

L1.LtrB-WT

<u>Gene name</u>	<u>5' flanking</u>		<u>additional nts</u>	<u>3' flanking</u>	
	EBS2 <u>GUGUA</u>	EBS1 <u>GUGUUG</u>		EBS2 <u>GUGUA</u>	EBS1 <u>GUGUUG</u>
5'nucleotidase	(2)	ACA CACUUC CACAAA	/AAAGAUCAUAUAG-253-CUU CACAUU UACUGG /	UGAGGAACUUCGGCA	
a-acetolactate decarboxylase	(1)	CAA UACAUU UAGCAC	/ACUCUCUAGCGGCUU-26--AAG CACUA ACACACG /	GCAAAGUCGGUAUCG	
<u>Ribosomal protein L13/S9</u>	(1)	ACC ACAU ACACAC	/ACAUGCUGAUACAGG-431-GCAGACA UGCGUCUU /GUU AUCAACCAACCA		
<u>Ribosomal protein L13/S9</u>	(1)	ACA CACACA UACAGC	/UCAAAAACUGAAGU-355-UA CACGU ACGCGCCG /	UAUGGUUGAACGUAA	
<u>Ribosomal protein L13/S9</u>	(2)	ACC ACAU ACACAC	/ACAUGCUGAUACAGG-303-ACAU U UAG CACAAG /	UACAAUAUGCCGGCA	
<u>L-lactate dehydrogenase</u>	(1)	UGC AGCUUA CUCAAU	/CAUCGCUAAAAAAGG-170-UU CACAUU CAUUGA /	ACGAUGCUGAAAUUGC	
NADH dehydrogenase	(1)	AAC AGCAC ACACAGC	/AGCGGC AAAUAUUCU-563-CU CAUAA AGCAAC UG/	AAUGGGGUCUCUCCA	
ABC transporter permease	(1)	AUG CACUUC CCCAAU	/UUUCAUCUGUCCGUA-150-UUUC ACUGG UUCUA /AC AGCCCUAGUACAU		
Glutamine synthetase	(1)	CAU CACAA CCCAAC	/AGUUAACUCAUACAA-386-CAA GUU U CACA AAU /	UUGUAUCGCAUUGGG	
Glyceraldehyde dehydrogenase	(1)	AA UGUCA UACAUUC	/AAACCUUGUUC-----GUA CACUUC CAUACU /	UCGCUAAAAUCGCUA	
Elongation factor Tu	(1)	ACUC CAGA CGCGAC	/ACUGACAAACCACUC-362-CUA CUUCC ACACAAC /	UGACGUUACUGGUUC	
Ribosomal protein L19	(1)	ACU GAU UCCUGAC	/UUCGGUCCUGGUGAC-153-GUUC ACACU CA CGU /GUU GAAAAAUCGAA		
<u>Ribosomal protein L11</u>	(4)	AUU CACAU CAUCAC	/AAAAACUCCUCCAGC-233-CCU CACCU ACUUAGG /	UAGGCUUUGGAACCU	
<u>N-acetylglucosaminidase</u>	(1)	GAC AAU UACAG	/UAAUACCGUUCAG-173-GCA CAU U CAU CAU /	UCGUUAAUACACAA	
Aldehyde dehydrogenase	(1)	CCA CAGUC CACAAA	/GCUUUCAUUGUUGCC-254-UUCC U UAG CGCAAG /	UUUGAAAGAGAUCUU	
Phosphotransferase	(1)	UAU CUGGU ACACAAC	/UUCAUGAGAUUCUUG-94--UAU CACAU UCCUUG /U AUGUGGUUGCUUAA		
Ferredoxin oxidoreductase	(1)	GAC CCCUU UACAAC	/AAAUGAACAGGAACA-107-AAG CACUUC CCAGAA /	UGUCGGAUGAUUUAC	
ABC transporter	(1)	UGAA CAU UUGCGAU	/UCUAGCGCGUUCAGG-136-UA UCU CUU UGCAAC A/	ACGUUUCGUUCUAAU	
g-aminobutyrate permease	(1)	ACUC AUU UUGCAAC	/UUAUGAUGAUAAAA-56--UU CACAU CACU UUG /	AUUUUUUUUACUCCA	
<u>Uncharacterized prot 1-2</u>	(1)	UA GACCU UCCACGAU	/UUAUAGUAUACCUAG-214-AAC CACAG AAACA AAU /	UAAAAUUAUUGCUACA	
Thioredoxin reductase	(1)	AAU CAU UUGCGAC	/UGGAGCUAAUACCCG-185-AAU UACGU GCACAAG /	AAAUAUUUACAACAA	
Arsenate reductase	(1)	CAC AAU CUUACAG	/ACCCUCUUGUACAAG-185-UUU CACUUC CACAAG /	CAAUUAAUACAUUU	
NADH dehydrogenase	(1)	AAC AGCAC ACACAGC	/AGCGGC AAAUAUUCU-173-CAG CACAU UUAUUUCU /	GGCGUUUAGCCUUGG	
Transcriptional regulator	(1)	UAA ACGCA CACGAC	/AACAAUGGUUUCAAU-173-AC AGCUU UGCAAC UG/	UAAAUUUUGAUGGUU	
Uncharacterized protein	(1)	AAA CACA AGUUAU	/GGAAAAUAUAUACAA-284-UGA CUCAG GCACCA /	AAAAGAAAGCAGCAA	
Ser/Thr protein kinase	(1)	AAC AAU ACAUCAAC	/AAAAUGCUCUUCUAG-51--UGG CUCU UUCACA UG /	GAAUUAUUCACCGUG	
a-acetolactate decarboxylase	(1)	CAA UACAU UACGAC	/ACUCUCUAGCGGCUU-50--GUAUC GGUA CGUUG /AU ACGGCAAUUGGCG		
Ribosomal protein S14	(1)	AA UACAU AAUUGGCU	/AAGAAAUCAUUGGUU-64--GUCCACA UUCAGUUU /ACC GCAAAUUUAAAC		
N-acetylglucosaminidase	(1)	UAA CUCAA CUCGAC	/UUCUUCUAAUCUAAA-50--AAA UACU GGCAGC C/	CAAUUGCUUCAAUCA	

Figure S3. mRNA fragments identified at the splice junction of L1.LtrB-WT circles. Additional nts are shown along with their flanking sequences (5' flanking) (3' flanking), their origin (Gene name) and frequency of identification between parentheses. The junctions between the additional nts and their flanking regions (/) as well as the IBS1- (yellow) and IBS2- (green) like sequences are denoted. Some IBS1/2-like sequences were adjusted to optimize their potential base pairing with the EBS1/2 sequences of the intron. The number of nts separating the IBS1/2-like sequences was fixed between 0-2 nts, and their maximum distance from the junction with the intron was fixed between -14, +4 nts. The bolded nts represent residues from the IBS1- and IBS2-like sequences that can potentially base pair with the intron's EBS1 and EBS2 sequences specified above. Sequences spanning two genes and including a short intergenic region are underlined while the mRNA sequence including 3' untranslated residues is italicized.