGSQNGGDKPSFPCFCGSRFKTKRGLGQHRSRCHPAEV NAERIAALPTRRLTWTHEDDEALRRHADQHWEKGMMKKVLITKLTPYFPGRSHEAIKKRLQL LKWDPLKENPERILYPQTHDCPERPHTDAPPTSPVAAQLTHELDPSAATTWRSALLTAAADH LKENKMGANELTQIARDLQAGVASRKQCRARLNQHAQSWFPHTWKPSVARPKTATHQWSKKA IRRANYATIQKLYRLRRRDAASTVLSGRWRNAYKGPTPVPPEMEEYWSTVFKTPSATDDRKP AEPSAVLWSLMVAIAPSEVSAALQGMKASAPGPDRIPAVELLTWDQPSLAAYYNLLLAVGGP PKHLS**CSRVVFLPKTEDPMTPGDFRPISIASTILRAFHKILARRLRDTLQLSPLQHAFLQRD GCLEATTLLHTIIRRIHEDRVPCAMLFLDIAKVFDTVSHQSLFRVAYEAGLPRPLVSYLRHL YKKSTVRLAEATTKCGRGVRQGDHLSPLLFIMVIEDIIKTALPDIGFDLNGQTIGSLAYADD LVLLAGKSSRLQEKLHLLSGALQQAGMALNAKKSHGLTIAKDGKRKCMSLVPTLYECKGDNI QPLGTDDSVRYLGLQF**NWKGRVIPKHTGKLDNLLNELTKAPLKPHQRLKLLKSHLVPKFTHE LVLGHAHRNTLKRLDWLMRVAVRHWLRLPTDTPLGYFHANVLDGGLGLPFFSTSIPLLQQKR LEKVALNPATLFQIVQRQNSFRSLGRRMDQPCRLNGTVVTSKKEVREEWQNMLSNSVDGKEL SIPEVYKASHEWLRRPDRVFPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRTCRGSCCAPE TLNHILQVCKVTHDARCTRHNRVLGYLEKLLRRAEIRTRVEPVIPSGSSFIKPDLIVETANQ LWVLDVSIVAGYRMEDTWDIKAAKYGRPDHVADIRRWAAVPATTPVTQLPVIVSNRGLFHQP SGDGLRALGLSRRDIMDMCIYAIRGSLKCYDLYMRGTSGWKNV	201
(51) R2SibCol RT	CSRVVFLPKTEDPMTPGDFRPISIASTILRAFHKILARRLRDTLQLSPLQHAFLQRDGCLEA TTLLHTIIRRIHEDRVPCAMLFLDIAKVFDTVSHQSLFRVAYEAGLPRPLVSYLRHLYKKST VRLAEATTKCGRGVRQGDHLSPLLFIMVIEDIIKTALPDIGFDLNGQTIGSLAYADDLVLLA GKSSRLQEKLHLLSGALQQAGMALNAKKSHGLTIAKDGKRKCMSLVPTLYECKGDNIQPLGT DDSVRYLGLQF	202
(51) R2SibCol 5'UTR	TTAAACAGCGTATAGTTGCAGGGCTTGGACGCCTCCCCGTGGCGACCGCTGGATATGGTACC CTAAGTGCCAGAACTAAGATCCAAGCCCCCCCTCTGCGAGGGCCCAAGTCTCGTAGAGGATT CGTTGCTGGTTATTAAA	203
(51) R2SibCol 3'UTR	TCTGTTTTGAAGGGCCTGTAACCTATTGACATTACTGTCTGATAGCAGTTTTATCTGCCTTA GAGGGCCTGGCGGGCGGGTTGGTATAGTGAGATGTGTGTACTGTATTGTACTGTATTGCCTA ATATTGTATCGTACTGTATTGTGCATTGCCTTACCTTGTACCGTACTGACTTGCTTTTCTAA AAATAAAATGTGACT	204

(52) R2LacBil ORF	MNHRGNAKRGGGSTDRTTGVRSSSRAAGPGVVPAVEPPVFVCEYPGCERTFNTKTGRGVHHR RAHQDWYDAQQNIHQVKRRWTPEETSLLAREEARLLTSGTRHMNLALMEKFPERTAEAIKAH RRQQAYRELVVAYTQEQLEVAARPAPQPVVEASNTQWRNRLIQHLQRLPRMTATTEHLERLY RLSATVELRSLTDVFDELSLVLRGMQGTRPWRAQRGERSSQETPTSNRRKRRKQFAKFQHLW HKSPTKCIKWALDKNRDAPIENGADLFPFWEKLMTRAPDCTPGPGPRRNEITSLWEPITAAE IKRALPSASTAPGPDGMTVEQVKRLPLGVLERLFNLIMVCGKCPQSML**GARTTLIPKKEYAR** **DPGDFRPITVSPVLTRVFHKVLAAQLTREVDLDPRQRAFQQFDGCSENIFLLDMVLRYHKQR** **FKPLYIASLDVAKAFDSVAHYTIKETLQMMGVPDPMTHYISDVYDRSTTVLSCGGATSQEFH** **PRCGVKQGDPMSPCIFNMIMDRLLRKIPREVGADLGEGVTENGVAFADDLVLLASTPVGLQT** **PIDIVADHLALCGLTVNSGKSLTVSLETVPREKKTVVNPRRVFTCNNRPLQALKRSDQWRYL** **GVPF**TPDGRMSARPLDRLRELLDTIGKAPLKPQQRMFGLRTVLLPSLYHQLSMGSTTVGLLN KVDRCVRTAVRKWLGLPDSTPVAYFHAKVGDGGLGIPAMRWMMPRLRRDQLLRLPLQRTEES TSGTFLRRELSLVEGRLLERGNSLFTSQAQDRRWASLLYQSVDGCALRESGKTRSQHDWVYG CTRFLSGKDYRNSIRMRINALPTRSRLARGRRADRLCRGGCQMPETLNHVLQVCYRSHAARV RRHDAVVKYISKGLRRQGYHVDLEPHFRTSEGLRKPDIVASIGRTARIIDAQVVGEQADLRR AHAQKVQYYAGNSSLLEMVKQRYQVDDLCTTSATLSCRGVWSDRSAEDLLAMGIVRKQDLKV LSSRVVLGGIYAHSVFNKTTASLSAGW	205
(52) R2LacBil RT	GARTTLIPKKEYARDPGDFRPITVSPVLTRVFHKVLAAQLTREVDLDPRQRAFQQFDGCSEN IFLLDMVLRYHKQRFKPLYIASLDVAKAFDSVAHYTIKETLQMMGVPDPMTHYISDVYDRST TVLSCGGATSQEFHPRCGVKQGDPMSPCIFNMIMDRLLRKIPREVGADLGEGVTENGVAFAD DLVLLASTPVGLQTPIDIVADHLALCGLTVNSGKSLTVSLETVPREKKTVVNPRRVFTCNNR PLQALKRSDQWRYLGVPF	206
(52) R2LacBil 5'UTR	GGGAATGACTGCTGATGATCTGTGATGTTAATCTGACGATCTTCGGAGGTCTTTTCCACGTG CCTATCCCTACCTCCTCGTGGTGAGAGTCGGGCAACGTCTTTAGGCATACGTCTTTAGGCGG CTAACAGGGAAGGCTTACCTGAGGCTCTGGGTCCTATACCAGAGAGTCTCAGTTCAACCTCA CCCAGATTCGGGCATCCCCCGAA	207
(52) R2LacBil 3'UTR	GCGCCCAGGAAATTGCCCAGGTGTAGCACGGGACTGCTAGAAAAATAGTCCGTACGTCTGCG GATATACAGACGCAATTTCTGAGGGGCGTGGCGCCCCGATGATGACGGAAA	208
(53) R2LiaOli ORF	MEAAEFRCACGRVWPSARSLAMHNHHRHPEVVNEARLQARPVSRLHWTGEEEEALRRLADRT WTAGMLKGELCSKLADSFPGRSAEAIKKRLQHIKWAPLSSGAISSEASTPRAALICDGFDPV EWRKKLLGDIVKNLNEPRMGSDTLVAISQRLRTGELSLVAAVECLEDHAKQWFPHKWRPSIP RSQKPTQPHSRKAIRRATYAAVQRLYYRRQKDAASAVLSGQWRDIYRQRKEVVGLNRHWAKV FEIPSVADTEPPIMPSNLLCSLLDRIHPSEVTAALGDMTASAPGLDRLTAGELLKWHQPSLS AYFNLLLAVEAPPRHLA**CSRTAFIPKVADPTSPEEFRPISIASVLLRAFHKILARRLRDHLD** **LSSLQFAFLLRDGCLEATTLLRSVLRRIHHERTSCAMLFLDISKAFDTVSHGTLSRVMRSAG** **FPPPLVSYIQHLYECSTVQLGDTMVSCNRGVRQGDPLSPLLFILVIDDILRKSLPEVGYNFG** **GLRLGAMAYADDIVLLADHPQRLQEKLDVFSQALHAAGMALNAKKSVGLTVVRDGKRKVVAL** **DPTCYNYRDGPIRSVGLRESVRYLGLQF**DWKGQVRFKHTRKIDSMLEEITRAPLKPHQRLRI LSAFLIPKFTHELVLGAAHRKTLLQLDRQVRVAVRSWLHLPKDTPLGFFHASVRDGGIGIPS FSTTIPILQKNRFEKLASQTSQWSVVVRDQAAFRVLSNIVNRPCKLNNMAVTTKKESRVEWQ RALSQSVDGNEVAIPGVDPVSHRWLRDPSRVFPRLFLRGIQLRGGVLPTRARSARGINREVT DKRCRGPCQASETINHILQACEITHDARCARHNRVLQRLDQCIRKGGAETWVEPIICSGSSF IKPDLIVSKLDHLWILDVSVVAAPRLPVTWDLKVAKYGSAVCARDISTWHASSVPLKHLPVV FSNRGLCYEPSGKGLRALGLSHRDIMDLCLLTIQGSLKCYDLYTRGTRSGFGGTV	209
(53) R2LiaOli RT	CSRTAFIPKVADPTSPEEFRPISIASVLLRAFHKILARRLRDHLDLSSLQFAFLLRDGCLEA TTLLRSVLRRIHHERTSCAMLFLDISKAFDTVSHGTLSRVMRSAGFPPPLVSYIQHLYECST VQLGDTMVSCNRGVRQGDPLSPLLFILVIDDILRKSLPEVGYNFGGLRLGAMAYADDIVLLA DHPQRLQEKLDVFSQALHAAGMALNAKKSVGLTVVRDGKRKVVALDPTCYNYRDGPIRSVGL RESVRYLGLQF	210
(53) R2LiaOli 5'UTR	AGGGCATAGTTGCACGGCCTGGGCGCCTCCACGTGGCGACCGCTGGATAGGGTGCATTTTTA CACCTGAACTAAGATCCAGGCCACCCTCCTTTTGGGCTTGATTTCCCAATAGGAGGTTCTCT GCTGGTTTCCGAACCAATTTGTAATTGTTGTCCTGTGCTCTCTGGTGGCAGGACTTCCAGCA GACATAGTGCCAGTAACTTCCTTGGCCAATTAGCAGTCCACC	211
(53) R2LiaOli 3'UTR	AGTTTAGTGGGGTGGGTGTGAGATCTGGGTCAAGGAATGTTCCTTTAGAACGGCGCTAGTTA ATTAGCGAGTTTTCGCCTTGGAGGATGATTATGCTTGGTTAACGTCCCCCAAGATATTGGTG GGTTCGGAACCTGTTTGATTATCTGTACCAATAATAATAAATGATAAAA	212
(54) R2MasLat ORF	MDPSNKDGNASSVGTQDNNTRPSYRCFCGSQFKTQRGLGKHQSHRHPAEVNAKRVAALPTKR LAWTQEDDEALKRHSDRLWQTGMTKTKLYPLLLPYFPERSLEAIKKRLLLTHWVPPNETPEL NPPPEHPERPGSPDRSQPPNQQPPMSPVAPPLNGLGSGNATSWRGNILASAADHLKESRMGS EELAQIARDIQTGVISRKQCRSRLNQHAKSWFPHKWKPSVNRAVETNRQWSKKAIRRANYAE VQKLYKLRRRDAASTVLSGHWRNAYKGPAPVPPEMEEYWAAVFKTPSAVDDRAPAEPPAVLW SLITPIAPSEVTAALNGMKASAPGLDRISAMELLSWDQPSLAAYYNLLLVAGGPPKHLS**CSR**	213

	VVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRDTLQLSPLQHAFLQRDGCLEASTL LHTILRRVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKSTVRL AETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAERST RLQEKLQLLSEALQRAGMALNAKKSHGLTVARDGKRKCMALVPTTYECKGETIKPLGTDDSV RYLGLQFNWKGRIIPKHTGMLHSLLNELSKAPLKPHQRLKLLSAHLVPKFTHELVLGHAHRN TLKRLDCLTRAAVRAWLRLPTDTPVGYFHAHVLDGGLGLPCFSTTIPLLQQKRLEKVASNPA TLFQIVQRQTSFQTLGRRIDRPCRLGGSVVTTKAEVREGWQIMLSNSVDGKELSISEVDKAS HEWLRHPARVYPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRTCRGACCAPETQNHILQVC EVTHDARCARHNRVLRCLEKLLRRKDLHPRLEPVIPSGASFIKPDLIVEVMNQLWILDVSIV AGYRVEETWDIKASKYGRPDRVADIRRWAAVPATTPVIQLPVVISNRGLCYRPSGDGLRALG LSRRDTMDLCILAIIGSLKCFDLYQRGTKDWNV	
(54) R2MasLat RT	CSRVVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTILRRVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAE RSTRLQEKLQLLSEALQRAGMALNAKKSHGLTVARDGKRKCMALVPTTYECKGETIKPLGTD DSVRYLGLQF	214
(54) R2MasLat 5'UTR	TTAAGAAGAGCGTAGTTGCAGGCTTGGGCACCTCCACGTGGTGACCGCTGGATATGGTACTT CGGTACCAGAACTAAGATCCAAGCCACCCCTCTACACGACTTTGCTTTGTAGAGGGTTAGTT ACTGGTTTACGTATTGGTTTTCGTGTTAGTCCAAGTCCCTCTGGAACT	215
(54) R2MasLat 3'UTR	TATGTCTTGTTATACCTATGCTTGTATGTTGTTACCTGTCTATTAGCGAGCAATCGCACTTA GAGGGCCTGGCGGGCGGGAAAGGTCTAGTCGGTTGTATTTGCTGTATTAACATGTATACAAT GTATATTGATACCATGCAATGTGATGTATCTTGTGATATCGTATTATGCACTGCTTTACCCT GTACTGTTTTGCCTTGCTTAATCTAAAAATAAAGTGTAAAA	216
(55) R2MasLat2 ORF	MDPSNTDGNASSVGTQDNNNARSSYLCFCGSKFNSERGLGKHQAHRHPAEVNAKRVAALPTK RLMWTQEDDDALKRHSDRLWQMGMTKSRLYPLLLPFFPERSIEAIKKRLLITNWAPPIITSE LNPHEHPERRTNPERRTNLERLAHPKELTNPDRAAPLTQQPPTSPVAHPPNGLGSGNATSWR GNILASAAVHLKENRMGDEELAQIARDLQTGAISRKQCRSRLNQHAKRWFPHTWKPSATKTV KANRQWSNKAIRRANYAQVQKLYKLRRRDAASTVLSGRWRNAYKGQAPVPPEMEEYWAKVFN TPSAIDNSAPAEPSAVLWSLITPIVPSEVTAALNGMKASAPGLDRLPATELLSWDQPSLAAY YNLLLAAGGPPEHLS**CSRVVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRDTIQLS PFQHAFLQRDGCLEASTLLHTILRRVHTERVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLP PPLVSYLRHLYKKSTVRLADTTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGLR VGSLAYADDLVLFAEKSTRLQEKLQLLSEALERAGMALNAKKSHGITIAKDGKRKCMALVPT TYECKGETIKPLGTDDSVRYLGLQF**NWKGRIIPKHTGMLDSLLNELSKAPLKPHQRLKLLSV HLVPRFTHELVLGHAHRNTLKRLDCLTRAAVRRWLRLPTDTPVGYFHAHVLDGGLGLPCFST TIPLLQQKRLEKVASDPATLFQIVQRQASFQTLGRRIDRPCRLGGSAVTTKAEVREGWQIML SNSVDGKELSIPEVDKASHEWLRHPARVFPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRT CRGACCAPETQNHILQVCELTHDARCARHNRVLRHTEKLLRRKGLRTRLEPVIPSGASFIKP DLIVEMTDQLWILDVSIVAGYRLEDTWDIKASKYGRPDRVTDIRRWAAVPDTTPVTQLPVVI SNRGLCYRPSGDGLRTLGLSRRDTMDLCLLAILGSLKCFDAYKRGTKDSNI	217
(55) R2MasLat2 RT	CSRVVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRDTIQLSPFQHAFLQRDGCLEA STLLHTILRRVHTERVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLADTTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGLRVGSLAYADDLVLFAE KSTRLQEKLQLLSEALERAGMALNAKKSHGITIAKDGKRKCMALVPTTYECKGETIKPLGTD DSVRYLGLQF	218
(55) R2MasLat2 5'UTR	TTAAAAAGAGTGTAGTTACAGGCTTGGACACCTCCCAGTGGTGACCGCCGGATATGGTGCTT CGGCACCAGAGCTAAGAGCCAAGCCACCCCTCTACTTGACGATCGCTCTTGCAGAGGGTCTG TTACTGGTCTACTAAATATCAATAGTAATAGTCCTAGTCCCCCCGGAACT	219
(55) R2MasLat2 3'UTR	TTATGTCTTGGACTTACGTTTTTGTGTTGTTACCTATCTCCTAGCGGAAATCGCACTTAGAG GGCCTGGCGGGCGGGATGGTCTTGCCGGTTGTATTTGTATTCCATATATGATGCACTGCTTT ACCATGTACCGTTTTGCCTTGCTTTTCTAAAAATAAACAGTCAAC	220
(56) R2OxyLep ORF	MSAKVSPSPSGPLDSSSIDNDAMNTVAEDRAQHPSFPCFCGLCFTTKQGLGQHQAHRHPAEV NAKRVAALPSKRLAWTLEDDETLRRHADGYWEKGMMKKTLYTLLAPRLPGRSQEAIKKRLQI LKWAPPSEHPERLDNPERETVQRAKGAPPSEHPERPDNPDRGAVHPECNTIPTPTASQPVAN PLTHEPDSTVDAAATWRRALLVAAAGRFRENKMGARELTQAARDLQNGVTSKKQCRARLNQH AEKWFPHKWKPSVARPETANRKWSKKAIRRANYANIQKLYKLRRRDAASTVLSGRWRNAYKG PAPMPAEMEEYWASVFETPSANDDSAPAEPSAVLWSLMDAIVPSEVTAALHGMKTSAPGLDR IPAEELLTWDQPSLAAYYNLLLAAGGPPKHLS**CSRVVFLPKTDTPSAPGDYRPISIASTILR AFHKIMARRLRDTLQLSPLQHAFLQRDGCLEATTLLHSIFRRIHMDRVPCAMLFLDVAKAFD TVSHKSLFRVAYEAGLPPPLVSYLRHLYSKSSVRLANKDTNCGRGVRQGDPLSPLLFILVID DIIRSSLPEVGFDFNGQRVGYLAYADDLVLLAEKSSRLQEKLDLLCGALQKAGMALNAKKSH GLTFAKDGKRKCMALLPTPYECKGEPINPLGTGDSVKYLGLQF**NWKGRVCPKHTGKLDFALQ ELTKAPLKPHQRLKLLTLHLVPKFTHELVLGFAHRKTIKKLDCLTRAAVRKWLRLPTDTPVS	221

	YFHAHVRDGGLGLPCFSTSIPLLQRKRFEKVASNPDTLFQIVQRQDSFRSSGRRVDQPCRLN GAVLTSKTEVREEWKAVLHQSVDGKELSAPEVDKASHEWLRRPDRMFPRLYIRGIQLRGGLL STKARSARGKNRQVTDRTCKGACRAPETLNHILQVCEITHNARCARHNRVLRCLEKELRRAD AQTRMEPVIPSGSTFIKPDLIVEAANQLWVLDVAVVAGYRMEETWTLKTTKYGRPDRVEDIC RWAARPATTQVHHLPVVISNRGLFYQPSAEGLKALGLSQRDLMDLCILAITGSLKCYDLYQH GSMTWGCK	
(56) R2OxyLep RT	CSRVVFLPKTDTPSAPGDYRPISIASTILRAFHKIMARRLRDTLQLSPLQHAFLQRDGCLEA TTLLHSIFRRIHMDRVPCAMLFLDVAKAFDTVSHKSLFRVAYEAGLPPPLVSYLRHLYSKSS VRLANKDTNCGRGVRQGDPLSPLLFILVIDDIIRSSLPEVGFDFNGQRVGYLAYADDLVLLA EKSSRLQEKLDLLCGALQKAGMALNAKKSHGLTFAKDGKRKCMALLPTPYECKGEPINPLGT GDSVKYLGLQF	222
(56) R2OxyLep 5'UTR	CTTGCCCCACGCCTCCCAGGGTTTGGACGCCTCCACGTGGCGACCGCTGGATAAGGGGTGGT GGTCCCAGACTATCCACACCCGAACCAAGAGCCAAACCACCCTCCTGCGGCCCCTAACACAC CGCAGGAGGACTGCTGCTGGTCAATAAA	223
(56) R2OxyLep 3'UTR	GGAGTACTCGGTGACGGCGTAACAGGCTGAAATTGTCTTCTAGCGATGCAAATCGCCTTAGA TAACTTGGTGGGAAGGGGGGCTTTGCGCGCGGAATGTGCTCTGCCATGTACCTGTACCGCCT TGTTACTCTAAAAATAAATGTTAC	224
(57) R2PanAll ORF	MDPSNKDGNASSAGTQDTNTRPSYRCFCGSQFNTQRGLGRHQAHRHPAQVNALRVAALPTKR LAWTQEDDEALKRHSDRLWQSGMTKTKLYPLLIPYFPERSLEAIKKRLLLTHWAPPIETSEQ NPPQEHPERPENPDRPPTIQQPPTSPVAPPQNGPGCGNATSWRGAILASAADRLKETKMGGE ELAQIARDIQTGVISRKQCRARLNQHAKSWFPHKWKPSVTRPAGANRQWSKKAIRRANYAEV QKLYKLRRRDAASTVLSGRWRNAYKGPAPVPPEMEEYWAAVFKTPSAIDDRVPAEPPAVLWS LITPIVPSEVTAALNGMKSSAPGLDRLSATELLSWDQPSLAAYYNLLLAAGGPPKHLS**CSRV** **VFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEASTLL** **HTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKSTVRLA** **ETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAERSTR** **LQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPYECKGDAIQPLGTDDSVR** **YLGLQF**NWKGRIIPKHTGMLDSLLNELSKAPLKPHQRLKLLSVHLVPKFTHELVLGHAHRNT LKRLDCLTRAAVRHWLRLPTDTPVGYFHAHVLDGGLGLPCFSTTIPLLQQKRLEKVASNPAT LFQIVQRQTSFQSLESRVDRPCRLGGTVVTTKAEVREGWQIMLSNSVDGSELNISEVDKASH EWLRHPARVYPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRTCRGACCAPETQNHILQVCE LTHDARCARHNRVLRCLDKLLRRKGLRTRLEPVIPSGASFIKPDLIVEVMNQLWILDVSIVA GYRVEETWVIKTTKYGRPDRVADIRRWAAVPATTPVVQHPVVISNRGLCYRPSGDGLRALGL SHRDTMDLCLLAIMGSLKCYDLYQRGTKDWNV	225
(57) R2PanAll RT	CSRVVFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAE RSTRLQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPYECKGDAIQPLGTD DSVRYLGLQF	226
(57) R2PanAll 5'UTR	TTAAGAAGAGCGTAGTTGCAGGCTTGGGCACCTCCTCGTGGTGACCGCTGGATATAGTACTT CATGTACTAGAACTAAGATCCAAGCCACCCCTCTACACGACTTTGCTTTAGTAGAGGGTTAG TTACTGGTTTACTTATTTATATATCGTACCAGTCCAAGTCCCCCTGGAACT	227
(57) R2PanAll 3'UTR	TATGTCTTGTAATGCTTATGCTTGTATGTTGTTACCTGTCTATTAGCGAGCAATCGCACTTG GAGGGCCTGGCGGGCGGGAAAGGTCTAGTCGGTTGTATTTGCTATATAAACTTGTATACCAT GTATATTGATACCATGCATATGTGATGTATCTTGTTGTATCGTATTATGCACTGCTTTTCCA TGTACTGTCTTGCCTTGCTTAAACCAAAAATAAACTGTATCC	228
(58) R2PanGut ORF	MDPSNKDGNASSAGTQDTNTRPSYRCFCGSQFNTQRGLGRHQAHRHPAQVNALRVAALPTKR LMWTQEDDEALKRHSDRLWQSGMTKTKLYPLLIPYFPERSLEAIKKRLLLTHWVPPIETSEQ NPPQEHPERPENPDRPPTIQQPPTSPVAPPQNGPGCGNATSWRGAILASAADRLKETKMGGE ELAQIARDIQTGVISRKQCRARLNQHAKSWFPHKWKPSVTRPAGANRQWSKKAIRRANYAEV QKLYKLRRRDAASTVLSGRWRNAYKGPAPVPPEMEEYWAAVFKTPSAIDDRVPAEPPAVLWS LITPIVPSEVTAALNGMKSSAPGLDRLSATELLSWDQPSLAAYYNLLLAAGGPPKHLS**CSRV** **VFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEASTLL** **HTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKSTVRLA** **ETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAERSTR** **LQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPYECKGDAIQPLGTDDSVR** **YLGLQF**NWKGRIIPKHTGMLDSLLNELSKAPLKPHQRLKLLSVHLVPKFTHELVLGHAHRNT LKRLDCLTRAAVRHWLRLPTDTPVGYFHAHVLDGGLGLPCFSTTIPLLQQKRLEKVASNPAT LFQIVQRQTSFQTLESRMDRPCRLGGTVVTTKAEVREGWQIMLSNSVDGSELNISEVDKASH EWLRHPARVYPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRTCRGACCAPETQNHILQVCE LTHDARCARHNRVLRCLDKLLRRKGLRTRLEPVIPSGASFIKPDLIVEVMNQLWILDVSIVA GYRVEETWDIKTAKYGRPDRVADIRRWAAVPATTPVVQHPVVISNRGLCYRPSGDGLRALGL SHRDTMDLCLLAIMGSLKCYDLYQRGTKDWNV	229

(58) R2PanGut RT	CSRVVFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAE RSTRLQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPYECKGDAIQPLGTD DSVRYLGLQF	230
(58) R2PanGut 5'UTR	GAAGAGCGTAGTTGCAGGCTTGGGCACCTCCTCGTGGTGACCGCTGGATATAGTACTTCATG TACTAGAACTAAGATCCAAGCCACCCCTCTACACGACTTTGCTTTAGTAGAGGGTTAGTTAC TGGTTTACTTATTTATATATCGTACCAGTCCAAGTCCCCCTGGAACT	231
(58) R2PanGut 3'UTR	TATGTCTTGTAATGCTTATGCTTGTATGCTGTTACCTGTCTATTAGCGAGCAATCGCACTTG GAGGGCCTGGCGGGCGGGAAAGGTCTAGTCGGTTGTATTTGCTATATAAACTTGTATACCAT GTATATTGATATCATGCATATGTGATGTATCCTGTTGTATCGTATTATGCACTGCTTTTCCA TGTACTGTCTTGCCTTGCTTAAACCAAAAATAAACTGTATCC	232
(59) R2VipUrs ORF	MIPCIDDPRPPGPPDRSSPGRRSRVACRESDAKDPGHVCPCGSEFQSKQGLAMHRYHRHMAE ENAKRLASLPSKKLAWSQEEDRALLRLATAQWKGGMAKVTLYGLLQPQFPGRSVEAIKKRLQ GLKWTAPADTPEREPLPMPPRAAQGDVEPATSPAPTADPTVFNQNTDDPTVSWRRALLTAVE GKLSEPKMGADTLRRIASQLKSGEINGEACRALMNEHAREWFPHMWKPSLVKDQTDTRCWSK RAVRRANYAKVQRLYKLQRRDAANTVLSGRWRNAYRGQAPLPKGMDQYWSKVFEAGSAKDEK AQVVPDKTLWETLAAVTPSEVTSALKGMKASAPGLDRISAEDLLAWNQPSLAAYFNLMLAAG GPPEQLA**CSRVTFIPKVDEPKTPGDYRPISIASVILRAFHKILAWRLRDTVQLSPFQHGFLQ RDGCLEASALLHTILRKSHRDRKPCAMLFLDVAKAFDTVSHQTLFRVATAMGLPPPLVDYLK HLYGRSTVQIAGQNTKCGRGVRQGDPLSPLLFIMVMEDIVRDSLPEVGFDLDGQRVDSLAYA DDLVLLAEKPSRLQEKLTLLTAALQNAGMELNARKSHGLTIEKASKRKCMVLSDTKYQCKGE GIQPMGTDDSVRYLGLHF**NWKGRLVPKHTGKLDSLLKELSAAPLKPFQRLQLLKLHVVPKFV HELVLGHAHRNTLQKLDCLTRAAVRQWLRLPTDTPLGYLHANIKDGGLGIPCFSTSIPLLQQ KRFEKIAQNPATIFQIIQRQDSFRTLGRRLDRPCRLQGAVLTSKAEAREEWGNVLSNSVDGR ELRYPDVDKASHEWLLRPSRVFPRLHMRGIQLRGGLLPTKARSARGNNRQVKDKTCRGACHQ PETLNHILQVCEVTHDGRCARHNRVLHRLAQLLRKKGVRTTLEPVIPSGSSFIKPDLIIETD TQTYVLDVSVVAGGRMKTSWDIKDEKYGRPDRVNDIRKWGAIPEASRVCQLPVIISNRGLCY QATGEGLRALGLTTREISDLCLLAISGSLKCFDLYMRGTRGVRRAPEHR	233
(59) R2VipUrs RT	CSRVTFIPKVDEPKTPGDYRPISIASVILRAFHKILAWRLRDTVQLSPFQHGFLQRDGCLEA SALLHTILRKSHRDRKPCAMLFLDVAKAFDTVSHQTLFRVATAMGLPPPLVDYLKHLYGRST VQIAGQNTKCGRGVRQGDPLSPLLFIMVMEDIVRDSLPEVGFDLDGQRVDSLAYADDLVLLA EKPSRLQEKLTLLTAALQNAGMELNARKSHGLTIEKASKRKCMVLSDTKYQCKGEGIQPMGT DDSVRYLGLHF	234
(59) R2VipUrs 5'UTR	TCTCCCGTCGCTTGACATTGGCGTTTTTTCGCAGTGCGTTTATCAATTGCGAGAGCTAGTCT TTGTCCTTGACGTGAGTTGATTGCATAGTTGCACGGGTTAGACACCTCCTTGTGGTGACCGC TGGATAAGGCCCCGGGCCTGAACCAAGATCTAACCCACCCTCCGTTGCAGAAACAGCGGAGG CTAAACTGCTGCTA	235
(59) R2VipUrs 3'UTR	GGAGCACCCTTAGGACTGCACTAGGCGGTGACCCCTCTGGTCGACACCTGGGCTAGCGACCA AGTCGCCTTAGAAGACCTGATGGGGGGGGTCTCCGTACCCTTTCTCGTCTTGTCTCTCTAAT AATACATGTGAAAT	236
(60) R2FasGig1 ORF	LLSGGGSNRSPAAPRGVGPGRSSGANIGSNANGFLCDCGRSFGSRIGLGQHRRHAHSTQVNE ERIASVISRTRRGTWSTPDDESLVRLADTTFQPGMLRKDLCATLTSYFDGRSAESIKKRLQF LHWSPPTSVHLVEASEAPPTMDSQHLPEMVDQSAIDWTQRLVSDITSGLASGPGGVGPWAGL LGIITDWRDGVLDLQSARSALETLLAATFPHTWQSTPARTVRCVPTTKRGIRRLQYAHVQKL YATRRKDCAQTVLSGDWKEAHLRRSAPPSDLMAYWKGIFSQPSHPDPRPVQRPSATQWSVLD PILASEVSEALRLMGHTAPGLDRLTVRDLLQMDQNCIAQLLNALLVLRTPTHHLSMARVTLV PKCSNPTGPGDYRPIAVTSVMLRLLHKILARRWRSQLALSPWQMAFLQRDGCSEASTVLHTI LRSVHQDCTPLASIFVDVSKAFDLVSHETIIRAATMFGLPTPMVDYLRSLYTGSMVRLGDES IACSRGVRQGDPLSPLLFIAVMEEVLSNSAPTLGVSIGGTPVDHLAYADDLVLFADNADRLR ERLVALSTALASAGMSINGDKSVGLTIRKDGKNKRLILWPTAYEVNGVRVQPIGVDQQIKYL GLHFTWKGRIQPRSTAKLESMLVNITKAPLKPHQRLDILARFALPKLYHELVLGMAHRRTIK SMDVMVRARVRAWLRLPKDTVLGFFYAKRASGGLGLPNLSTTVPMAQRTRLERLARSSLAPA RMATSASPFHQLVRQANIPIRVAASVITSKADAISAWSADLLNSNDGRGLRNFPMDRASLLW LGTGDSVPPRLFLRATWLRGGLLSTKVRRARGYVVSTEDLKCRGSCGTAESADHILQCCAAT HQVRCERHNKVVRWLASSLLKRGLRVMVEPIIPLVRSFCKPDLIVCCNDLVYVLDVCVVSGQ SAESAWRTKVDKYSATPVEESIKRTLGLPTARLKHCPVLISNRGLLFKRSGSGLRRLGLRPF VLARACQLVVEGSLKTYDVYMKGL	237
(60) R2FasGig1 RT	MARVTLVPKCSNPTGPGDYRPIAVTSVMLRLLHKILARRWRSQLALSPWQMAFLQRDGCSEA STVLHTILRSVHQDCTPLASIFVDVSKAFDLVSHETIIRAATMFGLPTPMVDYLRSLYTGSM VRLGDESIACSRGVRQGDPLSPLLFIAVMEEVLSNSAPTLGVSIGGTPVDHLAYADDLVLFA DNADRLRERLVALSTALASAGMSINGDKSVGLTIRKDGKNKRLILWPTAYEVNGVRVQPIGV DQQIKYLGLHF	238

(60) R2FasGig1 5'UTR	GGGTCATAGTTGCAGACCAGACCACCTCCTCGTGGTCCCGCTGGATAGGTCTTGTCTCGGCT TCGGCTGGGGCTTGGCTGAACCAAGTGTCTGGTCACCCCCCGTGTGGGCCAATCACCCACTC GGGGGTTTTGGGCTCGATTACAGAAACAAATACAAATACACAATCTCTCGACTCC	239
(60) R2FasGig1 3'UTR	TGAGAATGGCCTGACTCGCCAGGTGAACATGGACCGGCTCCACCTATGACGTTCTGTGCCTA AACCTGCCAGGAGGTACAGAGCCGATGCAGAGGAGACGTTGGACGCGTATATATCAATCTCC AATCAGATGATGCGTGCTCTCCCATGAAAGACTTGCTTGCTCCCCCCTATCTTTCTTCACTG AATTGTTTTCTGACCCCGTATTGCTGTATTTTCTGATTGGACAATAAATGGTATCA	240
(61) R2FasGig2 ORF	LLSGGGSNRCPPAPREVGPCRSSGANIGSNENGFQCDCGRSFGSRIGLGQHRRHAHPTQVNE ERIASVISRTRRGTWSTADDESLVRLANTTFQPGMFRKDLCATLTSSFDGRSAESIKKRLQF LHWSPPASINLVEASEASPSSPPSADTQHLPEMVDESANEWTQRLVADITSGLANGPVGAGP WADLLGIITNWRDRVFDLQSARSALETLLAATFPHTWQSTPARMVRCVPTSKRGIRRLQYAH VQKLYATRRKDCAQTVLSGDWKEAHLRRSAPPTDLLAYWKGIFSQPSHPDPRPVQRPSVAQW SVLDPILASEVSEALRLMGTTAPGLDRLTVRDLMQMDQNYLAQLLNALLVLRTPTQHLSMAR VTLVPKCSKPTGPGDYRPIAVTSIVLRLLHKILARRWRNQLALSPWQMAFLQRDGCSEASTV LHTILRSVHQDCTPLASIFVDISKAFDLVSHDTIIRAASMFGLPAPMVDYLRSLYSGSMVRL GDESIACSRGVRQGDPLSPLLFIAVMEEVLSNSAPALGVSIGGASVDHLAYADDLVLFADNA DRLRERLAALSTSLAAAGMCINGAKSVGLTIRKDGKNKRLILWPTTYEVNGVRVQPVGVDQQ IKYLGLHFTWKGRIQPRSTAKLESMLVNITKAPLKPHQRLDILARFALPKLYHELVLGMAHR RTLKSMDVMVRARVRAWLRLPKDTVLGFFYAKRASGGLGLPNLSTTVPIAQRTRLERLARSS LAPARMATSASPFHQLVRQANIPIRVGASVITSKDDAVSAWSADLFNSNDGRGLRNFPMDRA SLLWLGAGDSVPPRLFLRATWLRGGLLSTKVRRARGYVVSAEDLKCRGRCGAVESADHILQC CAATHQVRCERHNKVVRWLASSFLKRGLRVLVEPIIPLVPSFCKPDLIVCCDDLAYVLDVCV VSGQSAESAWRTKVEKYSTTPVEESIKRTLGLPTVRLKHCPVIISNRGFLFKRSGSGLRRLG LRPFVLARACQLVVEGSLKTYDVYMKGL	241
(61) R2FasGig2 RT	MARVTLVPKCSKPTGPGDYRPIAVTSIVLRLLHKILARRWRNQLALSPWQMAFLQRDGCSEA STVLHTILRSVHQDCTPLASIFVDISKAFDLVSHDTIIRAASMFGLPAPMVDYLRSLYSGSM VRLGDESIACSRGVRQGDPLSPLLFIAVMEEVLSNSAPALGVSIGGASVDHLAYADDLVLFA DNADRLRERLAALSTSLAAAGMCINGAKSVGLTIRKDGKNKRLILWPTTYEVNGVRVQPVGV DQQIKYLGLHF	242
(61) R2FasGig2 5'UTR	GGGTCATGGTTGCAGACCAGACCACCTCCTCGTGGTCCCGCTGGATAAGCCTTGTCTCGGCT TCGGCTGGGGCTTGGCCGAACCAAGTGTCTGGTCACCCCTCGTGTGGGCCAATCACCCACTC GGGGGTTTTGGGCACGATTACTGAAACTCATACAAATACACAATCTCTCGACTCC	243
(61) R2FasGig2 3'UTR	TAAGAAAGGCCTGACTCACCAGATGAACATGGACCAACTTCACCTATGACGTTCTGTGCCAA AACCTGCCAGGAGGTACAGAGCCGATGCAGAGGAGACGTTGGACGCGTATATAACAATCTCC AATCAGACGTTGTGTGCTCTACCATGAAAGACTTGCTTGCTCCCCCCATCTTTCTCTCTTGA ATTGTTTTCTGCCCCCATATTGCTGTAAATTCTGATTGGACAATAAATGGTATCA	244
(62) R2FasGig3 ORF	LESPDSSRIDANRRLADSEEHGMNTCKCRWCSGSFRTRRGLTQHMRHAHPNELNAARLGSLR PKRKRWSKEEDTRLYELVDYRLGYTSIAALCREIADHFPGRSVEAIEARVAGLRRPGGNWGG MMERDQDNIVTLQVDEEDEHAETKRWEAKVLEAISETANGAEPERVGVLQWLARRFCDGRIS REAAAGELEAIVASWFPHTLKPRPRSEPPNAGQLSKKRVRRLQYAVLQRMYRKRRKDAVDMV LTGDWRSSHMPKRRTVPGLTAYWKEVFEVESAVDNRPTPRPASSLWAVIDPIQEGEVRKAIR EIGKSSPGLDGLTAADVGKLDNRLVLALLNVLLLLRSPPEALTKARITLIPKVVNPEKPAEY RPIAIASVITRILHKVLASRWSRSHSFADLQVAFQKRDGCCEASEILQAVLRRVHEKTKGLA AAFVDISKAFDTVATSTIIRNAVAQGLPPPLVEYIRHMYDTSAVLIEQETVRPGRGVRQGDP LSPLLFLSVMDEMLKSSLPGVGFDQGGRLVDALAYADDLVLFAEDESRLRDKLEALNSALQA AGMSLNASKSCTLTIVPDGKRKTVVLAPTQYSVGGTRLRTMNVEDRVKYLGLEFNWKGKVRV RHTGILERYLIEVTRAPLKPYQRLTVLKQHLVPKLLHQLVLGAAHRNTLASMDRMIRQFVRG WLRLPADVPVAYLYTKIPDGGLGLPSLATLVPLAQRKRTEKVASSSAATAVLFKDTPCFKQF VARANMGIRIGGEAVCSKAEASQAWSEQLWKSHDGKALRGESVDPVSHDWVRDPTRLPGKVF LRGIKLRGGLLPTRTRASRGGRTPAHGVMCKGGCGHPESIDHILQSCQLTHNVRCERHNRVL RKINKVLNRKCEKVWLEPIVPCENSFRKPDLVVRERGVVYVVDVCVTTSARMEETRRLKVQK YGTESVTGAVRDLAGVVESARVVQAPVVITNRGLLSKESGSSLRNLGFTRSELSDLAMTVIL GSLKGYDLYMRGTRRY	245
(62) R2FasGig3 RT	KARITLIPKVVNPEKPAEYRPIAIASVITRILHKVLASRWSRSHSFADLQVAFQKRDGCCEA SEILQAVLRRVHEKTKGLAAAFVDISKAFDTVATSTIIRNAVAQGLPPPLVEYIRHMYDTSA VLIEQETVRPGRGVRQGDPLSPLLFLSVMDEMLKSSLPGVGFDQGGRLVDALAYADDLVLFA EDESRLRDKLEALNSALQAAGMSLNASKSCTLTIVPDGKRKTVVLAPTQYSVGGTRLRTMNV EDRVKYLGLEF	246
(62) R2FasGig3 5'UTR	CTGGGTCGGGCTAATTACCCCTCCTTGGTTCATGTCGCCTATCAATTCCGTTAAACAATAAA AGAATATTTTTATCTCAAAAC	247
(62) R2FasGig3 3'UTR	TGAGAGATCTGGGGCGGAACGTCACGTGGGTCTGTGAAGTTTGCTGTACCATGGCTGGTACC AGCCTCTGTTGAAAGAGTGCGCTGTTGGGGACAGCGGCGCACTTTCTGCTTCACGCTGGCAA	248

	CCGCATGAGTCCACCTATGGATCGGCCCACGTGTACTAACTTAAAGCGCGTCACACGAAAAT GCTCACTACGGGTTGAAGTCTCTTGCTCTGTAACTGTGCGAAACCGTCGGTAAGCAGCTTTG CTGCCTCCAGATGACGCGTAGAGTGTACCCACTGGAAAGTATCGCGAGTGCTTGCCGGATAG ATGTAGACCGGCTGGAATGATGTTTATTGAAGCCTAATGTCATCTGGAAAACTTCCCGTAGA C	
(63) R2FasGig4 ORF	MSTHNANGAASATPMASSSQPMEYICEICGKSCRSKAGLTLHKKVHPPPTHPPLPPTPPTPK AFQCVLCSDSFHTKMGLTQHKRHRHLQEYNGDKRQQIVTSARSRRWTTSEEAALIALANDLA ATHSEKQELCAALSTQFPNRSAESIKKRLQHLDWQYPDDVLPDSTSSDDDTNTSNLSSDSIS MLQPLPNSPTDSEPVAPQPTTSLCATSTALLEAACSLLRSTPHDVFGSRELLSIVEAGLTHS MDATALRCRLEAHAEANIPNLWNPSRPRSSRLPTRPLSRRPRKPPRPKILARRAQYAHLQCL FAQRKKDAAREVLSGRWRSAHENDSSWSPDFTRFWSEKLSLPPLNDPRPHRQLSHPPSLILH PITIPELQDALRDMRGSAPGIDRVTPETLLRSPLSVFTAYINILLALAHVPRRLNTARITFV PKSDNPTSPSDYRPISVSPVLLRSLHKVLQRRWTSLFPVDGLQFAFLKRDGAFEATALLHTL LRYAHTSFHSLAFASLDLSKAFDSISHDSILRSAATFGAPPILLDYVRSSYATASCVLPDGS TITPRRGVRQGDPLSPLLFIMAIDEILSLSHPQTSFLTPSGPVDAIAYADDLILFADSADGL QKKLSDCSVACGMSGLVFNTQKSFSASIITSPKQKISALDETTLFVNGHPIRTLSTSDTFSY LGTSFTYRGKAGVDYSNTLRIMLQDVISAPLRPFQRLYILRSHLISRLHHTLCLGVIHKKTL KRLDLQVRHCTRKFLRLPKDTPTAYFHARCSDGGLGIPHLSSQIPLIRRKRLERLLSSTAPL LRWAGTCPAASSSQAISHVPIAVQGAYIQSKEQAADAWKNSLYGTLDGKHISCQQFAAPSDL WITSPTRMFPRDFIHNIKLRCHLLSTKTRRCRGLRSADDILCRGGCGQPESLSHILQSCSIT HDSRCRRHDDVVNLLLKRLMRMGATCYNEPRIPLETTFCKPDVIVVKGNTAFVCDIAVCDPS LIQQTLTNKVAKYSTPVVNTAIQSFLSASGHATPNIVHKPIVITYNGFVHRSSASTLRQFHL SPLDILDLVSTTMRGSTKTYLRYMRGPN	249
(63) R2FasGig4 RT	IDRVTPETLLRSPLSVFTAYINILLALAHVPRRLNTARITFVPKSDNPTSPSDYRPISVSPV LLRSLHKVLQRRWTSLFPVDGLQFAFLKRDGAFEATALLHTLLRYAHTSFHSLAFASLDLSK AFDSISHDSILRSAATFGAPPILLDYVRSSYATASCVLPDGSTITPRRGVRQGDPLSPLLFI MAIDEILSLSHPQTSFLTPSGPVDAIAYADDLILFADSADGLQKKLSDCSVACGMSGLVFNT QKSFSASIITSPKQKISALDETTLFVNGHPIRTLSTSDTFSYLGTSF	250
(63) R2FasGig4 5'UTR	TTCCTTTCACTGCTAATTTTCGTGAAATTCCGTCTATATGAGTCACTCACTATGATGTCGCT CGCTACCTCTGGATTTGTTTTTTTTTTGTTTTTGTTTTGTTTGTTTTGTTGCTGTTTGTTTC TTTGCTCTTTTTAACTCACACACACATGTGCTGGTCCTCGCACCTATCACACACACACTACC CTATGCCAGTTACCCGCCAGTCTAAGGAATATAGACGCCGGCCGTAACTTGAACACCCTGTG ATTACCCCATATGCTTCGGGTTGGTCTTCGTGAAGAAGCCGAAGCTAGGGTCTGCTTCCAAA GCTGCCCTACACACCTGTGTTGGTGCCGTCCGTGATACTCACGTCCGCCACCTTTTTTTCTT TCGCTGGCC	251
(63) R2FasGig4 3'UTR	TGACACGACATCCCAACCCATCTCAACACCTCAATGTTCACACACTTGCACTTTTCCACCGA CTCGCACAACCATTCCGCGAGCGCACATATATATTCCGATTCGGTGACAAAATTACTGTACT GTGTACAGCGCTTGTCAAACGCAATGTCAAAACCGCTTTATTTGACCTATGTTTGACAAAAA TATATGTTTC	252
(64) R2FasHep ORF	LVFVKKPKLGSASKAALHTCVGAVRDTHVRHLFSLAMSTHNANGAASATPMASSSQPMEYIC EICGKSCRSKAGLTLHKKVHPPPTHPPLPPTPPTPKAFQCVLCSDSFHTKMGLTQHKRHRHL QEYNGDKRQQIVTSARSRRWTTSEEAALIALANDLSATHSEKQELCAALSTQFPNRSAESIK KRLQHLDWQYPDDVLPDSTSSDDDTNTSNLSSDSISMLQPLPNSPTDSEPVAHQPTASLSAT STALLEAACSLLRSTPHDVFGSSELLSIVEAGLTHSMDAAALRCRLEAHAAANIPNLWNPSR PRSSRLPTRPPSRHPRKPPRPKVLARRAQYAHLQCLFAQRKKDAAREVLSGRWRSAHANDSS WSPDFTRFWSEKLSLPPLNDPRPHRQLSHPPSLILHPITIPELQDALRDMRGSAPGIDRVTP ETLLRSPLSVFTAYINILLALAHVPRRLNTARITFVPKSDNPTSPSDYRPISVSPVLLRSLH KVLQRRWTSLFPVDGLQFAFLKRDGAFEATALLHTLLRYAHTSFHSLAFASLDLSKAFDSIS HDSILRSAATFGAPPILLDYVRSSYATASCVLPDGSVITPRRGVRQGDPLSPLLFIMAIDEI LSLSHPQTSFLTPSGPVDAIAYADDLILFADSADGLQKKLSDCFVACGMSGLMFNTQKSFSA SIITSPKQKISALDNTTLFVNGHPIRTLSTSDTFSYLGTSFTYRGKAGVDYSNTLRTMLQDV ISAPLRPFQRLYILRSHLISRLHHTLCLGVIHKKTLKRLDLQVRLCTRKFLRLPKDTPTAYF HARCSDGGLGIPHLSSQIPLIRRKRLERLLSSTAPLLRWAGTCPAASSSQAISHAPITVQGA YIQSKEQAADAWKNSLYGTLDGKHISCQQFAAPSDLWITSPTRMFPRDFIHNVKLRCHLLST KTRRSRGLRSADDILCRGGCGQPESLSHILQSCSITHDSRCRRHDDVVNLLLKRLMRMGATC YNEPRIPLETTFCKPDVIVVKGNTAFVCDIAVCDPSLIQQTLTNKVAKYSTPVVNTAIQSFL SASGHATPNIVHRPIVITYNGFVHRSSASTLRQFHLSPLDVLDLVSTTMRGSTKTYLRYMRG PN	253
(64) R2FasHep RT	IDRVTPETLLRSPLSVFTAYINILLALAHVPRRLNTARITFVPKSDNPTSPSDYRPISVSPV LLRSLHKVLQRRWTSLFPVDGLQFAFLKRDGAFEATALLHTLLRYAHTSFHSLAFASLDLSK AFDSISHDSILRSAATFGAPPILLDYVRSSYATASCVLPDGSVITPRRGVRQGDPLSPLLFI MAIDEILSLSHPQTSFLTPSGPVDAIAYADDLILFADSADGLQKKLSDCFVACGMSGLMFNT QKSFSASIITSPKQKISALDNTTLFVNGHPIRTLSTSDTFSYLGTSF	254

(64) R2FasHep 5'UTR	TTATTAAAATATCTTCGCCTCGAACGTTTACTGTGCGCTTAGCCTCATTTTGAGACTATTTG TGCAGTTTTCCATACTGTATTTCATTCCGTCGACCCTCTCATCCCACGGCCGAACCTCAACG AGGCTGCACAGGCCATTCTTCAGTACTTTCACCACTTTTTCTCCAAATTGTCCACTCTTGTG CCTCCGACTGGTCACTTTCAGTTTCTTTATCACTAATTCCGCCTAGAAACGGTCTTTTTGAC ATAACTTATATATATAAGTTTTGTGCAATTTCCATTTTTGTCCCTTTTTTCACCCTTGACGA GGGCACATCACGCCACTGCTCCTCCTTGCTAAACGCGGGAGTATTATTAAAATATCTTCGCC TCGAACGTTTACTGTGCGCTTAGCCTCATCTTGAGACTATTTGTGTAGTTTTCTATCGGATA CTGCATACAGGCATTTCCTTTCACTGCTAATTTTCGTGAAATTCCGTCTATATGAGTCACTC ACTATGATGTCGCTCGCTACCTCTGGATTTGTTTTTTTTTTTGTTTGTTTGTTTTGTTGCTG TTTGTTTCTTTGCTCTTTTTAACTCACACACACATGTGCTGGTCCTCGCACCTATCACACAC ACACAACCCTGTGCCAGTTACTCGCCAGTCTAAGGAATATAGACGCCGGCCGTAACTTGAAC ACCCTGTGATTACCCCATATGCTTCGGG	255
(64) R2FasHep 3'UTR	TAACACGACATCCCAACTAATCCCAACACTTCCAAGTCCACACATTTGCACTTTTCCATTGA CTCGCACAACCATTCCGCGAGCGCACAAATATATTCCGTTTTGGTGACAAAATTACTGTACT GTGTACAGCGCTTGTCAAACGCTATGTCAAAACCGCTTTATTTGATTATGTTTGACAAAAAT ATATGTTA	256
(65) R2OpiViv1 ORF	MGWHANDGNAVCNICGLSFRTVRGLGQHKRHRHPIELNDERTKLLPRRKGEWSEYESMKLAR LADEMASSFQTKAELYAALAAKFAGRSAEAIKKRLIKIGWSGITKYKTETGPKRGPESLPKI TEGGIDGLSDDDSDCTDWRNALLVCIADGIRQAEKHPDARALGDIAEGMLSGEMSVDQARMR LASVVDRMFPEVWTPHRGAMRRHVPKTKKAQRRATFARLQSLFKRKRKEAAKCVLDGQWQEL WKCDETKPEGTVDYWKGVLGAEAHLDARPVPTAVQKWSILSPITSAEVTGALKESVGSSPGM DKKDARELLTWNPIILAQLLNLFLVTELPPRTLSRARISLIPKTVSPASPGDYRPIAISSVV LRLLHKLLFRRWRKVVKLESIQFAFQERDGCLEASSLLNAILHSVHDTVRPIAAAFLDISKA FDRVSHDTILRAAQAHGAPMPLVRYLRRLYSDGVVNLNGEDIKSKRGVRQGDPLSPLLFVMV MEEVVKQCCPEFGYNLEGTRIGALAYADDLVLLAPSRFELEEKLNGLTSALAAAGMELNSAK SAVMIVEKSGKHKTLAIVPGSLQVGSEAIKCLGPTDAVKYLGLRFNWKGRLRTKAGTLFQEM LSNISRAPLKPYQKFEMLKEFATPRLQYELVLGSAHRNTLKALDVAARHAVRAWLRLPKDTP LGFFYAKVKDGGLGLTSFATTIPLLQRKKLQRIASSPDPVVRGLLKAKNVLSEQKLAYIPVR AKDTLVSSADEVSQAWRVDLKSSMDGRELVTDSVDRASFDWLANPRTVLPRLFIRGVQLRAG VLGCKARNARGRSVASGDLLCRGRCGVAETLSHILQRCQVTHDARCARHNRLVHNLAGRLSR RGYEVMTEPIIPVGRTFCKPDLVVRHEEHAYVLDVTVTRGQRLAESWNLKTTKYGLKEVETY IRSNLFGRTLGDVRHLPVVFTDRGLLCERSGKGLRSLGLTTKELSVLCLLAVRGSLACYDVY MRGT	257
(65) R2OpiViv1 RT	RARISLIPKTVSPASPGDYRPIAISSVVLRLLHKLLFRRWRKVVKLESIQFAFQERDGCLEA SSLLNAILHSVHDTVRPIAAAFLDISKAFDRVSHDTILRAAQAHGAPMPLVRYLRRLYSDGV VNLNGEDIKSKRGVRQGDPLSPLLFVMVMEEVVKQCCPEFGYNLEGTRIGALAYADDLVLLA PSRFELEEKLNGLTSALAAAGMELNSAKSAVMIVEKSGKHKTLAIVPGSLQVGSEAIKCLGP TDAVKYLGLRF	258
(65) R2OpiViv1 5'UTR	GGGAGCCATGGTTGCATGGTCTGTTTACCTCCTCGTGGTTCCCGCCGGGGTCAGAAGGGTGA GCCCTGAAAAGCGGTGATCCCTTCTCAGCCAAATTTCAGACCACCCGTCAGGGGCCGACCAC CCCTGTCGGGGTTATCCACTTGGTTACGCTACCTTCAATCGATGAAAAATCCTCCGGTACAT CAACGAGTGCCGGTTACCAAGTGGAGGATCTTCACAGCCGGACTGATCCACCGGTGAGGTTA GAG	259
(65) R2OpiViv1 3'UTR	TAACCGGTCCTGAAAGACCCCTGTCTGGTCGGTGGATGAACGCTGCGCGTGCGCTCCCTTCT TCTCTCTATCCTCATGCAGGAGATCTTTCCGACCTATGACGGTGGACCGCGGAAACCGAAAG GACGCGATTCCCTGATGGCTGGAAGACCCTCTGCGCAAGCCTGTTGGCCATGGATCAATGTC AATTAAAGACATTGATTCAAAATATGTTGTTATC	260
(66) R2OpiViv2 ORF	MKSFDPFSAKLQTTRQGPTTLVGFFHLVTKPPKSGKSSGTSTSTGTQVEDLHSRTDPPVSVE NRVHVEVGNVVCDICGLSYRTARGLGQHKRHRHPIELNDERLRLLPRRKGEWSDFESEKLVQ LADEMASTIPKKSELYAALAAKFAGRSAEAIKKRLIKMQWSGITKYKNGAERRPNPLPKTPE CERTTDDEEVTDGRLAENSDCANWRNSMLVCIADGARQAGQQPEARALGEIAERMLSGKLNV VQARERLTSMVERMFPEVWTPHKGAMRKAIPKSKKAQRRVTYARLQSLYKRKRKEAAKCVLN GQWKELWKCVEAKPKGTMEYWKGVLGAEAHLDARPVPTTVQEWSILSPITSAEVTGALKESV GSSPGMDKKDARELLTWNPSVLAQLLNLFLVTELPPKTLSRARISLIPKTVSPMSPGDYRPI AISSVVLRLLHKLLFRRWRRVVKLENIQFAFQERDGCLEASSLLNAILHSVHDTVRPIAAAF LDISKAFDRVSQDTILRVAQAHGAPMPLVRYLRCLYSDGVVNLNGEDVKSKRGVRQGDPLSP LLFMMVMDEVVKQCRPELGYNLGSARVGALAYADDLVLLAPSRSELEEKLKGLAPALAAAGM DLNGAKSAAMIVEKSGKHKTLAIVPGSLQVGSAEIECLGPTDTVKYLGLRFNWKGRVRTNAG NIFQEMLSNISRAPLKPYQKLEILKEFATPRLQYELVLGSAHRNTLKALDVAARHAVRAWLR LPKDTPLGFFYAKVKEGGLGLTSFATTIPLLQRKKFQRIASSPDPVVRGLLKAKNVVTEQRL AYIPVRAKETLVSSADEASQAWRADLKSSVDGRELATDSVDCASFDWLANPGTVLPRLFIRG IQLRAGVLGSKVRNARGRSVASDELLCRGRCGVAETLSHILQRCQVTHDARCARHNRLVQNL AGRLSRRGYQVMTEPIIPVGRTFCKPDLVIRHEENAYVLDVTVARGQRLAESWALKTAKYGL	261

	KEVETYIQSNLFGRPLGDVRHLPVVFTDRGLLCKRSGRGLRSLGLTTKELSVLCLLAIRGSL TCYDVYMRGT	
(66) R2OpiViv2 RT	RARISLIPKTVSPMSPGDYRPIAISSVVLRLLHKLLFRRWRRVVKLENIQFAFQERDGCLEA SSLLNAILHSVHDTVRPIAAAFLDISKAFDRVSQDTILRVAQAHGAPMPLVRYLRCLYSDGV VNLNGEDVKSKRGVRQGDPLSPLLFMMVMDEVVKQCRPELGYNLGSARVGALAYADDLVLLA PSRSELEEKLKGLAPALAAAGMDLNGAKSAAMIVEKSGKHKTLAIVPGSLQVGSAEIECLGP TDTVKYLGLRF	262
(66) R2OpiViv2 5'UTR	CCGGGAGCCATGGTTGCATGGTCTGTTTACCTCCTCGTGGTTCCCGCCGGGGTCAGAAGGGT GAGCC	263
(66) R2OpiViv2 3'UTR	TAACCGGTCCTGTAGGACACCTCTAGGGTCCTTATCTGTTTTTCAGATCCTGCCGGTTGGTT GATATTTTATCGTTTTTGACGAACGGATACTGCGTGTGCTCTCCCTTTTTTCTCTATCCTCA CGCAGGAGATCTTTCCGACCAATGACGGTGGACCGCGGAAACCGGATGGACGCGATTCCCTG ATGGCTGGGAAGAACCTCTGTGCAAGCCTAATGGCTCTGAATCGATGTCAACGACGCCTGAC CTTAGTATGGATCTTTCAAGCTGGAACTATTATTCAACCCGAATGCTTACTTTGAATGTGGC TTGACCGGTCGATTTTGCGACCATGTGAATGACTGAAACAAGTGGAGCCTCCTTTCTTTCCC GAATGCTTCTTGGCGTACTGAATGCTGGACGTGCTAAGAACGAGCCTGCTGGAGGCTTTCGT GCGACAGGAGGAACGACAGACTGTGTTTCCATGGGAGGTCTGAGATATGGCCCTGTTGCGGC GGAGCTGGAGTGCTCAACACGGGGAAGGGCTGCACAGGCGGATTCCTATCCGCCACGACGGC CTAAGGGGGTCCAGAAAATAGGGACCTACCTCCAGGTGTGATTAGACGCCTGGAAAAACTTT CCCAAAAAGGGGAGACCAATTCCGATGGTGTTACACTGTCGAAGCCCGTGAGCACGCAAATT CCATGCCTGCTAATCCATGGGTCCAGACACTTGGACCAACCTTGGTTCACGGGAAGTCGGGA GAACTGTCCATGGACTGCCGAAGGCGTTTTATGAACGAATGGAACTTCAAACTCCCCACATA TACGCACGTCCCCCCCCCTCTTTAATAAAGAGTTGACGCTTGTAAAAGACATTGATTCAAAA TATATTGTTATC	264
(67) R2OpiViv3 ORF	MVMKPISDITPGTTTSAGHQVEDLHSRTDPPVSTELGLHDSNRGVRRVVCDLCGLSYKSARG LGQHKRHRHPIALNEERLKLLPRRRGEWSEFESAKLVQLADELASSIPKKAELYAALAAKFA GRSAEAVKKRLIKMQWYGISKYRAKAEQRPELLPKTPGCESTTDDEGVADGRFVENVDYVNW RRSMLECIADGVRQAKKQPDAQALSDIAKGMLSGKLDVDQARIHLTSVVDQMFPTVWAPHRG VRRKGIPKSNKAQRKVTYARLQSLYKRKKKEAAKCVLDGQWKELWKCVEAKPEGMMDYWKGV LGAEAHLDARPAPTTAQEWSILSPITSAEVTGALRESVGSSPGMDKKDARELLTWNPSVLAQ LLNLCLVTELPPKTLSRARISLIPKTVSPASPGDYRPIAISSVVLRLLHKLLFRRWRKVVKL EGIQFAFQERDGCLEASSLLNAILHSVHETVRPIAVAFLDISKAFDRVSHDTILRVAQAHGA PMPLVHYLRRLYSDGIVNLNGEDIRSKRGVRQGDPLSPLLFMMVMEEVVKQCRPELGYDLEG TRVGALAYADDLVLLAPSRLELEEKLDRLASALTAAGMELNGAKSAVMIVERSGKHKTLAIV PGSLQVGSEVVKCLGPTDTVKYLGLRFNWKGRLKTNIGRVFQDMLVNISKAPLKPHQRLEVL KEFATSRLQYELVLGSAHRNTLKALDVAARHAVRAWLRLPKDTPLGFFYAKVKDGGLGLTSF ATAIPLLQRKKFQRLASSPDLVVRGLLKAKNVLSEQRLAYVPVRVKETLVSSADEVSQAWRG VLKSSVDGRELVTESVDHASFAWLRNPETVLPRLFIRGIQLRAGVLGSKVRNARGRSGASDD LLCRGRCGMAETLGHILQRCQITHDARCARHNRLVHNLAGRLSRRGYEVMTEPIIPVGGTFC KPDLVVRHGEQAYVIDVTVARGQRLAESWSLKTEKYGLREVETFIRSNLFGRPLGDVKHLPV VFTDRGLLSERSGKGLRSLGLSTKELSILCLLVVRGSLACYDVYMRGT	265
(67) R2OpiViv3 RT	RARISLIPKTVSPASPGDYRPIAISSVVLRLLHKLLFRRWRKVVKLEGIQFAFQERDGCLEA SSLLNAILHSVHETVRPIAVAFLDISKAFDRVSHDTILRVAQAHGAPMPLVHYLRRLYSDGI VNLNGEDIRSKRGVRQGDPLSPLLFMMVMEEVVKQCRPELGYDLEGTRVGALAYADDLVLLA PSRLELEEKLDRLASALTAAGMELNGAKSAVMIVERSGKHKTLAIVPGSLQVGSEVVKCLGP TDTVKYLGLRF	266
(67) R2OpiViv3 5'UTR	CCGGGGGGCATGGTTACATGGTCTGAATACCTCCTCGTGGTTCCCGCCGGGGTCAGAAGGGT GAGCCCTGAAAAGCGTTGATCCCTTCTCAGCTAAACTTCAGACCACCCGTCAGGGGCCGACC ACCCTTGTCGGGATTTCCAC	267
(67) R2OpiViv3 3'UTR	TGATCGGACCTATAAGACCCTTTAGGTTATTTTCTTGAGTGTTATTTGGTTGTTTATGTTGG TTTTATTGGTAGACAGATACTGCGCGTGGTCTCTCTTCTCTGTGCTCATGCAGGTGATCGCT CCGACCTATGACGGTGGGCCGCGGAAACCGAATGGACGTGGTTCCCTGATGGCTGGAAGAAC CTCTATGCAAGCCGTACGGCCCTGGACCAATGTCAACGATGCTGGCTCTTTGTCTGGATCTT ATATGACGGATTTCCTACTTGACCTGGATGCTTTCCACCTTTTGGCTGTTCTTGCGATCCGA AGGCTGGACGTGCAAAGGACGAGTTCGCCGGAGGCGCTCGTGTGACAGGGGGAACGACGGAC TGTGTTTCCTCGGGAGGTCTGAGGTATGGCCCTGATAATGCGGAGCTGGAGTGCTTAACACG GGAAAAGGCTGCATAGGCGGATTCCTATCTGCCACGATGGCCAAGGGGGGTCCCCAAATAAG GACCTACCTTCAGGTGTGATTAGACACCTGACCAAATTTCCCTGAAAAGGGGATAATTCCGA CGGTGTCTACTGTCGAAGCCCGTAAGCACGCGAATTCCATGCCTGCTGTTAGACGGGTTCAG ACCCTTGGACCAACCTTGGTTCACGGGAAGTTGGGAGAACTGTCCACGGACTGCCGAAGGCC TTGGAGAACGGACGGAACTCTGCTATATTCCAATACACACGCACGTCCCCCCCCTCTTTATT AAGAGTTGAAGATTGTACTGACATTGATCCAAAATAAATTGTTAAC	268

(68) R2CalDau1 ORF	LGSTAPWGVFSLVTDTYNKSPLGSSLPGHHQSDLGPAAPVEVVSFECSQPLKRASSSSPELL RPVGKRARTTGNSEGTTSGHVVERRMVQSTIGWGERVNANGGVKRVRRPEPETQGASAGEVS NSQWDDNTEGLELIMVEMVEKVEKVRPSCLKKVVAGPSNRRVRFEGVRSVDLDETPEERRAR GATLRPGSKKVLEHGEGGEPVDPSQYQQASGGEVRAFPCQFCALSFTTKAGLGQHMRQRHPL ELNAERLRLISKRPQWNDADDAMLIDKANSLYKSFDTKLAMCKFLKQVFKGRSVEAIKLRLK TLKWSKPTRAIVGGESVPSQSQPSIPSSTSLESGLTVTKSDERANWFVEFAKSLATRMGLAL ERCRDDEVDRLQGRALLDIAEKIRDGQNSLNTARKRISEHAARCFPHTRQPTDERKYGRSSV PSSNRKLKRANYAAIQALYAKSKKDAAEVVLSGKWPVAYKGRQKPPPGHLRYWNKLFSQEGV TDSRPVVERSNTSWALVDPITVNELNTALKGLGLASAAGLDRLTVGQVRSMNRNLLVEYLNL LLLTEAAPSQLSKARVTLVPKPGEGGEPVDPSGYRPIAVTSVLLRVLHKILAQRWTKSLLLT SFQVGFLPRDGCLEANLLLQTALRLSHERKIPLAAAFIDVSKAFDTVSHDSLLRAAQSFGAP PPLLNYLRALYTRSEVTVDGRRSVKCQRGVRQGDPLSPLLFIMVMEEVMEQSLPSLGVKLPN FNAHAVAYADDLVVFSSNGDCLLERLNHLEAALARSGMSLNTKKSRTLVIAACGKSKVTALL PTEFRVKDGQLRPIGVNDKVRYLGLEFTWKGQEKVKHTTKLEGFLQNLTKAPLKPAQRMELL RDFAIPSLLYELVLGDAHLNTLKKLDRMVRSKIRTWLRMPKDVPTAFIHANVWDGGLNVERL SVTVPLMRRKRLEKVLTRDTNVTTVVRTCSAFQNKLRQAARPVKVNGEVVMNRREYQGATKQ ALLSTLDGKELGMPDVDPESFRWVQDPSNIFPRLYIRGLQLRAGVLATKVRRGRGMRSGNRP TIAAREFLCRGGCGYPESLVHILQKCQVTHEARCARHNRVVRFLSAKFKRAGYEVLVEPIIP VTGSFRKPDLVVKLGNKAYVIDVSVVSGHSFDPSWGIKVRHYNTYDVHAGLYALFGNEIVLK HVPFLISYRGLVCKRSTAPWRMNVLTGRDIRDLSLLAVRGSLKAYDVYMRGT	269
(68) R2CalDau1 RT	KARVTLVPKPGEGGEPVDPSGYRPIAVTSVLLRVLHKILAQRWTKSLLLTSFQVGFLPRDGC LEANLLLQTALRLSHERKIPLAAAFIDVSKAFDTVSHDSLLRAAQSFGAPPPLLNYLRALYT RSEVTVDGRRSVKCQRGVRQGDPLSPLLFIMVMEEVMEQSLPSLGVKLPNFNAHAVAYADDL VVFSSNGDCLLERLNHLEAALARSGMSLNTKKSRTLVIAACGKSKVTALLPTEFRVKDGQLR PIGVNDKVRYLGLEF	270
(68) R2CalDau1 5'UTR	TGTCTGGCTATTCTAGGATGGCCAGAGGAGCTTTCGGGTTCCAGCCTTATGTGGGCCTTTCG GGGCTCTGCTTAGTGCTCTACACGCACCATGTGGCGTGCTTTGCAGACTCATAGTTACGGAA ATACCAGGGCCCCCTATTGGCCCCGCCGGGTCAGCTGTCTCTTCCTTCGGGTAAATGACAGT GAGCTAACGCTGGTACGGTTTCCATCTTCGGATGCGCGGTTCAGTCCCGTGTTGAAAAGCTA AATACTGAACTGGGGGAGCCTGCCATCACATCCCTCGTAAGCTGGATTGGTATCCGCGTTGC CCGATTTCGGTCGGGGATCCAGCGCCCTCTGGCAACGGAGGGTAATCGGCATACCAGGCGTG GGAGTAGGCATGTCCCGTATATATATGGGCCTGCTTGCTGAGGATTCCTAACCTGCCCTAGC TTCTGCTTGGGTGGGGACCGAGTCCTGAGGTTGGCTACGTTGGGAGGGTTCGCTCCTTCCCC CCGTCCATGAGCCCTGATGGGATTTCCCAAAAGGGTGAGGCCGACCGACCGATTCGTCGGTC TATCTGTCACAGAGGGCGCCTAGTTCGGGTGAGCTCTAGGGGTGACAGGGGGCCCCGGTGGG GCTCAATTGCTTCTCGCAAAACCTCTTTCCCACTGTTCTTTGTTTCTGCTTTCTCTTACATG ATCTTGCTCCGGGTCTCGTACTTGATCGGGCTAGTCGGTGCTGGCCTTAAAACGGTGGCAAA CTCCCTCCC	271
(68) R2CalDau1 3'UTR	TAGGACGTCTGGAGGAACGCGAGGACTGGTACGAAACGAGGATTGACAGTGTTGGACGTTGG CACGACCCAAAGGACCAATCGAGACCGCGGGATGACCTGGGCTGCCTGAAAGCCCATCAGGG AAGGGTAGTCTGCGATGACGAGTCAATGGAACTTGATCGCTCTAACACCCATGTCGGGGTGC CCACTCAGTGTCGAATTGTACGGGATGAATTGGTGCTAACAATTTGAAGACGGAATTTCATT GGACGTGCATACAATGATTGATTTTATGAACTGACAGTTTGACTTCTTATGAAAGAACATTA TGTGCCTCTGCTACTCTGTTACTTTCACAAAATTTTTCTTATCCTTGTATTTATTTTATGAA CTCCGCCCGTACGATACGTACTGAAAATAAACGATAAAGG	272
(69) R2CalDau2 ORF	MHSDPLSSPSGSPDAKRPRVEGPEVWEEGFDIEAIERALEGTVSTPQGVSEVGELAVVQLPG ESPQVERTVETGSCNLNAGDGAVEGQPQAVSLSNCTICHAVFHTARGVAQHMRFKHPEMHTH PKAITTVRGQQSSGGETCPCTHCNAVFITARGLGQHMRQRHPFELNEARRKLSVNRRFWTPQ DDAALLEHANKQCRNFATRMALCKCLCQLFTGRTREAIDLRLKTLRWSKPICPVGRDCHVIP QAESQSTMAGEMVREEAYDQWFSESASGLVERLSRAAKGDDAKVDQIQGHTLLDIAEKLKAG TTSIKGARKRMQQHATLCFPHKRRPTKVMEPSQTKAQVSKRKQRRANYAAIQALYAMSQKDA AETVLSGGWKDAYKGRARVPREHFDFWAKLFEGKAVADQRPVSQKVETNWDLIRPITSEEVS QALRELRNTATGLDRVSVSEVNSINRGMLVEYLNLLLLTEAAPGQLSKARVTLVPKTAEALE PGDYRPIAVTSILLRALHKILARRWSNVLKLDSLQVGFQPRDGCMEANLVLQTALRLSHEKT TPLAAAFVDVAKAFDTVSHETIIRAAISYGAPSPLVNYLRALYSRSEVTLDGKRTIVCNRGV RQGDPLSPLLFIMVMEEVLELALPRARFQVCDFSAHALAYADDLVVFANNADRLREKLVALE QALSLAGMAVNAGKSHTLTIDCNTKLKQCTLLPTSYTIHGEALRPLGVNDRVKYLGLWFTWK SRERVKHTGKLEDLLQNISRAPLKPQQRVCMLRDYALPRLQYELILGDAHSNTLKRMDLMVR RNVKAWLRFPRDTPSPLFYVGIGHGGLGIPQLSAMIPLQRRERLQKVLARDTAINVMIRPLG AFKAKLKVANRNVMVEKQAVTSKSEYHRAMETCLFGKVDCRELGVPDVDPDSFRWIRDPRGV FPRLYIRGLQLRAGVLGTKVRRARQKGIRSLARTMQYDTTIRCRGGCGQPESIGHILQNCSV THEARCARHNRIVHSVAKRLRQSKYQVTVEPIIPVTGSFRKPDIVAVKDGIAHVIDIAVVTG IGFEYPWRLKVEKYSTLEVHAGIKSLFGSNVQILHSPLLVTNRGLIYRRSAAMLRSMGLSGR	273

	DIGDICLLAIRGSLKEYDVYMKGT	
(69) R2CalDau2 RT	RVTLVPKTAEALEPGDYRPIAVTSILLRALHKILARRWSNVLKLDSLQVGFQPRDGCMEANL VLQTALRLSHEKTTPLAAAFVDVAKAFDTVSHETIIRAAISYGAPSPLVNYLRALYSRSEVT LDGKRTIVCNRGVRQGDPLSPLLFIMVMEEVLELALPRARFQVCDFSAHALAYADDLVVFAN NADRLREKLVALEQALSLAGMAVNAGKSHTLTIDCNTKLKQCTLLPTSYTIHGEALRPLGVN DRVKYLGLWF	274
(69) R2CalDau2 5'UTR	TACCTGGCCACCTTTGGGTGGTCAGAGGGGCCTTCGGGTTCCAGCCTTATGTGGACCTTTCG GGGTCCTGCCAGATTCCGTTTGCACTTACACTGCACTCGAGCTGGTAGACTTATAGTTACAG AAATACCAGGGCCTGCTATCGGCCTAGTTGGGTAAGCCACCTTCTTTTATGAGAGATTGGTG GTGAGCTAACACTGGTACGGGTTGCTTCCTTCGGGTGGCTGGTTCAATCCCGTGTTGAAAAA CTAAATTTGAACCGGGAGGGTCACCATGCCCTCTCTGAGTCGATCTTGGTGCGTGCGTTGCC CGACTGCTGTCGGGGGATCGGCGCCTGTCCTCTGGATGGGTAATCAGCCCACTTGCACGGGG GTGAGCATTTCCTGGTAGGTCTGCTCACTGTGGGACTCAGACCTTGCGCACCTTGGTGTGTG AGGTTTGGGTTCCACTGGTCGGCTACGCAGCGGAGGCCTTGGCTTCTGATCGTCCGTGGGCC TGATCTTGTGAACATTTTCAGGTGAGGCCGACTGGCAGACGCCCAAAAAGTCTGTCTCACTG CTACTGGAGACTCCTAGGAGGTGTAAGCATAGGAGTGGCAGCACTGCCCTCGGGTAGTTATA GTCTCCAAAATGCGGTCGGATGGCCACTCTTCGGTGGGGCCAGGCGCGCACCAGGTCTCGTA CCCCGGCCGGGGCTGGGTAACCAGCCCAAATTGGTGGCGTGCTAGGGCATATGGGTCAACAC CCCGTAGGCCTGAAATAGTTGAATCGTGCCAGTCTCTCAAATTAAAATACAGTTCCCTGCTA AGGCACGGGAACAGCACGTAAATGGGCTGTGCAGGAAGCCGCCCGTGGTAGTTGTGAGCTCT GGTGCCAATCTCCAGGGAGACACAGGA	275
(69) R2CalDau2 3'UTR	TGATGGAACAAACGCGCAACCGGGCCATTGGGGCTATTCTGGACTCTTAGACGACCGGCAGA TACCGGCGGGACAATTGGCATCGTGCAGGTCCGACCATGGGATTCGTACCACCTCCAACTAT ACACACCGATGAATGGGTTATGCTGAAAGCTCATCCAGAGAAGGTATAGCGCTGATTCAGGT GACCTGCATGGCCCATAGCGCTTCGGCGCCCCTCACTGCTGGCTTGGACAGCCTGACTGGTT CGTACGCATGAAGGATCTATCTGGACTGGAACATCACAACGATTGCATTTAGACACTAAGTC ACGGAACGCCTCTATCTCATGGTCATTCTGCGCGTATAAAAATTCAAACAAGACAAAAAACC CAACCAGCTGGATTCCTTACATATTCAAAGACCAACCTTGGAGGTCGCGCGCCCGAGCGGAT GGTTCATGTCTAAGGAACGGCACGCAATTACTTGATAACAAACACTTTGTGCGGAATGACTG TACCAACTCGTTATCTTGCATTTTCCTCTTATTTTATTATATTTTATAAGTAACACTTGCTG TCATAAGCCGCTTTGAAAATATATGTTACA	276
(70) R2SchJap ORF	MSAMLNITTECHPVSTGVNADPSFSYSSTCLCCFSSFASNSLLLDHATASHSANIVCPPSES GSFQLVCMLCSSHYLSSRGFTQHLRHVHPDAYNSLKCLRLNASPLTHRSWSSEEDVCLLSRA DELSASCRLKVDLYNRLHTVFPSRSPEAIKKRLRFLSSSSESSSSAPSPSDISSAASSVSSL PVDLDAPTRSLIPTRRPARSHSSSILSFFTRTPRDVSSSASPVVLHSSPSTTIVINNASVRL HLDSPVSSHALESSCMSSLATNPLHIIDNQHLAASPPLSGVHCPSVPSSAPAAAHVDQSFDN PALGSTPVDSCPPVTSLTLLPAQLSDVAVNLMTSCPELIKPASPSPPLMDDNLLQLILPSDS SNMSPHIVPPPDVVVAPMCTDSTRELISAALRLITQNSPLMHAPTLREFLQAALVNMPDMIE IQLFLNSHAELQFPTKWRPSKPRLPPTYRANTSRKHLRRLQYGHIQTLYNRCRRDAANTVLD GRWRSPHTSSPFSIPEFETFWKTIFTTPSTPDNRPVVPVLPMCPALLDPITPDEITWALKDM RNSAPGVDRLSAQHFLNFDVPSLAGYLNMVLAFKFLPTNLSISRVTFIPKGASPQQPNDFRP ISIAPVITRCLHKILAKRWMPLFPSSKLQFAFLQRDGCFEAINLLHSLLRHAHERHSGCSIA LLDISRAFDSVSHHSILRAAHRFGAPDGLCQYLQRVYNGASSLFNTVDCAPSRGVKQGDPLS PLLFIMSLDEALESIETDSPVIVDGLPISYIAYADDLVILAPNADLLQKKLDKLASLLQRSG LIINTSKSMSIDLIAGGHSKLTALKPTVFKIDGNQLQRLNVSDHFDFLGISFDYKGRSKMDH VETLSAYLLNLTQAPLKPQQRISILRENLEPRLLYPLTIGVVHKCTLRQMDRLIRSSVRKWL RLPSDTPTSFFHSSISTGGLGIPHLSSIIPLHRRKRAAKLLLSPCPIIRWVSQSPSFSNFLR ICNLPINVHRDLIHSFDEARCSWSKQLHSTCDGRGLSMSSRNTVSHLWLRYPEHIFPRLYIN AIKLRGGLLSTKVRRSRGRQENADLLCRGRCGHHESIQHILQHCSLTHDIRCRRHNDICRLV ASRLRRNNIRFFQEPCIPTPDSFCKPDFIIIRDSIAYVLDVSVCDDANVHLSRQLKINKYGC STVVSSIYNFLNATGLRISSVRQTPLIITYRGLIDPLSTTSLRRLSFSSRDISDLCVASIQG SMRIYNTYMRGTSPQDP	277
(70) R2SchJap RT	ISRVTFIPKGASPQQPNDFRPISIAPVITRCLHKILAKRWMPLFPSSKLQFAFLQRDGCFEA INLLHSLLRHAHERHSGCSIALLDISRAFDSVSHHSILRAAHRFGAPDGLCQYLQRVYNGAS SLFNTVDCAPSRGVKQGDPLSPLLFIMSLDEALESIETDSPVIVDGLPISYIAYADDLVILA PNADLLQKKLDKLASLLQRSGLIINTSKSMSIDLIAGGHSKLTALKPTVFKIDGNQLQRLNV SDHFDFLGISF	278
(70) R2SchJap 5'UTR	CGAATTTTGCTGCGTGAGAGCTTAGAAATATTTTCTTCGCTATACTTCAGTTTTTCAAAGCG GAGATAGCCAAATTATCGATTACTTCACTACGTGGGACGTGCACTATCTCTGGATATCATTT ACTTTGCTTTCCACACACATTTAGCTGTTTCGTCATCATGACCACCAACTCGACATTCGTCG CTTCTATATTCAGCATCTTTCGGGTTCTGAATCACCCTGCAGTTTTGTCTGTTCGTTATTCA CTATTCTGTGAAGTCATGGTGAGGGCCTTTTCTGGTCTGTGACGCGGGCAACTTAACCTGTT GCTCGTCCTCATCTGCACGACTGATCCTATGTGATCTACTCCTGC	279

(70) R2SchJap 3'UTR	TGATTCGAGCATTCCATATATATTATATTACATTATGCGTCGCAGGCTCATACTGCATCTGC TCTATCAAGCACCTATTGAATTTATATGACAGACAATATTTCTCTTATGTGTTATTTGAAAT TTGTTAACTTTAAAATTATTTGCATTTCAAATAATCTTTCTCTAAACACAATTAACATTGCA AACACCCATTTGGGATTGTATATGCTATTGATGTTTAAAAATAAATATGTTAC	280
(71) R2SchSol ORF	LQTCAGSPGVSDPEVHSNDPGPSAVGERPVSTGLFASRLPPCPGCDRILRSAAGLKLHVRSA HPELLSQINQSDVTRSAAILRTTNVSDVVQCQYCVRQFQTARGCSMHTKSAHSDEYHRSRVP LASGRRVWTPQEDNLLCSRAAALLVDGLLSLRELARKLQGAFPGRSAEAIYKRLSKLEKDEL LSRNLVTVIPTPVVGSPPNSASNPPTNQPEEEESTPDYRSQLLRTAATQLEASPCQALVTEE IVSIAIDLLMGEIGIADAQQKMEAHAKKVFPHRWRPGVNRRPKMERPCTRRQLRKAQYAEVQ ARLAKRPKLAAQFVLSGDWRSASSEEPPRPAGTLDYWKAISETPSGKDQRPVPSVVKHWSLI DPILSTDVTRELRRMRGSAAGVDGLNVSEVSTWPPAAIAGYLNVILAVGQLPGQHCVSRTTL IPKVPIPQTPSQYRPISVSNVLPRLLNRVLALRWSRCCPSARFQFGFQERDGTAEATGLLHG MLRYAVTAPKSVAVAVLDVSKAFDSVNHETILRAATSGGAPAPLLSFLESTYSRQKTCISDT AVKCLRGVRQGDPLSPLLFSLALSEALTHSNRQLGMDIANTTVDSIAYADDLVLFAETPFRL QERLNGLVTGLSKAGMVLNPAKCLSFYIQALGKAKSSCLSPCSVCIGGEPLRVLGPTDAFTY LGVPFTYRGKQPIRHRPLLYKMLTEISQAPLKPPQRITLLKRHCVPKIMHDLVLGSVYRQTL KRLDIQIRQAVRAWLRLPTDTPTSFFYASLRDGGIGLPCLLTSVPLMKRKRLDAQLSSTEEA VRVASALPEALTALQQASRPVLIHQAVINSKEDAKKMWREQLYRSVDGRPLAHIGTSASVHH WLSSPDRLFPWLYLRGIQLRAGVLSTKARRSRRKRLSDVLCHGRCGQIETLPHILQCCQVTK EARIWRHNSIMKSITAKMRRGKADVLLEPHVPEGSSYCKPDVVVCMKGSIQVMDVAVAGEMN MDRVYTGKVNRYCTEEVANNLRAIFRKPSDYPVLHKPVIFSSRGGLLRKSELNLKSVGLTDF DVSDLCVLAIRGSLKAYDVYTRGSNDRRGRQAPAVRSNRML	281
(71) R2SchSol RT	QYRPISVSNVLPRLLNRVLALRWSRCCPSARFQFGFQERDGTAEATGLLHGMLRYAVTAPKS VAVAVLDVSKAFDSVNHETILRAATSGGAPAPLLSFLESTYSRQKTCISDTAVKCLRGVRQG DPLSPLLFSLALSEALTHSNRQLGMDIANTTVDSIAYADDLVLFAETPFRLQERLNGLVTGL SKAGMVLNPAKCLSFYIQALGKAKSSCLSPCSVCIGGEPLRVLGPTDAFTYLGVPF	282
(71) R2SchSol 5'UTR	ACTGACGAATTCTCTTGTGTGCTCTTCTGTGAAAAATGTGTTCGCCAATGTTCTGCTGCCCA CCTGCGCTCTGGGTGCTCCGGGGTCGTCGCCGGTCCGTCTTCGGGCCGTTCTACGCCTTCCT CGGTGAGTCATAGTGATGGCACGTTACCTCCGCGTGGTACTCCTGGGGAGGCGGATGGACCC TCTCATACTGATGTGGACTCCGTCTTC	283
(71) R2SchSol 3'UTR	TAACTCTGACGTGAAATGTATGATATCTGTACTGACATTATGGAACATACCCCTTTGTCTTG TACTCTTGCACATTATGCATAACATTGATGAGTTCTGCCCAACTGTCTTGACAGTGGGTATA ATAAATACTTCC	284
(72) R2SpiEri ORF	MMSSNEVSPPASAGVSATSGLICVHCGRAYRSAAGLKCHLRRCQPESVLSPSVSVPAEPSPR NFVCDHCGRGFSSQRGRSLHVKSAHPAAYFAERSRLAPSRTTWTTQEDALFRRCASAVLQAS GPLALRALAREVGIQFPTRSVEAVFKRLSKLDKDEPLFTPPVSLSAVGPSSSSPSPAPPAVC EPDQVEFYRTSLLHSAVKLLTDSPCQALASGDLLSVASDLLMGEAQLSDSAARLDAHAAVVF PRRWRPGVNRRPTVVARPTASHKLRRAQYAEVQARLAKNAKAGAQFILSGNWRNAPCPEAPL PPGTLEFWRELTERPSEPDQRPIQPPTAVAWEILEPITATDVTGNLRLMQGSAAGIDGLSAK DLLKWSPSAIAGYLNIILAIGHLPAQHCTCRTTLVPKVSTPTLPSEYRPIAVSNIIPRLLTR ILAHRWSRLCPTARFQFGFQERDGTAEATSLLHGMLRYAVSAPRSIAVAVLDVAKAFDSVNH GTLLRAAEANGAPPLLTTLLASTYSRATTYILDAEVRCQRGVRQGDPLSPLLFNCALSEALS YSDRQLGFDLAGVTVDSIAYADDLVLFAESPHRLQQRLDGLANGLSLAGMVLNSAKCLSFYV QALGKEKSACLCPCDVSIGGSVLRSLGPTDTFKYLGVPFSYRGKVTVGHRSALSGMLEELTR APLKPQQRITLLKRHCVPKLLHKLVLGAVHRNTLKRLDVQVRQSVRRWLKLPADTPTSFLHA PIVDGGLGIPCLSVLVPFAKRRRLDSVLSSAEPVVRAAATMPSAISGLRLAAQPVRMGHTVL ASKDDARNYWRNELYNSADGRPLAQFAKSACASHWLSSPARVFPWLYLRGIQLRAGVLSTKS RRSRGAGITDVLCRGQCGQRETLFHILQCCQVTHRARVWRHNLAMKLLASKLIARGHEVLVE PHIPEGQTFKKPDIVVCADDRLTVVDVAVADEDLMDTVYAGKTRHYSTVEVEGNLRRILGKP ADFPVSHAPAIFSSRGALSPRSEASLKLLGLSKFDLSDLCLAVIRGSLKAYDVYVRGAGNGE NQ	285
(72) R2SpiEri RT	TCRTTLVPKVSTPTLPSEYRPIAVSNIIPRLLTRILAHRWSRLCPTARFQFGFQERDGTAEA TSLLHGMLRYAVSAPRSIAVAVLDVAKAFDSVNHGTLLRAAEANGAPPLLTTLLASTYSRAT TYILDAEVRCQRGVRQGDPLSPLLFNCALSEALSYSDRQLGFDLAGVTVDSIAYADDLVLFA ESPHRLQQRLDGLANGLSLAGMVLNSAKCLSFYVQALGKEKSACLCPCDVSIGGSVLRSLGP TDTFKYLGVPF	286
(72) R2SpiEri 5'UTR	CTGGGGCCTTGCACCGTCTTGGTCCCGAAGGGTTAGAGTCTTGGCCCCACTTATTGGGTCAG GGGTGCTGCCCTGGAGCATAGTTGGTACGCCTTGCCTTGTCGTGGCACTCCCCTTGACGGGT AGGGTCCACTCGTGGGCCCGATGGCTCAGCCGTACCCGGACTAATTGGCGGCCCCACTGGTC TTGCCGGTGGTTAAACTAGCAATTCCTACCA	287
(72) R2SpiEri 3'UTR	TGAGGTTGCAGTGTGGACCCCCGGTCGTTGTATGACGCAGGGTTACTAAACCGACTAGCAGC ACCCCAGTTTAGCTGGCCCTGAGTGGTTCAGCGGCTGAGGACCGTTTGTCTGAGGAGGTACC TTGCTGATGTGCACCGGTGGAGATTGCTGGATGTTGCCCTTCTCCTTCCCCCTGGGATTACG	288

	GACTGTTGTTCTGTATCAGCCCCCATGACTGCAAACTGGAATGTGTCCTTACTGTTGTTGCA AATGCTTGCTTACCCTGGGGCATGCTCTGCCCAAATTGTACTTTCAGTTCGGGTATAATAAA TATTATC	
(73) R2DipCan ORF	MVTITDTNDTTEASASANIPGPTGAVVLLCPVCSRPYKSRSGLSNHLRSHATVGEGHACQIC AASFRSASGLTQHMRHRHPTVANQVKLERLAAKRTYWRKTDTELLEHLLQRLPCELSPETIV ENIRESFPGKTTRQLTIRVRNLLTHSQPRPVERSGATPSPPCLSAEPPWREALLQATTSLLS ERADDFPGAPELLVTLDRLLSSDADGAREALEIHANTYFGRKWQPTERRSIPVGGFTSRHAL RRAQYAHMQRLYHLRRRDAAETILSGAWETAHVSGQGIPPNSENHWSDVLSLPAVSDTRPIQ KLQPTDWSLLRPIQPQEVKAALTQMKGSAAGPDRISVESVLSWHEGYFTGYLNLLLASSYLP RVLRSSRVTLVPKTPHRPSSPTDLRPISVSSVVTRCLHKILAARWSDSITLPALQFAFQKCD GTFEATNLLDGLLRERASHNRDLALASVDVSRAFDSVARDTIFRAASSFGAPAPLLEFIKCG YSEAQAFLGPRCLTPQRGVRQGDPLSPLLFIMAVDEAVRAASPELGASLNGATVDAIAYADD LILLAPTPEQLSHKLTQLRSALRAGGMTVNERKSFILTIRGDTKSKQSAVVPFPSVLAHGEP FRSLGPEDSFEYLGVNFSWRGRAPTNHLERLSVWLEHLTRAPLKPQQRVSLLRTSLIPRLHH SLVLGVAHRMTLKAMDRQVRQALRSWLRLPKDTATAFFHADVKDGGLGIPSLSSTIPIHKRK RLDRLLGSANPILQQLARSPVMARAIGQHSHPGLAMGRVVDSLQSSREAWRDLLWETKDGSY LSGIQSSRESSRWLDHPEKTFPRVYIRGVQLRANTLGTKVRSSRGRPIPAELLKCRGQCGAH ESLNHIVQGCAVTHEARCKRHNRVAGLLQTRLERPGRVVLGEPRVPLVHSFCKPDLVVLEND TAIVMDVAITSEQRMATAWREKEEKYGEESVARSILQHVRAMGYEASRIEQAPMIISMRGYW YRHSSDKARRLGITKLTRSDLSLLALTGTLRCYDVYMRGTGRARGSRME	289
(73) R2DipCan RT	SSRVTLVPKTPHRPSSPTDLRPISVSSVVTRCLHKILAARWSDSITLPALQFAFQKCDGTFE ATNLLDGLLRERASHNRDLALASVDVSRAFDSVARDTIFRAASSFGAPAPLLEFIKCGYSEA QAFLGPRCLTPQRGVRQGDPLSPLLFIMAVDEAVRAASPELGASLNGATVDAIAYADDLILL APTPEQLSHKLTQLRSALRAGGMTVNERKSFILTIRGDTKSKQSAVVPFPSVLAHGEPFRSL GPEDSFEYLGVNF	290
(73) R2DipCan 5'UTR	GGGGTCATGGTTACAGACCTGGGCTCCTCCTCGTGGTTCCCGCCGGGGTCGGACGGCTGGGT CTTGGGATCGGGAGTGGGATCTCCTGGTCCTGGTCCTGGTCCCCTGGCCTGGTCCTCCCAGC TAATGCCCGGGTCACCCCTCTGGAGAATACGGGCTAACCACCCACTCCAGAGGTATCA	291
(73) R2DipCan 3'UTR	TAGGTCCCGAATGACGGAGCTTTTTCCCGATAACCAAACCCGTAGTCGACGCACCCCCTCTT GGACTGGTATCTCCATTGCAGTGTACATCCCTTGTTTTCCGCGTGCTTATATGTGTAACTAT GACCCCCCCCTTGCTGTACACAAATGCTGTCCAAAAATAAATGTGTC	292
(74) R2MicRuf ORF	LLVPIKKYRDKPGPSGTQADSRDSKVEGAVAGPSGKDSDQVEFRCVCGRVFPSKRSLAMHRH HRHPQELNDERLATLPTKRLVWSKEDDNALIRYANNLWTPGLLKKDLCSKLVEHFPGRSVES VKKRLQLLKWAPPPNSPPTSQDSGLSLSPSGLSPVSVDLTSSWRDSILEAVISNLQEPKLGA EDLVKIANDLRSGMISPTQCAQLLDDHAKRYFPHLWKPTKRKPQRHHGLLSNKKIRRANYAA IQRLYKIQKKDAAKCVLSGQWRKAYLGRANLIEGLDDYWSKIFETASAPDSRDFGPPAEPKW SIIAAITADEVTTALRDMGKSATGLDRISATELLVWHQPSLAAYFNLLLVAELAPPHLSKSR VVYIPKTNDPVSPSDFRPISIASVILRAFHKVLAHRITNQLDFSPLQHAFLQRDGCLEASSL LQAVLRKVHDDVVPASMLFLDVSKAFDSISHYTLLTVIEASGLPPPLTKYLRHLYDHAFVQL ESRSVKCGQGVRQGDPLSPLLFILALEALLKDSIPEVGFDLDGPTIGYLAYADDLVLLAEKP NRLQEKLDGLVSSLAAAGLSLNAKKSIGLSIAKDGKKKNLVLLPKTYRSGDDIIKPMNINDS VRYLGLDFNWKGKITPKHTTKLQKMLEEVSMAPLKPYQRLELLKTFLVPKFTHELVLGSAHR NTLKKLDVLVRDHVRRWLRLPKDTPLGYLHANIKDGGIGIPCFSSSIPILQKDRFTKLMNNQ GPVLDAIKQQKTFSKMMKRINLPCRVNGTAVTSKVQVREAWAAALKDSIDGKELMVPDVDPA SHKWLKTPENIFPRLHLRGIQLRGGLLPTKARNGRGLNRTDINRICNGACNKPETLNHILQI CEITHDARCARHNRIVQKLEQFIRKRNMRTWIEPIIPTSNTFMKPDLLVETNNKLMVMDVSV VAGYRMTETWDLKKQKYGSPTNMDALQSWSESQTSIEQLPIVLSSRGLLFGPSGRGLRSIGL SERDLMDLCLLAVAGSLKCYDLYTRGTKSSQRSMSRKPVTSSL	293
(74) R2MicRuf RT	KSRVVYIPKTNDPVSPSDFRPISIASVILRAFHKVLAHRITNQLDFSPLQHAFLQRDGCLEA SSLLQAVLRKVHDDVVPASMLFLDVSKAFDSISHYTLLTVIEASGLPPPLTKYLRHLYDHAF VQLESRSVKCGQGVRQGDPLSPLLFILALEALLKDSIPEVGFDLDGPTIGYLAYADDLVLLA EKPNRLQEKLDGLVSSLAAAGLSLNAKKSIGLSIAKDGKKKNLVLLPKTYRSGDDIIKPMNI NDSVRYLGLDF	294
(74) R2MicRuf 5'UTR	AGGGCATAGTTGCACGGCCTGGGCGCCTCCTCGTGGCGACCGCTGGATAGGACATGGCTTCA TGCCATGTCTGAACTAAGATCCAGGCCAACCTCTGACTGGGCTAACCACCCGGTCGGAGGCT TACT	295
(74) R2MicRuf 3'UTR	TGACCTGCCTTCCCGGTGGCGATGCCTGAGACCTAGCTGTCACTAGCGGTTACGCCTCAAGG ACTCCTGATTGTCTTTCTTCCCCCATCTATCCTTTTATAACTATTGTTTTCTTCTCCCTATG GGCCTGCGTTGGCCACTCTGTAATCGGGATGTTCTTGATAATAAATGATACC	296
(75) R2Chabot ORF	MASGGANEFRCDCGTLWPSARSLAMHNFHRHTEKVMAERREKQPAKRLRWSAEEEAALHRQA DRLWSEGMLSKDLCLLLTDFFPGRSAEAIKKRLQSTRWIPPGQRPPVSPEGNEESSLAEQFD FAMWSKDLLEEAVGNLQEPRMGAELLSRIARGLLTGEFSKEEGVKRLEDHACQWFPYRWKSS VPRKSGNPQMHSNKKIRRMNYAAIQRMYHRRLKDAATCVLSGQWRTAYQPRPEVEGLNEYWT	297

	KIFETRSVEDTGCLRTPTRTLYSLIGRIIPSEVTAALKTMRSSAPGLDRITASELLAWHQPS IASYYNLLLVLEVCPRYLSCSRVAFLPKVPVPSAPDDLRPISITSVLLRAFHKILARRLNDE LHLTSLQFAFLQRDGCLDATTLLHSILRNIHHTRSPGAMLFLDISKAFDTVSHSTIARIITS AGFPIPLIHYICNLYSNSTVQLGEKMIKCGRGVRQGDPLSPLLFILVIEEILLASLPDVGYN FGGQRLGAIAYADDLVLLAEHPQRLQEKFKIFSEALHTTGMVLNPKKSIGLTISRDGKRKAI VWQPHCYNYLGTPIRPLKIDESVKYLGLQFNWKGRINPRHVGKLDEMLGEISAAPLKPYQRI RILTSFLIPKLTHSLVLGLAYRKTLKRADRLVYAAVRSWLRLPKDTALGYFHASVADGGLGI PCLSTTIPLLQRDQFAKLISQTDPLSVIIREQEAFKVVGNIINRPCRINSRVVTSKEEARAE WKTSLSQSVDGSELTIPDVDLASHGWLKYPVRVFPRLFLRGIQLRGGVLPTKARSARGNNRV VVDRSCKGLCGTSETLNHVLQSCEITHNARCSRHNKIVQQLEYLLNKIGIKIWVEPIIPSGS SFIKPDLVVMKDNQLWVLDVSIVSGMRLRESWNIKIVKYGSTERVNDIVCWFGSPIAVKQLP IIISNRGLCSTFSGSGLRALGLSRRDITDLCLLTVKGSLMCYDLYNRSTRSSGR	
(75) R2Chabot RT	LSCSRVAFLPKVPVPSAPDDLRPISITSVLLRAFHKILARRLNDELHLTSLQFAFLQRDGCL DATTLLHSILRNIHHTRSPGAMLFLDISKAFDTVSHSTIARIITSAGFPIPLIHYICNLYSN STVQLGEKMIKCGRGVRQGDPLSPLLFILVIEEILLASLPDVGYNFGGQRLGAIAYADDLVL LAEHPQRLQEKFKIFSEALHTTGMVLNPKKSIGLTISRDGKRKAIVWQPHCYNYLGTPIRPL KIDESVKYLGLQF	298
(75) R2Chabot 5'UTR	GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG GTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGCGGGAGTAACTATGACTCTCTTAACTTTCTT GTCATAGTTACAGGCCTGGACACCTCCGCGTGGTGCCCGCTGGATAGGGTATGGGATCTTAT ATCCCATGCCTGAACTAAGTTCCAGGCCACCCTCCTCTCGGGAAGACTTAATTACCCGTTAG GAGGTTGCTGCTGTTTTTTTGTTAAATTGATTAATTTATGTCCGGAGCCCTCTGGTTGACCG GTGATACAGCAGGTAAGTGAGGTAAGACAGTTGTCAAGAATCCACA	299
(75) R2Chabot 3'UTR	TGATTTTCGATCTGGATTCGATGAACGACGATTGTCTCCTGTTATGATTTGAATTATTTTCT TGGAGATGTTGTATTTATAAATTTGCTAAATTATAATTGTATCTTTTTGAACATACTCCAAT AATAAATGTAAA	300
(76) R2LiaOli ORF	MEAAEFRCACGRVWPSARSLAMHNHHRHPEVVNEARLQARPVSRLHWTGEEEEALRRLADRT WTAGMLKGELCSKLADSFPGRSAEAIKKRLQHIKWAPLSSGAISSEASTPRAALICDGFDSV EWRKKLLGDIVKNLNEPRMGSDTLVAISQRLRTGELSLVAAVECLEDHAKQWFPHKWRPSIP RSQKPTQPHSRKAIRRATYAAVQRLYYRRQKDAASAVLSGQWRDIYRQRKEVVGLNRHWAKV FEIPSVADTEPPIMPSNLLCSLLDRIHPSEVTAALGDMTASAPGLDRLTAGELLKWHQPSLS AYFNLLLAVEAPPRHLACSRTAFIPKVADPTSPEEFRPISIASVLLRAFHKILARRLRDHLD LSSLQFAFLLRDGCLEATTLLRSVLRRIHHERTSCAMLFLDISKAFDTVSHGTLSRVMRSAG FPPPLVSYIQHLYECSTVQLGDTMVSCNRGVRQGDPLSPLLFILVIDDILRKSLPEVGYNFG GLRLGAMAYADDIVLLADHPQRLQEKLDVFSQALHAAGMALNAKKSVGLTVVRDGKRKVVAL DPTCYNYRDGPIRSVGLRESVRYLGLQFDWKGQVRFKHTRKIDSMLEEITRAPLKPHQRLRI LSAFLIPKFTHELVLGAAHRKTLLQLDRQVRVAVRSWLHLPKDTPLGFFHASVRDGGIGIPS FSTTIPILQKNRFEKLASQTSQWSVVVRDQAAFRVLSNIVNRPCKLNNMAVTTKKESRVEWQ RALSQSVDGNEVAIPGVDPVSHRWLRDPSRVFPRLFLRGIQLRGGVLPTRARSARGINREVT DKRCRGPCQASETINHILQACEITHDARCARHNRVLQRLDQCIRKGGAETWVEPIICSGSSF IKPDLIVSKLDHLWILDVSVVAAPRLPVTWDLKVAKYGSAVCARDISTWHASSVPLKHLPVV FSNRGLCYEPSGKGLRALGLSHRDIMDLCLLTIQGSLKCYDLYTRGTRSGFGGTV	301
(76) R2LiaOli RT	CSRTAFIPKVADPTSPEEFRPISIASVLLRAFHKILARRLRDHLDLSSLQFAFLLRDGCLEA TTLLRSVLRRIHHERTSCAMLFLDISKAFDTVSHGTLSRVMRSAGFPPPLVSYIQHLYECST VQLGDTMVSCNRGVRQGDPLSPLLFILVIDDILRKSLPEVGYNFGGLRLGAMAYADDIVLLA DHPQRLQEKLDVFSQALHAAGMALNAKKSVGLTVVRDGKRKVVALDPTCYNYRDGPIRSVGL RESVRYLGLQF	302
(76) R2LiaOli 5'UTR	AGGGCATAGTTGCACGGCCTGGGCGCCTCCACGTGGCGACCGCTGGATAGGGTGCATTTTTA CACCTGAACTAAGATCCAGGCCACCCTCCTTTTGGGCTTGATTTCCCAATAGGAGGTTCTCT GCTGGTTTCCGAACCAATTTGTAATTGTTGTCCTGTGCTCTCTGGTGGCAGGACTTCCAGCA GACATAGTGCCAGTAACTTCCTTGGCCAATTAGCAGTCCACC	303
(76) R2LiaOli 3'UTR	TAAAGTTTAGTGGGGTGGGTGTGAGATCTGGGTCAAGGAATGTTCCTTTAGAACGGCGCTAG TTAATTAGCGAGTTTTCGCCTTGGAGGATGATTATGCTTGGTTAACGTCCCCCAAGATATTG GTGGGTTCGGAACCTGTTTGATTATCTGTACCAATAATAATAAATGATAAAA	304
(77) R2BotJar ORF	MLTRTDDPRPPGHPDPSCPGSCSRDCRESVAESPRHACPCGCEFPTKRGLAMHRFHRHMAEV NAERLATLPTKKLAWTQEEEAGLLRHASQAWKEGMTKTQIFAAVLPHFPGRSMEAIKKRLQS LKWVPPQNTPERNLQPLHATSGNVAPAALPSPDSSVKVDGQGQNTEDPAASWRRALLAAAVD HLEEPKLGAETLRQITQDLENGKISSKKCRVLMNEHAKHWFPHTWKPSSARPEPKARQWSNR EIRRGNYGAIQKLYKLRKRDAAQTVLSGRWRTAYRGRAPMPAKMEEYWSEVFQAPSATDAIP PGSPSDTLWSVMGKITSSEVTAALKGMRNSAPGLDRITAEELLTWHQPSLASYYNLMLVAGG PPKHLACSRVTFIQKVELPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQHGFLQR	305

	DGCLEATALLHAILRKVHNSRKPCAMLFLDVAKAFDTVSHQTLFRVASEMGLPPPLVNYLKY LYSRSTIKLADKVIKCGRGVRQGDPLSPLLFIMVMEVIVREALPEVGFNLDGQRVDSLAYAD DLVLLAEKSTRLQEKLERLSEALRKAGMALNARKSHGLTVTKASKRKQMVLTPTTYECEGDT IRPMGTDDSVRYLGLHFNWKGRIVPKHTGKLDTLLKELTRAPLKPYQRLELLRFHTVPKLTH ELVLGHAHRNTLKKLDCLTRAAVRQWLRLPQDTPLGYLHAHVKDGGLGIPCFSTSIPLLQQR RFEKIATNPATIFQVLQRQDSFRTQWRRLDKPCRLNSTVVTSKVEVRKEWENMLSNSVDGRE LRYPEVDKASHEWLLRPSRVFPRLHMRGIQLRGGLLPTKARSGRGKNRQVTSKTCRGACQTV ETLNHILQVCEVTHDARCARHNRVLHHLEKLLRKKGIRTTLEPVIPAGSSFIKPDLIIETDT KLFVLDVSVVAGYRMAESWSIKVEKYGRPDRVEAIRRWAAIPEEIRVEHLPVIVSNRGLCYR PSGDGLRALGLTSRDISDLCLLAINGSLRCYDLYMRGTRATRETERHLP	
(77) R2BotJar RT	CSRVTFIQKVELPQTPGDYRPISIASTILRALHKILAWRLRDTLRLSPFQHGFLQRDGCLEA TALLHAILRKVHNSRKPCAMLFLDVAKAFDTVSHQTLFRVASEMGLPPPLVNYLKYLYSRST IKLADKVIKCGRGVRQGDPLSPLLFIMVMEVIVREALPEVGFNLDGQRVDSLAYADDLVLLA EKSTRLQEKLERLSEALRKAGMALNARKSHGLTVTKASKRKQMVLTPTTYECEGDTIRPMGT DDSVRYLGLHF	306
(77) R2BotJar 5'UTR	CTACGGGCCTTAACTAAGATCCAACCCACCCCTCCGTTACTGACAGACCACAGTAACTGAGA TTAAACAGCTGGC	307
(77) R2BotJar 3'UTR	TGACAAAGAGGACCCCTCTCTGATTAGGCGGTGACCACCCTGGCCTCTACCGAAA	308
(78) R2BoaFul1 ORF	MDRNTNVKGPHASSPSNAFTCSCGRQFSTKRGLGMHRRHCQPVEVNIERMAVLPTKRLIWSL EEDEALRRCADHQWKTGMLKKDLYTLLLPHFPGRSQEAIKKRLQTIKWVPPTDPALVSHPPP GVDTGSASDRGRDSSAASDSDPDPAATWVSSLLTATTSNLQEAKMGAEELYVIARDLAMGVI SQTQSRALLQDHAQKWFPHLWKPSKPRPLSRTHQWSKRTVRRANYAVIQRLYKLRRRDAAST IFSGQWKNAYRGSPPLPPDMESYWSSVFQTPSAIDDRVPATPSEVYWTLLCAMTPSEVTAAL KSMASSAPGLDRITAATLLTWHQPSIAAYFNLLLAAGGPPEHLACSRVVFLPKTEAPSTPGE FRPISIASTILRAFHKILSWRLRDTLQLSPFQHGFLQQDGCLEATTLLHTILRRVHLKCEPC AMLFLDIAKAFDTVSHHTLFRVATAAGLPPPLVCYLRHLYERSTIQLAGTTTVCGRGVRQGD PLSPLLFILVIEDIVTSSLPSIGYDLDGHRIHSIAYADDLVLFAEKSPRLQEKLHLLSEALS KAGMTLNAKKSIGLTIAKDGKRKRTALLPTTYRCGIGNINPLGPGETLRYLGIQFNWKGRVT PKHTGKLETLLQELTQAPLKPQQRLLLLKFFLVPKFTHELVLGHAHRNTLRKMDWLIRAAVR RWLRLPKDTPVGYLHAHIKDGGLGVPCLSTTIPLLQASRLEKIAAHPTTLFQQIQRQDALGP LRRRMNQPCRINTTTVTTKAEAREEWKSTLVASFDGRELGVPKVDPVSHDWLLRPSRVFPRL FLRGIQLRGGVLPTKSRSARGANRKVVDKICRGACRATETINHTLQVCEVTHDARCARHNRI VRRLDQLLRRQGIQPSVEPIIPSGTSFIKPDLIAEINNRLWVMDVSVVAGHRMQETWKIKVD KYGKPEQVASILRWAAVDSSVPVHHLPIVITNRGLIFEPSAAGLRTLGLSRRDLVDLCLLTI MGSLKCYDVYMRGTF	309
(78) R2BoaFul1 RT	CSRVVFLPKTEAPSTPGEFRPISIASTILRAFHKILSWRLRDTLQLSPFQHGFLQQDGCLEA TTLLHTILRRVHLKCEPCAMLFLDIAKAFDTVSHHTLFRVATAAGLPPPLVCYLRHLYERST IQLAGTTTVCGRGVRQGDPLSPLLFILVIEDIVTSSLPSIGYDLDGHRIHSIAYADDLVLFA EKSPRLQEKLHLLSEALSKAGMTLNAKKSIGLTIAKDGKRKRTALLPTTYRCGIGNINPLGP GETLRYLGIQF	310
(78) R2BoaFul1 5'UTR	TGGTTACGCGGCCTGGACGCCTCCTAGTGGCGACCGCTGGATAAGGGGTAGATAGTTTAACT TCTGCTCCCGAGCTAAGAGCCAGGCCACCCCTCGGCGAGGTTAGTACCTTGTTGAGGTTTCA CTGCTGGTTTTCTAATTGAAAGGAATTGTCCCGGTCCCCCTGGACTTAAGGACTCCAGCAGC	311
(78) R2BoaFul1 3'UTR	TAAGGTTGAACTGGCTTCTATTGGTCCCCCCCCCAACACATTAGTGTACTACTCAATTAGCG ACTTGTCGCTCTCAGGGAGGCTCTCCTCCCTGATAATATATGTCACG	312
(79) R2BoaFul2 ORF	MSSDPPGSMDSSRTEFCTAPKGLESCAGINAYECVCGRRFPTKRGLGMHRRHCQPVEVNQER VAALPTKRQVWSKEEDDALRRHANCQWKKGMLKETLYTSLLPHFPGRSQEALKKRLQSIKWV PTADPKLPSCAPLEVDCSTASGQGREPPATAGVDPSDLWLTCLLEKALSDLREVKMGADELR LIAQNLLKGTVDQLRSRDLLQTHAQKWFPHLWKPSKPHPPPKNNSWSKRSIRRANYAIIQKL YKLRRKDAASTVLSGRWRSAYQGTPPLPSNMEDYWSSIFETPSEIDNKAPAAPLTTHWSLLC PMAPSEVTAALKSMQSSAPGLDRITAQTLLTWHQPSLAAYFNLLLAAGGPPEHLTCSRAVFI PKVDNPSAPGDYRPISIASTILRAFHKIMAWRLRDTLTLSPFQHGFLQQDGCLEATTLLHTI FRRVHLNCNPCAMLFLDIAKAFDTVSHHTLFRVATSAGLPPPLVCYLRQLYDRSTVHLAGKP ITCGRGVRQGDPMSPLLFILVIEDIITHSLPEIGYNIDGQRVHSIAYADDLVLFAEKPPRLQ DKLSLLSEALKNAGMSLNAKKSVAITIAKDGKRKCMALLPTVYRCDTGSINTLGPGDTIRYL GIHFNWKGRVTPKHTGKLDPFLQELTQAPLKPHQRLQLLKFFLVPKFTHELVLGHAHRNTLR KMDFLVRAATRRWLRLPKDTPIGYFHARIKDGGLGVPCFSTSIPLLQAARMEKIASHPTTLF QLIQRQDALGPLRRRMSRPCKINSVTVTTKAEVRDEWKHLLVTSNDGKELGIPKTDPSSHDW LLRPSRVFPRLFLRAIQLRGGVLPTKSRSARGNNRQVNDTICRGACKARETLNHTLQVCEKT HDARCARHNRVMRRLDQLLRRQNIHPTVEPIIPSGSSFIKPDLVLDTKDKIWIMDVSIVSGY RTQETWNLKTNKYGLPDRVASILQWANTTKPVPVCHLPIIITNRGLMYEPSATGLRTLGLSR	313

	RDIVDLCLLTIVGSLKCYDTYMRGTA	
(79) R2BoaFul2 RT	CSRAVFIPKVDNPSAPGDYRPISIASTILRAFHKIMAWRLRDTLTLSPFQHGFLQQDGCLEA TTLLHTIFRRVHLNCNPCAMLFLDIAKAFDTVSHHTLFRVATSAGLPPPLVCYLRQLYDRST VHLAGKPITCGRGVRQGDPMSPLLFILVIEDIITHSLPEIGYNIDGQRVHSIAYADDLVLFA EKPPRLQDKLSLLSEALKNAGMSLNAKKSVAITIAKDGKRKCMALLPTVYRCDTGSINTLGP GDTIRYLGIHF	314
(79) R2BoaFul2 5'UTR	TTGTGACGTGTAGTTACATGGTCTGGGCGCCTCCGAGTGGCGACCGCTGGATAAGGTGTGTG GATTTAATTCCCACACTGAACTAAGATCCAGACCCCCCCCCTACTTGGCCGACCACCTCGTA GCGGGTTTTCTCTGCTGGTTAATTCTGATATATTATA	315
(79) R2BoaFul2 3'UTR	TAAGCCGGTACGATTCTAAGACCGAATCCGCCCTAAGCGGTGACAGTTCAGTGCAAACTCTA ATAGCGATTCGTCGCCCTCGGAGGTGTTCTACCTCCGATAATTCATGTTACAA	316
(80) R2TriAlb1 ORF	MLTLHMDPRPPGSPDPSWPGSCPREVCQESAAEPPKHSCPCGSEFRTKRGLAMHRSHRHQAE VNAERLASLPTRKLVWTQEEEAGLLRHATQAWKKGMAKTQIFASLLPHFPGRSVEAIKKRLQ SLKWVPPRDTPERNSQPLNATSGNVAPATLPAPDSSVTIDDQGRVEDPAASWRRALLTAAAV RLSEPKLGAETLRQITQDLQDGKISSKKCRVLMNEHAQHWFPHKWKPTSVRPEPKTRQWSTR EIRRANYASVQKLYKLRKRDAAQTVFSGRWRNAYRGQAPLPAEMDEYWAKIFETPSAIDSVT PEVPSKVLWSVMGKIVPSEVTDALKGMRNSAPGLDRISAEELLTWHQPSLAAYYNLMLAAGG PPEHLACSRVTFIPKLENPQAPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQR DGCLEASALLHAILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVASAMGLPPPLVHYLKH LYSRSTVRLANKVTTCGRGVRQGDPLSPLLFIMVMEDIVREALPEVGFNLDGQRVDSLAYAD DLVLLAEKSPRLQEKLHRLAEALRKAGMSLNARKSHGLTVTKVSKRKQMVLSPTSYECEGET IKPMGTDDSVRYLGLHFNWKGRIVPKHTGLWDSLLKELKKAPLKPYQRLQLLRFHAVPKFTH QLVLGHAHRNTLKKLDCLTRAAVRQWLRLPPDTPLGYLHANVKDGGLGIPCFSTSIPLLQQK RFEKIATNPATIFQVIERQDSFRTQRRRLDRPCRLNKTVVTSKAEVREGWGNMLSNSVDGKE LRYPEVDKASHEWLLRPSRVFPRLHLRGIQLRGGLLPTKARSGRGKNRQVTNKTCRGACQNV ETLNHILQVCEVTHDARCARHNRVLHLLEKLIRKKGARATLEPVIPAGSSFIKPDLIIETDT QMFVLDVSVVAGYRMAESWDIKVEKYGRPDRVEAIRRWAAIPEKIRVEQLPVIVSNRGLCYR PSGVGLRALGLTSRDISDLCLLAISGSLRCYDLYMRGTRASKEN	317
(80) R2TriAlb1 RT	CSRVTFIPKLENPQAPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQRDGCLEA SALLHAILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVASAMGLPPPLVHYLKHLYSRST VRLANKVTTCGRGVRQGDPLSPLLFIMVMEDIVREALPEVGFNLDGQRVDSLAYADDLVLLA EKSPRLQEKLHRLAEALRKAGMSLNARKSHGLTVTKVSKRKQMVLSPTSYECEGETIKPMGT DDSVRYLGLHF	318
(80) R2TriAlb1 5'UTR	TTACCCATTGGTCTTGATTTTGATACTGGTGTTTTGCGAATTGGTGTTGATTGCAAAGTGTT GTTGCGGTGAATTGACGATTGGTTGCTAGGGTTGGGCACCTCCAAGTGGTACCCGCTGGATA GGGCCTTCGTGCCTGAACTAAGCCCCAACCCACCTCTCCGTTGCTGGCAGACCACAGTAACT GAGATTAAACAGCTGGT	319
(80) R2TriAlb1 3'UTR	TGAGCAGTGCCGAACGGGACCCAGAGGACCGCCTCGCCGATCAGGCGTGGGCGGTGACCACC CTGGCCTCTACCTGAAATAGCGATCTAATCGCCTTAGAAGGCTTGATGCTGGAGGGTCGAGC AGACACCAAGTGTCACGAATGCTCCTTACCCTACCTCTTCTTGCTTCTCTAATAATCAATGT TAACA	320
(81) R2TriAlb2 ORF	MVTLHMDPRPPGSPDPSWPGSCPREAFRESAAKLPKHVCPCGSDFKTKRGLAMHRFHRHPAE VNAERLASLPIKKLVWTQEEEGGLLRHANQAWKKGMAKTQIFAAILPHFPGRSVEAIKKRLQ SLKWVPPRDSPEGNSQPLYASSGNIAPTTLPAPDSSVPIDNQGTEADPAASWRKALLTASAG HLSEPKLGAETLRNICQELQDGKISSKKCRVLMNKHARHWFPHTWKPTSARPEPKTRQWSTR EIRRANYAAVQKLFKLRKRDAAQTVFSGSWKNAYRGHAPLPTNMEEYWTEVFGSPSAIDAAT PEVPSKVLWSVMGRIVPSEVTAALKGMRNSAPGLDRISAEELLTWNQPSLAAYYNLMLAAGG PPEHLACSRVTFIPKTDNPQTPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQR DGCLEASALLHAILRKVHNNRKSCAMLFLDVAKAFDTVSHETLFRVASGMGLPPPLVHYLKH LYSRSTVRIADKATKCGRGVRQGDPLSPLLFIMVMEDIVREALPEVGFNLDGQRVDSLAYADDLVLLA EKSPRLQEKLQRLTEALQKAGMSLNARKSHGLTLTKVSKRKQMVITPTAYECDGETIKPMGT DDSVRYLGLHFNWKGRIVPKHTGTWDNLLKELKKAPLKPYQRLQLLRFHAVPKFTH HLVLGHAHLNTLKKLDCLTRAAVRQWLRLPPDTPLGYLHANVKDGGLGIPCFSASIPLLQQK RFEKIATNPATIFQILQRQDSFRTQRRRLDKPCRLNTRVVTSKAEVREEWGNMLSNSVDGKE LRCPEIDKASHEWLLRPSRVFPRLHLRGIQLRGGLLPTKARSGRGNNRQVTIKTCRGACQNI ETLNHILQVCEVTHDARCARHNRVLHRLDKLIRKKGARATLEPIIPAGSSFIKPDLIIETDT HLFILDVSVVTGYRMAESWDIKVEKYGRPDRVEAIRRWATIPEKIRVEQLPVIVSNRGLCYR QSGAGLRALGLTSRDISDLCLLAISGSLKCYDLYMRGTRATREN	321
(81) R2TriAlb2 RT	CSRVTFIPKTDNPQTPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQRDGCLEA SALLHAILRKVHNNRKSCAMLFLDVAKAFDTVSHETLFRVASGMGLPPPLVHYLKHLYSRST VRIADKATKCGRGVRQGDPLSPLLFIMVMEDIVREALPEVGFNLDGQRVDSLAYADDLVLLA EKSPRLQEKLQRLTEALQKAGMSLNARKSHGLTLTKVSKRKQMVITPTAYECDGETIKPMGT	322

	DDSVRYLGLHF	
(81) R2TriAlb2 5'UTR	GAATTGTTGATTTTGATACTAGTGGCTTGAGAATTGGTATTGATTGAATAGGTTCATAGATG ATCTGTTGCTAGGGTTGGACACCTCCCCGTGGTACCCGCTGGATAGGGCATTCGTGCCTGAA CCAAGCTCCAACCCACCCCTCTGTTGCTGACAGACCACAGCAACTGAGGTTAAACAGCTGGT	323
(81) R2TriAlb2 3'UTR	TGAGCCGTGCCGAACGGAACCCAGAGGACCGTCTCTTGATTAGGCGGTGTCCACCCTGGCTT CTACCTGAATTAGCGAATTTTTATTCGCCTTAGATGGCCTGATATAGGAGGGCTGAGTAGAC ACCGCGTGTCACGAATGCTCTTTACCCTACCTCTTCTCGCCTCTAATAATCAATGTCAACA	324
(82) R2CroTig ORF	MLTDTVVLRPPGLPDPSCPGSCPRVASPKSDAECPKHSCPCGSEFRTKRGLAMHRFHCHPAE VNAERLTTLPTKKLVWTQEEMDGLLRHANGAWKEGMLKTQIFTSLMPHFPGRSAEAIKKRLQ SLKWVPARDNPERNSQPLHATSGNVAPATLPTPDPSRVDDQGANIEDPAASWRRALLTAAAD HFKEPKLGAGTLKQITQDLLNGKISSRECRVLMGEHAQHWFPHTWKPSSARPEPKARQWKAR EIRRANYAAVQKLYKLRRRDAAQAVFSGSWRNAYRGRAPQPDKMDQYWAEVFETASATDVTP PGSPSEIYWSVVSKITPSEVADALKGMRNSAPGLDRITAEELLTWDHPSLAAYYNLMLAAGG PPEHLACSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQR DGCLEATALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKC LYSRSTVRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYAD DLVLLAEKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEP IKPMGTDDSVRYLGLHFNWKGRIVPKHTGKLDSLLKELTKAPLKPYQRLQLLKFHAVPKFTH ELVLGHAHRNTVKKLDCLTRAAVRKWLRLPQDTPLGYLHANVKDGGLGIPCFSTSIPLLQKK RFEKIATNPATIFQIMQRQDSFRTQGRRLDKPCRLNSTVVTSKTEVREEWGNMLSNSVDGRE LRHPEVDKASHEWLLRPDRVFPRLHMRGIQLRGGLLPTKARSGRGNNRQVTSKTCRGACQNV ETLNHILQVCDVTHDARCARHNRVLHRLERLLHKKGLRTTLEPVIPAGSSFIKPDLIIETDT QTYVLDVSVVAGYRMTESWDIKVEKYGRPDRVEAIRKWAAIPATKRVEHLPVIVSNRGLCYQ PSGDGLRALGLTSRDISDLCLLAISGSLRCYDLYMRGTRAV	325
(82) R2CroTig RT	CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQRDGCLEA TALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKCLYSRST VRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYADDLVLLA EKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEPIKPMGT DDSVRYLGLHF	326
(82) R2CroTig 5'UTR	CGACCATTTGGTGTTGATATTGATACTGGCGACTGGCTAGTTGGTGTTGGTTGAAAGTGCTG GCGAGGTTGAGCTGATAATTGGTTGCACGGGTTGGGCACCTCCTAGTGGTGACCGCTGGATA GGGCTTCGGCCTGAGCCAAGATCCAACCCACCCTCCGTTGCTGGCAGACCCCAGTAACTGAG ATTAAACAGCTGGT	327
(82) R2CroTig 3'UTR	TAGGACTACCGAGCGCCACCGGAGGGGACCCAGAGGACCGTCTCGGTGATTAGGCGGTGACC GCCCTGGCCTCTACCTGAAATAGCGATCAAATCGCCTTAGAGGGCCTGATTTGGGGGGTTGG GCAGACACTCTGTGTCATTACTGCCCTTTACCTTTCCACTTCTTGCCTGTCTAATAATAAAT GTTAACA	328
(83) R2CroHel ORF	MLTDTVVLRPPGLPDPSCPGSCPRVASPKSDAECPKHSCPCGSEFGTKRGLAMHRFHCHPAE VNAERLTTLPTKKLVWTQEEMDGLLRHANGAWKEGMLKTQIFTSLMPHFPGRSAEAIKKRLQ SLKWVPARDNPERNSQPLHATSGNVAPATLPTPDPSRVDDQGANIEDPAASWRRALLAAAAD HFKEPKLGAGTLKQITQDLLNGKISSGECRVLMGEHAQHWFPHTWKPSSARPEPKARQWKAR EIRRANYAAVQKLYKLRRRDAAQAVFSGSWRNAYRGRAPQPDKMDQYWAEVFETASATDVTP PGSPSEIYWSVVSKITPSEVADALKGMRNSAPGLDRITAEELLTWDHPSLAAYYNLMLAAGG PPEHLACSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQR DGCLEATALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKC LYSRSTVRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYAD DLVLLAEKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEP IKPMGTDDSVRYLGLHFNWKGRIVPKHTGKLDSLLKELTKAPLKPYQRLQLLKFHAVPKFTH ELVLGHAHRNTVKKLDCLTRAAVRKWLRLPQDTPLGYLHANVKDGGLGIPCFSTSIPLLQKK RFEKIATNPATIFQIMQRQDSFRTQGRRLDKPCRLNSTVVTSKTEVREEWGNMLSNSVDGRE LRHPEVDKASHEWLLRPDRVFPRLHMRGIQLRGGLLPTKARSGRGNNRQVTSKTCRGACQNV ETLNHILQVCDVTHDARCARHNRVLHRLERLLHKKGLRTTLEPVIPAGSSFIKPDLIIETDT QTYVLDVSVVAGYRMTESWDIKVEKYGRPDRVEAIRKWAAIPATKRVEHLPVIVSNRGLCYQ PSGDGLRALGLTSRDISDLCLLAISGSLRCYDLYMRGTRAV	329
(83) R2CroHel RT	CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDNLQLSPFQHGFLQRDGCLEA TALLHTILRKVHNSRKSCAMLFLDVAKAFDTVSHQTLFRVAVELGLPPPLVNYLKCLYSRST VRLADKATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYADDLVLLA EKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVITPTTYECEGEPIKPMGT DDSVRYLGLHF	330
(83) R2CroHel 5'UTR	CGACCATTTGGTGTTGATATTGATACTGGCGACTGGCTAGTTGGTGTTGGTTGAAAGCGCTG GCGAGGTTGAGCTGATAATTGGTTGCACGGGTTGGGCACCTCCTAGTGGTGACCGCTGGATA GGGCTTCGGCCTGAGCCAAGATCCAACCCACCCTCCGTTGCTGGCAGACCCCAGTAACTGAG ATTAAACAGCTGGT	331

(83) R2CroHel 3'UTR	GACTACCGAGCGCCACCGGAGGGGACCCAGAGGACCGTCTCGGTGATTAGGCGGTGACCGCC CTGGCCTCTACCTGAAATAGCGATCAAATCGCCTTAGAGGGCCTGATTTGGGGGGTTGGGCA GACACTCTGTGTCATTACTGCCCTTTACCTTTCCACTTCTTGCCTGTCTAATAATAAATG	332
(84) R2OxyLep1 ORF	LLTEGRGIVTGLNMHIFFSPGPPGLSDSSDRDGNANSTGTQSEANNPSFPCFCGAQFKTKRG LGKHQSYRHPAQVNAERVKALPTKRLTWSHEDDEALRRLANNIWKDGMMKKVLITNLTPYFP GRSHEAIKKRLQVITWSPERAEHPTVRETPDTNPVERDTPGTTDPVTHGPDPSASWRDALLR AAAEHLTESNMGAEELTQIARDLQNGVSRKRIRTRLNQHAEKYFPHKWKPSVIRPKKTIASS KKAIRRANYATVQKLYKLRQRDAASTVLSGRWRSAYKGPAPIPQGMEEYWRSVFETPSAADN SVPVVPSTVLWSLMEPIAPSEVTAALHGMKASAPGLDRISAEELLTWDQPSLAAYYNVLLAV GGPPKHISCSRVVFLPKTEDPSAPGDYRPISIASTILRAMHKILAWRLRDTLQLSPLQHAFL QRDGCLEASTLLHTIFRGVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAREAGLPRPMVSYL RHLYQESTIQLAEATTQCGRGVRQGDPLSPLLFIMVMDDIIKSSLPDIGFDLNGQKIGSLAY ADDLVLLAGNPSRLQEKLNLLSEGLQRAGMALNAKKSHGITIARDGKRKCMALVPEAYKCKG DNIQPLGTDDSVKYLGLQFNWKGRITPKHTGKLATMLNELTKAPLKPHQRLKLLKIHLVPKL THELVLGHAHRNTLKKLDCLTRAAVRRWLRLPPDTPVGYFHANVKDGGLGMPCFSTSIPLLQ QKRLEKIALSPASLFQEVQRQDSFRTLGRRIERPCRLNGLEVATKAEVREGWRIALSSSVDG RELSIPKVDKASHEWLRNPARVFPRLYIRGIQLRCGLLPTKARSARGNNRQVTDRTCRGACR APETLNHVLQVCEVTHDARCARHNRVLRSLEKLLSRAGIQTRVEPIIPSGASFIKPDLIAEM ENQLWVLDVSVVAGYRLEESWEIKSAKYGRPDRVADIRRWAAVPDTTPVFQLPVIISNRGLF YEPSGKKLRALGLSCRDIMDLCVLAIGGSLKCYDWYTRGTKDWKC	333
(84) R2OxyLep1 RT	CSRVVFLPKTEDPSAPGDYRPISIASTILRAMHKILAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTIFRGVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAREAGLPRPMVSYLRHLYQEST IQLAEATTQCGRGVRQGDPLSPLLFIMVMDDIIKSSLPDIGFDLNGQKIGSLAYADDLVLLA GNPSRLQEKLNLLSEGLQRAGMALNAKKSHGITIARDGKRKCMALVPEAYKCKGDNIQPLGT DDSVKYLGLQF	334
(84) R2OxyLep1 5'UTR	GAAGATCATAGTTGCATGACTTGGACGCCTCCGCGTGGCGACCGCTGGATATGATACTACGG TATCAGAACTAAGATCCAAGTCACCCCTCTGCC	335
(84) R2OxyLep1 3'UTR	TGATTTGCTTTGCTTTGTATTGTCTGTATGACGTTGACGTTGTTGTCTAATAGCGGGCGACC GCCTTGAAAGGGGAGGCGGGCGGGGAAGGTACAAATATATGTATATATGTGTAAGTTTTGAC ATCCCTATACATTGCTGTAACCTGTAACTGTCTTGTTTTGCCTTTTCAAAAATAAAATGTAA C	336
(85) R2OxyLep2 ORF	LLTEGRGIVTGLNMHIFFSPGPPGLSDSSDRDGNANSTGTQSEANNPSFPCFCGAQFKTKRG LGKHQSYRHPAQVNAERVKALPTKRLTWSHEDDEALRRLANNIWKDGMMKKVLITNLTPYFP GRSHEAIKKRLQVITWSPERAEHPTVRETPDTNPVERDTPGTTDPVTHGPDPSASWRDALLR AAAEHLTESNMGAEELTQIARDLQNGVSRKRIRTRLNQHAEKYFPHKWKPSVIRPKKTIASS KKAIRRANYATVQKLYKLRQRDAASTVLSGRWRSAYKGPAPIPQGMEEYWRSVFETPSAADN SVPVVPSTVLWSLMEPIAPSEVTAALHGMKASAPGLDRISAEELLTWDQPSLAAYYNVLLAV GGPPKHISCSRVVFLPKTEDPSAPGDYRPISIASTILRAMHKILAWRLRDTLQLSPLQHAFL QRDGCLEASTLLHTIFRGVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAREAGLPRPMVSYL RHLYQESTIQLAEATTQCGRGVRQGDPLSPLLFIMVMDDIIKSSLPDIGFDLNGQKIGSLAY ADDLVLLAGNPSRLQEKLNLLSEGLQRAGMALNAKKSHGITIARDGKRKCMALVPEAYKCKG DNIQPLGTDDSVKYLGLQFNWKGRITPKHTGKLATMLNELTKAPLKPHQRLKLLKIHLVPKL THELVLGHAHRNTLKKLDCLTRAAVRRWLRLPPDTPVGYFHANVKDGGLGMPCFSTSIPLLQ QKRLEKIALSPASLFQEVQRQDSFRTLGRRIERPCRLNGLEVATKAEVREGWRIALSSSVDG RELSIPKVDKASHEWLRNPARVFPRLYIRGIQLRCGLLPTKARSARGNNRQVTDRTCRGACR APETLNHVLQVCEVTHDARCARHNRVLRSLEKLLSRAGIQTRVEPIIPSGASFIKPDLIAEM ENQLWVLDVSVVAGYRLEESWEIKSAKYGRPDRVADIRRWAAVPDTTPVFQLPVIISNRGLF YEPSGKKLRALGLSCRDIMDLCVLAIGGSLKCYDWYTRGTKDWKC	337
(85) R2OxyLep2 RT	CSRVVFLPKTEDPSAPGDYRPISIASTILRAMHKILAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTIFRGVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAREAGLPRPMVSYLRHLYQEST IQLAEATTQCGRGVRQGDPLSPLLFIMVMDDIIKSSLPDIGFDLNGQKIGSLAYADDLVLLA GNPSRLQEKLNLLSEGLQRAGMALNAKKSHGITIARDGKRKCMALVPEAYKCKGDNIQPLGT DDSVKYLGLQF	338
(85) R2OxyLep2 5'UTR	GAAGATCATAGTTGCATGACTTGGACGCCTCCGCGTGGCGACCGCTGGATATGATACTACGG TATCAGAACTAAGATCCAAGTCACCCCTCTGCC	339
(85) R2OxyLep2 3'UTR	TGATTTGCTTTGCTTTGTATTGTCTGTATGACGTTGACGTTGTTGTCTAATAGCGGGCGACC GCCTTGAAAGGGGAGGCGGGCGGGGAAGGTACAAATATATGTATATATGTGTAAGTTTTGAC ATCCCTATACATTGCTGTAACCTGTAACTGTCTTGTTTTGCCTTTTCAAAAATAAAATGTAA C	340
(86) R2AzeFea1 ORF	LLRTSTVGKLLNMLTRSKDPRPPGSPDLSSTGSCDGVASLESSPEFSRHPCPCGSEFRTKQG LAMHRYHRHPAEVNAERLVSLPTKKLVWTQEEEAALLRHASQAWKKGMMKVELFTCLLPHFQ GRSEEAIKKRLNRLQWVPPVDNPERDVPERNSLPLPIPAVSGHVAPATHPLPEAKMVEDQGR	341

	NTGSVESWRRDLLATVADHLSETKLGIEALSQILGGLKSGQISTDKCRALMNDHARQQFPHT WKPSSARPGAEARKWSQREMRRANYATVQKLYKLRRRDAAHTVLSGRWRTAYRGHAPLPSNM EEYWSAIFETPSETDAAIPEVPTEIFWSVMGRIAPSEVTSALKGMRNSAPGLDRISAEELLT WHQPSLAAYYNLMLAAGGPPEHLACSRVAFIPKNDNPQTPGDYRPISIASTILRALHKILAW RLRDTLQLSPFQHGFLQRDGCLEATALLHTILRKVHHNRKPCAMLFLDVAKAFDTVSHQTLF RVAAAMGLPPPLVDYLKNLYSRSTVRIADKATRCGRGVRQGDPLSPLLFIMVMEDIVRRVLP EVGFNLDGQRVDSLAYADDLVLLAEKSSRLQEKLHLLTEALQKAGMALNASKSHGLTLDKVS KRKQMVLTPTTYECKGGTIKPMGTDDSVRYLGLHFNWKGKIVPKHTGKLDSLLKELTKAPLK PHQRLQLLKGHLVPKFIHELVLGHAHRNTLKKLDCLTRAAVRHWLRLPQDTPIGYLHAHVKD GGLGIPCFSTSIPLLQQKRFEKIATNPATIFQIIQRQDSFRTQGRRLDRPCRLNGAVMTSKA EVREEWGNVLSNSVDGKELRCSEVDKASHGWLQHPGRVFPRLHMRGIQLRGGLLSTKARSAR GKDRQVTDKTCRGSCHCPETLNHILQVCEITHDARCARHNRVLHRLEQMLRKKGLRTTLEPV IPSGSSYIKPDLIIETGTRRYVVDVSVVAGYRMKESWDIKVEKYGRPDRVNNIRKWAAIPEQ DLVIQLPVIVSNRGLCYRPSGEGLRALGLTCRDISDLCLMSINGSLRCYDLYMRGTRSDKVR GRPRNTEGG	
(86) R2AzeFea1 RT	CSRVAFIPKNDNPQTPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQRDGCLEA TALLHTILRKVHHNRKPCAMLFLDVAKAFDTVSHQTLFRVAAAMGLPPPLVDYLKNLYSRST VRIADKATRCGRGVRQGDPLSPLLFIMVMEDIVRRVLPEVGFNLDGQRVDSLAYADDLVLLA EKSSRLQEKLHLLTEALQKAGMALNASKSHGLTLDKVSKRKQMVLTPTTYECKGGTIKPMGT DDSVRYLGLHF	342
(86) R2AzeFea1 5'UTR	TCGCCCCACAACTTGACCTTGGTTTTGATGTTGAATGTGAATTGTTGCAAAGCGAATTGTTG TGGACGCGAGATTGACGAACAGTTGCACGGGTTGGACACCTCCCAGTGGTGAACGCCGGATA GGGCCTTGTGCCTGAGCTAAGAGCCAACCCAAACCC	343
(86) R2AzeFea1 3'UTR	TAGGGGACTCCCACAGGACTGCATAGGCAGTGTCCCCTCCACGCACTATCCTGAAGTAGCGA TGAGTATCGCCTTAGAAGGCCTGATGGGGGAGGCTGAGTGGACACAATGTGTCACAAACGCT CTTTACCTCTCCTCGTCTTGCCTGTCTAATAATATATGTTAACA	344
(87) R2AzeFea2 ORF	LLRTSTVGKLLNMLTRSKDPRPPGSPDLSSTGSCDGVASLESSPEFSRHPCPCGSEFRTKQG LAMHRYHRHPAEVNAERLVSLPTKKLVWTQEEEAALLRHASQAWKKGMMKVELFTCLLPHFQ GRSEEAIKKRLNRLQWVPPVDNPERDVPERNSLPLPIPAVSGHVAPATHPLPEAKMVEDQGR NTGSVESWRRDLLATVADHLSETKLGIEALSQILGGLKSGQISTDKCRALMNDHARQQFPHT WKPSSARPGAEARKWSQREMRRANYATVQKLYKLRRRDAAHTVLSGRWRTAYRGHAPLPSNM EEYWSAIFETPSETDAAIPEVPTEIFWSVMGRIAPSEVTSALKGMRNSAPGLDRISAEELLT WHQPSLAAYYNLMLAAGGPPEHLACSRVAFIPKNDNPQTPGDYRPISIASTILRALHKILAW RLRDTLQLSPFQHGFLQRDGCLEATALLHTILRKVHHNRKPCAMLFLDVAKAFDTVSHQTLF RVAAAMGLPPPLVDYLKNLYSRSTVRIADKATRCGRGVRQGDPLSPLLFIMVMEDIVRRVLP EVGFNLDGQRVDSLAYADDLVLLAEKSSRLQEKLHLLTEALQKAGMALNASKSHGLTLDKVS KRKQMVLTPTTYECKGGTIKPMGTDDSVRYLGLHFNWKGKIVPKHTGKLDSLLKELTKAPLK PHQRLQLLKGHLVPKFIHELVLGHAHRNTLKKLDCLTRAAVRHWLRLPQDTPIGYLHAHVKD GGLGIPCFSTSIPLLQQKRFEKIATNPATIFQIIQRQDSFRTQGRRLDRPCRLNGAVMTSKA EVREEWGNVLSNSVDGKELRCSEVDKASHGWLQHPGRVFPRLHMRGIQLRGGLLSTKARSAR GKDRQVTDKTCRGSCHCPETLNHILQVCEITHDARCARHNRVLHRLEQMLRKKGLRTTLEPV IPSGSSYIKPDLIIETGTRRYVVDVSVVAGYRMKESWDIKVEKYGRPDRVNNIRKWAAIPEQ DLVIQLPVIVSNRGLCYRPSGEGLRALGLTCRDISDLCLMSINGSLRCYDLYMRGTRSDKVR GRPRNTEGG	345
(87) R2AzeFea2 RT	CSRVAFIPKNDNPQTPGDYRPISIASTILRALHKILAWRLRDTLQLSPFQHGFLQRDGCLEA TALLHTILRKVHHNRKPCAMLFLDVAKAFDTVSHQTLFRVAAAMGLPPPLVDYLKNLYSRST VRIADKATRCGRGVRQGDPLSPLLFIMVMEDIVRRVLPEVGFNLDGQRVDSLAYADDLVLLA EKSSRLQEKLHLLTEALQKAGMALNASKSHGLTLDKVSKRKQMVLTPTTYECKGGTIKPMGT DDSVRYLGLHF	346
(87) R2AzeFea2 5'UTR	TCGCCCCACAACTTGACCTTGGTTTTGATGTTGAATGTGAATTGTTGCAAAGCGAATTGTTG TGGACGCGAGATTGACGAACAGTTGCACGGGTTGGACACCTCCCAGTGGTGAACGCCGGATA GGGCCTTGTGCCTGAGCTAAGAGCCAACCCAAACCC	347
(87) R2AzeFea2 3'UTR	TAGGGGACTCCCACAGGACTGCATAGGCAGTGTCCCCTCCACGCACTATCCTGAAGTAGCGA TGAGTATCGCCTTAGAAGGCCTGATGGGGGAGGCTGAGTGGACACAATGTGTCACAAACGCT CTTTACCTCTCCTCGTCTTGCCTGTCTAATAATATATGTTAACA	348
(88) R2SisCat ORF	MLTGTVVLRPPGLPDPSCPGSCPRVASPKSDAECPKHSCPCGSEYGTKRGLAMHRFHCHPAE VNAERLATLPTKKLVWTQEEMDGLLRHANGAWKEGMSKTQIFTSLMPHFPGRSAEAIKKRLQ SLKWVAPRANPERNSQPLHATSGNVAPAALPTPDPSRVDDQGANMEDPAASWRRALLAAAAD HLKEPKLGAGTLKQITQDLLNGKISSEECRVLMCEHAQHWFPHTWKPSSARPEPKTRQWKAR EIRRANYAAVQKMYKLRRRDAAQAVFSGRWRNAYRGRAPQPDKMDEYWAEVFETASATDVTP PGSPSDIYWSVVSKITPSEVADALKGMRNSAPGLDRITAEELLTWDHPSLAAYYNLMMAAGG PPEHLACSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDTLQLSPFQHGFLQR DGCLEATALLHTILRKVHNSRKPCAMLFLDVAKAFDTVSHQTLFRVAVEMGLPPPLVNYLKH	349

	LYSQSTVRLADQATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYAD DLVLLAEKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVLTPTTYECEGET IKPMETDDSVRYLGLHFNWKGRIVPKHTGKLDSLLKELTKAPLKPYQRLQLLKFHAVPKYIH ELVLGHAHRNTLKKLDCLTRAAVRQWLRLPKDTPLGYLHANVKDGGLGIPCFSTSIPLLQTK RFEKIATNPATIFQIMQRQDSFRTQGRRLDKPCRLNSTVVTSKTEVREEWGNMLSNSIDGRE LRHPEVDKASHEWLLRPDRVFPRLHMRGIQLRGGVLPTKARSGRGNNRQVTSKTCRGACQNV ETLNHILQVCEVTHDARCARHNRVLHRLERLLHKKGLRTTLEPVIPAGSSFIKPDLIIETDT HTYVLDVSVVAGYRMTESWDIKVEKYGRPDRVEAICKWAAIPATKRVEQLPVIVSNRGLCYQ PSGDGLRALGLTSRDISDLCLLTINGSLRCYDLYMRGTRAV	
(88) R2SisCat RT	CSRVTFIPKVENPQTPGDYRPISVASTVLRAFHKILAWRLRDTLQLSPFQHGFLQRDGCLEA TALLHTILRKVHNSRKPCAMLFLDVAKAFDTVSHQTLFRVAVEMGLPPPLVNYLKHLYSQST VRLADQATKCGRGVRQGDPLSPLLFIMVMDDIVRKTLPEVGFDLDGQRVDSLAYADDLVLLA EKSPRLQDKLHLLSEALRKAGMSLNARKSRGLTITKVSKRKQMVLTPTTYECEGETIKPMET DDSVRYLGLHF	350
(88) R2SisCat 5'UTR	CGACCATTTGGTGTTGATTTTGATACTGGCGACTGGCTAGTTGGTATTGGTTGAAAGCGCTG GCGAGGTTGAGCTGATAATTGGTTGCACGGGTTGGGCACCTCCTTGTGGTGACCGCTGGATA GGGCTTCGGCCTGAACCAAGATCCAACCCACCCTCCGTTGCTGGCAGACCCCAGTAACTGAG ATTAAACAGCTGGC	351
(88) R2SisCat 3'UTR	TAGGACTACCGAGCGCCACCGGAGGGGACCCAGAGGACCGTCTCGGTGATTAGGCGGTGACC ACCTGGGCTCTACCTGAAATAGCGATCAAATCGCCTTAGAGGGCCTGATTTGGGGGGTTGGG CAGACACTCAGTGTCATAACTGCCCTTTACCTTTCCACTTCTTGCCTGTCTAATAATAAATG TTAACA	352
(89) R2VipLat ORF	LDKASGPEPRSNPPSVAETAEAKLLLMFPCIDDPRPPGPPDRSSPGRRSRVACRESDAKDPG HVCPCGSEFQSKQGLAMHRYHRHMAEENAKRLASLPSKRLAWSQEEDRALLRLATAQWTGGM AKVTLFGLLQPQFPGRSVEAIKKRLQGLKWTAPADNPERDPLPVPPSTAQGDVEPATSPAPT ADPRVFDQNTDDPTVSWRRALLTAVEGKLSEPKMGADTLCRIASQLKNGEINGEACRALMNE HAREWFPHTWKPSSVKDQTDTRRWSKRAIRRANYATVQRLFKLRRRDAAHTVLSGRWRNAYR GQAPLPKGMDQYWSKVFEAGSAKDEKAQVVPDKTLWETLAAVTPSEVTSALKGMKASAPGLD RISAEDLLAWNQPSLAAYFNLMLAAGGPPEQLACSRVTFIPKVDEPKTPGDYRPISIASVIL RAFHKILAWRLRDTVQLSPFQHGFLQRDGCLEASALLHTILRKSHRDRKPCALLFLDVAKAF DTVSHQTLFRVATAMGLPPPLVDYLKHLYGRSTVQIAGQNTKCGRGVRQGDPLSPLLFIMVM EDIVRDALPEVGFDLDGQRVDSLAYADDLVLLAEKSSRLQEKLTLLTAALQNAGMELNARKS HGLTIEKASKRKCMVLRDTKYQCKGEGIQPMGTDDSVRYLGLHFNWKGRLVPKHTGKLDSLL KELSAAPLKPFQRLQLLKLHVVPKFVHELVLGHAHRNTLQKLDCLIRAAVRQWLRLPTDTPL GYLHANIKDGGLGIPCFSTSIPLLQQKRFEKIAQNPATIFQIIQRQDSFRTLGRRLDRPCRL QGAVLTSKAEAREEWGNVLSNSVDGRELRYPDVDKASHEWLLRPSRVFPRLHMRGIQLRGGL LPTKARSARGNNRQVQDKTCRGACRQPETLNHILQVCEVTHDGRCARHNRVLHRLAQLLRKK GVRTTLEPVIPSGSSFIKPDLIAETDKQTYVLDVSVVAGGRMKTSWDIKDEKYGRPDRVNDI RKWGAIPEASRVCQLPIIISNRGLCYQATGEGLRALGLTTREISDLCLLAISGSLKCYDLYM RGTRGVRGAPEHRYGAPKGPH	353
(89) R2VipLat RT	CSRVTFIPKVDEPKTPGDYRPISIASVILRAFHKILAWRLRDTVQLSPFQHGFLQRDGCLEA SALLHTILRKSHRDRKPCALLFLDVAKAFDTVSHQTLFRVATAMGLPPPLVDYLKHLYGRST VQIAGQNTKCGRGVRQGDPLSPLLFIMVMEDIVRDALPEVGFDLDGQRVDSLAYADDLVLLA EKSSRLQEKLTLLTAALQNAGMELNARKSHGLTIEKASKRKCMVLRDTKYQCKGEGIQPMGT DDSVRYLGLHF	354
(89) R2VipLat 5'UTR	CTCCCGTCGCTTGACATTGGCGTTTTTTCGCAGTGCGTATATCAATTGCGAGAGCTAGTCTT TGTCCTTGGCGTGAGTTGATTGCATAGTTGCACGGGTTAGACACCTCCTAGTGGTGACCG	355
(89) R2VipLat 3'UTR	TAGGTGGTGACCCCTCTGGTCGACACCTGGACTAGCGACCAAGTCGCCTTAGAAGACCTGAT GGGGGGGGTCTCCGTACCCTTTCTCGTCTTGTCTCTCTAATAATACATGTGAAAA	356
(90) R2VipUrs ORF	LLLMIPCIDDPRPPGPPDRSSPGRRSRVACRESDAKDPGHVCPCGSEFQSKQGLAMHRYHRH MAEENAKRLASLPSKKLAWSQEEDRALLRLATAQWKGGMAKVTLYGLLQPQFPGRSVEAIKK RLQGLKWTAPADTPERELPLMPPRAAQGDVEPATSPAPTADPTVFNQNTDDPTVSWRRALLT AVEGKLSEPKMGADTLRRIASQLKSGEINGEACRALMNEHAREWFPHMWKPSLVKDQTDTRC WSKRAVRRANYAKVQRLYKLQRRDAANTVLSGRWRNAYRGQAPLPKGMDQYWSKVFEAGSAK DEKAQVVPDKTLWETLAAVTPSEVTSALKGMKASAPGLDRISAEDLLAWNQPSLAAYFNLML AAGGPPEQLACSRVTFIPKVDEPKTPGDYRPISIASVILRAFHKILAWRLRDTVQLSPFQHG FLQRDGCLEASALLHTILRKSHRDRKPCAMLFLDVAKAFDTVSHQTLFRVATAMGLPPPLVD YLKHLYGRSTVQIAGQNTKCGRGVRQGDPLSPLLFIMVMEDIVRDSLPEVGFDLDGQRVDSL AYADDLVLLAEKPSRLQEKLTLLTAALQNAGMELNARKSHGLTIEKASKRKCMVLSDTKYQC KGEGIQPMGTDDSVRYLGLHFNWKGRLVPKHTGKLDSLLKELSAAPLKPFQRLQLLKLHVVP KFVHELVLGHAHRNTLQKLDCLTRAAVRQWLRLPTDTPLGYLHANIKDGGLGIPCFSTSIPL LQQKRFEKIAQNPATIFQIIQRQDSFRTLGRRLDRPCRLQGAVLTSKAEAREEWGNVLSNSV	357

	DGRELRYPDVDKASHEWLLRPSRVFPRLHMRGIQLRGGLLPTKARSARGNNRQVKDKTCRGA CHQPETLNHILQVCEVTHDGRCARHNRVLHRLAQLLRKKGVRTTLEPVIPSGSSFIKPDLII ETDTQTYVLDVSVVAGGRMKTSWDIKDEKYGRPDRVNDIRKWGAIPEASRVCQLPVIISNRG LCYQATGEGLRALGLTTREISDLCLLAISGSLKCFDLYMRGTRGVRRAPEHR	
(90) R2VipUrs RT	CSRVTFIPKVDEPKTPGDYRPISIASVILRAFHKILAWRLRDTVQLSPFQHGFLQRDGCLEA SALLHTILRKSHRDRKPCAMLFLDVAKAFDTVSHQTLFRVATAMGLPPPLVDYLKHLYGRST VQIAGQNTKCGRGVRQGDPLSPLLFIMVMEDIVRDSLPEVGFDLDGQRVDSLAYADDLVLLA EKPSRLQEKLTLLTAALQNAGMELNARKSHGLTIEKASKRKCMVLSDTKYQCKGEGIQPMGT DDSVRYLGLHF	358
(90) R2VipUrs 5'UTR	CTCCCGTCGCTTGACATTGGCGTTTTTTCGCAGTGCGTTTATCAATTGCGAGAGCTAGTCTT TGTCCTTGACGTGAGTTGATTGCATAGTTGCACGGGTTAGACACCTCCTTGTGGTGACCGCT GGATAAGGCCCCGGGCCTGAACCAAGATCTAACCCACCCTCCGTTGCAGAAACAGCGGAGGC TAAA	359
(90) R2VipUrs 3'UTR	TAAGGAGCACCCTTAGGACTGCACTAGGCGGTGACCCCTCTGGTCGACACCTGGGCTAGCGA CCAAGTCGCCTTAGAAGACCTGATGGGGGGGGTCTCCGTACCCTTTCTCGTCTTGTCTCTCT AATAATACATGTGAAAT	360
(91) R2AhaPra ORF	MDSSDKDGNAIGAGTQNRADNPRYPCFCGSTFRSKRGLGLHRAHKHPAVLNAERLAVLPTKR LTWTQEDDEALKRLSDKHWERGMLKKDLITLLTSYFPDRSFEAVKKRLQLTKWAPPEENPER LERPHRAETPERVDAHLADPTETLQSHGPDPTTATSWRRALLVAAADRFKETKLGAEELNQM VRDIQTGELSRKQCRARLRQHAKQWFPHKWKPSKPKPANASCLKNKKAIRRANYAMVQKLYR LRRRDAASTVLSGRWRNAYRGPAPLPPEMEEYWASVFQTPSANDDRSPAEPAAVLWSLVAAI TPSEVTSALNGMKDSAPGLDKIPAEELLTWDQPSLAAYYNLLLAAGGPPKHLSCSRVVFLPK TDEPSAPGDYRPISIASTILRAFHKILAWRLRDTLQLSPLQHAFLQRDGCLEASTLLHSIFR RIHTSRVPCAMLFLDVAKAFDTVSHKSLFRVAYEAGLPPPLVSYLQHLYKRSSVKLAETTKC GRGVRQGDPLSPLLFILVIDDIVKRSLPDVGFDFDGQRVGSLAYADDLVLFAEKSTRLQEKL NLLSEALQQAGMALNAKKSHGLTIARDGKRKCMALVPTTYECKGDTIQPLGTDDQVRYLGLQ FNWKGRIVPKHTGKLDSLLNELTKAPLKPHQRIKLLTTFLVPKFTHELVLGHAHRNTLKRLD CLTRAAVRRWLRLPTDTPTGYFHANVLDGGLGLPCFSTSIPLLQQKRLEKIALNPATLFQIV QRQDSFRSLAQRMDRPCRLGGAVATTKAEVREGWQIMLSASVDGKELSIPQVDKASHEWLRR PARVFPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRTCRGACRAPETQNHILQVCELTHDA RCARHNRVLRTLEKQLRRTEARTRMEPVIPSGSSFIKPDLIVEAANQMWVLDVSIVAGYRME ETWGIKARKYGREDRVADIRRWADVPEHTPVFQMPVVISNRGLLYGPSGEGLRAMGLSRRDL TDLCLLAISGSLMCFDLYKRGTKDWKD	361
(91) R2AhaPra RT	CSRVVFLPKTDEPSAPGDYRPISIASTILRAFHKILAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHSIFRRIHTSRVPCAMLFLDVAKAFDTVSHKSLFRVAYEAGLPPPLVSYLQHLYKRSS VKLAETTKCGRGVRQGDPLSPLLFILVIDDIVKRSLPDVGFDFDGQRVGSLAYADDLVLFAE KSTRLQEKLNLLSEALQQAGMALNAKKSHGLTIARDGKRKCMALVPTTYECKGDTIQPLGTD DQVRYLGLQF	362
(91) R2AhaPra 5'UTR	TGATATAGGACAATAGTCATAGTTACAGGCTTGGACGCCTCCGCGTGGCGACCGCCGGGGAG GGTACTTGGTACCCAAGCCAAGATCCAAGCCACCCCTCTGCATGCCATACCGCCCTAGCAGA GGGATTGCCACTGGTTACTTAACCATAGTATATAGTCCTAGTCCCTCTGGACAC	363
(91) R2AhaPra 3'UTR	TGAACAATCTGTTTTGTGTGTTGCTTGTATCCGTTGACATAACTGTCTTCTAGCGATTGTAT CGCCTTAGAAAGCCTGGCGGGCGGGAAGGGGATAGAGAGATGTATTATGCTTTATATGTATT GTATACTGATGTATTTTTGGTGTCCGAATACTGATGTATCTTACTGTAAATGTATCATGCAC TGCTATAACCTGTACTGTTTTGCCTTGCCTTTCTAAAAATAAACAGTTAAA	364
(92) R2ParGut ORF	LLVYLFIYRTSPSPPGTMDPSNKDGNASSAGTQDTNTRPSYRCFCGSQFNTQRGLGRHQAHR HPAQVNALRVAALPTKRLMWTQEDDEALKRHSDRLWQSGMTKTKLYPLLIPYFPERSLEAIK KRLLLTHWVPPIETSEQNPPQEHPERPENPDRPPTIQQPPTSPVAPPQNGPGCGNATSWRGA ILASAADRLKETKMGGEELAQIARDIQTGVISRKQCRARLNQHAKSWFPHKWKPSVTRPAGA NRQWSKKAIRRANYAEVQKLYKLRRRDAASTVLSGRWRNAYKGPAPVPPEMEEYWAAVFKTP SAIDDRVPAEPPAVLWSLITPIVPSEVTAALNGMKSSAPGLDRLSATELLSWDQPSLAAYYN LLLAAGGPPKHLSCSRVVFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPL QHAFLQRDGCLEASTLLHTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPP LVSYLRHLYKKSTVRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVG SLAYADDLVLFAERSTRLQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPY ECKGDAIQPLGTDDSVRYLGLQFNWKGRIIPKHTGMLDSLLNELSKAPLKPHQRLKLLSVHL VPKFTHELVLGHAHRNTLKRLDCLTRAAVRHWLRLPTDTPVGYFHAHVLDGGLGLPCFSTTI PLLQQKRLEKVASNPATLFQIVQRQTSFQTLESRMDRPCRLGGTVVTTKAEVREGWQIMLSN SVDGSELNISEVDKASHEWLRHPARVYPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRTCR GACCAPETQNHILQVCELTHDARCARHNRVLRCLDKLLRRKGLRTRLEPVIPSGASFIKPDL IVEVMNQLWILDVSIVAGYRVEETWDIKTAKYGRPDRVADIRRWAAVPATTPVVQHPVVISN RGLCYRPSGDGLRALGLSHRDTMDLCLLAIMGSLKCYDLYQRGTKDWNV	365

(92) R2ParGut RT	CSRVVFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAE RSTRLQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPYECKGDAIQPLGTD DSVRYLGLQF	366
(92) R2ParGut 5'UTR	AAGAGCGTAGTTGCAGGCTTGGGCACCTCCTCGTGGTGACCGCTGGATATAGTACTTCATGT ACTAGAACTAAGATCCAAGCCACCCCTCTACACGACTTTGCTTTAGTAGAGGGTTAG	367
(92) R2ParGut 3'UTR	TAGTATGTCTTGTAATGCTTATGCTTGTATGCTGTTACCTGTCTATTAGCGAGCAATCGCAC TTGGAGGGCCTGGCGGGCGGGAAAGGTCTAGTCGGTTGTATTTGCTATATAAACTTGTATAC CATGTATATTGATATCATGCATATGTGATGTATCCTGTTGTATCGTATTATGCACTGCTTTT CCATGTACTGTCTTGCCTTGCTTAAACCAAAAATAAACTGTATCC	368
(93) R2MasLat1 ORF	LLCRGLVTGLRIGFRVSPSPSGTMDPSNKDGNASSVGTQDNNTRPSYRCFCGSQFKTQRGLG KHQSHRHPAEVNAKRVAALPTKRLAWTQEDDEALKRHSDRLWQTGMTKTKLYPLLLPYFPER SLEAIKKRLLLTHWVPPNETPELNPPPEHPERPGSPDRSQPPNQQPPMSPVAPPLNGLGSGN ATSWRGNILASAADHLKESRMGSEELAQIARDIQTGVISRKQCRSRLNQHAKSWFPHKWKPS VNRAVETNRQWSKKAIRRANYAEVQKLYKLRRRDAASTVLSGHWRNAYKGPAPVPPEMEEYW AAVFKTPSAVDDRAPAEPPAVLWSLITPIAPSEVTAALNGMKASAPGLDRISAMELLSWDQP SLAAYYNLLLVAGGPPKHLSCSRVVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRD TLQLSPLQHAFLQRDGCLEASTLLHTILRRVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAY EAGLPPPLVSYLRHLYKKSTVRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFD FDGQRVGSLAYADDLVLFAERSTRLQEKLQLLSEALQRAGMALNAKKSHGLTVARDGKRKCM ALVPTTYECKGETIKPLGTDDSVRYLGLQFNWKGRIIPKHTGMLHSLLNELSKAPLKPHQRL KLLSAHLVPKFTHELVLGHAHRNTLKRLDCLTRAAVRAWLRLPTDTPVGYFHAHVLDGGLGL PCFSTTIPLLQQKRLEKVASNPATLFQIVQRQTSFQTLGRRIDRPCRLGGSVVTTKAEVREG WQIMLSNSVDGKELSISEVDKASHEWLRHPARVYPRLYIRGIQLRGGLLPTKARSARGNNRQ VTDRTCRGACCAPETQNHILQVCEVTHDARCARHNRVLRCLEKLLRRKDLHPRLEPVIPSGA SFIKPDLIVEVMNQLWILDVSIVAGYRVEETWDIKASKYGRPDRVADIRRWAAVPATTPVIQ LPVVISNRGLCYRPSGDGLRALGLSRRDTMDLCILAIIGSLKCFDLYQRGTKDWNV	369
(93) R2MasLat1 RT	CSRVVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTILRRVHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAE RSTRLQEKLQLLSEALQRAGMALNAKKSHGLTVARDGKRKCMALVPTTYECKGETIKPLGTD DSVRYLGLQF	370
(93) R2MasLat1 5'UTR	AAGAGCGTAGTTGCAGGCTTGGGCACCTCCACGTGGTGACCGCTGGATATGGTACTTCGGTA CCAGAACTAAGATCCAAGCCACCCCTCTACACGACT	371
(93) R2MasLat1 3'UTR	TATGTCTTGTTATACCTATGCTTGTATGTTGTTACCTGTCTATTAGCGAGCAATCGCACTTA GAGGGCCTGGCGGGCGGGAAAGGTCTAGTCGGTTGTATTTGCTGTATTAACATGTATACAAT GTATATTGATACCATGCAATGTGATGTATCTTGTGATATCGTATTATGCACTGCTTTACCCT GTACTGTTTTGCCTTGCTTAATCTAAAAATAAAGTGTAAAA	372
(94) R2MasLat2 ORF	MDPSNTDGNASSVGTQDNNNARSSYLCFCGSKFNSERGLGKHQAHRHPAEVNAKRVAALPTK RLMWTQEDDDALKRHSDRLWQMGMTKSRLYPLLLPFFPERSIEAIKKRLLITNWAPPIITSE LNPHEHPERRTNPERRTNLERLAHPKELTNPDRAAPLTQQPPTSPVAHPPNGLGSGNATSWR GNILASAAVHLKENRMGDEELAQIARDLQTGAISRKQCRSRLNQHAKRWFPHTWKPSATKTV KANRQWSNKAIRRANYAQVQKLYKLRRRDAASTVLSGRWRNAYKGQAPVPPEMEEYWAKVFN TPSAIDNSAPAEPSAVLWSLITPIVPSEVTAALNGMKASAPGLDRLPATELLSWDQPSLAAY YNLLLAAGGPPEHLSCSRVVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRDTIQLS PFQHAFLQRDGCLEASTLLHTILRRVHTERVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLP PPLVSYLRHLYKKSTVRLADTTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGLR VGSLAYADDLVLFAEKSTRLQEKLQLLSEALERAGMALNAKKSHGITIARDGKRKCMALVPT TYECKGETIKPLGTDDSVRYLGLQFNWKGRIIPKHTGMLDSLLNELSKAPLKPHQRLKLLSV HLVPRFTHELVLGHAHRNTLKRLDCLTRAAVRRWLRLPTDTPVGYFHAHVLDGGLGLPCFST TIPLLQQKRLEKVASDPATLFQIVQRQASFQTLGRRIDRPCRLGGSAVTTKAEVREGWQIML SNSVDGKELSIPEVDKASHEWLRHPARVFPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRT CRGACCAPETQNHILQVCELTHDARCARHNRVLRHTEKLLRRKGLRTRLEPVIPSGASFIKP DLIVEMTDQLWILDVSIVAGYRLEDTWDIKASKYGRPDRVTDIRRWAAVPDTTPVTQLPVVI SNRGLCYRPSGDGLRTLGLSRRDTMDLCLLAILGSLKCFDAYKRGTKDSNI	373
(94) R2MasLat2 RT	CSRVVFIPKTEDPKTPGDYRPISIASTILRAFHKILAWRLRDTIQLSPFQHAFLQRDGCLEA STLLHTILRRVHTERVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLADTTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGLRVGSLAYADDLVLFAE KSTRLQEKLQLLSEALERAGMALNAKKSHGITIARDGKRKCMALVPTTYECKGETIKPLGTD DSVRYLGLQF	374
(94) R2MasLat2	AAAGAGTGTAGTTACAGGCTTGGACACCTCCCAGTGGTGACCGCCGGATATGGTGCTTCGGC	375

5'UTR	ACCAGAGCTAAGAGCCAAGCCACCCCTCTACTTGACGATCGCTCTTGCAGAGGGTCTGTTAC TGGTCTACTAAATATCAATAGTAATAGTCCTAGTCCCCCCGGAACT	
(94) R2MasLat2 3'UTR	TAGTTATGTCTTGGACTTACGTTTTTGTGTTGTTACCTATCTCCTAGCGGAAATCGCACTTA GAGGGCCTGGCGGGCGGGATGGTCTTGCCGGTTGTATTTGTATTCCATATATGATGCACTGC TTTACCATGTACCGTTTTGCCTTGCTTTTCTAAAAATAAACAGTCAAC	376
(95) R2DipCan ORF	MDPSNKDGNASSVGSQDTNTRPSYRCFCGSQFNTQRGLGRHQAHRHPAQVNASRVAALPTKR LAWTQEDDEALKRHSDRLWQSGMTKTKLYPLLIPYFPERSLEAIKKRLLLTHWVPPIETSEH NLPQEHPERPENPDRPQTIQQPPTSPVAPPQNGPGCGNATSWRGALLASAADRLKETKMGGE ELAQIARDIQTGVISRKQCRARLNQHAKSWFPHKWKPSATRPAGANRQWSKNAIRRANYAEM QKLYKLRRRDAASTVLSGRWRNAYKGPAPVPPEMEEYWAAVFKTPSAIDDRVPAEPPAVLWS LITPIVPSEVTAALNGMKSSAPGLDRLSATELLSWDQPSLAAYYNLLLAAGGPPKHLSCSRV VFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEASTLL HTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKSTVRLA ETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAERSTR LQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPYECKGDTIQPLGTDDSVR YLGLQFNWKGRIIPKHTGKLDSLLNELSKAPLKPHQRLKLLSVHLVPKFTHELVLGHAHRNT LKRLDCLTRAAVRHWLRLPTDTPVGYFHAHVLDGGLGLPCFSTTIPLLQQKRLEKVASNPAT LFQIVQRQTSFQTLESRMDRPCRLDGTVVTTKAEVREGWQIMLSNSVDGSELNISDVDKASH EWLRHPARVYPRLYIRGIQLRGGLLPTKARSARGNNRQVTDRTCRGACCAPETQNHILQVCE LTHDARCARHNRVLRCLDKLLRRKGLRTRLEPVIPSGASFIKPDLIVEVTNQLWVLDVSIVA GYRMEETWDIKTTKYGRPDRVADIRRWAAVPATTPVVQHPVVISNRGLCYRPSGDGLRALGL SRRDTMDLCLLAIIGSLKCYDLYKRGTKDWNI	377
(95) R2DipCan RT	CSRVVFLPKTEDPKTPGDYRPISIASTILRAFHKIMAWRLRDTLQLSPLQHAFLQRDGCLEA STLLHTIFRRIHTDRVPCAMLFLDVAKAFDTVSHMSLFRVAYEAGLPPPLVSYLRHLYKKST VRLAETTKCGRGVRQGDPLSPLLFIMVIDDIIKSSLPDIGFDFDGQRVGSLAYADDLVLFAE RSTRLQEKLDLLSEALQRAGMALNAKKSHGITIARDGKRKCMALVPTPYECKGDTIQPLGTD DSVRYLGLQF	378
(95) R2DipCan 5'UTR	AAGAGCGTAGTTGCAGGCTTGGGCACCTCCACGTGGTGACCGCTGGATATAGTACTTCATGT ACTAGAACCAAGATCCAAGCCACCCCTCTACACGACTTTTGCTTTAGTAGAGGGTTGTTACT GGTTTACTTACTTATATAGTGTATTAGTCCAAGTCCCCCTGGAACT	379
(95) R2DipCan 3'UTR	TAGTATGTCTTGTTATGCCTATGCTTGTATGTTGTTACCTGTCTATTAGCGAGCAATCGCAC TTGGAGGGCCTGGCGGGCGGGAAAGGTCTAGTCGGTTGTATTTGCTGTATTGACTTGTATAC CATGTATTGTGATATAAATGCATATGCGATGTATCCTGTTGTACTGTATTATGCACTGCTTT ACCATGTACTGTCTTGCCTTGCTTAAACCAAAAATAAATGTTAA	380
(96) R2SibCol ORF	MLVRFKAIMKCLGPSGVRDNSSNGSLVRAMEAPNEEPSSSQGGASPSPFLCACGRQFKSQRG LSTHRRHKHPAEVNAERVAVIPTKRLHWSSEEDEALRRHADRSWVEGMLKKDLYELLLPHFP GRSAEAIKKRLQTIKWAKPSPGLGPVDLPPRPPVLPLGAADPVTPHQSPCSSALPSPPSPVF DPAAWKSALLDSALRAFQEPKLGAGVLRAITSALRSGGITDSQARDLLHQHAQEWFPYKWRP SVVRHKVPGKRSNKEIRRANYAVVQRLYKLRRKDAAATVLSGRWRQAYRGPAPVPAGMADYW SSVFETPSAADTTPAAHPPVLLWSVLGLIAPSEVTAALQSMTASAVGLDRISARDFLAMHQP SLAAYFNLLLVAGGPPCQLVRSRVVFIPKTDNPTNPGDYRPISVASVLLRALHKILAKRLRD TLHLSPFQHAFLQRDGCLEATTLLHAILRRVHTDRVPCAMLFLDVAKAFDTVSHQSLYRVAS AAGFPPPLLSYLTNLYKNSTIQLADRVIQCGRGVRQGDPLSPLLFIMVIDDILRASVPTMGF TLGDHKIDCLAYADDLVLLAERPARLQEKLDCLADALHATGMSLNPSKSFGLTITKGGKRKY MALLPTAYTYGGASICPLGPDDSVRYLGLQFNWKGRVVPKHTGLMEGMLNELSQAPLKPHQR LRILKFHLVPKFTHELVLGHVHRNTLVKMDRLLRSAVRRWLRLPSDAPLGYFHAHIKDGGLG VPCFSTSIPLLQKARLEKVASGGTMVPPSILQQEAFRVVSHRINRPCMMASAIVTTKAEVRT EWRNTLAQAIDGRDLCIADVDGASHDWLLHPSRVFPRLFIRGIQLRGGVLRTKSRSARGSNR MVVDRSCRGACRAPETLNHLLQVCETTHDARCARHNRVLRHLAHLFQRDGARCWVEPIIPSG PSFIKPDLLVQCNRELWVLDVAIVSGYRLRESWDLKIAKYGREDRVRDILSWSDGVTKVEQL PVILSNRGLCYEPSGGGLRTLGLSRRDVMDLCLLTIRGSLQCYDVYQRGTRAARAVT	381
(96) R2SibCol RT	RSRVVFIPKTDNPTNPGDYRPISVASVLLRALHKILAKRLRDTLHLSPFQHAFLQRDGCLEA TTLLHAILRRVHTDRVPCAMLFLDVAKAFDTVSHQSLYRVASAAGFPPPLLSYLTNLYKNST IQLADRVIQCGRGVRQGDPLSPLLFIMVIDDILRASVPTMGFTLGDHKIDCLAYADDLVLLA ERPARLQEKLDCLADALHATGMSLNPSKSFGLTITKGGKRKYMALLPTAYTYGGASICPLGP DDSVRYLGLQF	382
(96) R2SibCol 5'UTR	CATCTTTGAAGCTTGGGTAGGATCATAGTTACACGGCCTGGGCGCCTCCGCGTGGCGACCGC TGGGTAGGGTGCAGGGGAGTTTTCTCTCTGCATCTTAGCCAAGACCCAGGCCAACCCCCTGT CGGGTTAACCTCCCTACAGGGGTTTA	383
(96) R2SibCol 3'UTR	TGATCCTTCGAGCGGACATGGAATATGACCTGCTTTTAGCGGCGAAGACCCTGAATGAACTA CTGACTAGCGACGTGAGTCGCCTTGGTGGGTGGGCAGAGGGGGGATTACCTTTGTGATTTCC CTGTATTGTTCAACCCCCAAAAATAAATGTAAA	384

(97) R2GraSmi ORF	MHWQIAVQNLLLIMLKSHDSCLDPSGMQDDSNNDIHTGGRHEIPAVPFVCTCGMQFGTQRSL SMHRFHRHQDELNAERVAALPTRKLVWSREEDLALIRQAETLWQEVMLKKDLYQRLLPLFPG RSVEALKKRLQILKWRPVDNRGSGQTTTAVPPTSVDVPNLTPGAPFVGMPIAIPSFCTVTAP TRLWSMEEDDILRKNADAFHGSAKTKKELGQRLMLYLPGRSADAILKRLRTIKWAPPKLTSK VPTPLTTLTASETSTSSVMHTSLSVASLATGAPILDGKMMTVPVTSEVALGISRSPARAAAD FASQATPYDPVVLPIDSSPRRNGVMKSASSEATEIDLSLQLLWRDRLLRLVINTLREQRLGS DELRIIAQDLLSGETSLQQAGILLQGHAQRWFPHKWRPSVVRTPPLIKSWSKNKIRRSNYAI IQRLYATRPKDAASTVFSGNWRDAYRGSAHLPSEMDSYWSSIFETPSSTDDRISFPLSELQW SLMEPIAPSEVTAALRGMGSTAPGLDRISAVDLLSWHQPSVAAYFNLLLATESPPQHLACSR VTFLPKTSNPTLPGDYRPISVASVLLRAFHKIMALRLQDSVNISPLQHAFQQRDGCLEATTL LHSIFRLIHTKQVPCAMVFLDVSKAFDTVSHQTLFRVAADAGLPPPLISYLRNLYGSATVKM TNRFVHCGRGVRQGDPLSPILFLLVIDEILKSCLPQVGFDFDGHLIGSLAYADDLVLLARTP ARLQEKLDCLNDALRRAGMSLNVKKSFGMTIAKAGKQKHMVLLPKPYIVSSDSIYPLGPTET VNYLGLQFNWKGRVSVKHTGKLERLLCELTQAPLKPYQRLDLLKRFLVPKLTYELVLGTAHR NTLLKLDRLIRSSIRCWLRLPPDTPTAYFYAPIKSGGMGVPCFGTSIPLLQRARLAKLAENS GQLYKTLLIQSFFRSLQQRVEKPCRLNSVVVSSKIEVKGEWTTLLHQSVDGRELGAAGIDPA SHGWLLQPSGVFPRLYLRGLQLRGGLLSTKSRNARGVCRQVTDKSCKGSCRMVETINHILQV CEITHDARCARHNRVVQLLEKSLRLHCVRTWVEPIIPSVSSFLKPDLILETNDNFWVLDVSV VSGHRLQTTWDLKTAKYGSPDRVADIAVWTDKPFPIHQLPVIISNRGLILKASSAGLRTLGL SRRDIMDLCLLAIRGSLKCYDMYMRGTWTGCK	385
(97) R2GraSmi RT	CSRVTFLPKTSNPTLPGDYRPISVASVLLRAFHKIMALRLQDSVNISPLQHAFQQRDGCLEA TTLLHSIFRLIHTKQVPCAMVFLDVSKAFDTVSHQTLFRVAADAGLPPPLISYLRNLYGSAT VKMTNRFVHCGRGVRQGDPLSPILFLLVIDEILKSCLPQVGFDFDGHLIGSLAYADDLVLLA RTPARLQEKLDCLNDALRRAGMSLNVKKSFGMTIAKAGKQKHMVLLPKPYIVSSDSIYPLGP TETVNYLGLQF	386
(97) R2GraSmi 5'UTR	ACATAGTTACATGGTCTGGACGCCTCCTCGTGGCGACCGCTGGATAGGGGTGAGAGTGCTCA CCCTGAACTAAGATCCAGACCACCCT	387
(97) R2GraSmi 3'UTR	TGATTTAATGTGTCTCATGTACTGCTTACTTGTAATGTCTATTAGCGACTTTGTTCGCCTTA GGGAGTCTGTGTGCGGGGTGGGGGGCCCTTGATGGGATTCCTATTCTTGTATGCTGACTGTC TAAAAATAAATGTTTCC	388
(98) R2DipCan ORF	MVHLFFSPSPPGPPDSRSTDGKASNTGKLPREGETGYLCECGARFKTKRGLGQHRSHRHPAE VNAERVAAMPSKRLAWSLEDDEALQRHAYRHWKEGMMKKTLIAHIASHFPGRSQEALKKRLQ IIKWAPLPERPNSDVPERLNFRPDHPETEQLIQPTTSKPVLSTQTPEPSDETAANRRRALLA AAADHLKEDKMGGEELAQIARDLQSGETSKKAARSRLNQHAQRWFPHKWKPSVSRPERETCQ WSKSKIRRANYAIVQKLYKLRRKDAASTVLSGRWRNAYKGPTPEPPQMVDYWAEVFNTPSAK DDRSPVEPSTVLWSLMAAVVPSEVTAALNGMKSSAPGLDRISAVELLTWDQPSLAAYYNLLL AAGGPPKHLACSRVVFLAKKEAPTTPGDFRPISIASAILRAFHKIMAWRFRDTLQFSPLQHA FLQRDGCLEATTLLNTIFRKIHRDRVPCSMLFLDVAKAFDTVSHQSLFRVAREAGLPQPLVS YLRHLYKHSSVQLAGATTKCGRGVRQGDPLSPLLFIMVIEDAIRSSLPDVGFDLDGLRIGSL AYADDLVLFAEKSSRLQEKLDLLSGALQRAGMALNANKSHGITIASDGKRKCVALLPTQYHC KGDTIHPLETGDKVRYLGLQFNWKGRIVPKHTGLLDNMLNEITKAPLKPHQRIELLRYHLVP KFTHELVLGHAHRNTLKKLDCLSRAAVRHWLRLPKDTPTGYFHAHVRDGGLGLPCFSTSIPL MQQKRLEKVASCPANLFQVVQRQDSFRSLGRRMDRPCRLNNSVVSTKAEVRKEWQGMLSNTV DGIELTVPCVDKASHEWLRNPARVFPRLYIRGIQLRCGLLSTKARSARGTNRQVVDRTCRGP CRAPETLNHLLQICEITHDARCARHNRVLRCLDKMLRRADAVTSVEPVIPSGTSFIKPDLIL STPNELWVLDVAVVAGYRMRDSWDIKSAKYGRPDRVADILRWAEKPDTTPVYQLPVIFTNRG LCYEPTGEGLRTLGLSCRDIMDLCLQAIGGSLKIYDVYTRGTLTS	389
(98) R2DipCan RT	CSRVVFLAKKEAPTTPGDFRPISIASAILRAFHKIMAWRFRDTLQFSPLQHAFLQRDGCLEA TTLLNTIFRKIHRDRVPCSMLFLDVAKAFDTVSHQSLFRVAREAGLPQPLVSYLRHLYKHSS VQLAGATTKCGRGVRQGDPLSPLLFIMVIEDAIRSSLPDVGFDLDGLRIGSLAYADDLVLFA EKSSRLQEKLDLLSGALQRAGMALNANKSHGITIASDGKRKCVALLPTQYHCKGDTIHPLET GDKVRYLGLQF	390
(98) R2DipCan 5'UTR	AAGAGCATAGTTGCACGGTTCAGACGCCTCCCCGTGGCGACCGCTGGATAAGGTAGCAATAC CGAACTAAGATCTGAACCACCTCTCTACACGACTGATGCTTTGGTAGAGAGTTTGTTGCGGG A	391
(98) R2DipCan 3'UTR	TAGAGACTTTTCTGGGCGGATTCCTGTCCGTCACTGGGGCTTCCTTCCAGTAGCGATTTATT CGCCTTAGATTTGCTTGGTGGGCGGGATGTTGGTTGGTGCATGCCTCTTACCTGTATCGCCT TGCATTCTAAAAATAAAATGTGATTA	392
(99) R2MyxAsi1 ORF	LLVTDKKNDKEKPGPSGSRVYSRSVALARAGTSTQRTPPRTSFTCKDCDASFPSLRSLGMHR YHRHPEELNEERIHLLTTKRRMWSPLEDEHLIKMANEEWRESMLRKQHLRILHLILPQRSTE AIKKRLQTLKWTPPLLEEQLRQSPFSEVTPINPMMEEEHRQSPLSEVTPISPPLEVSISQKK NEKRGVRGKIWDAKEDKDLLEYADKIWEDGMLKKDLARCLFKRMEGRSVEAIKKRLQTLGWK	393

	PPDTAEGTRQEHPSITSPMLCQEMATSGQEKELVWRGRMLHKALEMLEDPQMGSEMLKVIAN DLLEGRINIAQGAAKLNAHVQHLFPIKWKLHQSKKVKDQPLTSNKRIRRARYAYIQRLYRVK KKDAAHTILQGHWREAYRGLNRTIDGLEEYWEKVYKATTTSTLCNKTSVTGDMKWDLIAPIE ATEVTEALRSMGRSAPGMDKVTSLQLLKWHQPSVAGLFNLFLITEALPTTLSCARVTFIPKV EHPVTPGDYRPIAISSVLARALHKILAKRMRENLTFSPLQYAFLQRDGCLEASALLNAIIRR AHDETRPMTALFLDLTKAFDTVAHDAILRSAEGAGMPPPLLKYLSHLYSHSEVQLGERKTTC GRGVRQGDPISPILFILTMGDIIQQVTPQIGLEIGEQTVAAIAYADDLIMFAEHPHQLQEKL DCLGVALENAGLSLNIKKSVAMTLVKDKRRKHMVLLPHYYSTRDGVVPPLNIADKQRYLGIE FGWKGKIIPKQTVDLERMLTEVTAAPLKPYQRLEILKTFLVPKLTHELVIGCAHRNTLAKMD GMIRREVRKWLRIPTDTPLAFLYAPINTGALGIPHLGSTIPLYQKARFQKLLANDNIFNKAL TGQPAFKTLLRRINIPCRIGHEIICNTNDGKEKWVDLLKVSVDGRDLLQAEVDKASFNWLER PDNVFPRLHLRGIQLRGGLLHTKSRAARGRGKDPEEIRCRGACHERETLNHILQICEITHDA RCARHNRVVKLLSKLLKGKTDMTLIEPVIPTRTSFIKPDIMVIKDNRITIMDVTIVAGYRME ESYRLKTLKYGSPDNIAAITEWTAIKTPITQRPIIISSRGLIYGPSGRALRSLKLSSRDLSD ISLLTIIGSLKCYDLYKRGTSNI	
(99) R2MyxAsi1 RT	CARVTFIPKVEHPVTPGDYRPIAISSVLARALHKILAKRMRENLTFSPLQYAFLQRDGCLEA SALLNAIIRRAHDETRPMTALFLDLTKAFDTVAHDAILRSAEGAGMPPPLLKYLSHLYSHSE VQLGERKTTCGRGVRQGDPISPILFILTMGDIIQQVTPQIGLEIGEQTVAAIAYADDLIMFA EHPHQLQEKLDCLGVALENAGLSLNIKKSVAMTLVKDKRRKHMVLLPHYYSTRDGVVPPLNI ADKQRYLGIEF	394
(99) R2MyxAsi1 5'UTR	GTTACATGGCCCGGGCGCCTCCTCGTGGCGACCGCTGGATAGGGTGGGGGAATCCGGTGATA TATTCGCCGTCCCACCTGAACTAAGATCCGGGCCGCCCCTATATGGGCCAACCACCCATATG GGGTAAC	395
(99) R2MyxAsi1 3'UTR	TAATTGAGGAAAGTACTATGCAGTATGACCTGCCTAGCAGCGATAACCTTGCCAGAGCTCAC AACTAGCGCCGGAAAGGAGCGCCTCAAGCATCAGGTGGATAATTCTGCTAGCTTGCTTGATA ACTTATTGTAAA	396
(100) R2MyxAsi2 ORF	LLVTDKKNDKEKPGPSGSRVYSRSVALARAGTSTQRTPPRTSFTCKDCDASFPSLRSLGMHR YHRHPEELNEERIHLLTTKRRMWSPLEDEHLIKMANEEWRESMLRKQHLRILHLILPQRSTE AIKKRLQTLKWTPPLLEEQLRQSPFSEVTPINPMMEEEHRQSPLSEVTPISPPLEVSISQKK NEKRGVRGKIWDAKEDKDLLEYADKIWEDGMLKKDLARCLFKRMEGRSVEAIKKRLQTLGWK PPDTAEGTRQEHPSITSPMLCQEMATSGQEKELVWRGRMLHKALEMLEDPQMGSEMLKVIAN DLLEGRINIAQGAAKLNAHVQHLFPIKWKLHQSKKVKDQPLTSNKRIRRARYAYIQRLYRVK KKDAAHTILQGHWREAYRGLNRTIDGLEEYWEKVYKATTTSTLCNKTSVTGDMKWDLIAPIE ATEVTEALRSMGRSAPGMDKVTSLQLLKWHQPSVAGLFNLFLITEALPTTLSCARVTFIPKV EHPVTPGDYRPIAISSVLARALHKILAKRMRENLTFSPLQYAFLQRDGCLEASALLNAIIRR AHDETRPMTALFLDLTKAFDTVAHDAILRSAEGAGMPPPLLKYLSHLYSHSEVQLGERKTTC GRGVRQGDPISPILFILTMGDIIQQVTPQIGLEIGEQTVAAIAYADDLIMFAEHPHQLQEKL DCLGVALENAGLSLNIKKSVAMTLVKDKRRKHMVLLPHYYSTRDGVVPPLNIADKQRYLGIE FGWKGKIIPKQTVDLERMLTEVTAAPLKPYQRLEILKTFLVPKLTHELVIGCAHRNTLAKMD GMIRREVRKWLRIPTDTPLAFLYAPINTGALGIPHLGSTIPLYQKARFQKLLANDNIFNKAL TGQPAFKTLLRRINIPCRIGHEIICNTNDGKEKWVDLLKVSVDGRDLLQAEVDKASFNWLER PDNVFPRLHLRGIQLRGGLLHTKSRAARGRGKDPEEIRCRGACHERETLNHILQICEITHDA RCARHNRVVKLLSKLLKGKTDMTLIEPVIPTRTSFIKPDIMVIKDNRITIMDVTIVAGYRME ESYRLKTLKYGSPDNIAAITEWTAIKTPITQRPIIISSRGLIYGPSGRALRSLKLSSRDLSD ISLLTIIGSLKCYDLYKRGTSNI	397
(100) R2MyxAsi2 RT	CARVTFIPKVEHPVTPGDYRPIAISSVLARALHKILAKRMRENLTFSPLQYAFLQRDGCLEA SALLNAIIRRAHDETRPMTALFLDLTKAFDTVAHDAILRSAEGAGMPPPLLKYLSHLYSHSE VQLGERKTTCGRGVRQGDPISPILFILTMGDIIQQVTPQIGLEIGEQTVAAIAYADDLIMFA EHPHQLQEKLDCLGVALENAGLSLNIKKSVAMTLVKDKRRKHMVLLPHYYSTRDGVVPPLNI ADKQRYLGIEF	398
(100) R2MyxAsi2 5'UTR	GTTACATGGCCCGGGCGCCTCCTCGTGGCGACCGCTGGATAGGGTGGGGGAATCCGGTGATA TATTCGCCGTCCCACCTGAACTAAGATCCGGGCCGCCCCTATATGGGCCAACCACCCATATG GGGTAAC	399
(100) R2MyxAsi2 3'UTR	TAATTGAGGAAAGTACTATGCAGTATGACCTGCCTAGCAGCGATAACCTTGCCAGAGCTCAC AACTAGCGCCGGAAAGGAGCGCCTCAAGCATCAGGTGGATAATTCTGCTAGCTTGCTTGATA ACTTATTGTAAA	400
(101) R2MyxAsi3 ORF	MGQPPIWGNLLVTDKKNDKEKPGPSGSRVYSRSVALARAGTSTQRTPPRTSFTCKDCDASFP SLRSLGMHRYHRHPEELNEERIHLLTTKRRMWSTLEDEHLIKVANEEWRESMLRKQHLRILH LILPQRSTEAIKKRLQTLKWTPPLLEEQLRQSPFSEVTPINPLPDEEHRQSPLSEVTPISPP QEVSISQKKNEKRGVRGKIWDAKEDKDLLEYADKMWVDGMLKKDLARCLFERMKGRSVEAIK KRLQTLGWKPLDTAEGTRQEHPSITSPMLCQETATSGQEKELVWRGRMLHKALGMLEDPQMG SEMLKVIANDLLEGRINIAQGAAKLNAHVQHLFPIKWKLHQSKKVKDQPLTSNKRIRRARYA	401

	YIQRLYRLKKKDAAHTILQGHWRDAYRGLNRTIDGLEEYWEKVYKATTTSTPCNKTSVTGDM KWELIAPIEATEVTEALRTMGRSAPGMDKVTSLQLLKWHQPSVAGLLNLFLVTEALPTTLSC ARVTFIPKVEHPLTPGDYRPIAISSVLARALHKILAKRMREYLTFSPLQYAFLQRDGCLEAS ALLNAIIRRAHEETRPMTALFLDLAKAFDTVAHAAILRSAEGAGMPPPLLKYISHLYSHSEV QLGKRKTTCGRGVRQGDPISPILFILTMGDIIQQVTPQIGLEIGEQTVAAIAYADDLILLAE HPQQLQEKLDCLGVALEDAGLSLNIKKSVAMTIIKDGRRKHMVLLPHYYSTRDGVVPPLNIA DKQKYLGIEFGWKGKITPKQTVDLERMLTEVTAAPLKPYQRLEILKTFLVPKLTHELVIGCA HRNTLAKMDGMIRREVRKWLRIPTDTPLAFLYAPINTGALGIPHLGSTIPLYQKVRFQKLLS NDNIFNKALTGQPAFKTLLRRINIPCRIGQEIICNTNDVKEKWVELLKVSVDGRDLLQAEVD MASFNWLERPDNVFPRLHLWGIQLRGGLLHTKSRAARGRGKDSEEIRCRGACHERETLNHIL QICEVTHDARCARHNRVVKLLSKLLKGKTDMTLIEPVIPTRTSFIKPDIMVIKDSRITIMDV SIVAGYRMEESYRLKTSKYGSPDNIAAITDWTATKTPITQRPIIISSRGLIYGPSGRALRSL KLSSRDLSDISLLTILGSLKCYDLYKRGTSNI	
(101) R2MyxAsi3 RT	CARVTFIPKVEHPLTPGDYRPIAISSVLARALHKILAKRMREYLTFSPLQYAFLQRDGCLEA SALLNAIIRRAHEETRPMTALFLDLAKAFDTVAHAAILRSAEGAGMPPPLLKYISHLYSHSE VQLGKRKTTCGRGVRQGDPISPILFILTMGDIIQQVTPQIGLEIGEQTVAAIAYADDLILLA EHPQQLQEKLDCLGVALEDAGLSLNIKKSVAMTIIKDGRRKHMVLLPHYYSTRDGVVPPLNI ADKQKYLGIEF	402
(101) R2MyxAsi3 5'UTR	GTTACATGGCCCGGGCGCCTCCTCGTGGCGACCGCTGGATAGGGTGGGGGAATCCGGTGATA TATTCGCCGTCCCACCTGAACTAAGATCCGGGCCGCCCCTAT	403
(101) R2MyxAsi3 3'UTR	TAATTGAGGAAAGTACTATGCAGTATGACCTGCCTAGCAGCGATAACCTTGCCAGAGCTCAC AGCTAGCGCCGGAAAGGGGCGCCTCAAGCATCAGGTGGATAATTCTGCAAGATGGCTTGATA ACTTATTGTAAA	404
(102) R2AcrFas ORF	LGLALLVTESKERNKDPGPSGNQEFSSGEAVGDIGMTLEPSPCSSASMNICKDCGGVFPSSR SLGMHRYHRHPVEVNQERIVEFQGRRKGWTSGEDELLRVSADSAWRQGMKRKEHLILLHDTF LHRTTEALKKRLQILKWTPPETSVTLQHHHSPTTEMITDQGHQITRTKEPTLTLETEPPIQA SGHLANLPCSRTAIKRGKAVKSWLEVDDMRLIHEATGLWVEGMTKRELVDKLRPYFPGRTAE AIRKRLQKVDWTPRHNSQEETSTKSVPQKPEGSPTHPREPPTRDTIAQSWPLDQAISPRNDR VSSPTVPPVQVESWRREMLEKAYRSMKDPKLHSEKFRHLIQNVLEGRVTRTEGARLFNEHAA KVFPTRWQPVGPKRNPRIKLGTQKQIRRATYASIQKLYKMRRKDAARTILEGRWKDAYKGQQ RVEGMEDYWENIYAGGVHEEYRPPTTERDEQKWEIISPITAEEVTGVLKGMGHTAMGMDRVS AKDLLAWDQPSLASFCNVTLALEVMPPSLANARVTFIPKVDVPESPGDFRPIAVSSVMTRAL HKILARRMRDHLQFSPLQYAFLKRDGCLEASTLLQALFRKTHDEVNPIAMLFLDIAKAFDTV AHATILETAREAGIPDPLLNYLGNLYQNSEVNLGSRMTKCGRGVRQGDPLSPLLFIMVMDKA VKAACPEIGAVIGGQHIDAIAYADDLILIAQSSEELQLKLDGLCKALNGMGMTLNERKSKAI TIIKDKRRKRLLLMPQSYHTTDGAIGAMGVVDEQRYLGLNFTWKGKVIPKHTGPLEKMLHEV TSAPLKPFQRLDILKNFLVPRLVYELVLGGAHRNTLLRMDRMIRAAVRKWMRLPKDTPLAYL HTSHRAGGLDIPSLSTAIPIWQRSRAEKLLSGTSPIDKAVTELPAFQKSLERMNRPCRVGGE AVTSIKEAREEWAQQLALSTDGKDLTVTEVDEASHLWLEKPERVFPRLHLRGIQLRGGVLPT KARGARGRGIEPQELRCRGACEATETLNHILQRCNVTHDARCARHNRVMRRLDIMLRKRASK VWMEPIIPTSTSFIKPDLLVDRDKTSMVMDVSIVAGSRMEETWKLKTRKYSTPENVKAIKAW LKSPTPLKHTPVIISFRGLMYGPSGRALRGVGLTKRDISDLCLLTIAGSLKTYDLYMRGT	405
(102) R2AcrFas RT	NARVTFIPKVDVPESPGDFRPIAVSSVMTRALHKILARRMRDHLQFSPLQYAFLKRDGCLEA STLLQALFRKTHDEVNPIAMLFLDIAKAFDTVAHATILETAREAGIPDPLLNYLGNLYQNSE VNLGSRMTKCGRGVRQGDPLSPLLFIMVMDKAVKAACPEIGAVIGGQHIDAIAYADDLILIA QSSEELQLKLDGLCKALNGMGMTLNERKSKAITIIKDKRRKRLLLMPQSYHTTDGAIGAMGV VDEQRYLGLNF	406
(102) R2AcrFas 5'UTR	GTTACGCGGCCCGGGCGCCTCCTCGTGGCGACCGCTGGATAGGGTGGGAAGCGATGGTAATT CCAACTACCGGCGGAGGGTGTCACGTCGGCACCCCCGTCTGCAACCCACCTGAACTAAGATC CGGGTCATCCCAGTCGGGCCAACCATCCGG	407
(102) R2AcrFas 3'UTR	GCTATGTAAAGTCAAACCAAGAAGAATGACCTGCCAGAGATGGCGGCGAATACCCCCCCCCA ACATACTCACCATTAGCGTCTCCTTAGACGCGTAGATAATAAGTGGTATAACATCTCCACCG CTATTGCTAAGGCAATATAACAAGTTGTTAA	408
(103) R2RhiKla ORF	LLVMKKKSKEKPGPSGSQAKSKCEAMVKAGETEANQDDISCCDCYRIFPTLRSLGMHRHHQH PEEVNKERLDLIQKRRKFWSPTEDHHLLTTADSLWKDDNLKKDHLRTLHTTLPHRSIDAIKK RLQLLKWTPPQSINTPPTPPIAEETPTTSSRRSRRGATLQPTGREVDLPKQPTNEDTSTTSR GSKANSRATAWTQDEDAHLLHLANTQWLPGMLKKDHCKALEEITNGRSMEAIRKRLKKLSWS PPEHMITITTSPAPPDTVIDQGRETTESNKPSPHSTKETDEDSWRNGILEKALESLSEPRIG SEGLKSIAKGMLEGRISKEQGVEMLQTHISQHIPLNWRTRNIKKGPSRKPMSNKQIRRARYA HIQKLYRTKRKDAANTILQGQWRDAFKGLEHGVDGVERYWADVYATPSNGNGQEVSIQRSEN TNWNILAPITAEDVKDALRSMGNTSPGMDKLTAKEISSWHLPSIAAMFNLFLATETLPSTLA	409

	CARVVLLPKVERPANPSDFRPIAISSTLARALHKVLSRRMREEYVFSPLQYAFLQRDGCLEA SSLLHAVLRQTQDEKRPLAAAFLDLAKAFDTVSHNAILTAAGKAGTPPPLIRYLSQLYTHAD LHIGSLVTKGVKGVRQGDPISPILFILAMGEVMEKACPEIGVKVGDQLIATMAYADDLILLA DNSELMQRKLDGLCVGLEESGMALNTKKSTAITIVKDGRQKTVALLPHRYNTPSGTLPPMGI KDTQKYLGIEFSWKGRVTPKQTAELEQMLREITEAPLKPYQRLEMVRTFVIPRLTHKLVLGC AHRNTLTKMDRMVRKATRAWLRLPKDTPLGFLHATVKHGGLAIPNLGTAIPLLQKARFEKLL SSKNQVEKALTNMPSFNTILRKVNLPCRVGKETVCSTEEAKAEWRNQLESSADGRENSTVDI DDGSFHWLENPDKVFPRLHLRGIQLRGGVLNTKARAARGREIPPKEQLCRGACKMKETLNHI LQKCEITHNARCARHNRIMKGAEKLLRKKAMETWIEPIIPTENSFIKPDLIIHTNNGYTVLD VSVVNNPRMELTYKLKTDKYGSEVNTMEITKWLDTTEQIKHVPIIVSSRGLMYGPSGRGLRA LGLTKRDISDLCLLTVLGSIKCYDLYTRGT	
(103) R2RhiKla RT	CARVVLLPKVERPANPSDFRPIAISSTLARALHKVLSRRMREEYVFSPLQYAFLQRDGCLEA SSLLHAVLRQTQDEKRPLAAAFLDLAKAFDTVSHNAILTAAGKAGTPPPLIRYLSQLYTHAD LHIGSLVTKGVKGVRQGDPISPILFILAMGEVMEKACPEIGVKVGDQLIATMAYADDLILLA DNSELMQRKLDGLCVGLEESGMALNTKKSTAITIVKDGRQKTVALLPHRYNTPSGTLPPMGI KDTQKYLGIEF	410
(103) R2RhiKla 5'UTR	CTGGTGAGAAATGGTTACACGGCCCGGGCGCCTCCTCGTGGCGACCGCTGGAGAAGGTGGGG GACTATAACGGTAAGTACGTCTACCGGCGATAGGGGACCACGTCGGTACCCATCGTGTCCCC CACCCGAGCTAAGATCCGGGCCACCCCGGGAAGGGCTAACCACCCAACCCGGGGGTAAA	411
(103) R2RhiKla 3'UTR	TGAATATATCTTCGGGATGGTCAAGCACTATGACCTGCATTGCAGCGATAACTATACCATTG CCAAACAATAGCACCTCCCTAGGTGCCTCAAGAATCTTCTGATGGGTTCCCACACAGAAGAT AATCTTGATAACTAATGTAAA	412
(104) R2MobBir ORF	LSSNNENNPGASAPGVDPRPESQIGYVCPDCGRAFRSRSGMSNHRRTHAAAGAVGGRFSCGI CSESFLSKAGLAQHTRHRHPVEHNIRRLEELPARRGRWTDQEDLGLVRVANELWLPGLGLAE LHRKLVPLVPGRSAEAIKKRLKILGWTPPVASPRPAEFEEEPAISAASPVEGPGWREALLEV ATHQLHSCQGDTYGAGEIRAIAELLKTGGISDPEAARRLEEHACLHFPICWRPTVPRQAHSG TAGELGRKAVRRAQYAHIQRLYHTRRKDAAQTVITGAWRNAHLQERALPPGTFAYWSTILSA EGVSDDRPVREVCPVNWDLMRPVEPSDISTAVGGVSGAAAGPDRISVDQIRRMDPSFLAAYL NIILCLGELPSPLKHSRITLVPKHSGDLSSPSDFRPINISSIILRIFHRILASRWAAVLELP TMQFGFLQRDGTFEAVTLLDSVLRRARERYLNAAMASIDVSKAFDSVSMDTILRSAQAFGAP PPLVRYIRSIYTGATAQLGDISISARRGVRQGDPLSPLLFIMAVDEVVALADPSTGIPFADR TLDAIAYADDLIVFSESAVMLQQKLARVDAALRLAGMRINMGKSSVLVIKGSRKKKISALGE GGVTLGDGQIRALGPTDEFRYLGVQFDWRGRRPVKHRELLMRALENVTRAPLKPHQRLTILR MHLIPKLVHSLVLGRAHRNTLKALDKMTRAAVRLWLRLPSDTSVAFLHAGLKDGGLGIPSFA ASIPIQKAARMERLLASSCPVLRAMVGLPEFAKTISQFTIPGTIMGNPVGSKREAVMAWRSA LAANLDCKYLRHVQVTRESRSWIEAPDRVFPRLFIRGIQLRANTLGTKVRRCRGRPGSVDLN CRGLCGAPESLHHIMQSCNVTHLARCARHNRVARLLAKSIRRGGFEVLEEPRIPTAGTFCKP DLIVKKGQAAVVIDVAVAIEERLFESWDSKRLKYGEPSTQSAIEAYLRNIDWDIQEVRHEPA VFGYHGLLYIKSAKTLAAMGLSRLTRADAALLVVRGSLGCYDVYMRGAGGGRGRDWDT	413
(104) R2MobBir RT	HSRITLVPKHSGDLSSPSDFRPINISSIILRIFHRILASRWAAVLELPTMQFGFLQRDGTFE AVTLLDSVLRRARERYLNAAMASIDVSKAFDSVSMDTILRSAQAFGAPPPLVRYIRSIYTGA TAQLGDISISARRGVRQGDPLSPLLFIMAVDEVVALADPSTGIPFADRTLDAIAYADDLIVF SESAVMLQQKLARVDAALRLAGMRINMGKSSVLVIKGSRKKKISALGEGGVTLGDGQIRALG PTDEFRYLGVQF	414
(104) R2MobBir 5'UTR	CTCTCGGGGGGGATGGGTGCAGTATTTTGCTGCACTCACACGGCCTGGGCGCCTCCTCGTTG GCGACCCCTGGAGATGGGGTGACTTCGGTCCCCTTGAGCTAAGATCCAGGCCACCTTCGGTG TGGGTTAACCTCCCGCACCGTTGTTTTTCTCGGGTT	415
(104) R2MobBir 3'UTR	CAACTGGATGATGTGGGTCTTTGTAAAGATTCTTGATGGATCTGGCTCTCGGGTCTACTGTA CACCTTCGGGTGTCTCTTGGAGTGTCGTATATTTATGAGTTTTTGTATGTGCATATGGTTTA TTCAGTGTAAGCTTTTGTACTTTTATGATTGTTTTGCATTGATGCCTGTGTATATATGCGGC CTTAAATGACATTCCAAAAATAAATATGTTAA	416
(105) R2PetZak ORF	MMITTKEKPRKTGASASEQTGGRKVFGADALSPMFACACGRSFGSQRGLTQHRRKAHPVDHN KHQVQRVTSARTVWSELECARLIWNADELVGSCATRAKLCTELHRLFPDRSVESIRKRLQLL RWTPPKHGVKDCPGLIYLPATETTELTEGTNGGRAETIADLTERTIGPQLVGQTEYPKIADV DLDATLPGSGGLDLDETLTAPFTSWTCASADTMIPDQGQLDSQELAWRSSLLGAVIADREKL PDHLVEIASGLQSGTVTDEWAVKELNSFVRNRYPHLWKTKVVRKIRRPRVLSNRLIRRMNYA AIQRLYRVKRGDAAKVVLNGGWRTAYKFDQGLPKGSMDHWREVLKRPTHRDVRKVDVVRETD WTLIEPVTAAEIRCSLRDMGDSSPGLDKLTVERVRKTEVLTLTGLCNLMLALGRVPEHLCAA RVTLIPKVDQPASPSDLRPITIASVLLRCFHKVLASRWSAKVQLDRLQFAFQKSDGCLQATT LLNVILRNASRLNSGLACCFVDVAKAFDSVSHDSIIRAATSFGAPLPLVRYLRSTYDLASCR LDDNVIRPKRGVRQGDPLSPLLFVMVIDEVLRRSRPGIGVKLDGCLVDAIAYADDLVLFAEN AAGMDEKLSVLSTALRDCGMVVNARKSRGLLIGGNRKEKLSALLRANFMIGGERVMPVEVDE KLSYLGIEFNWKGVTSTNHRWLLAGCLEEIRAAPLKPQQRMEILRTYALSRLQHSLVLGTAH	417

	ANTLKGLDRMVRASVRCWLRLMKDVPVGFFHTSIKEGGLGLPCFRTCIPLMKKARVDGLLAV DREAVRVAVRNPAYKPVLRQVKAPVRICGEEVENKQEARSAWRNSLLDSSAGTGVAEVCKSP ASTNWLVHPERVFPNVYVRAVQLRCNSVSTKVRNARLKRGPSEVTCRGLCRQPESLNHILQK CAVTHGGRCQRHNRVAKLLGKQLRGKGFLVVEEPRVLLAHSFCKPDLLVIRGSECTVCDVTI VGDKNLRTGYDDKVRKYKTAEVEAAIRAYVRDMLDLDVQLKHQPIVISAKGVMYEKTEAALR FFGLSMWSMSDVCILTMLGSMKCYDMYMRGFS	
(105) R2PetZak RT	AARVTLIPKVDQPASPSDLRPITIASVLLRCFHKVLASRWSAKVQLDRLQFAFQKSDGCLQA TTLLNVILRNASRLNSGLACCFVDVAKAFDSVSHDSIIRAATSFGAPLPLVRYLRSTYDLAS CRLDDNVIRPKRGVRQGDPLSPLLFVMVIDEVLRRSRPGIGVKLDGCLVDAIAYADDLVLFA ENAAGMDEKLSVLSTALRDCGMVVNARKSRGLLIGGNRKEKLSALLRANFMIGGERVMPVEV DEKLSYLGIEF	418
(105) R2PetZak 5'UTR	CCGCCAGGTGTGGTTCTGATCTCTCAGATGGGTCCTAGTAGGACGAATCATATCAAATACAG GTCTGTGAACTAGGGCGATGCTCGTATGTTTTATATAGGCCAAGTCGTGGTGCTGCGTCAGG TGTAGTCCGCTTCGGCGGTTCTGCATCTGGTGTAGTCCGCGGCTTCGGCAAATGAAGTTGAC AAAAGTGGCGGGCTACAGCTACTGCAGGTCAACAAGAAAAGCGGACAACAGAGGTGTCAAAC ATAGTTAGGAAATGAATCAGTCCGCTAAACGAACGCCCATTGAAGGGCTGGCTTGAGTGCTA ATCCATCTTGCCTCATACCAGGGTGTACGCATGAAAATGGATTAACGCCCGATGGTAGCCTC TTGCTGGTAATTAAACAGGTCTGTGTGTAAGGCCAACCTACGAGCTTGTTAGAGTAGTGGAT CCATTCTGAGGTGTGAAGTGCTGGCCTGGTGTGTGGGATGTCTGGTTGCTCAAGAGACCACC TCCTCGAGGTACCGCTGGATAAACCCAGGCTTGCCTGGGCGAACCAAGAGTCTCAGCATGTC CAGTAGCAAATCATGGTATATATATAATTGAGCCGACGTGCGATTGAACGAGGGTCCGGGGC TTGCCCTGTTAGCTAGCCTTCTCGATAGGCTGGTGTGACCCCTGTCTGTTGGGATCCTGTTT TGAGGGGTGCGGTGACGCAAATGGATCTCGGGTGTCGAGCTTGTGTTCTGGGAGGAGATTTT CTCCTTTTCATAGAGGGGGTTATTGTCGCTGAATATGCCCTACAGTGTCCGATGCCCTCCGT AAACACCGAATTGGCGGTGTCCGGGACATAGGCTTGGCAATGCCTTAGATCTAGCCCTTCTT AGCAGCCCGATGGGCTTTGACTGTTATTCGAGACCCCTTCCTTCCTCTCATTTTCATTCCTT GAATTCCAAGTGGATGCCAATCGAGGCGTTCTCTGACTGTGGTGGTTCTGATCTCTCAGGCA GGTCCTATACCTAGGGCCGAGGGAGCATATGCTCAAAATTACAGGTCTGTGTGTAGGGAATT GCTCGTGGGACAAGTTCTGGGTTGTGGATTTGTATTGGGTGCGGGCTCTGGGCGACTCTGCA TCTGGTGTGAGTCTGTAGCTTGGAAATGAGTTAGGCGGGCTTGTGTGTATACCTGCAACTTG TGCAACCAGGCGGTGGGGTATACGCAAGTTCGCAAAAGTGTGAAGTATGCGGACAATCAGTT GGGACCAAAGAGCTAGTTAGGATTGAATAAAGTCCGCTTTATAAATCTGCCCTTACGCTTGG CTGGTGAGATTGCCCTCCAGCTTGTGCTTCCGGACGGCCAAGGTCTTACCTTATATCAGTGT GCGTCCTCTAGCGCTTGATGGAATTCCGATCGAACTGGCTTTGTTGTACAGTCTGTGTTTAG GGTGTACCCACGAGTTCTAGAGTAGTGGAACTGTCCTGAGATGTGATGAGCTGCCACGGAAA GAGTCCGATTGGCTGGAGAGTGCACCTTCCTAAGGTACCACTGGAGATCTCGGCCTAGTCCG AGTGATCCAAGTCTCTCTCCGTGGGTCTCGCTTCCCGGTGGAGGGTGTCAAGCCTTCTGCCA AGCATCATGCTTGGGTTGCGAGATCAAGTGGACTGGCAATTGAAGCTGACGCTGATGCCGGA CTGCGCCTGCAGGGTGGTCATACCTGCTGCCGTGTGTGGTGACACACATGGCGAGGCCCTTA ACACTGGTGGCAGTGGGGTCTCTAGTGCATGCGGCGGGATTGATACCCCGTCCGGGCCACGC CCATATACTCTGTGGGAATTGTGCAGGCTGCTAAGCATTAGTAGCTCGGTGGTCTAGTCTCC TTGCTGGTTAGGGGAGGCTCCCTCGGGGGCTGAACCTGGCTGGGACGTTGTACCCGGAAAAG TACCCACCCGGGAGAGCTTAACTCCGGGGGTATGGCGTTGGGACCGGCAGTGGGGGGCGGGA CATTTGCCACCTATTGCGGAGCGTTATGAACGGGTGCTTAGTTGGGGGCACCCGTATTCATG TATCTAGCCCTTGCAACGCCCCTTAGTGGGCGGGTTAGAAATGCGTTGCAAACCTCCCCGAG GGCTCCAATTGCCCTGAGCGGAGACCACCACACTAATT	419
(105) R2PetZak 3'UTR	GCTCGTTGCATTTGTTCCCTCGTGATACGAGCAGCCGAAAGCCTTCGGGAAAGCGGCTTGTT GAATAAGAGACCGTACACTTGCCATTATAAAGATATACCGTCCATCAGACGGTGGGGCTCAA GTGGCTGTGTCGAAAGTTTCTTATATTATACTCTTTTCTTTGTTATTTATTGTCCCTATTAC AAAGTCAGAGACGAATGCCTCTAATGTTTGCATTGGCTGCTTGATAACAAGTTAAAC	420
(106) R2PetZak ORF	LKDPPSFGLTSLWWGKTLVTIQINAKAGVSTSGITSVGCGGTGENSGSFVCECGRGFGSQRG LSQHRRKCQAILYNSERRVLLAPKRGIRWSCQEDERLGALATDLSAECNTKRDLYKKLMEHF PGRTEDALKKRLQKVDWGGKRVTPPGGKCCGTRAGDVRIGSRNWGGSVSTPYSLDEGVPAEA SFDASSVAGDVCPDWKKCLLRSVLNCQGKPTEAEPKGGNLVDLAQVILNGEITNEDACCRLN EYVAAAFPLSWRPPKVRKPKQVPGLSRKKAIRRANYAVIQRLYRTRRKDAAHAILSGNWQKA YKMDRGLPGGSTGYWSNTLKRESFMDGRAIAAPQEVAWSLLAPITKSEVVSALKAMGKSSPG PDRVTPSAALRMNLQGITALLNLMLALESAPPHLCKSRVTLVPKGGTASSPGDLRPITVSSV IIRCMHKILAERWSSRLKLVNLQFGFLQRDGCLEATTVLNAILRYSSRTHADLACVMLDVSK AFDSVSHDSICRAAEANGAPLPLVSYLRSTYENAYALLDQEWIKTERGVRQGDPLSPLLFIM VVDEVLSLAKPELSASLNGQSVGAIAYADDLILFTENEARMREKLDAIRSALALAGMSVNCL KSRSLTLKGFRKRKTSVLEEVEYMLDGVKVPSVGPSDSFEYLGLLFDWKGVLKSPCKGRLAG MLDEVTKAPLKPQQRVTILREYLLPKLGYSLELCNIYGKSLKQMDKSVRKALRGWLRLASDT PVAYFYTSVRDGGLGIPCLSTQIPLRRRNRLDKVIASTCPAVQAAVTDSSYVPILRIVNGPV	421

	VVQGEVVRNKAEEREGWCKALYRTHDGAGLAGMRGSTASVQWLRAPERVFPRLFIRFIKLRG GLLSTKVRASRGPRKGDVACRGQCKHPESLAHILQKCDVTHDVRVARHNRVADKLAGMLKGL CFSVFCEPRVPLRGSFCKPDLLVVRDGIGYVMDVAVMGDDLGPTTWQRKVNKYGSVDCAAAI TSWLASVGQEVSCVRQLPVCISWRGVMFGQTERQLRLIGLRRYDLSDLVTLTMMGSLKTYDA YMRGAGGLPL	
(106) R2PetZak RT	KSRVTLVPKGGTASSPGDLRPITVSSVIIRCMHKILAERWSSRLKLVNLQFGFLQRDGCLEA TTVLNAILRYSSRTHADLACVMLDVSKAFDSVSHDSICRAAEANGAPLPLVSYLRSTYENAY ALLDQEWIKTERGVRQGDPLSPLLFIMVVDEVLSLAKPELSASLNGQSVGAIAYADDLILFT ENEARMREKLDAIRSALALAGMSVNCLKSRSLTLKGFRKRKTSVLEEVEYMLDGVKVPSVGP SDSFEYLGLLF	422
(106) R2PetZak 5'UTR	AGAGTCATTGGTCTCCTTAATCCCTTCCATTACACAAAGGTTGTGGTATCCAGGCTAACCGT CAATAATTAGGCAGGACCACCTCCACGTGGTGTGAGGCGGGCAACTGTTGGGTGCTTCGGTG CCTGGCAGGCCAAGAGTCCTGTTCTTAGCGGGGTACCTCTCCACCCGCTCTGCCTAGTGAGA TAGAGAGGACCTCTGAGGAGTCACGGCCCTCGGAGGTGCCCCCAAGTCCGTGACACCCCCAC TTGCCTATACTCGAGTTGTTGTGCCTTCTTACCCTCAGCGGTGCAGCTCGGGTGGCAAA	423
(106) R2PetZak 3'UTR	TGTGCCAGGCTGAAAGCCCTCGGGAAAGTCTGGTGTTGATGTGTACCCTTGTGACTCACCCT AGCTGGCGGTTTTCGAATCGCTTGCCCTCATGGGTGAAACAGGTCTGTGATCTTTGATGGTT GCCGGGCTGTTAGCCCATGAACAAGACGGATAC	424

Example 2. PCR Verification of Integration

[0171] This example used PCT to confirm that the newly identified R2 elements were capable of integrating themselves to the human genome. The tested R2 elements included R2AspMar (No. 23 in **Table A**), R2ATLH (No. 39), R2AspTig (No. 2), R2Vila (No. 45), R2PoEx (No. 12), R2Hal (No. 14), R2Age (benchmark), R2MelGeo (No. 8), R2MelCri (No. 35), R2PoAt (No. 37), R2TyAl (No. 44), R2HarHar (No. 32), R2ChPi (No. 38), R2Boja (No. 47), R2CapEur (No. 17), R2Bofu (No. 48), R2ColStr (No. 5), R2CroAda (No. 28), R2Tral (No. 46), R2CucCan (No. 29), R2CarCar (No. 10).

[0172] The mRNA templates derived from the full-length wild type R2 constructs were transfected into human HEK293T cell lines. Junctions at the 28S rDNA target site were analyzed by dPCR to assess genome integration activities. Both 5' and 3' junctions were measured. The Y axis reflects integrated R2 copy numbers per haploid genome (**FIG. 1**). As shown in **FIG. 1**, all of these constructed successfully inserted themselves to the HEK293T genome.

Example 3. Testing of new R2 Sequences

[0173] This example tested a few newly identified R2 retrotransposons for their ability in delivering transgene to a target cell, via *trans* or *cis* configurations.

[0174] In a *cis* configuration (**FIG. 2**, panel above), a transgene (e.g., GFP) is inserted downstream of the R2 ORF within the R2 retrotransposon. Upon insertion into the target genome along with the R2 ORF, the transgene can be expressed by the host. In a *trans* configuration

(**FIG. 2**, panel below), the transgene is flanked by the UTR sequences of the R2 (forming a "donor mRNA") for insertion into the target genome, while the R2 ORF is provided as a separate mRNA ("helper" mRNA) which is not inserted.

[0175] In a first batch, R2AtLH (No. 39), R2AspMar (No. 23), and R2AspTig (No. 2), were tested in the *cis* configuration and R2AtLH (No. 39), R2AspTig (No. 2), R2AspMar (No. 23), R2Vila (No. 45), and R2Boja (No. 47) were tested in the *trans* configuration. As shown in **FIG. 3**, all of them exhibited high integration efficiency.

[0176] In a second batch, R2AspTig (No. 2) and R2AtLH (No. 39) were tested. HEK cells were transfected with the mRNA in *cis* with a CMV-GFP-PolyA cassette carrying a cargo, or in *trans* with a CMV-GFP-PolyA cassette carrying a cargo and an R2 cassette encoding the R2 protein.

[0177] Cells were imaged and subjected to flow cytometry analysis at Day 3 post transfection, with no observed reduction in viability relative to transfection of GFP mRNA. As shown in **FIG. 4** (*cis*) and **FIG. 5** (*trans*), all tested R2 elements exhibited excellent gene delivery efficiency, in particular R2AtLH and R2MicRuf.

[0178] In a third batch, 14 new identified R2 elements, including R2ATLH (No. 39), R2MicRuf (No. 74), R2FasGig2 (No. 61), R2MobBir (No. 104), R2AriEle (No. 22), R2FasGig1 (No. 60), R2MasLat (No. 54), R2VipLat (No. 89), R2LiaOli (No. 53), R2OpiViv1 (No. 65), R2SchSol (No. 71), R2FasGig3 (No. 62), R2FasHep (No. 64), R2MyxAsi1 (No. 99), and R2Ptesp (No. 106), were tested for gene delivery. Cis constructs with GFP cargo were tested, and %GFP+ cells were measured by flow cytometry. As shown in **FIG. 6**, these R2 elements would effectively deliver GFP to the target cells.

Example 4. Gene Delivery with Templates Including Pseudouridine-Modified Nucleotides

[0179] In this example, six additional novel R2 elements were tested for gene delivery, including R2MicRuf (No. 74), R2ATLH (No. 39), R2FasGig2 (No. 61), R2FasGig1 (No. 60), R2MobBir (No. 104), R2VipLat (No. 89), and R2AriEle (No. 22). *Trans* constructs with GFP cargo were tested, and %GFP+ cells were measured by flow cytometry.

[0180] Here, all the RNA templates by *in vitro* transcription used pseudouridine-modified nucleotides. With such modifications, it appeared that all of the tested *trans* R2 constructs had excellent gene delivery efficiencies (**FIG.** 7).

* * *

[0181] Unless otherwise defined, 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.

[0182] The inventions illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising", "including," "containing", etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the invention claimed.

[0183] Thus, it should be understood that although the present invention has been specifically disclosed by preferred embodiments and optional features, modification, improvement and variation of the inventions embodied therein herein disclosed may be resorted to by those skilled in the art, and that such modifications, improvements and variations are considered to be within the scope of this invention. The materials, methods, and examples provided here are representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention.

[0184] The invention has been described broadly and generically herein. Each of the narrower species and subgeneric groupings falling within the generic disclosure also form part of the invention. This includes the generic description of the invention with a proviso or negative limitation removing any subject matter from the genus, regardless of whether or not the excised material is specifically recited herein.

[0185] In addition, where features or aspects of the invention are described in terms of Markush groups, those skilled in the art will recognize that the invention is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[0186] All publications, patent applications, patents, and other references mentioned herein are expressly incorporated by reference in their entirety, to the same extent as if each were incorporated by reference individually. In case of conflict, the present specification, including definitions, will control.

[0187] It is to be understood that while the disclosure has been described in conjunction with the above embodiments, that the foregoing description and examples are intended to illustrate and not limit the scope of the disclosure. Other aspects, advantages and modifications within the scope of the disclosure will be apparent to those skilled in the art to which the disclosure pertains.

CLAIMS:

1. A method for inserting an exogenous sequence to a genomic DNA of a cell, comprising introducing to the cell

(a) a polypeptide or a first polynucleotide encoding the polypeptide, wherein the polypeptide comprises

(i) a DNA binding domain or a CRISPR Cas protein,

(ii) a reverse-transcriptase comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 6, 2, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118, 122, 126, 130, 134, 138, 142, 146, 150, 154, 158, 162, 166, 170, 174, 178, 182, 186, 190, 194, 198, 202, 206, 210, 214, 218, 222, 226, 230, 234, 238, 242, 246, 250, 254, 258, 262, 266, 270, 274, 278, 282, 286, 290, 294, 298, 302, 306, 310, 314, 318, 322, 326, 330, 334, 338, 342, 346, 350, 354, 358, 362, 366, 370, 374, 378, 382, 386, 390, 394, 398, 402, 406, 410, 414, 418 and 422, or an amino acid sequence having at least 85% sequence identity to any of the amino acid selected from the group, and

(iii) an endonuclease or a nickase, and

(b) an RNA comprising (i) a template for the exogenous sequence and (ii) a 3' UTR, or a second polynucleotide encoding the RNA.

2. The method of claim 1, which comprises introducing the first polynucleotide and the second polynucleotide to the cell.

3. The method of claim 2, wherein the first polynucleotide and the second polynucleotide are on a same vector or on separate vectors.

4. The method of any preceding claim, wherein the 3'UTR is an R2 retrotransposon 3' UTR.

5. The method of claim 4, wherein the 3'UTR comprise a sequence selected from the group consisting of SEQ ID NO: 8, 4, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76,

80, 84, 88, 92, 96, 100, 104, 108, 112, 116, 120, 124, 128, 132, 136, 140, 144, 148, 152, 156, 160, 164, 168, 172, 176, 180, 184, 188, 192, 196, 200, 204, 208, 212, 216, 220, 224, 228, 232, 236, 240, 244, 248, 252, 256, 260, 264, 268, 272, 276, 280, 284, 288, 292, 296, 300, 304, 308, 312, 316, 320, 324, 328, 332, 336, 340, 344, 348, 352, 356, 360, 364, 368, 372, 376, 380, 384, 388, 392, 396, 400, 404, 408, 412, 416, 420 and 424.

6. The method of any preceding claim, wherein the reverse-transcriptase comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 6, 2, 10, 14, 18, 22, 26, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78, 82, 86, 90, 94, 98, 102, 106, 110, 114, 118, 122, 126, 130, 134, 138, 142, 146, 150, 154, 158, 162, 166, 170, 174, 178, 182, 186, 190, 194, 198, 202, 206, 210, 214, 218, 222, 226, 230, 234, 238, 242, 246, 250, 254, 258, 262, 266, 270, 274, 278, 282, 286, 290, 294, 298, 302, 306, 310, 314, 318, 322, 326, 330, 334, 338, 342, 346, 350, 354, 358, 362, 366, 370, 374, 378, 382, 386, 390, 394, 398, 402, 406, 410, 414, 418 and 422.

7. The method of any preceding claim, wherein the polypeptide comprises the endonuclease.

8. The method of any one of claims 1-7, wherein the polypeptide comprises the nickase, which optionally is Cas nickase.

9. The method of any preceding claim, wherein the CRISPR Cas protein is selected from the group consisting of Cas9, Cas12 and Cas13.

10. The method of any one of claims 1-8, wherein the DNA binding domain is selected from the group consisting of a transcription activator-like effector (TALE) DNA binding domain, a homeodomain protein, a zinc finger, a helix-turn-helix, a leucine zipper, and a DNA-binding domain of a homing endonuclease, a transposon or retro-transposon.

11. The method protein of claim 10, wherein the DNA binding domain is a TALE.

12. The method of claim 10, wherein the homing endonuclease is selected from the group consisting of I-SceI, I-CreI and I-DmoI.

13. The method of claim 10, wherein the retro-transposon is a long interspersed nuclear element (LINE), optionally selected from the group consisting of R1, R2, R4, R5, R6, R7, R8, and R9.

14. The method of any one of claims 1-5, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 5, 1, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69, 73, 77, 81, 85, 89, 93, 97, 101, 105, 109, 113, 117, 121, 125, 129, 133, 137, 141, 145, 149, 153, 157, 161, 165, 169, 173, 177, 181, 185, 189, 193, 197, 201, 205, 209, 213, 217, 221, 225, 229, 233, 237, 241, 245, 249, 253, 257, 261, 265, 269, 273, 277, 281, 285, 289, 293, 297, 301, 305, 309, 313, 317, 321, 325, 329, 333, 337, 341, 345, 349, 353, 357, 361, 365, 369, 373, 377, 381, 385, 389, 393, 397, 401, 405, 409, 413, 417 and 421, or an amino acid sequence having at least 85% sequence identity to any amino acid sequence selected from the group.

15. The method of claim 14, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 5, 1, 9, 13, 17, 21, 25, 29, 33, 37, 41, 45, 49, 53, 57, 61, 65, 69, 73, 77, 81, 85, 89, 93, 97, 101, 105, 109, 113, 117, 121, 125, 129, 133, 137, 141, 145, 149, 153, 157, 161, 165, 169, 173, 177, 181, 185, 189, 193, 197, 201, 205, 209, 213, 217, 221, 225, 229, 233, 237, 241, 245, 249, 253, 257, 261, 265, 269, 273, 277, 281, 285, 289, 293, 297, 301, 305, 309, 313, 317, 321, 325, 329, 333, 337, 341, 345, 349, 353, 357, 361, 365, 369, 373, 377, 381, 385, 389, 393, 397, 401, 405, 409, 413, 417 and 421.

16. The method of any preceding claim, wherein the RNA further comprises a 5'UTR.

17. The method of claim 16, wherein the 5'UTR comprises a sequence selected from the group consisting of SEQ ID NO: 7, 3, 11, 15, 19, 23, 27, 31, 35, 39, 43, 47, 51, 55, 59, 63, 67, 71, 75, 79, 83, 87, 91, 95, 99, 103, 107, 111, 115, 119, 123, 127, 131, 135, 139, 143, 147, 151, 155, 159, 163, 167, 171, 175, 179, 183, 187, 191, 195, 199, 203, 207, 211, 215, 219, 223, 227,

231, 235, 239, 243, 247, 251, 255, 259, 263, 267, 271, 275, 279, 283, 287, 291, 295, 299, 303, 307, 311, 315, 319, 323, 327, 331, 335, 339, 343, 347, 351, 355, 359, 363, 367, 371, 375, 379, 383, 387, 391, 395, 399, 403, 407, 411, 415, 419 and 423.

18. The method of any preceding claim, wherein the cell is a mammalian cell, preferably a human cell.

19. The method of any preceding claim, wherein the exogenous sequence encodes a protein, preferably a therapeutic protein.

20. The method of any preceding claim, wherein the RNA comprises pseudouridine-modified nucleotides.