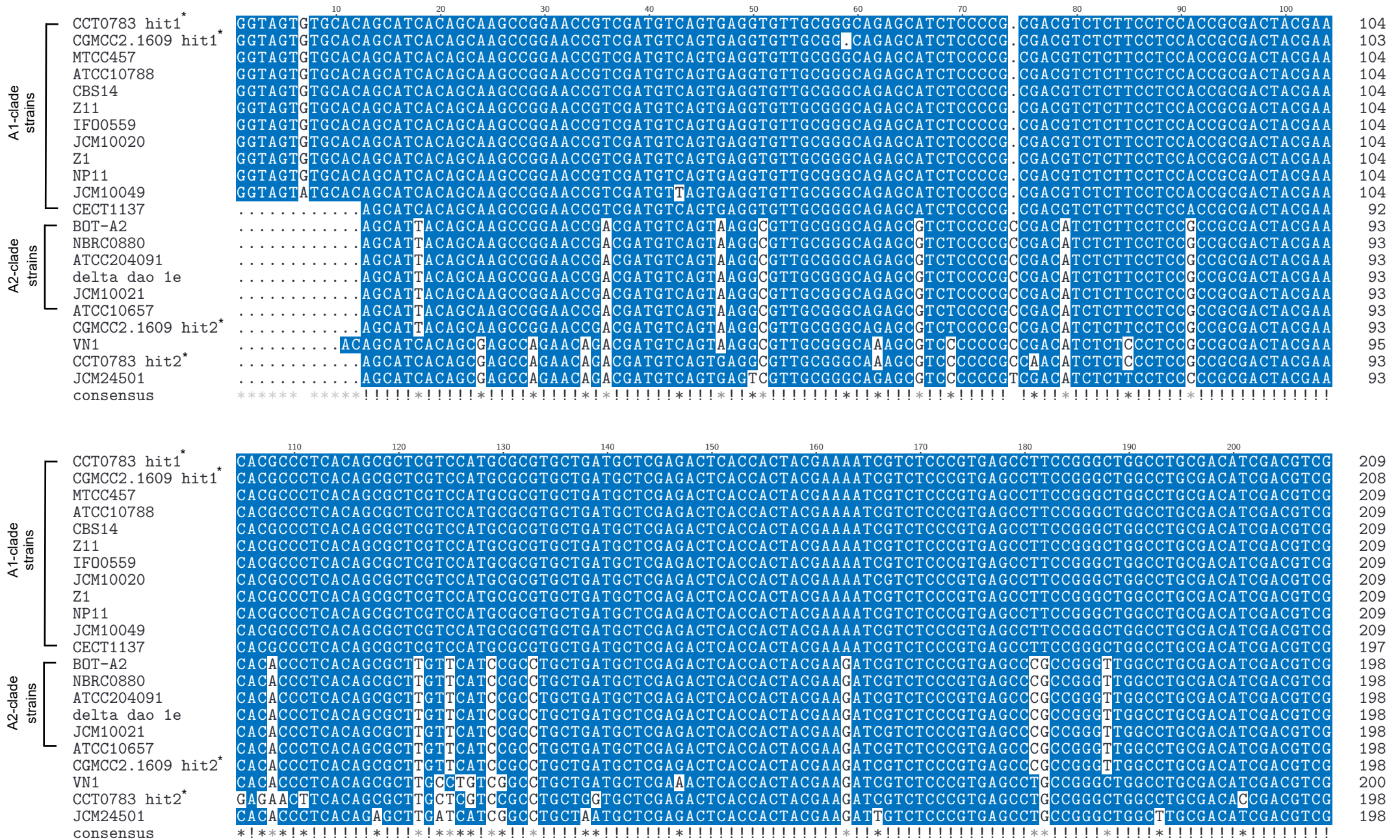




Multiple Sequence Alignment (MSA) of the putative 3877 promoter sequences from all the currently available *R. toruloides* genomes. The candidate sequence from strain BOT-A2 was identified as the 1000 kb directly upstream of the start codon of the gene, and this sequence was used as input for a nucleotide Blast analysis to identify the corresponding sequences in the other genomes. No Blast hits were found in the NRBC10032 assembly; two hits were found in each of the CGMCC2.1609 and CCT0783 assemblies and are marked with *. See the Materials and Methods section for details on the MSA. The brackets indicate the established clades from the Average Nucleotide Identity analysis (see Figure 1 in the main text).

		210	220	230	240	250	260	270	280	290	300	310	
A1-clade strains	CCT0783 hit1*	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	CGMCC2.1609 hit1*	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		312
	MTC457	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	ATCC10788	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	CBS14	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	Z11	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	IF00559	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	JCM10020	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	Z1	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
	NP11	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313
JCM10049	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313	
CECT1137	AGGCGACAGGGACC	ATGTCGACAGAAGGGCG	TTTGC	GGGTGTCGTAGACGGTGGAGAGGCCTTGTACAGATGCGCAGGGTCTTGAAG						AAGTGCTCTTGTCTGAGG		313	
A2-clade strains	BOT-A2	AGGCGACAGGGACC	ATGTCGAGG	GAAACGCGGTTCGGGGTTCGTGAAGGTTGGG	GAGGTTGGCTGGTGATGTGCAGGGA	ACTCTGAGG	AAATGCTCTGGGCGAGC						303
	NBRC0880	AGGCGACAGGGACC	ATGTCGAGG	GAAACGCGGTTCGGGGTTCGTGAAGGTTGGG	GAGGTTGGCTGGTGATGTGCAGGGA	ACTCTGAGG	AAATGCTCTGGGCGAGC						303
	ATCC204091	AGGCGACAGGGACC	ATGTCGAGG	GAAACGCGGTTCGGGGTTCGTGAAGGTTGGG	GAGGTTGGCTGGTGATGTGCAGGGA	ACTCTGAGG	AAATGCTCTGGGCGAGC						303
	delta dao 1e	AGGCGACAGGGACC	ATGTCGAGG	GAAACGCGGTTCGGGGTTCGTGAAGGTTGGG	GAGGTTGGCTGGTGATGTGCAGGGA	ACTCTGAGG	AAATGCTCTGGGCGAGC						303
	JCM10021	AGGCGACAGGGACC	ATGTCGAGG	GAAACGCGGTTCGGGGTTCGTGAAGGTTGGG	GAGGTTGGCTGGTGATGTGCAGGGA	ACTCTGAGG	AAATGCTCTGGGCGAGC						303
	ATCC10657	AGGCGACAGGGACC	ATGTCGAGG	GAAACGCGGTTCGGGGTTCGTGAAGGTTGGG	GAGGTTGGCTGGTGATGTGCAGGGA	ACTCTGAGG	AAATGCTCTGGGCGAGC						303
	CGMCC2.1609 hit2*	AGGCGACAGGGACC	ATGTCGAGG	GAAACGCGGTTCGGGGTTCGTGAAGGTTGGG	GAGGTTGGCTGGTGATGTGCAGGGA	ACTCTGAGG	AAATGCTCTGGGCGAGC						303
	VN1	AGGCGACAGGGACC	ATGTCGAGG	GAAAGACCGTGGCGGGC	GTTCAAGAGGTTGGAGAAGTATGGCGG	AGATGTGCAGGATCTTTAAC		AAGCGCTCTGGGCGAGC					304
	CCT0783 hit2*	AGGCGACAGGGACC	ATGTCGAGG	GAAAGACCGTGGCGGGC	GTTCAAGAGGTTGGAGAAGTATGGCGG	AGATGTGCAGGGTCTTTAAC		AAGCGCTCTGGGCGAGC					302
	JCM24501	AGGCGACAGGGACC	ATGTCGAGG	GAAAGCGGCTACGGGTTCGTGGAGGTTGGAGAGG	CTGTGGAGATGTGGGAGGTGTTTCGAG		GAGTGCTGTGGGCGAGC						302
consensus													

[illegible]

[illegible][illegible]

		630	640	650	660	670	680	690	700	710	720	730	
A1-clade strains	CCT0783 hit1*	C	C	T	C	G	T	C	A	C	C	G	732
	CGMCC2.1609 hit1*	C	C	T	C	G	T	C	A	C	C	G	731
	MTCC457	C	C	T	C	G	T	C	A	C	C	G	732
	ATCC10788	C	C	T	C	G	T	C	A	C	C	G	732
	CBS14	C	C	T	C	G	T	C	A	C	C	G	732
	Z11	C	C	T	C	G	T	C	A	C	C	G	732
	IF00559	C	C	T	C	G	T	C	A	C	C	G	732
	JCM10020	C	C	T	C	G	T	C	A	C	C	G	732
	Z1	C	C	T	C	G	T	C	A	C	C	G	732
	NP11	C	C	T	C	G	T	C	A	C	C	G	732
	JCM10049	C	C	T	C	G	T	C	A	C	C	G	732
	CECT1137	C	C	T	C	G	T	C	A	C	C	G	720
	BOT-A2	T	C	C	A	G	A	C	G	A	A	G	722
A2-clade strains	NBRC0880	T	C	C	A	G	A	C	G	A	A	G	722
	ATCC204091	T	C	C	A	G	A	C	G	A	A	G	722
	delta dao 1e	T	C	C	A	G	A	C	G	A	A	G	722
	JCM10021	T	C	C	A	G	A	C	G	A	A	G	722
	ATCC10657	T	C	C	A	G	A	C	G	A	A	G	722
	CGMCC2.1609 hit2*	T	C	C	A	G	A	C	G	A	A	G	722
	VN1	T	C	C	A	G	A	C	G	A	A	G	723
	CCT0783 hit2*	T	C	C	A	G	A	C	G	A	A	G	721
	JCM24501	T	C	C	A	G	A	C	G	A	A	G	721
	consensus	*	!	*	!	*	!	*	!	*	!	*	

		740	750	760	770	780	790	800	810	820	830	
A1-clade strains	CCT0783 hit1*	T	C	C	T	T	C	T	G	C	G	837
	CGMCC2.1609 hit1*	T	C	C	T	T	C	T	G	C	G	836
	MTCC457	T	C	C	T	T	C	T	G	C	G	837
	ATCC10788	T	C	C	T	T	C	T	G	C	G	837
	CBS14	T	C	C	T	T	C	T	G	C	G	837
	Z11	T	C	C	T	T	C	T	G	C	G	837
	IF00559	T	C	C	T	T	C	T	G	C	G	837
	JCM10020	T	C	C	T	T	C	T	G	C	G	837
	Z1	T	C	C	T	T	C	T	G	C	G	837
	NP11	T	C	C	T	T	C	T	G	C	G	837
	JCM10049	T	C	C	T	T	C	T	G	C	G	837
	CECT1137	T	C	C	T	T	C	T	G	C	G	825
	BOT-A2	T	C	C	T	T	C	T	G	C	G	827
A2-clade strains	NBRC0880	T	C	C	T	T	C	T	G	C	G	827
	ATCC204091	T	C	C	T	T	C	T	G	C	G	827
	delta dao 1e	T	C	C	T	T	C	T	G	C	G	827
	JCM10021	T	C	C	T	T	C	T	G	C	G	827
	ATCC10657	T	C	C	T	T	C	T	G	C	G	827
	CGMCC2.1609 hit2*	T	C	C	T	T	C	T	G	C	G	827
	VN1	T	C	C	T	T	C	T	G	C	G	827
	CCT0783 hit2*	T	C	C	T	T	C	T	G	C	G	826
	JCM24501	T	C	C	T	T	C	T	G	C	G	826
	consensus	!	!	!	*	!	!	!	*	!	!	

[illegible][illegible]