Regulation of the vitamin B_{12} metabolism and transport in bacteria by a conserved RNA structural element

ALEXEY G. VITRESCHAK,^{1,2} DMITRY A. RODIONOV,^{1,3} ANDREY A. MIRONOV,^{1,3} and MIKHAIL S. GELFAND^{1,3}

¹Integrated Genomics-Moscow, 117333, Russia

²Institute for Problems of Information Transmission, Moscow, 101447, Russia

³State Scientific Center GosNIIGenetika, Moscow, 113545, Russia

ABSTRACT

Cobalamin in the form of adenosylcobalamin (Ado-CBL) is known to repress expression of genes for vitamin B_{12} biosynthesis and be transported by a posttranscriptional regulatory mechanism, which involves direct binding of Ado-CBL to 5' untranslated gene regions (5'UTR). Using comparative analysis of genes and regulatory regions, we identified a highly conserved RNA structure, the *B12*-element, which is widely distributed in 5'UTRs of vitamin B_{12} -related genes in eubacteria. Multiple alignment of approximately 200 *B12*-elements from 66 bacterial genomes reveals their common secondary structure and several extended regions of sequence conservation, including the previously known B12-box motif. In analogy to the model of regulation of the riboflavin and thiamin biosynthesis, we suggest Ado-CBL-mediated regulation based on formation of alternative RNA structures including the *B12*-element. In Gram-negative proteobacteria, as well as in cyanobacteria, actinobacteria, and the CFB group, the cobalamin biosynthesis and vitamin B_{12} transport genes are predicted to be regulated by inhibition of translation initiation, whereas in the *Bacillus/Clostridium* group of Gram-positive bacteria, these genes seem to be regulated by transcriptional antitermination. Phylogenetic analysis of the *B12*-elements reveals a large number of likely duplications of *B12*-elements in several bacterial genomes. These lineage-specific duplications of RNA regulatory elements seem to be a major evolutionary mechanism for expansion of the vitamin B_{12} regulon.

Keywords: Bacteria; comparative genomics; regulatory RNA; B12-element; cobalamin

INTRODUCTION

Synthesized only by prokaryotic organisms, vitamin B_{12} or cobalamin (CBL) is an essential cofactor for several important enzymes that catalyze a variety of transmethylation and rearrangement reactions (Martens et al. 2002). Expression of the *Salmonella typhimurium cob* operon, encoding the CBL biosynthetic pathway, and of the *btuB* gene of *Escherichia coli* and *S. typhimurium*, encoding the vitamin B_{12} by a post-transciptional regulatory mechanism (Lundrigan et al. 1991; Richter-Dahlfors and Andersson 1992). As shown by deletion analysis, this regulation requires unusually long 5'-untranslated leader sequences of the corresponding mRNAs, which contain several conserved elements. The leader mRNAs of the *cob* and *btuB* genes contain an evolutionarily

conserved sequence known as the B12-box (Franklund and Kadner 1997). Translational regulation of these genes requires also a conserved RNA hairpin that would mask the ribosome-binding site (RBS), thus inhibiting translation initiation and gene expression (Richter-Dahlfors et al. 1994; Nou and Kadner 1998). In addition, a *cis*-acting translational enhancer element in the cob leader mRNA is absolutely required to unfold the inhibitory RBS hairpin in the absence of coenzyme B₁₂ (Ravnum and Andersson 2001). In the same work, a secondary structure model for the cob leader mRNA was obtained from the chemical probing expriments with single-stranded RNA-modifying agents and combined with the output of a RNA folding computer program. No cobalamin-regulatory genes were identified in bacteria, but it was shown that adenosylcobalamin (Ado-CBL) is an effector molecule involved in the regulation of CBL genes (Nou and Kadner 2000). Recently, the structuredependent spontaneous cleavage of RNA technique was applied to the *E. coli btuB* leader sequence in the presence and absence of Ado-CBL, and it was shown that this sequence fragment can directly bind Ado-CBL, consequently undergoing conformational changes in the secondary and tertiary

Reprint requests to: Mikhail Gelfand, GosNIIGenetika, 1st Dorozhny pr., 1 Moscow 113545 Russia; e-mail: gelfand@ig-msk.ru; fax: +7-095-3150501.

Article and publication are at http://www.rnajournal.org/cgi/doi/10.1261/rna.5710303.

structure of the RNA (Nahvi et al. 2002). The investigators suggested that the mechanism of the *btuB* regulation involves formation of two alternative RNA structures, repressing and antirepressing, in the presence and absence of Ado-CBL, respectively.

The comparative analysis is a powerful approach to identification of regulatory patterns in bacterial genomes (Gelfand et al. 2000). Comparative sequence analysis was used for prediction of conserved RNA secondary structures (Eddy and Durbin 1994; Marck and Grosjean 2002) and detection of novel regulatory RNA elements, for instance, iron-responsive elements in E. coli (Dandekar et al. 1998) or S-boxes in Gram-positive bacteria (Grundy and Henkin 1998). In such studies, analysis of complementary substitutions in aligned sequences is used to construct a single conserved structure. The number of known noncoding RNA families is expanding rapidly. This resulted in development of a number of specialized databases, in particular the RNA family database Rfam (Griffiths-Jones et al. 2003), the database of known RNA structures RNABase (Murthy and Rose 2003), and the noncoding RNAs database (Szymanski et al. 2003). Current availability of multiple complete genomes provides an opportunity to identify consensus RNA elements upstream of co-regulated genes. In particular, highly conserved RFN and THI elements were identified in the regulatory regions of genes involved in the biosynthesis of two different vitamins, riboflavin and thiamin, respectively (Vitreschak et al. 2002; Rodionov et al. 2002). Recently it was confirmed that these RNA elements control expression of the target genes through a post-transciptional regulatory mechanism (Mironov et al. 2002; Winkler et al. 2002a,b).

Here we applied the comparative approach to analysis of 5'-untranslated regions (5'UTRs) of vitamin B_{12} -related genes in ~100 prokaryotic genomes. We report identification of a novel conserved RNA element involved in regulation of B_{12} -related genes in eubacteria. This new 5'UTR regulatory RNA, named the *B12*-element, is highly conserved on the sequence and structural levels and includes the previously defined B12-box motif. The *B12*-elements are widely distributed in bacteria: ~200 elements were identified in 5'UTRs of B_{12} -related genes in 67 bacterial genomes. The common structure of the *B12*-element was inferred and a possible mechanism of the *B12*-element-mediated regulation involving either transcriptional or translational attenuation was proposed for different groups of bacteria.

RESULTS AND DISCUSSION

Conserved structure of the B12-element

The *btuB* genes of *E. coli* and *S. typhimurium* have extensive regulatory regions including the conserved B12-box sequence (Ravnum and Andersson 1997; Nou and Kadner 2000). We started with identification of orthologs of *btuB* in

related bacteria. The upstream regions of btuB orthologs were aligned by the RNAMultAln program and the conserved RNA secondary structure was identified. This novel RNA structure, named the B12-element, consist of a number of helices and conserved sequence motifs, including the known B12-box. We constructed a pattern, which corresponded to the identified B12-element and scanned 107 genomic sequences using the RNA-PATTERN program. As a result, we found ~200 B12-elements in 67 bacterial genomes. Multiple alignment of these elements is shown in Figure 1. The B12-element is widely distributed in eubacteria, but it has not been observed in Archaea and Eukaryota. Most eubacterial genomes, containing the CBL biosynthesis or transport genes, have 1 to 13 B12-elements. The distribution of the elements is different in various taxonomic groups. On the average, Gram-negative α -proteobacteria have five B12-elements per genome, γ - and β -proteobacteria have two elements, Gram-positive bacteria have three elements, and cyanobacteria have only one B12-element per genome. All of these elements are located upstream of B₁₂-related genes. The detailed functional, positional, and phylogenetic analysis of the CBL genes is submitted elsewhere (D.A. Rodionov, A.G. Vitreschak, A.A. Mironov, and M.S. Gelfand, in prep.).

Previously, we described highly conserved 5'UTR structures, namely the RFN and THI elements, involved in regulation of riboflavin- and thiamin-related genes, respectively (Rodionov et al. 2002; Vitreschak et al. 2002). Similarly to these elements, the B12-element has a set of helices closed by a single base stem, and regions of high sequence conservation that are distributed along the entire element (Fig. 2). The conserved core of B12-element consists of seven helices (P0 to P6) and single-stranded regions with a high degree of sequence conservation. Existence of the conserved helices is confirmed by compensatory substitutions in base-paired positions. In addition to the conserved core, the B12-element has a number of facultative nonconserved stem-loops, designated Add-I and Add-II, and one internal variable structure, named VS. The previously defined B12-box is situated within the long internal loop between P0 and P4, slightly overlapping the base stem P0. In a number of cases a part of the B12 box can fold into a facultative stem-loop structure (Fig. 2). Other identified conserved regions of the B12-element are located upstream of the B12-box. The first one stretches from P1 to P4 and contains an AG-rich internal loop between P1 and P2 with consensus 5'-AAN AGGGAA-3'. Interestingly, AG-rich conserved regions were also observed in internal loops of the RFN and THI elements. Two other conserved regions, with consensi 5'-GC CACTG-3' and 5'-YGGGAAGGC-3', are located between P5 and P6, partially overlapping these helices.

Beside the conserved core structure, the *B12*-element includes additional nonconserved stem-loops, Add-I and Add-II. The former seems to be correlated with phylogeny, as it occurs only in genomes of proteobacteria. In contrast,

0,	L CICAC GAACAAC TCTGAC GAACAAC TCTGAC GAACAAC TCTGAC GAACAAC TCTGAC GAACAAAA TCTGAC GAACAAAA TCTGATCGATTCA TCCACTTCAGATTCA TCCCGTCAAAA TCCCGTCAAAA TCCCGTCAAAA TCCCGTCAAAA TCCCGTCAAAAA TCCCGTCAAAAA TCCCGTCAAAAA TCCCGTCAAAAA TCCCGTCAAAAA TCCCGTCAAAAAA TCCCGTCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	TCACCGTCAAC TTTCACCGTCAAC TTTCACCGAAAA TTTCACCGAAAA TCACCGCAAAAT TCACCGAAAAT CCAAAAATGCAA TCAAAATGCAA TCAAAATGCAA TCAAAATGCAA	Technic Trans Transcranser Cococcancer Cococcase Cococcase Cococcase Coccase Accestance Accestance Transcrand Transcrand Transcrand Transcrand Transcrand	Treat	Secendertor Teacrancertor Teacrancertor Adancestar Adancestar Teacrance Teacrance Adartocata Adarto	AGACACGGGGC ACGCGGGGCG GGACGTTTCGT GGACGTTTCGT GGAACGGACG GAACNATTTC CAACTGATTTT TAGTTCATTTTT TAGTTCCTTTTTT GGAACGAAG TGCAACGAAG GAAACGAAG TTCCTTCGTT
	er Chegy age Consol loc transformer loc consequence of loc consequence	ACCI, ESTANCET OPE ACCI, ACCI, ACCI, ACCI, OPE ACCI, ACCI, ACCI, ACCI, OPE ACCI, ACCI, ACCI, ACCI, OPE ACCI, ACCI, AC	acconstant accordent acconstant accordent accordent actor accordent accordent actor actor accordent actor actor accordent actor actor accordent actor actor accordent accord	ar Chiestata Comedia ac Chiestata Comedia ac Chiestata Comedia a Comestata Comedia a Comestata Comedia ac Chiestatata Comedia ar Chiestatata Comedia ar Chiestatata Comedia ar Chiestatata Comedia ac Chiestatatata Comedia ac Chiestatatatatatatatatatatatatatatatatatata	LIST CLAREAL ALCORDER LIST CLAREAL ALCORDER	In COLGEANIA COEGÓC IN COCGEANIA COEGÓC IN COCGEANIA COEGOC INCOCGEANIA COEGOC INCOCGEANAA COEGOC INCOCGEANAA COEGOC INCOCGEANAA COEGOC INCOCGEANIA COEGOC INCOCGEANIA COEGOC INCOCGEANIA COEGOCI
4	TEACODECIES TEACODECIES TEACODECIES AGACCOSECIES AGACCOSECIES AGACCOSECIES ACAPTOCALO ACAPTOCALO ACAPTOCALO COSECIES ATEATOCOSE ATEATOCOSE ATEATOCOSE ACACOCOSE AGACCOSE	TEAGCOGTGA AGACCOGGCA ATAGCOGCAA TGACCOGCAA TGACCOGCAA TGACCOGTGA TCACCOGTGA -TGACCOGTGA -TGACCOGTGA -TGACCOGCGA -TGACCOGCGA	TOTAL CONCEPTION TOTAL CONCEPTION TOTAL CONCEPTION TOTAL CONCEPTION AGAIN CONTRACT CONTRACTOR CONCEPTION AGAIN CONCEPTION AGAIN CONCEPTION CONCEPTION CONCEPTION CONCEPTION	CGAACCEC M AGACCEGCOGCIA AGACCEGCOGCIA TEACCEGCOAC ATTEACCEGCIA ATTEACCEGCIA AGACCEGCIA AGACCEGCIA AGACCEGCIA TEACCEGCIA TEACCEGCIA TEACCEGCIA TEACCEGCIA	TGACCOGTICA CGACCOGTICA CGACCOGTICA ATATICOGTICA TGACCOCOCICA TGACCOCOCICA ATATICOGCICA ATATICOGCICA ATATICOGCICA ATATICOGCICA	GTTTTTGFTC3 GACGTOGCG3 CGACGTOGCG3 CGACGTOGCG3 CGACGTGATA AATTCTTGCA3 AATTCTTGCA3 GATGTCTGGG3 GATGTCGGG3 GATGTCGGCG3 GATGTCGGCCG3 CGGTTCGGCC3
SA 1	C 2000 C	CT 10 CT 10 CT 10 CT 10 CT 10 CC 10	GG 10 GG 10 GG 10 GG 10 GG 11 CG 11 CG 11 10 CT 18 CG 24 CT 18 CG 24 CT 18 CG 24 CT 18 CG 24 CT 28 CT 18 CG 28 CT 28 CG	AC 10 (GA 9 (AG 9 (AG 9 (AG 9 (AG 9 (AG 10) (AG 10) (A	AC 12 AA 9 6GT 9 6GT 7 6GT 7 6GA 10 6A 10 6A 8 6A 8 6A 8 6A 8 6A 8 6A 8 6A 8 6A 8	Add 5 Add 5 Add 5 Add 5 Add 5 Add 5 Add 10 Add 10 A
is .	MGGG MGGG MGGG MGGG MGGG MGGG MGGG MGG	VAGGCO VAGGCO VAGGCO VAGGCO VAGGCO VAGGCO VAGGCO	MGGCG MGGCG MGGCG MGGCG MGGCG MGGCG MGGCG MGGCG MGGCG	Masso Maso Ma	VAGECG VAGECG VAGECG VAGECG VAGECG VAGECG VAGECG VAGECG VAGECG	MGGCG MGGCG MGGCG MGGCG MGGCG MGGCG MGGCG MGGCG
9						
ļ	CC660CC 			AACGG 		CGCCT CGCCC CGCCC CGCCC CCTTC CCTTC CCTTC CCTTC CCTTC CCTTC CCTTC CCTTC CCTTC CCCTTC CCCTTC CCCCCC
AddII	AA 69 - 119 A	GA-TGA A-TGA 29 - 29 -		10011111111111111111111111111111111111	118111	16 - 15 T
٦,	ACCENT CONTRACTION	GALLA GALCA GALLA GALLA GALLA GALLA GALLA GALLA			CCCAA CAGCAA CTTT- CTTT- CTTT- AA- CTTT- CAGA- CAGA-	CCCCACACACACACACACACACACACACACACACACAC
9	CT656 CT656 CT6566 CT6566 CT6566 CT6566 CT6566 CT6566 CT6567 CT6567 CT6567	CTGGA CTGGA CTGGA CTGGA CTGGA CTGGA CTGGA	C1666 C1666 C1666 C1666 C1666 C1666 C1666 C1666 C1666 C1666 C1666		creed creed creed creed creed creed creed creed creed creed creed	CT666A CT676 CT676 CT676 CT676 CT676 CT66A CT66A CT66A CT66A CT66A
5	gccaa gccaa	TGCCA TGCCA TGCCA TGCCA TGCCA	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GACCA A COCA A COCA GGCCA GGCCA A COCA A COC	TGCCA GGCCA GGCCA AGCCA AGCCA AGCCA AGCCA TGCCA TGCCA TGCCA	Tecca Tecca Tecca Tecca Tecca Tecca Tecca
SAI	11 66 11 66 10 CAA 10 CAA 10 CAA 11 67 12 CAA 11 07 10 000 0000000000	224 GCA 224 GCA 110 AA 111 CA 111 CA 112 CA 112 CA 112 CA 112 CA 112 CA 112 CA 112 CA 113 CA 114 CA	111 69 69 69 69 69 69 69 69 69 69 69 69 69	10 AA 9 CA 9 CA 9 CA 9 CA 10 AA 110 CA 110 CA 110 CA 111 AT 111 AT 111 GG 6 CA 6 CA 111 GG 111 GG 111 GG 111 GG 113 CA 113 CA 113 CA 113 CA 114 CA 11	10 CA 11 TC 11 TC 10 CA 11 GT 8 GA 8 GA 8 GA 11 TC 11 TC 11 CA 11 CA 11 CA 11 CA 11 CA 11 CA 11 CA 11 CA	553 CC 253 CC 253 CC 258 AC 258 AC 258 AC 355 AA 355 AC 356 CC 357 CC 35
1	CGAG AGAG AGAG AGAG AGAG GTTC GTTC CGAG AGAG A	CGAG CGAG CGAG CGAG CGAG CGAG CCGAG ATTG CATG	TGAA ATAA AGAG CGAG CGAG CGAG CGAG CGAG	CGAG CGAG CCAAG CCAAG CCAAG CGAG CCAAG CCCAAG CC	CGAG TGAG CGAG CGAG CGAG CGAG CGAG CGAG	AGGC CCTG CCTG CCTG CCTG CCCG CCCG CCCG
4	Carlos Contraction of the contra	AGCGG AGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	MGCGG MGCG MGCG MGCG MGCG MGCG MGCG MGCG MGCG MGCG MGCG MGC MGC	AGCGC AGCCC AGCGC AGCCCC AGCCCC AGCCCC AGCCCCC AGCCCC AGCCCCCCC AGCCCC AGCCCC AGCCCCCCCC	MACCAC MACCAC MACCAC MACCAC MACCAC MACCAC MACCAC MACCAC	MGCGG MGCCG MGCCG MGCAA MGCGA MGCGG MGCGG
		ACT 61 ACT 61 ACT 61 ACT 61 ACT 61 ACT 61 ACT 61	ACTER		ACTGT ACTGT ACTGT ACTGT ACTGT ACTGT ACTGT ACTGT ACTGT ACTGT	ACSET ACSET ACSET ACSET ACSET ACSET ACSET ACSET ACSET ACSET ACSET
ē (ecece coce coce coce coce coce coce coc					
-				77775555555555555555555555555555555555		55555555555555555555555555555555555555
ā.	RCdd CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TGCC TGCC TGCC TGCC TGCC TGCC GAAC				
ļ	total total	TATCG TATCG GGCCG GGCCG GGCCG ATTCG GGCCG GGCCG GGCCG		000000000000000000		
Tbbr	x x x x x x x x x x x x x x x x x x x 	********	GAAGGAAAA	ATCC GATCC CTTCC CTTCC CTTCC ATCC ATCC A	ATCC60 ATCC60 ATCC60 ATCC60 ATCC60 ATCC60 GGCC60 GGCC60 GGCC60 GGCC60 GGCC60	AAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
í		23 23 20 20 20 23 23 23 23 23 23 23 23 23 23 23 23 23	24 AAA 25 AAA 27 AAA 27 AAA 27 AAA 28 AA 29 AA 21 AA 20 AA 21 AA 20 AA 21 AA 21 AA 20 AA 21 AA 20 AA 21 AA 20 AAA 20 AA 20 AA 20 AA 20 AA 20 AAA 20 AA 20 AA 20 AA 20 AA 20 AA	221 AATCC 22 ACTCC 22 ACTCC 22 AATCC 22 AATCC 217 AAGCC 117 AAGCC 126 AATCC 116 AATCC 116 AATCC 116 AATCC 116 AATCC 116 AATCC 22 AATCC 116 AATCC 117 AATCC 116 AATCC 116 AATCC 117 AATCC 116 AATCC 117 AATCC 116 AATCC 116 AATCC 117 AATCC 116 AATCC 117 AATCC 116 AATCC 117 AATCC 117 AATCC 117 AATCC 118 AATCC 1	2 AGACCTO 28 AATCOCSO 16 ATTCOCSO 15 AATCOCSO 15 AATCOCSO 19 AGECCOSO 16 AGTCCSO 15 ATGCCOSO 15 ATGCCOSO 15 ATGCCOSO 15 ATGCCOSO 15 ATGCCOSO 15 AGGCCOSO 15 AGGCCO	22 CAACC 22 CAACC 22 CAACC 9 CCAACC 9 CCAACC 23 ATGC 23 ATGC 16 AATGC 34 CAACC 36 CAACC 2 ATGC 2 ATGC 2 CAACC 2 CAACC 2 CAACC 2 CAACC 2 CAACC 2 CAACC 2 CAACC 3 CAACC 2 CAACC 2 CAACC 3 CAACC 3 CAACC 3 CAACC 2 CAACC 3 CAACC
2	Contraction of the second of t	CCGNC 23 CCGNC 23 CCGNC 23 CCGNC 23 CCGNC 23 CCGNC 23 CCGNC 23 CCGNC 20 CCGNC 20 CCC	6664T 24 AAP C664T6 12 AAP C664T6 12 AAP C664T6 12 AAP C664T6 12 AAP C664T6 15 6A C664T6 15 6A C664T6 15 6A C664T6 15 6A C664T6 2 AAP C664T6 21 AAT C664T6 21 AAT	Constraint of a constraint of	Addition of the second of the	Modelan 22 CAARD Modelan 22 CAARD Modelan 27 CAARD Modelan 27 CAARD CAARD CAARD 22 ATTCA CAARD 22 ATTCA AGOTS 22 ATTCA CONTR 16 AATTCA CONTR 16 AATCA CONTR
	A coord A-coor	ATCC66T5 24 ATCC66T5 24 ATCC66T5 22 ACAC66T7 18 ATTC66T5 23 ATTC66T5 23 ATTC66T5 20 ATTC66T5 20 ATTC66T5 20 ATCC66T5 20 ACC66T5 20 ACC665	A	A CACGERATE 21 ANTOR A CACGERATE 21 ANTOR A CACGERATE 2 ANTOR A CACGERATE 2 ANTOR A CACGERATE 2 ANTOR A CACGERATE 2 ANTOR A CACGERATE 1 A AGE A CACGERATE 1 A AGE A CACGERATE 1 A ANTOR A TACGERATE 1 A ANTOR A ANTOR A ANTOR A ANTOR A ANTOR A ANTOR A ANTOR A ANTOR A ANTOR A ANTOR A ANTOR	ACategory 2 Adsocn AScotaris 2 Adsocn AScotaris 28 AATOOR ATAcentris 15 AATOOR ATacentris 15 AATOOR ABoografs 19 AATOOR ABoografs 16 AATOOR ATacestris 15 AGOOR ATacestris 16 AGOOR ATacestris 12 AGOOR ATacestris 12 AGOOR ATacestris 12 AGOOR ATacestris 12 AGOOR CagaA 20 GOC COCAACOR	AACKGR at 22 CAAC ACOCKGR 22 CAAC AACKGR 22 CAAC AACKGR 29 COCK AACCCKR 29 D200 AACCKR 21 A.COCK AACKGR 21 A.COCKGR 21 A.COCK AACKGR 21 A.COCKGR 22 A.COCK AACKGR 21 A.COCKGR 21 A.COCKGR 21 A.COCKGR 22 A.COCKGR 23 A.COCKGR 24 A.COCKGR
	addah contraction of the second of the secon	(ceedah - TCoedre 2 2 (ceedah - TCoedre 2 2 Adecah - TCoedre 2 2 (ceedah - TTCoedre 2 1 Adecah - TTCoedre 2 Adecah - CACerrh 12 Adecah - CCOerre 2 (ceedah - TCoedre 2 (ceedah - TCoedre 2 (ceedah - TCoedre 3) (adecah - CC)(cerre 3)	COSEGAA-TECCERT 24 AND COSEGAA-TECCERT 24 AND AGGEAA-TECCERT 24 AND AGGEAA-TECCERT 27 AND AGGEAA-TECCERT 12 AND AGGEAA-TECCERT 12 AND AGGEAA-TECCERT 2 AST AGGEAA-TECCERT 2 AST AGGEAA-TECCERT 2 AND AGGEAA-TECCERT 2 AND ADD AGGEAA-TECCERT 2 AND ADD ADD AGGEAA-TECCERT 2 AND ADD ADD ADD ADD ADD ADD ADD ADD ADD AD	Additional and a second and additional and additional a	MadeanACCategers 2 AdBACTR AddataCCategers 28 AATCOS MADEANCCCeders 15 AATCOS MADEANCCCeders 15 AATCOS MADEANCCCEGRS 15 AATCOS AddataCCCEGRS 15 AATCOS AddataCCCEGRS 15 AATCOS MADEANPLOCES 15 AGCOS MADEANPLOCES 15 AGCOS MADEAN-PLOCES 15 AGCOS	CoocanKcheera 22 chack CoocanCcheera 22 chack CoocanCcheera 21 chack CoocanCcheera 21 chack CoocanCcheera 21 Arroc NocanhCcheera 21 Arroc NocanhCcheera 21 Arroc NocanhCcheera 21 Arroc CoocanhCcheera 21 chack CoocanhCcheera 22 chack CoocanhCcheera 22 chack CoocanhCcheera 22 chack
	aa addatxa aa addatxa Criityakidaedaa - erooarro (Criityakidaedaa - erooarro (Criityakidaedaa - erooarro - axartosedaa - erooarro - axartosedaa - erooarro - eraatrosedaa - erooarro	-AAAAGGGAATOCGATG 22 -AAAAGGGAATOCGATG 22 -AAAAGGGAATOCGATG 22 -AAAAGGGAATOCGATG 22 -AAAAGGGAACACGATT 12 -AAAAGGGAACACGATG 20 -TAAAAGGGAACACGATG 20 -TAAAAGGGAACACGATG 20 -TAAAAGGGAACACGATG 20 4 AATCGGGAACOTGATG 20 8 AAAACGGGAACOTGATG 30		-TryAnylescia, -Cacegris 21 Auro -TryAnascia, -Cacegris 21 Auro -TryAnascia, -Cacegris 2 Auro -Cananoscia, -Cocegris 2 Auro -Cananoscia, -Cocegris 2 Auro -Cananoscia, -Cocegris 2 Auro -Cananoscia, -Cocegris 1 Aaso -Cananoscia, -Teccorgis 17 Auro -Cananoscia, -Teccorgis 1 Auro -TraAnoscia, -Teccorgis 1 Auro -Tananoscia, -Teccorgis 1 Auro -Tananoscia, -Teccorgis 1 Auro -Tananoscia, -Teccorgis 1 Auro	-GAAAAGGGAACCRAGET 2 AGACOTT 	-CMACGGGAA - ACI, 66GAN 22 CAAR -TMACGGGAA - CCC 66GAN 22 CAAR -TMACGGGAA - CCC 66GAN 21 CAAR - CAAGGGGAA - ACC 66GAN 21 CAAR - MAGAGGGAA - CCC 76GAN 21 CAAR - MATAGGGAA - CCC 76GTG 22 ATTGC - AAATAGGGAA - TCC 66TG 22 ATTGC - AAACGGGAA - CCG 66GAN 22 CAAR - AAACGGGAA - CGC 66GGAN 22 CAAR - AAACGGGAA - CGC 75 CAACGAA - AAACGACGAA - CGC 75 CAACGAA - AAACGGGAA - CGC 75 CAACGAA - AAACGGGAA - CGC 75 CAACGAA - AAACGACGAA - CGC 75 CAACGAA - AAACGACGAAA - CGC 75 CAACGAA - AAACGAAACGAAA - CGC 75 CAACGAA - AAACGACGAAA - CGC 75 CAACGAAACGAAACGAAACGAAACGAAA - AAACGAAAA - CGC 75 CAACGAAACGAAACGAAACGAAACGAAACGAAAACGAAAACGAAAAACGAAAACGAAAACGAAAACGAAAACGAAAAACGAAAACGAAAACGAAAAACGAAAAACGAAAACGAAAACGAAAACGAAAAACGAAAACGAAAAACGAAAAACGAAAAAA
1,	xg aa addatya cgfggg xg aa addatya coreans conscrattleaceata - translaterata - translaterata addracata - translaterata - translaterata	ANCT-PARAGOGANTOCOGFIC 22 ANCT-PARAGOGANTOCOGFIC 22 ANGC-TANANGGGANTOCOGFIC 22 ANGC-TANANGGGANTOCOGFIC 23 ANGC-TANANGGGANCACGEFIC 13 GGTGANANGGGANCACGEFIC 22 GGC-TANANGGGANCACGEFIC 22 GGC-TANANGGGANCACGEFIC 22 GGC 4 ANTCAGGGANCOTOGFIC 2 GGC 4 ANTCAGGGANCOTOGFIC 3 GGC 4 ANTCAGGANCOTOGFIC 3 GGC 4 ANTCAGGGANCOTOGFIC 3 GGC 4 ANTCAGGANACOTOGFIC 3 GGC 4 ANTCAGGANACOT	642-TAMTGGGAAA-TGGGGAT 24 ANY 6424-TAMTGGGGAAA-TGGGGT 24 ANY 6627-GAAAGGGAAA-TGGGGT 27 ANY 6687T-GAAAGGGAAA-TGGGFT 27 ANY 6687T-GAAAGGGAAA-TGGGFT 12 ANY 6684T-GAAAGGGAAA-GCGGTT 2 14 ANY 6684C-GAAAGGGAAA-GCGGTT 2 15 ANY 6684-GAAAGGAAA-CGCGGTT 2 15 ANY 6684-GAAAGGGAAA-CCCGGTT 2 15 ANY 6684-GAAAGGGAAA-TCGGTT 2 12 ANY 6684-GAAAGGAAA-TCGGTT 2 2 ANY 6684-GAAAGGGAAA-TCGGTT 2 2 ANY 6684-GAAAGGGAAA-TCGGTT 2 2 ANY 6684-GAAAGGAAA-TCGGTT 2 2 ANY 6684-CGAAAGGAAA-TCGGTT 2 2 ANY 6684-CGAAAGAAA-TCGGTT 2 2 ANY 6684-CGAAAGAAA-TCGGTT 2 2 ANY 6684-CGAAAGAAA-TCGGTT 2 2 ANY 6684-CGAAAGAAA-TCGGTT 2 2 ANY 6684-CGAAAGAAA-TCGTT 2 2 ANY 6684-CGAAAAGAAA-TCGTT 2 2 ANY 6684-CGAAAGAAA-TCGTT 2 2 ANY 6684-CGAAAGAAA-TCGTT 2 2 ANY 6684-CGAAAAGAAA-TCGTT 2 2 ANY 6684-CGAAAAAAGAAA-TCGTT 2 2 ANY 6684-CGAAAAAAGAAA-TCGTT 2 2 ANY 7000000000000000000000000000000000000	eschr-travitseischuchoretris 11, Arging eschr-travitseischchoretris 14, Arging esch-travitseischchoretris 2, Arthol esch-travitseischacorestris 2, Arthol esch-travitseischacorestris 2, Arthol esch-travitseischabroestris 11, Alson esch-travitseischafroestris 11, Alson esch-travitseischafroestris 11, Alson esch-travitseischafroestris 11, Alson esch-travitseischafroestris 11, Alson esch-travitseischafroestris 16, Arthol esch-travitseischafroestris 16, Arthol esch-travitseischa-froestris 16, Arthol esch-tra	early canadesta career 2 Adactive care canadesta career 2 Adactive early canadesta creer 18 Arroos early canadesta tacer 15 Arroos Ader-canadesta troegre 15 Arroos Ader-canadesta troegre 19 Arroos Ader-canadesta troegre 19 Arroos early Andoseaa troegre 14 Arroos early Andoseaa troegre 24 Accord early Andoseaa troegre 24 Accor	a. AddrCAAACGGGAAFOXGGGAA 22 CAACG 6 6677TAAACGGGAAFOXGGGGA 22 CAACG 6 6572CAACGGGAACOCCGGGA 9 D02AC 6 6572CAACGGGAACOCCGGGG 9 D02AC 0 5 G572CAACGGGAACOCCGGGG 9 D02AC 0 5 G572AACGGGGAACOCCGGGG 9 D02AC 0 5 G672AAACGGGAACOCGGGGG 9 D02AC 8 100-100-100-100-100-100-100-100-100-100
1 1'	CY Tg aa adddAA cgtg CC 12 cacc -TAALacdarA -PCCeRTIG CC 15 cacc -TAALacdarA -PCCeRTIG CC 15 cacc -PACAAAGGGAA -PCCeRTIG CC 15 cacc -PACAAGGGAA -PCCeRTIG CC 15 cacc -PACAAGGGAA -PCCeRTIG CC 11 cacc -CAATYCGGGAA -PCCeRTIG CC 11 cacc -CAATYCGGCAA -PCCCGGTIC CC 11 cacc -CAATYCGGCAA -PCCCGGTIC CC 11 cacc caATYCGGCAA -PCCCGGTIC CC 11 cacc caATYCGGCAA -PCCCGAAT CC 11 cacc caATYCGGCAA -PCCCGAAT CC 11 cacc caATYCGGCAA -PCCCGAAT CC 11 cacc cacccaaTA -PCCGAAT CC 11 cacccaaTA -PCCGAAT C	CUT 19 ACT-AAAGGGAA-PCCGAFG 22 TT 19 ACT-AAAGGGAA-PCCGAFG 22 TT 19 ACT-AAAGGGAA-PCCGAFG 22 TT 16 AIG-TAAAGGGAA-PCCGAFG 22 CUT 20 ACT-AAAGGGAA-PCCGAFF 12 CUT 15 GFT-CAAAGGGAA-CACGAFG 22 CUT 15 GFT-TAAAGGGAA-CACGAFG 22 CUT 15 GFT-TAAAGGGAA-CACGAFG 22 CUT 15 GFT 8 AAGAGGGAA-CACGAFG 22 CUT 12 GFT 8 AAGAGGGAA-CACGAFG 22 CUT 12 GFT 8 AAGAGGGAA-CACGAFG 22 CUT 12 GFT 8 AAGAGGGAA-CACGAFG 23 CUT 12 GFT 8 AAGAGGGAA-CACGAFG 23	CC 22 64 "TANTCGGGAL - TGGGGAT 24 AN TO 12 66 CAAAAGGGAA TGGGGAT 24 AN CC 12 66 CAAAAGGGAA TGGGAT 21 AN CC 12 66.T- CAAAAGGGAA TGGGAT 27 AN CC 12 66.T- CAAAAGGGAA TGGGAT 12 AN CC 13 66.T- CAAAAGGGAA CCGGTG 2 AN TC 14 66 CAAAGGGAA CCGGTG 2 AN CC 14 66.C- CAAATAGGGAA CCGGTG 2 AN CC 24 66.GGAAAAGGGAA CCGGTG 2 AN CC 24 66.GGAAAAGGGAA TANGGTG 2 AN CC 24 66.GGAAAAGGAAA CCGGTG 2 AN CC 24 66.GGAAAAGGAAA CCGGTG 2 AN CC 24 66.GGAAAAGGAAA CCGGTG 2 AN CC 24 66.GGAAAAGGAAA TANGGTG 2 AN CC 24 66.GCAAAAGGAAA TANGGTG 2 AN CC 24 66.GCAAAAGGAAA TANGGTG 2 AN CC 24 66.CCAAAGGAAAA TANGGTG 2 AN CCAAAGGAAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAAA TANGGTG 2 AN CCAAAGGAAAAGGAAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAAA TANGGAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAAA TANGGAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAA TANGGAAA TANGGAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAAA TANGGAAA TANGGTG 2 AN CCAAAGGAAAA TANGGAAAA TANGGAAA TANGGAAA TANGGAAA TANGGTG 2 AN CCAAAGGAAAA CCAAAGGAAA TANGAAAA TANGGAAA TANGGAAAA TANGGAAAA TANGGAAAA TANGGAAAA TANGGAAAA TAN	CC 10 GGT-TANTAGEAAA-CARGEGTS 21 AATTOC CC 12 GGTT-TGTAAAGGGAAA-CARGEAAA 14 AGA CC 12 GGTT-TGTAAAGGGAAA-CARGEAAA 14 AGA CC 14 GGG-TANTAGGGAAA-CARGEATS 2 ACTOC BGC 17 GCTT-GAATAGGGAAA-CARGEGTS 14 AGG CC 14 GGT-CAAAAGGGAAA-CARGEGTS 14 AGG CC 14 GGT-CAAAAGGGGAAA-CARGEGTS 14 AGG CC 14 GGT-CAAAAGGGGAAA-CARGEGTS 17 AAGC CC 14 GGT-CAAAAGGGGAAA-TGGOGTS 10 - CAA CC 14 GGT-CAAAAGGGGAAA-TGGOGTS 10 - CAA CC 14 GGT-CAAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-CAAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-CAAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-CAAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-CAAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-TAAAAGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-TAAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-TAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTT-TAAAAGGGGAAA-TGGOGTS 16 ACTOC CC 19 GGTTTS 10 CC 10 CGTTS 10 CC 1	COC 8 GANT-GANANGGGANCGARGET 2 AGACOTT Re 15 GALAGAGGANCGCARGET 28 AATOGG COC 10 GGATCANANGGGANCGCGGTG 18 AATOGG COC 10 AGGT-CANANGGGANCGCGGTG 13 AATOGG COC 10 AGGT-CANANGGGANCGCGGTG 13 AATOGG COC 12 AGCT-CANANGGGANCGCGGTG 15 AATOGG COC 13 GGTCANANGGGANTGCGGTG 15 AATOGG COC 13 GGTCANANGGGANTGCGGTANTGCGGTG 15 AATOGG COC 13 GGTCANANGGGANTGCGGTG 15 AATOGG COC 13 GGTCANANGGGANTGCGGTG 15 AATOGG COC 13 GGTCANANGGGANTGCGGTG 15 AATOGG COC 14 COC 15 CGCGGTANTGCGGTG 15 AATOGG COC 14 COC 15 CGCGGTANTGCGGTG 15 AATOGG COC 15 CGCG COC 15 CGCCG COC 15 CGCCG COC 15 C	CCT 23 AGTCAAACGGGAAACAGGGA 22 CAAAC CCT 16 GTTAAACGGGAAACAGGGA 22 CAAAC CC 22 SCCCAAACGGGAAAACAGGGAA 17 CAAC CC 6 GGCCAAACGGGAAACGCCAGGG 9 DCSC CC 6 GGCCAAACGGGAAACGCCAGGG 9 DCSC TTC 22 GATCAAACGGGAAACGCCAGGT 23 ATGC CT 23 AATTAAACGGGAAACCCGGGTG 2 ATGC CT 24 AATTAAACGGGAAACGCGGGTG 2 ATGC CC 24 AATTCAAACGGGAAACGCGGGTG 2 ATGC CC 24 AATTCAAACGGGAAACGCGGTGG 2 ATGC CC 24 AATT-CAAACGGGAAACGCGGTGG 2 ATGC CC 24 AATT-CAAACGGGAAACGCGGGTG 2 ATGC CC 24 AATT-CAAACGGGAAACGCGGGTA 2 2 ATGC CC 24 AATT-CAAACGGGAAACGCGGGAAACGCAGGTG 2 ATGC CC 24 AATT-CAAACGGGAAAGCCGGGTAACGCAGGTG 2 ATGC CC 24 AATT-CAAACGGGAAAGCCGGGTAACGCAGGTG 2 ATGC CC 24 AATT-CAAACGGGAAAGCCAGGTAAACGCAGGTAAACGCAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAACGCAGGTAACGCAGGTAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAACGCAGGTAAAAACGGTAAACGCAGGTAAAAAAAAAA
1 1'	Agtky rg aaddaA ggtg adrucr 22 Mc-ThAnsdedA-coeffo darroc 15 casorThAndedA-coeffo adrucr 15 casorThAndedA-coeffo adrucr 16 casorThAndedA-coeffo adrucr 16 cas-chAndedA-coeffo rgroco 11 cas-chAndedA-coeffo adrucr 16 cas-chAntescaA-coeffo darroc 16 cas-chAntescaA-coeffo adrucr 18 cash-chAntescaA-coeffo adrucr 11 cash-chAntescaA-ffocaffo adrucr 11 cash-chAntescaA-ffocaffo	<pre>serrer 19 ActT-AndAcedaAProcerce 22 serrer 19 ActT-AndAcedAProcerce 22 serrer 19 ActAcAcedAProcerce 22 defrer 20 ActT-AndAcedAProcerce 23 defrer 20 ActT-AndAcedAProcerce 20 serrer 17 GeT-GAAAGGGAAProcerce 20 serrer 18 GAAT-TAAAGGGAAProcerce 20 serrer 9 GAC 4 AATCGGAAProcerce 30 serrer 9 GAT 9 AATAGGGAACOTGGTE 30 serrer 17 GAT 8 AATAGGGAACOTGGTE 30 serrer 10 GAT 9 GAT 9 GAT 9 GAT 90 serrer 10 GAT 9 GAT</pre>	<pre>reprint in the image of th</pre>	efforce 10 GeAT-TANTARGEAAACACCERTS 21 AATOR METTYCE 12 GEAT-TERANAGGEAAACACCERTA 14 AGA FEAT-CC 12 GEAT-TERANAGGEAAACACCERTA 24 AGA FEAT-CC 14 AGAC-CAAGAAGGEAAACACCERTS 2 AATOR METTEGEC 11 AGAC-CAAGAAGGEAAACACCERTS 2 AATOR METTEGEC 10 GEAT-GAATOGEGAAACACCERTS 2 AATOR METTEGEC 10 GEAT-GAATOGEGAAACACCERTS 2 AATOR METTEGEC 10 GEAT-GAATOGEGAAACACCERTS 2 AATOR METTEGEC 10 GEAT-GAATAGGEAAATGCCERTS 2 AATOR METTEGEC 10 GEAT-GAATAGGEAAATGCCERTS 2 AATOR METTEGEC 10 GEAT-GAATAGGEAAATGCCERTS 2 AATOR METTEGEC 10 GEAT-GAATAGGEAAATGCCERTS 1 AAGA METTEGE 1 GEAT-GAATAGGEAAATGCCERTS 1 AAGA METTEGE 1 GEAT-GAATAGGEAAATGCCERTS 1 AAGA METTEGE 1 GEAT-TGAAAGGEAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAGAGGAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAGAGGAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAGAGGAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAGAGGAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAAGAGGAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAAGAGAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAAGAGAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAAGAGAAAATGCCERTS 1 AATOR METTEGE 1 GEAT-TAAAAGAAAATGCCERTAATGCCERTAATGCCERTS 1 AATOR METTEGEATAAAAGAAAAAAAAAAAAAAAAAAAAAAAAAA	MARTICE 8 GANT-GAMANGEGAMCGERERT 2 AGLOUT MERTICE 15 0.0GALAG-CGERERTE 22 ANTOGR MERTICE 16 0.0	Refrect 23 AGTCAAACGGGAAACAGGGGA 22 CAAAC BETCER 16 GGTTAAACGGGAAACAGGGGA 22 CAAAC BETCER 22 GGTTAACGGGAAACAGGGAA 17 CAAC BETCER 22 GGTAAGGAGGAAACAGGGAA 17 CAAC BETCER 2 GATAAGAGGGAAACAGGGAA 27 CAC BETCER 19 AAATAAGAGGAAACAGGGG 22 AFGC BETCER 19 AAATAAACGGGAAFCCGGTG 16 AAGC BETCER 19 AAATAAACGGGAAFCCGGTG 16 AAGC BETCER 22 GHAAACGGGAAFCCGGTG 16 AAGC BETCER 21 AAATAAACGGGAAFCCGGGTG 22 AFGC BETCER 21 AFTTCAAACGGGAACCCGGGTG 2 AFGC BETCER 21 AFTTCAAACGGGAACCCGGGTG 2 AFGC BETCER 21 AATTAAACGGGAACCCGGGTG 2 AFGC
	hgetkey rg aa adda. ggtg hrgnes geruch 2 Mo-ThAlasdacacocomo prodaa geruch 12 Mo-ThAlasdacacocomo soffta faerou 16 Gast-charlasdacocomo adda.geruch 16 Gast-charlesdacocomo adda.geruch 18 Gast-charlesdacocomo Arrota geruch 18 Gast-charlesdacocomo Arrota geruch 18 Gast-charlesdacocomo adda.geruch 18 Gast-rhanresdacocomo arrota 18 Gast-rhanresdacocomo arrota 18 Gast-rhanresdacocomo arrota 18 Gast-rhanresdacocomo arrota 18 Gast-rhanresdacocomo arrota 18 Gast-rhanresdacocomo arrota 18 Gast-rhanresdacocomo	TETTCS GETTER 19 MATT ALABOREDA - FOCGETE 2 TETTCS GETTER 19 MATT ALABOREDA - FOCGETE 2 TETTCS GETTER 19 ANTACTAAAAGGGAA - FOCGETE 22 TETTCS GETTER 19 ANTACTAAAAGGGAA - FOCGETE 22 TETTCS GETTER 19 ANTACTAAAAGGGAA - FOCGETE 22 TOTTCS GETTER 17 GAT - GAAAAGGGAA - FOCGETE 22 COTOG GETTER 17 GAT - GAAAAGGGAA - FOCGETE 22 COTOG GETTER 17 GAT - GAAAAGGGAA - FOCGETE 22 COTOG GETTER 17 GAT - CAAAAGGGAA - FOCGETE 22 COTOG GETTER 18 GAT - TAAAAGGGAA - FOCGETE 22 COTOG GETTER 19 GAAA - COTOGETE 22 COTOG GETTER 19 GAAAAGGGAA - COTOGETE 22 COTOG GETTER 19 GAAA - AAAAGGGAA - COTOGETE 22 COTOG GETTER 19 AAAAAGGGAA - COTOGETE 22 COTOG GETTER 19 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	Coogon derror 22 GJA - TATTOGGAAA - TOSGENT 24 ANY COOSON derror 22 GJA - TATTOGGAAA - TOSGENT 24 ANY COOSON derror 12 GOA - CANANGGGAAA - TOSGENT 12 ANA COGGAT GETTOC 12 GGAT - GAAAAGGGAAA - TOSGENT 12 ANA COGGAT GETTOC 12 GGAT - GAAAAGGGAAA - COCGENT 12 ANA COCGAT GETTOC 13 GGAT - GAAAAGGGAAA - COCGENT 2 AN COCCAT GETTOC 14 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 14 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 14 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 14 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 14 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 14 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 14 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 13 GGAA - CANANGGGAAA - COCGENT 2 AN COCCAT GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGAAA - COCGENT 2 AN COCCAC GETTOC 13 GCAA - CANANGGAAA - COCGENT 2 AN COCCAC GCAA - CANANGGAAA - COCGENT 2 AN COCCAC GCAC COCCAC C	Accord actron 10 card-manifestaladvancedadvanced 1 and offerror agrinor 12 cant-renainseauadvanta 14 advanced cases of card-or 15 cash-renainseauadvanta 14 advanced cases of card-or 16 cash-renainseauadvanta 14 advanced cases of card-or 16 cash-renainseauadvanta 14 advanced cases of card-or 16 cash-cantaced card card card 16 cash-cantaced card card 16 cash-cantaced card card 16 cash-cantaced card card 16 cash-cantaced card 16 cash-cantaced card 17 card 18 card-cantaced card 18 card 18 card cantaced card card 18 card-cantaced card card 18 card-cantaced card card 18 card-cantaced card 18 c	CIGERCA METTICO 8 GATT CANANGGRANCOTRETT 2 AGLOCHT CONTRIMINATION 8 GATT CANANGGRANCOTRETT 2 AGLOCHT CANCON GATTICO 8 GATT-CANTAGGRANCOCRETTE 12 ANTOGO CANCON GATTICO 16 GATT-CANTAGGRANTANGGRANTANGGRAN 15 ANTOGO AAGCAN GATTICO 10 GATT-CANANGGRANTANGGRAN 15 ANTOGO COTTGE GATTICO 10 GATTICO 10 GATTICO COTTGE GATTICO 10 GATTICO COTTGE CANANGGRANTANGGRANTANGGRAN 15 ANTOGO COTTGE GATTICO 13 GATTICO COTTGE CANANGGRANTANGGRANTANGGRAN 20 GOCO COTTGE CANANGGRANTANGGRANTANGGRAN 20 GOCO CANANGGRAN 14 ANCOANANGGRANTANGGRAN 20 GOCO COTTGE CANANGGRANTANGGRANTANGGRAN 20 GOCO COTTGE CANANGGRANTANGGRANTANGGRAN 20 GOCO CANANG 20 COTTGE CANANGGRANTANGGRAN 20 GOCO COTTGE CANANG 20 COTTGE CANANGGRANTANGGRAN 20 GOCO CANANG 20 COTTGE CANANGGRANTANGGRANTANGGRAN 20 GOCO CANANG 20 COTTGE CANANGGRANTANGGRAN-TANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRANTANGGRAN-TANGGRAN-TANGGRANTANGGRAN-TANGGRANTANGGRAN-TANGGRAN-TANGGRANTANGGRAN-TANGGRANTANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRAN-TANGGRA	CCETT GETTER CONTECT 23 MET-CANACGEGANRCMEREA 22 CANACCECCT FERGET 15 6677-TAACGEGAANRCMERAA 17 CANACCECCT FERTER SETTER SETTER SETTER SETTER 25 6676-1AAGAGEGAANRCMERAA 17 CANACCTCAA 667678 15 6776-1AAGAGEGAANRCMERT 23 ATTER SETTER SETTER 22 6676-1AAATGAGEAANRCMERT 23 ATTER SETTER SETTER 22 6676-1AAATGAGEAANRCMERT 23 ATTER SETTER 22 6676-1AAATGAGEAANRCMERT 23 ATTER 22 6760-1AAATGAGGAAN-RCMERT 24 CANACCTCAAGAGAAN-RCMERT 24 CANACCTCAAGAGAAA-RCMERT 24 CANACCTCAAGAGAAA-RCMERT 24 CANACCTCAAGAGAAA-RCMERT 24 CANACCTCAAGAGAAA-RCMERT 24 CANACCTCAAGAGAAA-RCMERT 24 CANACCTCAAGAAAA-RCMERT 24 CANACCTCAAGAAA-RCMERT 24 CANACCTCAAGAAAA-RCMERT 24 CANACCTCAAGAAA-RCMERT 24 CANACCTCAACAAAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAGAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAACAAA-RCMERT 24 CANACCTCAACAAA-RCMERT 24 CANACCTCAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAAAA-RCMERT 24 CANACCTCAAAA-RCMERT 24 CANACCTCAAAAAA-R
0 1 1	A cocontranse derivery zg aa adoda. copte cocontranse derivery zg aa adoda. copte cocontranse derivery zg accommon contranse cocontranse derivery zg accommon contranse cocontranse and contranse and and accommon and accommon contranse and and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon and accommon accommon and accommon accommon and accommon accommon and accommon accommon and accommon	CTTUTUTE CONTRACT AND A CONTRACT AND	Addational and a set of the set o	CAAAASSEAN SETTICS 10 6547-TANATIGGEAAACAAGESTE 21 AATIO TEACGFFFFFF ISTTICS 12 6477-TSAAASSEAAACAAGESTE 21 AATIO CAACGSGSCTFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF	GCTTC BOERT ABETTICC 8 GATT-GAAAAGGGAACGAGEET 2 AGACUT GCCCOSTO GATTICC 15 GALAAAGGGAAACGCAGET 2 AATOGG TTAACCAACAA GATTICC 10 GATT-CAAAAGGGAAACGCGGTG 13 AATOGG TTAACCAACAA GGAA GATTICC 10 GATT-CAAAAGGGAAACGCGGTG 13 AATOGG TTGACCATGG GATTICT 10 AGTT-CAAAAGGGAAACGCGGTG 13 AATOGG ATGGCCGTG GATTICT 10 AGTT-CAAAAGGGAAACGCGGTG 13 AATOGG ATGGCCGTG GATTICT 10 AGTT-CAAAAGGGAAACGCGGTG 13 AATOGG AATGGCCGTGA GATTICT 10 AGTT-CAAAAGGGAAACGCGGTGA AATGCG AATGGCCGTGA GATTICT 10 AGTT-CAAAAGGGAAACGCGGTGA AATGCG AATGGCCGTGA GATTICT 10 AGTT-CAAAAGGGAAACGCGGTGA AATGCG AATGGCCGTGA GATTICT 10 AGTT-CAAAAGGGAAACGCGGTGA AATGCG AATGGCCGTGAAAA GATTICT 10 AGTT-CAAAGGGAAACGCGGTGAAAAGGGAAACGCGGTGAAAAGGGAAACGCGGGAAAA-GGGGAAACGCGGGAAAAGGGAAACGCGGGAAAAGGAAACGCGGAAAAGGGAAACGCGGAAAAGGGAAACGCGGGAAAAGGGAAACGCGGGAAAA-CGCGGGAAAAGGGAAACGCGGGAAAA-CGCGGAAAAAGGGAAACGCGGGAAAA-CGCGGAAAA-CGCGGAAAAAGGGAAAACGCGGGAAAA-CGCGGAAAAAGGGAAAA-CGCGGGAAAAAGGGAAAA-CGGGAAAAAGGGAAAA-CGCGGAAAAAGGGAAAACGGAAAA-CGCGGGAAAAAGGAAAA-CGCGGAAAAAGGGAAAA-CGCGGGAAAAAGGGAAAAACGGGAAAAAGGAAAAACGGAAAA-CGCGGGAAAAAGGGAAAACGGAAAAAGGGAAAAACGGAAAAAGGGAAAACGGAAAAAGGAAAAAGGGAAAAACGGGAAAAAA	acrossement server 23 AstCAACessaAKCAssa 22 CAAC seconderry server 23 AstCAACessaAKCAssa 22 CAAC recordentry server 16 astrCAACessaACCCAssa 21 CAAC recordentry server 2 accCAACessaACCCASSA 20 CAC recordentry server 2 accCAACessaACCCASSA 20 CAC cCCFFFITTIG server 2 actCAACessaACCCASSA 20 CAC cCCFFFITTIG server 2 actCAACessaACCCASSA 20 CAC ACACCANTCH server 2 actAACessAAFCCASTC 16 AAC ACACCANTCH server 2 AATAACessAAFCCASTC 16 AAC ACACCANTCH server 2 AATAACessAAFCCASTC 16 AAC CCCCASTCASTC Server 2 A CAAC ACACCANTCH server 2 A CAAC ACACCANTCH server 2 A CAAC ACACCANTCH server 2 A CAAC CCCCASTCASTCASTCASTCASTCASTCASTCASTCASTC
0 1 1'	CLEATA AGGAN TATCS AGTOR 22 AGC - TAAAAGGGAAA - COCGATG -266 GOCCA TATCS AGTOR 12 AGC - TAAAAGGGAAA - COCGATG -278 COCCTCORAAA AGTOR 12 AGC - TAAAAGGGAAA - TCOCGATG -290 GGCTCORAAA AGTOR 15 GGGCTTAAAGGGAAA - TCOCGATG -231 AATCATCAAA TATCS - TAAAAGGGAAA - TCOCGATG -231 AATCATCAAA TATCTCO 16 GGC - TAAATGGGAAA - CACGATG -300 ATTGATCAAATTGAAG - CACATTGGGGAAA - CACGATG -301 ATTGATCAAATTGAAG - CACATTGGGAAA - CACGATG -301 ATTGATCAATTGAAG - CACATTGGGAAA - CACGATG -301 ATTGATCAATTCAAGT 16 GGA - GAAATTGGGAAA - CACGATG -301 ATTGATCATTCA 16 GGAA - CAAATTGGGAAA - CACGATG -331 AGTGATCATCA 16 GGAA - CAAATTGGGAAA - CACGATG -331 ATTGATCATTCA 16 GGAA - CAAATTGGGAAA - CACGATG -331 ATTGATCAATTCA 16 GGAA - CAAATTGGGAAA - CACGATG -251 TGCACGACGAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 ATTGATGGAAGAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 ATTGATTGAAGAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACGACAGAAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACGACGAAGAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACGACAGAAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACTTCAAGAAAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACTTCAAGAAAAA GGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACTTCAAAAAAAAAAGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACTTCAAAAAAAAGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACTTCAAAAAAAAGTTCC 11 GGAA - TAAATTGGGAAA - TGTGALCG -338 CACTTCAAAAAAAAAACTTCAAAAAAAAAAAAAAAAAA	 558 CTTUTIETTECE GETTER 19 ANTALAGEGGAA TOCGERE 23 -229 ANGEGATER GETTER 19 ANTACANAAGGGAA TOCGERE 23 -229 ANGEGATER 19 ANTACANAAGGGAA TOCGERE 23 -257 CTUGTIETTECE GETTER 19 ANTACAGAA TOCGERE 21 -257 CTUGTIETTECE GETTER 19 ANTACAGAA TOCGERE 21 -258 CTUGTIETTECE GETTER 19 ANTACAGGAA TOCGERE 21 -258 CTUGTIETTECE GETTER 19 ANTACAGGAA TOCGERE 21 -238 CANAACCEA GETTER 10 GAGC TAAAGGGGAA CANAGGGAA COCGERE 21 -238 CANAACCEA GETTER 11 GGT CANAAGGGAA COCGERE 21 -231 TTCTTCCGERE GETTER 21 GAG AAAAGGGAA COCGERE 21 -231 TTCTTCCGERE GETTER 21 GAT TAAAGGGGAA COCGERE 21 	 250 AdaAGGGRA Garter 22 GA-TANTGGGAATGGGRA 24 AM -308 Arrectaces Adarter 22 GA-TANTGGGAATGGGRA 24 AM -308 GGCCGGRQA GARTER 12 AAA-GAAAAGGGAATGGGRA 12 AM -313 TFCCGRFC TATTOR 12 GGATGAAAAAGGAAA-TGCGRA 12 AM -314 GCCGGRC TATTOR 14 GGATGAAAAAGGAAA-GCCGRA 27 AM -364 GCCGGGCGCG GGATTOR 14 GGATGAAAAAAGGAAA-GCCGGRA 2 AM -364 GCCGGGCGC GGATTOR 14 GGATGAAAAAAAGGAAA-GCCGGRA 2 AM -364 GCCGGGCCG GGATTOR 14 GGATGAAAAAAAGGAAA-GCCGGRA 2 AM -364 GCCGGGCCG GGATTOR 14 GGATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	-222 CAACASEGRA SERVER 10 SEAT-TANTRAGEAAACACEENS 14 AGA -222 TRAACASEGRATE LISTICE 12 SEAT-TRAAASEGAAACACEENS 14 AGA -226 CATEGRETTE LISTICE 12 SEAT-TRAAASEGAAACACEENS 14 AGA -276 CATEGRETTERSET 61 SCIC-CAAAAASEGAAACACEENS 2 AATUC -276 CATEGRETTERSET 17 SCIT-CAAAAASEGAAACACEENS 2 AATUC -271 SCICASEGRETTE SEATOC 10 SCAT-CAAAAASEGAAACACEENS 2 AATUC -271 SCICASEGRETTE SEATOC 11 SCIT-CAAAAASEGAAACACEENS 2 AATUC -271 SCICASEGRETTE SEATOC 11 SCIT-CAAAAAASEGAAACACEENS 2 AATUC -271 SCICASEGRETTE SEATOC 11 SCIT-CAAAAAASEGAAAFOCCACIS 0	 318 GCTTC GOETCA GETTOC 8 GENT-GAMANGEGAACCERERT 2 AGLOCTT 2 281 GCGCOCTTC GETTOC 8 GENT-CAMANGEGAACCERERTE 22 AGLOCTT 448 TEACCAACCTA GETTOC 6 GENT-CAMANGEGAACCERERTE 15 ANTOOG -333 TTGAAGCAI GETTOC 10 AGT-CAMANGEGAATACGEGTTE 15 ANTOOG -346 TEGCAAGCAI GETTOC 10 AGT-CAMANGEGAATACGEGTTE 15 ANTOOG -2412 CCGTTGGTTP GETTOC 10 AGT-CAMANGEGAATACGEGTTE 15 ANTOOG -3412 CCGTTGGTTP GETTOC 10 AGT-CAMANGEGAATACGEGTTE 15 ANTOOG -340 ANTGCCTTG GETTOC 10 AGT-CAMANGEGAATACGEGTTE 15 ANTOOG -340 ANTGCCTTG GETTOC 10 AGT-CAMANGEGAATACGEGTTE 15 ANTOOG -340 ANTGCCTTG GETTOC 13 GETTGAMANGEGAATACGETTE 15 ANTCOOG -340 ANTGCCGTTB GETTOC 13 GETTGAMANGEGAATACGETTE 15 ANTCOOG -340 ANTGCCTTB GETTOC 13 GETTGAMANGEGAATACGETTE 15 ANTCOOG -340 ANTGCCCTTB GETTOC 13 GETTGAMANGEGAATACGETTE 15 ANTCOOG -340 ANTGCCTTB GETTOC 13 GETTOCAMANGEGAATACGETTE 15 ANTCOOG -340 ANTGCCTTB GETTOC 13 GETTOCAMANGEGAATACGETTE 14 ANTCO -340 ANTOCACCTTB GETTOC 13 GETTOCAMANGEGAATACGETTE 14 ANTCO -340 ANTOCACCTTB GETTOC 13 GETTOCAMANGEGAATACGETTE 14 ANTCO -340 ANTOCACCTTB GETTOCAAAMANGEGAATACGETTE 14 ANTCOCAA- -340 ANTOCACCTTB GETTOCAAAMANGEGAATACGETTE 24 ANTCOCAA- -340 ANTCOCACTTB GETTOCAAAMANGEGA	Control Con

FIGURE 1. (Continued on next page)

.0	1 GCCAACGTCGC	GCTAACGTCGC GTCGGTGATAA	GCCGGTAACTG GAATACGTCGC	GTATTACGTCG	GAATACGTCGC	GCAACTGAGTT	GCAACGCCCTG	CGTCGATCCCG	GGAACCTTTCG	GATACTGCCAA	GGAGCTTCACT	GGTCACGTTTT GGCGTTCATGA	GATCCATCCAG	GCAACACACACAG	CGAACCGGGCT	GACATTTTTCC	ACCCGATTCAA	TAAAGGTGTTTT	GTGGCGATCCA	GAGGGATTGAC	IGTTIGA TCAGC	GTTTTAACAAC AATGTATGCAC	TTCAGTTTGAG	GTTCTAACAGC	TACTTCCACAA	TGCTTGGAACG	TGTTTGGATCG	TTATTGTGTGAGG AATAAGACGGA	TTATATTCGAT	TATTCTAGTAT	ATTTTAAAAC	TTTATTTAAGT	AGTATGCTATT	AAATATATAAA	AGTATTAATTTA	TAAGGTTTTAG	GA TAAGTCTAA	GTTACCACGTC	GTCGCGGACAAA	GTAACGCTCGG	GTTAATAAAAC	TGAACATTGCG	TTTAGAAGCT	ATGTGTATACC CATTCCGGAAA
	EAGYCM69AgaCCk6Co	TAAGCCIGAAGACCTGCC	TAAGCCAGAAGACCTGCC	PAAGCCCGAAGACCTGCC DAAGCCCGAAAACCTGCC	PAGCCAGNAGACCAGCC	AAGTCCGAAGACCTGCC	AGCCOGGAGACCGGCC	TGAGCCAGGAGACCTGCC	GAGCCCGGAGACCCGCC	CAAGCCCGGAGACCGGCC	AAGCCIGGAGACCCGCC	A AGC CAGGAGACCT GCC A AGC COGGAGACCOGCC	CAGCCUBGAGACCUBCC	CAGCCCGGGAGACCCGGCC	CAGCCAGGAGACCTGCC	GAGCCCGGGAGACCGGCC	DOAGT CARGAGACCTGCC	A AGIIC GGAGACCIGCC	FEAGC COGGAGACCEGCC	AGCCCGGAGACCCGCC	GAGCCAGGATACCTGCC	PAGE CAGEAGACCEGCC	PAGE CAGGAGACCEGC	AGAGOC//GGAGACC//GCC	GAGTCAGGAGACCTGCC	GAGE CAGGAGACCEGCC GAGCCAGE AGACCEGCC	D AGT C G GAGACCT GCC	A AGT CAGGAGACCTGTC	TAAGT CAGGAGACC-GCT	PERSOCAGGAGACCTRACC	SAGCO GAGACCI GCC	AAGTCGGGATACCTGCC	TAGCCAGGAGACCTGCC	A AGLICAGGAGACCUGCC	PAGE CAGGATACCTGCC	CAGC CAGGAGACCUGCC	TAGT CAGGATACCGGCT	CAGCCAGGAGACCTGCC	CAGCCAGAAGACCTGCC CAGCCAGGAGACCTGCC	CAGT CAGE AGACCTIGCC	GAGCCAGGAGACCIGCC	PAGE CAGGAGACCTGCC	CAGCCAGA AGACCTGCC	PAGECAGNAGAC-1600
VS 4	c 19 Accounce	C 24 AACCCTCC G 6 AGGACGCT	T 11 CAGCCTCC	C 67 GAGATCCT C 32 GAGATCCC	C 8 GTGACGCT C 45 CGGCATCC	G 14 GTGCTCTC	G 9 ACCUTCED	G 16 CAGACCGI		G 17 CCACTCAC	A 9 CCCCTCGC	C 20 TTCCTCGC	G 14 CCGCTCAT	C 23 CCCCTCGT	G 39 GCTGCCGT	C 30 CCCCTCGC	G 11 -GACATGC	TATAGCC	G 25 TCGCTCGT	G 19 CCACTCC	GCGCCA	A 10 TGAAGCAT	C 9 ATGAACAT	T 8 TGACGCTA	C 8 TAAAGCAC	C 9 CGATGCAC	C 8 ATGATCCT	T 9 TGATACAC G 8 ATGAAGCC	G 8 ATGATGTT	A 10 -GATTTT	C 9 TGATACTT	G 7 AAGAAGCA	A B TCAATCAT	A 8 ATGAAGCT	T 9 ATGAAGTT	G 8 TGAGTCCC	G 8 AGGAGCT	A 8 ATGAACCO	C 9 TGATTCCT	G 7 ATGATTOC	T 4 AAAGGCGC	G 7 AAGACTGC	G 9 TGACACGA	G 8 ATGATCCT G 8 ATGAGTTC
ا و	(GGGAAGGC	GGGAAGTCAT	COCCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	GGGAAGT CAT	GGGAAGCTAG	GGGAAGT CGA	GGGAAGGCGA	GGGAAGGCGG		GGGAAGGCGC	GGGAAGGT GA	GGGAAGGCGT	GGGAAGGCGC	GGGAAGGTGA	GOGAAGGCGC	GGGAAGGTGA	GGGAAGGC GG		GGGAAGGCGC	GGGAAGGCGC		GGGAAGGTAA	GGGAAGGC GG	GGGAAGACTG	GGGAAGGCTT	GGGAAGGCCC	CCCAACA CCC	GGGAAGGT TC	GGGAAGACTG	GGGAAGGATIG	GGGAAGGCAC	GGGAAGA CGA	GGGAAGGATT	GGGAAGGAAA	GGGAAGGTTA	GGGAAGGCAG	TGGAAGGATT	GGGAAGACGG	GGGAAGGLAD	GGGAAGGCGT	GGGAAGGCGA	GGGAAGGCAA	GGGAAGGCGT	GGGAAGG AA
AddIT	19901	CGG	AGGAT	Adgan	80CGA	CGAG	CGGCA	ACG		AGCATC	AAGCA	T-AGTACGCC	ACAG	CGGGCAI		TGGCA	AGATOC		TCGGACA	TCGCA		AA	GCGAC	ATTGGT	14 AAATGTAS	14 -CGCCGC	TGCTTC	TAATCAC	AATTCC	ACT	TTTATA	TTTGTGTGT	ATAGTCAL	13CTCAI	TTUCAU	60CCGGG	TTTC	AAGCGC	DID		AGACT	AGOC	TAGAGA	GTTTC
9	CACTG	CACTGACCTC	CACTGOGGCT	CACTGTCCGC	CACTGATCCAT-	CACTGCTTTT	CACTGTGTGCTC	CACTGCGTG		CACTGGATGCC-	CACTGTGCTC	CACTGCGCCA	CACTGCTTA	CACTGTGCCCA-	CACTGCTG	CACTGTGTGT	CACTGGATTC		CACIGTGTGTGC-	CACTGTGCAG		CACTGTGTGA	CACTGTCTA	CACTGACCGTTC	CACTGTACATC-	CACTGGCGGCA-	CACTGAAGGA	CACTGTGAGAAA-	CACTGGAAGC	CACTGGGA	CACTGGAATATA	CACTGCATTT	CACTGTGATAA-	CACTATGATG	CACTGTCAGT-	CACTGTCC6G	CACAGAGATT	CACTGGCGTTT-	CACTGGACTG	CACTGGGA	CACTGGCGA	CACTGGGATG	CACIGTOTTTTT	CACTGAAAGG
SA 1	91 16 GCG <mark>G</mark>	T 15 GCAG	G 19 GCAG	ST 18 ACAG	C 17 GCAG	C 9 AACG	A 12 AIG	NT 8 AATG		A 10 AGGG	A 9 GATAC	AC 11 ACAG	A 11 GATGO	A 6 AGAT	AT 9 ACAG	A 11 CGIG	G 11 ACTO		G 10 ATAG	C 11 TACC	A	C 10 GAAAC	A 8 AGA	G 10 ATTG	G 10 TTTGC	C 8 ATGGC	C 10 CAAG	AC 8 ATAAC	G 11 TATG	G 11 TATG	G 12 TAAG	G 15 TATG	C 8 AAAG	G 9 TTAG	C 7 AAAT	C 10 TAAG	C TO TANG	NG 10 AATCH	IC 12 GAATC	TT B TTGGC	TT 17 CATC	C 9 AGGCC	C 10 AATG	TA 11 ATTRC
4	KGTra	DEGTARGEAAGE	CCTAAGGAAGGA	CET A A G G G G A A C	CGTAAGGAATGC	CGT AGAGAGA	COLUMN COUNT	TIGT GAACGGCGP	CULAGCGCATC	COLOCATCACTA	GGTAAGCGAGCG	CIGTAAGCACCGP	SGTAAATGAGTP	COLLA AGCORAGO	CIGTAAACGGTCA	CGTAAGCGAGAG	UIGTAGGCATCGP	SUGTOAGTAGTTP	SGETAGACGAGTC	COL COCCENECT	CIATTGGCCP	CIPATGCTGP	CIAAATGGGAG	C GTAGTAGCTGP	CFGT CAGTGCTAC	TIGTAT GTGCGGP	C GT GTGGGGAG	TIGTAAGTGTGGAG	CI GTATACGGAC	CIGTALACAGTIA	CIATAGAGGA	AGT AACCGTGAA	CITCATAGGA	GTGATGCAGAC	CIGTGAAACCAAC	GT AGGGAGGA	CIATGGTGGTGGAT	CLAAGGGGAAT	CIAL GGGAAGP	CTARAGAGAGC	CLAABCGCCGT	CTCAACCACGA	TETATCGGAGA	CIPAGGGAAAT
3 3'	dG-yccccccha	GACGCGC	6ACCCCCC	6ACGCGC	GACCCCCC	GACGCGC	6- CCCCBC A	G- CCCCGC AC	G- CCCCGC A	G- CCCCGC A	6- CCCCGC AC	6- CCCCGC AC	G- CCCCGC A	6- CCCCGC A	6- CCCCBC A	6- CCCCBC AC	G- CCCCGC A	G- CCCCBC A	6- CCCCGC A	16- CCCCBC V	GCCCGCCM	6CCCGCCA	GCCCGCCAC	G CCCGCCA	6- CCCCCC A	6- CCCCGC A	6- GCCCGC M	6- CCCCGC AC	GGCCCTAGCTA	6CCCGC M	B-T-CCCCCCCC	A-00T	6-CCCCCCCCCC	6-CCCCCCC	6-CCCCCCCCCC	6-CCCCCCCCC	G-ACCTACCTA	G CCCGCCA	-T-CCCCCCCCTA	G-A-CCCACCA	A-CCCCCCTTA	G CCCGCCA	6-A-TCCCC	G-T-CCCGCCG
dī 2'	a tCcg RC	2 AATCIGGA C	2 AATCCCCC C	2 AATCCGGA C	2 AATCCCCC C	2 ACTCCCCAAC	6 AGTCCCCCCC	1 ATGCCGCGCGCGC	4 AGTCCGGCGC	7 ATTCCGGCGC	9 AGTCCGGTGC	1 AAGCCGCGGGC 9 ATGCCGGTGC	0 ATCCC66C6C	9 AGTCCGGTGC	1 ATGCCGCAGC	8 AGTCC6GT6C	7 AAAACGAAAC	2 ATTCREGCAC	4 AGGCCGGCGC	4 ATTCC66AGC	1 AAGCCGGT_C	2 ACTCCAAC C	2 AATCCEAC	2 AATCCGGA C	2 AATCCGGAGC	2 AATCCGGAAC	3 AATCCEGC C	2 AGACCAGTAC 2 ATGCCGAAAC	2 AATCCTCT	2 AAGCCEGA C	2 AATCCCAC C	2 AGTOTTCA C	2 AAACTC	2 AFTCCITT C	2 ATGCCTT C	2 AATCCCGC C	2 AATCOTOT C	2 AATCCGAC	2 AGACCTCA	2 AATCOCGG C	2 AATCOTGA C	2 AATCCGGC C	2 AGTCCGGT C	2 AGGCCGGC C
2 Ad	egGtg	-TCCAGTG	-cccaere	-TCCGGTG		-TCCGGTG	- ecceerce 1	-CGCGGGTC	-GCCGGTG 1	- CCCGGTG 1	-ACCOGIG 1	-CGCGGTA	-GCCGGTG 3	-ACCGGTG 2	-TGCGGTA	-Acceere 1	-TTCGCCA 1	-rceeec 1	-GCCGGTG 2	-TCCGGTG 2	-ACCGGTA	-CTTGGTG	-GTCGGTG	GTTCCGGTG	-TCT GGTG	-TCCGGCG	-ccceere	-ACTGGTG	-AGAGGTG	-TCAGGTG	- TTGGGTT	-TGAAGTT	AGTT	-AAGGIT	AAAGGUT	-ACGGGTG	-AGAGGTG	-TTCGGTG	-TGAGGTG	-CTGGGTG	-TCAGGTG	-CCGGGTG	-ACCGGTA	- GCCGGTT - TCCGGTG
	TAATAGGGAA-	TAATAGGGAAAGGGAA-	TAGGGAA-	GTTAAAAGGGAA- GTTAAAAGGGAA-	TC-AAAGGGAA-	T-GAAAGGGAA-	G-AMACAGGGAA-	TAAGAGGGAA-	T-GAAACGGGAA-	G-AAACTGGGAA-	GAAACGGGAA-	T-AAGACGGGGAA- GAAACGGGGAA-	G-AAACAGGGAA-	GAAAOGGGAA-	T-AGACGGGAA	GAAACGGGAA-	C-TAACAGGGAA-	TAATCGGGAA-	G-AAACAGGGAA-	TAAACGGGAA-	T-TAAAGGGAA-	T-TAAAAGGGAAG T-GAAAAGGGAA-	T-GNAAGGGAA-	T-TAAAAGGGAA	T-TAAAGGGAA-	T-TAATAGGGAA- T-TAATAGGGAA-		T-TAATAGGGGAA- T-GAAGAGGAA-	ATCCGGGAA-	T-TAATAGGGAA-	T-TAAAGGGAAG	T-TAATAGGGAA-	T-TAATAGGGAA- T-GAAAGGGGAA-	T-TAATAGGGAA-	4 AAAAGGGAA- T-TAATAGGGGAA-	T-TAAAGGGAAG	AAAAAAAAAAAAAAAAAAA	TTAGAAAGGGAAG	T-GAAAAGGGAA-	GAATAGGGAA-	T-GAAAGGGAA-	A-GAAAGGGAA-	GAATAGGGAA	GAAAGGGAA- T-TAATAGGGAAC
1 1	Gtkcy rg	GTUPG 12 T	GTTCG 15 C	GC-CT 22 AGA	GTATG 15 T	GTOCT 26 660	GTGCC 18 GGT	GTTCC 16 6GC	GTGTC 17 AGG	Greece 21 667	GTGCC 18 GGT	GTICE 13 AGC	GTOCE 19 AGT	GTSCC 8 GGT	GTICE 25 AGC	CTGCC 8 66T	GTITC 6 GAM	GTCCT 18 AGA	erecc 5 Ger	GTGAC 16 GTT	GTGCC 18 GGC	GTGGT 16 TGT GTGGT 16 ATC	Greer 11 ece	GTACG 22 TGT	GTECC 6 GGC	GTETC 6 GAC	GCTTGGGT-AG-	GTGCC 5 GGC GTGGA 12 TTC	GTGAA 4 TTA	GTICA 16 TGA	CTUAA 20 AUA	CTOTO 5 COC	GTGTT 4 AAC	GTICE 7 AGA	GITCI 5 AGA	GTCCC 5 660	Soo C TOTA	GTCGC 20 GCG	GOSCC 5 GGC	GTGCC 8 GGA	ANT 3 ATTE	CTUCC 8 6GA	ercou 5 ceA	GTIGC 12 GTT
0	48 ATCCACTTGC	152 ATCCGTGGGC	145 ATTCGCCTAC	24 CATTGTGGTC	82 ATACTGAACA	126 AGCGCCAAGCI	197 GGCCCGTTCC	105 TGCCGGTTCG	45 GGGTTGTCCC	02 THE CONTRACT	119 CGCCAGTTTC	09 CCTCGCGTTCG	100 GGCTTGTTTC	35 AACTTGTTACE	(31 TGCCGGTTCGA	03 ACCTTGTTTC	121 TGCCGGCCGT	32 CAACCTTCTGT	102 GCCTCGCTTC	Clostridium arc	37 AATGTCAAAT	77 CTCAAGGAAAT	101 TCGCGCTGAAG	47 TAGGCTTCTT	145 GTGGACGGTA	33 ATTCGCAGCAAL	146 AGCCTGCTTA	18 GCAACAGTAA	132 ACAACTAAAT	07 TAAAATTTGTA	94 TTTAGAAAT	82 AACTAATAATT	17 AAAAGGATT	167 ATAATATTAT	287 GATTAAATT 93 AATAATACTAGG	137 TAAAGCCTTA	40 TGTACTTATG	196 TTGCTGGAAC	11 CIACGETTAC	106 TAACCGTTTC	97 AACTATTGAC	98 ATAGTATTCA	52 ACATAGCTTA	25 AGGCGGGAATAGC
	-proteobact	SY_BTUB -2	(P_BTUB -2	(E_BTUB -2	ZCBIA -2 20 BTUB -3	AC_BTUB -3	A_BTUB2 -2	A COBW -C	A CBTAB -2	P BTUR -:	P_BTUB2 -3	PP_COBW -2	PU_BTUR -0	U CBTAB -3	PY COBW -3	Y_CBTAB -3	PY BTUF -5	ION BTUB -3	W BTUB -3	The Bacillus/	IS BTUF -2	ID ACHX -3	ID_BTUF -4	ID_METE -2	ID_NRDA -3	BE NRDA -:	SE_CBIW -3	MOX -5	MOCBIA -3	A_CBIM -3	DPE_CBIM -4	TPE_BTUF -4	THE CBLT -5	DE_CBIM -3	DE CBIP -2	CHT_BTUR -3	TE BTUE -3	INO CBIM -5	IMO CBIQ -3	THO CELS -2	HABTUF -2	NHA CBLET -2	DHA CNOA -3	NHA_NRDD -0 0HA05379 -3

www.rnajournal.org 1087

	Sectors Sectors and sectors and sectors an	AGATTTG AGATTTG AAGCST AAGCST CHACTAA CAATTAA ACTAATTAA ACTAAAGG ACCAAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAGG ACCAAAAAAAA	TTACAACA TTGACGAT SAGTTAACA SAGTTAACC SATCAACC SAGTCAACC SACTTACC TTCCATCG	AAATTCCT AAAAAAC TTTTTTTTTTTTTTTTTTTTTT	e second e-paired cleotides
•0		CCL TACY CCL TACY CCL TTACY CCL CCL CCL CCL CCL CCL CCL CCL CCL CCL	CC AAA TTTCCAAA CC AAAT CC ACT CC ACT CC ACG CC ACC CC ACC CC ACC CC ACA CC AAAT CC CC ACG CC ACAC CC CC	CT TGA AC AGTA CA AAT CCATTA CCATTA CCCCCA CCCCCA CCCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCCA CCCC	es. The e. Base us nuc
	ga CCCC Garcercerce Garcercerce Garcerce Garcerce Garcercercerce Garcerce G	Caccino Caccin	TGCCCG TGCCCG TGCCCCG TGCCCCG TGCCCCG TGCCCCG TGCCCCG TACCCCG CGCCCG CGCCCG CGCCCG CGCCCG	TACGAC GACTTA GACTTA GACTTA AACCTG CCCCTG CCCCTG CCCCTG GACCTG GACCTG GACCTG	n gene <i>per</i> lin insens
	VICTOR CONSTRUCTION CONSTRUCTIC	CCAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN FICAGAN	II CAGAA II CCAGAA II CCAGAA II CCAGAA II CCAGAA II CCAGAA II CCAAA II CCGAA	TCAGAA TCAGAA TCAGAA TCAGGA TCAGGA TCAGGA TCAGGA	nstreau the <i>up</i> noncc
1	AND ACC CARL CONCERNING CONCERNIN CONCERNIN CONCERNING CONCERNING CONCERNING CONCERNING	ICTAAC AATAAC SCCAAC ACTAAC AACTAAC AATAAC SCTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC ACTAAC	TCCAAA TCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	TTTCAAE CGGAAAE CGGAAAE CGGGAAE CGGGAAE CGGGAAE CGGGAAE CCGTGAE	l dowi ws in 1 black;
10	 Gavage Control of Contro of Control of Control of Control of Control of Control of Con	A AGAA A COLTEG A COLTEG A A A A A A A A A A A A A A A A A A A	-Argoo -Actra -Actra -Actra -Cacto -Actro -Actro -Actro	AGAI AGAI B TTGAZ B TTGAZ D TGATG CAGGG CAGGG D TGATG D TGATG D TGATG D TGATG D TGATG D TGATG	oxima y arro 1s, in l
Þ		A CCC CCC III A CCC CCC III A CCC CCC III A CCC CCC		GGTG TGC TGC TGC CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGCG 10 CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CGTG 10 C CG 10 C CG 10 C C CG 10 C C C CG 10 C C C C C C C C C C C C C C C C C C	the pr own b ositioı
uti		CALAGE CALAGE CALAGE CALAGE CALAGE CALAGE CALAGE CALAGE CALAGE CALAGE CALAGE		SGAAGG SGAAGG SGAAGG SGAAGG SGAAGG SGAAGG SGAAGG	nes of are sho ved p
9		OCTATIO GATCHO GATCHO FIGALIO FIGALIO FIGACIO ACCTO GACCHO GACCHO GACCHO CCTCO CACCHO CCTCO CCTCO CCTCO CCTCO CCTCO CCTCO CCTCO CCTCO CCTCO CCTATIO CC		AACCTG TATTTG AACCTG -AACTG -AACTG -AACCTG -AATG AATG AATG AATG CG	ne nan cture ; consei
LIPP	Y	11 -0 19 -0 10 -0 10 -0 10 -0 10 -0 11 -0 12 -0 13 -0 14 -0 15 -0 16 -0 16 -0 16 -0 16 -0 16 -0 16 -0 17 -0 18	-10	2	and th 'y stru 1; non
A	AG TATA AG TATA AG TATA AT	AGA CAGA CAGA CATT COGTTOT COGTTOT AAAA AAAA AAAA AAAA		AC	ıble 3) condar 1 greeı
9	Acriectory Acriectory	Activity Activity Activity Activity Activity Activity Activity Activity Activity		ACTGGG ACTGGG ACTGGG ACTGGG	l in Ta VA sec ons, ir
ι Ω	Gradie Tradec Grade Gerade Gerade Grade Constructure Annor Annor Constructure Const	GATGO GATGO GATACO GATALO CATALO CCTTO GATACO GATACO GATACO GATACO AATACO AATACO AATACO AATACO CAGGTI TGTAT		TAAGC CATGC TATAC ATTAC AATGC GATGC GATGC GATGC GATGC TAAGC TATGC	(listec the Rl positi
ΔS	<pre>> * * * * * * * * * * * * * * * * * *</pre>	CGA 45 GAT 19 GAT 19 GAT 19 GAT 19 CGA 15 CAT 15 CAT 26 CAT 27 CAT 27 CA	AGA	ACG 10 GAA 16 GAA 16 GAA 16 ACG 8 GAC 11 TGT-11 TGT-11 TAG 14 AGA 14 AGA 14 AGA 14	ations ms of served n.
4	A Trong to the test of test of the test of tes	ATSCIT AGTICA AGTICA AGTICA AGTICA AGTICA AGTICA AGTICA AGTICA ATTCA ATTCA ATTCA ATTCA ATTCA ATTCA ATTCA ATTCA ATTCA ATTCA CAACGG	ATGGGA ATGGAAA AGGAAAA AGGGAGA ATGGGAGA ATGGGGC AAGGGGC AAGGGGC AAGGGGC AAGCGC	AGA AGG ATT GT GC ATT GT GC ATT GT GG ACC GG ATT GT GG AGT GAG AGT GAG	bbrevi ary ste e cons re give
	A COOT A	ACTION CONTRACTORY CONTRACTORY CONTRACTORY CONTRACTORY CONTRACTORY CONTRACTORY ACTION	ACTERIC ACTERIC CONTECTION CONTEC		ome al ementi enerat ops al
ση.					ne gen comple d; deg em–lo
m					ains th The c t in re able st
2,		rocoda contraction rocoda roco	rcceec rcceec rcreec rcreec rcreec rcceec rcceec rcceec	TCCGTC TCCCGC TCCCGC TCCCGC TCCCGC TCCCGC TCCCGC TCCCGC TCCCGC TCCCGC TCCCGCC	n cont codon. are se d variá
IPPY			V V V V V V V V V V V V V V V V V V V	N N N N N N N N N N N N N N N N N N N	colum start c itions nal ano
3		CCCCIC TTCCCIC TTCCCIC CCCCCC	CCGGTG CCGGTG CCCGGTG CCCGGTG CCCGGTG CCCGGTG CCCGGTG CCCGGTG	ACGGTT TT GGTT TT GGTT TT GGTT TT GGTT CC CC	e first e to the ed pos Iditior
	560.1	563.A	SGAAGT SGAAGT SGAAAT SGAAAT SGAAAT SGAAAGT SGAAAGT SGAAGT SGAAGT SGAAGT	66446 66441 96446 96446 96446 96447 96447 96447	es. The lative inserve the ac
	a. alo Arrest ar	AND CONTRACT	raaggee raagge raagge raagge raag raag	AGGAAGC PAAGAGG VATRAAG VATRAAG CAAGGTG PAAAGGTG TAAAAGGTG TAAAAGGTG TAAAAGGTG TAAAAGGTG TAAAAGGTG TAAAAGGTG	quence ase re- ors. Co chs of
Ĥ	and	01160- AAT- 0.0164	AGAGG' AGAGG' AGAGG' GAAGG GAAGG	ATG	ent sec first b ig colc e leng
-	Kory Kory 0.00 12 0.01 0.01	AXC 19 AXC 19 CCC 23 CCC 23 CCC 23 CCC 23 CCC 22 CCC 22 CCC 22 CCC 12 CCC 22 CCC 12 CCC 22 CCC 12 CCC 22 CCC 12 CCC 22 CCC 22 CCC 22 CCC 22 CCC 22 CCC 22 CCC 22 CCC 22 CCC 23 CCC 22 CCC 23 CCC 23 CC	rcr 17 cr 17 cr 17 rrcr 17 rrc	TAT 7 TAT 7 TTOC 11 66CT 4 66AC 28 66AC 28 66AC 28 66AC 14 14 14 14 14 14 14 14 14 14 14 14 14 1	-elem of the atchir ie. Th
		T	CT GGT CT GGT CT GGT GCT GGT GCT GGT GCT GGT AT GGT AT GGT GCT GGT CT CT GGT C	CA AGA CT CT CT TT CT CT CT CC CC CC CC CC CC CC CC CC CC CC CC CC	of <i>B12</i> sition 1 in m in blu
0	CATTCA CATTCA CAGCCCA CAGGCCA CAGGCCA CAGGCCA CAGGCA CAGGCA CAGCA	AAAGAC CGTGTC CGTGTC CATCGT CATCGT CATCGGA AGCCGGA AGCCGCA CGCGCAG CGCGGAT GCCATG GCCATG GCCATG	TATCAA TAAATA CTAAATA TAAATA TAAATA TAATTA TAATTA AGTTAA AGTTAA AGGCCA	GTTTCA GTTTCA TGTAAA TGTCAAA AAACAA TTCGT TTCGT TTCGGT AACCAT	ment he po lighted tions,
	The The 5 <td>526 GJ 5556 GY 3452 GG 3454 CC 344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3354 GG 752 S51 TT 772 CS 256 TT 772 CS 777 TT 722 CS 777 TT 722 CS 777 TT 722 CS 777 TT 772 CS 777 TT 772 CS 777 TT 772 CS 777 TT 772 CS 772 CS 772 TT 772 CS 772 CS</td> <td>153 A7 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 153 T3 152 T3 153 T3 15</td> <td></td> <td>Align tains 1 e high 1 posi</td>	526 GJ 5556 GY 3452 GG 3454 CC 344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3344 GG 3354 GG 752 S51 TT 772 CS 256 TT 772 CS 777 TT 722 CS 777 TT 722 CS 777 TT 722 CS 777 TT 772 CS 777 TT 772 CS 777 TT 772 CS 777 TT 772 CS 772 CS 772 TT 772 CS 772 CS	153 A7 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 152 T3 153 T3 152 T3 153 T3 15		Align tains 1 e high 1 posi
		6114 6114 6114 6114 6114 6114 6114 7114 7	CBLJ CBLJ CBLJ CBLJ CBC CBLJ CBC CBC CBC CBC CBC CBC CBC CBC CBC CB	TUF TUF	IRE 1. In con ons ar
	A C C C C C C C C C C C C C C C C C C C	PG BT PG BT PG BT PG BT PG BT PG BT BZ BT BZ BT BZ BZ BZ BZ BZ BZ BZ BZ BZ BZ BZ BZ BZ BZ BZ BZ B		Other LI CB LI BT FN BT FN BT FN BT FN BT FN BT FN BT CAU BT CAU BT CAU BT CAU BT CAU BT CAU BT CAU BT CAU BT CAU BT FN F	FIGU colun positi in cor

1088 RNA, Vol. 9, No. 9



FIGURE 2. Conserved structure of the *B12*-element. Capital letters indicate invariant positions. Lowercase letters indicate strongly conserved positions. Degenerate positions: R = A or G; Y = C or U; K = G or U; M = A or C; H = not G; D = not C; N = any nucleotide. Dashes and black dots indicate obligatory and facultative base pairs, respectively. The conserved helices are numbered P0 to P6. Stem–loops of variable and constant lengths are shown by broken and sold lines, respectively. Additional stem–loops are designated Add-I and Add-II. The sequence of the conserved B12-box motif is shaded in gray. The variable structure VS separates two conserved parts of the *B12*-element, BI and BII. Taxonomic variations of the VS topology are shown in the *inset*.

the presence of the latter stem-loop does not depend on the phylogenetic position of a genome and occurs in <10% of cases. The maximal observed lengths of the Add-I and Add-II stem-loops are 34 and 122 nucleotides, respectively.

The *B12*-elements can be classified into two major types based on the existence of a highly conserved stem-loop region, named BII. This part of the *B12*-element includes

conserved regions, 5'-GCCACTG-3' and 5'-YGGGAAGGC-3', which partially overlap the P5 and P6 helices (Fig. 2). Although most B12-elements are complete, the BII part is absent in a number of genomes: in all B12-elements of cyanobacteria, bacteria Deinococcus radiodurans, Bacillus subtilis, Shewanella oneidensis, Chloroflexus auranatiacus, as well as in some B12-elements of actinobacteria and Pseudomonas aeruginosa (Table 1). Short B12-elements without the BII part were found upstream of some B₁₂-related genes and, therefore, are functional. In complete B12-elements, the BII part is separated from other conserved parts of the B12-element by a variable linker, which can fold into a nonconserved variable structure, named VS. This structure has different topologies in various taxonomic groups of bacteria. For example, the VS structures are represented by one helix in bacteria from the Bacillus/Clostridium group, whereas VS of γ -proteobacteria consist of two adjacent helices. In other cases the VS structure is more complex

(see the lower frame in Fig. 2). Thus, the additional highly conserved BII part can play either an auxiliary role in the function of the *B12*-element or perform some other function.

The previously proposed secondary structures for the *cbiA* (Ravnum and Andersson 2001) and *btuB* (Nahvi et al. 2002) leader mRNAs from enterobacteria differ in some

TABLE 1.	Phylogenetic	distribution of	f B12-elements	in bacteria
----------	--------------	-----------------	----------------	-------------

	Conomos with	Number of	Type of B	12-elements	
Taxonomic group	B12-elements	B12-elements	complete	without BII	Proposed type of regulation
α-proteobacteria	11	58	all	no	Translational
β-proteobacteria	5	11	all	no	Translational
γ-proteobacteria	14	31	yes	3	Translational
δ-proteobacteria (GME)	1	1	all	no	Translational/transcriptional
The Bacillus/Clostridium group	14	37	yes	1	Transcriptional
Actinobacteria	7	22	yes	6	Translational
Cyanobacteria	5	8	no	all	Translational
The Thermus/Deinococcus group	1	3	no	all	Translational
The CFB group (BX, PG, CL)	3	17	all	no	Translational
Spirochetes (LI, TDE)	2	5	all	no	?
Chloroflexaceae (CAU)	1	2	no	all	Translational
Fusobacteriaceae (FN)	1	2	all	no	Transcriptional
Thermotogales (TM)	1	1	all	no	Transcriptional
Total	66	198	174	23	Transcriptional ~20%; translational ~80%

details in their topology. Here we propose a conserved secondary structure for 5'UTRs of B₁₂-regulated genes. Similarly to the riboflavin- and thiamin-specific 5'UTR regulatory RNAs, RFN and THI elements, this new B₁₂-specific RNA has a compact secondary structure, consisting of a set of conserved helices closed by a single base stem. The previously proposed RNA secondary structure of the cbiA regulatory region (Ravnum and Andersson 2001) is not compact, although a number of helices in both structures coincide (namely, P1, P2, P3, and P6). Moreover, despite of compactness of the predicted secondary structure of the btuB regulatory region (Nahvi et al. 2002) and similarity in topology of both structures, this structure differs from the B12-element in regions of base pairing of the base stem. However, the suggested conserved structure of the B12element is mostly consistent with the chemical probing data for the *btuB* and *cbiA* leader regions (Fig. 3).

Regulation of transcription/translation mediated by *B12*-element

Analysis of the leader sequences of *B12*-regulated genes allowed us to predict additional RNA regulatory hairpins



FIGURE 3. Predicted RNA secondary structure of the *B12*-elements upstream of *btuB* from *E. coli* and *cbiA* from *S. typhimurium* genes. Chemical modifications of the B12-loop region of the *btuB* (*A*) and *cbiA* (*B*) 5'UTRs are shown by filled triangles (Ravnum and Andersson 2001; Nahvi et al. 2002).

downstream of all B12-elements. In some cases these additional hairpins are followed by runs of thymidines and therefore are candidate p-independent terminators of transcription. In other cases the hairpins overlap the ribosomebinding site (RBS) of the first gene in the B12-regulated operon and are candidate translational sequestors that prevent ribosome binding to the RBS. Previously, it has been shown that RBS-sequestering hairpins are involved in the regulation of translation of the btuB gene of E. coli and S. typhimurium, as well as the cbiA gene of S. typhimurium. (Ravnum and Andersson 1997; Nou and Kadner 2000). Most Gram-positive bacteria from the Bacillus/Clostridium group, Chloroflexus aurantiacus, Fusobacterium nucleatum, and Termotoga maritima have a candidate terminator hairpin, whereas most Gram-negative bacteria (proteobacteria and the CFB group), as well as actinomycetes, cyanobacteria, and D. radiodurans have candidate RBS-sequestering hairpins downstream of B12-elements. Therefore, the regulation of B₁₂-related genes is likely to operate mainly at the level of transcription in the former group of bacteria and at the level of translation in the latter group (Table 1). Thus, the phylogenetic distribution of the proposed terminators and sequestors in Gram-positive and Gram-negative bacte-

> ria, respectively, is similar to previously observed distribution of the regulatory hairpins in the riboflavin and thiamin regulons (Vitreschak et al. 2002; Rodionov et al. 2002). Analysis of the 5'UTR of the cobalamin biosynthetic operon from *Geobacter metallireducens* reveals two possibilities of regulation. In this case the predicted terminator hairpin overlaps the RBS sequence of the first gene in the operon, and therefore, it can function as both a terminator and a sequestor.

> The mechanism of the Ado-CBL-dependent regulation involves formation of two alternative RNA conformations. It has been shown in experiments that the RBS sequestor of the *S. typhimurium cbi* operon forms in the presence of Ado-CBL, leading to inhibition of the translation initiation, whereas the antisequestor conformation formed in the absence of Ado-CBL opens the RBS sequence and allows translation to initialize (Ravnum and Andersson 2001).

Previously we proposed the riboswitch mechanism involving the *RFN* and *THI* elements. These elements are stabilized by effector molecules (flavin mononucleotide or thiamin pyrophosphate, respectively), allowing formation of downstream regulatory hairpins (transcriptional terminator or a hairpin sequestering the RBS site) that leads to repression of transcription or translation, respectively. In the absence of the effector, a more energetically favorable RNA structure that is alternative to parts of the vitamin-specific element and the regulatory hairpin can fold, thus releasing gene expression (Rodionov et al. 2002; Vitreschak et al. 2002). This riboswitch mechanism was experimentally confirmed for thiamin and riboflavin regulons (Mironov et al. 2002; Winkler et al. 2002a,b). In analogy with the riboflavin- and thiamin-specific riboswitches, we propose here the cobalamin-specific riboswitch.

It seems to have some specific properties. It has been experimentally shown that an additional regulatory element is located between the B12-element and the sequestor hairpin of the cbiA gene of S. typhimurium (Ravnum and Andersson 2001). This element, a translational enhancer, plays the key role in the regulation of translation initiation of cbiA. In the absence of cobalamin it interacts with the region, corresponding to the left stem of the RBS sequestor, and thus releases the translation initiation of cbiA. Moreover, it has been suggested (Ravnum and Andersson 2001) that in the presence of cobalamin, the enhancer element interacts with the B12-element, thus the RBS sequestor forms and repression the translation initiation is repressed.

To elucidate the mechanism of regulation, we analyzed multiple alignments of upstream regions of close ortologs of *btuB* genes from enterobacteria and some α -proteobacteria. It turned out that these alignments contained islands of conserva-

tion, corresponding to *B12*-elements, sequestor hairpins, and some additional elements in between (Fig. 4). The first additional conserved element is complementary to the left stem of the RBS sequestor. Thus, it can form an antisequestor structure whose possible role is similar to the translational enhancer. Another conserved region can form a pseudoknot by interacting with the loop region of the P3 helix of the *B12*-element. Similar additional elements in 5'UTRs of *B12*-regulated genes were found in most Gram-positive and Gram-negative bacteria (Fig. 5). In some cases the pseudoknot element overlaps the left base stem of the antisequestor/antiterminator, in other cases (including the cases shown in Fig. 4), it is located within the antisequestor/antiterminator. Thus, the predicted mechanism of the co-



FIGURE 4. Multiple alignment of the *btuB* upstream regions from enterobacteria (*A*) and some α -proteobacteria (*B*). Gray background denotes the P0 and P3 stems of the *B12*-element and the proposed RBS sequestors. Bold black denotes the main stem of the anti-sequestor. Arrows show the complementary stems of RNA secondary structures. Candidate pseudoknots are underlined and designated Pkn. RBSs and start codons are shown in bold letters.

balamin riboswitch differs from the thiamin and riboflavin riboswitches, as it involves formation of the pseudoknot (Fig. 6). The *B12*-element stabilized by Ado-CBL interacts with the antisequestor/antiterminator region, forms the pseudoknot, and promotes the sequestor formation. In the absence of Ado-CBL, the nonstabilized *B12*-element cannot form the pseudoknot and the more favorable of antisequestor/antiterminator structure folds, thus allowing for the translation initiation or transcriptional readthrough.

Phylogeny of the B12-elements

Global analysis of the *B12*-elements in available bacterial genomes demonstrates that this conserved RNA regulatory

		B12-el	ement					regulatory hairpin (terminator/RBS-sequestor)
	P0	P3	P3'	PO		antiterminatoriantisequestor		
	~~~~>	>	<	<			<===	
BIA BTUB	TCGGTGACGT	-GCTGCCCC	GCAACTGTAA	CORTCAGCOG	2	GTCACACGCGAA 9 CGGGG	46	ACCELOGTTCCCTCTCACCTCCCCCCCCCCCCCCCCCCCCCCCCC
RPA BTUB	GCGGTGACGGT	-GCTGTTCCC	GCAACTGTAA	CCGTCATTCG	2	GTCACACGCGAA 9 CGGGA	41	ACCTCGTTCGCTGTGACGTGCCACAGACGTCTGCCGAGGTTGAACCTTATG
BJA_BTUB3	CTCGCGACGGT	-GC <mark>TGC</mark> CCCC	C <mark>GCA</mark> ACTGTAA	CCGTCAGCCG 2	2	GTCACACGCGAA 9 CGGGG	46	GACCTCGTTCGCTGTGACGTGCCACTGACGTCATACCGACGTCCCGCTGTG
RPA BTUB3	CCAGCGACGGT	-GCTGCCCCC	GCAACTGTAA	CCGTCAGTCG 2	2	GTCACACGCGA- 10 CGGGG	53	TCGCCGCTGACGTGCCACGTTAGCCGCGAGCCAAGAAAATG
SAR BTUBE	ACGCCAGAGGT	-GCTGCCCCC	GCAACTGTAA	CCCTTGGT (	0	TGTCGTTCCGCG 6 CGGGG	58	CTCCCGCCATGAACGCCAGTTCCGTTCATGGGCCCGCCAGTCATG
CO BTUB	TCGCGTCAGGT	-GCTGCCCC	GCAACTGTAA	CCCGGCGCA- (	õ	GTCGTTCATCGC 5 CGGGG	20	TCCCGAGAAACGACAGTC60GTCGTTCATAACGCGAAGACTGTATTATG
CO_METE	GT <mark>TGTCGTGG</mark> T	-GCTGCCCC	GCAACTGTGA	CCTCGACAGA 11	1	GGCCCCCTGTCC	69	AGCCTCGCCAAAACACCGGCAGAGGCTTTTCAAAGATG
RPA_METE	GCCGTCGAGGT	-GC <mark>TGC</mark> CCCC	GCAACTGTTA	CCTCGTCGAA 12	2	GGG <mark>CCCCC</mark> TGCC	64	CCTTGCCCCCAACATTTTGGGGCTTGAGCCATG
BUA METE	CATGTCGTCGTCGT-	-TCTGCCCCC	GCAACIGIGA	CCACGACGAA 1	2	CCCCTCTCTCC	50	CACCCCCCACCCCCCCCCCCCCCCCCCCCCCCCCCCCC
MLO CFRX	CTCCAGACGGT	-GCTGTCCCC	GCAACTGTAA	CCGTCTGCGA 15	5	GCCCGGCGGGTT	1	CCGGGCAAGGAGCCCGGCTTTTGGAC
MLO_BTUD	TGC <mark>GTGATGG</mark> T	-GC <mark>TGC</mark> CCCC	GCAACTGTAA	CCATCACTGA 20	0	CGCGAGGAGACG	33	TCCTTCCGTCGCCAGTTTTGGAAAGTGACGGAAAGTTGGACG
RC_CFRX	CGG <mark>GCGCTGG</mark> T	-GC <mark>CGC</mark> CCCC	GCGACCGTGA	CCAGCGCATG 10	0	GGA <u>CGGGG</u> TGTT	37	CCGTTCCCCCAACCCCAAGAGGACGGGCGATCGCCACCACG
RS_CFRX	GGCGCGCTGGT	-GCCGCCCCC	GCGACCGTGA	CCAGCGCTGA	6	GGGCGGA <u>CGGG</u>	53	ATCGCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
BME BTUF	GTCATGACGGT	-GCTGCCCC	GCAACIGIAA	CCGTCTTACG 1	7	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	60	ATATCCCCGGA95TTCGGGGATATTTCATGAAA
AU_BTUF	AAAGCGACGGT	-GCTGCCCC	GCAACTGTAA	CCGTCAAAAT 18	8	CGGGGAAGCCCG	53	TTCTCCGTACAAGAC45GTTTTACTTTCGACTGGAGAACACATGT
MLO_ARDX	TCATCTCAGGT	-GC <mark>TGC</mark> C <u>CCC</u>	GCAACGGTGG	CCCGAGATTT 18	8	G <mark>CTCCC</mark> CGGTCA	48	CCTCAATCACCACCAGTTCAGTCCTTGAAGTATGAGGACAGGAC
SM_ARDX	ACACTCAAGGT	-GCTGCCCC	CAACGGTGG	CCTTGAAGCA 30	6	TCGGAGCCAATT 6 CGGG	30	CCTCTCCGAAACGGCGGAAAAACGAGGACAAGGGCATG
BIA CERX	CCGGCGTTGGT	-GCCGCCCC	GCGACCGTGA	CCAGCGCGGA 12	2	GCCGCACGCGCGT	49	TTCCCCCCCCCCCCCCTCT
RC04759	GTGGCGATGGT	-GCTGCCCC	GCAACTGTGA	CCATCGCAGA	6	GGGCGGGGTGAACCGGGG	37	GCCGTTTCCCCCGCCCCTGTTCAGGCAGGCGCGCGCATGGCGGTG
beta-prote	obacteria.							
BPS_HOXN	CGCGTTTCGGT	-GC <mark>TGC</mark> CCCC	GCAACGGTAA	CCGAGACA (	0	CCCCGTTC	37	CCCCAGGGTTCAGCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGC
NE DUID	CGTCGACCGGC	-GCTGCCCCC	GCAACTGTGA	COGGACGTT 9	9	CAAGCCTTCATC 3 CGGGG	65	CECTTERTC
MFL BTUB	AGTTTTGAGGT	-GCTGCCCCC	GCAACGGTAA	CCTGAAAACG	5	GGAAGTGCTGCGGGGA	35	TTCCCGCACACTCCATT32CTTTGCACTCATTAACTAATG
MFL BTUB2	TCACTTACGGT	-ACTGCCCCC	GCAACTGTAA	CCGCAAGTGA 15	5	AGGCCCCCCTGT J	121	CTCCGACCCTATAGCTGATTTA22AACAATCCACCAGGCAGCAATCCATG
MFL_NRDA	CCATCTACGGT	-GC <mark>TGC</mark> CCCC	CCAACTGTGA	CCGTAGATCA 14	4	GAGCCCCCCTGT 1	105	ACCCTGCTTCCTGACTGGGAGCCTAAGATGAC
RSO_HOXN	ACGATGATGGT	-GCTGCCCC	GCAACTGTAA	CCATCTCCTT 23	3	GCGGGGTGCACCGGGGGA	28	GTCCTCACTCGCGGCGATGACCCTTGTGGGGGACCTATCCGTGG
damma-prot	eobacteria	-GUIGULL	ATCAACGGTAA	CCAGGACAGI 43	3	ATCCGGCIGCCA 8 CGGGGG	54	accacaccaaAracaccasaccac
EC BTUB	CACTTGCCGGT	-GCTGACGC-	GCAGCGGTAA	CCGGCCAAC- (	0	GTCGCATCTG 11 CGCGT	15	TGCGGCATCCTTCTTCTATTGTGCATGCTTTACAATGA
YP_BTUB	TGTGGT <mark>CCGG</mark> C	-GCTGACGC-	-G <mark>CAGC</mark> GGTAA	CCGGTATTAC (	0	GTCGCAATAT 10 CGCGT	17	TGCGGCATCCGCCAATTGAGTTTGGATGCTCTCTTAATG
YE_BTUB	TGCGGT <mark>CCGG</mark> C	-GCTGACGC-	GCAGCGGTAA	CCGGTATAC- (	0	GTCGCAATAT 10 CGCGT	17	TGCGGCATCCGCTATTTGAGTTTGGATGCTTTCTTAATG
SY_BTUB	CGTGGGCCGGT	-GCTGACGC-	GCAGCGGTAA	CCCGGCTAAC- (	0	GTCGCATCTG 12 CGCGT	15	TGCGGCATCUTTCTTATATTGTGCATGCTTTACAATGA
EO BTUB	GAGCGTCCGGC	-GCTGACGC-	-GCAGCGGTAA	CCGGAATAC- (	0	GTCGCAATTG 11 CGCGT	15	TGCGGCATCCTTATTATCTCTCCGGATGCATTACAATG
PA BTUB	TTGCGACAGGT	-GCTGCCCC	GCAACGGTAG	CCTGACGCA- 12	2	CGGAGGGCGC	25	CGTCCTCCGTCTGCCTCG17GAGAGCGACGGACGCAGAGCGACGAGAGCATG
PA_BTUB2	CCGTTCCAGGT	-GC <mark>TGC</mark> C <u>CCC</u>	GCAACGGTAA	CCTGCAACG- 23	3	GCGGGAGCGA	40	GCGTTCCGTCCTGCCCGGTCATGAGATGGAACGCCCATGCCCCG
PP_BTUB2	CAGTTTCAGGT	-GCTGCCCC	GCAACGGTAA	CCTGGAGCT- 15	5	CGGTGCGCGGG	53	TTCCACCGGGATCCATTCATTCCTGCAGCAGTGGAACGACATGTCC
PY_COBW	GCTTTCGAGGT	-GCTGCCCCC	GCAACTGTAA	CETCGTCACA 18	3 8	GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	51	ATGCCCGCCACATTGCCAAAGGCACCTCATGAAAA
PP COBW	CACTTCGAGGT	-GCTGCCCC	GCAACTGTAA	CCTCGTCACG 15	5	CGGGCGGGGT	52	TGCCCGCGCCACACTGCCAACGGCAATGCGACATG
PA_COBW	CGGTTCGAGGT	-GCTGCCCC	GCAACTGTGA	CCTCGTCGAT 39	9	CGGGCGGGGT	50	GCATGCCCGCCGAGCTGAACAACGCCTGCCCATGAAAAC
SY_CBIA	AACCAACAGGT	-GCAGCCCC	GCTGCTGTGA	CCTGTCGGTG 14	4	CGET GEGAAG	59	CCACCCTGAACAGGATCAGGGTAATGGCGGCA
KP_CBIA	AACCGACAGGT	-GCAGCCCCC	GCTGCTGTGA	CCTGCCGGTA 20	0	GCCTGGCTTA	37	ATACCCTGACCGCCAGAACAGGACAGGGTATCGCAACCAGG
PP CBTAB	CGCGTTCAGGT	-GOTGOCCCC	GCAACGGTAA	CCTGGCGTTC 11	1	GCGGTGGGCGGGGGGG	37	SCCCTCCCCTATTCCCCACAAAAAGCAGACGATCGCGTCATG
PU CBTAB	TTGTTACGGGT	-GCTGCCCC	GCAACGGTAA	CCCGCAACAC 38	8	CGGTGGGCGGG	54	CCCGCTGACACTTCAGAGGGAAGCGCCATG
PY_CBTAB	TTGTT <mark>TCGGG</mark> T	-GC <mark>TGC</mark> CCCC	GCAACGGTAA	CCCGACATTT 20	0	TGGGCGGGCG	40	CATGGCCCATTTGCCAACACTCAAGACGCAACGCCATGTCCACCATCA
PA CBTAB	TTGTCCCAGGT	-GCTGCCCCC	GCAACGGTAA	CCTGGAACCT 20	0	TGGGCCCGCG	34	GTGCGTACTCCTCTGCCACGCCCATGCCAATCCAGAGGGAACGCACCATGTCCA
AV BTUB	TCCCTTCACCT-	-CTGTACCC	CAACIGIGAC	CTATTGCTG 29	6	CCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	3/0	CCCTTTTTCCCTCCCCCTTTCCCA23ATGCCTGTCGCGCCCACACGACCATGCCCCTTCGCCCCCCTTCGCCCCCCCC
The Bacill	us/Clostridium	group	ACCOTAC	COTO TO COM	U	6036160684666000 <u>0000</u>		
BS_BTUF	GTCAAATAGGT	-GCGGTCCC-	GCCACTGTAA	CCTGTTTGAT 25	5	GATGTGTAAACAATAGGC	6	GTGTTGTTTACAGCATCTTTACCGTCGTAGAGATGCTTTTTAGTT
BE_BTUF	CGCAGCAAGGT	-GCTGTCCCC	GCAACTGTCA	CCTTGCTTGC 1	7	CGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	45	CGCGCTCATTCCCAGATTTAGGCGGAATGGGCGTTTTTTATT
BE_CBIW	ACAGE AAGGC	-ACTGCCCC	CGCACTGTGG	TCTTTATTGT 2	0	GGCAGACGT <u>GCGGG</u> AGAATG	44	CATTTTTCTGCACCTCTGCTAGAGAGGTGCTTTTTATTT
HD NRDA	GACGGTAAGGT	-GCTGTCCCC	GCAACTGTGA	CCTTACTTCC 15	5	TCGGGGGTTG	85	AAGCCTTTCTCTCGAAGGAAGGTTTTTTATTTGG
EF_BTUF	ACTTATGAAGT	-ACAGACCTA	CCTACTGTAT	CTTGATAAGT 14	4	GTAGGAAAAA	51	ACTGCTTTTTTTCTATTGGAGAAAAAAGTGGTTTTTTTGTA
DF_CBIM	ATATTATAGGT	-ACAGCCCCC	<u>G</u> CTA <mark>CTGT</mark> GA	CCTAAAATAT 14	4	CGCGGATGGA	68	AGAAGCCTAACCATCGAGTTAGGCTTTTTTATTTTA
DF_CBIP	TAAAATTAGGT	-GCAGCCCCC	GCTACTGTGA	CCTAATTTAA 15	5	CTT <u>CGGGG</u> TGAAAGATTTTT	13	AAAAGTCCTATCTTTAGAAGTATTAGGACTTTTTTTAATTTG
DF BTUF	AATACTAGGT	-ACAGCCCCC	GCTACTGTAT	CCCAGTATTA 15	5	GGGTGGAGA	54	ACTTCTCCTTATTAGIAGCACAGAGGAGGAGTTTTTTATTTT
CPE BTUF	TAATAATTGGT	-ACTACCTC-	AGTAAC	CTTTTATT 15	C	TGAGGGTA	33	AACATACCTTCTTAAAATGGTATGTTTTTTATTT
CABTUF	ATATTTTAGGC	-GCAACCCCC	G TACTGTAT	CCTAAAATAT 10	0	CGGAGGGAA	82	CTCCGCAAAAGCGGAGTTTTTTCTT
CPE_CBLT	GAGC <mark>ATTAGG</mark> T	-GCAGCCCCC	GCTACTGTTG	CCTAGTATG- (	0	CTATTCTTA 6 CGGAGGG	37	AACTGGTGGGGGATAGTTATCCGTACCAGTTTTTTATTTT
CB_CBIP	AAGATTTAGGT	-GCAGCCCCC	GCTACTGTAA	CCTAGATTTT 2	2	TAGGGAAAC 3 CGGAGGG	59	ATTGCCTATAACTCTAAGTTTATAGCCAATTTTTATATT
HM001408	CGGTTACAGGT	-GCGGTCCC-	GCCGCTGTAT	CCTGTACACT 3	7	GTGGGATGA	3	AAAAGTCCCCAGATCTGATCATGCAGATCTGGGGACTTTTTGCTTTTCA
HMO CBID	ATGCAAAGGGT	-GCGGTGCC-	GCCACTGTGA	CCTG (	0	TCCCCGCACCGG	31	CTGGGGAAGATTAGGTTCCCTAGATTTTCATT
DHA_NRDD	ATCTCATAGGT	-ACGGTCCC-	-GCCG <mark>CTGT</mark> AA	CCTATGTGTA 27	7	ATGTTGGG	21	AGCTTCTCGGCATCTCTGTCGGGAAGTTTTGTTTCGG
THT_BTUR	AGCCTTATGGT	-GCAGCCCCC	GCTACTGTGA	CCATAAGGTTT10	0	CCTTCGGAGGGAAG	25	AAAAAGGTCTTCCTCTGAAGGGAAGGCCTTTTTTATATAAAA
Actipobact	AGCCTTATGGT	- OLAG <u>UUUUU</u>	SCIACIGIGA	CCATAAGGTTT11	T	CC11CleteAlater	54	AAAAAGTOTAGGTOGGAAGG
ML CBTG	GGCGATGACGA	-GCGGTCCC-	-GCCACTGTAA	CGTCATCGCG 11	1	CEE <mark>EEC</mark> EEC	58	GTCGCCACTATGGAGCGTTG-AATCGCATCCGATTACGCCTCGCCGCGTG
SX12454	GGCGACGACGG	-GCGGTCCC-	GCCACTGTGA	CGTCGTCGCC 14	4	CTGGGGCGG	4	CCCCGGAGAGAGGGGTGTTTCGGCGTG
DI_BTUC	ATTGATTAGGT	-GCTGACGC-	GCAGCGGTGA	CCTAATCAGA	1	ACATGTGCT 8 CGCGT	12	GACCGATATGTAGTTGAAGTTGGTCATGTGATC
SX_CBIM	CACCCCTTGGT	-GCGGCCCTC	GCCACTGTGA	CCAAGGCGCG 23	3	CCCCCCC	1.GC(	CEASTCATECACATAGCCGAGGGCTTTCTGCCACCGGCGCACGCGATCGCCTGGG
SX PDUX	CCGCAGTTGGT	-ACTGTCCC-	GCAGCGGTGA	CCACTGCGCC 3	7	GACTGGG 13 CGGGGC	26	CCTCCTTCCCGGTCGCATGCCAGGACGCGGGACAGGACGACGGTTTGTG
PI_CBIB	GCGCGTCCGCA	-GCTGTCCC-	-GCAACGGTGA	CTGTGGACGTG1	6	GCGCAGGG	3	CCACGAGCTGCCTGCGCGTGCACCGCGAAGGACGGCCCGTGGATTTGA

**FIGURE 5.** Conserved RNA elements upstream of some *B12*-regulated genes. The P0 and P3 stems of the *B12*-element are highlighted in green. Proposed regulatory hairpins (sequestors in proteobacteria and actinobacteria; terminators in the *Bacillus/Clostridium* group) are highlighted in blue. Red denotes the base stem of the antiterminator/antisequestor. Arrows in the *upper* line show the complementary stems of RNA secondary structures. Candidate pseudoknots that overlap the P3 loop and the antiterminator/antisequestor are underlined. RBSs and the start codons are set in blue.



FIGURE 6. Predicted mechanism of the B12-mediated regulation of CBL genes: (A) transcriptional attenuation; (B) translational attenuation (inhibition of translation initiation).

element is widely distributed in eubacteria and regulates most genes required for the CBL biosynthesis and some other genes (Table 2; for details see D.A. Rodionov, A.G. Vitreschak, A.A. Mironov, and M.S. Gelfand, in prep.). First, various *cbi* and *cob* biosynthetic genes are regulated by B12-elements in most CBL-synthesizing bacteria. These

genes either form single CBL gene clusters or are scattered along the chromosome. Moreover, genes for most known and predicted cobalt transporters, as well as cobalt chelatases and reductases, are often preceded by candidate B12elements. Indeed, cobalt ions are required for the de novo CBL biosynthesis. Second, the vitamin  $\mathrm{B}_{12}$  transport sys-

Gene cluster	Function	Taxonomic group
1. CBL biosynthesis		
cbi and cob	cobalamin biosynthesis	proteobacteria, the Bacillus/Clostridium group
cbt, hoxN, cbiMNQO, hupE	cobalt transporters	all CBL-synthesizing bacteria
orfl-cobW-cobN-chIID	cobalt chelation	$\alpha$ -, $\beta$ -proteobacteria, <i>Pseudomonadaceae</i> , actinobacteria
bluB	cobalt reduction	α-proteobacteria
btuR	CBL adenosyltransferase	α-, β-proteobacteria, <i>Pseudomonadaceae</i>
2. Vitamin $B_{12}$ transport		
btuB	vitamin B ₁₂ receptor	proteobacteria
btuFCD	vitamin B ₁₂ transporter	$\alpha$ -, $\beta$ -proteobacteria, <i>Pseudomonadaceae</i> , the
	(ABC components)	Bacillus/Clostridium and CFB groups, Deinococcus
		radiodurans, actinobacteria, spirochetes, Fusobacteriacea
		Thermotogales, Chloroflexaceae
3. B ₁₂ -dependent or alternative meta	abolic pathways	
metE	methionin synthase	various groups
nrd	ribonucleotide reductase	various groups
ardX-frdX	predicted enzymes	α-proteobacteria
achX	predicted enzymes	Deinococcus radiodurans and some other species

tems, *btuBFCD* in Gram-negative bacteria and *btuFCD* in Gram-positive bacteria, are mostly *B12*-regulated. Third,  $B_{12}$ -regulons of various bacteria are predicted to include enzymes from known  $B_{12}$ -dependent or alternative pathways, as well as several hypothetical enzymes from unknown pathways. In particular,  $B_{12}$ -independent izozymes of the methionine synthase and ribonucleotide reductase are *B12* regulated in bacteria that have both  $B_{12}$ -dependent and  $B_{12}$ -independent izozymes.

Using multiple alignment of identified *B12*-element sequences without additional nonconserved interior loops and the BII part, we constructed the maximum likelihood phylogenetic tree for these RNA elements (Fig. 7). The BII part was not used as it is not obligatory for all *B12*-elements.

The tree of *B12*-elements has a number of branches that correspond to taxonomic groups, for instance, the *Bacillus/Clostridium* group or cyanobacteria. Comparison of the *B12*-element phylogenetic tree with the standard trees for ribosomal proteins (Wolf et al. 2001) reveals both lineage-specific and gene-specific branches, as well as recent genome-specific duplications and horizontal transfer events.

The branch of *B12*-elements found upstream of the *cbi* operons of enterobacteria is the most obvious example of possible horizontal transfer, as it clusters with various *B12*-elements from the *Bacillus/Clostridium* group (Fig. 7), and the same holds for phylogenetic trees constructed for each gene from the *cbi* operon (data not shown). These observations allow us to suggest that the complete transcriptional



**FIGURE 7.** Phylogenetic tree of *B12*-elements. The names of the proximal genes of the *B12*-regulated operons are given. The genome abbreviations are listed in Table 3. *B12*-elements without the BII part are set in bold and shaded in gray. Groups of genome-specific *B12*-elements are circled by dotted lines.

unit, including the regulatory *B12*-element and the *cbi* operon, has been likely transferred from the *Bacillus/Clostridium* group to three enterobacterial genomes.

The phylogenetic tree of B12-elements contains a number of genome-specific branches. In particular, four B12-elements found upstream of the pccC, cbtD, nrdA, and metE genes from Bacteroides fragilis form a separate branch (Fig. 7). Furthermore, in D. radiodurans, Burkholderia pseudomallei, Rhodopseudomonas palustris, Rhodobacter capsulatus, Heliobacillus mobilis, Sphingomonas aromaticivorans, Streptomyces colelicolor, Clostridium difficile, Porphyromonas gingivalis, Chlorobium tepidum, and Chloroflexus aurantiacus, we observed other organism-specific branches containing two or three B12-elements. In all of these cases, the regulatory elements are located upstream of various nonhomologous genes. Thus, it seems likely that the evolution of the B12-elements often involved independent lineage-specific duplications with subsequent transfer to a position upstream of a new gene. On the other hand, the phylogenetic tree has branches that correspond to B12-elements occurring upstream of orthologous genes, for example, the btuB genes from enterobacteria, the cobW genes from pseudomonads, and the *bluB* and *metE* genes from  $\alpha$ -proteobacteria. These observations indicate that these B12-elements have co-evolved with the corresponding B12-regulated genes.

Finally, the constructed phylogenetic tree allows us to propose a possible origin of the B12-elements lacking the highly conserved BII part (see bold elements in Fig. 7). Most B12-elements without the BII part, for example, all B12elements from early-diverged bacteria (D. radiodurans, Chloroflexaceae, and cyanobacteria), form a separate branch in the phylogenetic tree. The BII part seems to appear after divergence of these taxonomic groups (recall that the BII part was not used in constructing the tree). However, a B12-element from B. subtilis lacking the BII part is clustered with several complete B12-elements from other Bacillus species. Moreover, another BII-deficient B12 element from Pseudomonas aeruginosa clusters with other Pseudomonas B12-elements with both BI and BII parts. The absence of BII parts in these two B12-elements likely is a consequence of late deletions in these species. Thus, although the additional BII part is highly conserved in a large number of B12elements, it is possibly not obligatory and can play an auxiliary role in the functioning of the B12-element.

### CONCLUSIONS

Regulation of most  $B_{12}$ -related genes in bacteria appears to operate through a unique RNA structural element. The *B12*-element is characterized by its compact secondary structure with a number of conserved helices and extended regions of sequence conservation, which could be necessary for specific metabolite binding. Recently, it has been shown that Ado-CBL specifically binds the leader mRNAs of CBL- related genes from enterobacteria (Nahvi et al. 2002). However, the *B12*-element also contains the highly conserved BII structure, which is not obligatory for all bacterial *B12*-elements, being absent in deeply branching bacterial groups and in some other cases. The role of this additional part of the *B12*-element is not clear.

On the whole, the mechanism of regulation of vitamin B₁₂-related genes is similar to the previously proposed mechanisms of regulation of riboflavin- and thiamin-related genes (Rodionov et al. 2002; Vitreschak et al. 2002). At that, a highly conserved RNA element (RFN, THI, or B12) is stabilized by direct binding of the effector (flavin mononucleotide, thiamin pyrophosphate, or adenosylcobalamin, respectively). Thus, an adjacent regulatory hairpin, terminator or sequestor, can fold that leads to transcriptional or translational repression of vitamin-related genes. In the absence of the effector, the unstable vitamin-specific RNA element is replaced by an alternative antiterminator or antisequestor RNA conformation allowing for transcription readthrough or translation initiation. Interestingly, the observed transcriptional and translational types of regulation have similar taxonomic distribution for each of the analyzed vitamin regulons. At that, the transcriptional termination occurs mostly in Gram-positive organisms, whereas the inhibition of translation initiation happens in Gram-negative proteobacteria, D. radiodurans, the CFB group and cyanobacteria.

This study once again demonstrates the power of the simultaneous positional and phylogenetic analysis of regulatory elements and genes. This approach provides an opportunity for identification of new RNA regulatory elements in bacterial genomes and prediction of the regulation. Moreover, it uncovers the traces of evolutionary events such as horizontal transfer and lineage-specific duplication of genes and regulatory elements.

### MATERIALS AND METHODS

Complete and partial sequences of bacterial genomes were downloaded from GenBank (Benson et al. 2003). Preliminary sequence data were also obtained from the www sites of the Institute for Genomic Research (http://www.tigr.org), the University of Oklahoma's Advanced Center for Genome Technology (http://www.genome.ou.edu/), the Wellcome Trust Sanger Institute (http://www.sanger.ac.uk/), the DOE Joint Genome Institute (http://jgi.doe.gov), and the ERGO Database (http://ergo. integratedgenomics.com/ERGO) (Overbeek et al. 2003). The genome abbreviations from the ERGO database are used throughout and listed in Table 3.

The conserved secondary structure of the *B12*-element was derived using the RNAMultAln program (A.A. Mironov, unpubl.). This program simultaneously creates a multiple alignment and a conserved secondary structure for a set of RNA sequences. The RNA-PATTERN program (Vitreschak et al. 2001) was used to search for new *B12*-elements in bacterial genomes. The input RNA pattern described both the RNA secondary structure and the se-

Tax	Genome	Abbreviation
α	Agrobacterium tumefaciens	AU
	Bradyrhizobium japonicum	BJA
	Brucella melitensis	BME
	Caulobacter crescentus	CO
	Mesorhizobium loti	MLO
	Rhodobacter capsulatus	RC
	Rhodobacter sphaeroides	RS
	Rhodopseudomonas palustris	RPA
	Sinorhizobium meliloti	SM
	Sphingomonas aromaticivorans	SAR
3	Burkholderia pseudomallei	BPS
	Methylobacillus flagellatus	MFL
	Nitrosomonas europaea	NE
	Ralstonia eutropha	REU
	Ralstonia solanacearum	RSO
v	Escherichia coli	FC
1	Salmonella typhimurium	SY
	Klebsiella preumoniae	KP
	Versinia enterocolítica	VE
	Versinia eneroconica	VP
	Fruinia carotovora	50
	Nibrio Cholerae Decudomonae acruginosa	
	Pseudomonas aeruginosa	rA DD
	Pseudomonas putida	PP
	Pseudomonas fluorescens	PU
	Pseudomonas syringae	PY
	Shewanella oneidensis	SON
	Azotobacter vinelandii	AV
	Xanthomonas axonopodis	XAX
6	Geobacter metallireducens	GME
	Chloroflexus aurantiacus #	CAU
	Fusobacterium nucleatum	FN
	Thermotoga maritima	TM
3/C	Bacillus subtilis	BS
	Bacillus cereus	ZC
	Bacillus megaterium	BI
	Bacillus halodurans	HD
	Bacillus stearothermophilus	BE
	Listeria monocytogenes	LMO
	Clostridium acetobutylicum	CA
	Clostridium perfringes	CPE
	Clostridium botulinum	CB
	Clostridium difficile	DF
	Thermoanaerobacter tengcongensis	THT
	Enterococcus faecalis	EF
	Heliobacillus mobilis	HMO
	Desulfitobacterium balfniense	DHA
Act	Corvnebacterium dinhtheriae	DI
ici	Mycobacterium tuberculosis	MT
	Mycobacterium Ioprao	N41
	Thormohifida fusca	
	Phodosoccus str	
	Stroptomycos coolicolor	
	Propionicibastorium champer ii	
~	A selectoria cierium snermanii	PI
_ya	Anabaena sp.	AN
	Prochlorococcus marinus.	PMA
	Synechocystis sp.	CY
	Synochococcus sn	SN
	Synechococcus sp.	514

**TABLE 3.** Taxonomy and abbreviations for bacterial genomes used in this work

Tax	Genome	Abbreviation
CFB	Porphyromonas gingivalis	PG
	Bacteroides fragilis	BX
	Chlorobium tepidum	CL
T/D	Deinococcus radiodurans	DR
SP	Treponema denticola	TDE
	Leptospira interrogans	LI

The names of taxonomic groups in Tax column ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , B/C, Act, Cya, CFB, T/D and SP) stand for  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\delta$ -proteobacteria, the *Bacillus/Clostridium* group, actinobacteria, cyanobacteria, the CFB group, the *Thermus/Deinococcus* group and spirochetes, respectively.

quence consensus motifs. The RNA secondary structure was described as a set of the following parameters: the number of helices, the length of each helix, the loop lengths, and the description of the topology of helix pairs. Additional RNA secondary structures, in particular antiterminators and antisequestors, were predicted using Zuker's algorithm of free energy minimization (Lyngso et al. 1999) implemented in the Mfold program (http://bioinfo.math. rpi.edu/~mfold/rna).

Protein similarity search was done using the Smith-Waterman algorithm implemented in the GenomeExplorer program (Mironov et al. 2000). Orthologous proteins were initially defined by the best bidirectional hit criterion (Tatusov et al. 2000) and if necessary, confirmed by construction of phylogenetic trees. The phylogenetic trees of the *B12*-elements and  $B_{12}$ -related proteins were created by the maximum likelihood method implemented in PHYLIP (Felsenstein 1981). Multiple sequence alignments were constructed using CLUSTALX (Thompson et al. 1997).

Complete FASTA sequences of *B12*-elements, including the additional and variable regions are available from the authors (L.V. at l_veter@mail.ru or M.G. at gelfand@ig-msk.ru).

### ACKNOWLEDGMENTS

We are grateful to Andrei Osterman for attention, advice, and encouragement. This study was partially supported by grants from the Ludwig Institute for Cancer Research and the Howard Hughes Medical Institute (55000309).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received April 16, 2003; accepted June 3, 2003.

### REFERENCES

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.I. 2003. GenBank. *Nucleic Acids Res.* **31**: 23–27.

Dandekar, T., Beyer, K., Bork, P., Kenealy, M.R., Pantopoulos, K., Hentze, M., Sonntag-Buck, V., Flouriot, G., Gannon, F., and Schreiber, S. 1998. Systematic genomic screening and analysis of mRNA in untranslated regions and mRNA precursors: Combining experimental and computational approaches. *Bioinformatics* 14: 271–278.

- Eddy, S.R. and Durbin, R. 1994. RNA sequence analysis using covariance models. *Nucleic Acids Res.* 22: 2079–2088.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17: 368–376.
- Franklund, C.V. and Kadner, R.J. 1997. Multiple transcribed elements control expression of the *Escherichia coli btuB* gene. J. Bacteriol. 179: 4039–4042.
- Gelfand, M.S., Novichkov, P.S., Novichkova, E.S., and Mironov, A.A. 2000. Comparative analysis of regulatory patterns in bacterial genomes. *Brief Bioinform.* 1: 357–371.
- Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. 2003. Rfam: An RNA family database. *Nucleic Acids Res.* 31: 439–441.
- Grundy, F.J. and Henkin, T.M. 1998. The S box regulon: A new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol. Microbiol.* **30**: 737–749.
- Lundrigan, M.D., Koster, W., and Kadner, R.J. 1991. Transcribed sequences of the *Escherichia coli btuB* gene control its expression and regulation by vitamin B₁₂. *Proc. Natl. Acad. Sci.* **88**: 1479–1483.
- Lyngso, R.B., Zuker, M., and Pedersen, C.N. 1999. Fast evaluation of internal loops in RNA secondary structure prediction. *Bioinformatics* 15: 440–445.
- Marck, C. and Grosjean, H. 2002. tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* 8: 1189–1232.
- Martens, J.H., Barg, H., Warren, M.J., and Jahn, D. 2002. Microbial production of vitamin B₁₂. *Appl. Microbiol. Biotechnol.* **58**: 275–285.
- Mironov, A.A., Vinokurova, N.P., and Gelfand, M.S. 2000. Genome-Explorer: Software for analysis of complete bacterial genomes. *Mol. Biol.* 34: 222–231.
- Mironov, A.S., Gusarov, I., Rafikov, R., Lopez, L.E., Shatalin, K., Kreneva, R.A., Perumov, D.A., and Nudler, E. 2002. Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell* 111: 747–756.
- Murthy, V.L. and Rose, G.D. 2003. RNABase: An annotated database of RNA structures. *Nucleic Acids Res.* **31:** 502–504.
- Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., and Breaker, R.R. 2002. Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9:** 1043–1049.
- Nou, X. and Kadner, R.J. 1998. Coupled changes in translation and transcription during cobalamin-dependent regulation of *btuB* expression in *Escherichia coli. J. Bacteriol.* **180:** 6719–6728.
- ———. 2000. Adenosylcobalamin inhibits ribosome binding to *btuB* RNA. *Proc. Natl. Acad. Sci.* **97:** 7190–7195.
- Overbeek, R., Larsen, N., Walunas, T., D'Souza, M., Pusch, G., Selkov Jr., E., Liolios, K., Joukov, V., Kaznadzey, D., Anderson, I., et al.

2003. The ERGO(TM) genome analysis and discovery system. *Nucleic Acids Res.* **31:** 164–171.

Ravnum, S. and Andersson, D.I. 1997. Vitamin  $B_{12}$  repression of the *btuB* gene in *Salmonella typhimurium* is mediated via a translational control which requires leader and coding sequences. *Mol. Microbiol.* **23**: 35–42.

— 2001. An adenosyl-cobalamin (coenzyme-B₁₂)-repressed translational enhancer in the cob mRNA of *Salmonella typhimurium. Mol. Microbiol.* **39:** 1585–1594.

- Richter-Dahlfors, A.A. and Andersson, D.I. 1992. Cobalamin (vitamin B₁₂) repression of the *cob* operon in *Salmonella typhimurium* requires sequences within the leader and the first translated open reading frame. *Mol. Microbiol.* 6: 743–749.
- Richter-Dahlfors, A.A., Ravnum, S., and Andersson D.I. 1994. Vitamin B₁₂ repression of the *cob* operon in *Salmonella typhimurium*: Translational control of the *cbiA* gene. *Mol. Microbiol.* 13: 541–553.
- Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. 2002. Comparative genomics of thiamin biosynthesis in procaryotes: New genes and regulatory mechanisms. *J. Biol. Chem.* 277: 48949–48959.
- Szymanski, M., Erdmann, V.A., and Barciszewski, J. 2003. Noncoding regulatory RNAs database. *Nucleic Acids Res.* **31**: 429–431.
- Tatusov, R.L., Galperin, M.Y., Natale, D.A., and Koonin, E.V. 2000. The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**: 33–36.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The CLUSTAL_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**: 4876–4882.
- Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. 2001. The RNApattern program: Searching for RNA secondary structure by the pattern rule. In *Proceedings of the 3rd International Conference* on "Complex Systems: Control and Modeling Problems.", September 4–9, 2001, pp. 623–625. The Institute of Control of Complex Systems, Samara, Russia.
- Vitreschak, A.G., Rodionov, D.A., Mironov, A.A., and Gelfand, M.S. 2002. Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res.* **30**: 3141–3151.
- Wolf, Y.I., Rogozin, I.B., Grishin, N.V., Tatusov, R.L., and Koonin, E.V. 2001. Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol. Biol.* 1: 8.
- Winkler, W., Nahvi, A., and Breaker, R.R. 2002a. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952–956.
- Winkler, W., Cohen-Chalamish, S., and Breaker, R.R. 2002b. An mRNA structure that controls gene expression by binding FMN. *Proc. Natl. Acad. Sci.* 99: 15908–15913.