Comparative genomic analysis of T-box regulatory systems in bacteria

Alexey G. Vitreschak, Andrei A. Mironov, Vassily A. Lyubetsky, et al.

RNA 2008 14: 717-735
Access the most recent version at doi:10.1261/rna.819308

Comparative genomic analysis of T-box regulatory systems in bacteria

ALEXEY G. VITRESCHAK,[1] ANDREI A. MIRONOV,[1,2,3] VASSILY A. LYUBETSKY,[1] and MIKHAIL S. GELFAND[1,2]

[1]Institute for Information Transmission Problems (The Kharkevich Institute), Russian Academy of Sciences, Moscow 127994, Russia
[2]Department of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Moscow 119992, Russia
[3]State Scientific Center GosNIIGenetika, Moscow 117545, Russia

ABSTRACT

T-box antitermination is one of the main mechanisms of regulation of genes involved in amino acid metabolism in Gram-positive bacteria. T-box regulatory sites consist of conserved sequence and RNA secondary structure elements. Using a set of known T-box sites, we constructed the common pattern and used it to scan available bacterial genomes. New T-boxes were found in various Gram-positive bacteria, some Gram-negative bacteria (δ-proteobacteria), and some other bacterial groups (Deinococcales/Thermales, Chloroflexi, Dictyoglomi). The majority of T-box-regulated genes encode aminoacyl-tRNA synthetases. Two other groups of T-box-regulated genes are amino acid biosynthetic genes and transporters, as well as genes with unknown function. Analysis of candidate T-box sites resulted in new functional annotations. We assigned the amino acid specificity to a large number of candidate amino acid transporters and a possible function to amino acid biosynthesis genes. We then studied the evolution of the T-boxes. Analysis of the constructed phylogenetic trees demonstrated that in addition to the normal evolution consistent with the evolution of regulated genes, T-boxes may be duplicated, transferred to other genes, and change specificity. We observed several cases of recent T-box regulon expansion following the loss of a previously existing regulatory system, in particular, arginine regulon in *Clostridium difficile* and methionine regulon in Lactobacillaceae. Finally, we described a new structural class of T-boxes containing duplicated terminator–antiterminator elements and unusual reduced T-boxes regulating initiation of translation in the Actinobacteria.

Keywords: comparative genomics; amino acid biosynthesis and transport; T-box; antitermination; regulatory 5′-UTR mRNAs; bacteria

INTRODUCTION

The bacteria use a wide range of regulatory mechanisms to control gene expression. While the most common regulatory mechanism seems to be regulation of transcription by DNA-binding proteins (van Nimwegen 2003; Rodionov 2007), there are other important mechanisms, in particular, regulation of transcription (by premature termination) and translation (by interference with initiation) via formation of alternative RNA structures in 5′-untranslated gene regions (Henkin and Yanofsky 2002; Gutierrez-Preciado et al. 2005; Merino and Yanofsky 2005; Gelfand 2006; Weinberg et al. 2007). One particular set of genes that are frequently regulated by RNA-mediated mechanisms are the genes involved in metabolism of amino acids. In Gram-negative bacteria, the concentration of amino acids is measured indirectly by the rate of translation of the leader peptide in transcription attenuators (Yanofsky 1988; Merino and Yanofsky 2005). In the Firmicutes, these genes may be regulated by riboswitches: S-boxes, also known as SAM riboswitches, regulate methionine and cysteine genes in response to concentration of a methionine derivative, *S*-adenosylmethionine (Grundy and Henkin 1998; Winkler et al. 2003; Rodionov et al. 2004); lysine riboswitches (LYS-elements, or L-boxes) regulate lysine metabolism (Grundy et al. 2003; Rodionov et al. 2003; Sudarsan et al. 2003); glycine riboswitches regulate the glycine cleavage *gcvT* operon (Mandal et al. 2004). Regulation of tryptophan biosynthesis and transport genes in *Bacillus subtilis* and a number of related bacteria is mediated by a RNA-binding protein, TRAP, but again this regulation involves formation of alternative RNA structures (Babitzke 2004).

However, the most frequent mechanism of RNA-dependent regulation of amino acid operons in the Firmicutes seems to be the T-box regulatory system (Grundy and Henkin 2003; Merino and Yanofsky 2005). The T-box is an

Reprint requests to: Alexey G. Vitreschak, Institute for Information Transmission Problems (The Kharkevich Institute), Russian Academy of Sciences, Bolshoj Karetny pereulok 19, Moscow 127994, Russia; e-mail: l_veter@mail.ru; fax: 10-095-6506579.

Article and publication are at http://www.rnajournal.org/cgi/doi/10.1261/rna.819308.

RNA structure that is capable of binding uncharged tRNA via an interaction between the highly conserved 5′-UGGN-3′ sequence of the T-box and the complementary 5′-NCCA-3′ end of the tRNA (Putzer et al. 1995; van de Guchte et al. 1998; Grundy and Henkin 1993). The specificity of this binding is defined by base-pairing of the tRNA anticodon and the so-called specifier (anti-anti)-codon in the T-box structure (Grundy et al. 1994, 1997b, 2002b; Marta et al. 1996). The bound uncharged tRNA stabilizes the antiterminator hairpin, which in turn prevents formation of the terminator and allows the gene to be expressed (Fig. 1A); the structural and kinetic details of this process are now understood in considerable detail (Luo et al. 1998; Pelchat and Lapointe 1999; Grundy et al. 2000; van de Guchte et al. 2001; Grundy and Henkin 2004; Fauzi et al. 2005; Yousef et al. 2005).

Initially identified in *B. subtilis* as a mechanism regulating expression of some aminoacyl-tRNA synthetase genes (Henkin et al. 1992; Putzer et al. 1992), T-boxes were then demonstrated to regulate genes encoding amino acid biosynthetic enzymes and transporters (Henkin 1994; Pelchat and Lapointe 1999; Sarsero et al. 2000), both in *B. subtilis* and other Firmicutes (Grundy and Henkin 1994; Grundy et al. 1997a; van de Guchte et al. 1998; Delorme et al. 1999; Panina et al. 2003). An important role in these studies was played by the comparative genomics approaches, since the relatively large size and high level of sequence and structural similarity of T-boxes made possible highly reliable predictions (Henkin et al. 1992; Putzer et al. 1992; Grundy and Henkin 1994; Chopin et al. 1998; Delorme et al. 1999; Grundy et al. 2002a; Mwangi and Siggia 2003; Panina et al. 2003; Rodionov et al. 2004; Gutierrez-Preciado et al. 2005). In particular, these studies demonstrated that T-boxes may regulate not only premature termination of transcription, but initiation of translation, since in the Actinobacteria, T-box alternative structures overlap with the Shine-Dalgarno boxes (Fig. 1C; Seliverstov et al. 2005).

Using a sample of known T-boxes, we constructed a search pattern and used it to scan available bacterial genomes. We predicted the specificity of amino acid transporters analyzing the specifier codons in T-boxes that regulate genes encoding these transporters and use other methods of comparative genomics to obtain additional, independent evidence. The regulation by T-boxes seems to be rather flexible and labile: in some cases, only a few genes in a set of orthologs are regulated by this mechanism. This may be caused by the diversity of systems regulating amino acid biosynthesis (Gutierrez-Preciado et al. 2005; Panina et al. 2003; Rodionov et al. 2004). In particular, we observed several cases of T-box regulon expansion caused by apparent loss of other regulatory systems. T-box sequences are sufficiently long to construct phylogenetic trees, although of course the order of deep branching events cannot be resolved. Still, analysis of reliable terminal branches revealed a number of interesting features, in particular, T-box duplications both in situ and accompanied by moving to another region, as well as changes in specificity. Finally, we described a new class of T-boxes, where a single specifier hairpin is followed by two repeated antiterminator/terminator structures (Fig. 3), and an unusual form of reduced T-boxes in the Actinobacteria (Fig. 1D).

RESULTS

T-box regulons

Initial scanning using the RNA-pattern program identified 805 T-boxes in 96 bacterial genomes. T-boxes were widely distributed in Gram-positive bacteria, mainly in the Firmicutes, but also in the Actinobacteria. Moreover, T-boxes were found in some Gram-negative bacteria (δ-proteobacteria) and other groups (Deinococcales/Thermales, Chloroflexi, Dictyoglomi). The specifier codon was clearly identifiable in 765 T-boxes (95%).

Preliminary functional annotation of T-box-regulated genes (including parts of candidate operons) was done by similarity search. Genes regulated by T-boxes (Supplemental Table S1) encode aminoacyl-tRNA synthetases, biosynthetic enzymes (Table 1), and amino acid transporters (Table 2). T-boxes responding to 19 amino acids (excluding glutamate) were identified. Functional and taxonomic distribution of T-boxes varied between amino acids. For example, only one GLN-T-box was observed, upstream of the *gltX* gene in *Clostridium perfringens*.

Aminoacyl-tRNA synthetases

In most Firmicutes, T-boxes regulate aminoacyl-tRNA synthetases for aromatic and branched-chain amino acids, serine, asparagine, aspartate, and glycine. The least diverse taxonomic distribution was observed for T-boxes regulating prolyl- and lysyl-tRNA synthetases. On the other hand, in the Bacillales, T-boxes regulate aminoacyl-tRNA synthetases for most amino acids. Relatively fewer T-boxes regulating aminoacyl-tRNA synthetases were observed in *Streptococcus* spp. Outside Firmicutes, T-boxes regulate some aminoacyl-tRNA synthetases in the Actinobacteria (isoleucine), *Atopobium minutum* (phenylalanine), Deinococcales/Thermales (isoleucine, valine, glycine), Chloroflexi (isoleucine, leucine), Dictyoglomi (threonine), and *Thermomicrobium roseum*.

The specificity of T-boxes suggested by the specifier codon is largely consistent with the functional specificity of regulated aminoacyl-tRNA synthetases predicted from homology. One case where the T-box specificity is different from the annotated aminoacyl-tRNA synthetase specificity is the *metS2* gene of *Clostridium beijerinckii*. The initial reason for the functional annotation of this gene, present in the Clostridiales as the methyonyl-tRNA synthetase, probably was weak similarity (17% identity) to *metG* of

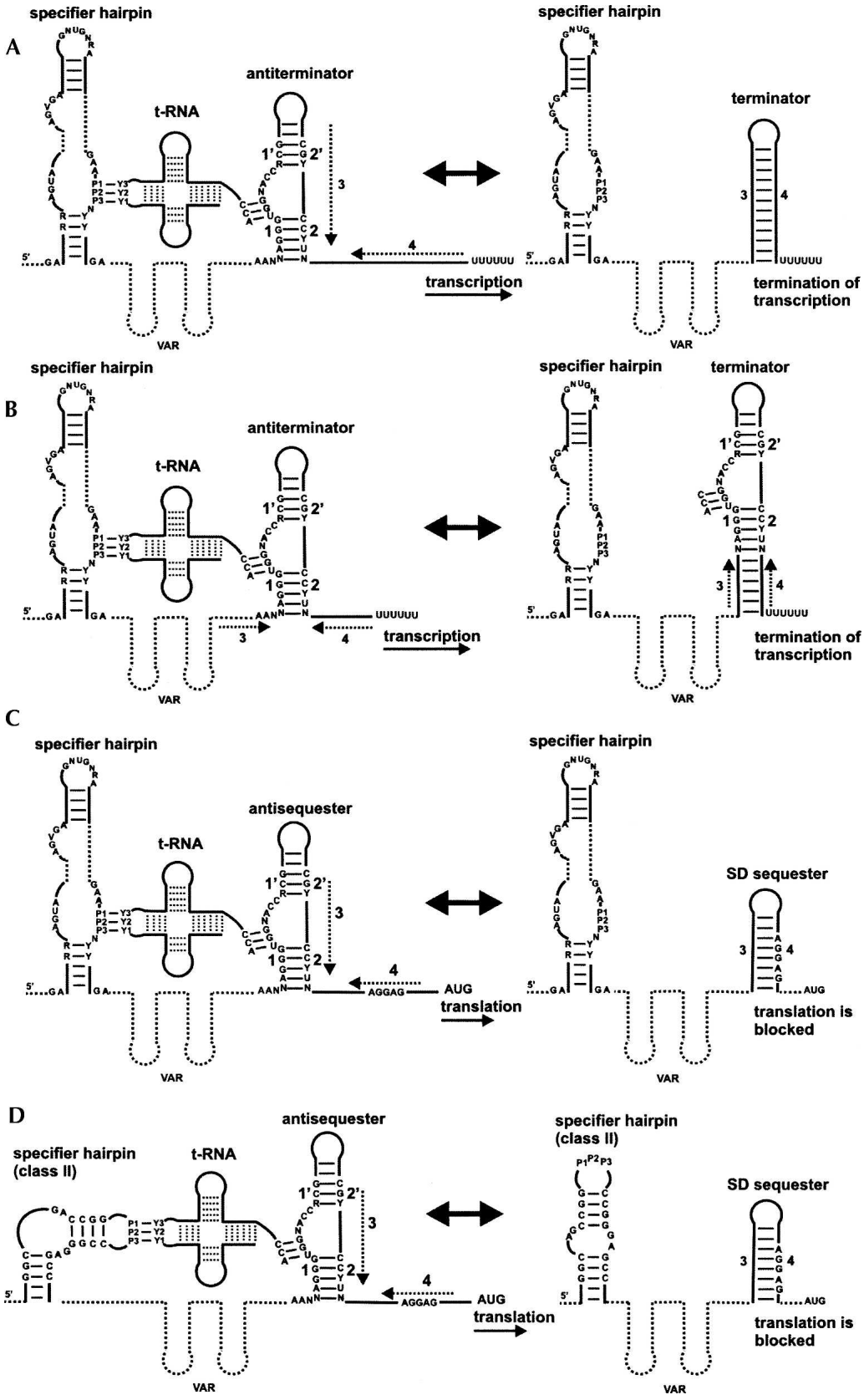

FIGURE 1. T-box regulation mechanism. (*A*,*B*) Transcription antitermination; (*C*,*D*) regulation of translation initiation. (*D*) The noncanonical variant of T-boxes in the Actinobacteria. The T-box structure contains (solid line) conserved base-paired regions and (VAR, dotted line) variable or facultative ones. The alternative antiterminator/antisequestor and terminator/sequestor stems are numbered 1 (1′)–2 (2′) and 3–4, respectively. (P1-P2-P3) The T-box specifier codon; (Y1-Y2-Y3) the tRNA anticodon; (AGGAG) the ribosome-binding site.

TABLE 1. Regulation of aminoacyl-tRNA synthetases and amino acid biosynthetic genes by T-box antitermination in bacteria

T-box specificity	Phylogenetic distribution of T-boxes
Aminoacyl-tRNA synthetases	
Aromatic amino acids	Most Firmicutes, *Atopobium minutum*
Branched-chain amino acids	Most Firmicutes, Actinobacteria (*ileS*), Deinococcales/Thermales (*ileS, valS*), Chloroflexi (*ileS*), *Thermomicrobium roseum* (*leuS*)
Methionine	Bacillales, Clostridiales, *Thermoanaerobacter tengcongensis*
Proline	Some Bacillales, Clostridiales
Cysteine	Bacillales, some Lactobacillales, Clostridiales, Thermoanaerobacteriales
Histidine	Bacillales, Lactobacillales (except *Streptococcus* spp.), some Clostridiales, *T. tengcongensis*
Arginine	Bacillales, Lactobacillales (except *Streptococcus* spp.), Clostridiales
Threonine	Bacillales, Lactobacillales, Clostridiales, Dictyoglomi, *T. roseum*
Serine	Most Firmicutes
Alanine	Bacillales, Lactobacillales, Clostridiales
Asparagine and aspartate	Most Firmicutes (except *Streptococcus* spp., Mycoplasmatales, Entomoplasmatales)
Glycine	Most Firmicutes, Deinococcales/Thermales
Lysine	*Bacillus cereus, Clostridium thermocellum*
Amino acid biosynthetic genes	
Aromatic amino acds	Most Firmicutes, Chloroflexi and Dictyoglomi (*trp* operon); some Firmicutes (*aro* genes, *pheA, pah*)
Branched-chain amino acids	Bacillales, Clostridiales, *Syntrophomonas wolfei*, δ-proteobacteria (*leu*), Dictyoglomi, *T. roseum*
Methionine	Lactobacillales (except *Streptococcus* spp.), *Desulfotomaculum reducens*
Proline	Bacillales, *Desulfitobacterium hafniense, D. reducens*
Cysteine	Bacillales, *Enterococcus faecalis, Clostridium acetobutylicum*, Dictyoglomi
Histidine	Some Lactobacillales
Arginine	*Clostridium difficile*
Threonine	*B. cereus, C. difficile*
Serine	Some Firmicutes
Alanine	—
Asparagine and aspartate	Some Firmicutes
Glutamine	*Clostridium perfringes*
Glycine	—
Lysine	—

The T-box specificity and phylogenetic distribution of T-boxes are shown in the first and the second column, respectively.

Escherichia coli (*metS2* does not show significant similarity to *metS* from *B. subtilis*). However, in *C. beijerinckii*, this gene is preceded by a LYS-T-box, which is also most similar to other LYS-T-boxes. This suggests that *metS2* may be a lysyl-tRNA synthetase (see Discussion).

Amino acid metabolism

T-box regulation of amino acid biosynthetic operons is rather uneven (Table 1). The *trp* and *his* operons and the *asnA* genes are regulated by T-boxes in most families of the Firmicutes, but not in all representatives of these families (*trp* being more, *his* and *asnA* less frequent). Branched-chain and cysteine biosynthetic operons are regulated by T-boxes in most Bacillales and Clostridiales. MET-T-boxes regulating methionine biosynthesis genes are restricted to the Lactobacillaceae. T-box regulation of serine and proline biosynthesis genes is restricted to the Firmicutes and rather rare. At that, PRO-T-boxes upstream of the biosynthetic *pro* genes clearly form a branch distinct from prolyl-tRNA synthetase PRO-T-boxes.

In some cases, T-box regulation of biosynthetic genes is restricted to one or two genomes. Thus, the *arg* genes are regulated by ARG-T-boxes only in *Clostridium difficile*, whereas the *hom* and *thrCB* genes involved in threonine biosynthesis are regulated by THR-T-boxes in *C. difficile* and *Bacillus cereus* (see the section on the evolution of T-box regulation).

T-boxes also regulate the *aro* genes (common pathway of aromatic amino acid biosynthesis), *pah* (phenylalanine-to-tyrosine conversion), *pheA* (phenylalanine/tyrosine biosynthesis), *ubiG*, *yrhB*, *yrhA* (reverse synthesis of cysteine from methionine), and phenylacetate-coenzyme A ligase (phenylalanine catabolism, in the incomplete genome of *Desulfotomaculum reducens*).

Outside the Firmicutes, T-box regulation of biosynthetic operons is rare: the *leu* operons in δ-proteobacteria, the *cysE* gene and the *leu* operon in *Dictyoglomus thermophilum*, the *leu* operon in *T. roseum* (Chloroflexi), and the *trp* operon in the Chloroflexi and *D. thermophilum*. Notably, operons regulated by T-boxes in these taxonomy groups are also most frequently regulated by T-boxes in the Firmicutes. It looks like there is a tendency of certain amino acid biosynthesis pathways to be regulated by T-boxes.

In several cases, T-boxes were observed upstream of genes whose involvement in amino acid metabolism is not straightforward (Table 3). The *mdh* gene (putative metal-dependent hydrolase) or the *met* operon including this gene is

TABLE 2. Regulation of amino acid transporters by T-box antitermination in bacteria

Gene	T-box specificity	Predicted function	Bacteria	P/C
ycbK (COG0697)	TRP	Putative efflux protein (Valbuzzi and Yanofsky 2001)	*Bacillus subtilis, Bacillus licheniformis*	
yhaG	TRP	Tryptophan transporter (Gutierrez-Preciado et al. 2005)	Clostridiales	
2.A.3 The amino acid-polyamine-organocation (APC) superfamily				
yvbW (COG1113)	LEU	Leucine-specific permease	*B. subtilis, B. licheniformis*	P
ykbA (*steT*) (COG0531)	THR	Threonine-specific permease (Reig et al. 2007)	*Bacillus cereus*	P
2.A.1 MFS superfamily (the major facilitator superfamily)				
lysX* (COG0477)	LYS	Lysine transporter	Some Bacillales	P
argX* (COG1457)	ARG	Arginine permease	*Clostridium difficile* (*CD1716*)	P
leuX* (COG1139)	LEU	Leucine transporter	*Desulfotomaculum reducens* (*Dred_0433*)	P
2.A.23 DAACS transporter family (the dicarboxylate/amino acid:cation [Na^+ or H^+] symporters)				
cysP1* (COG1823)	CYS	Cysteine transporter	*Clostridium perfringens* (*CPE0947, CPE0967*), *Clostridium botulinum*	P
cysP2* (COG1301)				
2.A.35 The NhaC Na^+:H^+ antiporter (NhaC) family				
yheL (COG1757)	TYR	Tyrosine transporter	Some Bacillales and Lactobacillales	P
2.A.26 The branched-chain amino acid:cation symporter family				
brnQ/braB (COG1114) (Tauch et al. 1998)	ILE	Branched-chain amino acid carrier protein	Some Bacillales, Lactobacillales, and Clostridiales	P
	THR		*B. cereus, Clostridium tetani*	P
	VAL		Some Lactobacillales	P
3.A.1 The ATP-binding cassette (ABC) superfamily				
metNPQ (COG1135, COG2011, COG1464)	MET	Methionine ABC transporter (Zhang et al. 2003; Hullo et al. 2004)	Lactobacillales, *Enterococcus faecalis*	
yqiXYZ (COG0834, COG765, COG1126)	ARG	Arginine ABC transporter	*C. difficile* (*CD0750, 51, 52*)	P
hisXYZ* (COG0765, COG1126, COG0834)	HIS	Histidine ABC transporter	Lactobacillales, *C. difficile* (*CD1774, 75, 76*), *L. monocytogenes* (*lmo1740,39,38*), *E. faecalis*	P
yckKJI (COG0834, COG765, COG1126)	CYS	Cysteine ABC transporter (Burguiere et al. 2004)	*Clostridium acetobutylicum* (*CAC3325, 26, 27*), Lactobacillales	P
	MET			
ytmKLMN (COG0834, COG0765, COG1464)	MET	Cysteine ABC transporter (Burguiere et al. 2004)	*Leuconostoc mesenteroides* (*LEUM_0136, 37, 38, 39*)	P
aspQHMP (COG0834, COG1126, COG0765)	ASP	ASP\ASN ABC transporter	*Lactobacillus johnsonii* (*LJ0752, 53, 54, 55*)	P
trpXYZ* (COG2984, COG0559, COG1101)	TRP	Tryptophan ABC transporter	Peptococcaceae, *Streptococcus spp.* (e.g., *SPD_0954, 55, 56*), *P. larvae*	P
mtsABC (–, COG1122, COG0619)	MET	Uptake of MET/or MET-precursors	Some Lactobacillales (e.g., *LEUM_1974, 73, 72*)	P, C
opp family (COG0747)	TRP	Transporters of oligopeptides	Some Lactobacillales (e.g., *lp_0092*),	P
	MET		*E. faecalis* (*EF3081*)	P
	ILE		*L. johnsonii* (*LJ1574*)	P
	LEU?		*Lactobacillus brevis* (*LVIS_0014*)	—
	LYS?		*Lactobacillus plantarum* (*lp_3686*)	—
	?		*Lactobacillus brevis LVIS_0024, 2175*	?
2.A.22 The neurotransmitter:sodium symporter (NSS) family				
yocR/yhdH(COG0733)	TRP	TRP sodium-dependent transporter	*B. cereus* (*BC1430*)	P
	PHE	PHE sodium-dependent transporter	*B. cereus* (*BC1767*)	P
	LEU	LEU sodium-dependent transporter	*B. cereus* (*BC2170*)	P
	Met?	Sodium-dependent transporter	*Clostridium tetani* (*CTC01948*)	—
2.A.3 The amino acid-polyamine-organocation (APC) superfamily				
ydgF\aapA (COG1113)	?	?	*Lactobacillus reuteri*	?

All predicted transporters were divided into groups according to the classification from Saier et al. (2006). Newly identified candidate amino acid transporters are set in bold (genes named in this paper are marked by asterisks; COGs are shown in parentheses). The predicted specificities of T-boxes are indicated in the second column. The fifth column: (P) T-boxes regulating transporter genes that were shown to coevolve with T-boxes regulating corresponding biosynthetic/aminoacyl-tRNA synthetase genes (data not shown). (C) Genes were shown by positional analysis to be colocalized (co-regulated) with corresponding biosynthetic/aminoacyl-tRNA synthetase genes (Supplemental Table S1).

TABLE 3. Regulation of enzymes by T-box antitermination in bacteria

Gene	T-box specificity	Possible function	Bacteria	P/C
Putative aminotransferases (2.6.1.-)				
alaRT (*yugGH*)(*COG1522, COG0436*)	ALA	Putative alanine transaminase(alaT) and transcriptional regulator alaR	Some Clostridiales, Peptococcaceae, *Moorella thermoacetica*	
yurG (COG0075)	SER	Putative phosphoserine aminotransferase	Some Clostridiales, Peptococcaceae, *M. thermoacetica*,	C
lp_2751 (COG1168)	MET	Putative methionine aminotransferase	*Lactobacillus plantarum*	P
OB1271–OB1272 (COG0454, COG1670)	LEU	Putative acetyltransferases	*Oceanobacillus iheyensis*	P
Hypothetical proteins				
yxjI *(Bacillus subtilis)*	ALA	Conserved hypothetical protein	Streptococcaecae	C
lmo2587 (COG9112)	TRP	Conserved hypothetical protein	*Listeria monocytogenes*	P
068_0017	TYR	Conserved hypothetical protein	*Ruminococcus albus*	P
079_0081	GLU	Conserved hypothetical protein	*Clostridium beijerinckii* (*cbei_2130*)	P
Other genes				
lp_3666 (COG0179)	ILE	2-Hydroxypenta-2,4-dienoate hydratase	Some Lactobacillales [e.g., *L. plantarum* (*lp_3666*)]	P
Cthe_0234 (COG0318)	LEU	AMP-dependent synthetase and ligase	*Clostridium thermocellum*	P
panE1 (COG1893)	MET?	2-Dehydropantoate	*L. reuteri, L. plantarum* (*panE1*)	C
panE2 (COG1893)	ILE	2-reductase	*L. reuteri, L. plantarum* (*panE2*)	P
Swol_2237 (COG1014)	LEU	Putative oxidoreductase	*Syntrophomonas wolfei*	P
yngI (COG0318)	THR?	Fatty-acid-CoA ligase\acyl-CoA synthase	*Heliobacillus mobilis*	—
yngI (COG1541)	PHE	Phenylacetate-coenzyme A ligase	*D. reducens* (*Dred_2884*)	P
LEUM_0134 LEUM_0135	MET	Putative uroporphyrinogen-III decarboxylase	*Leuconostoc mesenteroides*	P
mdh (COG1878)	MET	Metal-dependent hydrolase	*S. aureus* (*SAV0355*), *L. mesenteroides (LEUM_0115)*	P,C
114_0008 (COG1619)	LEU	Microcin C7 resistance protein	*Ruminococcus albus*	C
114_0009		Conserved hypothetical protein		

Newly identified regulon members are set in bold; COGs are shown in parentheses. The fifth column is as in Table 2.

preceded by MET-T-boxes in *Leuconostoc mesenteroides* and *Staphylococcus aureus*, respectively. Notably, in the Streptococcaceae, *mdh* also has been predicted to be co-regulated with methionine biosynthesis genes by a transcription factor (Rodionov et al. 2004). The role of the *mdh* gene product in the methionine metabolism is not clear. Even less clear cases are *SMU.932* (putative uroporphyrinogen-III decarboxylase) preceded by a MET-T-box in *L. mesenteroides*; *lp_3666* (annotated as 2-oxo-hept-3-ene-1,7-dioate hydratase/2-oxo-hept-4-ene-1,7-dioate hydratase) preceded by a ILE-T-box in some Lactobacillaceae (*Lactobacillus reuteri* [two regulated paralogs]; *Lactobacillus plantarum*, *Oenococcus oeni*); the pantothenate biosynthesis gene *panE* encoding 2-dehydropantoate 2-reductase preceded by a ILE-T-box in *L. reuteri* and *L. plantarum*; a Phe-T-box-regulated phenylacetate-coenzymeA-ligase in the unfinished genome of *D. reducens*; and the *yngI* gene annotated as fatty-acid-CoA ligase/acyl-CoA synthase and preceded by a T-box with unclear specificity (possibly a threonine one) in *Heliobacillus mobilis*. Finally, the incomplete genome of *Ruminococcus albus* has a LEU-T-box upstream of a candidate three-gene operon encoding a microcin C7 resistance protein, a conserved hypothetical protein and 2-isopropylmalate synthase (*leuA*, leucine biosynthesis) (Table 3).

Genes encoding enzymes from common pathways may be regulated by different T-boxes in different genomes. For instance, the *aro* genes encoding enzymes from the common part of the aromatic amino acid synthesis pathway are regulated by TYR-T-boxes in *B. cereus*, by PHE-T-boxes in the Peptococcaceae, *H. mobilis*, *Moorella thermoacetica*, *Clostridium thermocellum*, and by TRP-T-boxes in the Chlorophlexales.

Similarly, in the Bacillales, the branched-chain amino acid operon *ilv* may be preceded by a LEU-T-box (in most

Bacillus species: *B. subtilis*, *Bacillus licheniformis*, *Bacillus halodurans*, *Bacillus stearothermophilus*, *Bacillus clausii*) or by a ILE-T-box (*B. cereus*, *Oceanobacillus iheyensis*). In the Clostridiales, the *ilv* genes are regulated mainly by LEU-T-boxes, although the separate gene *ilvC* in *Clostridium acetobutylicum* is regulated by a VAL-T-box. The most interesting case was observed in *D. reducens*, where the *ilv* operon is preceded by a tandem formed by recently duplicated ILE- and LEU-T-boxes; the ILE-T-box is the original one (see the section on the evolution of T-box regulation).

Aminotransferases (EC 2.6.1.-) seem to be a functional group of enzymes most often regulated by T-boxes. MET-T-boxes were found upstream of the operon *ykrUV-panE* from *L. reuteri*; in *B. subtilis*, the orthologous aminotransferase YkrV(MtnE) catalyzes the final step of a long methionine salvage pathway (recycling of 5′-methylthioadenosine to methionine) (Sekowska and Danchin 2002). Another, nonhomologous putative aminotransferase *lp_2751* is preceded by a MET-T-box in *L. plantarum*. Putative phosphoserine aminotransferase *yurG*, nonhomologous to the classical gene *serC*, is colocalized with *serA* and regulated by a SER-T-box in the Clostridiales, *Syntrophomonas wolfei*, and *M. thermoacetica*. Putative alanine transaminase *yugH/alaT* forms a candidate operon with a transcriptional regulator *alaR* (the Subtilist database, http://genolist.pasteur.fr/SubtiList/, A.L. Sonenshein, pers. com.) and in addition is preceded by an ALA-T-box in the Clostridiales, the Peptococcaceae, and *M. thermoacetica*. Finally, in *O. iheyensis*, a LYS-T-box is found upstream of the *OB1271-OB1272* genes encoding possible acetyltransferases from the GNAT family.

Transporters

Identification of a T-box with a recognizable specifier codon upstream of a candidate transporter gene is a strong indicator to the transporter specificity toward an amino acid or maybe its precursor. We observed a large number of known and new amino acid transporters likely regulated by T-boxes (Table 2). They represent several superfamilies of both secondary transporters and ABC transporters and are involved in transport of three branched-chain amino acids, threonine, lysine, histidine, arginine, cysteine, methionine, three aromatic amino acids, aspartate, or their precursors.

In several cases, closely related, likely orthologous transporters are regulated by different T-boxes in different genomes. This may represent either changes in transporter specificity or, more likely, nonselective transport of several related amino acids. The examples of this type are the transporters of branched-chain amino acids *brnA/braB* regulated by ILE-T-boxes, THR-T-boxes, or VAL-T-boxes in various Firmicutes and the *yocR/yhdH* transporters from the NSS superfamily preceded by TRP-, PHE-, and LEU-T-boxes in *B. cereus* and possible MET-T-boxes in *Clostridium tetani*. Probably the most diverse regulation was observed for the *opp* ABC-transporters involved in oligopeptide transport (TC 3.A.1.5._): representatives of this family are regulated by TRP-, MET-, ILE-, and also possibly LEU- and LYS-T-boxes, mainly in the *Lactobacillus* spp. The latter observation is not surprising, given that these bacteria live in protein-rich media.

ABC-type transporters *yckKJI* and *ytmKLMN* have been shown experimentally to mediate L-cysteine uptake in *B. subtilis* (Burguiere et al. 2004). Interestingly, in *B. subtilis*, *yckKJI* is regulated by a CYS-T-box, whereas both *yckKJI* and *ytmKLMN* are regulated by MET-T-boxes in *C. acetobutylicum* and some Lactobacillales, respectively. Cysteine may be synthesized via cysteine recycling, whereas methionine is synthesized from homoserine and cysteine. Thus, candidate cysteine transporters seem to be regulated by two different amino acids, cysteine and methionine, which are close on the metabolic map, although one cannot definitely rule out changes in the transporter specificity.

Evolution of T-box regulation

Regulatory mechanisms

In most cases, T-boxes regulate premature termination of transcription (Fig. 1A). The T-box-containing hairpin, stabilized by interaction with an uncharged tRNA, is an antiterminator, alternative to an intrinsic transcriptional terminator of transcription (a hairpin followed by a run of uridines). Transcriptional terminators alternative to T-box hairpins were observed in most T-boxes of the Firmicutes as well as δ-proteobacteria, Deinococcales/Thermales, and Chloroflexi, Dictyoglomi. However, some T-boxes have a different mechanism of the terminator formation. The terminator of the *S. aureus ileS* T-box has been shown to contain an additional region capable of base-pairing that extends the antiterminator structure (Fig. 1B; Grundy et al. 1997a). In this case, T-box-tRNA interaction would be expected to prevent the base-pairing extending the antiterminator rather than promoting formation of an alternative structure. Such unusual T-boxes were observed in some bacteria: Streptococcaecae, Lactobacillaceae, Leuconostocaceae (ALA-T-box, *alaS*), Clostridiaceae (GLY-T-box, *glyS*), Streptococcaecae (SER-T-box, *serS*), and *S. aureus* (ILE-T-box, *ileS*) (data not shown).

However, this is not the only possibility, as the T-box-containing hairpin may be alternative to a hairpin that masks the Shine–Dalgarno box and thus interferes with the initiation of translation (Fig. 1C, sequester hairpin; Seliverstov et al. 2005). Such T-boxes were found upstream of the *ileS* genes in some Actinobacteria (all Actinomycetales and *Bifidobacterium longum*). In the latter case, the sequester hairpin is located ~100 nucleotides (nt) downstream from the T-box hairpin, and a complicated

rearrangement of alternative hairpins is predicted to lead to sequestering of the Shine–Dalgarno box (data not shown).

Notably, the translation-regulating T-boxes of the classical type were observed only in *Thermobifida fusca* and the *Streptomyces* spp. In other actinobacterial species, a new variant of the specifier hairpin was observed (Fig. 1D). It is much shorter and the specifier codon is located in the loop of the hairpin, and not in the bulge, as in the classic T-box structure. The predicted structure is supported by several cases of compensatory substitutions retaining base-pairing (Fig. 2).

This is consistent with the observation that in Actinobacteria (Actinomycetales), other RNA regulatory elements, such as riboswitches, tend to regulate initiation of translation and sometimes represent reduced versions of the riboswitches found in other species (Vitreschak et al. 2004). On the other hand, in some other actinobacteria, *Rubrobacter xylanophilus* (Rubrobacterales, regulation of *ileS*), *A. minutum* (Coriobacteriales, regulation of *pheST*), the transcription antitermination mechanism was predicted.

Double and partially double T-boxes

In some cases, T-boxes are arranged in tandem (Table 4). Predominantly this happens upstream of biosynthetic and transport genes, as it has been previously described for the *trp* operon (Panina et al. 2003; Gutierrez-Preciado et al. 2005). The two major types of such tandems are double T-boxes, that is, repeats of complete T-box structures (Fig. 3A), and partially double T-boxes, formed by a single specifier hairpin followed by two repeated terminator/antiterminator structures closely located to each other (Fig. 3B).

In most cases, formation of double T-boxes is due to recent duplications, so that the T-boxes forming such pairs are the closest relatives and have the same specificity (Table 4). An interesting exception is the tandem of ILE- and LEU-T-boxes upstream of the *ilv* operon in *D. reducens*; the enzymes encoded by this operon form the common biosynthetic pathway of isoleucine, leucine, and valine. The phylogenetic analysis demonstrates that the original T-box had ILE specificity (see below about changes of T-box specificity). In most cases, the tandem T-box duplications

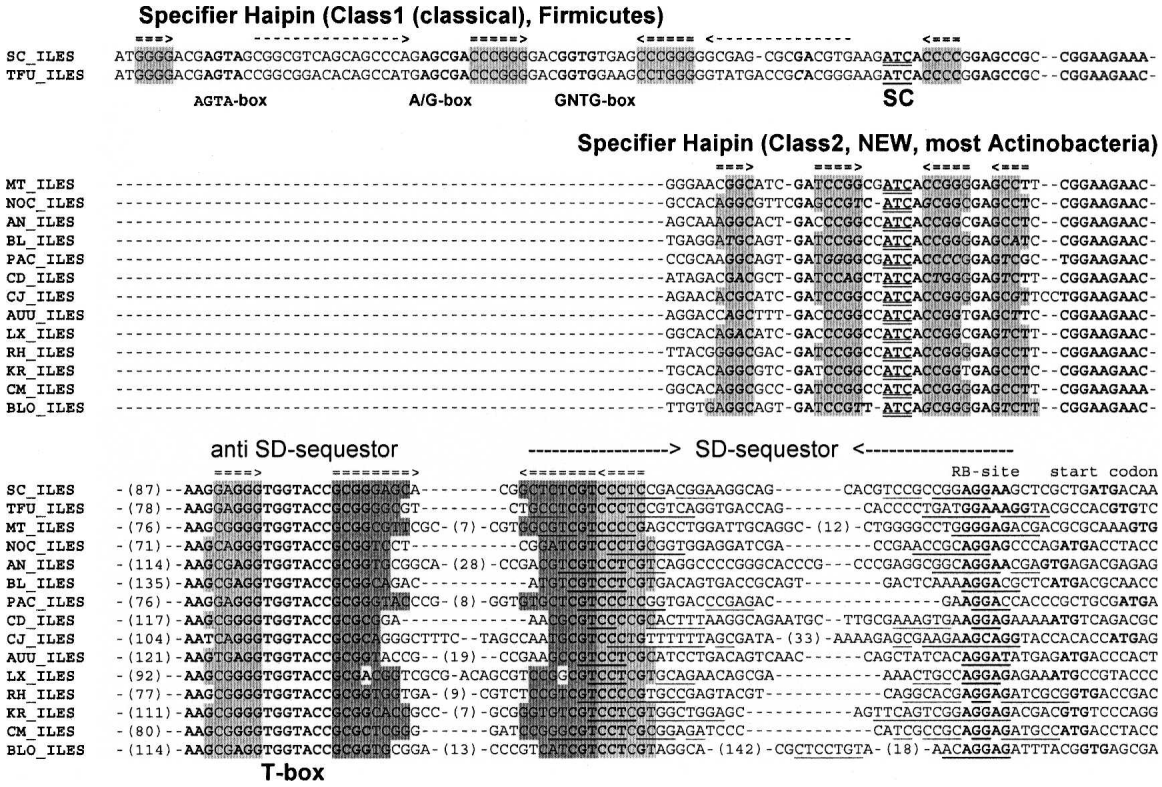

FIGURE 2. Alignment of the T-boxes regulating initiation of translation of the *ileS* gene in the Actinobacteria. (*Top*) Two different versions of the specifier hairpin region (the classical one as in Fig. 1C and the reduced one as in Fig. 1D). (*Bottom*) The sequestor/antisequestor region. Specifier codons (SC) are double-underlined. The complementary stems of the mRNA structure are shown by arrows *above* the alignment. (Gray) Based-paired positions; (underline) the SD-sequester hairpin; (bold) conserved positions; (bold italic) non-consensus nucleotides; (light) non-conserved positions. The lengths of variable regions are given in brackets. (SC) *Streptomyces* spp.; (TFU) *Thermobifida fusca*; (MT) *Mycobacterium* spp.; (NOC) *Nocardioides* sp.; (AN) *Actinomyces naeslundii*; (BL) *Brevibacterium linens*; (PAC) *Propionibacterium acnes*; (CD, CJ) *Corynebacterium* spp.; (AAU) *Arthrobacter aurescens*; (LX) *Leifsonia xyli*; (RH) *Rhodococcus* str.; (KR) *Kineococcus radiotolerans*; (CM) *Clavibacter michiganensis*; (BLO) *Bifidobacterium longum*.

TABLE 4. The distribution of tandem T-boxes in bacteria

Regulated genes	T-box specificity	Gene function	Species
Double T-boxes			
trp operon	B	TRP-TRP	Bacillales, *C. beijerinckii*, *D. hafniense*
pah	B	TYR-TYR	*Bacillus cereus*
ilvEBCDA	B	ILE-ILE	*B. cereus*
aroA	B	TYR-TYR	*B. cereus*
hom	B	THR-THR	*Clostridium difficile*
leuA	B	LEU-LEU	*Clostridium thermocellum*
ilvDBNCB-leuACDBA	B	ILE-LEU	*Desulfotomaculum reducens*
hisXYZ	T	HIS-HIS	*Enterococcus faecalis*
thrZ	A	THR-THR	Bacillales
Partially double T-boxes (two terminator–antiterminator structures)			
trp operon	B	TRP	*Thermoanaerobacter tengcongensis*
aroIA-pheA	B	PHE	*Desulfotomaculum reducens*, *Syntrophomonas wolfei*
serCA	B	SER	*Desulfitobacterium hafniense*
hisXYZ	T	HIS	*Lactobacillales*
trpXY2	T	PHE	*D. reducens*
yheL	T	TYR	*B. cereus*
yqiXYZ	T	ARG	*C. difficile*
brnQ/braB2	T	THR	*Clostridium tetani*
yngI	U	PHE	*D. reducens*
Partially double T-boxes (three terminator–antiterminator structures)			
yocR/yhdH2	T	PHE	*B. cereus*

The regulated genes, the T-box specificity, gene function, and species are shown in the first, second, and third columns, respectively. Second column: (B) biosynthetic genes; (T) transporters; (A) aminoacyl-tRNA synthetases; (U) unknown.

are restricted to single lineages; the exceptions are the TRP-T-boxes upstream of the *trp* operon in some Bacillales and also *C. beijerinckii* and *Desulfitobacterium hafniense* (cf. Gutierrez-Preciado et al. 2005, 2007) and THR-T-boxes upstream of the threonyl-tRNA synthetase gene *thrZ*, also in several Bacillales.

Again, such arrangements are similar to the ones observed in the analysis of riboswitches. The 5′ regulatory region of the *B. clausii metE* mRNA is formed by the S-box and B12 riboswitches (Sudarsan et al. 2006). Most glycine riboswitches possess two ligand-binding RNA domains that function cooperatively (Mandal et al. 2004). Recently, tandemly arranged THI (thiamin pyrophosphate) riboswitches were studied in detail (Sudarsan et al. 2006). Based on analogy with the riboswitches, one could expect that tandem arrangement of T-boxes would lead to a sharper (digital-like) response to uncharged tRNA concentrations. However, the glycine and THI riboswitches demonstrate different behavior: the ligand binding to the former is cooperative (Mandal et al. 2004), whereas the latter structures are bound independently (Welz and Breaker 2007). Thus the exact molecular mechanisms involved in the regulation by tandem T-boxes of the same or different specificity remain to be elucidated by experiment and by kinetic modeling of RNA structures.

The role of the partially double T-boxes is much less clear. A straightforward modeling of RNA secondary structures leads to an absurd conclusion that such T-boxes will always form at least one terminator hairpin, and thus the genes would never be expressed. A trivial explanation that such T-boxes are not functional is refuted by the observation that they are conserved. In particular, the candidate histidine ABC-transporter *hisXYZ* is regulated by partially double T-boxes in the Lactobacillaceae/Leuconostocaceae (*L. plantarum*, *Lactobacillus sakei*, *L. reuteri*, *L. mesenteroides*). Multiple alignment of the antiterminator/terminator regions (two for each genome) shows strong conservation of the RNA structures with some compensatory nucleotide substitutions (Fig. 4). In *B. cereus*, the *yocR-yhdH2* operon is preceded by an even more complicated structure: the specifier hairpin followed by three adjacent antiterminator/terminator sites.

Whatever is their function, partially double T-boxes seem to evolve by duplication followed by deletion of the downstream specifier hairpin. A convincing example is provided by the ARG-T-box upstream of the candidate arginine ABC-transporter *yqiXYZ* from *C. difficile*. This T-box is a fairly recent result of a rapid genome-specific expansion (see below). The antiterminator/terminator sites (~60 nt) of this T-box are highly similar to each other (~90% identity).

Evolution of T-boxes

In most cases, T-boxes coevolve with the regulated genes by vertical descent. In such cases, the branches on the phylogenetic tree correspond well to the regulated operons, and the topology of the branches is consistent with the taxonomy. Such relationships may survive minor operon rearrangements: for example, *yxjI* was inserted upstream of the *alaS* gene in seven *Streptococcus* spp., forming the operon *yxjI-alaS* preceded by the ALA-T-box instead of the ancestral ALA-T-box–*alaS* locus (Fig. 5A). Similarly, genes were inserted between the VAL-T-box and the *valS* gene in most *Streptococcus* spp. T-boxes may be lost, for example, ASP-T-boxes upstream of the *hisS-aspS* in the Clostridiales; most CYS-T-boxes and MET-T-boxes upstream of the *metS* gene in the Lactobacillales; LEU-, ARG-, ASP-, and TYR-T-boxes upstream of the respective aminoacyl-tRNA synthetase genes in the *Streptococcaeae*, etc. (Supplemental Table S1).

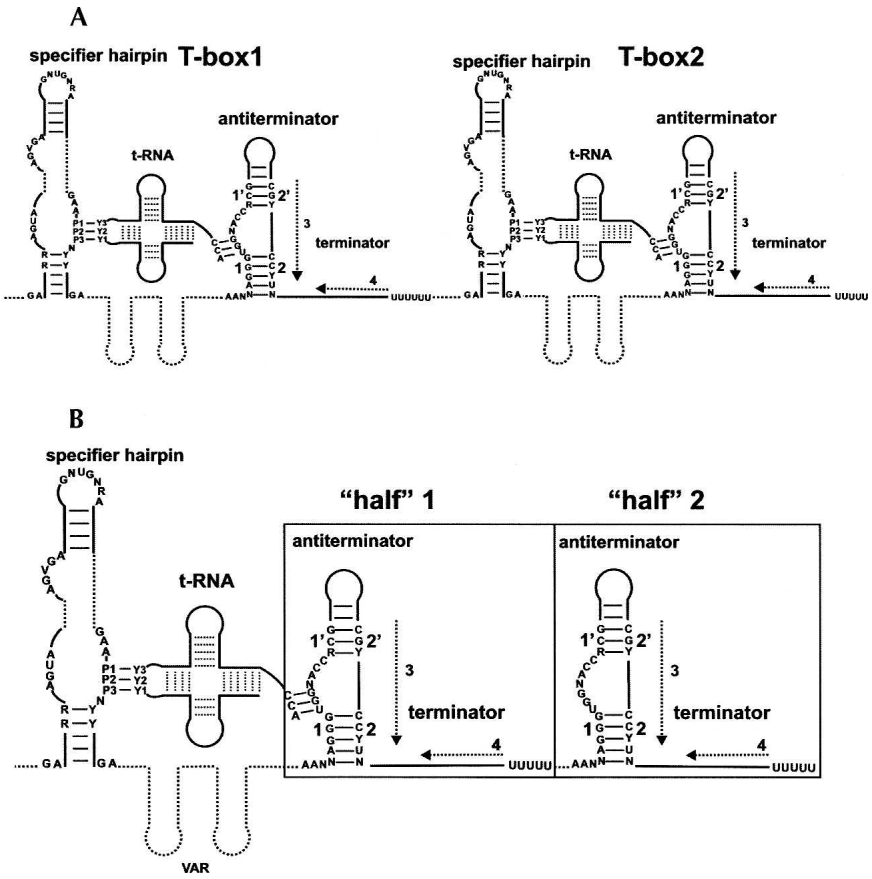

FIGURE 3. Two structurally different types of tandem T-boxes. (*A*) Double T-boxes (tandemly arranged complete T-boxes) and (*B*) partially double T-boxes (a specifier hairpin followed by two adjacent antiterminator/terminator structures).

Duplications and expansion of T-box regulons

T-boxes are often duplicated. In simple cases, close duplicates of aminoacyl-tRNA synthetase T-boxes are observed upstream of biosynthetic or transporter genes. For instance, in *C. perfringens* and *C. beijerinckii*, the ALA-T-box upstream of the *yugGH* (*alaRT*) operon encoding a candidate alanine transaminase and a transcriptional regulator is likely the result of duplication of the ALA-T-box regulating the *alaS* gene in the common ancestor of these species (Fig. 5A), whereas ALA-T-boxes upstream of the *alaRT* operon in other species clearly form a distinct branch (Fig. 5A). It is possible that in this case the duplication caused in situ substitution of the original ALA-T-box (cf. a similar phenomenon on the gene level in the operon dynamics) (Omelchenko et al. 2003).

Sometimes multiple duplications of T-boxes lead to rapid expansion of T-box regulons. Such expansions were observed in the genome of *B. cereus* and, independently, *C. difficile*. The duplication of the THR-T-box originally upstream of the *thrS* gene has led to emergence of new THR-T-boxes upstream of the transporter genes *brnQ*, *ykbA* in the former and the biosynthesis genes *hom*, *thrCB* in the latter (Fig. 5B). A specific case of T-box duplication is a duplication leading to double or partially double T-boxes. In most cases, the tandem T-boxes are closest relatives of each other (e.g., the tandem THR-T-boxes upstream of the *hom* gene in *C. difficile* or THR-T-boxes upstream of *thrZ* in *B. cereus*) (Fig. 5B).

In another case, we detected the recent expansion of T-box regulon in *C. difficile* that had occurred after the separation of the Clostridia into individual lineages (Fig. 5C). Multiple duplications of ARG-T-boxes following the loss of the arginine transcription regulator AhrC in *C. difficile* created five ARG-T-boxes upstream of three biosynthetic and two transporter operons; at that, the original ARG-T-box upstream of the *argS* gene was lost (Fig. 5D). Similarly, the loss of the S-box riboswitch regulating methionine biosynthesis in the Lactobacillales has lead to the expansion of MET-T-boxes in the Lactobacillaceae lineage (whereas in the *Streptococcus* spp. these genes are regulated by a new transcription factor, MtaR/MetR) (Rodionov et al. 2004; Kovaleva and Gelfand 2007; Sperandio et al. 2007).

In some cases, it does not make sense to say which of two duplicated T-boxes is the original one. In several cases, this seems to be caused by disruption of the original operon, where both parts are preceded by descendants of the original T-box. This may be the case for PRO-T-boxes upstream of the *proBA* and *proI* operons in *B. stearothermophilus*, as well as *B. subtilis* and *B. licheniformis*. These operon pairs are the closest relatives in the phylogenetic tree in the *B. stearothermophilus* branch and in the node corresponding to the latest common ancestor of *B. subtilis* and *B. licheniformis* (Fig. 5E). Other genomes have a PRO-T-box upstream of the complete biosynthetic *proIBA* operon, and this is a likely ancestral state. Duplication of this operon together with the preceding T-box and subsequent loss of the *proI* gene in one copy and *proBA* in the other copy would lead to the present state in these three genomes. An alternative scenario is simple operon disruption leading to separation of the *proBA* and *proI* genes and the loss of regulation by the latter (as was likely the case for *B. cereus*). The evolutionary signal does not seem to be sufficient to decide which scenario holds for *O. iheyensis*.

A more complicated example is provided by the *hisS-aspS* operon. In Lactobacillales, it is regulated by ASP-T-boxes,

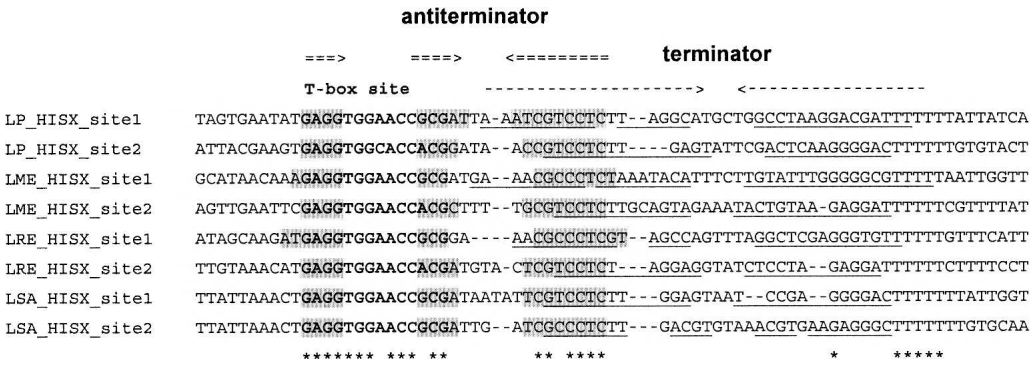

FIGURE 4. Alignment of antiterminator/terminator regulatory regions of the *hisXYZ* partially double T-boxes in the Lactobacillales. (LP) *Lactobacillus plantarum*; (LSA) *Lactobacillus sakei*; (LRE) *Lactobacillus reuteri*; (LME) *Leuconostoc mesenteroides*. (Arrows in the *upper* lines) The complementary stems of alternative antiterminator and terminator structures; (gray background) base-paired positions; (*) conserved positions; (bold) the conserved T-box site.

whereas in *L. reuteri*, there are two operons, *hisS* and *aspS*, regulated by the HIS- and ASP-T-boxes, respectively; the former is a copy of the latter with a changed specificity (see below for further discussion). Again, the most plausible scenario seems to be operon duplication with subsequent loss of *hisS* in one copy and *aspS* in the other copy.

Changes of specificity

Both the vertical mode of evolution and the duplications may be accompanied by changes in the T-box specificity. Here we consider only clear cases, when T-boxes with different specificity are closest relatives of each other. Such changes occur in the subgroups where the amino acids synthesized by adjacent metabolic pathways also have similar codons: the ASP/ASN/HIS group and the branched-chain amino acids group. In these cases, changes are caused by single nucleotide substitutions in the specifier codons. Interestingly, in all cases these substitutions occurred in the first codon positions.

In the Bacillales, the ASP ↔ ASN specificity changes in the T-box upstream of the *hisS-aspS* operon are caused by the **G**AC↔**A**AC transitions (Fig. 6B). Based on the analysis of other groups and gene functionality, the ASP specificity seems to be the ancestral one, but such changes seem to have occurred more than once and in both directions. In *Lactobacillus johnsonii*, the duplicated T-box upstream of an ABC-transporter annotated as a glutamine transporter is an ASP-T-box, although its likely ancestor is the ASN-T-box upstream of the *asnA* gene, observed also in other Lactobacillales (Fig. 6A). This also means that the transporter is more likely to be an asparagine (or aspartate) transporter than a glutamine one, and should be renamed *aspQHMP*. In *Pediococcus pentosaceus*, a duplicated copy of the ASP-T-box upstream of the annotated *asnS2* gene was inserted upstream of a candidate histidine ABC-transporter *hisXYZ* and changed specificity by the substitution **A**AC → **C**AC, becoming a HIS-T-box gene (Fig. 6A). In the already mentioned case, in *L. reuteri*, disruption of the ASP-T-box-regulated *hisS-aspS* operon has generated two operons; the *hisS* operon is now regulated by a HIS-T-box with the **G**AC → **C**AC change of the specifier codon (Fig. 6A).

Another group of T-boxes with obvious specificity changes is the ILE-T-boxes. There, the *brnQ* subbranch consists of three ILE-T-boxes (*Lactobacillus casei*, *L. johnsonii*, *L. reuteri*) and two VAL-T-boxes (*L. casei*, *L. plantarum*) (Fig. 6C). The phylogenetic analysis of the protein sequences of the branched-chain amino acid transporter BrnQ shows that genes regulated by ILE-T-boxes form two independent branches. Two VAL-T-box-regulated genes are immediately adjacent to one of these branches. In the T-box tree, all T-boxes on the same branch are ILE-T-boxes, which is likely the ancestral specificity. Thus it seems that simultaneous duplication of the *Lactobacillus brnQ* genes (creating the genes denoted *brnQ2* in Fig. 7) and the upstream T-boxes was followed by a change in the T-box specificity by the **A**TC → **G**TC (ILE → VAL) substitution in the specifier codon.

Experimentally studied BrnQ proteins have been shown to transport all three chain amino acids (leucine, isoleucine, and valine) (Stucky et al. 1995; Tauch et al. 1998). It will be interesting to see whether the substrate specificity or affinity of transporters encoded by VAL-T-box-regulated genes is different from that of ILE-T-box-regulated ones. An additional relevant observation is that two *brnQ* homologs belonging to different branches are regulated by THR-T-boxes (Fig. 7). In *B. cereus*, the THR-T-box upstream of *brnQ* results from duplication of the THR-T-box regulating the *thrS* gene. A more interesting case is the partially double THR-T-box upstream of the *brnQbraB1* operon of *C. tetani*. This T-box is closely related to ILE-T-boxes upstream of *brnQ* and some other genes in other *Clostridium* spp. (Figs. 6C 7). The most parsimonious scenario is that this T-box arose by partial duplication of the ancestral ILE-T-box accompanied by a specificity

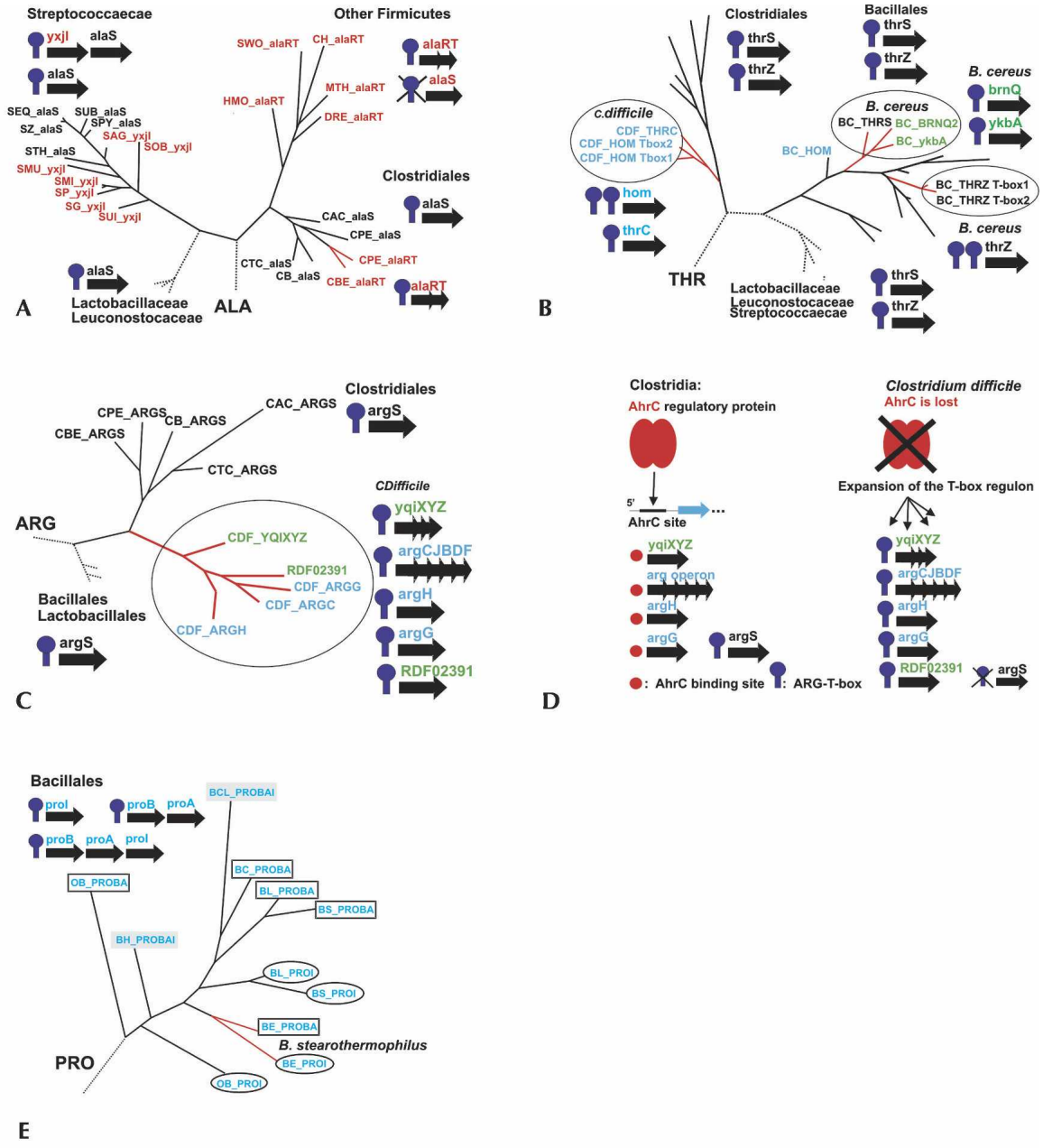

FIGURE 5. Duplications of T-boxes. (Filled rectangle with a filled circle above) Schematic RNA structure with T-boxes and (filled black arrows) regulated genes/operons. (Black) Aminoacyl-tRNA synthetase, (blue) biosynthetic, (green) transporter, and (red) other (unknown) genes. (Red or circled) The sub-branches of interest. (*A*) A fragment of the ALA-T-box phylogenetic tree demonstrating duplication in the Clostridiales. (*B*) A fragment of the THR-T-box phylogenetic tree. (Circled) The expansions of the T-box regulon in *C. difficile, B. cereus*, and the T-box duplication leading to double T-boxes (denoted "box1" and "box2"). (*C*) A fragment of the ARG-T-box phylogenetic tree. (Circled) The branch showing the rapid expansion of the ARG-T-box regulon. (*D*) The detailed picture illustrating the loss of the regulator AhrC and expansion of the ARG-T-box regulon in *C. difficile*. (*E*) A fragment of the PRO-T-box phylogenetic tree. T-boxes regulating (rectangles) *proBA*, (circles) *proA*, and (filled rectangles) *proBAI* operons. (Red) The sub-branch corresponding to the recent PRO-T-box duplication in *B. stearothermophilus*.

change. Again, it is interesting to see what is the substrate specificity of the THR-T-box-regulated transporters.

The final example of this type is the already mentioned duplicated ILE-LEU-T-box upstream of the *ilv* operon of *D. reducens*, where the LEU-T-box was created by the ATC → CTC substitution (Fig. 6C).

Variability of regulation of common biosynthetic pathways

The fact that homologous transporters may be regulated by T-boxes with different specificity is not surprising, since transporter specificity also is very flexible in evolution. A

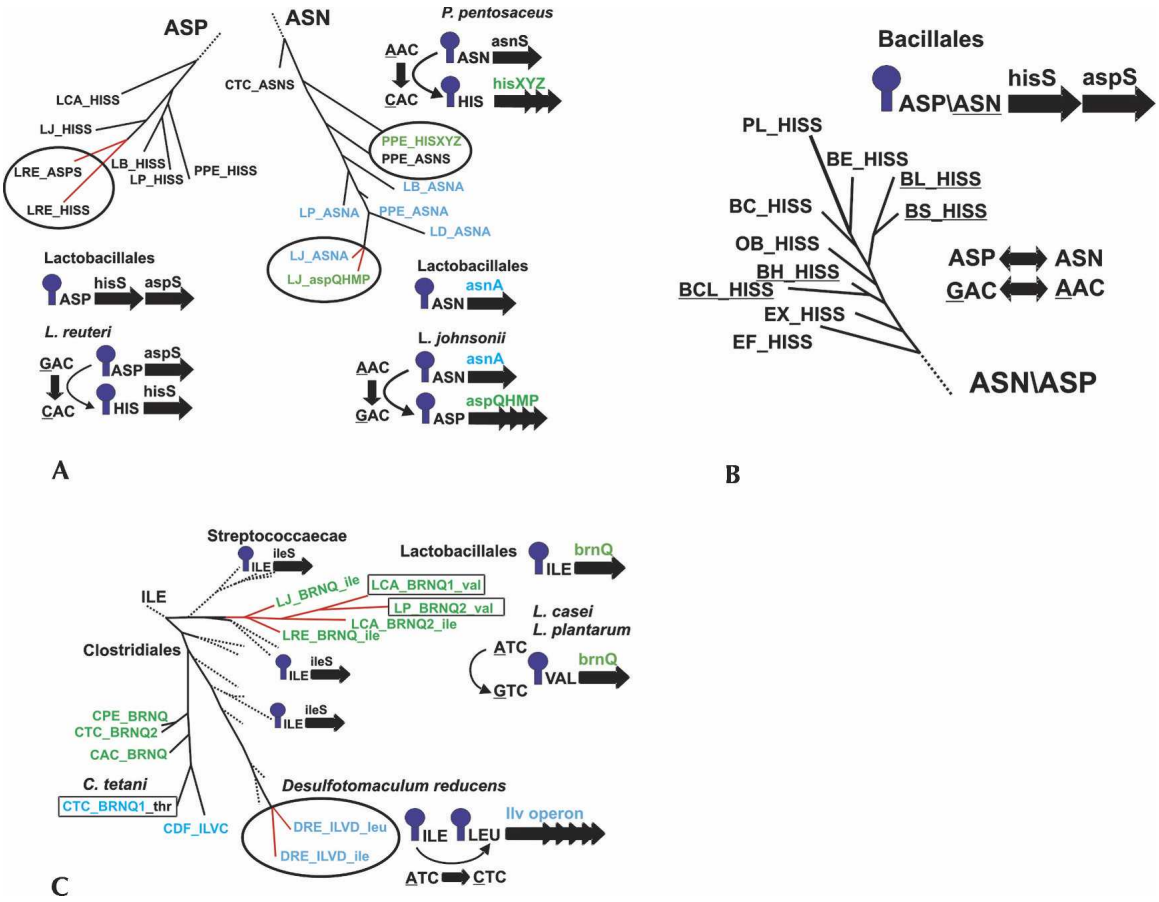

FIGURE 6. Changes of specificity of T-boxes. The notation is as in Figure 5. (*A*) Fragments of the ASP- and ASN-T-box phylogenetic trees. (Circled) Recent duplications accompanied by changes in the T-box specificity in the same genome or closest relatives. (*B*) Fragments of the ASP- and ASN-T-box phylogenetic tree. ASN-T-boxes are underlined, ASP-T-boxes are not. (*C*) A fragment of the ILE-T-box phylogenetic tree. (Red) The branch of T-boxes regulating *brnQ* genes. Recent changes of T-box specificity (ILE → VAL) in two closest *Lactobacillus* relatives (*L. casei*, *L. plantarum*). (Rectangles) Two VAL-T-boxes; (circled) the recently duplicated T-boxes (with changed specificity ILE → LEU) upstream of the *ilv* operon of *D. reducens*.

somewhat more interesting observation is that orthologous genes encoding enzymes common to several biosynthetic pathways may be regulated by T-boxes with specificity corresponding to all terminal products. Again, the clearest examples come from the aromatic and branched-chain amino acid pathways.

The regulation of the *aro* genes (aromatic amino acid common pathway) depends on the phylogenetic group: the *aro* genes are regulated by a TYR-T-box in *B. cereus*, PHE-T-boxes in the Peptococcaceae, *H. mobilis*, *M. thermoacetica*, *C. thermocellum*, and a TRP-T-box in the Chloroflexales (within the *trp* operon). These T-boxes are not similar and thus seem to result from independent acquisitions (data not shown).

Simiarly, the *ilv* genes (branched-chain amino acid biosynthesis) are regulated by LEU-T-boxes (some Bacillales, Clostridales, *S. wolfei*), ILE-T-boxes (some Bacillales), a VAL-T-box (*C. acetobutylicum*), and a double ILE-LEU-T-box in *D. reducens*.

DISCUSSION

The fact that T-boxes are fairly easy to identify in genomic sequences, and in most cases their specificity can be reliably predicted, allowed us to make a number of tentative functional annotations of hypothetical proteins. In particular, in several cases the existing function was reassigned, for instance, for several transporters whose specificity had been initially annotated purely by sequence similarity to relatively distant homologs.

A more interesting case is that of annotated *metS2* of *C. beijerinckii*, which is preceded by a LYS-T-box. The latter is most similar to other LYS-T-boxes, and thus its specificity is likely correctly predicted (data not shown). The encoded protein is only weakly similar to experimentally established methyonyl-tRNA synthetases, and the phylogenetic analysis of the MetS proteins demonstrated that the clostridial MetS2 branch is a distant deeply rooted outgroup. At the same time, *C. beijerinckii*, like all Clostridia, has the classical

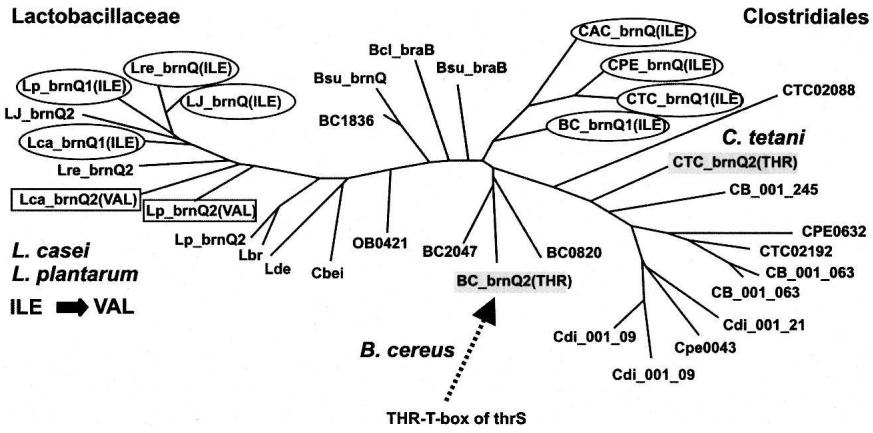

FIGURE 7. Protein phylogenetic tree of branched-chain amino acid transporters (*brnQ*/*braB*). (Circles) Genes regulated by ILE-, (rectangles) VAL-, and (filled rectangles) THR-T-boxes.

metS gene. Thus, we predict that *metS2* may encode a new lysyl-tRNA synthetase not homologous to known lysyl-tRNA synthetases and distantly related to methyonyl-tRNA synthetases. Of course, this is a very tentative prediction that requires experimental verification.

The relationships between T-box specificity and the regulated gene specificity may not be straightforward, as shown by rapid changes of the ASP ↔ ASN specificity of T-boxes upstream of the *hisS-aspS* operon. Given that in many species asparaginyl-tRNA is formed by transamidation of aspartyl-tRNA, and asparagynyl-tRNA synthetase is missing (Wolf et al. 1999; Woese et al. 2000), a more detailed study of the relationships between tRNA, aminoacyl-tRNA synthetases, and T-boxes would be interesting.

Another interesting aspect is the choice of the specifier codon among all codons corresponding to a given amino acid. We compared the codon distribution both in T-boxes and in gene-coding regions in the Firmicutes. Predominantly one or two codons are preferred in T-box regulatory sites (Table 5). In most cases, the results showed anticorrelation between the choice of the regulatory codon in T-boxes and the codon frequency in genes. For example, the rarest codon (NNC or NNG) is used in all T-boxes corresponding to all amino acids with two codons (PHE, TYR, ASN, ASP, HIS, LYS, CYS) (Table 5). Branched-chain amino acids (ILE, LEU) as well as SER-, THR-, and GLY-T-boxes also use the rarest or one of the rarest codons (Table 5). Two exceptions are ALA- and PRO-T-boxes, which use more frequent codons (NNU) in gene regulation. Thus, rare codons were chosen in T-boxes to regulate most genes in response to the concentration of corresponding uncharged tRNA. One possible explanation is that this makes the regulation more sensitive to changes of amino acid concentrations, similarly to translation-dependent attenuators (Harms et al. 1985; Harms and Umberger 1987; Fang et al. 2000; Elf et al. 2003). An alternative explanation is that the observed preference for C in the third codon positions provides for stronger interaction between the tRNA and the specifier codon.

A combination of the tree building approach, analysis of T-box specifier codons, and functional analysis of regulated operon allowed us to reconstruct the events that created the existing T-box regulons in the Firmicutes, using in each case the most parsimonious explanation of the observed T-box distribution (Fig. 8). Of course, this reconstruction is rough and tentative, and its details will need to be reconsidered when more genomes are available and the dynamics of the T-box evolution are better understood. Still, some observations seem to be firmly established, such as the T-box-dependent regulation of most aminoacyl-tRNA synthetase genes as well as tryptophan and branched-chain biosynthesis and transport genes in the last common ancestor of the Firmicutes, and secondary loss of many T-boxes in the Lactobacillales, and specifically, in the Streptococcaceae.

The evolutionary processes shaping the existent set of T-boxes are duplications providing raw material for evolution, and changes in specificity. As it has been mentioned above, all cases of identifiable changes of T-box specificity involve substitutions in the first codon position. One possible reason for that could be that codons differing in the first position often encode chemically related amino acids synthesized by homologous enzymes or imported by homologous transporters. Thus subtle changes in gene functional specificity following a duplication event may be accompanied by parallel changes in the specificity of the T-box regulating this gene.

The duplications with subsequent fixation may occur at a very fast rate, as demonstrated by the expansion of the ARG-T-box regulon in *C. difficile* that occurred simultaneously with lineage-specific loss of transcriptional regulation by the arginine repressor AhrC. The mechanism of duplication and selection should be very efficient. Indeed, assuming that duplicated T-boxes do not target upstream regions of specific genes (no mechanism of such targeted T-box insertion is feasible), they are inserted in random regions and then eliminated by selection if incorporated in a wrong place. On the other hand, we have not observed intermediate stages of this process, namely, cases of T-boxes or their identifiable remnants upstream of obviously irrelevant genes. It might be interesting to attempt to estimate the evolutionary parameters of this process, although the available data may be insufficient for that.

Another promising area of further research is modeling of the mechanism of T-box regulation on the level of RNA

TABLE 5. The distribution of codons in T-boxes and genes in the Firmicutes

A\A	PHE	TYR	ASN	ASP	HIS	LYS	CYS	THR
T-box	**UUC (47, 100%)**	**UAC (41, 100%)**	**AAC** (27, 100%)	**GAC** (21, 100%)	**CAC (16, 100%)**	**AAG (6, 100%)**	**UGC (29, 100%)**	**ACC (31, 55%)**
	UUU (0)	UAU (0)	AAU (0)	GAU (0)	CAU (0)	AAA (0)	UGU (0)	ACU (24, 43%)
								ACA (1, 2%)
Genes	UUU (73%)	UAU (69%)	AAU (69%)	GAU (71%)	CAU (68%)	AAA (68%)	UGU (62%)	ACA (34%)
	UUC (27%)	**UAC** (31%)	**AAC** (31%)	**GAC** (29%)	**CAC** (32%)	**AAG (32%)**	**UGC (38%)**	ACU (29%)
								ACC (19%)
								ACG (18%)
A\A	ILE	LEU	VAL	SER	ALA	PRO	GLY	ARG
T-box	**AUC (68, 100%)**	**CUC (48, 82%)**	**GUA (24, 51%)**	**UCC (28, 63%)**	**GCU (49, 91%)**	**CCU (17, 82%)**	**GGC (34, 83%)**	**AGA (9, 43%)**
	Others (0)	CUU (10, 18%)	GUC (11, 23%)	UCU (15, 33%)	GCC (5, 8%)	CCC (2, 9%)	GGA (4, 10%)	**CGC (7, 34%)**
		Others (0)	GUU (12, 26%)	AGC (2, 4%)	Others (0)	CCA (2, 9%)	GGG (3, 7%)	CGU (3, 14%)
			GTG (0)	Others (0)		CCG (0)	GGT (0)	AGG (2, 9%)
								Others (0)
Genes	AUU (53%)	UUA (35%)	GUU (38%)	UCA (24%)	**GCU (34%)**	CCA (36%)	GGT (32%)	**AGA (26%)**
	AUA (24%)	CUU (19%)	**GUA (26%)**	AGU (22%)	GCA (29%)	**CCU (31%)**	GGA (31%)	CGU (24%)
	AUC (23%)	UUG (18%)	GUG (19%)	UCU (21%)	GCC (22%)	CCG (21%)	**GGC (21%)**	**CGC (15%)**
		CUG (12%)	GUC (17%)	AGC (15%)	GCG (15%)	CCC (12%)	GGG (16%)	CGG (14%)
		CUA (9%)		**UCC (10%)**				CGA (11%)
		CUC (7%)		UCG (8%)				AGG (10%)

The average codon percentage in T-boxes and protein-coding regions is shown in the first and the second lines, respectively. Codons were arranged by decreasing frequency. Codons that are preferred in T-boxes are set in bold.

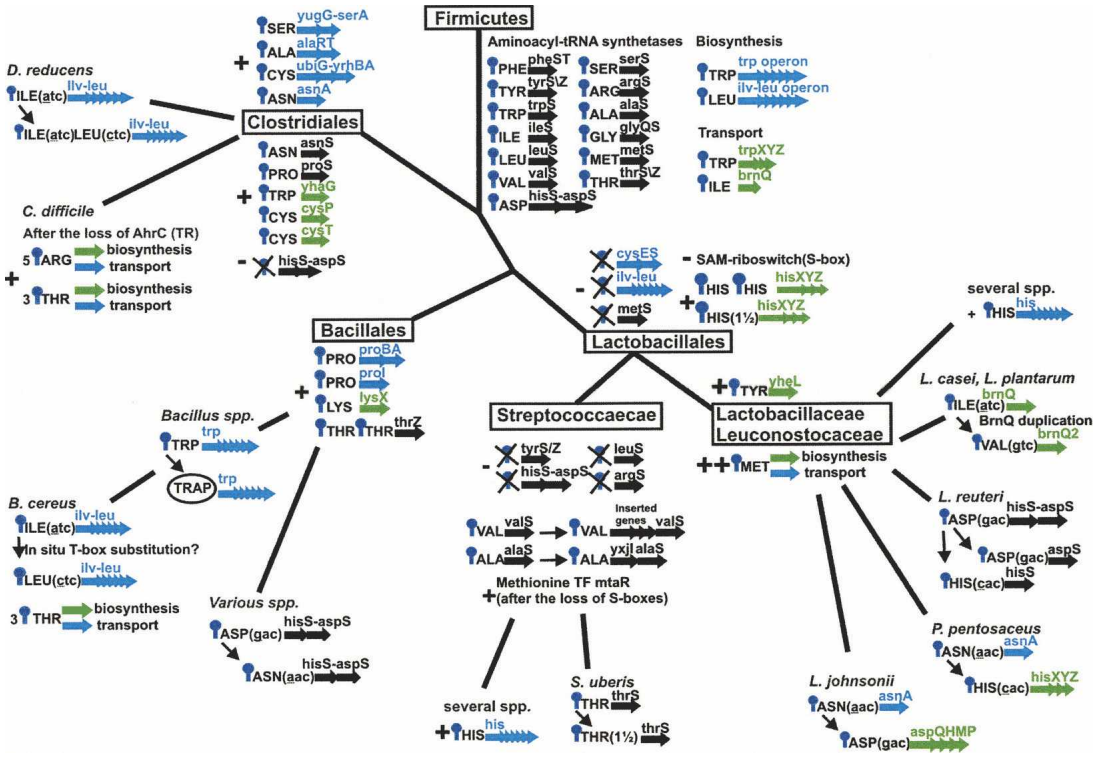

FIGURE 8. The taxonomy tree of most T-box-containing Firmicutes and the evolutionary events reconstructed as the most parsimonious explanation of the observed T-box distribution and phylogenetic relationships. The notation is as in Figure 5. This figure combines the events described in the text and illustrated in Figures 5 and 6. Pluses and minuses denote, respectively, gains and losses of T-boxes and other regulators. Numerous aminoacyl-tRNA synthetase genes, two biosynthetic operons (tryptophan *trp* and branched-chain amino acids *ilv-leu*), and two transporter operons (tryptophan *trpXYZ* and branched-chain amino acids *brnQ*) were regulated by T-boxes in the ancestral genome. In the ancestor of the Clostridiales, the T-box regulon expanded by several biosynthetic and transporter operons, with additional expansion in *C. difficile*. In the Bacillales, proline biosynthesis and lysine transporter genes become T-box-regulated, and subsequent lineage-specific specificity changes occurred. In the ancestor of the Lactobacillales, T-box regulation of the *cysES*, *ilv-leu*, and *metS* operons and S-box regulation of methionine metabolism genes was lost. A variety of changes occurred in the Streptococcae, Lactobacillaceae, and Leuconostocaceae, and in specific lineages.

kinetics. At that, specifically interesting are consequences of duplications in situ that may be followed by partial deletion: double and partially double T-boxes. As mentioned above, simple yes–no models of the latter based on discrete models of alternative RNA secondary structure yield a paradoxical conclusion that the gene is never expressed.

Unlike the majority of transcription-regulating T-boxes, the translation-regulating ones may function not only cotranscriptionally, but also regulate expression of already transcribed mRNAs. Thus the kinetic and thermodynamic parameters of such regulation might be different. To our knowledge, none of the translation-regulating T-boxes have been studied in experiment. At that, the new, reduced class of candidate actinobacterial T-boxes is a particularly interesting subject for experimental structural and functional studies.

The problem of relationship between T-boxes and other regulatory systems also deserves detailed scrutiny. In some cases, T-boxes coexist with other systems, regulating different sets of genes. For instance, LYS-T-boxes regulate only some transporters (*lysX* from Bacilalles) and, in rare cases, lysyl-tRNA synthetases, whereas LYS-element (L-box) riboswitches regulate most biosynthesis and transporter genes in Firmicutes (*lysP*, *lysXY*, *lysW*, *yvsH*, *ycgA*) (Grundy et al. 2003; Rodionov et al. 2003; Sudarsan et al. 2003; Winkler et al. 2003). T-boxes may co-regulate genes with transcription factors, for example, ALA-T-boxes and the transcription factor AlaR regulating the *alaRT* operon (our observations and data in the Subtilist database communicated by A.L. Sonenshein; http://genolist.pasteur.fr/SubtiList/), or with translational regulators, for example, TRP-T-box and TRAP regulating the putative tryptophan transporter *ycbK* (Valbuzzi and Yanofsky 2001; Babitzke 2004). Moreover, T-boxes may also participate in regulatory cascades, for example, in *B. subtilis* and *B. licheniformis*, TRP-T-boxes regulate the *yczA* gene encoding an anti-TRAP protein (Valbuzzi and Yanofsky 2001; Chen and Yanofsky 2003).

However, a more interesting situation arises when a regulatory system completely substitutes an ancestral system. Two clear examples of rapid T-box regulon expansions following the loss of original regulatory systems have been observed: MET-T-boxes in Lactobacillaceae (loss of

S-box riboswitches) (Rodionov et al. 2004), and ARG-T-boxes in *C. difficile* (loss of the AhrC repressor) (above). It is likely that other cases of lineage-specific T-box expansions, such as THR-T-box expansions in *B. cereus* and *C. difficile*, also could have been caused by the loss of the original regulatory systems. In other cases, T-box regulation gave way to other regulatory systems. For instance, in some *Bacillus* spp., the original TRP-T-boxes regulating the *trp* operon were substituted by the TRAP RNA-binding protein (Sarsero et al. 2000).

In all cases, the phylogenetic analysis clearly indicates that the source of expanded T-box families is T-boxes regulating aminoacyl-tRNA synthetase genes. Indeed, this seems to be the original role of the T-box regulation. Similarly, given the scattered appearance of T-boxes outside the Firmicutes and T-box propensity to duplication, it is likely that T-boxes in other taxonomic groups are the result of horizontal transfer from the Firmicutes. However, the phylogenetic trees (data not shown) indicate that the T-box-regulated genes in Gram-negative bacteria (δ-proteobacteria) and other groups outside the Firmicutes were not horizontally transferred from the Firmicutes, and thus the simultaneous transfer of the T-boxes and the currently regulated genes seems unlikely even for aminoacyl-tRNA synthetase genes that are often horizontally transferred (Wolf et al. 1999). At present, it is not clear, whether these T-boxes have entered the new genome alone or they are remains of old amino acid T-box regulons.

Overall, due to the relative ease of recognition and establishing specificity of T-boxes, they provide a unique possibility to study the evolutionary dynamics of regulatory systems in bacteria. We believe that many open problems left by our study might be resolved when more genomic data become available.

MATERIALS AND METHODS

Complete and partial sequences of bacterial genomes were downloaded from GenBank (Benson et al. 2007). Preliminary sequence data were obtained from 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 Sanger Center (http://www.sanger.ac.uk), and the DOE Joint Genome Institute (http://www.jgi.doe.gov).

The RNApattern program (Vitreschak et al. 2001) was used to identify candidate T-boxes. This is a pattern-matching program that allows one to characterize the input pattern as a combination of the RNA secondary structure and sequence consensus motifs. The input RNA pattern described an RNA motif as a set of the following parameters: the number of helices, the length of each helix, the loop lengths, and description of the topology (mutual arrangement) of helices. These definitions are similar to the approach implemented in the Palingol algorithm (Billoud et al. 1996). The T-box pattern (Fig. 9) was constructed using a training set of known T-boxes (Henkin et al. 1992, 1994; Vander Horn and Zahler 1992; Grundy et al. 1997a; Luo et al. 1997, 1998; Frenkiel et al. 1998; Pelchat and Lapointe 1999; Panina et al. 2003; Rodionov et al. 2003, 2004; Grundy and Henkin 2004; Gutierrez-Preciado et al. 2005) and includes the most conserved structures in T-boxes, namely, the specifier hairpin and the alternative antiterminator/terminator structures. The identified T-boxes were aligned using the sequence/structure alignment program MultAl (A.A. Mironov, unpubl.) with manual correction, where necessary. The specificity of T-boxes was assigned by inspection of anti-anti-codon regions (specifier hairpin) and validated by construction of phylogenetic trees and functional analysis of regulated genes.

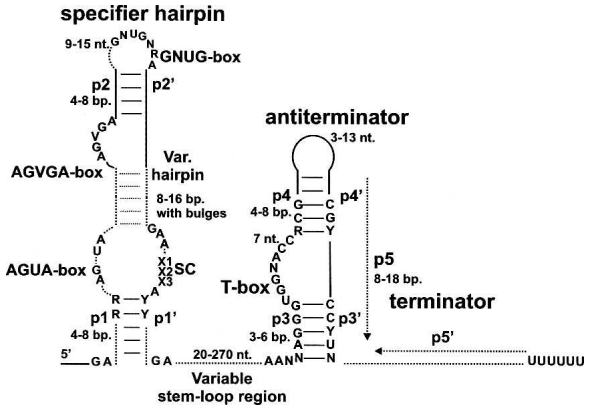

FIGURE 9. The T-box search pattern. The RNA secondary structure stems are numbered P1–P3 (the specifier hairpin), P4–P5 (the antiterminator), and P6 (the terminator). The conserved positions/boxes and structural-element lengths are shown.

Orthologous proteins were defined by the best bidirectional hits criterion (Tatusov et al. 2001) implemented in the GenomeExplorer program (Mironov et al. 2000). Distant protein homologs were identified using PSI-BLAST (Altschul and Koonin 1998). Multiple protein sequence alignments were constructed using CLUSTALX (Thompson et al. 1997).

Phylogenetic trees of T-boxes and proteins were constructed using the maximum likelihood algorithm implemented in PHYLIP (Felsenstein 1981) and drawn using the GeneMaster program (A.A. Mironov, unpubl.). Only highly conserved regions of T-box alignments, the specifier hairpin, and the terminator–antiterminator region were used.

Candidate operons were defined as sets of genes transcribed in the same direction, with short intergenic spacers and potentially belonging to the same functional subsystem.

SUPPLEMENTAL DATA

Supplemental material can be found at http://www.rnajournal.org.

ACKNOWLEDGMENTS

We are grateful to Dmitry Rodionov, Konstantin Severinov, and Alexander Spirin for useful discussions. This study was partially supported by grants from the Howard Hughes Medical Institute (55005610), INTAS (05-8028), and the Russian Academy of Sciences (program "Molecular and Cellular Biology").

Received September 6, 2007; accepted December 31, 2007.

REFERENCES

Altschul, S.F. and Koonin, E.V. 1998. Iterated profile searches with PSI-BLAST—A tool for discovery in protein databases. *Trends Biochem. Sci.* **23:** 444–447.

Babitzke, P. 2004. Regulation of transcription attenuation and translation initiation by allosteric control of an RNA-binding protein: The *Bacillus subtilis* TRAP protein. *Curr. Opin. Microbiol.* **7:** 132–139.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. 2007. GenBank. *Nucleic Acids Res.* **35:** D21–D25. doi: 10.1093/nar/gkl986.

Billoud, B., Kontic, M., and Viari, A. 1996. Palingol: A declarative programming language to describe nucleic acids' secondary structures and to scan sequence database. *Nucleic Acids Res.* **24:** 1395–1403. doi: 10.1093/nar/24.8.1395.

Burguiere, P., Auger, S., Hullo, M.F., Danchin, A., and Martin-Verstraete, I. 2004. Three different systems participate in L-cystine uptake in *Bacillus subtilis*. *J. Bacteriol.* **186:** 4875–4884.

Chen, G. and Yanofsky, C. 2003. Tandem transcription and translation regulatory sensing of uncharged tryptophan tRNA. *Science* **301:** 211–213.

Chopin, A., Biaudet, V., and Ehrlich, S.D. 1998. Analysis of the *Bacillus subtilis* genome sequence reveals nine new T-box leaders. *Mol. Microbiol.* **29:** 662–664.

Delorme, C., Ehrlich, S.D., and Renault, P. 1999. Regulation of expression of the *Lactococcus lactis* histidine operon. *J. Bacteriol.* **181:** 2026–2037.

Elf, J., Nilsson, D., Tenson, T., and Ehrenberg, M. 2003. Selective charging of tRNA isoacceptors explains patterns of codon usage. *Science* **300:** 1718–1722.

Fang, P., Wang, Z., and Sachs, M.S. 2000. Evolutionarily conserved features of the arginine attenuator peptide provide the necessary requirements for its function in translational regulation. *J. Biol. Chem.* **275:** 26710–26719.

Fauzi, H., Jack, K.D., and Hines, J.V. 2005. In vitro selection to identify determinants in tRNA for *Bacillus subtilis tyrS* T box antiterminator mRNA binding. *Nucleic Acids Res.* **33:** 2595–2602.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17:** 368–376.

Frenkiel, H., Bardowski, J., Ehrlich, S.D., and Chopin, A. 1998. Transcription of the trp operon in *Lactococcus lactis* is controlled by antitermination in the leader region. *Microbiol.* **144:** 2103–2111.

Gelfand, M.S. 2006. Bacterial *cis*-regulatory RNA structures. *Mol. Biol.* **40:** 541–550.

Grundy, F.J. and Henkin, T.M. 1993. tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell* **74:** 475–482.

Grundy, F.J. and Henkin, T.M. 1994. Conservation of a transcription antitermination mechanism in aminoacyl-tRNA synthetase and amino acid biosynthesis genes in Gram-positive bacteria. *J. Mol. Biol.* **235:** 798–804.

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.

Grundy, F.J. and Henkin, T.M. 2003. The T box and S box transcription termination control systems. *Front. Biosci.* **8:** 20–31.

Grundy, F.J. and Henkin, T.M. 2004. Kinetic analysis of tRNA-directed transcription antitermination of the *Bacillus subtilis glyQS* gene in vitro. *J. Bacteriol.* **186:** 5392–5399.

Grundy, F.J., Rollins, S.M., and Henkin, T.M. 1994. Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the *Bacillus subtilis tyrS* gene: A new role for the discriminator base. *J. Bacteriol.* **176:** 4518–4526.

Grundy, F.J., Haldeman, M.T., Hornblow, G.M., Ward, J.M., Chalker, A.F., and Henkin, T.M. 1997a. The *Staphylococcus aureus ileS* gene, encoding isoleucyl-tRNA synthetase, is a member of the T-box family. *J. Bacteriol.* **179:** 3767–3772.

Grundy, F.J., Hodil, S.E., Rollins, S.M., and Henkin, T.M. 1997b. Specificity of tRNA–mRNA interactions in *Bacillus subtilis tyrS* antitermination. *J. Bacteriol.* **179:** 2587–2594.

Grundy, F.J., Collins, J.A., Rollins, S.M., and Henkin, T.M. 2000. tRNA determinants for transcription antitermination of the *Bacillus subtilis tyrS* gene. *RNA* **6:** 1131–1141.

Grundy, F.J., Moir, T.R., Haldeman, M.T., and Henkin, T.M. 2002a. Sequence requirements for terminators and antiterminators in the T box transcription antitermination system: Disparity between conservation and functional requirements. *Nucleic Acids Res.* **30:** 1646–1655.

Grundy, F.J., Winkler, W.C., and Henkin, T.M. 2002b. tRNA-mediated transcription antitermination in vitro: Codon–anticodon pairing independent of the ribosome. *Proc. Natl. Acad. Sci.* **99:** 11121–11126.

Grundy, F.J., Lehman, S.C., and Henkin, T.M. 2003. The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc. Natl. Acad. Sci.* **100:** 12057–12062.

Gutierrez-Preciado, A., Jensen, R.A., Yanofsky, C., and Merino, E. 2005. New insights into regulation of the tryptophan biosynthetic operon in Gram-positive bacteria. *Trends Genet.* **21:** 432–436.

Gutiérrez-Preciado, A., Yanofsky, C., and Merino, E. 2007. Comparison of tryptophan biosynthetic operon regulation in different Gram-positive bacterial species. *Trends Genet.* **23:** 422–426.

Harms, E. and Umbarger, H.E. 1987. Role of codon choice in the leader region of the *ilvGMEDA* operon of *Serratia marcescens*. *J. Bacteriol.* **169:** 5668–5677.

Harms, E., Hsu, J.H., Subrahmanyam, C.S., and Umbarger, H.E. 1985. Comparison of the regulatory regions of *ilvGEDA* operons from several enteric organisms. *J. Bacteriol.* **164:** 207–216.

Henkin, T.M. 1994. tRNA-directed transcription antitermination. *Mol. Microbiol.* **13:** 381–387.

Henkin, T.M. and Yanofsky, C. 2002. Regulation by transcription attenuation in bacteria: How RNA provides instructions for transcription termination/antitermination decisions. *Bioessays* **24:** 700–707.

Henkin, T.M., Glass, B.L., and Grundy, F.J. 1992. Analysis of the *Bacillus subtilis tyrS* gene: Conservation of a regulatory sequence in multiple tRNA synthetase genes. *J. Bacteriol.* **174:** 1299–1306.

Hullo, M.F., Auger, S., Dassa, E., Danchin, A., and Martin-Verstraete, I. 2004. The *metNPQ* operon of *Bacillus subtilis* encodes an ABC permease transporting methionine sulfoxide, D- and L-methionine. *Res. Microbiol.* **155:** 80–86.

Kovaleva, G.Y. and Gelfand, M.S. 2007. Transcriptional regulation of the methionine and cysteine transport and metabolism in streptococci. *FEMS Microbiol. Lett.* **276:** 207–215.

Luo, D., Leautey, J., Grunberg-Manago, M., and Putzer, H. 1997. Structure and regulation of expression of the *Bacillus subtilis* valyl-tRNA synthetase gene. *J. Bacteriol.* **179:** 2472–2478.

Luo, D., Condon, C., Grunberg-Manago, M., and Putzer, H. 1998. In vitro and in vivo secondary structure probing of the *thrS* leader in *Bacillus subtilis*. *Nucleic Acids Res.* **26:** 5379–5387. doi: 10.1093/nar/26.23.5379.

Mandal, M., Lee, M., Barrick, J.E., Weinberg, Z., Emilsson, G.M., Ruzzo, W.L., and Breaker, R.R. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306:** 275–279.

Marta, P.T., Ladner, R.D., and Grandoni, J.A. 1996. A CUC triplet confers leucine-dependent regulation of the *Bacillus subtilis ilv-leu* operon. *J. Bacteriol.* **178:** 2150–2153.

Merino, E. and Yanofsky, C. 2005. Transcription attenuation: A highly conserved regulatory strategy used by bacteria. *Trends Genet.* **21:** 260–264.

Mironov, A.A., Vinokurova, N.P., and Gelfand, M.S. 2000. GenomeExplorer: Software for analysis of complete bacterial genomes. *Mol. Biol.* **34:** 222–231.

Mwangi, M.M. and Siggia, E.D. 2003. Genome wide identification of regulatory motifs in *Bacillus subtilis*. *BMC Bioinformatics* **4:** 18.

Omelchenko, M.V., Makarova, K.S., Wolf, Y.I., Rogozin, I.B., and Koonin, E.V. 2003. Evolution of mosaic operons by horizontal gene transfer and gene displacement in situ. *Genome Biol.* **4:** R55. doi: 10.1186/gb-2003-4-9-r55.

Panina, E.M., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. 2003. Regulation of biosynthesis and transport of aromatic amino acids in low-GC Gram-positive bacteria. *FEMS Microbiol. Lett.* **222:** 211–220.

Pelchat, M. and Lapointe, J. 1999. In vivo and in vitro processing of the *Bacillus subtilis* transcript coding for glutamyl-tRNA synthetase, serine acetyltransferase, and cysteinyl-tRNA synthetase. *RNA* **5:** 281–289.

Putzer, H., Gendron, N., and Grunberg-Manago, M. 1992. Coordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: Control by transcriptional antitermination involving a conserved regulatory sequence. *EMBO J.* **11:** 3117–3127.

Putzer, H., Laalami, S., Brakhage, A.A., Condon, C., and Grunberg-Manago, M. 1995. Aminoacyl-tRNA synthetase gene regulation in *Bacillus subtilis*: Induction, repression and growth-rate regulation. *Mol. Microbiol.* **16:** 709–718.

Reig, N., del Rio, C., Casagrande, F., Ratera, M., Gelpí, J.L., Torrents, D., Henderson, P.J., Xie, H., Baldwin, S.A., Zorzan, A., et al. 2007. Functional and structural characterization of the first prokaryotic member of the L-amino acid transporter (LAT) family: A model for APC transporters. *J. Biol. Chem.* **282:** 13270–13281.

Rodionov, D.A. 2007. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem. Rev.* **107:** 3467–3497.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. 2003. Regulation of lysine biosynthesis and transport genes in bacteria: Yet another RNA riboswitch? *Nucleic Acids Res.* **31:** 6748–6757. doi: 10.1093/nar/gkg900.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. 2004. Comparative genomics of the methionine metabolism in Gram-positive bacteria: A variety of regulatory systems. *Nucleic Acids Res.* **32:** 3340–3353. doi: 10.1093/nar/gkh659.

Saier Jr., M.H., Tran, C.V., and Barabote, R.D. 2006. TCDB: The Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res.* **34:** D181–D186. doi: 10.1093/nar/gkj001.

Sarsero, J.P., Merino, E., and Yanofsky, C. 2000. A *Bacillus subtilis* operon containing genes of unknown function senses tRNATrp charging and regulates expression of the genes of tryptophan biosynthesis. *Proc. Natl. Acad. Sci.* **97:** 2656–2661.

Sekowska, A. and Danchin, A. 2002. The methionine salvage pathway in *Bacillus subtilis*. *BMC Microbiol.* **2:** 8. doi: 10.1186/1471-2180-2-8.

Seliverstov, A.V., Putzer, H., Gelfand, M.S., and Lyubetsky, V.A. 2005. Comparative analysis of RNA regulatory elements of amino acid metabolism genes in Actinobacteria. *BMC Microbiol.* **5:** 54. doi: 10.1186/1471-2180-5-54.

Sperandio, B., Gautier, C., McGovern, S., Ehrlich, D.S., Renault, P., Martin-Verstraete, I., and Guédon, E. 2007. Control of methionine synthesis and uptake by MetR and homocysteine in *Streptococcus mutans*. *J. Bacteriol.* **189:** 7032–7044.

Stucky, K., Hagting, A., Klein, J.R., Matern, H., Henrich, B., Konings, W.N., and Plapp, R. 1995. Cloning and characterization of *brnQ*, a gene encoding a low-affinity, branched-chain amino acid carrier in *Lactobacillus delbruckii* subsp. *lactis* DSM7290. *Mol. Gen. Genet.* **249:** 682–690.

Sudarsan, N., Wickiser, J.K., Nakamura, S., Ebert, M.S., and Breaker, R.R. 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & Dev.* **17:** 2688–2697.

Sudarsan, N., Hammond, M.C., Block, K.F., Welz, R., Barrick, J.E., Roth, A., and Breaker, R.R. 2006. Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314:** 300–304.

Tatusov, R.L., Natale, D.A., Garkavtsev, I.V., Tatusova, T.A., Shankavaram, U.T., Rao, B.S., Kiryutin, B., Galperin, M.Y., Fedorova, N.D., and Koonin, E.V. 2001. The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **29:** 22–28. doi: 10.1093/nar/29.1.22.

Tauch, A., Hermann, T., Burkovski, A., Krämer, R., Pühler, A., and Kalinowski, J. 1998. Isoleucine uptake in *Corynebacterium glutamicum* ATCC 13032 is directed by the *brnQ* gene product. *Arch. Microbiol.* **169:** 303–312.

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. doi: 10.1093/nar/25.24.4876.

Valbuzzi, A. and Yanofsky, C. 2001. Inhibition of the *B. subtilis* regulatory protein TRAP by the TRAP-inhibitory protein, AT. *Science* **293:** 2057–2059.

van de Guchte, M., Ehrlich, D.S., and Chopin, A. 1998. tRNATrp as a key element of antitermination in the *Lactococcus lactis trp* operon. *Mol. Microbiol.* **29:** 61–74.

van de Guchte, M., Ehrlich, S.D., and Chopin, A. 2001. Identity elements in tRNA-mediated transcription antitermination: Implication of tRNA D- and T-arms in mRNA recognition. *Microbiol.* **147:** 1223–1233.

van Nimwegen, E. 2003. Scaling laws in the functional content of genomes. *Trends Genet.* **19:** 479–484.

Vander Horn, P.B. and Zahler, S.A. 1992. Cloning and nucleotide sequence of the leucyl-tRNA synthetase gene of *Bacillus subtilis*. *J. Bacteriol.* **174:** 3928–3935.

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,"* 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. 2004. Riboswitches: The oldest mechanism for the regulation of gene expression? *Trends Genet.* **20:** 44–50.

Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N., et al. 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res.* **35:** 4809–4819. doi: 10.1093/nar/gkm487.

Welz, R. and Breaker, R.R. 2007. Ligand binding and gene control characteristics of tandem riboswitches in *Bacillus anthracis*. *RNA* **13:** 573–582.

Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E., and Breaker, R.R. 2003. An mRNA structure that controls gene expression by binding *S*-adenosylmethionine. *Nat. Struct. Biol.* **10:** 701–707.

Woese, C.R., Olsen, G.J., Ibba, M., and Soll, D. 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64:** 202–236.

Wolf, Y.I., Aravind, L., Grishin, N.V., and Koonin, E.V. 1999. Evolution of aminoacyl-tRNA synthetases—Analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.* **9:** 689–710.

Yanofsky, C. 1988. Transcription attenuation. *J. Biol. Chem.* **263:** 609–612.

Yousef, M.R., Grundy, F.J., and Henkin, T.M. 2005. Structural transitions induced by the interaction between $tRNA^{Gly}$ and the *Bacillus subtilis glyQS* T box leader RNA. *J. Mol. Biol.* **349:** 273–287.

Zhang, Z., Feige, J.N., and Chang, A.B. 2003. A transporter of *Escherichia coli* specific for L- and D-methionine is the prototype for a new family within the ABC superfamily. *Arch. Microbiol.* **180:** 88–100.