

Table S1. *B. pseudofirmus* OF4 mobile element related gene content^a

Chromosome			pBpOF4-1			pBpOF4-2		
Locus Tag	Type	IS (family/group)	Locus Tag	Type	IS (family/group)	Locus Tag	Type	IS (family/group)
00230	R,Ig		19959	T	IS116/IS110/IS902	21449	T,Ig	IS116/IS110/IS902
00875	T		20034	T	IS200 family	21554	T	
00880	T		20039	T	IS605 family	21579	Ig	IS650-01
01380	T		20049	T,Ig		21604	T	IS4/IS641-02
01385	T		20054	T,Ig		21729	R,Ig	
01465	T		20059	T		21794	R	
01470	T		20139	R,Ig		21799	R,Ig	
01760	En		20144	R,Ig		21819	T	
02705	T	IS653-01	20174	T		21824	R,Iv,Res	
03680	T,Ig		20179	T	IS204/IS1001/IS1096/IS1165	21974	En	
03690	T,Ig	IS116/IS110/IS902	20209	T	IS4/IS641-02	21989	R,Iv	
03730	En		20219	In,RT	group II intron	22024	R,Iv	
03755	T	IS4/IS641-02	20224	RT,Ret	Bh.Int-03 (group II intron)			
03770	R,Ig		20229	RT,Ret	Bh.Int-03 (group II intron)			
03775	R,Ig		20244	T,Ig	IS116/IS110/IS902			
04275	Res		20274	T				
04310	En,Res		20524	T	IS643-02			
04420	R,Iv,Res		20529	T	IS643-02			
04550	R		20539	R,Ig				
04620	T	IS4/IS641-02	20659	RT,Ret	group II intron			
05240	En		20684	RT,M,In,Ret	Bh.Int-03 (group II intron)			
06065	T		20709	T	IS643-02			
06070	T		20714	T	IS643-02			
06175	En		20849	RT,Ret	group II intron			
06390	Res		20889	T				
06405	T		20949	T	IS605 family			
06430	T,R,Iv		20954	T	IS200 family			
06455	T		20989	T	IS116/IS110/IS902			
06485	T		20994	T	IS116/IS110/IS902			
06495	T		21004	T	IS116/IS110/IS902			
06500	T		21164	T				
07430	En	IS66	21339	R,Ig				

07525	T		21424	T	IS116/IS110/IS902
07530	T				
08130	En				
08855	T				
08860	T				
09820	T	IS653-01			
09825	T	IS653-01			
10075	T	IS4/IS641-02			
10640	T				
10805	Ig	ISBma2			
11875	En,Res				
11935	T				
11940	T				
12295	Ig				
13035	T	IS4/IS641-02			
13535	T	IS653-01			
14295	T				
14300	T				
14750	R,Ig				
15680	T	IS653-01			
16060	R,Ig				
16155	En				
16590	RT	group II intron			
17130	T,Ig	IS116/IS110/IS902			
17245	T				
19435	R,Iv				
19490	Ig				

a. Annotations from PGAAP were searched for Transposons (T), Recombinases (R), Reverse Transcriptases (RT), Maturases (M), Endonucleases (En), Invertases (Iv), Integrases (Ig), Introns (In), Retrons (Ret) and Resolvases (Res). For the chromosome, there were 59 total, representing 1.5% of the total ORFS; for pBpOF4-1, there were 33 total, or 11.3% of total ORFS; and for pBpOF4-2, there were 12 total, or 9.8% of total ORFS.

Table S2. Sigma factors in the *Bacillus pseudofirmus* OF4 chromosome^a

<i>A. Growth, motility, general stress</i>	Accession #	Locus Tag
1. Sig70	YP_003426643	BpOF4_08465
2. Sig70 (rpoS/rpoD domains)	YP_003428371	BpOF4_17190
3. SigD	YP_003424974	BpOF4_00065
4. SigD2	YP_003426183	BpOF4_06145
5. SigD3	YP_003427689	BpOF4_09205
6. SigB	YP_003426789	BpOF4_09205
<i>B. ECF</i>		
7. SigC	YP_003427382	BpOF4_12185
8. SigC2	YP_003428211	BpOF4_16370
9. SigL	YP_003426015	BpOF4_05305
10. SigM	YP_003428452	BpOF4_17595
11. Sig V	YP_003428217	BpOF4_16400
12. SigW	YP_003426710	BpOF4_08800
13. SigX	YP_003427643	BpOF4_13500
14. SigY	YP_003427988	BpOF4_15235
15. SigZ	YP_003427375	BpOF4_12150
16. Sig54 (Fis)	YP_003426988	BpOF4_10215
17. Sig54 (Fis)	YP_003426213	BpOF4_06305
<i>C. Sporulation</i>		
18. SigE	YP_003425086	BpOF4_00625
19. SigF	YP_003427908	BpOF4_14835
20. SigG	YP_003425085	BpOF4_00620
21. SigK	YP_003427542	BpOF4_12995

a. Assignments are based on BLASTP analyses; the first listed Sig70 is the closest homologue of Sig70 from *B. subtilis*

Table S3. Two-component signaling systems of *B. pseudofirmus* OF4

	Accession # [YP_00342..]	Locus tag [BpOF4_..]	Protein	Predicted Function
Chromosome				
1. HK ^a	5408	02245	LytS	Cell autolysis
RR	5407	02240	LytR	
2. HK	5442	02415	AtoS	Acetolactate metabolism
RR	5441	02410	AtoC	
3. HK	5464	02525	CheA	Chemotaxis
RR	5462	02515	CheY	
4. HK	5628	03350	PhoP	Phosphate regulon
RR	5627	03345	PhoR	
5. HK	5817	04305	YhcZ	Nitrate respiration
RR	5816	04300	YhcY	
6. HK	6057	05515	DcuS	Dicarboxylate metabolism
RR	6058	05520	DcuR	
7. HK	6094	05700	DegS	Degradative enzyme reg.
RR	6093	05695	DegU	
8. HK ^b	6490	07700	KinE	Entry into stationary phase
9. HK	6515	07825	WalK	Peptidoglycan formation
RR	6516	07830	WalR	
10. HK	6828	09405	CitS	Citrate metabolism
RR	6827	09400	CitT	
11. HK	6943	09980	BceS	Cell wall antibiotic sensing
RR	6944	09985	BceR	
12. HK	6953	10030	DesK	Phospholipid desaturation
RR	6954	10035	DesR	
13. HK	6961	10070	---	----
RR	6960	10065	---	
14. HK ^b	7140	10975	KinE	Entry into stationary phase
15. HK	7193	11240	YkoH	----
RR	7192	11235	YkoG	
16. HK	7230	11425	CesS	Secretion stress
RR	7229	11420	CesR	
17. HK	7257	11560	LiaS	Cell wall stress
RR	7258	11565	LiaR	
18. HK ^b	7306	11805	KinA	Sporulation
19. HK ^b	7363	12090	---	GGDEF domain
20. HK	7410	12325	---	---
RR	7409	12320	---	
21. HK	7441	12490	YufL	Citrate/Malate metabolism
RR	7440	12485	YufM	
22. HK	7537	12970	YycF	Cell wall stress
RR	7538	12975	YvrG	
23. HK ^b	7549	13030	---	---
24. HK	7551	13040	---	Autolysin regulation
RR	7552	13045	---	
25. HK ^b	7700	13785	---	---
26. HK	7781	14190	YclJ	---
RR	7782	14195	YclK	---
27. HK ^b	7788	14235	---	---
28. HK	7848	14535	---	---
RR	7847	14530	---	
29. HK	7954	15065	ResE	Respiration
RR	7953	15060	ResD	
30. HK ^b	8549	18080	NtrB	Nitrogen metabolism
31. HK ^b	8550	18085	KinA	Sporulation
32. HK	8558	18125	DcuS	Citrate/malate
RR	8557	18120	DcuR	
33. HK	8629	18480	---	---
RR	8630	18485	---	
Plasmid pBpOF4-01				
34. HK	8949	20299	----	---
RR	8950	20304	---	

35. HK	8982	20474	ResE	Respiration
RR	8983	20479	ResD	

a. HK=sensor histidine kinase; RR=response regulator; **b.** No RR identified, **c.** --- indicates no proposed protein and/or functional designation.

Table S4. Chemotaxis signaling sensors of *B. pseudofirmus* OF4^a

Protein Name	Uniprot #	Tag	Amino acid (aa)	TMs	Motif	Homologue of <i>B. halodurans</i> C-125
<i>A. Mcp and TarH</i>						
1. TarH1	D3G0X0	BpOF4_19984	660	2	Cache_1, marine_sort_HK, HAMP, MA	Q9K617(BH3915), Identity 33.0% Q9KFC0 (BH0565), Identity 36% (271aa)
2. TarH2	D3FZJ3	BpOF4_05865	659	2	Cache_2, marine_ HAMP, MA	Q9KFC0 (BH0565), Identity 31% (385aa)
3. TarH3	D3FZ89	BpOF4_14545	493	5	MA, TarH	Q9KCR0 (BH1509), Identity 47.0%
4. TarH4	D3G0G8	BpOF4_06925	431	2	HAMP, MA, TarH	Q9K6G1 (<i>mcpA</i> , BH3768), Identity 56.0%
5. TarH5	D3FQ85	BpOF4_07495	579	2	Cache_1, HAMP, Tar, MA	Q9KAL2 (BH2275), Identity 38.0% Q9KBU7 (<i>mcpC</i> , BH1827), Identity 34.0%
6. TarH6	D3FW63	BpOF4_02640	580	2-3	MA, TarH	Q9KEH6 (BH0876), Identity 29.0%(367aa)
7. TarH7	D3FYY8	BpOF4_04785	556	2	HAMP, MA, TarH	no hit
8. TarH8	D3FUJ9	BpOF4_10580	565	2	CHASE3, Ma, Tar, NtrY, NarQ, COG5278	Q9KEH6 (BH0876), Identity 47.0% Q9K632 (BH3900), Identity 33.0% Q9K669 (BH3863), Identity 31.0%
9. TarH9	D3FXE5	BpOF4_13025	424	2	HAMP, MA, TarH	Q9K669 (BH3863), Identity 39.0% (381aa)
10. TarH10	D3G0H7	BpOF4_06970	583	2	Cache_2, MA, TarH	no hit
11. TarH11	D3FXF1	BpOF4_13055	536	2	MAMP, Tar	no hit
12. TarH12	D3FTM6	BpOF4_10230	278	0	Tar, MA	Q9KE52 (BH1006), Identity 29.0%
<i>B. HemAT</i>						
13. HemAT1	D3FSA7	BpOF4_18505	434	0	sensor_globin, heme-binding site, TarH	Q9KFH4 (BH0505), Identity 45.0% (403aa)
14. HemAT2	D3FW31	BpOF4_12050	439	0	sensor_globin, heme-binding site, TarH	Q9KFH4 (BH0505), Identity 33.0%
15. HemAT3	D3FWD3	BpOF4_02990	434	0	sensor_globin, heme-binding site, TarH	Q9KFH4 (BH0505), Identity 25.0%

a. Assignments are based on BLASTp analyses; MCP: Methyl-accepting chemotaxis protein, TarH: Taxis toward Aspartate and Related amino acids and Homologs, HeamAT: heme-based aerotactic transducer. *Bacillus halodurans* C-125 has 11 Mcp-type and one HemAT-type proteins.

Legends Supplementary Figures

Fig. S1. Predicted secondary structures of the tmRNA and associated group I intron in the chromosome of *B. pseudofirmus* OF4. The structures were predicted by alignment with the homologous sequences in *Clostridium botulinum* (Williams, 2002)) followed by manual rearrangements. Nucleotides that are conserved in both sequence and structural position in *B. pseudofirmus* and *C. botulinum* are colored in red. The various subdomains of the tmRNA (P1-P12) and group I intron (P1-P9) are indicated. In addition, for the tmRNA, the regions forming pseudoknots are labeled (Ψ 1 and Ψ 4) the nucleotides encoding the proteolysis tag are underlined; for the intron, flanking exon sequences are in lower case letters.

Fig. S2. Schematic representation of the 4.5-kb group II twintron in the pBpOF4-01 plasmid of *B. pseudofirmus* OF4. The twintron consists of two nested group II intron retroelements, where the innermost intron (in grey) encoding the reverse-transcriptase (RT) BpOF4_20224 is inserted at the catalytic site of the RT (BpOF4_20229+BpOF4_20219) encoded by the outermost intron (in black). The catalytic site corresponds to the RYADD motif at the amino-acid level. The introns are delimited by arrows. Genes are drawn as boxes containing the gene names, with genes flanking the twintron drawn as open boxes.

Fig. S3. Multiple sequence alignment of the *bcr3* repeat elements from the chromosome of *B. pseudofirmus* OF4. *bcr3* is flanked by a 25-nucleotide direct repeat (DR) motif which is boxed. In addition, *bcr3* copies form two groups of sequences, *bcr3a* and *bcr3b*, that are homologous to the *bcr3* elements originally identified in bacteria from the *Bacillus cereus* group (Kristoffersen *et al.*, 2011). A *bcr3a* and a *bcr3b* copy from the *B. cereus* group (*bcr3a*_NC_009674_5F-II and *bcr3b*_NZ_ABDK02000047_1R, respectively) are included in the alignment for comparison. A plot of the conservation of each site was generated by CLUSTALX (Larkin *et al.*, 2007) and is shown below the alignment. Asterisks above the alignment indicate sites where all sequences have an identical nucleotide.

Fig. S4. Multiple sequence alignment of the *bpr1* repeat elements from the chromosome of *B. pseudofirmus* OF4. *bpr1* is made up of a CGCAATAA/CGCATAAA/CGCAATTA motif, which is boxed,

and that is repeated seven times. A plot of the conservation of each site was generated by CLUSTALX (Larkin et al., 2007) and is shown below the alignment.

Fig. S5. Multiple sequence alignment of the *bpr2* repeat elements from the chromosome of *B. pseudofirmus* OF4. *bpr2* is made up of two complementary regions, the left and right end, respectively, which are boxed. Partial copies corresponding either to the left or to the right end are also included in the alignment and their names start with the letter P. A plot of the conservation of each site was generated by CLUSTALX (Larkin et al., 2007) and is shown below the alignment.

Fig. S6. Top: Filled arrows indicate the location of perfect match *dnaA* boxes (TTATCCACA). Open arrows indicate *dnaA* boxes with a maximum of 1 mismatch. Bottom: nucleotide-level comparison of the AT-rich region in *B. subtilis* 168, *B. halodurans* C-125 and *B. pseudofirmus* OF4 that is unwound by DnaA during initiation of replication.

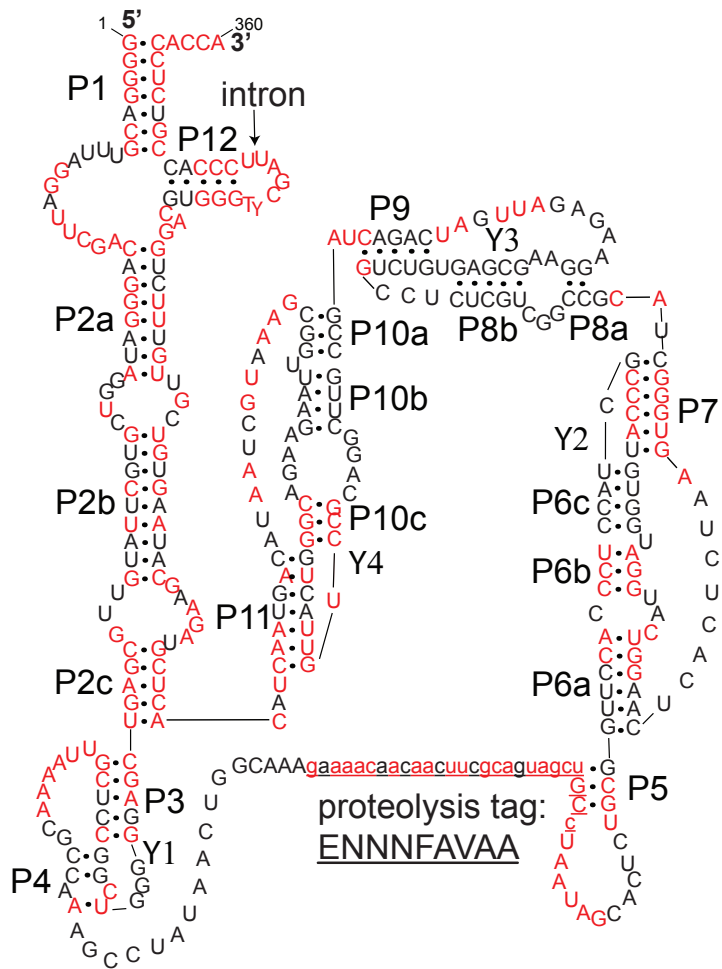
Fig. S7. Comparison of the *B. pseudofirmus* OF4 chromosome with *B. halodurans* C-125 at the gene level. Each dot represents predicted protein homology with a *B. halodurans* C-125 gene that has greater than 50% identity (the top hit). Dots lying on the X-axis represent <50% identity or no match.

Kristoffersen, S. M., N. J. Tourasse, A. B. Kolstø & O. A. Økstad, (2011) Interspersed DNA repeats *bcr1-bcr18* of *Bacillus cereus* group bacteria form three distinct groups with different evolutionary and functional patterns. *Mol. Biol. Evol.* **28**: 963-983.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettingan, H. McWilliam, P. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & O. G. Higgins, (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.

Williams, K. P., (2002) The tmRNA website: invasion by an intron. *Nucleic Acids Res*, **30**: 179-182.

A) tmRNA



B) group I intron

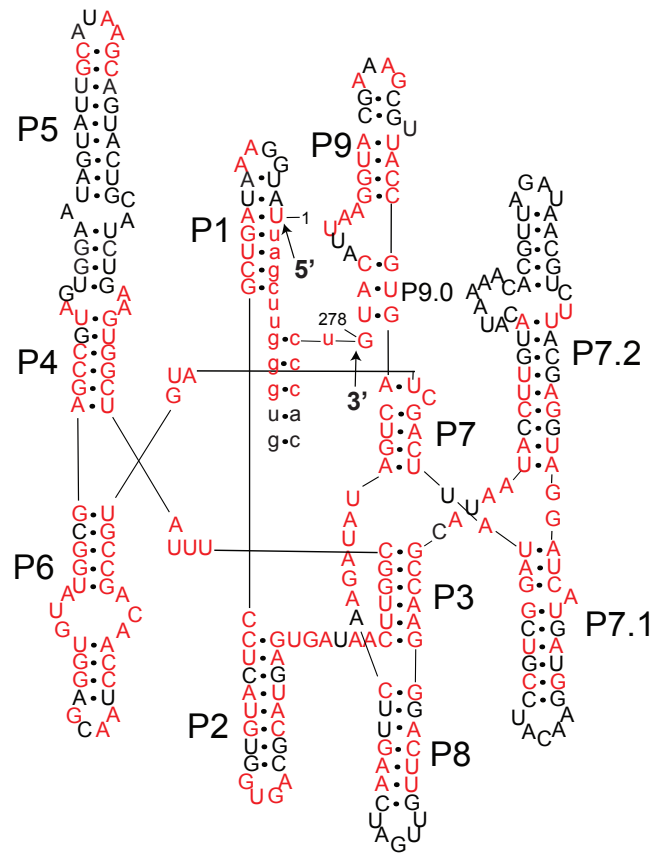


Fig. S1

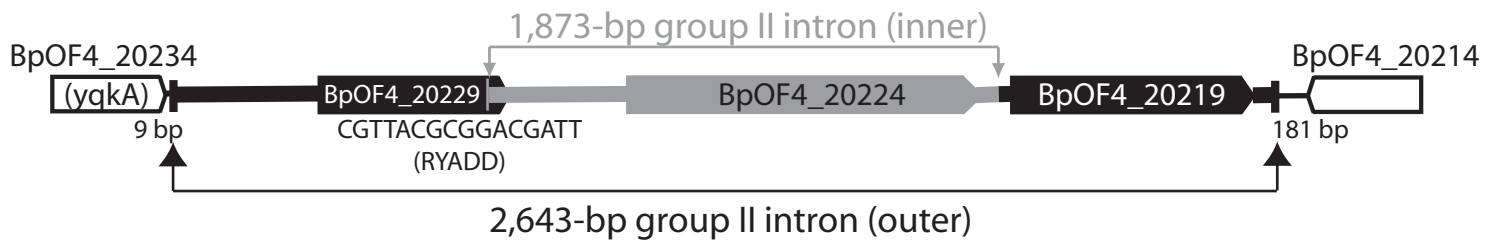


Fig. S2

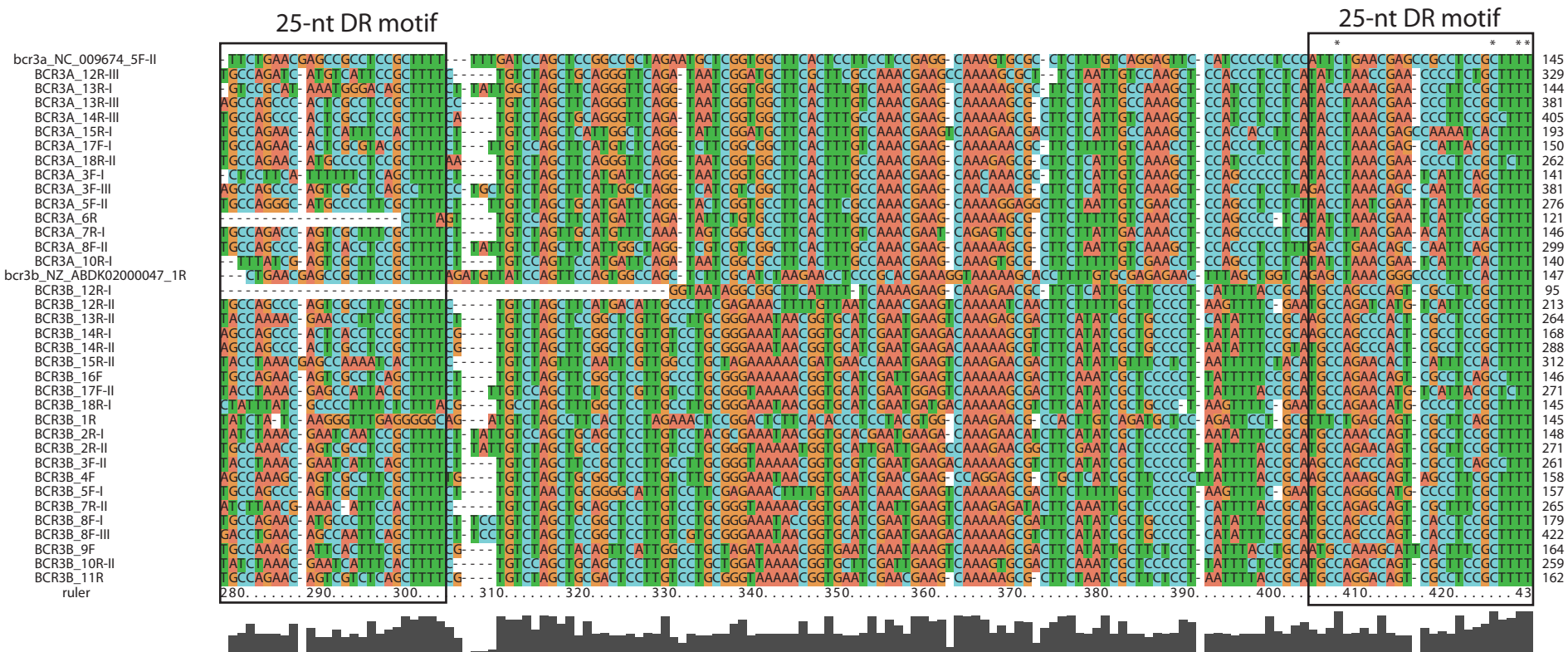


Fig. S3

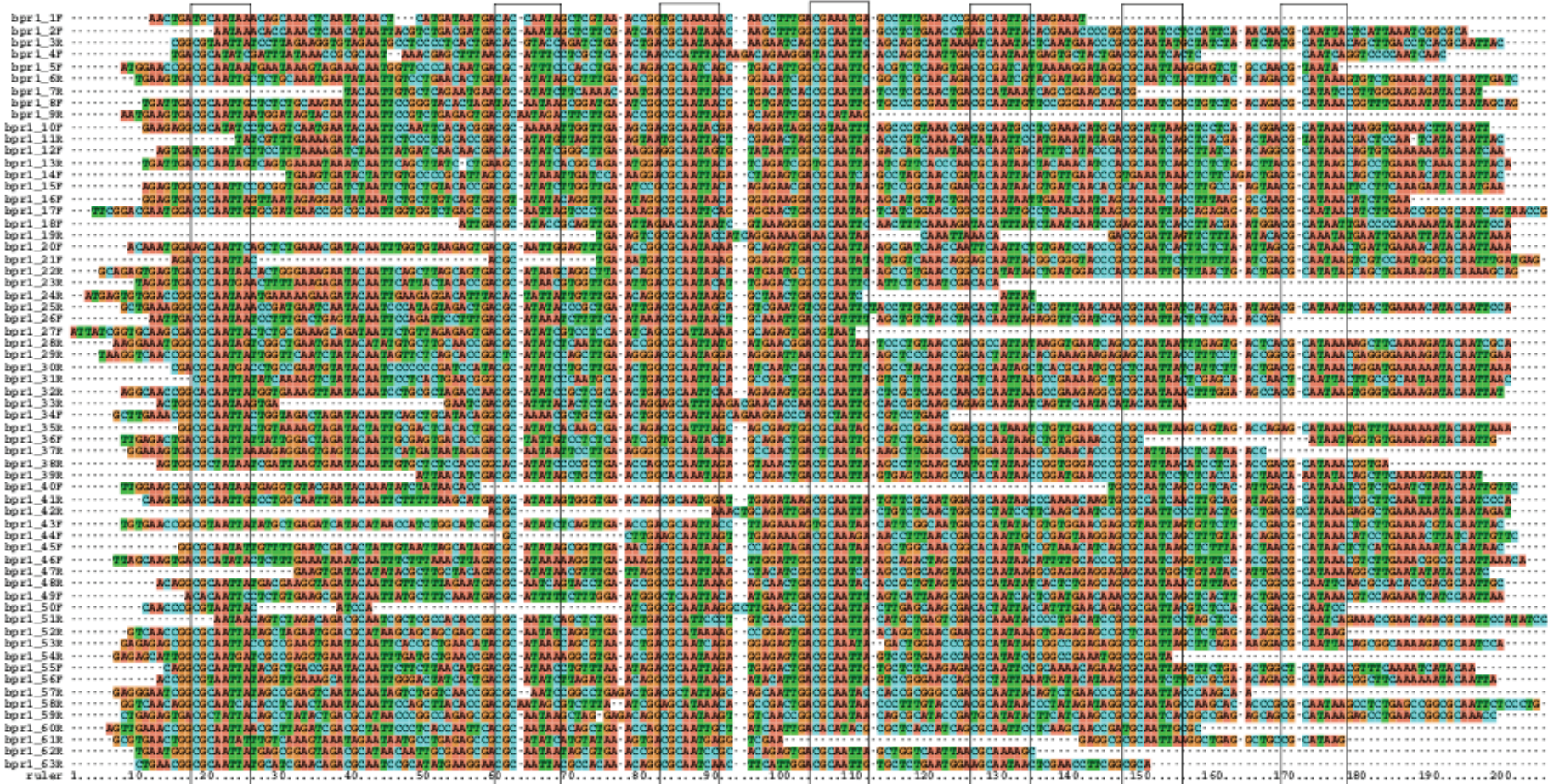


Fig. S4

Left End



Fig.S5



Fig.S5

Right End



Fig.S5

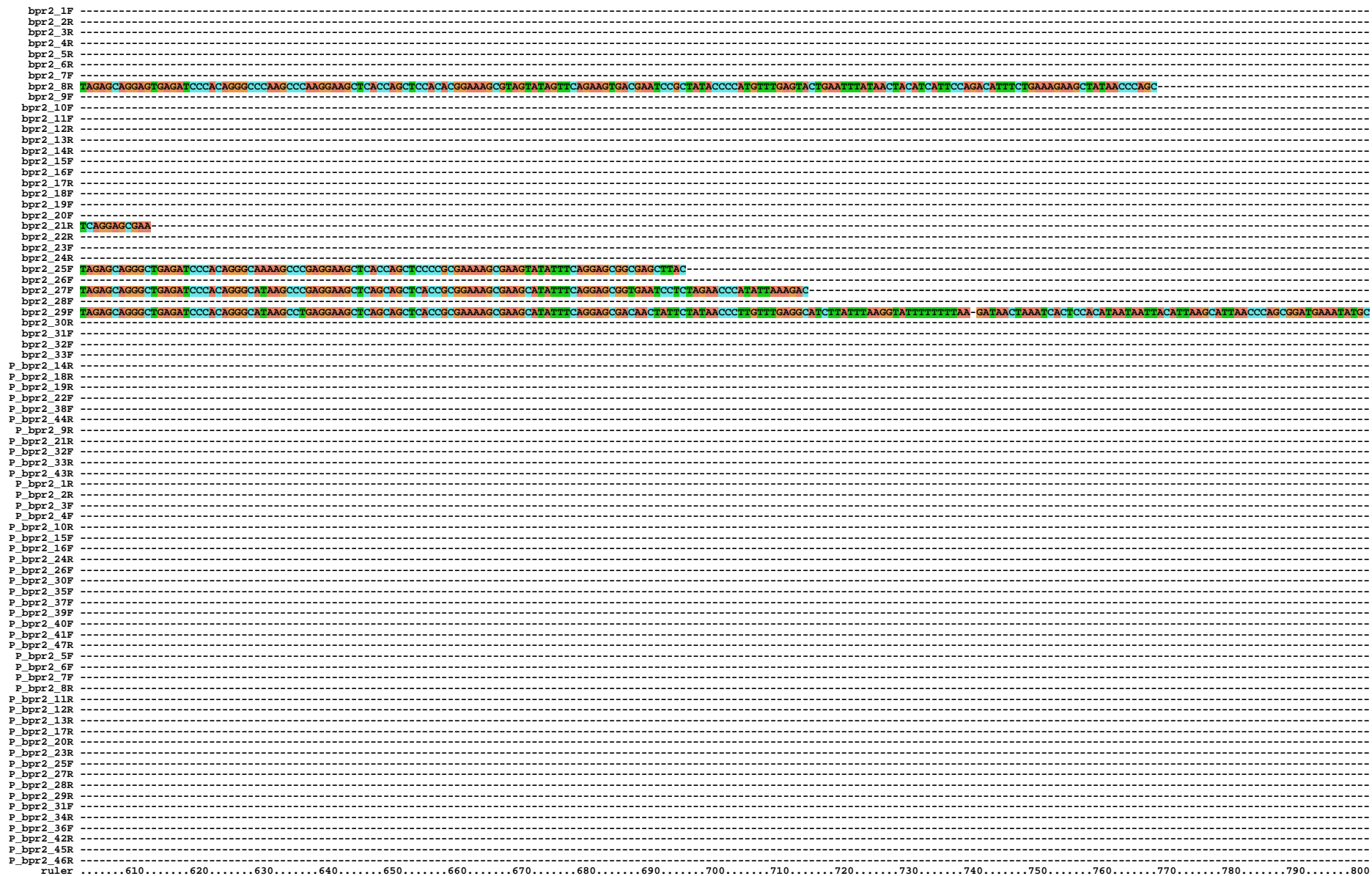


Fig.S5

bpr2_1F -----
 bpr2_2R -----
 bpr2_3R -----
 bpr2_4R -----
 bpr2_5R -----
 bpr2_6R -----
 bpr2_7F -----
 bpr2_8R -----
 bpr2_9F -----
 bpr2_10F -----
 bpr2_11F -----
 bpr2_12R -----
 bpr2_13R -----
 bpr2_14R -----
 bpr2_15F -----
 bpr2_16F -----
 bpr2_17R -----
 bpr2_18F -----
 bpr2_19F -----
 bpr2_20F -----
 bpr2_21R -----
 bpr2_22R -----
 bpr2_23F -----
 bpr2_24R -----
 bpr2_25F -----
 bpr2_26F -----
 bpr2_27F -----
 bpr2_28F -----
 bpr2_29F **GAAGACCTTCGGTGCCTAGAGCAGGGCTGAGATCCCA CAGGGCTAAAGCCCGAGGAAGCTCAC TAGCTCCCCACGGAAAGCGAAGCATATTCAGGAGCGGCAGTCCGCTATTACCCATGTTTA**
 bpr2_30R -----
 bpr2_31F -----
 bpr2_32F -----
 bpr2_33F -----
 P_bpr2_14R -----
 P_bpr2_18R -----
 P_bpr2_19R -----
 P_bpr2_22F -----
 P_bpr2_38F -----
 P_bpr2_44R -----
 P_bpr2_9R -----
 P_bpr2_21R -----
 P_bpr2_32F -----
 P_bpr2_33R -----
 P_bpr2_43R -----
 P_bpr2_1R -----
 P_bpr2_2R -----
 P_bpr2_3F -----
 P_bpr2_4F -----
 P_bpr2_10R -----
 P_bpr2_15F -----
 P_bpr2_16F -----
 P_bpr2_24R -----
 P_bpr2_26F -----
 P_bpr2_30F -----
 P_bpr2_35F -----
 P_bpr2_37F -----
 P_bpr2_39F -----
 P_bpr2_40F -----
 P_bpr2_41F -----
 P_bpr2_47R -----
 P_bpr2_5F -----
 P_bpr2_6F -----
 P_bpr2_7F -----
 P_bpr2_8R -----
 P_bpr2_11R -----
 P_bpr2_12R -----
 P_bpr2_13R -----
 P_bpr2_17R -----
 P_bpr2_20R -----
 P_bpr2_23R -----
 P_bpr2_25F -----
 P_bpr2_27R -----
 P_bpr2_28R -----
 P_bpr2_29R -----
 P_bpr2_31F -----
 P_bpr2_34R -----
 P_bpr2_36F -----
 P_bpr2_42R -----
 P_bpr2_45R -----
 P_bpr2_46R -----
 ruler810.....820.....830.....840.....850.....860.....870.....880.....890.....900.....910.....920.....

Fig.S5

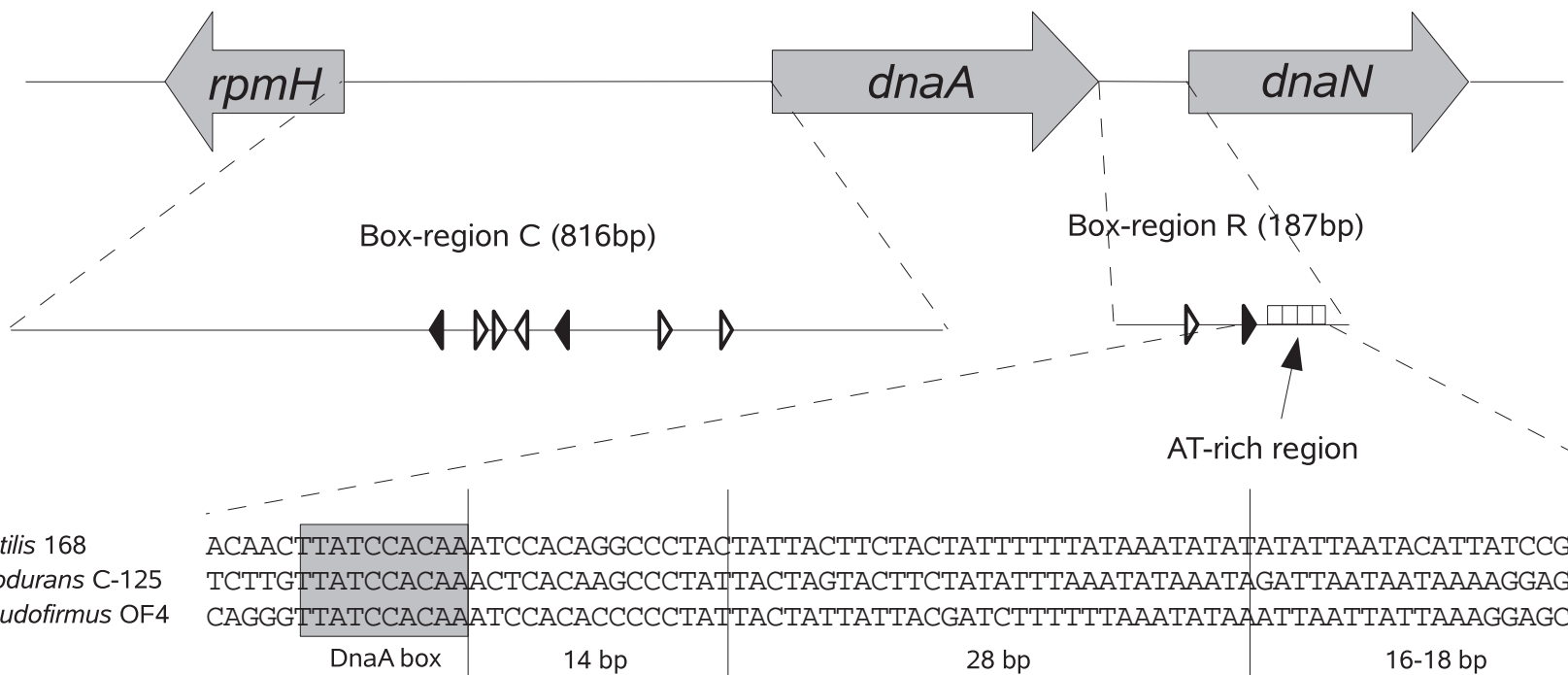


Fig. S6

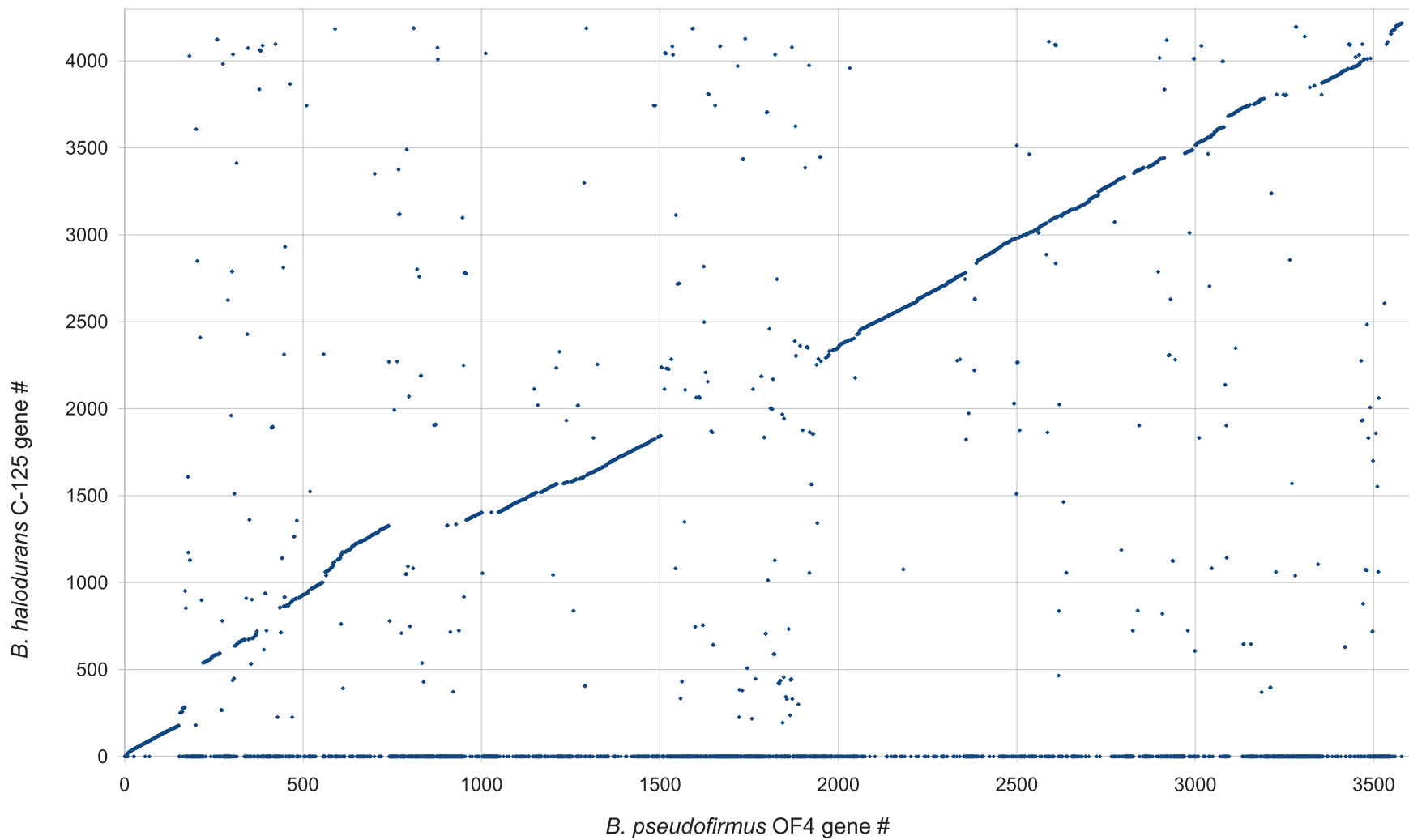


Fig. S7