

Supplementary Material for:
**Industrial fuel ethanol yeasts contain adaptive copy number changes in genes
involved in vitamin B1 and B6 biosynthesis**

B. U. Stambuk*, B. Dunn, S. L. Alves Jr, E. H. Duval, G. Sherlock*

*To whom correspondence should be addressed. E-mail: bstambuk@mbox1.ufsc.br (B.U.S.);
sherlock@genome.stanford.edu (G.S.)

This PDF file includes

Supplementary Tables S1 to S2

Supplementary Figures S1 to S5

Supplementary Table S1. Yeast strains analyzed in this study

Yeast Strains	Relevant genotype or description:	Source or reference:
<u>Industrial Fuel Ethanol Yeasts:</u>		
CAT-1	Isolated in 1998/1999 from <i>Usina VO Catanduva</i> , Sao Paulo (Brazil)	H.V. Amorim ^a
EHD04	Isogenic to CAT-1, but <i>SUC2/suc2::kanMX6</i>	This study
PE-2	Isolated in 1993/1994 from <i>Usina da Pedra</i> , Sao Paulo (Brazil)	H.V. Amorim
VR-1	Isolated in 1993/1994 from <i>Usina Vale do Rosario</i> , Sao Paulo (Brazil)	”
BG-1	Isolated in 1989/1990 from <i>Usina Barra Grande</i> , Sao Paulo (Brazil)	J. Finguerut ^b
SA-1	Isolated in 1989/1990 from <i>Usina Santa Adelia</i> , Sao Paulo (Brazil)	”
<u>Other Industrial Yeasts:</u>		
GSY924	Leinenkugel Ale Brewing yeast	T. Pugh ^c
GSY2	Montrachet wine yeast	(Dunn et al. 2005)
GSY3	”	”
GSY10	Prise de Mousse wine yeast	”
GSY11	”	”
GSY149	Baker’s yeast	M. Ettayebi ^d
GSY150	”	”
GSY154	”	”
GSY155	”	”
<u>Laboratory Yeasts:</u>		
S288C	<i>MATα mal gal2 mel flo1 flo8-1 hap1 SUC2</i>	(Mortimer and Johnston 1986)
FY1679	<i>MATa/MATα GAL2/GAL2 ura3-52/ura3-52</i>	(Winston et al. 1995)
GSY157	<i>MATa/MATα GAL2/GAL2</i>	This study

^aFermentec Ltda., Piracicaba, Brazil. ^bCentro de Tecnologia Canavieira, Piracicaba, Brazil. ^cMiller Brewing Co, WI, USA.

^dUniversity Sidi Mohamed Ben Abdallah, Fez, Morocco.

Supplementary Table S2. Oligonucleotide primers used in this study

Primer	Sequence (5'→3')
Seq-SNZ-F1	GGGCCACACCAGTTCTTTGGG
Seq-SNZ-R1	CCCAAAGAACTGGTGTGGCCC
Seq-SNZ-F2	GCACCTCTTCTATTACTGGGGC
Seq-SNZ-R2	GCCCCAGTAATAGAAGAGGTGC
Seq-SNZ-F3	CCTTCACTGTCATCACAGATAG
Seq-SNZ-F3B	GAAATCCCTGTTTTTCGACGATGC
Seq-SNZ-R3	CTATCTGTGATGACAGTGAAGG
Seq-SNZ-F4	ATGTTTCTCCACCATAGTCACA
Seq-SNZ-R4	TGTGACTATGGTGGAGAAACAT
Seq-SNZ-F5	GACACATCACATTGAGAAGCA
Seq-SNZ-R5	TGCTTCTCAATGTGATGTGTC
Seq-SNZ-F6	CAGACGCTGCTCTATTGATGC
Seq-SNZ-R6	GCATCAATAGAGCAGCGTCTG
Seq-SNZ-F7	CCTTCGAAGGGGCAGTCGG
Seq-SNZ-R7	CCGACTGCCCCTTCGAAGG
SNO2/3-For	GCGACACGTAGAAAAATGCATCGTC
SNO2/3-Rev	CCCATCTAATTTGTACAGGACCTGC
SNO/SNZ-For	CGGAGATCTTTGTACGTCACCTGCTTACTTCATG
SNO/SNZ-Rev	GGCAGATCTGAACTGTACTCATTGCATCATAGCC
SNZ2/3-For	GGCGGAGATCCAGCAGTATAAAGAG
SNZ2/3-Rev	TTGCACCGTTTTTGCCTCCTGCTTC
SUC2-F1	CAAGCAAAACAAAAAGCTTTTCTTTTCACTAACGTATA TGCCAGCTGAAGCTTCGTACGC
SUC2-R1	CTTTTGAAAAAATAAAAAAGACAATAAGTTTTATAAC CTGCATAGGCCACTAGTGGATC

Supplementary Figure legends:

Supplementary Figure S1. Microarray karyotyping of the 9 other non-fuel industrial yeast strains. A consensus karyoscope plot pooling the results of the 4 baker's, 4 wine and the ale brewing yeast strains relative to the lab strain S288C was generated by the program CGH-Miner. The plot displays statistically significantly altered copy number variations (red for amplified genes, green for depleted) along each chromosome, and the chromosomes are shown in numerical order from top to bottom.

Supplementary Figure S2. Genes with significant CNV (**A**) exclusive to the 5 fuel ethanol yeast strains, (**B**) in all 14 industrial yeast strains examined (both fuel and non-fuel), and (**C**) exclusive to the 9 other (4 baker's, 4 wine and the ale brewing yeast) non-fuel industrial strains. To generate these figures the average data for all strains (in duplicate for the fuel ethanol strains) were sorted from the highest $\log_2$ R/G values (red) to the lowest $\log_2$ R/G values (green), using the Java TreeView program. Those genes defined as significantly amplified or deleted in the industrial yeast strains are indicated by a red or green line at the side of each figure, respectively, while the other genes are classified as polymorphic (see Methods)

Supplementary Figure S3. Growth of yeast cells in the absence of pyridoxine and presence or not of thiamin. Typical growth curves of the industrial fuel ethanol strains PE-2 (**A**), SA-1 (**B**), VR-1 (**C**) and BG-1 (**D**) strains, as well as the diploid laboratory *ura3/ura3* yeast strain FY1679 transformed with a multi-copy 2μ *URA3*⁺ plasmid containing a cloned *SNO/SNZ* gene pair from

strain PE-2 (pBD1002) (E), or a cloned *SNO/SNZ* gene pair from strain CAT-1 (pBD1003) (F), in synthetic defined media containing 200 g/L glucose and lacking both pyridoxine and thiamin (green), or the same media supplemented with 1 mg/L of thiamin (blue).

Supplementary Figure S4. CLUSTALW sequence alignment of cloned *SNO/SNZ* loci. The insert sequences in pBD1001 (from strain S288C), pBD1002 (PE-2), and pBD1003 (CAT-1) were aligned using CLUSTALW with sequences derived from the *Saccharomyces* Genome Database present in the *SNO2/SNZ2* and *SNO3/SNZ3* loci on chromosomes XIV and VI, respectively. The cloned fragment we obtained from the S288C strain is identical to the chromosome VI sequence, including all of *SNO3*, up to bp 1532; however, starting with the nucleotide difference at bp 1882 (which is located in between the two genes), the insert becomes identical to the chromosome XIV sequence, yielding a *SNZ2* gene identical to that on chromosome XIV. There is also one change in the pBD1001 insert relative to the SGD sequence, where a C is deleted at bp 2644 outside of the *SNO* and *SNZ* coding sequences; this may represent a sequencing error present in the SGD consensus sequence, a SNP present in our particular S288C strain, or a PCR-induced sequence change. Our sequencing results demonstrate that in S288C *SNO3* is actually physically paired with *SNZ2*, with both presumably residing on chromosome XIV (since *SNZ2* is the more centromeric-proximal gene); likewise *SNO2* is presumably paired with *SNZ3* on chromosome VI; it is likely that due to the very high sequence identity between these 2 regions resulted in an assembly error of the S288C genome sequence. The assembly difficulties may also explain why there is a 1 bp difference between our sequence and the SGD sequence, i.e., that it is not necessarily due to PCR artifact. The inserts from the

two fuel ethanol strains diverge slightly from that of the SGD (either from chromosome VI or XIV) and the pBD1001 sequences, with the insert from PE-2 (pBD1002) showing 12 nucleotide differences, and the insert from CAT-1 (pBD1003) showing 31 nucleotide differences, both out of 2887 bp total. Of these differences, only 2 are shared by the PE-2 and CAT-1 inserts.

Supplementary Figure S5. (A) Amino acid changes between the *SNO* and *SNZ* genes cloned from strains S288C, CAT-1 and PE-2, compared with the sequences present in the chromosomes VI (*SNO3* and *SNZ3*) and XVI (*SNO2* and *SNZ2*) from the *Saccharomyces* Genome Database (SGD). (B) Parsimony trees for the *SNO* and *SNZ* DNA sequences corresponding to the ORF coding regions. The trees were generated by first aligning the sequences using MUSCLE (version 3.7; Edgar 2004), and then using the DNAPars program that is part of the PHYLIP package (version 3.68; <http://evolution.genetics.washington.edu/phylip.html>). For visualization purposes, the trees were rooted at their midpoints using retree, also part of the PHYLIP package.

Supplementary Figures

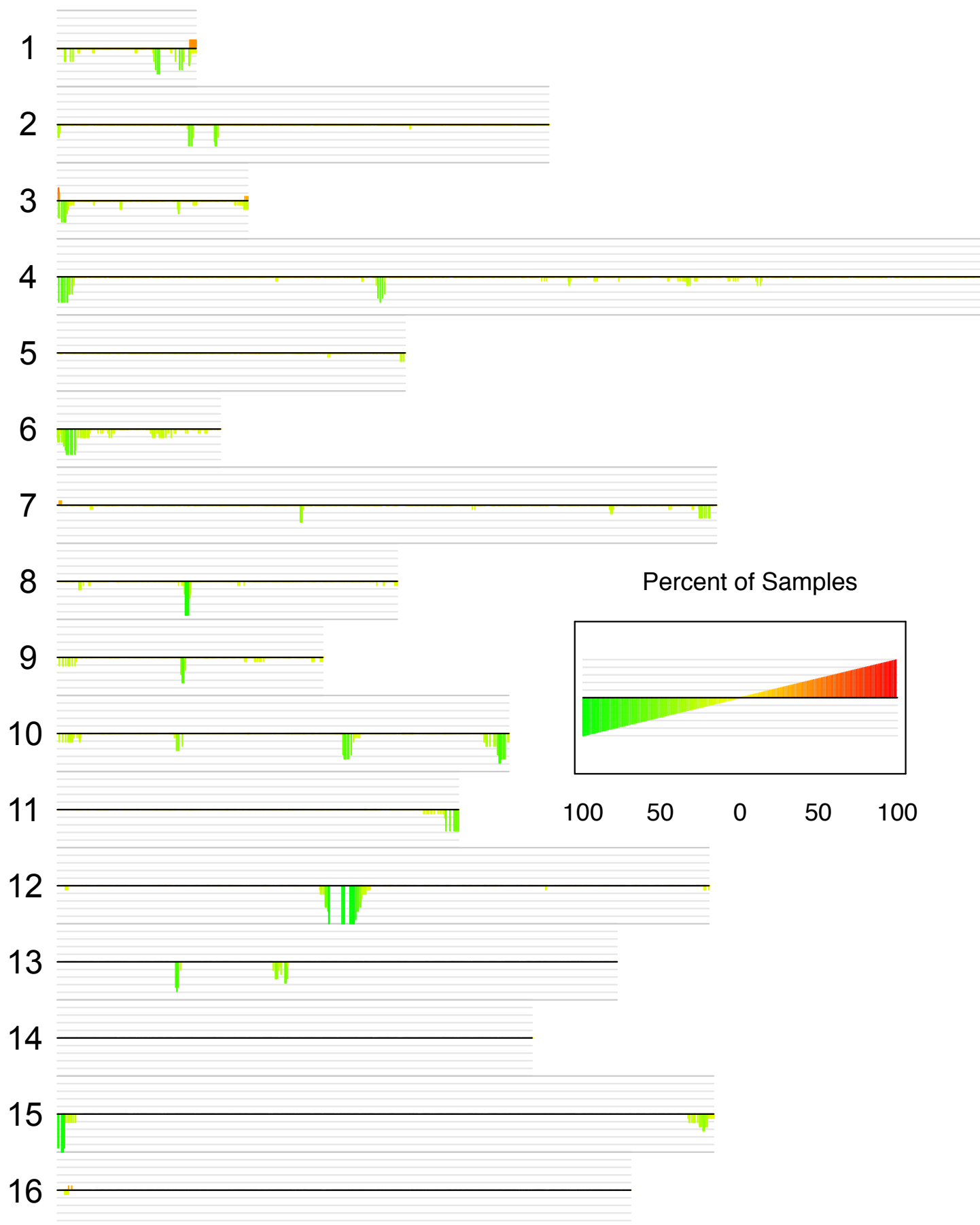
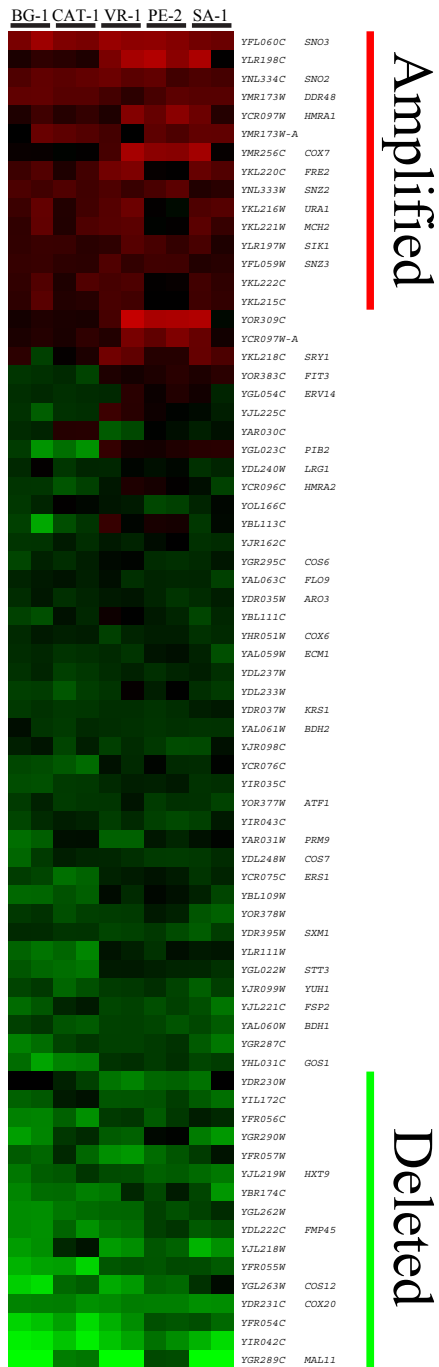
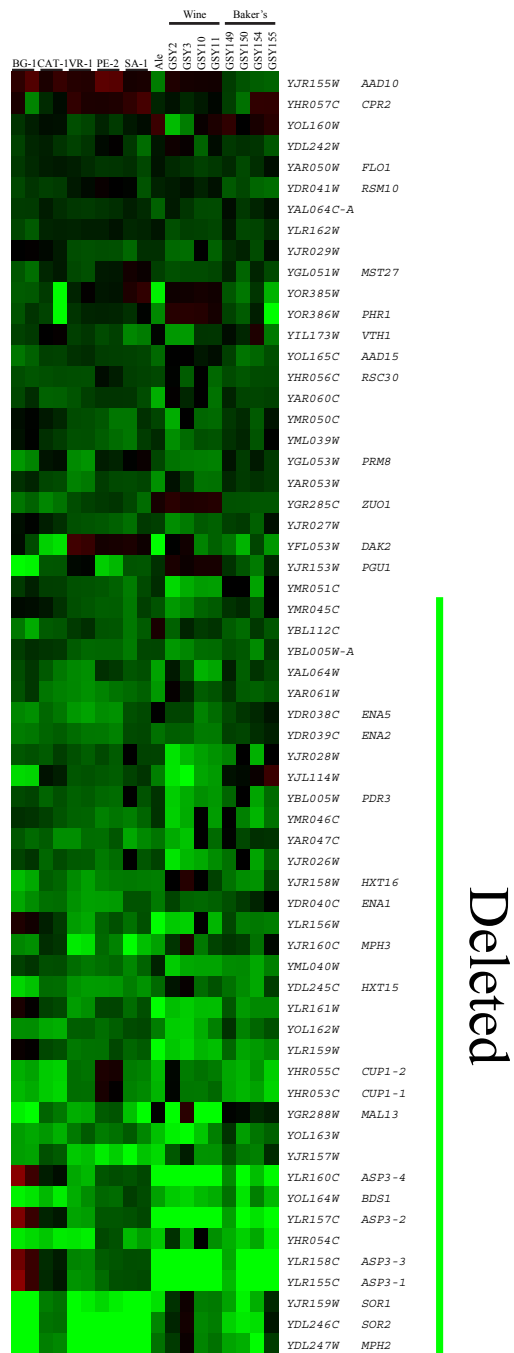


Figure S1

A



B



C

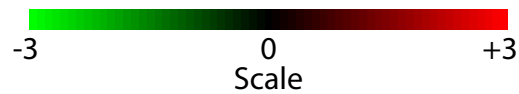


Figure S2

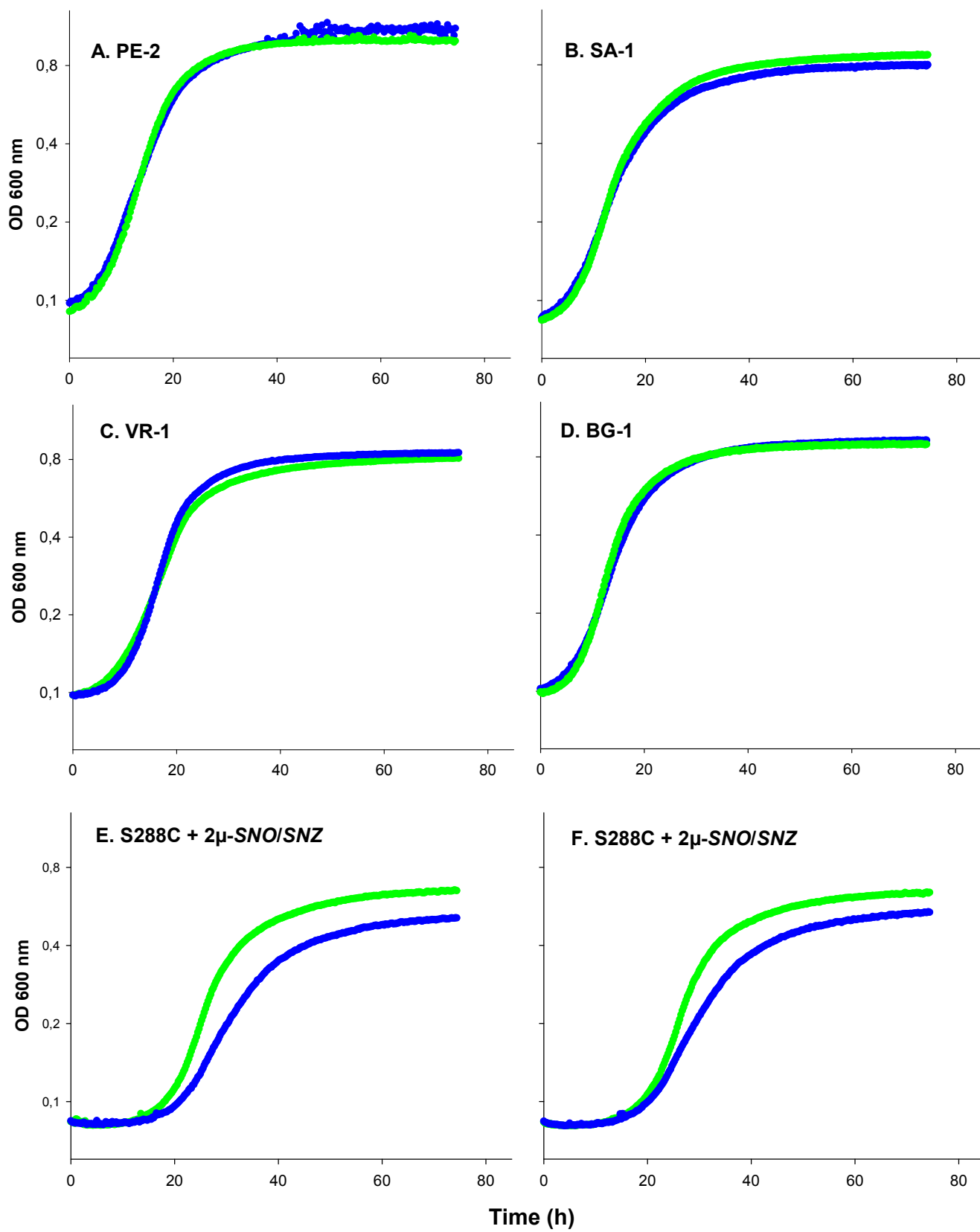


Figure S3

pBD1001- TTTGTACGTCACCTGCTTACTTCATGATATTGCAACAACAGATAAGAATA 50
ChrVI- TTTGTACGTCACCTGCTTACTTCATGATATTGCAACAACAGATAAGAATA 50
ChrXIV- TTTGTACGTCACCTGCTTACTTCATGATATTGCAACAACAGATAAGAATA 50
pBD1002- TTTGTACGTCACCTGCTTACTTCATGATATTGCAACAACAGATCAGAATA 50
pBD1003- TTTGTACGTCACCTGCTTACTTCATGATATTGCAACAACAGATAAGAATA 50

pBD1001- TGAGAGCCACGAAGATGTCATTTGAGTATTATGGTGGCATACTTTCAAGG 100
ChrVI- TGAGAGCCACGAAGATGTCATTTGAGTATTATGGTGGCATACTTTCAAGG 100
ChrXIV- TGAGAGCCACGAAGATGTCATTTGAGTATTATGGTGGCATACTTTCAAGG 100
pBD1002- TGAGAGCCACGAAGATGTCATTTGAGTATTATGGTGGCATACTTTCAAGG 100
pBD1003- TGAGAGCCACGAAGATGTCATTTGAGTATTATGGTGGCATACTTTCAAGG 100

pBD1001- GAGCTTGTATTTAATGCGACAGGTGGAAATCAGGACTATGCAGATGCAGT 150
ChrVI- GAGCTTGTATTTAATGCGACAGGTGGAAATCAGGACTATGCAGATGCAGT 150
ChrXIV- GAGCTTGTATTTAATGCGACAGGTGGAAATCAGGACTATGCAGATGCAGT 150
pBD1002- GAGCTTGTATTTAATGCGACAGGTGGAAATCAGGACTATGCAGATGCAGT 150
pBD1003- GAGCTTGTATTTAGTGCACAGGTGGAAATCAGGACTATGCAGATGCAGT 150

pBD1001- AACTGAGGCCATCATTCGTCACCAGGATTGACTGGGACTGGGTACATTA 200
ChrVI- AACTGAGGCCATCATTCGTCACCAGGATTGACTGGGACTGGGTACATTA 200
ChrXIV- AACTGAGGCCATCATTCGTCACCAGGATTGACTGGGACTGGGTACATTA 200
pBD1002- AACTGAGGCCATCATTCGTCACCAGGATTGACTGGGACTGGGTACATTA 200
pBD1003- AACTGAGGCCATCATTCGTCACCAGGATTGACTGGGACTGGGTACATTA 200

pBD1001- CCACCTTGGGGCTCATCTGCAGATTGCTACTACGCTTGACAATGTCGGA 250
ChrVI- CCACCTTGGGGCTCATCTGCAGATTGCTACTACGCTTGACAATGTCGGA 250
ChrXIV- CCACCTTGGGGCTCATCTGCAGATTGCTACTACGCTTGACAATGTCGGA 250
pBD1002- CCACCTTGGGGCTCATCTGCAGATTGCTACTACGCTTGACAATGTCGGA 250
pBD1003- CCGCCTTGGGGCTCATCTGCAGATTGCTACTACGCTTGACAATGTCGGA 250
** *****

pBD1001- TCCAATACCGATCTGATTTCATATCGATACAGTTAGTGCCATTAAACGAGCA 300
ChrVI- TCCAATACCGATCTGATTTCATATCGATACAGTTAGTGCCATTAAACGAGCA 300
ChrXIV- TCCAATACCGATCTGATTTCATATCGATACAGTTAGTGCCATTAAACGAGCA 300
pBD1002- TCCAATACCGATCTGATTTCATATCGATACAGTTAGTGCCATTAAACGAGCA 300
pBD1003- TCCAATACCGATCTGATTTCATATCGATACAGTTAGTGCCATTAAACGAGCA 300

pBD1001- ATTTCCAAGACTGCACTGGTTATCATGTTTTGCTACGGTGGTGGACACTG 350
ChrVI- ATTTCCAAGACTGCACTGGTTATCATGTTTTGCTACGGTGGTGGACACTG 350
ChrXIV- ATTTCCAAGACTGCACTGGTTATCATGTTTTGCTACGGTGGTGGACACTG 350
pBD1002- ATTTCCAAGACTGCACTGGTTATCATGTTTTGCTACGGTGGTGGACACTG 350
pBD1003- ATTTCCAAGACTGCACTGGTTATCATGTTTTGCTACGGTGGTGGACACTG 350

pBD1001- AAAACTCGAGAAAACCGTGGGGCCACACCAAGTTCTTTGGGTGATGATTTT 400
ChrVI- AAAACTCGAGAAAACCGTGGGGCCACACCAAGTTCTTTGGGTGATGATTTT 400
ChrXIV- AAAACTCGAGAAAACCGTGGGGCCACACCAAGTTCTTTGGGTGATGATTTT 400
pBD1002- AAAACTCGAGAAAACCGTGGGGCCACACCAAGTTCTTTGGGTGATGATTTT 400
pBD1003- AAAACTCGAGAAAACCGTGGGGCCACACCAAGTTCTTTGGGTGATGATTTT 400

pBD1001- TCAAAGAAAGTCATATGCAATACATTGGGTATAACTAAATAGGTAAGTT 450
ChrVI- TCAAAGAAAGTCATATGCAATACATTGGGTATAACTAAATAGGTAAGTT 450
ChrXIV- TCAAAGAAAGTCATATGCAATACATTGGGTATAACTAAATAGGTAAGTT 450
pBD1001- TCAAAGAAAGTCATATGCAATACATTGGGTATAACTAAATAGGTAAGTT 450
pBD1002- TCAAAGAAAGTCATATGCAATACATTGGGTATAACTAAATAGGTAAGTT 450
pBD1003- TCAAAGAAAGTCATATGCAATACATTGGGTATAACTAAATAGGTAAGTT 450

pBD1001- TCAATGAGGCGAGTTTCAAGATGAACAGTGATGATCTCCTTTCTCCATAAC 500
ChrVI- TCAATGAGGCGAGTTTCAAGATGAACAGTGATGATCTCCTTTCTCCATAAC 500
ChrXIV- TCAATGAGGCGAGTTTCAAGATGAACAGTGATGATCTCCTTTCTCCATAAC 500
pBD1002- TCAATGAGGCGAGTTTCAAGATGAACAGTGATGATCTCCTTTCTCCATAAC 500
pBD1003- TCAATGAGGCGAGTTTCAAGATGAACAGTGATGATCTCCTTTCTCCATAAC 500

pBD1001- TATACTATACGTCCCTTTTATTACTGTAGTTTTTAAGAACAATTTCTCT 550
ChrVI- TATACTATACGTCCCTTTTATTACTGTAGTTTTTAAGAACAATTTCTCT 550
ChrXIV- TATACTATACGTCCCTTTTATTACTGTAGTTTTTAAGAACAATTTCTCT 550
pBD1002- TATACTATACGTCCCTTTTATTACTGTAGTTTTTAAGAACAATTTCTCT 550
pBD1003- TATACTATACGTCCCTTTTATTACTGTAGTTTTTAAGAACAATTTCTCT 550

pBD1001- GATGAACCAAGTCGTGAAACCGTATATCGTTTTTCTGCCAATTCGGGATGAA 600
ChrVI- GATGAACCAAGTCGTGAAACCGTATATCGTTTTTCTGCCAATTCGGGATGAA 600
ChrXIV- GATGAACCAAGTCGTGAAACCGTATATCGTTTTTCTGCCAATTCGGGATGAA 600
pBD1002- GATGAACCAAGTCGTGAAACCGTATATCGTTTTTCTGCCAATTCGGGATGAA 600
pBD1003- GATGAACCAAGTCGTGAAACCGTATATCGTTTTTCTGCCAATTCGGGATGAA 600

pBD1001- ATGATGTCGCAAGAATATTGTTTTTTGCTTAGCGGCAACAATTAGTTCT 650
ChrVI- ATGATGTCGCAAGAATATTGTTTTTTGCTTAGCGGCAACAATTAGTTCT 650
ChrXIV- ATGATGTCGCAAGAATATTGTTTTTTGCTTAGCGGCAACAATTAGTTCT 650
pBD1002- ATGATGTCGCAAGAATATTGTTTTTTGCTTAGCGGCAACAATTAGTTCT 650
pBD1003- ATGATGTCGCAAGAATATTGTTTTTTGCTTAGCGGCAACAATTAGTTCT 650

pBD1001- TGACCACCATTATCCTTCCCATCTAATTTGTACAGGACCTGCACATGTTT 700
ChrVI- TGACCACCATTATCCTTCCCATCTAATTTGTACAGGACCTGCACATGTTT 700
ChrXIV- TGACCACCATTATCCTTCCCATCTAATTTGTACAGGACCTGCACATGTTT 700
pBD1002- TGACCACCATTATCCTTCCCATCTAATTTGTACAGGACCTGCACATGTTT 700
pBD1003- TGACCACCATTATCCTTCCCATCTAATTTGTACAGGACCTGCACATGTTT 700

pBD1001- AGGATCCAGCACCTCTTCTATTACTGGGGCTCTTATAAAAGTAGCAGGAA 750
ChrVI- AGGATCCAGCACCTCTTCTATTACTGGGGCTCTTATAAAAGTAGCAGGAA 750
ChrXIV- AGGATCCAGCACCTCTTCTATTACTGGGGCTCTTATAAAAGTAGCAGGAA 750
pBD1002- AGGATCCAGCACCTCTTCTATTACTGGGGCTCTTATAAAAGTAGCAGGAA 750
pBD1003- AGGATCCAGCACCTCTTCTATTACTGGGGCTCTTATAAAAGTAGCAGGAA 750

pBD1001- AATCATTGCAGTGAGGAATAAAGTTTGAAGTTCGCAATCCGGGTAGAA 800
ChrVI- AATCATTGCAGTGAGGAATAAAGTTTGAAGTTCGCAATCCGGGTAGAA 800
ChrXIV- AATCATTGCAGTGAGGAATAAAGTTTGAAGTTCGCAATCCGGGTAGAA 800
pBD1002- AATCATTGCAGTGAGGAATAAAGTTTGAAGTTCGCAATCCGGGTAGAA 800
pBD1003- AATCATTGCAGTGAGGAATAAAGTTTGAAGTTCGCAATCCGGGTAGAA 800

Figure S4

pBD1001- GACTGAGCTTGTCTCCCAAATGCATTTCTTTTACTTTAACCTTTAGTAA 850
 ChrVI- GACTGAGCTTGTCTCCCAAATGCATTTCTTTTACTTTAACCTTTAGTAA 850
 ChrXIV- GACTGAGCTTGTCTCCCAAATGCATTTCTTTTACTTTAACCTTTAGTAA 850
 pBD1002- GACTGAGCTTGTCTCCCAAATGCATTTCTTTTACTTTAACCTTTAGTAA 850
 pBD1003- GACTGAGCTTGTCTCCCAAATGCATTTCTTTTACTTTAACCTTTAGTAA 850

pBD1001- ATTCAGCGTCTTGACCAGTTTTGCTTCGTTAGATAAATGTTGTGAAATAT 900
 ChrVI- ATTCAGCGTCTTGACCAGTTTTGCTTCGTTAGATAAATGTTGTGAAATAT 900
 ChrXIV- ATTCAGCGTCTTGACCAGTTTTGCTTCGTTAGATAAATGTTGTGAAATAT 900
 pBD1002- ATTCAGCGTCTTGACCAGTTTTGCTTCGTTAGATAAATGTTGTGAAATAT 900
 pBD1003- ATTCAGCGTCTTGACCAGTTTTGCTTCGTTAGATAAATGTTGTGAAATAT 900

pBD1001- AAATCAAACCTGCACAAGTACCCAGGTTACCTTGCTTGGGTTGTGTACG 950
 ChrVI- AAATCAAACCTGCACAAGTACCCAGGTTACCTTGCTTGGGTTGTGTACG 950
 ChrXIV- AAATCATACCTGCACAAGTACCCAGGTTACCTTGCTTGGGTTGTGTACG 950
 pBD1002- AAATCATACCTGCACAAGTACCCAGGTTACCTTGCTTGGGTTGTGTACG 950
 pBD1003- AAATCATACCTGCACAAGTACCCAGGTTACCTTGCTTGGGTTGTGTACG 950

pBD1001- AATGCGTAGAGATCGTCGTAATAATCCTGTTCTTTCTGCAATAAGGGACAT 1000
 ChrVI- AATGCGTAGAGATCGTCGTAATAATCCTGTTCTTTCTGCAATAAGGGACAT 1000
 ChrXIV- AATGCGTAGAGATCGTCGTAATAATCCTGTTCTTTCTGCAATAAGGGACAT 1000
 pBD1002- AATGCGTAGAGATCGTCGTAATAATCCTGTTCTTTCTGCAATAAGGGACAT 1000
 pBD1003- AATGCGTAGAGATCGTCGTAATAATCCTGTTCTTTCTGCAATAAGGGACAT 1000

pBD1001- TGCAGTCGACTCTCCCCAGGTATGATCAATGCATCACATTGAGCTAGTT 1050
 ChrVI- TGCAGTCGACTCTCCCCAGGTATGATCAATGCATCACATTGAGCTAGTT 1050
 ChrXIV- TGCAGTCGACTCTCCCCAGGTATGATCAATGCATCACATTGAGCTAGTT 1050
 pBD1002- TGCAGTCGACTCTCCCCAGGTATGATCAATGCATCACATTGAGCTAGTT 1050
 pBD1003- TGCAGTCGACTCTCCCCAGGTATGATCAATGCATCACATTGAGCTAGTT 1050

pBD1001- GATTTTTATCCTTCACTGTCATCACAGATAGTTTTTTTTCATAGAAATCC 1100
 ChrVI- GATTTTTATCCTTCACTGTCATCACAGATAGTTTTTTTTCATAGAAATCC 1100
 ChrXIV- GATTTTTATCCTTCACTGTCATCACAGATAGTTTTTTTTCATAGAAATCC 1100
 pBD1002- GATTTTTATCCTTCACTGTCATCACAGATAGTTTTTTTTCATAGAAATCC 1100
 pBD1003- GATTTTTATCCTTCACTGTCATCACAGATAGTTTTTTTTCATAGAAATCC 1100

pBD1001- CTGTTTTTCGACGATGCATTTTTCTACGTGTCGCACATGTTCAATGAACGC 1150
 ChrVI- CTGTTTTTCGACGATGCATTTTTCTACGTGTCGCACATGTTCAATGAACGC 1150
 ChrXIV- CTGTTTTTCGACGATGCATTTTTCTACGTGTCGCACATGTTCAATGAACGC 1150
 pBD1002- CTGTTTTTCGACGATGCATTTTTCTACGTGTCGCACATGTTCAATGAACGC 1150
 pBD1003- CTGTTTTTCGACGATGCATTTTTCTACGTGTCGCACATGTTCAATGAACGC 1150

pBD1001- ACCCTGTAATGCCAAGACTCCGATAACGACGGTCATTTTTGAGATATTAC 1200
 ChrVI- ACCCTGTAATGCCAAGACTCCGATAACGACGGTCATTTTTGAGATATTAC 1200
 ChrXIV- ACCCTGTAATGCCAAGACTCCGATAACGACGGTCATTTTTGAGATATTAC 1200
 pBD1002- ACCCTGTAATGCCAAGACTCCGATAACGACGGTCATTTTTGAGATATTAC 1200
 pBD1003- ACCCTGTAATGCCAAGACTCCGATAACGACGGTCATTTTTGAGATATTAC 1200

pBD1001- TCTTGTCTTGGCCCCAGCAACCAGGGTCATGGTAAGTGTGCTTGCTACTC 1250
 ChrVI- TCTTGTCTTGGCCCCAGCAACCAGGGTCATGGTAAGTGTGCTTGCTACTC 1250
 ChrXIV- TCTTGTCTTGGCCCCAGCAACCAGGGTCATGGTAAGTGTGCTTGCTACTC 1250
 pBD1002- TCTTGTCTTGGCCCCAGCAACCAGGGTCATGGTAAGTGTGCTTGCTACTC 1250
 pBD1003- TCTTGTCTTGGCCCCAGCAACCAGGGTCATGGTAAGTGTGCTTGCTACTC 1250

pBD1001- AAACAGAAAAAATGGGCTTTTATCAATGAATAACTTTTTTTTTTTTTT 1300
 ChrVI- AAACAGAAAAAATGGGCTTTTATCAATGAATAACTTTTTTTTTTTTTT 1300
 ChrXIV- AAACAGAAAAAATGGGCTTTTATCAATGAATAACTTTTTTTTTTTT--TG 1297
 pBD1002- AAACAGAAAAAATGGGCTTTTATCAATGAATAACTTTTTTTTTTTT-TG 1299
 pBD1003- GAACAGAAAAAATGGGCTTTTATCAATGAGTAACCTTTTTTTTTTTT--TG 1298

pBD1001- TAGCTGAATGATGAAAGATGGCTCAACTATTCTTACAGTTTCCAATAAAG 1350
 ChrVI- TAGCTGAATGATGAAAGATGGCTCAACTATTCTTACAGTTTCCAATAAAG 1350
 ChrXIV- TAGCTGAATGATGAAAGATGGCTCAACTATTCTTACAGTTTCCAATAAAG 1347
 pBD1002- TAGCTGAATGATGAAAGATGGCTCAACTATTCTTACAGTTTCCAATAAAG 1349
 pBD1003- TAGCTGAATGATGAAAGATGGCTTAATTATTCTTACAGTTTCCAATAAAG 1348

pBD1001- TGACGCGCCGCCCAATGGCATAACAGTAGTAGGCACATTATATCGTTTGA 1400
 ChrVI- TGACGCGCCGCCCAATGGCATAACAGTAGTAGGCACATTATATCGTTTGA 1400
 ChrXIV- TGACGCGCCGCCCAATGGCATAACAGTAGTAGGCACATTATATCGTTTGA 1397
 pBD1002- TGACGCGCCGCCCAATGGCATAACAGTAGTAGGCACATTATATCGTTTGA 1399
 pBD1003- TGACGCGCCGCCCAATGGCATAACAGTAGTAGGCACATTATATCGTTTGA 1398

pBD1001- ATGGCAAAATGTTGTGGCAAGGAAAGAGCAGAATTATTTGTTTCATGTGGA 1450
 ChrVI- ATGGCAAAATGTTGTGGCAAGGAAAGAGCAGAATTATTTGTTTCATGTGGA 1450
 ChrXIV- ATGGCAAAATGTTGTGGCAAGGAAAGATCAGAATTATTTGTTTCATGTGGA 1447
 pBD1002- ATGGCAAAATGTTGTGGCAAGGAAAGATCAGAATTATTTGTTTCATGTGGA 1449
 pBD1003- ATGGCAAAATGTTGTGGCAAGGAAAGATCAGAATTATTTGTTTCATGTGGA 1448

pBD1001- ATGAATTT--CACACACACACACACATATATAAACTAGTATGTTTCT 1498
 ChrVI- ATGAATTT--CACACACACACACACATATATAAACTAGTATGTTTCT 1498
 ChrXIV- ATGAATTT--CACACACACACACACATATATAAACTAGTATGTTTCT 1495
 pBD1002- ATGAATTT--CACACACACACACACATATATAAACTAGTATGTTTCT 1497
 pBD1003- ATGAATTTAACACACACACACACATATATAAACTAGTATGCTTCT 1498

pBD1001- CCACCATAGTCACATTATTATTCGACCCTTCTGGGAAAGGAACATAATA 1548
 ChrVI- CCACCATAGTCACATTATTATTCGACCCTTCTGGGAAAGGAACATAATA 1548
 ChrXIV- CCACCATAGTCACATTATTATTCGACCCTTCTAGGAAAGGAACATAATA 1545
 pBD1002- CCACCATAGTCACATTATTATTCGACCCTTCTAGGAAAGGAACATAATA 1547
 pBD1003- CCACCATAGTCACATTATTATTCACCCTTCTGGGAAAGGAACATAATA 1548

pBD1001- GAAAAATAAGTATATCGTAAAAAGACAAAAATGTCAGAATTCAAGGTTA 1598
 ChrVI- GAAAAATAAGTATATCGTAAAAAGACAAAAATGTCAGAATTCAAGGTTA 1598
 ChrXIV- GAAAAATAAGTATATCGTAAAAAGACAAAAATGTCAGAATTCAAGGTTA 1595
 pBD1002- GAAAAATAAGTATATCGTAAAAAGACAAAAATGTCAGAATTCAAGGTTA 1597
 pBD1003- GAAAAATAAGTATATCGTAAAAAGACAAAAATGTCAGAATTCAAGGTTA 1598

Figure S4 (cont.)

pBD1001- AAACCTGGGCTTGCCCAAATGTTAAAGGGCGGTGTGATTATGGACGTCGTC 1648
 ChrVI- AAACCTGGGCTTGCCCAAATGTTAAAGGGCGGTGTGATTATGGACGTCGTC 1648
 ChrXIV- AAACCTGGGCTTGCCCAAATGTTAAAGGGCGGTGTGATTATGGACGTCGTC 1645
 pBD1002- AAACCTGGGCTTGCCCAAATGTTAAAGGGCGGTGTGATTATGGACGTCGTC 1647
 pBD1003- AAACCTGGGCTTGCCCAAATGTTAAAGGGCGGTGTGATTATGGACGTCGTC 1648

pBD1001- ACACCTGAACAGGCTATTATCGCAGAAAGAGCGGGCGCTTGTGCTGTAAT 1698
 ChrVI- ACACCTGAACAGGCTATTATCGCAGAAAGAGCGGGCGCTTGTGCTGTAAT 1698
 ChrXIV- ACACCTGAACAGGCTATTATCGCAGAAAGAGCGGGCGCTTGTGCTGTAAT 1695
 pBD1002- ACACCTGAACAGGCTATTATCGCAGAAAGAGCGGGCGCTTGTGCTGTAAT 1697
 pBD1003- ACACCTGAACAGGCTATTATCGCAGAAAGAGCGGGCGCTTGTGCTGTAAT 1698

pBD1001- GGCATTGGAACGCATTCCAGCTGACATGCGCAAGTCTGGCCAAGTATGCC 1748
 ChrVI- GGCATTGGAACGCATTCCAGCTGACATGCGCAAGTCTGGCCAAGTATGCC 1748
 ChrXIV- GGCATTGGAACGCATTCCAGCTGACATGCGCAAGTCTGGCCAAGTATGCC 1745
 pBD1002- GGCATTGGAACGCATTCCAGCTGACATGCGCAAGTCTGGCCAAGTATGCC 1747
 pBD1003- GGCATTGGAACGCATTCCAGCTGACATGCGCAAGTCTGGCCAAGTATGCC 1748

pBD1001- GTATGTCAGATCCTCGCATGATCAAAGAAATTTATGGAAGCTGTTTCAATT 1798
 ChrVI- GTATGTCAGATCCTCGCATGATCAAAGAAATTTATGGAAGCTGTTTCAATT 1798
 ChrXIV- GTATGTCAGATCCTCGCATGATCAAAGAAATTTATGGAAGCTGTTTCAATT 1795
 pBD1002- GTATGTCAGATCCTCGCATGATCAAAGAAATTTATGGAAGCTGTTTCAATT 1797
 pBD1003- GTATGTCAGATCCTCGCATGATCAAAGAAATTTATGGAAGCTGTTTCAATT 1798

pBD1001- CCGGTGATGGCAAAGGTCGGTATTGGACACTTCGTGGAGGCACAGATCCT 1848
 ChrVI- CCGGTGATGGCAAAGGTCGGTATTGGACACTTCGTGGAGGCACAGATCCT 1848
 ChrXIV- CCGGTGATGGCAAAGGTCGGTATTGGACACTTCGTGGAGGCACAGATCCT 1845
 pBD1002- CCGGTGATGGCAAAGGTCGGTATTGGACACTTCGTGGAGGCACAGATCCT 1847
 pBD1003- CCGGTGATGGCAAAGGTCGGTATTGGACACTTCGTGGAGGCACAGATCCT 1848

pBD1001- GGAAGAGCTGCAAGTAGACTACATTGACGAAAGCGAGGTTTGTACTCCAG 1898
 ChrVI- GGAAGAGCTGCAAGTAGACTACATTGACGAAAGCGAGGTTTGTACTCCAG 1898
 ChrXIV- GGAAGAGCTGCAAGTAGACTACATTGACGAAAGCGAGGTTTGTACTCCAG 1895
 pBD1002- GGAAGAGCTGCAAGTAGACTACATTGACGAAAGCGAGGTTTGTACTCCAG 1897
 pBD1003- GGAAGAGCTGCAAGTAGATTACATTGACGAAAGCGAGGTTTGTACTCCAG 1898

pBD1001- CTGATTGGACACATCACATTGAGAAGCATAAATTCAAGGTGCCATTGTGTT 1948
 ChrVI- CTGATTGGACACATCACATTGAGAAGCATAAATTCAAGGTGCCATTGTGTT 1948
 ChrXIV- CTGATTGGACACATCACATTGAGAAGCATAAATTCAAGGTGCCATTGTGTT 1945
 pBD1002- CTGATTGGACACATCACATTGAGAAGCATAAATTCAAGGTGCCATTGTGTT 1947
 pBD1003- CTGATTGGACACATCACATTGAGAAGCATAAATTCAAGGTGCCATTGTGTT 1948

pBD1001- TGCGGTGCCAAGGATCTAGGTGAGGCTTTGAGAAGAATAAACGAAGGTGC 1998
 ChrVI- TGCGGTGCCAAGGATCTAGGTGAGGCTTTGAGAAGAATAAACGAAGGTGC 1998
 ChrXIV- TGCGGTGCCAAGGATCTAGGTGAGGCTTTGAGAAGAATAAACGAAGGTGC 1995
 pBD1002- TGCGGTGCCAAGGATCTAGGTGAGGCTTTGAGAAGAATAAACGAAGGTGC 1997
 pBD1003- TGCGGTGCCAAGGATCTAGGTGAGGCTTTGAGAAGAATAAACGAAGGTGC 1998

pBD1001- TGCAATGATCCGTACCAAAGGTGAAGCAGGTACCGGTGACGTTTCCGAGG 2048
 ChrVI- TGCAATGATCCGTACCAAAGGTGAAGCAGGTACCGGTGACGTTTCCGAGG 2048
 ChrXIV- TGCAATGATCCGTACCAAAGGTGAAGCAGGTACCGGTGACGTTTCCGAGG 2045
 pBD1002- TGCAATGATCCGTACCAAAGGTGAAGCAGGTACCGGTGACGTTTCCGAGG 2047
 pBD1003- TGCAATGATCCGTACCAAAGGTGAAGCAGGTACCGGTGACGTTTCCGAGG 2048

pBD1001- CCGTCAAGCACATCACCAGATTAAGGCGGAGATCCAGCAGTATAAAGAG 2098
 ChrVI- CCGTCAAGCACATCACCAGATTAAGGCGGAGATCCAGCAGTATAAAGAG 2098
 ChrXIV- CCGTCAAGCACATCACCAGATTAAGGCGGAGATCCAGCAGTATAAAGAG 2095
 pBD1002- CCGTCAAGCACACACCAAGATTAAGGCGGAGATCCAGCAGTATAAAGAG 2097
 pBD1003- CCGTCAAGCACATCACCAGATTAAGACGCGAGATCCAGCAGTATAAAGAG 2098

pBD1001- AATTTGAAGACCGAATCCGATTTTGCAGCTAAGGCCACAGAATTACGCGT 2148
 ChrVI- AATTTGAAGACCGAATCCGATTTTGCAGCTAAGGCCACAGAATTACGCGT 2148
 ChrXIV- AATTTGAAGACCGAATCCGATTTTGCAGCTAAGGCCACAGAATTACGCGT 2145
 pBD1002- AATTTGAAGACCGAATCCGATTTTGCAGCTAAGGCCACAGAATTACGCGT 2147
 pBD1003- AATTTGAAGACCGAATCCGATTTTGCAGCTAAGGCCACAGAATTACGCGT 2148

pBD1001- CCCTGTCGACTTGCTGAAGACAACACTATCAGAGGGAAGCTACCTGTAG 2198
 ChrVI- CCCTGTCGACTTGCTGAAGACAACACTATCAGAGGGAAGCTACCTGTAG 2198
 ChrXIV- CCCTGTCGACTTGCTGAAGACAACACTATCAGAGGGAAGCTACCTGTAG 2195
 pBD1002- CCCTGTCGACTTGCTGAAGACAACACTATCAGAGGGAAGCTACCTGTAG 2197
 pBD1003- CCCTGTCGACTTGCTGAAGACAACACTATCAGAGGGAAGCTACCTGTAG 2198

pBD1001- TCAATTTTGCTGCTGGTGGAGTTGCTACTCCAGCAGACGCTGCTCTATTG 2248
 ChrVI- TCAATTTTGCTGCTGGTGGAGTTGCTACTCCAGCAGACGCTGCTCTATTG 2248
 ChrXIV- TCAATTTTGCTGCTGGTGGAGTTGCTACTCCAGCAGACGCTGCTCTATTG 2245
 pBD1002- TCAATTTTGCTGCTGGTGGAGTTGCTACTCCAGCAGACGCTGCTCTATTG 2247
 pBD1003- TCAATTTTGCTGCTGGTGGAGTTGCTACTCCAGCAGACGCTGCTCTATTG 2248

pBD1001- ATGCAATTTGGGTTGTGAAGGTGTTTTCGTGCGGTCAGGTATATTCAAATC 2298
 ChrVI- ATGCAATTTGGGTTGTGAAGGTGTTTTCGTGCGGTCAGGTATATTCAAATC 2298
 ChrXIV- ATGCAATTTGGGTTGTGAAGGTGTTTTCGTGCGGTCAGGTATATTCAAATC 2295
 pBD1002- ATGCAATTTGGGTTGTGAAGGTGTTTTCGTGCGGTCAGGTATATTCAAATC 2297
 pBD1003- ATGCAATTTGGGTTGTGAAGGTGTTTTCGTGCGGTCAGGTATATTCAAATC 2298

pBD1001- GTCAGATCCTGAGAAGTTAGCATGTGCTATTGTTGAAGCCACGACTCACT 2348
 ChrVI- GTCAGATCCTGAGAAGTTAGCATGTGCTATCGTTGAAGCCACGACTCACT 2348
 ChrXIV- GTCAGATCCTGAGAAGTTAGCATGTGCTATTGTTGAAGCCACGACTCACT 2345
 pBD1002- GTCAGATCCTGAGAAGTTAGCATGTGCTATCGTTGAAGCCACGACTCACT 2347
 pBD1003- GTCAGATCCTGAGAAGTTAGCATGTGCTATTGTTGAAGCCACGACTCACT 2348

pBD1001- ACGATAATCCAGCAAAACTGTTGCAAAATTTCCAGCGATTGGGTGACTTG 2398
 ChrVI- ACGATAATCCAGCAAAACTGTTGCAAGTTTCCAGCGATTGGGTGACTTG 2398
 ChrXIV- ACGATAATCCAGCAAAACTGTTGCAAAATTTCCAGCGATTGGGTGACTTG 2395
 pBD1002- ACGATAATCCAGCAAAACTGTTGCAAGTTTCCAGCGATTGGGTGACTTG 2397
 pBD1003- ACGATAATCCAGCAAAACTGTTGCAAGTTTCCAGCGATTGGGTGACTTG 2398

Figure S4 (cont.)

```

pBD1001- ATGGGTGGTATTTCCATCCAATCAATTAATGAAGCAGGAGGCAAAAACGG 2448
ChrVI- ATGGGTGGTATTTCCATCCAATCAATTAATGAAGCAGGAGGCAAAAACGG 2448
ChrXIV- ATGGGTGGTATTTCCATCCAATCAATTAATGAAGCAGGAGGCAAAAACGG 2445
pBD1002- ATGGGTGGTATTTCCATCCAATCAATTAATGAAGCAGGAGGCAAAAACGG 2447
pBD1003- ATGGGTGGTATTTCCATCCAATCAATTAATGAAGCAGGAGGCAAAAACGG 2448
*****

pBD1001- TGCAAGACTTTCTGAAATCGGATGGTAGAACTAGCGATGCTCACAACAG 2498
ChrVI- TGCAAGACTTTCTGAAATCGGATGGTAGAACTAGCGATGCTCACAACAG 2498
ChrXIV- TGCAAGACTTTCTGAAATCGGATGGTAGAACTAGCGATGCTCACAACAG 2495
pBD1002- TGCAAGACTTTCTGAAATCGGATGGTAGAACTAGCGATGCTCACAACAG 2497
pBD1003- TGCAAGACTTTCTGAAATCGGATGGTAGAACTAGCGATGCTCACAACAG 2498
*****

pBD1001- CTAATTTGTTTCCTTCGAAGGGGCAGTCGGCTATTGTAGTTTCTATATAT 2548
ChrVI- CTAATTTGTTTCCTTCGAAGGGGCAGTCGGCTATTGTAGTTTCTATATAT 2548
ChrXIV- CTAATTTGTTTCCTTCGAAGGGGCAGTCGGCTATTGTAGTTTCTATATAT 2545
pBD1002- CTAATTTGTTTCCTTCGAAGGGGCAGTCGGCTATTGTAGTTTCTATATAT 2547
pBD1003- CTAATTTGTTTCCTTCGAAGGGGCAGTCGGCTATTGTAGTTTCTATATAT 2548
*****

pBD1001- TATGTAGCACAAACATGGAAACCATATTCAGCCAGTTTGTATATATGAGAA 2598
ChrVI- TATGTAGCACAAACATGGAAACCATATTCAGCCAGTTTGTATATATGAGAA 2598
ChrXIV- TATGTAGCACAAACATGGAAACCATATTCAGCCAGTTTGTATATATGAGAA 2595
pBD1002- TATGTAGCACAAACATGGAAACCATATTCAGCCAGTTTGTATATATGAGAA 2597
pBD1003- TATGTAGCACAAACATGGAAACCATATTCAGCCAGTTTGTATATATGAGAA 2598
*****

pBD1001- TTAAACGTTAAAACTTTTTTCATCCTATCAGTCCTACACATTGGCCCTTGCC 2648
ChrVI- TTAAACGTTAAAACTTTTTTCATCCTATCAGTCCTTACACATTGGCCCTTGCC 2648
ChrXIV- TTAAACGTTAAAACTTTTTTCATCCTATCAGTCCTTACACATTGGCCCTTGCC- 2644
pBD1002- TTAAACGTTAAAACTTTTTTCATCCTATCAGTCCTTACACATTGGCCCTTGCC 2647
pBD1003- TTAAACGTTAAAACTTTTTTCATCCTATCAGTCCTTACACATTGGCCCTTGCC 2648
*****

pBD1001- TAAAGGGCACAAATCGTTCACGTATATACATATGTATGTTTACGTATATAT 2698
ChrVI- TAAAGGGCACAAATCGTTCACGTATATACATATGTATGTTTACGTATATAT 2698
ChrXIV- TAAAGGGCACAAATCGTTCACGTATATACATATGTATGTTTACGTATATAT 2694
pBD1002- TAAAGGGCACAAATCGTTCACGTATATACATATGTATGTTTACGTATATAT 2697
pBD1003- TAAAGGGCACAAATCGTTCACGTATATACATATGTATGTTTACGTATATAT 2698
**** *****

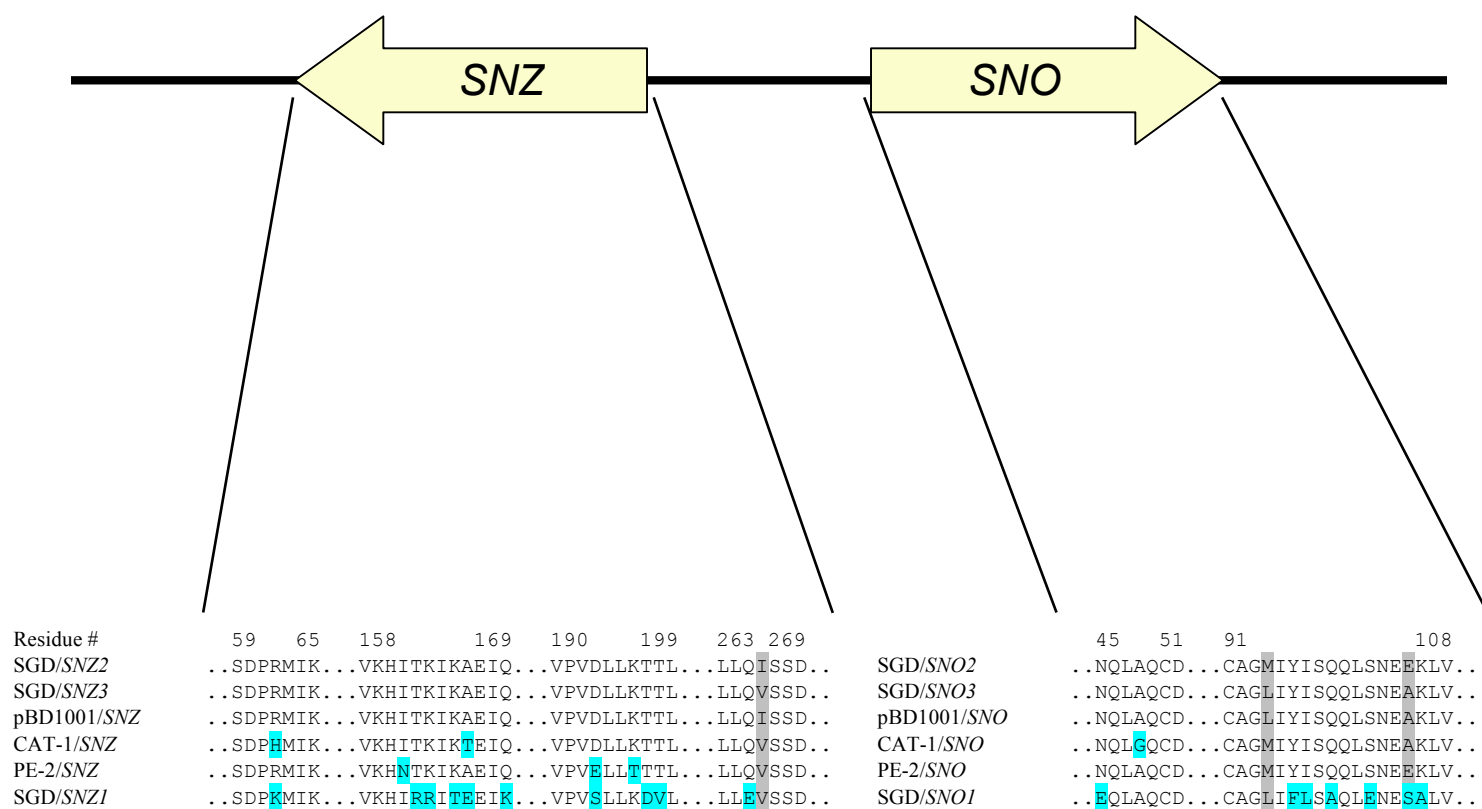
pBD1001- ATAT--GTGTTTACAAATATGCATGTATATCTTTTCGAGGAAGATTATT 2746
ChrVI- ATAT--GTGTTTACAAATATGCATGTATATCTTTTCGAGGAAGATTATT 2746
ChrXIV- ATAT--GTGTTTACAAATATGCATGTATATCTTTTCGAGGAAGATTATT 2742
pBD1002- ATATATGTGTTTACATATATGCATGTATATCTTTTCGAGGAAGATTATT 2747
pBD1003- ATAT--GTGTTTACAAATATGCATGTATATCTTTTCGAGGAAGATTATT 2746
**** *****

pBD1001- ATATGCTATTAGGCGATGAGTCGGCTTAATTTTTTAGGGGGACGAGATTG 2796
ChrVI- ATATGCTATTAGGCGATGAGTCGGCTTAATTTTTTAGGGGGACGAGATTG 2796
ChrXIV- ATATGCTATTAGGCGATGAGTCGGCTTAATTTTTTAGGGGGACGAGATTG 2792
pBD1002- ATATGCTATTAGGCGATGAGTCGGCTTAATTTTTTAGGGGGACGAGATTG 2797
pBD1003- ATATGCTATTAGGCGATGAGTCGGCTTAATTTTTTAGGGGGACGAGATTG 2796
*****

```

Figure S4 (cont.)

A



B

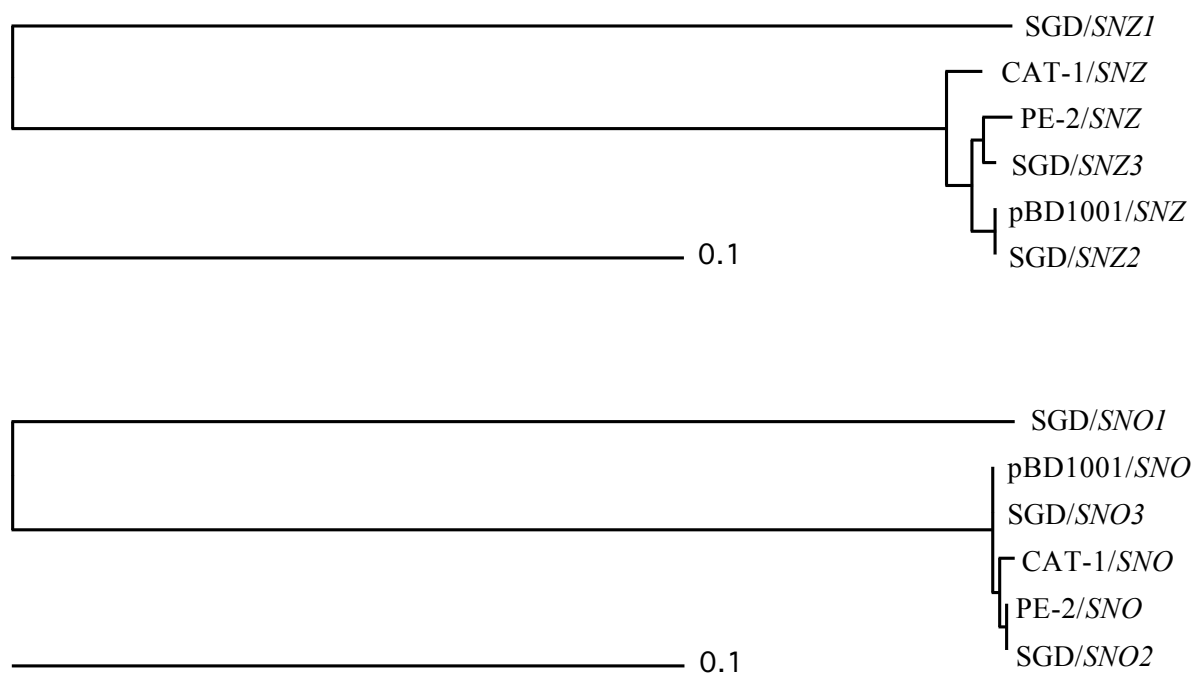


Figure S5