

RESEARCH ARTICLE

Open Access



De novo assembly of genomes from long sequence reads reveals uncharted territories of *Propionibacterium freudenreichii*

Paulina Deptula^{1*}, Pia K. Laine², Richard J. Roberts³, Olli-Pekka Smolander², Helena Vihinen², Vieno Piironen¹, Lars Paulin², Eija Jokitalo², Kirsi Savijoki¹, Petri Auvinen² and Pekka Varmanen^{1*} 

Abstract

Background: *Propionibacterium freudenreichii* is an industrially important bacterium granted the Generally Recognized as Safe (the GRAS) status, due to its long safe use in food bioprocesses. Despite the recognized role in the food industry and in the production of vitamin B12, as well as its documented health-promoting potential, *P. freudenreichii* remained poorly characterised at the genomic level. At present, only three complete genome sequences are available for the species.

Results: We used the PacBio RS II sequencing platform to generate complete genomes of 20 *P. freudenreichii* strains and compared them in detail. Comparative analyses revealed both sequence conservation and genome organisational diversity among the strains. Assembly from long reads resulted in the discovery of additional circular elements: two putative conjugative plasmids and three active, lysogenic bacteriophages. It also permitted characterisation of the CRISPR-Cas systems. The use of the PacBio sequencing platform allowed identification of DNA modifications, which in turn allowed characterisation of the restriction-modification systems together with their recognition motifs. The observed genomic differences suggested strain variation in surface piliation and specific mucus binding, which were validated by experimental studies. The phenotypic characterisation displayed large diversity between the strains in ability to utilise a range of carbohydrates, to grow at unfavourable conditions and to form a biofilm.

Conclusion: The complete genome sequencing allowed detailed characterisation of the industrially important species, *P. freudenreichii* by facilitating the discovery of previously unknown features. The results presented here lay a solid foundation for future genetic and functional genomic investigations of this actinobacterial species.

Keywords: *Propionibacterium freudenreichii*, Comparative genomics, Bacteriophage, Pilus, Vitamin B12, PacBio, Complete genome, Mobile elements, Restriction and modification, CRISPR-Cas, Gras

* Correspondence: paulina.deptula@helsinki.fi; pekka.varmanen@helsinki.fi

¹Department of Food and Environmental Sciences, University of Helsinki, 00014 Helsinki, Finland

Full list of author information is available at the end of the article



Background

Propionibacteria belong to the phylum Actinobacteria with high GC content (64–70%) genomes. They have a peculiar metabolism [1], characterised by the utilization of lactate and the production of propionate, acetate and carbon dioxide through the Wood-Werkmann cycle [2]. *Propionibacterium freudenreichii* is an industrially important species with Generally Recognized as Safe (GRAS) status, granted due to its long, safe use in dairy fermentations. *P. freudenreichii* is used as a secondary starter culture in the production of Swiss-type cheeses where it plays a crucial role in the formation of “eyes” by CO₂ production and the development of the typical flavour attributed to lipolysis, release of amino acids, especially proline, and to the production of short-chain fatty acids (SCFAs): propionate and acetate [3]. Due to their antimicrobial activity, propionate or strains of *Propionibacterium* species are commonly used as food and grain preservatives to prolong the shelf-life of many products by suppressing the growth of mold and spoilage microorganisms [4]. SCFAs are among the most abundant dietary metabolites produced by the gut microbes during dietary fermentation [5] with implications in e.g. controlling inflammatory responses and appetite [6–8]. Notably, the SCFAs produced by *P. freudenreichii* as well as milk fermented with this species were recently shown to specifically induce apoptosis of colon cancer cells, thereby opening new avenues for microbial-based therapies [9]. In addition to SCFAs, *P. freudenreichii* produce a wide variety of compounds with implications for human health and well-being, like conjugated linoleic acid [10], vitamins [11–14], exopolysaccharides [15] and trehalose [16], and have thus potential application as cell factories for natural enrichment of food with nutraceuticals. There is an increasing amount of evidence that strains of *P. freudenreichii* and other dairy propionibacteria have probiotic properties (reviewed recently [17]). In clinical studies concerning probiotic activity, *P. freudenreichii* strains have mainly been used as components of complex bacterial mixtures and rarely as monocultures [17]. The first step required for a probiotic to interact with a host and produce any particular response is adhesion to mucus bound to gastrointestinal epithelia [18]. While *P. freudenreichii* strains have revealed only weak and nonspecific adhesion to the mucus, the adhesion was increased by the presence of other probiotic bacteria [19, 20].

Despite the recognized role of *P. freudenreichii* in the food industry, its capability to produce appreciable amounts of active vitamin B12 and short-chain fatty acids as well as its well-documented probiotic potential, the bacterium remained poorly characterised on the genetic and genomic level. The first genome sequence was announced only in 2010 [21] shedding light on the

crucial characteristics of *P. freudenreichii* such as its unique metabolism, its hardiness and probiotic potential. In addition, some misconceptions about the species have been brought to light, for example the presence of all the genes necessary for aerobic respiration led to questioning of the anaerobic status of the species. Also, it was discovered that the features used for subdivision of the species into subspecies *shermanii* and *freudenreichii*, namely lactose utilisation and nitroreductase activity, result from acquisition through horizontal gene transfer and loss due to a frameshift, respectively. This led to questioning the validity of the subdivision [21], which was proven not warranted [22]. Sequencing projects resulted in 22 draft genomes [23, 24] and two additional complete genomes [25, 26] available for the species. Although the draft genomes proved valuable and were used in a number of comparative and functional studies [24, 27–29], they do not permit studies of genome organization or mobile elements absent from the reference genome [30]. In addition, due to the nature of the short-read sequencing itself, draft genomes do not give an insight into additional regions of sequences rich in repeats such as CRISPR-Cas systems, transposed mobile elements or gene duplications [31].

Here, we report complete genome sequences of 17 additional *P. freudenreichii* strains and a re-sequenced whole genome of the strain DSM 4902. Additionally, we performed a comparative genomics study of the 20 whole genomes available to date and, owing to the long sequence reads produced by the PacBio platform, we identified several thus far unknown features of these bacteria. We report the highly variable genome organization of the strains sharing high level of sequence identity, in addition to two putative conjugative plasmids and three active temperate phages discovered as circular molecules. Genome data mining revealed complete CRISPR-Cas systems, novel restriction-modification systems, complete pili operons, the presence of putative Integrative and Conjugative Elements (ICEs) and the active transposable elements, potentially playing an important role in species adaptation.

Results

Among the studied strains were the 14 strains from the collection of dairy company Valio Ltd., four isolated from barley grains by the malting company Polttimo Ltd., and two type strains originating from Swiss cheese (Table 1). Eighteen of the strains were sequenced with a PacBio RSII instrument, followed by assembly using Hierarchical Genome Assembly Process (HGAP3) in SMRT Analysis software (Table 2). The two remaining strains: the type strain JS16 (DSM 20271, CP010341) and JS (LN997841) were published before [25, 26]. The other type strain, JS15 (DSM 4902), has been sequenced

Table 1 *P. freudenreichii* strains included in this study. Summary of the genome sequences

Strain	Sequence name	Accession number	Genome size (bp)	Genome coverage	GCC%	No. of predicted genes	Note
JS	PFREUDJS001	LN997841	2,675,045	691	67%	2382	Sequenced previously [26]; 261 ^a
JS2	PFRJS2	LT576032	2,655,351	317	67%	2310	257 ^a
JS4	PFRJS4	LT576033	2,654,663	158	67%	2366	259 ^a
JS7	PFRJS7-1	LT618776	2,738,418	300	67%	2412	Genome with prophage; 263 ^a
	PFRJS7-2	LT618777	2,700,482	50	67%	2355	Genome without phage
	PFRJS7-ph	LT618778	37,936	60	65%	59	Circular phage
JS8	PFRJS8	LT576042	2,655,373	392	67%	2324	264 ^a
JS9	PFRJS9-1	LT618785	2,720,049	219	67%	2405	Additional transposase gene; 265 ^a
	PFRJS9-2	LT618786	2,718,592	219	67%	2401	
JS10	PFRJS10	LT576035	2,626,110	208	67%	2329	266 ^a
JS11	PFRJS11	LT576038	2,537,402	501	67%	2200	274 ^a
JS12	PFRJS12-1	LT604998	2,615,181	250	67%	2275	Additional transposase gene; 275 ^a
	PFRJS12-2	LT576787	2,613,734	250	67%	2274	
	PFRJS12-3	LT604882	24,909	400	64%	32	Putative conjugative plasmid
JS13	PFRJS13-1	LT618779	2,537,370	210	67%	2201	11-gene insertion; 276 ^a
	PFRJS13-2	LT618780	2,520,651	90	67%	2189	
JS14	PFRJS14	LT593929	2,507,188	330	67%	2180	277 ^a
JS15	PFRJS15-1	LT618787	2,621,081	94	67%	2322	3 additional transposase genes; 281 ^a
	PFRJS15-2	LT618788	2,616,005	94	67%	2320	Transposase gene disrupting a Type III RM methylase
JS16	RM25	CP010341	2,649,163	190	67%	2321	Sequenced previously [25]; 282 ^a
JS17	PFRJS17-1	LT618789	2,755,516	192	67%	2455	Duplicated transposase gene; 283 ^a
	PFRJS17-2	LT618790	2,754,069	192	67%	2454	
JS18	PFRJS18	LT576034	2,661,974	190	67%	2358	284 ^a
JS20	PFRJS20-1	LT618791	2,678,207	106	67%	2384	286 ^a
	PFRJS20-2	LT618792	2,682,327	106	67%	2376	3 additional transposase genes; one disrupting another transposase gene
JS21	PFRJS21-1	LT618781	2,659,993	222	67%	2330	Additional transposase genes; 287 ^a
	PFRJS21-2	LT618782	2,658,550	222	67%	2329	
JS22	PFRJS22-1	LT599498	2,633,661	190	67%	2326	Genome with prophage; 288 ^a
	PFRJS22-ph	LT615138	39,309	102	66%	61	Circular phage
JS23	PFRJS-23	LT618793	2,630,698	210	67%	2335	Genome with prophage; 289 ^a
	PFRJS-23-ph	LT618794	42,723	25	65%	66	Circular phage
JS25	PFRJS25-1	LT618783	2,666,517	400	67%	2336	291 ^a
	PFRJS25-2 pl	LT618784	35,640	800	64%	46	Putative conjugative plasmid

^a Strain number in the previous study [48]

previously [21], but it was re-sequenced with PacBio for this study.

We assembled 31 complete and circular sequences from the eighteen strains. For 11 of the strains, assembly resulted in more than one genome. In five of the strains circular elements were found: in JS12 and JS25 putative conjugative plasmids and in JS7, JS22 and JS23 bacteriophage genomes. In eight of the strains, the additional

genomes resulted from duplication and relocation (copy and paste) of transposable elements (Table 3).

Genome organization

Average Nucleotide Identity (ANI) calculated through pairwise BLAST alignments revealed that genomes of the *P. freudenreichii* strains are highly collinear, with an ANI value of nearly 99% on average (Fig. 1a). The whole

Table 2 Sequencing summary

Sample Name	Sequencing Chemistry	SMRTcells	Movie Time (min) ^a	Total Number of Subreads	Total Number of Bases	Mean Subread Length (bp)	N50 Subread Length (bp) ^b
JS4	P4/C2	2	120	142,306	465,129,578	3268	4131
JS10	P4/C2	2	120	205,113	606,802,054	2958	3733
JS15	P4/C2	2	120	179,642	540,206,310	3007	3653
JS18	P4/C2	2	120	194,246	550,199,736	2832	3679
JS20	P4/C2	2	120	230,371	632,635,668	2746	3456
JS22	P4/C2	2	120	181,361	563,541,807	4090	3107
JS23	P4/C2	2	120	192,373	617,310,336	3208	4222
JS2	P5/C3	2	240	100,864	719,516,424	7133	8940
JS7	P5/C3	2	240	107,194	1,002,978,417	9356	13,043
JS8	P5/C3	2	240	139,418	1,149,552,255	8245	11,148
JS9	P5/C3	2	240	160,808	1,316,084,462	8184	11,023
JS11	P5/C3	2	240	184,744	1,422,323,823	7698	10,498
JS12	P5/C3	2	240	207,391	1,490,421,212	7186	9765
JS13	P5/C3	2	240	103,466	793,823,369	7672	10,156
JS14	P5/C3	2	240	125,130	966,521,587	7724	10,326
JS17	P5/C3	2	240	151,002	1,149,295,071	7611	10,232
JS21	P5/C3	2	240	173,843	1,298,367,718	7468	9906
JS25	P5/C3	2	240	192,104	1,293,994,542	6735	8805

^a run time during which nucleotides added by the polymerase are recorded in real-time

^b minimum length of subreads in which half of the sequencing data is found

genome alignments show that despite genome-wide colinearity, large regions of inversions and other types of re-organisations are present even among the most closely related strains (Fig. 1b).

In eight of the sequenced strains we observed translocation of mobile elements, either with transposase genes alone or, in strain JS13, as a part of larger gene cluster. The gene cluster consists of 12 coding sequences: four transposase genes and eight hypothetical proteins, one with similarity to “Helicase conserved C-terminal domain” (PF00271.25). Additionally, we observed a transposase-mediated duplication in strain JS17, which was confirmed by PCR to eliminate the possibility of an assembly error. The duplication spans 35 genes: PFR_JS17-1_676-PFR_JS17-1_710 and PFR_JS17-1_711-PFR_JS17-1_745, located between genes coding for an Uma4 type transposase and an aspartate ammonia lyase. The duplication region included genes coding for, among others, thiamine biosynthetic proteins, transporters and glycerol metabolism.

Comparative genomics

The pangenome of the 20 *P. freudenreichii* strains was analysed with Roary [32] revealing 4606 ortholog groups. The core genome, defined as ortholog groups found in all of the isolates, consisted of 1636 orthologs. The soft core, ortholog groups found in 19 out of 20 isolates, consisted of 80 additional orthologs, while the 1251 ortholog groups

found in three to 18 strains made up the shell genome. The remaining 1639 ortholog groups were assigned to the cloud genome consisting of the ortholog groups which were found in either one or two strains only (Fig. 2).

The numbers of accessory genes in individual strains and numbers of unique genes varied between genomes (Fig. 3). To better visualise the differences between genomes a presence-absence matrix was created from the orthologs assigned to accessory genome (Fig. 4). The strains are organised into a phylogenetic tree based on the accessory genome alignments. The unique gene clusters accounting for the most obvious differences between genomes are highlighted (more detailed results in Additional file 1). The core genome size needs to be addressed with caution, as out of 1636 genes 457 differed in predicted size among the strains, 200 of which differed by a minimum of 90 nucleotides (see Additional file 1). The frequent co-localisation of such genes with the genes coding for short hypothetical proteins may be indicative of evolutionary events, which resulted in splitting of the coding sequence, mis-annotation or sequencing errors.

To characterise the individual genomes further, bioinformatics analyses were performed, including searches for prophages, genomic islands, CRISPR-Cas systems and restriction- modification (RM) systems. The cumulative results are summarised in Fig. 5 and the details

Table 3 *P. freudenreichii* strains included in this study. The details of differences between the genome sequences within strains

Strain	Accession number	GC%	Note	Details
JS7	LT618776	67%	Genome with prophage	PFR_JS7-1_1810 HTH-type transcriptional regulator KmtR;PFR_JS7-1_1869 Transcriptional regulator MtrR (preceded by tRNA-Lys)
	LT618777	67%	Genome without phage	
JS9	LT618785	67%	Genome with an additional transposase gene	PFR_JS9-1_62 Uma4 protein
	LT618786	67%		
JS12	LT604998	67%	Genome with an additional transposase gene	PFR_JS12-1_615 ^a Transposase of ISAar20, ISL3 family and PFR_JS12-1_616 Hypothetical protein (ahead of Carbon starvation protein)
	LT576787	67%		
JS13	LT618779	67%	Genome with 11-gene insertion or deletion ^b	JS13_289 Hypothetical protein: JS13_299 Transposase for insertion sequence element IS1001
	LT618780	67%		
JS15	LT618787	67%	3 transposase genes absent from the other genome	PFR_JS15-1_878 Uma4 protein and PFR_JS15-1_879 Hypothetical protein; PFR_JS15-1_1737 Transposase of ISAar20, ISL3 family (Transposase gene disrupting gene coding for CitT); PFR_JS15-1_2045 ^a Transposase of ISAar20, ISL3 family
	LT618788	67%	Transposase gene disrupting a Type III restriction enzyme	PFR_JS15-2_359 Transposase of ISAar43, IS3 family, IS407 group, orfA and PFR_JS15-2_360 ^c Insertion sequence IS407 orfB
JS17	LT618789	67%	A duplicated transposase gene	PFR_JS17-1_657 and PFR_JS17-1_658 ^a Transposase of ISAar20, ISL3 family
	LT618790	67%		
JS20	LT618792	67%	3 transposase genes absent from the other genome	PFR_JS20-2_544 Insertion sequence IS407 OrfB and PFR_JS20-2_545 ^c Transposase of ISAar43, IS3 family, IS407 group, orfA (disrupting Transposase IS30 family gene); PFR_JS20-2_568 Transposase of ISAar20, ISL3 family (ahead of Carbon starvation protein); PFR_JS20-2_1234 Transposase of ISAar20, ISL3 family (ahead of LuxS)
	LT618791	67%		
JS21	LT618781	67%	Transposase gene ahead of putative aminotransferase biotin synthesis related protein	PFR_JS21-2_248 Transposase of ISAar20, ISL3 family
	LT618782	67%		

^a Transposase genes PFR_JS12-1_615, PFR_JS15-1_2045, PFR_JS17-1_657 and PFR_JS17-1_658 found in strains JS12, JS15 and JS17, respectively, are 100% identical

^b The same element is present in the strain JS11 in the same location, but in all of the sequences

^c Transposase genes PFR_JS15-2_359 and PFR_JS15-2_360 are identical to PFR_JS20-2_544 and PFR_JS20-2_545. The transposase genes do not appear to be acquired through horizontal transfer, as each of them is present in their respective genomes in at least one more location. All of the transposase genes have counterparts in multiple strains, suggesting their intrinsic character for the species

can be viewed in Additional files 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 and 21.

Mobile elements

Bacteriophages

In this study, three bacteriophages were discovered as a circular DNA within the strains JS7, JS22 and JS23 (Fig. 6). The phage found in the strain JS7 (LT618778) has a total genome size of 37,936 bp and 59 predicted open reading frames. When integrated into the chromosome as a prophage, it was located between the sequence coding for the transcriptional regulator KmtR, immediately downstream of a tRNA-Ala (agc) and a tRNA-Lys(ttt), immediately upstream of a transcriptional

regulator MtrR. A BLAST search against the known Proionibacteria phages showed that PJS7 is 99% identical with the 38,071 bp-long PFR1 phage (NC_031076.1). The difference can be found in the gene coding for the minor tail protein, where PFR-JS7_47 is 135 nucleotides shorter than BI042_gp13. The strain JS7 differed from the other phage-carrying strains, since sequencing revealed the coexistence of three types of genomes in its DNA-sample: a bacterial genome carrying a prophage (LT618776), a circular phage genome (LT618778), and unlike in samples of JS22 or JS23, also a bacterial genome cleared of prophage (LT618777). Replication of the circular phage genome in JS7 was followed by PCR after sub-culturing, which revealed successive integration of

a

ANiB	JS	JS2	JS4	JS7_ph	JS8	JS9	JS10	JS11	JS12	JS13	JS14	JS15	JS16	JS17	JS18	JS20	JS21	JS22	JS23	JS25
JS	---	99.02	98.71	98.97	98.71	98.38	99.96	98.65	98.52	98.64	98.66	98.89	98.63	98.71	98.53	98.13	98.66	98.96	98.92	98.62
JS2	99.03	---	98.74	98.91	98.88	98.43	99.01	98.82	98.73	98.82	98.74	98.93	98.64	98.92	98.82	98.44	98.72	98.97	98.94	98.72
JS4	98.64	98.66	---	98.7	98.4	98.53	98.67	98.73	98.55	98.72	98.75	98.75	98.74	98.49	98.87	98.47	99.98	98.69	98.65	99.95
JS7_ph	99.12	98.99	98.87	---	98.84	97.96	99.12	98.66	98.49	98.65	98.83	98.87	98.79	98.87	98.6	98.26	98.83	98.78	98.84	98.82
JS8	98.72	98.87	98.54	98.77	---	98.33	98.74	98.96	98.79	98.96	98.79	98.86	98.53	98.92	98.55	98.2	98.49	99.02	98.95	98.5
JS9	98.57	98.56	98.77	98.23	98.45	---	98.56	98.64	98.4	98.64	98.4	98.43	98.45	98.34	98.34	98.93	98.75	98.49	98.45	98.65
JS10	99.93	98.96	98.65	98.91	98.66	98.31	---	98.61	98.48	98.6	98.61	98.78	98.53	98.64	98.47	98.08	98.6	98.87	98.81	98.55
JS11	98.57	98.7	98.63	98.43	98.71	98.31	98.56	---	99.03	100	99.03	98.52	98.45	98.61	98.48	98.28	98.59	98.57	98.47	98.65
JS12	98.46	98.65	98.44	98.31	98.66	98.07	98.54	99.15	---	99.14	98.98	98.45	98.4	98.54	98.21	98.07	98.36	98.52	98.47	98.38
JS13	98.57	98.7	98.63	98.43	98.71	98.31	98.56	100	99.03	---	99.03	98.51	98.45	98.61	98.48	98.28	98.59	98.57	98.47	98.65
JS14	98.5	98.57	98.67	98.54	98.62	98.09	98.52	99.01	98.92	99.02	---	98.69	98.62	98.56	98.69	98.41	98.64	98.49	98.68	98.67
JS15	98.87	98.89	98.8	98.78	98.85	98.21	98.82	98.69	98.55	98.7	98.84	---	98.76	99.18	98.8	98.35	98.75	99.26	99.62	98.72
JS16	98.67	98.64	98.86	98.72	98.51	98.23	98.64	98.68	98.56	98.67	98.88	98.85	---	98.69	98.77	98.36	98.84	98.72	98.79	98.81
JS17	98.89	98.96	98.67	98.84	98.94	98.28	98.86	98.79	98.7	98.8	98.76	99.27	98.72	---	98.7	98.2	98.66	99.31	99.34	98.64
JS18	98.71	98.85	98.98	98.62	98.64	98.2	98.69	98.75	98.54	98.74	98.9	98.94	98.96	98.72	---	98.35	98.96	98.72	98.87	98.91
JS20	98.29	98.45	98.55	98.22	98.21	98.72	98.28	98.57	98.2	98.55	98.62	98.43	98.42	98.14	98.3	---	98.48	98.37	98.42	98.48
JS21	98.65	98.68	99.99	98.7	98.4	98.53	98.68	98.73	98.55	98.73	98.76	98.75	98.75	98.5	98.86	98.48	---	98.69	98.65	99.96
JS22	98.93	98.98	98.73	98.75	99.03	98.29	98.92	98.72	98.65	98.73	98.6	99.29	98.68	99.23	98.55	98.32	98.66	---	99.32	98.68
JS23	98.87	98.87	98.77	98.73	98.9	98.17	98.79	98.64	98.51	98.64	98.85	99.67	98.77	99.19	98.78	98.35	98.69	98.99	---	98.68
JS25	98.59	98.63	99.97	98.67	98.4	98.54	98.63	98.73	98.55	98.62	98.77	98.7	98.76	98.42	98.8	98.47	99.98	98.68	98.64	---

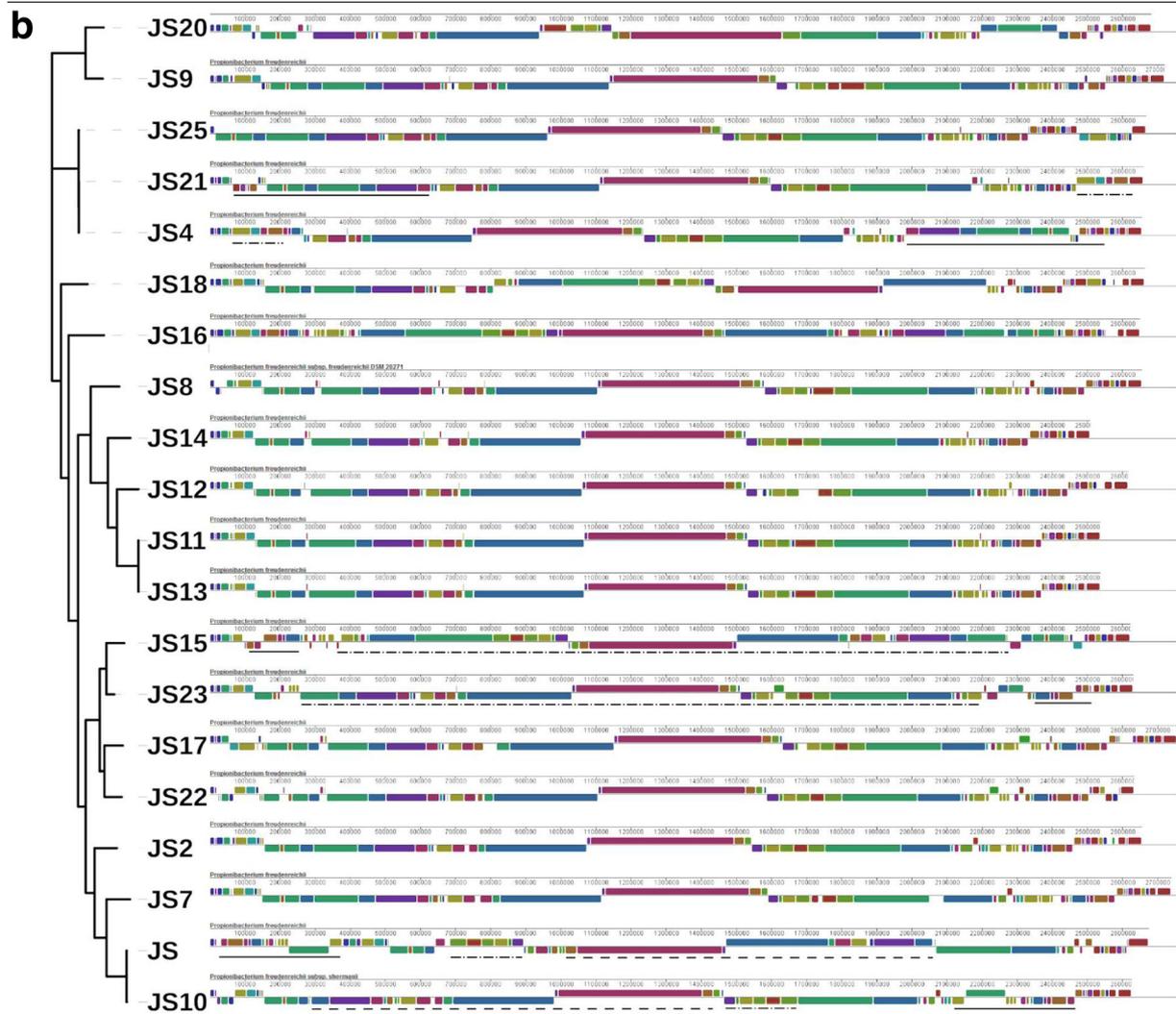


Fig. 1 (See legend on next page.)

(See figure on previous page.)

Fig. 1 Genome composition and organisation. Panel **a**) Average Nucleotide Identity (%) calculated based on pairwise BLAST alignment (ANIb). The levels of similarity are highlighted by coloring from green for the most similar to red for the most dissimilar. The strains JS4, JS15 and JS17 are on average the most similar to all other strains, while the strains JS9 and JS20 are the most dissimilar to all other strains and only slightly more similar to one another. The strains of cereal origin (JS11-JS14) are more similar to each other than to other strains. Panel **b**) Whole genome alignments generated with ProgressiveMauve. The genomes are arranged according to the phylogenetic tree generated from core genome alignments (see below). The distinct organisation of the genomes of closely related strains can be observed, most clearly between strains JS and JS10, JS15 and JS23 as well as JS4 and JS21. The regions of genome rearrangements in these strains are indicated with matching lines (solid, dot-dash or dash)

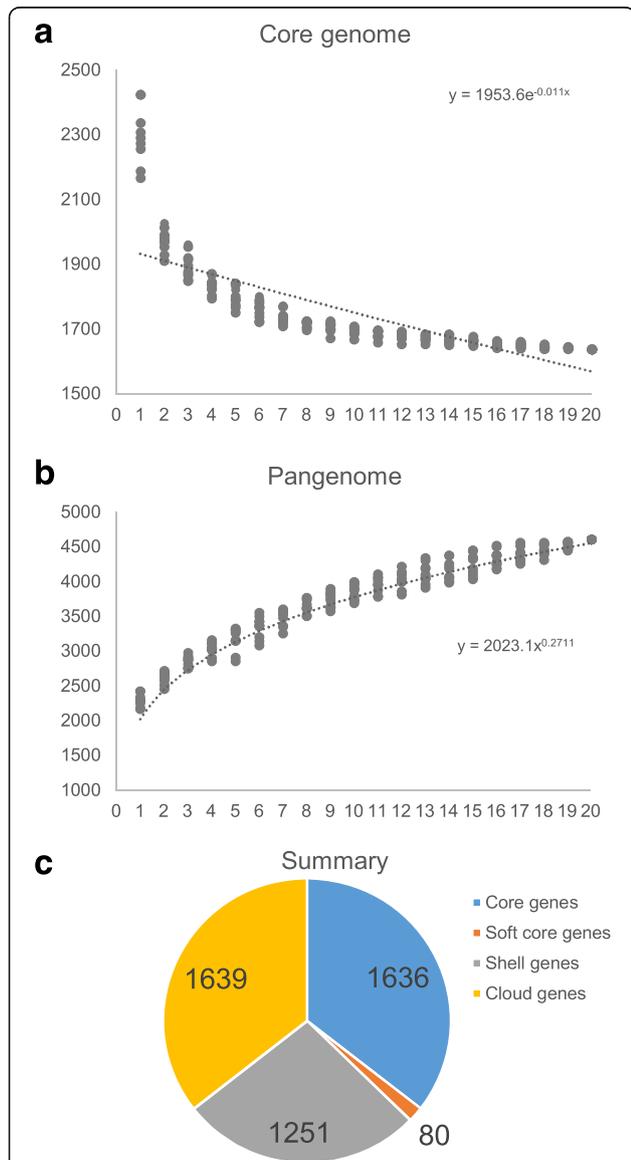
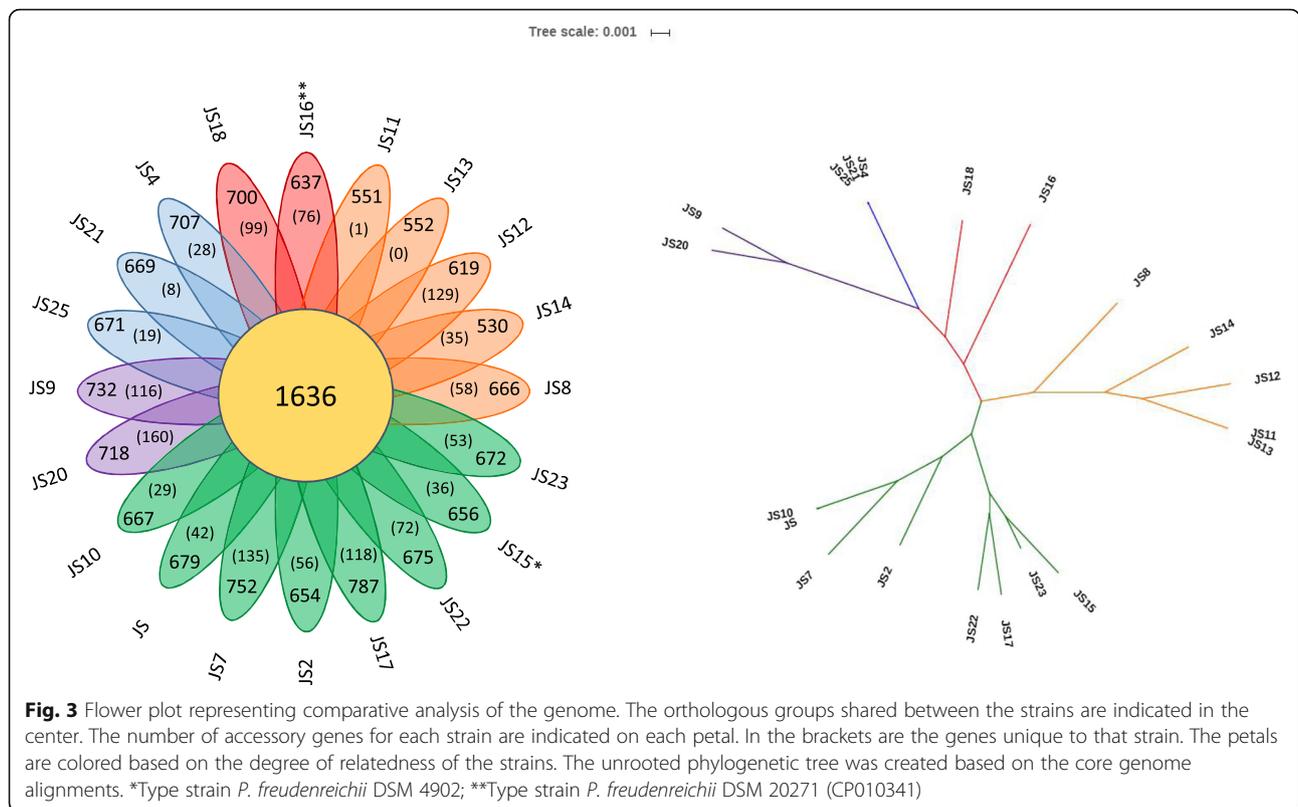


Fig. 2 Core genome and pan genome of the *P. freudenreichii* species. The core genome (**a**) and the pan genome (**b**) are represented as a variation of the gene pools after sequential addition of 20 *P. freudenreichii* genomes. The summary of ortholog group distribution between the strains is presented in a pie chart (**c**). Core genes- present in all of the strains; Soft core genes- present in 19 of the strains; Shell genes- present in 3-18 of the strains; Cloud genes- present in one or two strains only

phage genomes after five passages in the PPA medium (Additional file 5: Phage integration). The other two strains, JS22 and JS23 carried the prophage on all the copies of the chromosome as well as the circular phages. The phages PJS22 and PJS23 are 97% identical over 68% of their sequences. PJS22 shows 99% identity over 81% sequence to previously sequenced phage B22 (KX620750.1), PJS23 is the most similar to the phage Doucette (KX620751.1) with 97% identity over 64% of the sequence. The PJS22 phage is inserted between sequence coding for tRNA-Gly (ccc) and a DNA protection during starvation protein 2 (PFR_JS22-1_1997) while the PJS23 prophage is inserted between tRNA-Pro(tgg) and a Quaternary ammonium compound-resistance protein SugE (PFR_JS23_1469).

All the genomes were checked for additional prophage sequences with two dedicated programs: Phaster and Prophinder. From the candidate prophages, only the prophage from the strain JS17 appeared complete. The prophage JS17 is located between the tRNA-Ser(tga) gene and a transposase gene (PFR_JS17-1_2095). A BLAST search revealed 96% identity over 61% and 64% of the sequence to Propionibacterium phages Doucette and G4, respectively. Similarly, a BLAST analysis against phages PJS22 and PJS23 showed 97% identity over 62% and 65% of the sequence, respectively.

In the sequences of the prophage of the strain JS17 and the bacteriophage PJS23 a number of transposase genes were found. The PFR_JS17-1_2038 was identical to eight (PFR_JS17-1_341, PFR_JS17-1_394, PFR_JS17-1_676, PFR_JS17-1_711, PFR_JS17-1_2205, PFR_JS17-1_2341, PFR_JS17-1_2347 and PFR_JS17-1_2416), while PFR_JS17-1_2067 was identical to six (PFR_JS17-1_13, PFR_JS17-1_46, PFR_JS17-1_72, PFR_JS17-1_657, PFR_JS17-1_658 and PFR_JS17-1_1466) transposase genes found in other locations in the same strain. In addition, the PFR_JS17-1_657 and PFR_JS17-1_658 were the ones observed as duplicated in only part of the genome sequences of the strain JS17. These transposase genes were identical to the ones found in only part of the genome sequences of the strains JS12 and JS15 (see Table 3). Within prophage PJS23 sequence, there were four transposon-like elements, PFR_JS23_1432-PFR_JS23_1435 (PH1_40-PH1_43 on the phage genome). The PFR_JS23_1432 and PFR_JS23_1435 were both unique to the phage region of the genome, while



PFR_JS23_1433 (integrase) and PFR_JS23_1434 (transposase) were each found in two additional, co-located copies on the bacterial chromosome (PFR_JS23_368 and PFR_JS23_369; PFR_JS23_2185 and PFR_JS23_2184).

Plasmid-like elements

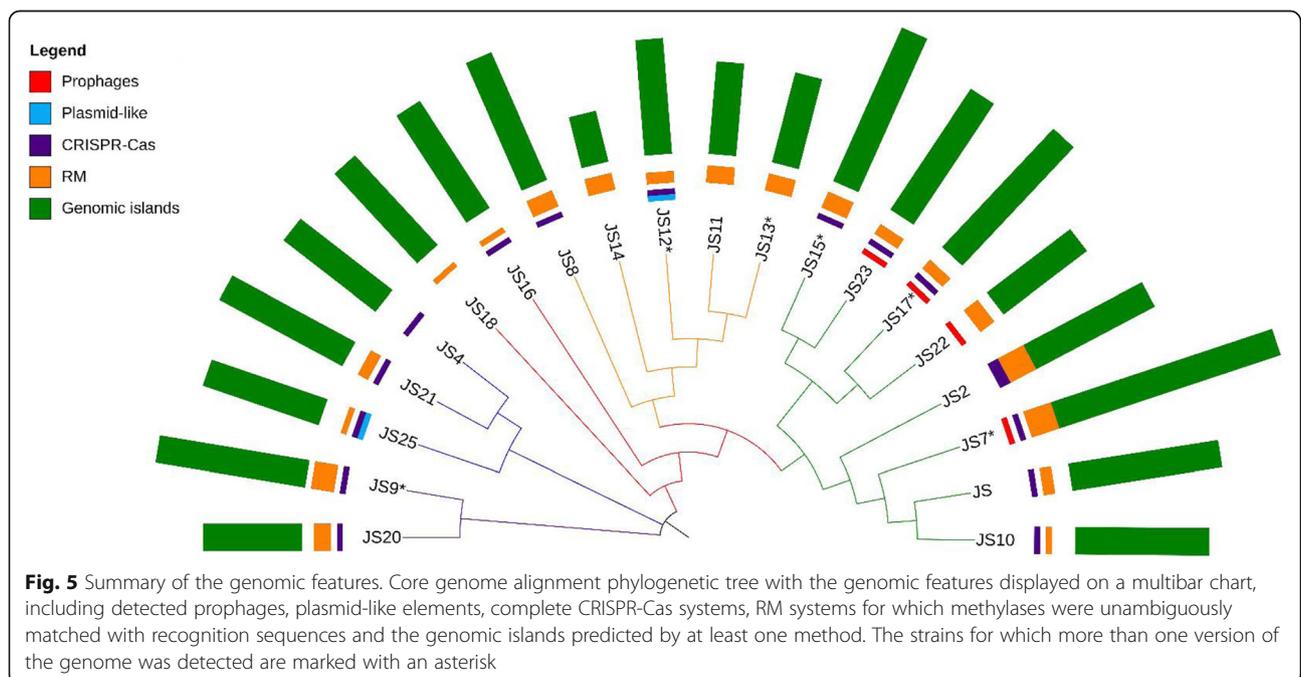
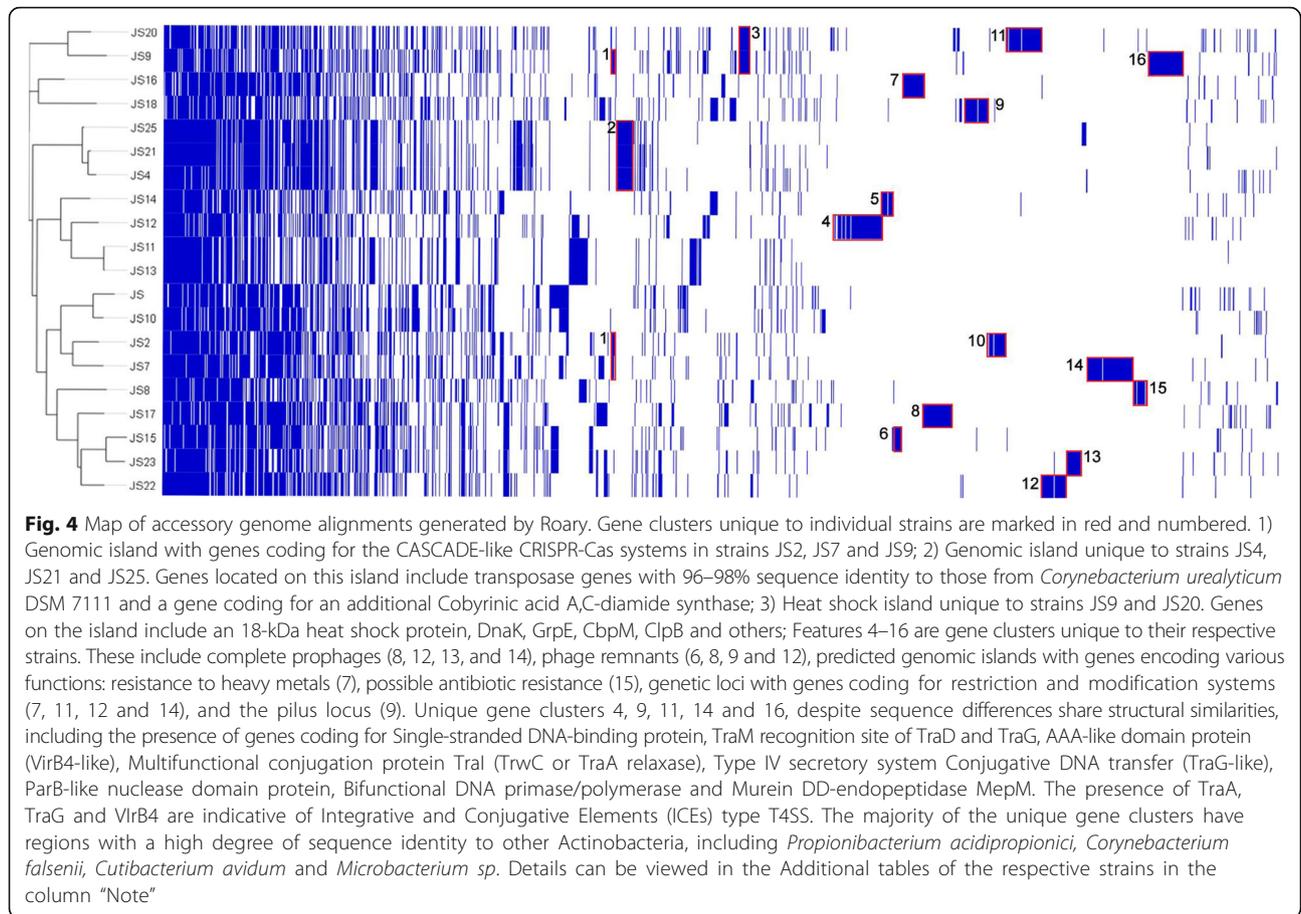
Two plasmid-like elements PFRJS12-3 (LT604882) and PFRJS25-1 (LT618784) were detected from strains JS12 and JS25, respectively. PFRJS12-3 and PFRJS25-1 are of 24.9 kbp and 35.6 kbp in size and include 32 and 46 predicted open reading frames, respectively (Additional file 10: LT604882 (plasmid) and Additional file 21: LT618784 (plasmid)). According to homology searches PFRJS12-3 and PFRJS25-1 sequences share no significant similarity with reported *P. freudenreichii* plasmids. In addition, no similarity to plasmids pIMPLE-HL096PA1 [33] or PA_15_1_R1 from the closely related species, *Cutibacterium acnes*, was found. A BLASTn search of the PFRJS12-3 revealed that the gene PFR_JS12-3_15 encoding a transposase is 93-95% identical with the transposase genes of *P. freudenreichii*, *Acidipropionibacterium acidipropionici*, *Micrococcus luteus* and *Corynebacterium variabile* at positions 8594-9669. The transposase gene PFR_JS12-3_12 in PFRJS12-3 is 90% identical to *A. acidipropionici* and *Micrococcus luteus* sequences at position 5799-7130, and the gene

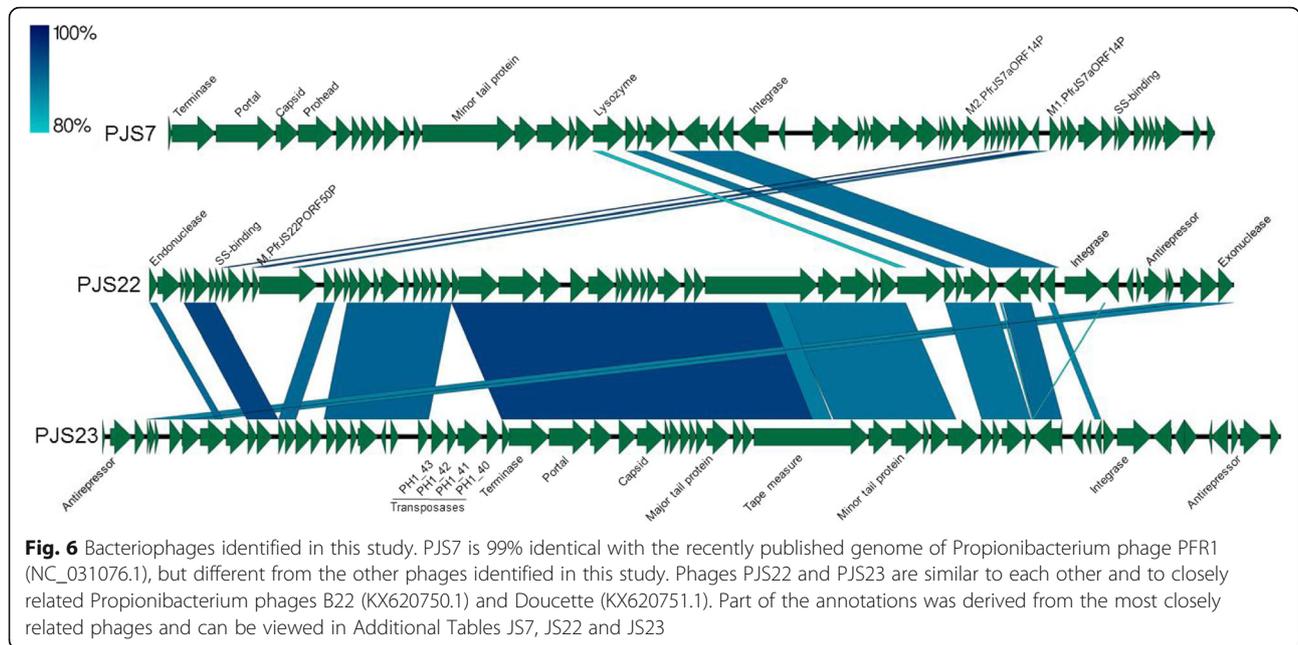
PFR_JS12-3_22is 92% identical to a resolvase gene from *A. acidipropionici* at positions 12,361-12,930. BLASTn search of PFRJS25-1 revealed a stretch of 88% identity to the *Propionibacterium* phage PFR1 over the stretch of its genes PFR1_23, PFR1_24 and PFR1_25, all encoding hypothetical proteins. Additionally, the 5' end of this sequence showed 98% identity with a 47 nt stretch in the non-coding region in *Burkholderia pyrrocinia* plasmid p2327 and *Burkholderia cenocepacia* plasmid pBCJ2315. A BLASTp search using the predicted proteins of PFRJS25-1 against those from p2327 and pBCJ2315 revealed negligible sequence similarity.

Further analysis showed that the PFRJS25-1 was 99% identical over 31% of the sequence of PFRJS12-3 (Fig. 7). The analysis comparing sequences against the Conserved Domains Database (CDD) [34] revealed multiple regions of similarity to conjugative plasmids from both elements. These regions of similarity included those with conserved domains of conjugal transfer proteins TrwC, TraC, TraG and TrbL as well as plasmid partition protein ParA. As no characteristic replication origin loci were found it remains to be elucidated whether the circular elements found in strains JS12 and JS25 are plasmids.

Genomic islands

The genomes were assessed for the presence of genomic islands by the integrative online tool IslandViewer 3



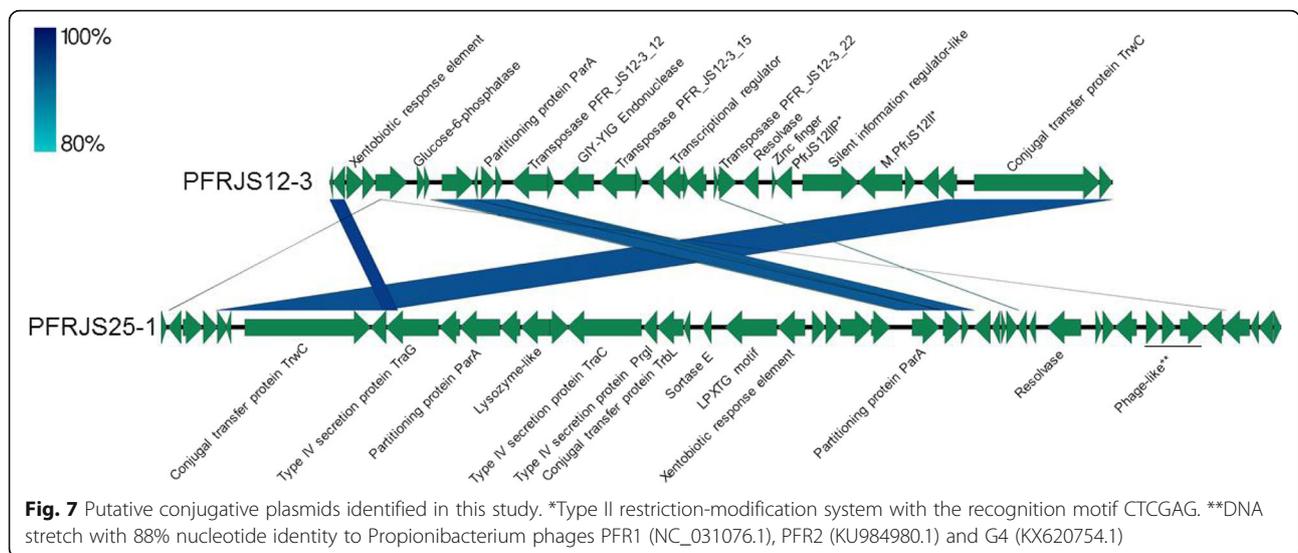


[35], which performs the analysis with three independent genomic island prediction methods: IslandPick, IslandPath-DIMOB, and SIGI-HMM.

The ability to utilise lactose, a historically important trait in *P. freudenreichii*, has been previously tied to a genomic island on which genes coding for UDP-glucose 4-epimerase (*galE1*), Sodium:galactoside symporter (*galP*) and Beta-galactosidase (*lacZ*) are located [21]. In our study, besides the type strain JS15, the same island was found in nine other strains: JS, JS2, JS7, JS8, JS10, JS17, JS18, JS22 and JS23, with JS23 possessing two copies of the region (PFR_JS23_160-PFR_JS23_162 and PFR_JS23_2069 -PFR_JS23_2071) (Additional files 2, 3,

4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 and 21). The presence of the island correlates with the ability to utilise lactose by these strains in vivo (see Additional file 22).

Another feature potentially giving a competitive advantage in the dairy environment is the ability to degrade D-lactate. Eight strains, including JS2, JS7, JS8, JS10, JS15, JS17, JS20 and JS23 were found to be equipped with a D-lactate dehydrogenase encoding gene located on a genomic island, while the gene encoding D-lactate dehydrogenase in strain JS18 is located just downstream of a predicted genomic island. For other traits important in food production see Additional file 23.



In strain JS4, a genomic island with an alternative pathway for the biosynthesis of rhamnose consisting of genes for dTDP-4-dehydrorhamnose reductase (*rmlD*), a putative dTDP-4-dehydrorhamnose 3, 5-epimerase (*rfbC*) and dTDP-glucose 4, 6-dehydratase (*rmlB*) was found. Finally, an island on which genes coding for pilus components was found in strain JS18, including Sortase SrtC1 (PFR_J18_2247), Type-2 fimbrial major subunit (PFR_J18_2248) and a Surface-anchored fimbrial subunit (PFR_J18_2249) (Additional file 16).

P. freudenreichii were previously reported to possess anti-inflammatory properties [29, 36]. Those properties were associated with a range of S-layer proteins: SlpE [29], SlpA and SlpB [36]. Genes coding for SlpA (RM25_1747 in the reference strain) and another Slp protein (RM25_1746) were found in all of the strains, in seven of the strains (JS, JS2, JS4, JS10, JS17, JS18 and JS23) identified as a part of a genomic island. The complete genes coding for the SlpE protein (Hypothetical protein) were found in 12 of the strains included in this study (PFR_JS2_13, PFR_JS4_13, PFR_JS8_26, PFR_JS9-1_12, PFR_JS10_12, PFR_JS12-1_12, PFR_JS14_12, PFR_JS15-1_12, PFR_JS17-1_12, PFR_JS22-1_12, PFR_JS23_12, PFR_JS25-1_2272); SlpB was found in two strains (PFR_JS14_229 and PFR_JS17-1_279). In addition, a 220 aa long S-layer protein precursor (encoded by the *ctc* gene) was found in 13 of the strains (PFREUDJS001_001526, PFR_JS2_711, PFR_JS8_732, PFR_JS10_644, PFR_JS11_672, PFR_JS12-1_664, PFR_JS13-1_672, PFR_JS14_687, RM25_1523, PFR_JS17-1_772, PFR_JS18_1933, PFR_JS22-1_727, PFR_JS23_662).

We surveyed the genomes for known antibiotic resistance genes. In nearly all of the *P. freudenreichii* strains these genes are not located on a predicted genomic island, however, strain JS8 appears to be diverging from the group. The genomic island unique to strain JS8 (see Fig. 4, feature 15) includes three genes coding for putative antibiotic resistance-related proteins: mitomycin radical oxidase, tetracyclin repressor domain-containing protein and a puromycin resistance protein Pur8. Furthermore, the edges of the island are flanked by the genes coding for hypothetical proteins which have 98 and 99% sequence identity to genes from *Brevibacterium linens* strain SMQ-1335 coding for mobile element proteins (see Additional file 6).

Immunity

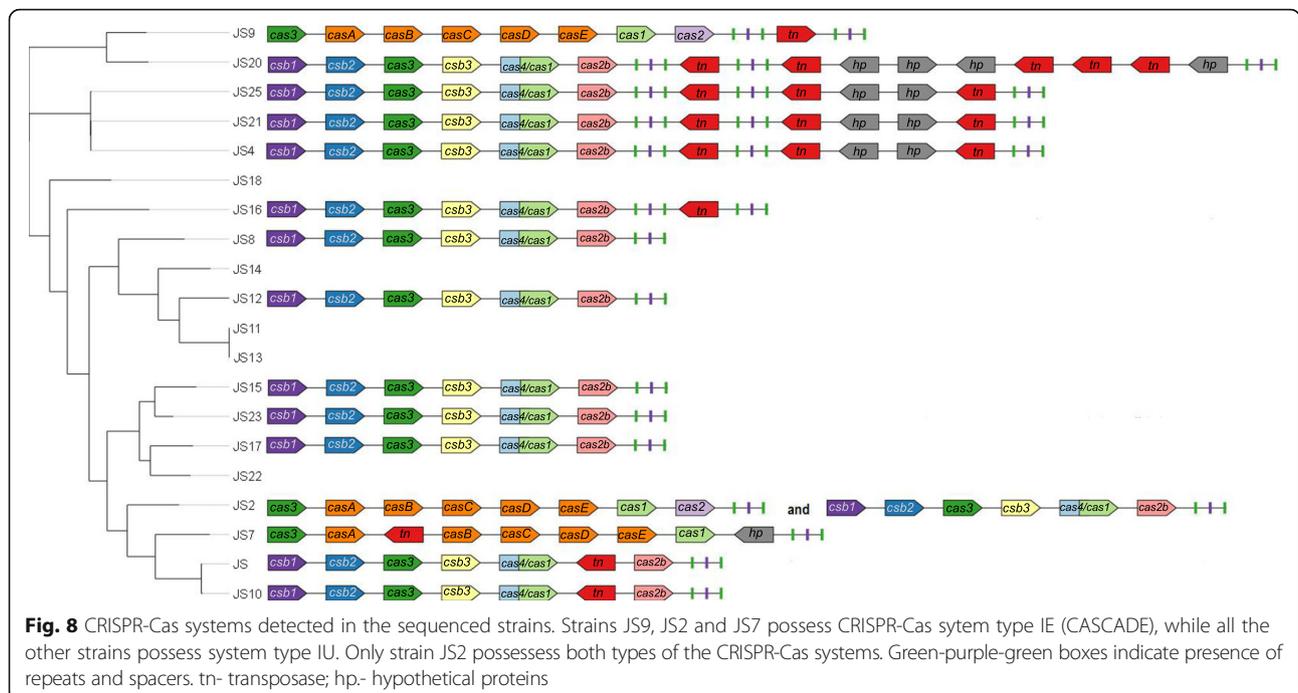
CRISPR-Cas systems

The clustered regularly interspaced short palindromic repeats (CRISPR) together with CRISPR associated proteins (Cas) form the adaptive immunity systems protecting their hosts against invasion by foreign DNA. The

function of the adaptive immunity system can be divided into two phenomena: CRISPR adaptation and CRISPR interference. The CRISPR adaptation results from spacer acquisition upon exposure to invading DNA, while the CRISPR interference involves recognition of the specific spacers on foreign DNA, which in turn allows for introduction of breaks in of the invading DNA and its resulting destruction reviewed by Savitskaya [37]. Currently, the CRISPR-Cas systems are subdivided into two classes, five types and 16 subtypes. Following this classification, we identified two systems in *P. freudenreichii* which, based on the presence of the Cas3 protein, we classified as belonging to the class1, type I CRISPR systems [37] (Fig. 8). The first of the systems, with the direct repeat consensus GGATCACCCCCGCGTATGCGGGGAGAAC, may be classified as the subtype IE based on sequence homology and gene organisation corresponding to the CASCADE system, which is well-characterised in *E. coli* [38, 39]. The second system, with the direct repeat consensus ATTGCCCTCCTTCTGGAGGGGCCCTTCATTGAG GC, bears similarity to the subtype IU (previously IC) [38], which is strengthened by the presence of the fusion protein Cas4/Cas1 found in several variants of the subtype IU [40, 41]. However, the atypical gene organisation suggests that it is a new variant of the subtype IU.

The CRISPR system IE was found in strains JS2, JS9 and JS7 and carried 96, 65 and 105 spacers, respectively (Table 4). These systems were located on genomic islands in all strains, which suggested relatively recent acquisition; however, the lack of sequence identity between the spacers suggested independent acquisition of immunity in each strain. The CRISPR system of strain JS7 had a transposase gene inserted between the *cse1* and *cse2* genes and only a fragment of the *cas2* gene comprising a part of a larger hypothetical protein. In strain JS9 the first 9 CRISPR spacers were separated from the following 96 spacers by an integrase. BLAST search of the spacers indicated immunity to all of the previously sequenced phages infecting *P. freudenreichii*, apart from the filamentous phage phiB5, for which immunity was found only in strain JS9. In addition, strain JS2 carried immunity against all three phages found in this study, and to the plasmid pJS25. The strain JS7 carried immunity to the plasmid pJS12, the phage PJS22 and to the phage by which it is infected (PJS7), suggesting that the presence of either the transposase gene or the incomplete *cas2* gene may have resulted in inactivity of the CRISPR-Cas system in this strain. JS9 carried markers of immunity against all the three phages found in this study and to the plasmid-like element PFRJS25-1.

The CRISPR-Cas system IU is more widespread in *P. freudenreichii* and can be found in 13 of the sequenced strains, including only one strain of cereal origin-JS12.



In strains JS4, JS16, JS20, JS21 and JS25 the systems are located on genomic islands predicted by Island Viewer. The number of spacers in IU CRISPR-Cas systems ranged from 25 in the strain JS2 to 64 in strain JS17 (Additional file 24). The spacers of the strains JS and JS10 and of the strains JS4, JS20, JS21 and JS25 are for the most part identical, which is in line with their phylogenetic relatedness. In other strains, only a few spacers are identical suggesting early diversification. Only strain JS2 carries both types of CRISPR systems, although strain JS9 possesses an additional stretch of 83 spacers separated by a distinct repeat sequence (GGGCTCACCCCGCATATGCGGGGAGCAC), indicating they would belong to a separate CRISPR-Cas system. Nevertheless, the lack of Cas genes in the vicinity and location of the CRISPR on a genomic island may mean that the system was acquired through incomplete horizontal gene transfer.

In strains JS11, JS13, JS14, JS18 and JS22 no complete CRISPR-Cas systems were found, although strain JS22 had a short stretch of CRISPRs. No immunity to known phages was found on that stretch. For each of the strains, 2–4 additional “Possible CRISPRs” were identified, most of which mapped within sequences coding for DNA topoisomerase 1, a hypothetical protein or fell between coding sequences. None of them showed homology to known Cas genes. Still, the “Possible CRISPR1” from strain JS14 carries a spacer with 100% identity to a fragment of the gene coding for a tape measure protein in phages Anatole and E1.

Interestingly, strain JS23 appears to have an intact CRISPR-Cas system and one spacer with 100% identity to the sequence of the prophage the strain carries. This could mean that the system is not functional, that one spacer is not sufficient to destroy the phage DNA, or that the phage possesses a mechanism of countering the strategies employed by the host. We explored the possibility that the mobile elements might carry anti-CRISPR genes [42]. To this end we performed a conserved domain search [34] of the mobile elements for which self-immunity was found, namely phages PJS7, PJS23, prophage of the strain JS17 and the plasmid pJS25, to identify candidate proteins with a typical helix-turn-helix domain, or the ability to bind DNA. The candidate proteins were then compared by the BLASTp algorithm with the previously identified anti-CRISPR genes acting on systems type IE and IF (Bondy-Denomy et al., 2013; Pawluk et al., 2015) (Additional file 25), however no similarities were found. Determining the activity of these putative anti-CRISPR genes requires further experiments, which are out of the scope of this study.

Restriction-modification systems

To gain some insights into the possible restriction-modification (RM) systems present in the 20 strains studied here, we first analyzed the genome sequences for the presence of genes that could be identified as components of RM systems. This was accomplished using SEQWARE and the REBASE database as outlined previously [43]. In this way, 216 different RM system genes could be identified associated with 127 different systems. For many of them, putative recognition sequences could

Table 4 CRISPR-Cas systems identified in the studied strains by CRISPR-finder

Strain	Predicted CRISPR	Start	End	Size	DR (nt)	No. spacers	Type	Immunity to identified phages	Immunity to mobile elements identified in this study	Locus ID	Note
JS	Crispr_2	2,424,099	2,427,021	2922	36	40	IU	Anatole, B3, B22, E1, E6, G4, PFR1, PFR2	PJS22, PJS23, PJS7, pJS12	PFREUDJS001_002141- PFREUDJS001_002147	-
JS2	Crispr_3	2,082,550	2,088,376	5826	28	95	IE	B22, B3, E1, E6, G4, Anatole, Doucette, PFR1, PFR2	PJS22, PJS23, PJS7, pJS25	PFR_JS2_1825-PFR_JS2_1832	-
JS4	Crispr_4	2,120,759	2,122,610	1851	36	25	IU	PFR1, PFR2	PJS7	PFR_JS2_1843-PFR_JS2_1849	-
JS4	Crispr_3	1,825,170	1,827,746	2576	36	35	IU	Doucette, E6, G4, PFR1, PFR2	PJS22, PJS7, pJS25	PFR_JS4_1650-PFR_JS4_1655	One system, negative strand, spacers separated by a transposase gene and then by a four-gene integron.
JS4	Crispr_2	1,823,162	1,823,697	535	36	7	-	-	-	-	-
JS4	Crispr_1	1,818,598	1,819,215	617	36	8	-	-	-	-	-
JS7	Crispr_2	2,203,337	2,207,335	3998	28	65	IE	B22, B3, E1, E6, G4, Anatole, Doucette, PFR1, PFR2	PJS22, PJS7, pJS12	PFR_JS7-1_1964-PFR_JS7-1_1971	No <i>cas2</i>
JS8	Crispr_3	2,192,257	2,196,491	4234	36	58	IU	Anatole, Doucette, B3, B22, E1, E6, G4, PFR1, PFR2, phiB5	PJS22, PJS23, PJS7, pJS25	PFR_JS8_1930-PFR_JS8_1935	-
JS9	Crispr_3	2,260,079	2,260,655	576	28	9	IE	-	-	PFR_JS9-1_2018-PFR_JS9-1_2025	One system, negative strand, spacers separated by an integrase
JS9	Crispr_2	2,253,046	2,258,937	5891	28	96	-	B22, B3, E1, E6, G4, Anatole, Doucette, PFR1, PFR2, phiB5	-	-	-
JS9	Crispr_4	2,288,755	2,292,746	3991	29	65	N/A	Doucette, B3, B22, G4, PFR1, PFR2	PJS23, PJS7	-	Remnant of a system, no Cas. Spacers separated by an integrase.
JS9	Crispr_5	2,293,887	2,295,016	1129	29	18	-	PFR1, PFR2	PJS7	-	-
JS10	Crispr_2	2,029,533	2,032,236	2703	36	37	IU	Anatole, B3, B22, E1, E6, G4, PFR1, PFR2	PJS22, PJS23, PJS7, pJS12	PFR_JS10_1797-PFR_JS10_1803	-
JS12	Crispr_1	2,181,462	2,185,411	3949	36	54	IU	Doucette, B3, B22, E6, G4, PFR1, PFR2, phiB5	PJS22, PJS23, PJS7	PFR_JS12-1_1906-PFR_JS12-1_1911	-
JS14	PossibleCrispr_1	1,595,184	1,595,290	106	36	1	N/A	Anatole, E1	-	PFR_JS14_1397	-
JS15_1	Crispr_1	441,841	444,266	2425	36	33	IU	Anatole, B3, B22, E1, E6, G4, phiB5	PJS23, pJS12	PFR_JS15-1_371-PFR_JS15-1_376	-
JS16	Crispr_2	417,820	419,522	1702	36	23	IU	B22, E6, G4, PFR1, PFR2	PJS22, PJS23, PJS7	RM25_0342-RM25_0347	One system, negative strand, spacers separated by a transposase gene. Both CRISPR stretches carry immunity to known phages.
JS16	Crispr_1	413,342	416,347	3005	36	41	-	Anatole, B3, B22, E1, E6	PJS22	-	-

Table 4 CRISPR-Cas systems identified in the studied strains by CRISPR-finder (Continued)

Strain	Predicted CRISPR	Start	End	Size	DR (nt)	No. spacers	Type	Immunity to identified phages	Immunity to mobile elements identified in this study	Locus ID	Note
JS17_1	Crispr_2	2,220,735	2,225,402	4667	36	64	IU	Anatole, Doucette, B3, B22, E1, PFR1, PFR2	PJS22, PJS23, pJS12, pJS25	PFR_JS17-1_1953-PFR_JS17-1_1958	-
JS20_1	Crispr_3	2,034,463	2,037,183	2720	36	37	IU	Doucette, B22, E6, G4, PFR1, PFR2	PJS23, PJS7	PFR_JS20-1_1815-PFR_JS20-1_1820	One system, spacers separated by a transposase gene and then by an eight-gene integron.
	Crispr_4	2,038,656	2,039,048	392	36	5		-	-		
	Crispr_5	2,044,247	2,044,936	689	36	9		-	-		
JS21_1	Crispr_5	2,189,758	2,192,624	2866	36	39	IU	Doucette, E6, G4, PFR1, PFR2	PJS22, PJS7, pJS25	PFR_JS21-1_1963-PFR_JS21-1_1968	One system, negative strand, the spacers separated by a transposase gene and then by a four-gene integron.
	Crispr_4	2,187,750	2,188,285	535	36	7		-	-		
	Crispr_3	2,183,185	2,183,803	618	36	8		-	-		
JS22	Crispr_2	2,150,779	2,151,318	539	36	7	IU	-	-		Remnant of the same IU system, but no Cas. No immunity to known phages found.
JS23	Crispr_2	2,124,520	2,126,512	1992	36	27	IU	Anatole, B3, B22, E1, E6, G4, phiB5	PJS23	PFR_JS23_1889-PFR_JS23_1894	-
JS25	Crispr_2	2,042,081	2,045,018	2937	36	40	IU	Doucette, E6, PFR1, PFR2	PJS22, PJS7, pJS25	PFR_JS25-1_1807-PFR_JS25-1_1812	One system, spacers separated by a transposase gene and then by a four-gene integron.
	Crispr_3	2,046,491	2,046,954	463	36	6		-	-		
	Crispr_4	2,050,901	2,051,519	618	36	8		-	-		

be assigned based on similarity to well-characterized RM systems in other organisms. Next, we took advantage of the fact that PacBio sequencing can detect the methylated bases, m6A and m4C, and the motifs in which they occur can be assigned [44]. Most strains contained more than one motif, although one strain, JS4, was devoid of apparent methylase activity and one strain, JS10, had an unusual motif characteristic of a Type I RM system, but with only one of the two sub-motifs methylated. The reason for this, as well as its significance, are unknown. Among the remaining eighteen strains, forty-nine motifs were found.

To assign the methylase genes responsible for each of the motifs we used a combination of direct assignment when a gene had very high similarity, usually greater than 90%, to a known gene or by noting when there was only candidate for a gene of a given Type of RM system. For instance, Type I RM systems have recognition sequences which are split into two sub-motifs containing 2–5 specific bases separated by a spacer of four to nine non-specific bases. Finally, once the easily identified motifs were matched to the genes encoding the methylases responsible a process of elimination was used to assign a few of the remaining matches. In this way, all but four of the motifs could be matched unambiguously to the genes encoding the responsible methylases (Table 5, Additional file 26). Among the strains, JS2 and JS7 had three Type I systems, while six strains had two such systems and nine strains had a single system. In all of these strains, except JS10, the R gene responsible for restriction was intact and the level of methylation was close to complete. This suggests that the systems were active as RM systems. It should be noted that many of the specificities were unique or newly found in this genus.

BLASTp analysis of the RM genes revealed that one variant of the Type I (group 1 in Table 5) system is widespread among the *P. freudenreichii* strains tested, located in 18 of the sequenced strains, with the exception of the strains JS9 and JS20. However, in six strains: JS (M.PfrJSII), JS4 (PFR_JS4_139), JS10 (M.PfrJS10ORF2151AP), JS15 (M.PfrJS15ORF135P), JS17 (M.PfrJS17ORF2252P) and JS23 (M.PfrJS23ORF2177P) the systems were inactive, most likely due to transposon-mediated inactivation of methylases (JS and JS10) or genes coding for specificity proteins (JS15, JS17 and JS23) or due to other, undetermined reasons (JS4 and JS23). Interestingly, even though the methylases from this group of RM systems were 97–100% identical on the amino acid level, the differences in specificity proteins resulted in differing recognition sequences. Strikingly, the amino acid sequences of the specificity proteins PFR_JS18_2128 and PFR_JS18_2129 aligned 100% to regions of the specificity protein PFR_JS8_2158. DNA sequence alignment of the coding regions showed that an insertion of cytosine at

Table 5 Methylation motifs and the responsible methylases identified in *P. freudenreichii*

Organism	locus	Enzyme	Recognition sequence	Type	group
<i>P. freudenreichii</i> JS	921	M.PfrJSII	AGTNNNNNCTT	I	2
	566	M.PfrJSI	AGCAGY	III β	4
<i>P. freudenreichii</i> JS2	PFR_JS2_249	M.PfrJS2I	ATCNNNNNGTCG	I y	3
	PFR_JS2_1297	M.PfrJS2II	AGTNNNNNCTT	I y	2
	PFR_JS2_2113	M.PfrJS2III	GATNNNNNITTT	I y	1, unique motif unique in PF
	PFR_JS2_2250	M.PfrJS2V	GCCGCG	II α	
	PFR_JS2_1977	PfrJS2IV	CTTCNAC	II y	
<i>P. freudenreichii</i> JS4	no hits				
<i>P. freudenreichii</i> JS7	PFR_JS7-1_251	M.PfrJS7I	ATCNNNNNGTCG	I y	3
	PFR_JS7-1_659	M.PfrJS7II	CAANNNNNCTG	I y	unique in PF
	PFR_JS7-1_2249	M.PfrJS7III	CGANNNNNITTT	I y	1, unique motif
	PFR_JS7-1_2117	M.PfrJS7IV	CCWGG	II	unique in PF
	PFR_JS7-1_1519	M.PfrJS7V	CTGAG	II y	unique in PF
<i>P. freudenreichii</i> PJS7 phage	PFR_JS7-PH_14	M1.PfrJS7aORF14P	(not annotated)	II	
	PFR_JS7-PH_22	M2.PfrJS7aORF14P	(not annotated)	II	
<i>P. freudenreichii</i> JS8	PFR_JS8_242	M.PfrJS8I	ATCNNNNNGTCG	I y	3
	PFR_JS8_2157	M.PfrJS8III	CAGNNNNNRTCG	I y	1, unique motif
	PFR_JS8_2013	M.PfrJS8II	CCWGG	II	6
<i>P. freudenreichii</i> JS9	PFR_JS9-1_274	M.PfrJS9I	GCAANNNNNGTGG	I y	3
	PFR_JS9-1_1487	M.PfrJS9II	CAGNNNNNCTG	II y	unique in PF
	PFR_JS9-1_705	M.PfrJS9III	CTGGAG	II	unique in PF
	PFR_JS9-1_1590	M.PfrJS9IV	ARCCCG	II	unique in PF
<i>P. freudenreichii</i> JS10	PFR_JS10_1239	M.PfrJS10I	AGTNNNNNCTT	I y	2
<i>P. freudenreichii</i> JS11	PFR_JS11_224	M.PfrJS11I	ACNNNNNCTC	I y	3
	PFR_JS11_2032	M.PfrJS11II	GGANNNNNCTC	I y	1
	PFR_JS11_1917	M.PfrJS11III	CCWGG	II	6
<i>P. freudenreichii</i> JS12	PFR_JS12-1_2109	M.PfrJS12I	GCAANNNNRTGA	I y	1
	PFR_JS12-1_1387	M.PfrJS12III	GCGCC (GGCGGAG) (TANAG)	II β (II y) (II y)	unique in PF
<i>P. freudenreichii</i> PJS12	PFR_JS12-3_27	M.PfrJS12II	CTGAG	II y	unique in PF
<i>P. freudenreichii</i> JS13	PFR_JS13-1_224	M.PfrJS13I	ACNNNNNCTC	I y	3
	PFR_JS13-1_2034	M.PfrJS13II	GGANNNNNCTC	I y	1
	PFR_JS13-1_1919	M.PfrJS13III	CCWGG	II	6
<i>P. freudenreichii</i> JS14	PFR_JS14_217	M.PfrJS14I	ATCNNNNNGTCG	I y	3
	PFR_JS14_1997	M.PfrJS14III	CAGNNNNNCTG	I y	1
	PFR_JS14_1883	M.PfrJS14II	CCWGG	II	6
<i>P. freudenreichii</i> JS15	PFR_JS15-1_281	M.PfrJS15II	CCWGG	II	6
	PFR_JS15-1_2040	M.PfrJS15I	AGCAGY	III β	4
	PFR_JS15-1_303	PfrJS15III	CTTCNAC	II y	7
<i>P. freudenreichii</i> JS16	RM25_0134	M.Pfr20271I	GGANNNNNCTT (TCGCGA)	II	1
<i>P. freudenreichii</i> JS17	PFR_JS17-1_893	M.PfrJS17I*	GAANNNNNCTT	I y	5
	PFR_JS17-1_273	M.PfrJS17II	AGCAGY	III β	4
<i>P. freudenreichii</i> JS18	J18_2127	M.PfrJS18I	CAGNNNNNCTG	I y	1
<i>P. freudenreichii</i> JS20	PFR_JS20-1_520	M.PfrJS20I	GACGTC	II β	unique in PF
	PFR_JS20-1_402	M.PfrJS20II	ACGNNNNNGTGG	I y	unique in PF
	PFR_JS20-1_2119	M.PfrJS20III	GCAANNNNNGTGG	I y	3
<i>P. freudenreichii</i> JS21	PFR_JS21-1_2269	M.PfrJS21I	ACANNNNNRTAY	I y	1
	PFR_JS21-1_2026	M.PfrJS21II	CCWGG	II	6
<i>P. freudenreichii</i> JS22	PFR_JS22-1_2165	M.PfrJS22I**	GAANNNNNCTT	I y	1
	PFR_JS22-1_1883	M.PfrJS22II	GRGCYC	II	unique in PF
	PFR_JS22-1_271	M.PfrJS22III	AGCAGY (CCGCTC) (m5C) (not annotated)	III β (II y)	4
<i>P. freudenreichii</i> JS22 phage	PFR_JS22-PH_50	M.PfrJS22PORF50P	(not annotated)	II	
<i>P. freudenreichii</i> JS23	PFR_JS23_2000	M.PfrJS23I	AGCAGY	III β	4
	PFR_JS23_1965	PfrJS23II	CTTCNAC	II y	7
<i>P. freudenreichii</i> JS25	PFR_JS25-1_2023	M.PfrJS25I	ACANNNNNRTAY	I y	1

Recognition sequences in parentheses indicates the methylase responsible cannot be assigned unambiguously. Modified bases are marked in colour; “not annotated” indicates there is no locus tag as the gene is not annotated in the GenBank file. Color coding for the methylation types Blue: 6 mA; Orange: 4mC; Purple: 5mC *It is possible that it is really M.PfrJS17ORF2252P that is the active methylase in this organism. M.PfrJS17ORF2252P shares 99% aa sequence identity with M.PfrJS22I, but its corresponding S protein is truncated **It is possible that it is really M.PfrJS22ORF850P that is the active methylase in this organism. M.PfrJS22ORF850P is 100% identical to M.PfrJS17I, as are their corresponding S proteins while the R proteins differ by one amino acid (475 Ala:Gly)

position 2,404,971 in JS18 DNA caused a frameshift, splitting the specificity gene into two otherwise 100% identical genes. It is worth noting that the recognition sequences associated with the corresponding RM systems in strains JS18 and JS8 are CAGNNNNNCTG and CAGNNNNNRTCG, respectively. The other variants of Type I systems (group 2, 3 and “unique in PF” in Table 5) differed from each other on the sequence level, but were highly conserved within the groups, with identical recognition sequences within groups.

In addition to the Type I systems, there were also examples of Type II methylases including examples of the same specificity in several strains, for instance seven strains contained a methylase recognizing CCWGG, identical in 6 of the strains (group 6 Table 5) and distinct in strain JS7 (M.PfrJS7IV), which is commonly found in many different genera. The RM system located on the putative plasmid JS12 (PfrJS12II) was active and unique among the *P. freudenreichii* strains included in this study. However, a BLASTn search revealed that the contig NZ_CDAG01000006.1 from the draft genome of the strain CIRM-BIA 456 showed 99% identity over 21% of the JS12 putative plasmid sequence (13756–18,773), spanning 48% of the contig sequence (150–5167) and encompassing the restriction-modification system PfrJS12II.

While some of the methylase motifs identified here were identical with previously known ones, the Type IIG enzymes were all new and unique. One Type III RM system, with the unique recognition sequence AGCAGY, was found in five of the strains.

Vitamin B12 biosynthetic pathway

The vitamin B12 biosynthetic pathway in *P. freudenreichii* has been resolved previously [21, 45–47] and the organisation of the genes has been described earlier [21]. All the strains included in this study demonstrated an ability to produce active vitamin B12 [48] and we confirmed that all the strains possess the previously identified genes, in similar organisation, and highly conserved (protein alignments can be found in Additional file 27). Strain JS4 is an exception since *hemL* and *cbiD* genes are shorter, and the gene *cbiX* appeared to be missing. However, it was determined that in this strain a one-nucleotide-shorter spacer region resulted in a frameshift, which in turn led to the formation of a fusion gene of *cbiX* with the preceding gene *cbiH*. This result was confirmed by visual assessment of ten consensus sequence reads, arising from separate sequenced molecules from the PacBio assembly, that were aligned to the region. Nine out of these ten reads supported the observed deletion of a guanine base in the region, which causes the frameshift. In addition, *cbiX* and *cbiH* have 18 and 15 nucleotides variation in predicted sizes between strains,

respectively, pointing to the variable character of the spacer region.

The B12 biosynthetic pathway is known to be regulated at the translational level by the cobalamin riboswitches [49]. In *P. freudenreichii* three of those riboswitches have been found upstream of genes *cbiL*, *cbiB* and *mutA* [50]. The B12-riboswitches in *P. freudenreichii* are not well characterized and the actual span of the element is not known but all the elements are expected to possess the conserved B12 binding region termed the B12-box, which is characterized by a consensus sequence rAGYCMGgAgaCCkGCcd [50]. Based on the previous reports [49, 50], we retrieved the predicted sequences for the three putative riboswitches and compared them among strains. All the *P. freudenreichii* strains possess the expected three riboswitches, which are highly conserved between the strains, with the B12-box consensus for the species: SAGYCMSAMRMB-CYGCCD (Additional file 28). The actual effect of the riboswitches on the expression of the downstream genes is yet to be addressed. What may be of interest is that the riboswitches of the genes *cbiL* and *mutA* are located very close to each other, in opposite orientations, and can therefore interact.

Pili and mucus binding

The search for pilus gene clusters using the LOCP tool [51] identified putative pilus operons in the genomes of JS18, JS20 and JS14 consisting of three, four and five ORFs, respectively (Fig. 9a). The first genes of each operon (PFR_J18_2249, PFR_JS20-1_1986, PFR_JS14_352) are predicted to encode surface-anchored fimbrial subunits, whereas class C sortases are the predicted function of the last genes (PFR_J18_2247, PFR_JS20-1_1983, PFR_JS14_357) in these clusters. PFR_J18_2248, PFR_JS20-1_1985 and PFR_JS14_354 located in the middle of the operons are predicted to encode type-2 fimbrial major subunits, in JS18, JS20 and JS14, respectively. The putative pilus operons in JS14 and JS20 are similarly organized and the predicted pilus proteins share 99–94% amino acid identity. ORF prediction and location of functional domains in predicted proteins suggest that in the JS14 genome the genes coding for the putative surface-anchored fimbrial subunit and the Type-2 fimbrial major subunit have been split, possibly due to frameshifts causing mutations. A BLAST search revealed the presence of highly conserved gene clusters in the pilus operons of JS14 and JS20, with similar structural organisation, in all the *P. freudenreichii* genomes studied here. The exception is strain JS9, where only partial genes coding for the surface-anchored fimbrial subunit (PFR_JS9-1_404) and the sortase (PFR_JS9-1_414) are found with a genomic island (Unique gene cluster 1) inserted between them (see Additional file 7).

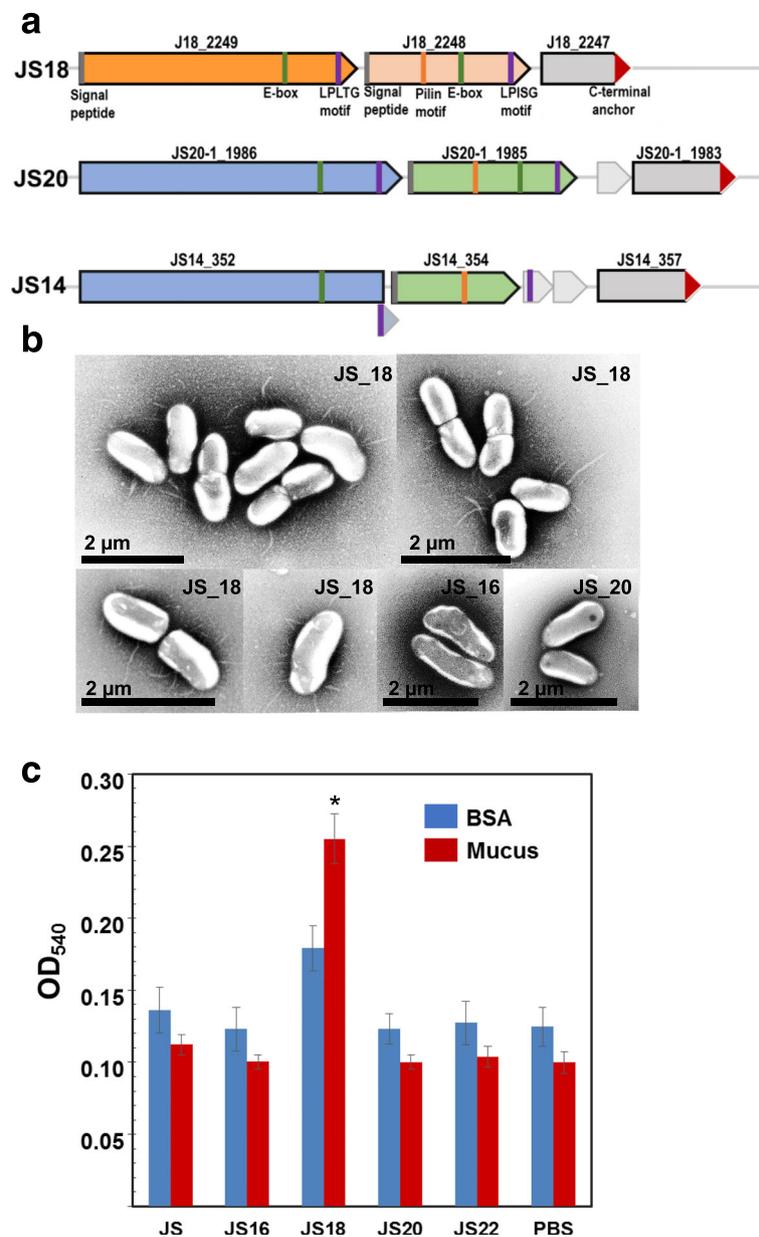


Fig. 9 Pilus and mucus binding of *P. freudenreichii*. **a** Pilus operons predicted by LOCP. **b** Transmission electron microscopy (TEM) images of the strains with intact operons (JS18 and JS20) and control (JS16). **c** Adhesion assay of specific binding to porcine mucus and non-specific binding to BSA with cell-free PBS as control. The difference in specific binding of strain JS18 to mucus compared to non-specific binding to BSA and background PBS was statistically significant ($p < 0.05$) and is marked with asterisk

In contrast, homology searches revealed that the PFR_J18_2249-PFR_J18_2248-PFR_J18_2247 operon located in a genomic island that is unique to the JS18 strain, since no counterpart of the intact operon was identified in the other genomes. Based on BLAST searches, the genome of JS9 carries genes encoding the putative surface-anchored fimbrial subunit (PFR_JS9-1_546) and the Type-2 fimbrial major subunit (PFR_JS9-1_547) with 100% identity to the gene products of JS18 genome, but the third gene encoding the

putative sortase is not present. The predicted pilus proteins encoded by JS18 operon share 32–54% identity with their counterparts encoded by the JS14/JS20 operon and BLASTp search against the NCBI non-redundant protein database revealed the highest amino acid identity (39–55%) with proteins of *Haematococcus sanguinis*.

Since in silico searches suggested that the genomes of JS18 and JS20 carry intact pilus operons, these strains were chosen for electron microscopic (EM) analyses

together with the *P. freudenreichii* type strain JS16. Transmission electron microscopic images of negatively stained cells showed that the surfaces of JS18 cells contain pilus-like appendages, which were not observed in JS20 and JS16 cells (Fig. 9b).

Since EM revealed pili-like structures on the surface of the JS18 cells and pili contributes to probiotic properties by mucin binding in other bacteria [52], we next tested the adherence of the JS18 strain to mucus and to bovine serum albumin. For comparison also strains JS, used previously in mucus adhesion assays [19, 20], JS16 (type strain), JS20 with intact pilus operon and JS22 showing variable adherence in biofilm assay were included in the experiment. The results revealed adherence ability in strain JS18, which unlike other *P. freudenreichii* strains shows more efficient binding to mucus than to bovine serum albumin (Fig. 9c).

Discussion

In this study we determined whole genome sequences of 17 new *P. freudenreichii* strains and re-sequenced whole genome of the strain DSM 4902. Using a comparative genomics approach, we identified several thus far unknown features of this species.

The comparative analyses revealed that despite genome-wide collinearity, large regions of inversions and other types of re-organisations are present even among the most closely related *P. freudenreichii* strains. This finding is of interest, as it was recently reported that in the closely related *Cutibacterium acnes* (previously *Propionibacterium acnes*) the gene synteny is highly conserved between strains [53]. These rearrangements could serve as an explanation for the observed adaptability and hardiness of *P. freudenreichii*, as the resulting genome instability was suggested as a driving force of adaptation and evolution in bacteria [54]. These genome reorganisations in *P. freudenreichii* appear to be transposon-mediated as transposase genes are found on the edges of many of the locally collinear blocks. The fact that in eight of the strains population diversity due to the translocation of transposase genes was observed suggests that transposons play an important role in genome plasticity in *P. freudenreichii* and explain this organism's ability to adapt to various environments.

The genomes of three active temperate bacteriophages were discovered as circular molecules from the *P. freudenreichii* strains studied here. Relatively little is known about phages infecting *Propionibacteria* in general and *P. freudenreichii* in particular. At the time of writing of this study there are ten complete bacteriophage genome sequences available: nine tailed phages belonging to the order Siphoviridae and one Inoviridae (filamentous) phage. The filamentous phage B5 [55, 56] and the tailed phages: B3, B22, E1, E6 and G4 [57, 58], isolated in

France from Swiss-type cheeses as well as previously uncharacterised Doucette and Anatole were sequenced in 2015 at the UCLA Genotyping and Sequencing Core Facility [59, 60]. The tailed PFR1 and PFR2 were induced directly from a Christian Hansen strain FD DVS PS-1 and a Swiss-type cheese purchased in Australia, respectively [61]. The phages identified in this study, namely PJS7, PJS22 and PJS23 are each the most similar to the previously reported PFR1, B22 and Doucette, respectively, but it is the first study reporting the presence of *P. freudenreichii* bacteriophage sequence present both as a prophage and also in free, circular form. In addition, an apparently complete prophage was identified from the strain JS17, but its circular form was not observed. In the previous study, the bacteriophage PFR1 was shown to infect *C. acnes* strain as well, while the PFR2 differing from PFR1 only by the presence of a transposable element could not [61]. The transposable element found in the PFR2 genome shares 100% DNA sequence identity to the PFR_JS12-1_615, PFR_JS15-1_2045 and PFR_JS17-1_657 found in strains JS12, JS15 and JS17, respectively and also to the prophage-encoded PFR_JS17-1_2067 of the strain JS17. This suggests that the insertion of the transposable elements intrinsic to the strain into the prophage-coding region could serve as a strategy for better control of the prophages.

At present, there are only a few published reports available concerning plasmids of *P. freudenreichii* [62–67]. Currently, four *P. freudenreichii* plasmid sequences are accessible at NCBI: p545 (AF291751.1), pRGO1 (NC_002611.1), pLME106 (NC_005705.1) and pLME108 (NC_010065.1). Here, we report sequences of two additional plasmid-like elements PFRJS12-3 and PFRJS25-1. It is noteworthy that the circular element PFRJS25-1 appears to be widespread among *P. freudenreichii* strains. We compared the recently published draft genomes [21, 23, 24] and detected sequences with 99% identity over 90% of the sequence of the putative plasmid PFRJS25-1 to the contigs in *P. freudenreichii* strains ITG P20 (CIRM-BIA 129) (NZ_HG975461 and CCBE010000014), ITG P1 (CCYV01000031), CIRM-BIA 135 (CCYZ01000006) and CIRM-BIA 1025 (NZ_CCYV01000031). Because of the limited size of the contigs, it is impossible to determine whether in these strains the elements were circular or existed as a part of the chromosome. We explored the possibility that the circular elements are a type of Integrative and Conjugative Elements (ICEs), which are widely distributed mobile genetic elements existing normally within the host's chromosome. Under certain conditions those elements can become activated, excise from the chromosome and transfer to a new recipient [68]. Although no such elements have been described

in *Propionibacteria* so far, they are widespread among other *Actinobacteria* in two types [69]. The FtsK/SpoIIIE-type relies on a single FtsK/SpoIIIE-like protein for DNA translocation, while the T4SS-type requires an assembly of a complex type IV translocation system for mobility [69]. The previously mentioned Unique gene clusters 4,9,11,14 and 16 (see Fig. 4), found in strains JS12, JS18, JS20, JS7 and JS9, respectively, share structural similarities indicative of ICEs type T4SS that are also present in PFRJS12–3 and PFRJS25–1. Thus, it is possible that the extra-chromosomal circular DNA elements in JS12 and JS25 represent mobilised ICEs instead of plasmids. Nevertheless, further studies are necessary to determine the true nature of these new, *P. freudenreichii*-specific elements.

Complete CRISPR-Cas systems were identified in 15 out of 20 sequenced strains and for the first time they were classified for *P. freudenreichii* as type IE, found in three strains and a novel type IU, found in 13 strains. The activity of CRISPR-Cas systems in *P. freudenreichii* should be addressed in further studies.

Very little is known about restriction-modification (RM) systems in *P. freudenreichii*. It has been shown that such systems are present based on the observed interference with transformation efficiencies [66] and the host range dependent on the compatibility of the RM systems between the sources and intended hosts of the plasmids [65]. In this study, the most striking feature of the RM systems distributed among the *P. freudenreichii* strains was the variability of the systems and the genomic locations found from one strain to another. This contrasts with the more usual situation where, at the very least, there are one or more common methylases found in all strains of a particular species (see REBASE [70]).

Pili or fimbriae are surface adhesive components and well-documented virulence factors for many harmful opportunistic pathogens [71]. On the other hand, the role of these non-flagellar proteinaceous hair-like fiber appendages in probiotics and/or commensal bacteria and as niche-adaptation factors has only recently been recognized. Among *Actinobacteria*, pili have been reported to be crucial for establishing both host-microbe and microbe-microbe interactions in probiotic *Bifidobacteria* [72, 73], while in *Propionibacterium* pili have not been described. Here, we identified a unique pilus operon from a genomic island in a *P. freudenreichii* strain and show that cells of this strain are decorated by pili-resembling structures. Furthermore, the strain with pili-like appendages was capable of specific mucus binding, while *P. freudenreichii* strains generally bind similarly to mucus and bovine serum albumin through non-specific interactions [19, 20].

Conclusions

The whole genome alignments showed that, despite genome-wide collinearity, large regions of transposon-mediated inversions and other types of re-organisation are present in *P. freudenreichii* genomes. The fact that in eight of the strains we observed population diversity due to the translocation of transposase genes suggests that transposable elements play an important role in genome plasticity in *P. freudenreichii* and explain this organism's ability to adapt to various environments, while the additional role of the transposons in control of the prophages and CRISPR-Cas systems needs to be explored further.

The utilization of long read technology enabled us to correctly assemble the genomic elements, such as phages, that were found both within and outside the genomes. These separate sub-populations could have been missed with short read data, even when using mate-pair or other long fragment-based methods. The long reads additionally enabled detailed analysis of the CRISPR arrays and allowed characterisation and classification of the CRISPR-Cas systems. The use of the PacBio sequencing platform, which detects methylation patterns, also permitted the detection of potentially active restriction-modification systems through matching of the recognition motifs with the methylases responsible. Many of the recognition sequences of these RM systems identified in this study are being reported for the first time. Finally, we report the first evidence of a *P. freudenreichii* strain being decorated by pili appendages and showing specific mucus binding. Taken together, the whole genome sequencing from long reads proved to be a useful method for improving the characterisation of *P. freudenreichii* by allowing the discovery of previously uncharted territories for the species. The amassed data provides a firm foundation for further, more in-depth studies of the species.

Methods

Bacterial growth and extraction of DNA

The strains were grown in propionic medium (PPA) [19] or the whey-based liquid medium (WBM) [48]. The PPA composition was: 5.0 g tryptone (Sigma-Aldrich), 10.0 g yeast extract (Becton, Dickinson), 14.0 ml 60% w/w DL-sodium lactate (Sigma-Aldrich) per liter, with pH adjusted to 6.7 prior to autoclaving. The industrial-type medium, WBM, was composed of 60.0 g of filtered whey powder (Valio Ltd., Finland), 10.0 g of yeast extract (MERCK, KGaA), 0.1 g Tween 80 (MERCK, KGaA), 0.2 g magnesium sulphate (MERCK, KGaA), 0.05 g manganese (II) sulphate (MERCK, KGaA), 100 mM potassium phosphate buffer (MERCK, KGaA) and was prepared as previously described [48].

For the phenotypic tests the strains were grown in YEL [74] medium composed of 10 g of tryptone (Sigma-Aldrich), 10 g yeast extract (Becton Dickinson), 16.7 g of

60% w/w DL-sodium lactate (Sigma-Aldrich), 2.5 g K₂HPO₄, 0.005 g MnSO₄.

The cultures were prepared from 15% glycerol stocks stored at -80 °C by streaking on a PPA agar plate and incubation at 30 °C in anaerobic jars (Anaerocult, Merck, Germany) for 4 days, unless stated otherwise. For the preparation of liquid cultures, colonies from the plate were picked and transferred to 15 mL Falcon tubes containing 10 mL of the liquid medium.

For the DNA extraction, the cells were harvested from liquid cultures incubated for 72 h by centrifugation for 5 min at 21000 g at 4 °C and washed with 0.1 M TRIS pH 8.0. The DNA extraction was performed with ILLUSTRATE™ bacteria genomicPrep Mini Spin Kit (GE Healthcare) with 10 mg of lysozyme and the incubation time 30 min.

Genome sequencing and assembly

Nineteen *Propionibacterium freudenreichii* samples were sequenced with the Pacific Biosciences RS II Instrument using either P4/C2 or P5/C3 chemistries (listed in Table 2). Two SMRT cells were used for each sample. Movie times varied from 120 to 240 min. The total number of obtained bases and subreads and their mean and N50 lengths are listed in Table 2. The Hierarchical Genome Assembly Process (HGAP) V3 implemented in the SMRT Analysis package (v.2.3.0) was used to generate de novo genome assemblies with default parameters, excluding the genome size estimate parameter which was set to 3,000,000 bp. Obtained circular sequences were polished using SMRT Analysis RS Resequencing protocol and the Quiver consensus algorithm. Chromosomal genome sequences were set to begin from the chromosomal replication initiator protein (dnaA). The sequences were then annotated with Prodigal v. 2.6.2. All sequences were deposited in the European Nucleotide Archive (ENA). Complete genome, phage and plasmid sequence sizes, sequencing coverages, GC percentages, number of predicted genes and ENA accession numbers are listed in Table 1. Base modifications and motifs were detected using RS Modification and Motif analysis protocol (SMRT Analysis package v.2.3.0).

Bioinformatics analyses

Average nucleotide identities (ANI) were calculated using JSpecies V1.2.1 [75]. Genome organizations were visualized with the Mauve alignment tool using the Progressive Mauve algorithm [76] with GenBank input files generated by conversion of the EMBL files obtained after the submission of genomes to the European Nucleotide Archive (ENA) with the use of Seqret, a part of the EMBOSS package. Like many other packages used in this study EMBOSS was a part of the BioLinux 8 workstation [77].

The core and pan-genome was estimated with Roary [32]) at standard settings with GFF3 annotation files generated by PROKKA [78] used as input files. The genomic islands were detected with the help of IslandViewer 3 software [35]. For the prediction of prophages, Prophinder [79] and Phaster [80] online software were used. Predicted bacteriophages were then reviewed visually for structural completeness.

Prediction of CRISPR loci was aided by the CRISPR-Finder [81]. The obtained results were then reviewed manually for co-localisation with Cas genes. The immunity to known bacteriophages was tested by searching the spacer sequences against NCBI tailed bacteriophages and also whole nucleotide collection excluding Propionibacteria (txid1743) with the aid of the BLASTn suite. Restriction-Modification systems were identified using SEQWARE [43] and the REBASE database [70] followed by manual matching of methylase genes with predicted recognition sequences against the methylation profiles generated by PacBio sequencing. In some cases, matches could be unambiguously inferred when only a single methylase gene and a single motif were present or left unmatched. Automatic pilus cluster search was performed using LOCP v. 1.0.0 [51]. The results were visualised with iTOL [82], Phandango 0.8.5 [83], EasyFig [84] and PigeonCad [85].

PCR reactions

All of the PCR reactions were performed with the Phusion (ThermoFisher Scientific) mastermix with 0.3% DMSO and the primers prepared by Oligomer Oy (Helsinki, Finland). The results were visualised by 0.8% agarose (BioRad) gel electrophoresis with ethidium bromide (0.5 µg/ml) (Sigma-Aldrich) staining.

Phage integration analysis The phage detected in the strain JS7 was found both in a free, circular form as well as integrated into the chromosome as was the case for strains JS22 and JS23. However, in strain JS7 a phage-free bacterial genome was detected as well, which allowed us to study the dynamics of phage integration and release from the chromosome. For that purpose, PCRs were designed to amplify the regions of both the region of phage integration in the bacterial chromosome as well as the attachment sites on the bacteriophage genome. The primers used were: PB5 CGCATACGCAGA TATTAAG complementary to the 5' end of the gene coding for the KmtR transcriptional regulator, PB6 GAGGTGCTGGCGGATAC complementary to the 3' end of the transcriptional regulator located just downstream of the CmtR regulator-coding gene in the phage-free chromosome, PB7 CTTCCCGCAGTGTCTTG and PB8 GAAGCAGGGCGTTTATG both complementary to the phage-encoded Integrase. The reaction mix PB5 and

PB6 would therefore detect a phage-free bacterial chromosome with the 815 nt long product; PB6 and PB8 would detect a chromosome with the phage integrated with the 691 nt product; PB7 and PB8 would detect a circular phage with the 850 nt product. The reactions were performed on bacterial cells picked from separate colonies grown on PPA agar for 4 days in three separate experiments. For one set of colonies PCRs were repeated on after 7 days. The same colonies were then picked and propagated every 7 days for 10 generations. The PCR was repeated after 5 and after 10 generations.

Duplication in the strain JS17 In order to eliminate the possibility that the duplication observed in the strain JS17 is a result of a sequencing error we analysed the edges of the region by PCR. The primers used were: PD6 fwd CTGGTTGCGTCATCTCTAAGCCT, PD7 rev CGCTCTTTTAGGGAATCGCTCAT and PD8 fwd TCTTCTTCTGTACGCGTGACAT. The PCRs were performed with PD6 and PD7, PD8 and PD6 and also single PD6 and PD8 as negative controls. Because of the high melting temperatures of the primers, the PCRs were performed with a 2-step protocol (annealing and extension combined at 72 °C for 1:30 min). The reactions were deemed positive when the products of sizes 1888 nt for the PD6-PD7 and 1738 nt for PD7-PD8 were seen on the 0.8% agarose gel.

Electron microscopy for detection of pili The strains were grown on YEL agar at 30 °C for 7 days under near-anaerobic atmosphere (Anaerocult, Merck) Single colonies were picked and suspended in 0.1 M PIPES buffer, pH 6.8. An aliquot of 3 µl was added to Pioloform-coated 200-mesh copper grids previously glow discharged (Emitech K100X, Emitech Ltd., UK) to ensure even adhesion of the bacterial cells to the grids. After 1 min incubation, the excess suspension fluid was removed by soaking with a filter paper. The grids were negatively stained with 1% neutral uranyl acetate for 15 s and air-dried. Images were acquired at 120 kV with a Jeol JEM-1400 microscope (Jeol Ltd., Tokyo, Japan) using an Orius SC 1000B CCD-camera (Gatan Inc., USA).

Mucus adhesion assay Adhesion of the *P. freudenreichii* strains (JS, JS16, JS18, JS20 and JS22) on immobilized porcine mucin (Sigma-Aldrich) in 96-well Polysorp microplates (Nunc Immuno plates, Nunc, Denmark) was conducted according to Lecesse Terraf [86] with the following modifications. Shortly, plates were covered with 300 µl of 0.2 mg ml⁻¹ mucin in phosphate buffered saline (PBS, pH 7.5) (Thermo Fischer Scientific) and incubated for 30 min at 37 °C (250 rpm, PST-60HL Thermo-Shaker (Biosan) and then overnight at 4 °C. Wells were

washed twice with PBS and blocked with 1% bovine serum albumin (BSA) in PBS, at room temperature. Replica microplates treated with PBS without mucin and then 1% BSA as above were also prepared to exclude BSA-specific binding. After 2 h incubation, mucin- and BSA-coated wells were washed twice with PBS and allowed to dry. For adhesion, 200 µl of cells suspended in PBS (OD₆₀₀ = 2.0) from each strain was added into the mucus- and BSA-coated wells, and plates were incubated at 37 °C for 2 h (250 rpm). PBS alone was included in all experiments to subtract mucin binding by PBS. Non-adherent cells were removed and wells were washed with PBS. Then, adherent cells were stained with 200 µL of the crystal violet solution (0.1%, w/v) (Sigma-Aldrich, Munich, Germany) for 30 min at room temperature. Excess stain was washed off with deionized H₂O and the stained cells were suspended in 30% acetic acid by shaking (400 rpm) at room temperature, and recorded at 540 nm using an ELISA reader (Labsystems Multiskan EX). Up to four independent experiments were performed, each with at least sixteen technical replicates.

Significant differences between sample means of independent experiments and the mean of PBS as test value were determined by one-sample t-test. *P* values <0.05 were considered statistically significant. The calculations were performed with statistical package program (IBM SPSS Statistics v24 for Windows, IBM, USA).

Additional files

Additional file 1: Gene-presence-absence table generated by ROARY analysis. (XLSX 893 kb)

Additional file 2: Annotation file of the strain JS with the cumulative results of bioinformatics analyses. (XLSX 331 kb)

Additional file 3: Annotation file of the strain JS2 with the cumulative results of bioinformatics analyses. (XLSX 346 kb)

Additional file 4: Annotation file of the strain JS4 with the cumulative results of bioinformatics analyses. (XLSX 401 kb)

Additional file 5: Annotation file of the strain JS7 (LT618776) and of the phage PJS7 (LT618778) with the cumulative results of bioinformatics analyses and the results of the phage integration study. (XLSX 360 kb)

Additional file 6: Annotation file of the strain JS8 with the cumulative results of bioinformatics analyses. (XLSX 329 kb)

Additional file 7: Annotation file of the strain JS9 with the cumulative results of bioinformatics analyses. (XLSX 406 kb)

Additional file 8: Annotation file of the strain JS10 with the cumulative results of bioinformatics analyses. (XLSX 315 kb)

Additional file 9: Annotation file of the strain JS11 with the cumulative results of bioinformatics analyses. (XLSX 298 kb)

Additional file 10: Annotation file of the strain JS12 (LT604998) and of the putative conjugative plasmid PFRJS12-3 (LT604882) with the cumulative results of bioinformatics analyses. (XLSX 386 kb)

Additional file 11: Annotation file of the strain JS13 with the cumulative results of bioinformatics analyses. (XLSX 307 kb)

Additional file 12: Annotation file of the strain JS14 with the cumulative results of bioinformatics analyses. (XLSX 304 kb)

Additional file 13: Annotation file of the strain JS15 (DSM 4049) with the cumulative results of bioinformatics analyses. (XLSX 323 kb)

Additional file 14: Annotation file of the strain JS16 (DSM 20271) with the cumulative results of bioinformatics analyses. (XLSX 357 kb)

Additional file 15: Annotation file of the strain JS17 with the cumulative results of bioinformatics analyses. (XLSX 391 kb)

Additional file 16: Annotation file of the strain JS18 with the cumulative results of bioinformatics analyses. (XLSX 379 kb)

Additional file 17: Annotation file of the strain JS20 with the cumulative results of bioinformatics analyses. (XLSX 413 kb)

Additional file 18: Annotation file of the strain JS21 with the cumulative results of bioinformatics analyses. (XLSX 335 kb)

Additional file 19: Annotation file of the strain JS22 (LT599498) and of the phage PJS22 (LT615138) with the cumulative results of bioinformatics analyses. (XLSX 347 kb)

Additional file 20: Annotation file of the strain JS23 (LT618794) and of the phage PJS23 (LT618793) with the cumulative results of bioinformatics analyses. (XLSX 327 kb)

Additional file 21: Annotation file of the strain JS25 (LT618783) and of the putative conjugative plasmid PFR_JS25-1 (LT618784) with the cumulative results of bioinformatics analyses. (XLSX 336 kb)

Additional file 22: Phenotypic characterization. The characteristics of the studied strains were assessed and included: carbohydrate utilization patterns, nitroreductase activity as well as growth and biofilm formation in various growth conditions. (DOCX 612 kb)

Additional file 23: Traits important in food production. Genes coding for traits found as important in food production are explored. (PDF 8 kb)

Additional file 24: The CRISPR spacers within CRISPR arrays identified in *P. freudenreichii*. The collection of all CRISPR spacers predicted in the *P. freudenreichii* strains together with their predicted immunity. (PDF 75 kb)

Additional file 25: Self-immunity: the table of spacers conferring self-immunity; Candidate proteins: the summary of the candidate proteins for Anti-CRISPR elements on the invading DNA elements. (XLSX 13 kb)

Additional file 26: Summary schematics of the RM systems in *P. freudenreichii* strains. (PDF 1680 kb)

Additional file 27: Protein sequence alignments of the proteins involved in biosynthesis of vitamin B12. (XLSX 131 kb)

Additional file 28: The alignment of the three predicted B12 riboswitches with the conserved B12-box highlighted. (PDF 28 kb)

Abbreviations

ANI: Average nucleotide identity; BSA: Bovine serum albumin; CAS: CRISPR associated proteins; CDD: Conserved Domains Database; CRISPR: Clustered regularly interspaced short palindromic repeats; ENA: European Nucleotide Archive; GRAS: Generally Recognized as Safe; HGAP3: Hierarchical Genome Assembly Process; ICEs: Integrative and Conjugative Elements; OD₆₀₀: Optical density measured at 600 nm; PBS: Phosphate buffered saline; PPA: Propionic medium; RM: Restriction-modification; SCFAs: Short-chain fatty acids

Acknowledgements

We thank Harri Kangas for PacBio sequencing. The authors wish to acknowledge CSC - IT Center for Science, Finland, for computational resources. We thank Dr. Mirko Rossi for the advice on selection of bioinformatics tools and for the help with PROKKA and Roary analyses. We acknowledge Dr. Patrik Koskinen for his initial involvement in the project, and especially for annotating the genomes with Prodigal. We thank Pasi Perkiö for the statistical analysis of the mucus binding assay and for the help with Excel-based managing of the CRISPR spacers.

Funding

The study was funded by the Academy of Finland (grant numbers 257,333, 272,363 and 267,755).

Availability of data and materials

All genome sequences have been deposited in the European Nucleotide Archive with accession numbers: LT576032, LT576033, LT576034, LT576035, LT576038, LT576042, LT576787, LT593929, LT599498, LT604882, LT604998, LT615138, LT618776, LT618777, LT618778, LT618779, LT618780, LT618781, LT618782, LT618785, LT618786, LT618787, LT618788, LT618789, LT618790, LT618791, LT618792, LT618793, LT618794, LT61878, LT618784. Datasets used for phylogenetic analyses are available from the Dryad Digital Repository (<https://doi.org/10.5061/dryad.kg5s5>). The datasets generated and/or analysed during the current study are available from the corresponding authors on reasonable request.

Authors' contributions

PV, KS, PA, PD and VP designed and conceived the study. PD performed majority of the experiments. KS, RJR, and HV designed and performed individual experiments. LP organized genome sequencing. PKL and O-PS assembled the genomes and participated in the bioinformatic analyses. PV, PA, EJ, and VP provided materials and facilities. PD, PV, RJR, PKL and KS wrote the manuscript with input from PA, O-PS, and LP. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

RJR works for a company that sells restriction enzymes and DNA methylases. The other authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Food and Environmental Sciences, University of Helsinki, 00014 Helsinki, Finland. ²Institute of Biotechnology, University of Helsinki, 00014 Helsinki, Finland. ³New England Biolabs, Ipswich, MA 01938-2723, USA.

Received: 7 June 2017 Accepted: 5 October 2017

Published online: 16 October 2017

References

1. Thierry A, Deutsch S-M, Falentin H, Dalmaso M, Cousin FJ, Jan G. New insights into physiology and metabolism of *Propionibacterium freudenreichii*. *Int J Food Microbiol.* 2011;149:19–27.
2. Wood HG. Metabolic cycles in the fermentation by propionic acid bacteria. *biol. Cycles. Curr Top Cell Regul.* 1981; <https://doi.org/10.1016/B978-0-12-152818-8.50021-9>
3. Vorobjeva L. Economic and medical applications. In: *Propionibacteria*. Springer Netherlands; 1999. p. 209–43.
4. Zárte G. Dairy Propionibacteria: Less conventional probiotics to improve the human and animal health. In: *probiotic in animals*. Intech 2012. Chapter 8, p.153–202.
5. Kasubuchi M, Hasegawa S, Hiramatsu T, Ichimura A, Kimura I. Dietary gut microbial metabolites, short-chain fatty acids, and host metabolic regulation. *Nutrients.* 2015;7:2839–49.
6. Chambers ES, Viardot A, Psichas A, Morrison DJ, Murphy KG, Zac-Varghese SE, et al. Effects of targeted delivery of propionate to the human colon on appetite regulation, body weight maintenance and adiposity in overweight adults. *Gut.* 2015;64:1744–54.
7. Psichas A, Sleeth ML, Murphy KG, Brooks L, Bewick GA, Hanyaloglu AC, et al. The short chain fatty acid propionate stimulates GLP-1 and PYY secretion via free fatty acid receptor 2 in rodents. *Int J Obes.* 2015;39:424–9.
8. Nastasi C, Candela M, Bonefeld CM, Geisler C, Hansen M, Krejsgaard T, et al. The effect of short-chain fatty acids on human monocyte-derived dendritic cells. *Sci Rep.* 2015;5:16148.
9. Cousin FJ, Jouan-Lanhouet S, Theret N, Brenner C, Jouan E, Le Moigne-Muller G, et al. The probiotic *Propionibacterium freudenreichii* as a new

- adjuvant for TRAIL-based therapy in colorectal cancer. *Oncotarget*. 2016;7:7161–78.
10. Wang LM, Lv JP, Chu ZQ, Cui YY, Ren XH. Production of conjugated linoleic acid by *Propionibacterium freudenreichii*. *Food Chem*. 2007;103:313–8.
 11. LeBlanc JG, Rutten G, Bruinenberg P, Sesma F, de Giori GS, Smid EJA. Novel dairy product fermented with *Propionibacterium freudenreichii* improves the riboflavin status of deficient rats. *Nutrition*. 2006;22:645–51.
 12. Hojo K, Watanabe R, Mori T, Taketomo N. Quantitative measurement of tetrahydromenaquinone-9 in cheese fermented by propionibacteria. *J Dairy Sci*. 2007;90:4078–83.
 13. Burgess CM, Smid EJ, van Sinderen D. Bacterial vitamin B2, B11 and B12 overproduction: an overview. *Int J Food Microbiol*. 2009;133:1–7.
 14. Deptula P, Chamlagain B, Edelmann M, Sangsuwan P, Nyman TA, Savijoki K, et al. Food-like growth conditions support production of active vitamin B12 by *Propionibacterium freudenreichii* 2067 without DMBI, the lower Ligand Base, or cobalt supplementation. *Front Microbiol*. 2017;8:368.
 15. Nordmark E-L, Yang Z, Huttunen E, Widmalm G. *Propionibacterium freudenreichii*. *Biomacromolecules*. 2005;6:521–3.
 16. Cardoso FS, Castro RF, Borges N, Santos H. Biochemical and genetic characterization of the pathways for trehalose metabolism in *Propionibacterium freudenreichii*, and their role in stress response. *Microbiology*. 2007;153:270–80.
 17. Altieri C. Dairy propionibacteria as probiotics: recent evidences. *World J Microbiol Biotechnol*. 2016;32:172.
 18. van Tassell ML, Miller MJ. *Lactobacillus* adhesion to mucus. *Nutrients*. 2011;3:613–36.
 19. Ouwehand AC, Suomalainen T, Tölkö S, Salminen S. In vitro adhesion of propionic acid bacteria to human intestinal mucus. *Lait*. 2002;82:123–30.
 20. Collado MC, Meriluoto J, Salminen S. Development of new probiotics by strain combinations: is it possible to improve the adhesion to intestinal mucus? *J Dairy Sci*. 2007;90:2710–6.
 21. Falentin H, Deutsch SM, Jan G, Loux V, Thierry A, Parayre S, et al. The complete genome of *Propionibacterium freudenreichii* CIRM-BIA1T, a hardy actinobacterium with food and probiotic applications. *PLoS One*. 2010;5:e11748.
 22. Scholz CFP, Kilian M. The natural history of cutaneous propionibacteria, and reclassification of selected species within the genus *Propionibacterium* to the proposed novel genera *Acidipropionibacterium gen. nov.*, *Cutibacterium gen. nov.* and *Pseudopropionibacterium gen. nov.* *Int J Syst Evol Microbiol*. 2016;66:4422–32.
 23. Falentin H, Deutsch S-M, Loux V, Hammani A, Buratti J, Parayre S, et al. Permanent draft genome sequence of the probiotic strain *Propionibacterium freudenreichii* CIRM-BIA 129 (ITG P20). *Stand Genomic Sci*. 2016;11:6.
 24. Loux V, Mariadassou M, Almeida S, Chiapello H, Hammani A, Buratti J, et al. Mutations and genomic islands can explain the strain dependency of sugar utilization in 21 strains of *Propionibacterium freudenreichii*. *BMC Genomics*. 2015;16:296.
 25. Koskinen P, Deptula P, Smolander O-P, Tamene F, Kammonen J, Savijoki K, et al. Complete genome sequence of *Propionibacterium freudenreichii* DSM 20271T. *Stand Genomic Sci*. 2015;10:83.
 26. Ojala T, Laine PKS, Ahlroos T, Tanskanen J, Pitkänen S, Salusjärvi T, et al. Functional genomics provides insights into the role of *Propionibacterium freudenreichii* ssp. *shermanii* JS in cheese ripening. *Int J Food Microbiol*. 2017;241:39–48.
 27. Aburjaile FF, Madec M-N, Parayre S, Miyoshi A, Azevedo V, Le Loir Y, et al. The long-term survival of *Propionibacterium freudenreichii* in a context of nutrient shortage. *J Appl Microbiol*. 2016;120:432–40.
 28. de Freitas R, Madec MN, Chuat V, Maillard MB, Mukdsi MCA, Falentin H, et al. New insights about phenotypic heterogeneity within *Propionibacterium freudenreichii* argue against its division into subspecies. *Dairy Sci Technol*. 2015;95:465–77.
 29. Deutsch S-M, Mariadassou M, Nicolas P, Parayre S, Le Guellec R, Chuat V, et al. Identification of proteins involved in the anti-inflammatory properties of *Propionibacterium freudenreichii* by means of a multi-strain study. *Sci Rep*. 2017;7:46409.
 30. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods*. 2013;10:563–9.
 31. Treangen TJ, Abraham AL, Touchon M, Rocha EPC. Genesis, effects and fates of repeats in prokaryotic genomes. *FEMS Microbiol Rev*. 2009;33:539–71.
 32. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*. 2015;31:3691–3.
 33. Kasimatis G, Fitz-Gibbon S, Tomida S, Wong M, Li H. Analysis of complete genomes of propionibacterium acnes reveals a novel plasmid and increased pseudogenes in an acne associated strain. *Biomed Res Int*. 2013;2013:918320.
 34. Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res*. 2016;45:1129.
 35. Dhillon BK, Laird MR, Shay JA, Winsor GL, Lo R, Nizam F, et al. IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. *Nucleic Acids Res*. 2015;43:W104–8.
 36. Le Maréchal C, Peton V, Plé C, Vroland C, Jardin J, Briard-Bion V, et al. Surface proteins of *Propionibacterium freudenreichii* are involved in its anti-inflammatory properties. *J Proteome*. 2015;113:447–61.
 37. Savitskaya EE, Musharova OS, Severinov KV. Diversity of CRISPR-Cas-mediated mechanisms of adaptive immunity in prokaryotes and their application in biotechnology. *Biochemistry (Mosc)*. 2016;81:653–61.
 38. Koonin EV, Makarova KS. CRISPR-Cas: evolution of an RNA-based adaptive immunity system in prokaryotes. *RNA Biol*. 2013;10:679–86.
 39. Jore MM, Lundgren M, van Duijn E, Bultema JB, Westra ER, Waghmare SP, et al. Structural basis for CRISPR RNA-guided DNA recognition by Cascade. *Nat Struct Mol Biol*. 2011;18:529–36.
 40. Makarova KS, Aravind L, Wolf YI, Koonin EV. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct*. 2011;6:38.
 41. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*. 2015;13:722–36.
 42. Rauch BJ, Silvis MR, Hultquist JF, Waters C, McGregor MJ, Krogan NJ, et al. Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. *Cell*. 2017;150–8. <https://doi.org/10.1016/j.cell.2016.12.009>
 43. Murray IA, Clark TA, Morgan RD, Boitano M, Anton BP, Luong K, et al. The methylomes of six bacteria. *Nucleic Acids Res*. 2012;40:11450–62.
 44. Flusberg BA, Webster DR, Lee JH, Travers KJ, Olivares EC, Clark TA, et al. Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods*. 2010;7:461–5.
 45. Murakami K, Hashimoto Y, Murooka Y. Cloning and characterization of the gene encoding glutamate 1-Semialdehyde 2, 1-aminomutase, which is involved in 5-Aminolevulinic acid synthesis in *Propionibacterium freudenreichii*. *Appl Environ Microbiol*. 1993;59:347–50.
 46. Roessner CA, Huang KX, Warren MJ, Raux E, Scott AI. Isolation and characterization of 14 additional genes specifying the anaerobic biosynthesis of cobalamin (vitamin B12) in *Propionibacterium freudenreichii* (*P. shermanii*). *Microbiology*. 2002;148:1845–53.
 47. Sattler I, Roessner CA, Stolowich NJ, Hardin SH, Harrishaller LW, Yokubaitis NT, et al. Cloning, sequencing, and expression of the uroporphyrinogen-III methyltransferase cobA gene of *Propionibacterium freudenreichii* (*shermanii*). *J Bacteriol*. 1995;177:1564–9.
 48. Chamlagain B, Deptula P, Edelmann M, Kariluoto S, Grattepanche F, Lacroix C, et al. Effect of the lower ligand precursors on vitamin B12 production by food-grade Propionibacteria. *LWT - Food Sci Technol*. 2016;72:117–24.
 49. Nahvi A, Barrick JE, Breaker RR. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res*. 2004;32:143–50.
 50. Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA*. 2003;9:1084–97.
 51. Plyusnin I, Holm L, Kankainen M. LOCP - Locating pilus operons in gram-positive bacteria. *Bioinformatics*. 2009;25:1187–8.
 52. Kankainen M, Paulin L, Tynkkynen S, von Ossowski I, Reunanen J, Partanen P, et al. Comparative genomic analysis of *Lactobacillus rhamnosus* GG reveals pili containing a human- mucus binding protein. *Proc Natl Acad Sci U S A*. 2009;106:17193–8.
 53. Scholz CFP, Brüggemann H, Lomholt HB, Tettelin H, Kilian M. Genome stability of *Propionibacterium acnes*: a comprehensive study of indels and homopolymeric tracts. *Sci Rep*. 2016;6:20662.
 54. Darmon E, Leach DRF. Bacterial genome instability. *Microbiol Mol Biol Rev*. 2014;78:1–39.
 55. Chopin MC, Rouault A, Dusko Ehrlich S, Gautier M. Filamentous phage active on the gram-positive bacterium *Propionibacterium freudenreichii*. *J Bacteriol*. 2002;184:2030–3.

56. Gautier M, Rouault A, Hervé C, Sommer P, Leret V, Jan G, et al. Bacteriophages of dairy propionibacteria. *Lait*. 1999;79:93–104.
57. Gautier M, Rouault A, Lortal S, Leloir Y, Patureau D. Original article characterization of a phage infecting *Propionibacterium freudenreichii*. *J Dairy Sci*. 1992;72:431–5.
58. Gautier M, Rouault A, Sommer P, Briandet R. Occurrence of *Propionibacterium freudenreichii* bacteriophages in Swiss cheese. *Appl Environ Microbiol*. 1995;61:2572–6.
59. Modlin RL, Cheng LS, Marinelli LJ, Grosset N, Gautier M, Fitz-Gibbon S, Pellegrini M, Bowman CA, Russell DA, Jacobs-Sera D, Hatfull GF. *Propionibacterium* phage Doucette, complete genome - Nucleotide - NCBI. 2016. <https://www.ncbi.nlm.nih.gov/nuccore/1072745899>. Accessed 23 May 2017.
60. Modlin RL, Cheng LS, Marinelli LJ, Grosset N, Gautier M, Fitz-Gibbon S, Pellegrini M, Bowman CA, Russell DA, Jacobs-Sera D, Hatfull G. *Propionibacterium* phage Anatole, complete genome - Nucleotide - NCBI. 2016. <https://www.ncbi.nlm.nih.gov/nuccore/1072745723>. Accessed 23 May 2017.
61. Brown T L, Tucci J, ZA D, Lock P, Adda CG, Petrovski S. Dynamic interactions between prophages induce lysis in *Propionibacterium acnes*. *Res Microbiol*. 2016;168:103–12.
62. Perez Chaia A, Sesma F, Pesce de Ruiz Holgado A, Oliver G. Screening of plasmids in strains of propionibacterium and mesophilic lactobacilli isolated from Swiss type cheeses. *M.A.N. Microbiol. Aliment. Nutr*. 1983;6:171–4.
63. Rehberger TG, Glatz BA. Characterization of *Propionibacterium* plasmids. *Appl Environ Microbiol*. 1990;56:864–71.
64. Dasen GH. Molecular identification and applied genetics of *Propionibacteria*. PhD thesis. 1998.
65. Kiatpapan P, Hashimoto Y, Nakamura H, Piao YZ, Ono H, Yamashita M, et al. Characterization of pRGO1, a plasmid from *Propionibacterium acidipropionici*, and its use for development of a host-vector system in propionibacteria. *Appl Environ Microbiol*. 2000;66:4688–95.
66. Jore JPM, Van Lwijk N, Luiten RGM, Van der Werf MJ, Pouwels PH. Efficient transformation system for *Propionibacterium freudenreichii* based on a novel vector. *Appl Environ Microbiol Am Soc Microbiol*. 2001;67:499–503.
67. Stierli MP. DNA transformation of *Propionibacteria* based on plasmids pLME106 and pLME108. PhD thesis. 2002.
68. Delavat F, Mitri S, Pelet S, van der Meer JR. Highly variable individual donor cell fates characterize robust horizontal gene transfer of an integrative and conjugative element. *Proc Natl Acad Sci U S A*. 2016;113:1–9.
69. Ghinet MG, Bordeleau E, Beaudin J, Brzezinski R, Roy S, Burrus V. Uncovering the prevalence and diversity of integrating conjugative elements in actinobacteria. *PLoS One*. 2011;6:e27846.
70. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*. 2015;43:D298–9.
71. Proft T, Baker EN. Pili in gram-negative and gram-positive bacteria - structure, assembly and their role in disease. *Cell Mol Life Sci*. 2009;613–35.
72. O'Connell Motherway M, Zomer A, Leahy SC, Reunanen J, Bottacini F, Claesson MJ, et al. Functional genome analysis of *Bifidobacterium breve* UCC2003 reveals type IVb tight adherence (tad) pili as an essential and conserved host-colonization factor. *Proc Natl Acad Sci U S A*. 2011;108:11217–22.
73. Turrone F, Serafini F, Foroni E, Duranti S, O'Connell Motherway M, Taverniti V, et al. Role of sortase-dependent pili of *Bifidobacterium bifidum* PRL2010 in modulating bacterium-host interactions. *Proc Natl Acad Sci U S A*. 2013;110:11151–6.
74. Malik A, Reinbold G. An evaluation of the taxonomy of *Propionibacterium*. *Can J*. 1968;14:1185–91.
75. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci U S A*. 2009;106:19126–31.
76. Darling AE, Mau B, Perna NT. Progressivemaue: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 2010;5:e11147.
77. Field D, Tiwari B, Booth T, Houten S, Swan D, Bertrand N, et al. Open software for biologists: from famine to feast. *Nat Biotechnol*. 2006;24:801–3.
78. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30:2068–9.
79. Lima-Mendez G, Van Helden J, Toussaint A, Leplae R. Prophinder: a computational tool for prophage prediction in prokaryotic genomes. *Bioinformatics*. 2008;24:863–5.
80. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res*. 2016;44:1–6.
81. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspace short palindromic repeats. *Nucleic Acids Res*. 2007;35:52–7.
82. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res*. 2016;44:W242–5.
83. Hadfield J, Croucher NJ, Goater RJ, Abudahab K, Aanensen DM, Harris SR. Phandango: an interactive viewer for bacterial population genomics. *bioRxiv*. 2017; <https://doi.org/10.1101/119545>.
84. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualiser. *Bioinformatics*. 2011;27:1009–10.
85. Bhatia S, Densmore D. Pigeon: a design visualizer for synthetic biology. *ACS Synth Biol*. 2013;2:348–50.
86. Leccese Terraf MC, Mendoza LM, Juárez Tomás MS, Silva C, Nader-Macías MEF. Phenotypic surface properties (aggregation, adhesion and biofilm formation) and presence of related genes in beneficial vaginal lactobacilli. *J Appl Microbiol*. 2014;117:1761–72.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

