

RESEARCH ARTICLE

# Comparative genomic and methylome analysis of non-virulent D74 and virulent Nagasaki *Haemophilus parasuis* isolates

Tracy L. Nicholson <sup>\*</sup>, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore

National Animal Disease Center, Agricultural Research Service, USDA, Ames, Iowa, United States of America

\* [tracy.nicholson@ARS.USDA.GOV](mailto:tracy.nicholson@ARS.USDA.GOV)



## Abstract

*Haemophilus parasuis* is a respiratory pathogen of swine and the etiological agent of Glässer's disease. *H. parasuis* isolates can exhibit different virulence capabilities ranging from lethal systemic disease to subclinical carriage. To identify genomic differences between phenotypically distinct strains, we obtained the closed whole-genome sequence annotation and genome-wide methylation patterns for the highly virulent Nagasaki strain and for the non-virulent D74 strain. Evaluation of the virulence-associated genes contained within the genomes of D74 and Nagasaki led to the discovery of a large number of toxin-antitoxin (TA) systems within both genomes. Five predicted hemolysins were identified as unique to Nagasaki and seven putative contact-dependent growth inhibition toxin proteins were identified only in strain D74. Assessment of all potential *vtaA* genes revealed thirteen present in the Nagasaki genome and three in the D74 genome. Subsequent evaluation of the predicted protein structure revealed that none of the D74 VtaA proteins contain a collagen triple helix repeat domain. Additionally, the predicted protein sequence for two D74 VtaA proteins is substantially longer than any predicted Nagasaki VtaA proteins. Fifteen methylation sequence motifs were identified in D74 and fourteen methylation sequence motifs were identified in Nagasaki using SMRT sequencing analysis. Only one of the methylation sequence motifs was observed in both strains indicative of the diversity between D74 and Nagasaki. Subsequent analysis also revealed diversity in the restriction-modification systems harbored by D74 and Nagasaki. The collective information reported in this study will aid in the development of vaccines and intervention strategies to decrease the prevalence and disease burden caused by *H. parasuis*.

## OPEN ACCESS

**Citation:** Nicholson TL, Brunelle BW, Bayles DO, Alt DP, Shore SM (2018) Comparative genomic and methylome analysis of non-virulent D74 and virulent Nagasaki *Haemophilus parasuis* isolates. PLoS ONE 13(11): e0205700. <https://doi.org/10.1371/journal.pone.0205700>

**Editor:** Feng Gao, Tianjin University, CHINA

**Received:** May 10, 2018

**Accepted:** September 28, 2018

**Published:** November 1, 2018

**Copyright:** This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. Additionally, whole-genome sequences were deposited in DDBJ/ENA/GenBank with the accession numbers CP018034 for Nagasaki and CP018032 for D74 genome and CP018033 for the plasmid sequence. These data have also been deposited in the REBASE database. The *Haemophilus parasuis* D74 are available at: <http://rebase.neb.com/cgi-bin/pacbioget?20940> and the *Haemophilus parasuis* Nagasaki data are available at: <http://rebase.neb.com/cgi-bin/pacbioget?20939>.

## Introduction

*Haemophilus parasuis* is a small, Gram negative, non-motile, pleomorphic rod-shaped, and nicotinamide adenine dinucleotide (NAD)-dependent bacterium of the *Pasteurellaceae* family [1, 2]. *H. parasuis* is a respiratory pathogen affecting swine and is the etiological agent of Glässer's disease, a systemic infection resulting in arthritis, polyserositis (inflammation of serous

**Funding:** Funding was provided by the United States Department of Agriculture, Agriculture Research Service (Project# 5030-32000-119-00D). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

membranes), and meningitis [2–4]. Additionally, *H. parasuis* infections can lead to pneumonia without signs of systemic disease in swine [5–7]. The morbidity and mortality caused by *H. parasuis* is a significant source of economic loss to the swine industry worldwide.

Serotyping based on the production of heat-stable antigens, in which capsular polysaccharide is presumed to be the dominant component of the serotyping antigen, is routinely used for isolate classification and epidemiological purposes as well as for guidance in regards to vaccination strategies. Fifteen serovars of *H. parasuis* have been defined, however, a substantial percentage of clinical isolates are identified as nontypeable (NT) using conventional indirect hemagglutination (IHA) methods [8, 9]. Progress to alleviate this problem has been made with the determination of the nucleotide sequence of the capsule locus from fifteen serovar reference isolates, which has been used to develop molecular serotyping methods [10–12].

*H. parasuis* isolates can exhibit different virulence capabilities ranging from lethal systemic disease to subclinical carriage. Numerous studies have focused on the identification of virulence factors that enable some isolates to cause systemic disease, distinguishing them from isolates that remain colonizers of the upper respiratory tract. Examples of potential virulence factors that have been evaluated to date include capsule production, outer membrane proteins (OMPs), trimeric autotransporters, and regulatory proteins QseC and OxyR [13–22]. Despite the advancement in our understanding of the pathogenic mechanisms used by *H. parasuis*, a direct link between specific virulence factors and the ability to cause systemic disease has not been demonstrated. Accordingly, virulence is thought to be multifactorial [2–4]. Data directly linking specific genes to disease outcomes has been hindered by several substantial complications, the most notable being difficulties in genetic modification of the chromosome due to low transformation efficiencies attributed to strain specific restriction modification barriers, as well as difficulties in consistently reproducing Glässer's disease in conventionally raised pigs due to confounding factors such as age, health status, differences in maternal antibody titers towards *H. parasuis*, and coinfection with other respiratory pathogens [23–26].

There are no effective approaches to eradicate *H. parasuis* from pig herds and controlling outbreaks has proven difficult [2, 27]. Although vaccines have been developed, most are comprised of bacterins, resulting in poor heterologous protection. Consequently no broadly protective vaccines or intervention strategies exist [28–30]. The current treatment for *H. parasuis* is broad spectrum antibiotics, which are expensive and are believed to increase the risk of resistant strain development [29, 31–33]. Additionally, with increased pressure to limit antibiotic use in agriculture, alternative approaches are desperately needed to reduce disease burden and economic losses caused by *H. parasuis*.

In a previous effort to link genomic differences to disease outcome, draft genome sequence data was obtained for ten genetically distinct isolates along with the evaluation of virulence in Caesarean-derived, colostrum deprived (CDCD) pigs [34]. These results demonstrated that strain D74 is a non-virulent colonizer of the upper respiratory tract, while in contrast, strain Nagasaki was highly virulent and capable of causing systemic disease [34]. Many genomic differences, including gene content and/or nucleotide variation, were identified that could account for the phenotypic difference between the strains [34]. Unfortunately many genes or regions of interest within each strain were incomplete, preventing a reliable one-to-one assignment and subsequent comparison of any predicted protein structure. In order to definitively identify and characterize the genomic differences between the highly virulent Nagasaki strain and the non-virulent D74 strain, the goal of our report was to obtain the closed whole-genome sequence and genome-wide methylation patterns between these phenotypically distinct strains.

## Materials and methods

### Genome sequencing and annotation

*H. parasuis* strain Nagasaki is a Serotype Type 5 reference Strain and a Multilocus sequence typing (MLST) Type 24 strain. *H. parasuis* strain D74 is a Serotype Type 9 reference Strain and a MLST Type 25 strain. Strains were cultured in Brain Heart Infusion (BHI) Broth (BD Biosciences, Sparks, MD) supplemented with 5% filtered heat-inactivated horse serum (GIBCO, Life Technologies, Grand Island, NY) and 0.01% (w/v) nicotinamide adenine dinucleotide (NAD) (Sigma-Aldrich, St. Louis, MO) at 37°C in 5% CO<sub>2</sub> for 24 hours and total genomic DNA was extracted using the High Pure PCR Template Preparation Kit (Roche Applied Science, Indianapolis, IN). Whole genome sequencing was performed using both the Pacific Biosciences (PacBio) and Illumina MiSeq platforms. Library preparation for PacBio sequencing was performed following the PacBio 10-kb insert library preparation protocol available online at (<http://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-10-kb-Template-Preparation-and-Sequencing.pdf>). The 10 kb library for each strain was sequenced using the PacBio RSII platform with two SMRT cells for each isolate. Indexed libraries for the MiSeq protocol were generated with the Nextera XT DNA sample preparation and index kits (Illumina, San Diego, CA), pooled, and sequenced using MiSeq v2 500-Cycle reagent kit yielding 2 x 250-bp paired-end reads (Illumina, San Diego, CA).

Whole genome assemblies were generated using the PacBio smrtanalysis v. 2.3.0 (<https://www.pacb.com/products-and-services/analytical-software/smrt-analysis/>) and CANU v. 1.3 [35] software. The average PacBio coverage for the assembled genomes was 805x for Nagasaki and 1,284x for D74. Assembling the PacBio data for each strain resulted in a fully sequenced closed circular chromosome, which was subsequently oriented to start at the *dnaA* gene and trimmed by removing any overlapping sequence. The genomes were then polished and error corrected using the Broad Institute's Pilon v 1.18 [36] and Illumina data 103x and 120x average coverage for Nagasaki and D74, respectively. The closed genome for each strain was then annotated using NCBI's Prokaryotic Genome Annotation Pipeline (PGAP) and additional curation was performed using the Prokka annotation software (version 1.12) [37] along with a *Haemophilus*-specific custom database. To compare putative protein sequences between Nagasaki and D74, the RAST prokaryotic genome annotation server [38] (<http://rast.nmpdr.org/>) was used to map annotated protein coding sequences (CDS) to functional subsystems and performed a one-to-one BLASTP comparison between the strains. Genome organization was evaluated using the Artemis Comparison Tool [39] and Mauve [40].

### Plasmid sequencing and annotation

Assembled PacBio data for pD74 did not initially result in a circularized plasmid sequence. The complete nucleotide sequence was subsequently determined using a primer walking strategy employing five separate plasmid preparations isolated from strain D74 using a Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, WI) according to the manufacturer's protocol. The resulting circularized plasmid sequence was then polished and error corrected using the Broad Institute's Pilon v 1.18 [36] and Illumina data 104x average coverage for strain D74. The assembled pD74 sequence was then annotated using NCBI's Prokaryotic Genome Annotation Pipeline (PGAP). Further sequence analysis was carried out using BLASTN, BLASTX, and Tandem Repeats Finder [41].

### IS and CRISPR analysis

Genomes were submitted to ISfinder (<https://www-is.biotoul.fr/>) [42] for the identification of bacterial IS elements using default parameters. Genomes were submitted to CRISPRFinder

(<http://crispr.i2bc.paris-saclay.fr/Server/>) [43] for the identification CRISPR elements using default parameters.

### Phenotypic analysis

Phenotypic antibiotic resistance was determined using the broth microdilution method by Iowa State University Veterinary Diagnostic Laboratory following standard operating procedures. Each isolate was tested using the Trek BOPO6F plate (Thermo Fisher Scientific Inc., Oakwood Village, OH) and minimum inhibitory concentrations (MICs) were determined. MICs were evaluated in accordance with Clinical Laboratory Standards Institute (CLSI) recommendations for resistance interpretations.

### Genomic antimicrobial resistance (AMR) analysis

ResFinder 2.1 from the Center for Genomic Epidemiology (<http://www.genomicepidemiology.org/>) and the Comprehensive Antibiotic Resistance Database (CARD) (<https://card.mcmaster.ca/home>) were employed for AMR determinant identification. Genomes submitted to ResFinder 2.1 were evaluated for AMR determinants using starting parameters of a threshold ID of 90% and a minimum length of 60% and final parameters of a threshold ID of 30% and a minimum length of 20%. Genomes submitted to CARD were evaluated for AMR determinants using the criteria “default–perfect and strict hits only”.

### Capsular loci analysis

The nucleotide sequences of genes within the capsule loci of Nagasaki and D74 reported by Howell et al. [10] were obtained from NCBI. Alignments of individual gene and protein sequences, as well as calculation of percent identities, were performed using the Geneious Alignment tool in Geneious 10.1.3 (Biomatters Ltd., Auckland, New Zealand). Global alignment with free end gaps parameters were used for both nucleotide and protein sequence alignments followed by determining the percent identity relative to Howell et al. [10]. Nucleotide and amino acid insertions are reported using HGVS nomenclature [44].

### *cdiA* gene analysis

The nucleotide sequences of genes within the *cdiA* region of D74 were evaluated using BLASTX. Translated coding sequences were extracted and evaluated for the occurrence of any domain using the Pfam 31.0 database (<https://pfam.xfam.org/>) [45].

### *vtaA* gene analysis

To ensure that all potential *vtaA* genes were identified in both Nagasaki and D74 genome annotations, the translated coding sequences were extracted and batch queried using the Pfam 31.0 database (<https://pfam.xfam.org/>) [45] for all occurrences of a YadA\_anchor domain (PF03895) [15]. Geneious 10.1.3 (Biomatters Ltd., Auckland, New Zealand) was employed to evaluate and compare the genome location of *vtaA* genes in both Nagasaki and D74, including flanking upstream and downstream genes. The domain architecture and content of each *vtaA* gene from both Nagasaki and D74 was further evaluated for the occurrence of any domains using the Pfam 31.0 database (<https://pfam.xfam.org/>) [45]. BLASTN was employed to search sequences upstream of *vtaA* genes in both Nagasaki and D74 for the occurrence of the Nagasaki *vtaA4* promoter sequence identified by Pina et al. [15].

### Methylation analysis

Detection of modified bases (m6A, m4C, m5C) and clustering of modified sites to identify methylation associated motifs was performed using the RS\_Modification\_and\_Motif\_analysis.1 tool from the SMRT analysis package version 2.3.0. Briefly, raw reads were aligned to the complete genomes of D74 and Nagasaki and interpulse duration (IPD) ratios were measured for all pulses aligned to each position in the reference sequence ([http://www.pacb.com/pdf/TN\\_Detecting\\_DNA\\_Base\\_Modifications.pdf](http://www.pacb.com/pdf/TN_Detecting_DNA_Base_Modifications.pdf)). [46]

### SSR analysis

The nucleotide sequences of all putative RM genes identified in D74 and Nagasaki were evaluated for the occurrence of SSRs, tandem repeats or homopolymeric tracts of consisting of five or more bases, within the coding region and in the region encompassing 150 bp upstream of the putative start codon using Geneious 10.1.3 (Biomatters Ltd., Auckland, New Zealand).

### Accession number(s)

The whole-genome sequences for these isolates were deposited in DDBJ/ENA/GenBank with the accession numbers CP018034 for Nagasaki and CP018032 for D74 genome and CP18033 for the plasmid sequence. The sequence data, target sequences and associated details for methylation enzymes, used for analyses in this report have been deposited in the REBASE database ([www.rebase.neb.com](http://www.rebase.neb.com)) [47]. RM-system and methylation motifs for both strains can be accessed via the index of the REBASE database (<http://tools.neb.com/genomes/>) or directly via this link: <http://rebase.neb.com/cgi-bin/pacbioget?20940> for D74 or <http://rebase.neb.com/cgi-bin/pacbioget?20939> for Nagasaki.

## Results and discussion

### Genome features of *H. parasuis* strains D74 and Nagasaki

The complete genome assembly and annotation of *H. parasuis* strain Nagasaki contains a single circular chromosome 2,348,962 base pairs (bp) in length, encodes a total of 2,268 predicted protein coding sequences (CDSs), and a G+C content of 40.0% (Table 1). Of the 2,268 CDSs, 95 were predicted to be pseudogenes. The complete genome assembly and annotation of *H. parasuis* strain D74 encompasses a single circular chromosome 2,467,568 base pairs (bp) in

**Table 1. General features of the genomes of *H. parasuis* strains D74 and Nagasaki.**

	D74	Nagasaki
Sequence Type (ST)	9	5
<b>Chromosome Size (bp)</b>	2,467,568	2,348,962
G + C Content (%)	39.7%	40.0%
<b>Total CDSs</b>	2,252	2,268
Pseudogenes <sup>a</sup>	95	95
<b>Functional CDSs</b>	2,157	2,173
rRNA (16S-23S-5S)	6-6-7	6-6-8
tRNA	57	60
Phage Regions	6	9
Plasmid	1	0

<sup>a</sup>Encoding either an incomplete predicted protein sequence, a frameshift, or an internal stop codon.

<https://doi.org/10.1371/journal.pone.0205700.t001>

**Table 2. Phage regions identified in *H. parasuis* strains D74 and Nagasaki.**

Strain	Region #	Start <sup>a</sup>	End <sup>a</sup>	Length <sup>b</sup>	Classification <sup>c</sup>
Nagasaki	1	75,146	95,579	20,434	Incomplete
	2	578,665	613,918	35,254	Intact
	3	930,933	982,262	51,330	Intact
	4	978,209	1,022,173	43,965	Intact
	5	1,146,622	1,187,091	40,470	Intact
	6	1,837,359	1,870,441	33,083	Intact
	7	1,890,250	1,901,023	10,774	Incomplete
	8	2,053,720	2,085,576	31,857	Intact
	9	2,113,966	2,153,632	39,667	Intact
D74	1	1,021,542	1,030,153	8,612	Incomplete
	2	1,179,090	1,197,266	18,177	Incomplete
	3	1,186,455	1,216,786	30,332	Incomplete
	4	1,303,971	1,339,245	35,275	Intact
	5	1,390,803	1,407,651	16,840	Incomplete
	6	1,868,288	1,889,812	21,525	Incomplete

<sup>a</sup>Basepair chromosomal location.

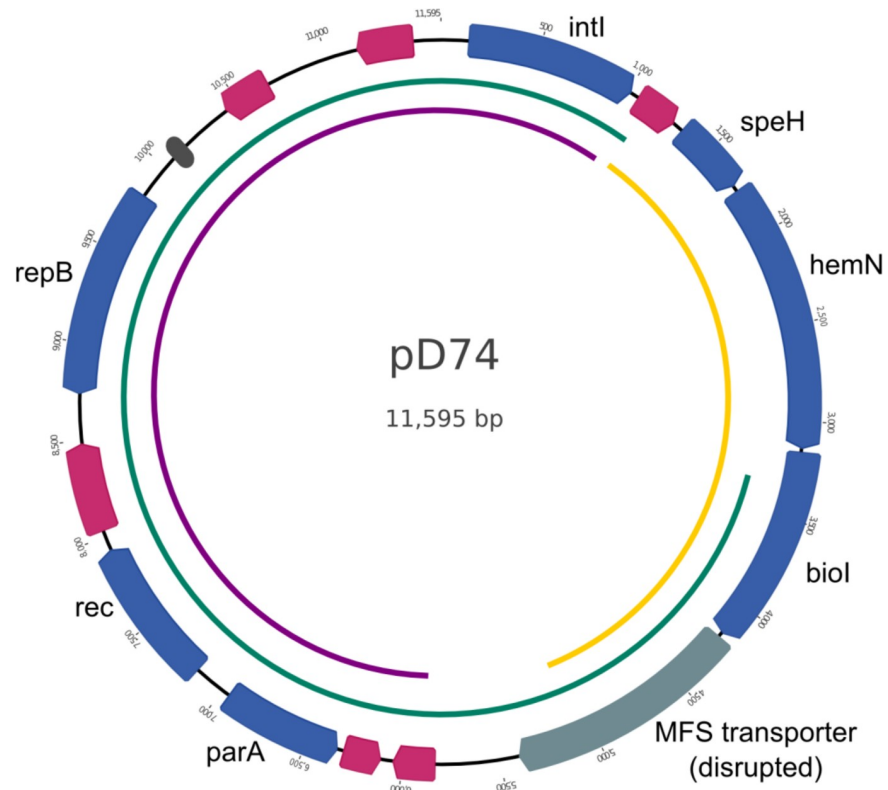
<sup>b</sup>Length of region depicted in base pairs.

<sup>c</sup>Classification determined by PHASTER analysis (<http://phaster.ca/>) [48].

<https://doi.org/10.1371/journal.pone.0205700.t002>

length, a total of 2,252 total predicted protein coding sequences (CDSs), and a G+C content of 39.7% (Table 1). Similar to Nagasaki, 95 out of the 2,252 CDSs were predicted to be pseudogenes. Despite the equivalent number, none of the predicted pseudogenes were similar between the stains (S1 Table and S2 Table). The difference in rRNA numbers between D74 and Nagasaki is consistent with other closed *H. parasuis* genomes available in GenBank (Table 1). The genomes of *H. parasuis* Nagasaki and D74 were assessed using PHASTER (<http://phaster.ca/>) [48] for the occurrence of phage regions. Nine phage regions, seven intact and two incomplete, were identified along the Nagasaki chromosome (Table 2). Six phage regions, one intact and five incomplete, were also identified along the D74 chromosome (Table 1). The phage regions identified are unique to each genome. The chromosomal location, including start and end nucleotide positions, length of phage region, and classification of phage regions for both strains are summarized in Table 2.

An 11,595-bp circular plasmid was additionally identified in the complete genome assembly and annotation of *H. parasuis* strain D74. With the addition of the plasmid, the total nucleotides, including both the circular chromosome and plasmid, is 2,479,163 bp for strain D74. The plasmid harbored by strain D74, designated as pD74, contains seven CDSs with predicted functions based on sequence homology, including *parA*, *rec*, and *repB*, which have predicted functions in plasmid replication. Additionally, pD74 harbors six CDSs of unknown function and one predicted pseudogene (Fig 1). These CDSs lacked a reciprocal match in Nagasaki. A region containing four copies of a 22-bp tandem repeat sequence was identified upstream of the *repB* CDS, which could potentially comprise the origin of replication (Fig 1). The sequence spanning from 3,328 to 1,243 bp comprises the entire 9,462 bp sequence of *H. parasuis* plasmid pHS-Rec (accession no. AY862436) with 99.2% sequence identity (Fig 1) [49]. The sequence spanning from 5,847 to 1,025 bp shares 99.4% sequence identity with *Pasteurella trehalosi* plasmid pCCK13698 (accession no. AM183225) (Fig 1) [50]. pD74 sequencing from 1,246 to 5,350 bp shares 73.9% sequence identity with a chromosomal region of *Gallibacterium anantis* strain UMN179 (accession no. CP002667) (Fig 1) [51].



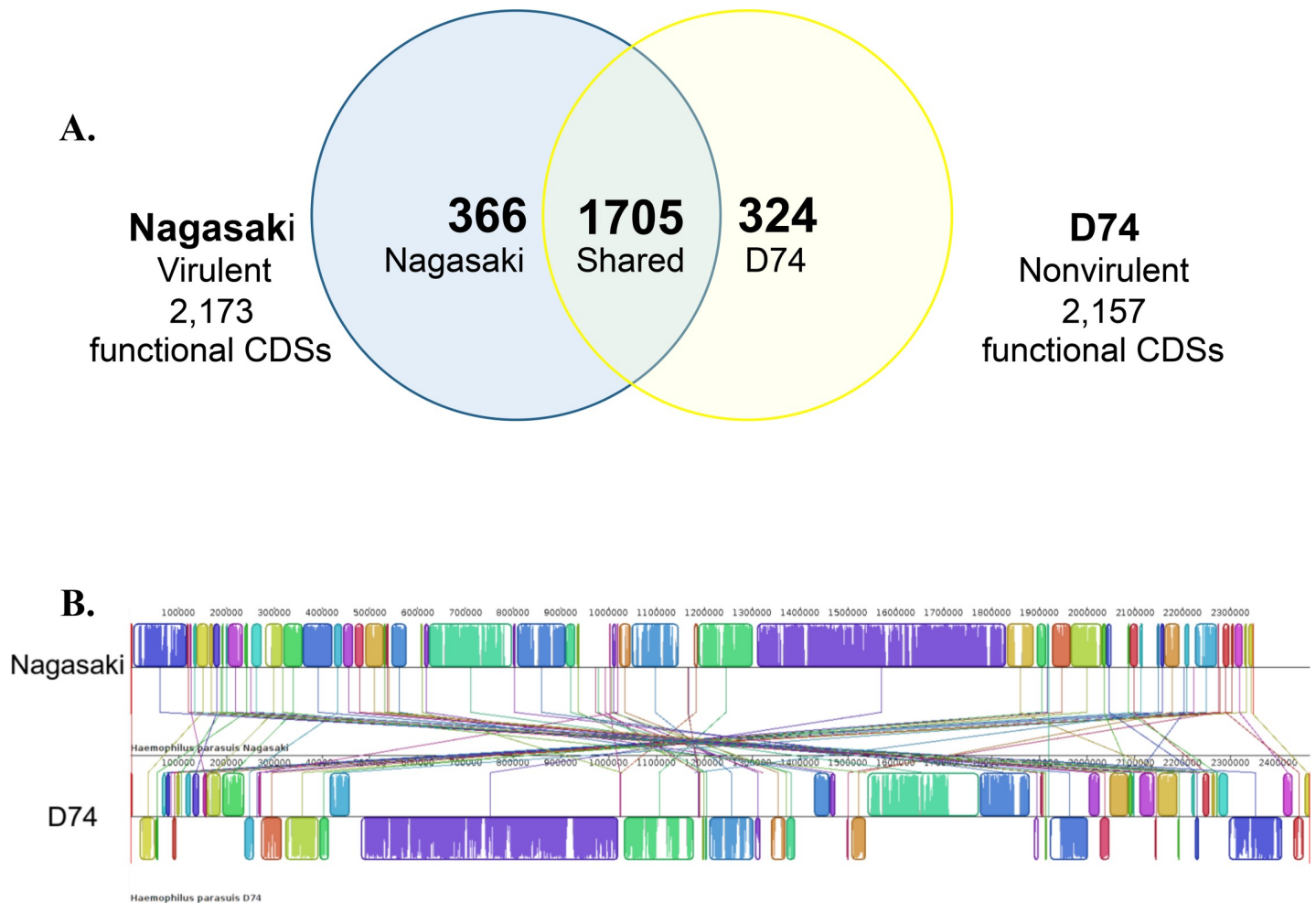
**Fig 1. Map of plasmid pD74.** Arrows indicate annotated CDSs; blue arrows represent CDSs with predicted functions based on sequence homology, pink arrows represent predicted CDSs of unknown function. The MFS transporter gene, labeled with a blue-grey arrow, is predicted to be a pseudogene. Dark grey box indicates region containing 4 copies of a 22-bp tandem repeat sequence identified using the Tandem Repeats Finder tool [41]. Arcs inside map indicate sequence with similarity to other sources; green arc represents sequence similar to *H. parasuis* plasmid pHS-Rec [49], purple arc represents sequence similar to a portion of the *Pasteurella trehalosi* plasmid pCCK13698 [50], and yellow arc represents sequence similar to a region from the genome sequence of *Gallibacterium anantis* strain UMN179 [51].

<https://doi.org/10.1371/journal.pone.0205700.g001>

## Comparison of the genome sequences of *H. parasuis* strains D74 and Nagasaki

A reciprocal or one-to-one BLASTP comparison of the protein coding sequences in Nagasaki and D74 identified 1,705 shared CDSs between the strains (Fig 2). Of the 2,173 functional CDSs in strain Nagasaki, 366 CDSs lacked a reciprocal match in D74 and were designated unique or Nagasaki-specific (Fig 2A). Conversely, 324 CDSs, out of the 2,157 functional CDSs in strain D74, lacked a reciprocal match in Nagasaki and were designated unique or D74-specific (Fig 2A). Comparison of the linear organization of the genomes of Nagasaki and D74 revealed many genome re-arrangements and inversions (Fig 2B). The alignment additionally resulted in 73 locally collinear blocks (LCBs), the largest of which is 520,607 bp (Fig 2B).

The genomes of *H. parasuis* Nagasaki and D74 were assessed for the presence of insertion sequence (IS) elements using ISfinder (<https://www-is.biotoul.fr/>) [42]. IS elements found previously in other *H. parasuis* and *Pasteurellaceae* genomes were identified in both D74 and Nagasaki genomes and these IS elements contained the greatest similarity to sequences from the IS1595 family. Specifically, four ISHps3 and three ISHps4 IS elements were identified in D74, while one ISHps3 and four ISHps4 IS elements were identified in Nagasaki. *In silico* analysis revealed the occurrence of additional transposases annotated as belonging to other families, such as IS256 and IS110, in both D74 and Nagasaki genomes that were not identified by



**Fig 2. Comparison of the genome sequences of *H. parasuis* strains D74 and Nagasaki.** (A) Distribution of protein-encoding genes in *H. parasuis* strains Nagasaki and D74. Venn diagram demonstrating the unique 2,173 protein-encoding genes in Nagasaki (blue), the 2,157 protein-encoding genes D74 (yellow), and the shared coding sequences defined as bi-directional best hits (center circle, green) via 1-to-1 reciprocal BLASTP comparison using RAST [38] (excluding frame-shifted genes). (B) Comparison of the linear organization of the *H. parasuis* strains Nagasaki and D74 chromosomes. Mauve [40] was used to compare the chromosomes of Nagasaki and D74. Locally collinear blocks (LCBs) representing regions of sequence that align in each genome are illustrated as colored rectangles connected by lines. Nagasaki was used as the reference sequence. In D74, LCBs placed above the center line are in the same orientation as in Nagasaki and LCBs placed below the center line are in the reverse orientation relative to Nagasaki. Blank sections are regions that did not align and are likely to contain unique or strain-specific sequence. Many of the larger blank regions not in LCBs contain loci predicted to encode products of phage origin.

<https://doi.org/10.1371/journal.pone.0205700.g002>

ISfinder analysis, given that it typically does not identify frameshifted or truncated IS units. These transposases were annotated as pseudogenes and therefore are likely nonfunctional. Of the five complete IS units identified in Nagasaki, two are within regions of conserved sequence order and have homologs in D74. Three of the intact IS units identified in Nagasaki are located at the border of sequence rearrangement, with one located at the border of one of the phage regions. Six complete IS units were identified in D74 with two located at the border of rearranged regions. Of the four IS units that are present within regions of conserved sequence order, two are within regions of conserved sequence order and have homologs in Nagasaki and two are small sequence insertions located within a conserved region in which only the IS unit is the only sequence not conserved within the region. Clusters of regularly interspaced short palindromic repeats (CRISPR) and spacer sequences in the genomes of both D74 and



Nagasaki were screened using CRISPRFinder (<http://crispr.i2bc.paris-saclay.fr>) [43]. No evidence for CRISPR sequences were found in either genome.

Phenotypic antibiotic resistance was determined for *H. parasuis* D74 and Nagasaki and results are summarized in [S3 Table](#). Nagasaki exhibited phenotypic resistance to clindamycin while D74 exhibited “intermediate” limited phenotypic resistance to clindamycin; no other resistance was identified. The genomes of *H. parasuis* Nagasaki and D74 were screened for the presence of acquired resistance genes and known chromosomal mutations conferring clindamycin resistance. No *erm*, *lnu*, or other known resistance determinants were identified in either genome.

### Capsule loci in *H. parasuis* strains D74 and Nagasaki

Due to the importance of capsular polysaccharide in serotyping or strain classification, virulence, and vaccine related research areas, the CDSs located within the capsule loci of Nagasaki and D74 were compared to nucleotide sequences reported by Howell et al. [10]. Several nucleotide and amino acid differences were identified and are summarized in [S4 Table](#) and [S5 Table](#). Eleven out of sixteen predicted proteins within the D74 capsule locus were found to contain amino acid differences compared to the sequences reported by Howell et al. [10]. The genes harboring the most or more noteworthy changes for D74 include *wzx2*, *astA*, *gltL*, and *wzs* ([S4 Table](#)). The predicted amino acid changes in *AstA* relative to the Howell et al. [10] sequence included a two amino acid insertion and are: N98\_H99insHI, Q134R, S157A, V177I, V183I, I189V, S195N, and V2056M. In contrast, the predicted amino acid sequence for *GltL* is shorter than that reported in Howell et al. [10] due to a different predicted start codon. The predicted amino acid changes in *GltL* are M1\_K8del and V9M. The predicted amino acid changes in *Wzx2* relative to the Howell et al. [10] sequence are: V176A, I180V, A187V, S192N, R278H, P291S, and R373S ([S4 Table](#)). The predicted amino acid changes in *Wzs* relative to the Howell et al. [10] sequence are: E3K, L27I, V230A, V613A, and V622A ([S4 Table](#)).

In *H. parasuis* Nagasaki, twelve out of fourteen predicted proteins within the capsule locus were found to contain amino acid differences compared to the sequences reported by Howell et al. [10]. The genes harboring the more notable changes for Nagasaki include *funA*, *wcfQ*, and *wbgX* ([S5 Table](#)). *FunA* is predicted to result in a longer protein sequence compared to Howell et al. [10] due to a different predicted start codon and five amino acid changes within the shared region, along with 37 additional amino acids on the N-terminus. The predicted amino acid changes in *FunA* are M1ext-37, A2V, A6V, D27V, S28N, and V138I ([S5 Table](#)). The predicted amino acid changes in *WcfQ* relative to the Howell et al. [10] sequence are: E43K, I58V, S61T, I64T, F106S, D131G, F132S, N194K, Y262C, R269N, F270N, and L271F ([S5 Table](#)). *WbgX* is predicted to encode a shorter protein sequence compared to Howell et al. [10] due to a different predicted start codon. The predicted amino acid changes in *WbgX* are: M1del, K2M, E3N, F4Y, T177A, S343N, P344A, and A349I ([S5 Table](#)).

### Virulence-associated genes identified in *H. parasuis* strains D74 and Nagasaki

To identify factors that could support host colonization and/ or virulence, the D74 and Nagasaki annotations were searched for CDSs encoding predicted functions in adhesion, hemolysis, secretion, toxin production, or other virulence-associated roles. Twenty-one CDSs encoding predicted adhesins were identified in D74 including five outer membrane genes (*ompA*, *ompP1*, *ompP2*, *ompP5*, and *ompD15*), two *fimD* fimbrial usher genes located adjacent to each other along the chromosome, pertactin family virulence factor *aidA*, filamentous hemagglutinin transporter *fhaC*, adhesin autotransporter *bmaC*, and two type IV pili genes ([Table 3](#)).

**Table 3. Predicted virulence-associated genes identified in *H. parasuis* D74.**

Group	D74 locus_tag	D74 Name	Product/ Function	Hit <sup>a</sup>	Nagasaki locus_tag	% Identity <sup>b</sup>	
Adhesin	A2U20_02695	<i>pilW</i>	Type IV pilus biogenesis/stability protein PilW	bi	A2U21_08925	100	
	A2U20_03215	<i>pulG</i>	Type II secretory pathway, pseudopilin PulG	bi	A2U21_08415	97	
	A2U20_03220	<i>pulJ</i>	Type II secretory pathway, component PulJ	bi	A2U21_08410	96	
	A2U20_03425	<i>pilM</i>	Type IV pilus biogenesis protein PilM	bi	A2U21_08215	98	
	A2U20_03430	<i>pilN</i>	Type IV pilus biogenesis protein PilN	bi	A2U21_08210	98	
	A2U20_03445	<i>pilQ</i>	Type IV pilus biogenesis PilQ	bi	A2U21_08195	94	
	A2U20_03490	<i>ompP5</i>	Outer membrane protein P5	bi	A2U21_08150	92	
	A2U20_03710	<i>ompP1</i>	Outer membrane protein precursor P1	bi	A2U21_07950	90	
	A2U20_03770	<i>pilT</i>	pilT domain-containing protein	uni	A2U21_06460	56	
	A2U20_04380	<i>ompP2</i>	Outer membrane protein P2 precursor	bi	A2U21_07315	85	
	A2U20_04655	<i>ompD15</i>	Surface antigen (D15), outer membrane protein	bi	A2U21_07040	99	
	A2U20_04855	<i>ompA</i>	Outer membrane protein A	bi	A2U21_06820	98	
	A2U20_06825	<i>fhaC</i>	Filamentous hemagglutinin transporter protein FhaC	uni	A2U21_11135	52	
	A2U20_07210	<i>pilA</i>	Type IV pilin PilA	bi	A2U21_10860	87	
	A2U20_07215	<i>pilB</i>	Type IV fimbrial assembly ATPase PilB	bi	A2U21_10855	98	
	A2U20_07220	<i>pilC</i>	Type IV fimbrial assembly protein PilC	bi	A2U21_10850	99	
	A2U20_07225	<i>pilD</i>	Tfp pilus assembly pathway, fimbrial leader peptidase	bi	A2U21_10845	94	
	A2U20_08340	<i>aidA</i>	Type V secretory pathway, adhesin AidA	bi	A2U21_04065	72	
	A2U20_08910	<i>bmaC</i>	Adhesin BmaC autotransporter	uni	A2U21_04065	51	
	A2U20_08920	<i>fimD</i>	Putative F17-like fimbrial usher	uni	A2U21_03995	98	
	A2U20_08925	<i>fimD2</i>	Putative F17-like fimbrial usher	bi	A2U21_03995	99	
	Hemolysin	A2U20_02515	<i>prtC</i>	Serralyisin C	uni	A2U21_00470	58
		A2U20_06885		Hemagglutinin/hemolysin-related protein	-		
A2U20_07380		<i>hlyD</i>	Hemolysin secretion protein D	-			
A2U20_08300		<i>shlB</i>	Hemolysin transporter protein ShlB	bi	A2U21_11135	53	
A2U20_08375		<i>osmY</i>	Osmotically-inducible protein OsmY; putative hemolysin	bi	A2U21_04100	100	
A2U20_09890		<i>ahpA</i>	Hemolysin regulation protein AhpA	bi	A2U21_02620	99	
A2U20_10965		<i>prtB</i>	Serralyisin B, hemolysin-type calcium-binding region	uni	A2U21_00470	56	
A2U20_10970		<i>prtB2</i>	Serralyisin B, hemolysin-type calcium-binding region	bi	A2U21_00470	69	
Secretion	A2U20_01865	<i>esiB</i>	Putative secretory immunoglobulin A-binding protein	bi	A2U21_08055	51	
	A2U20_09675		Putative periplasmic/secreted protein	bi	A2U21_02375	93	
Toxin	A2U20_00255	<i>ebgC</i>	EbgC protein_Toxin-antitoxin biofilm protein TabA	bi	A2U21_00620	98	
	A2U20_01495	<i>higA</i>	Antitoxin HigA_mRNA interferase antitoxin	uni	A2U21_00995	63	
	A2U20_01500	<i>hicA</i>	Addiction module toxin HicA	uni	A2U21_01000	59	
	A2U20_01850	<i>vapC</i>	VapC toxin family PIN domain ribonuclease	bi	A2U21_10000	78	
	A2U20_02115	<i>hipA</i>	Serine/threonine-protein kinase toxin HipA	bi	A2U21_09500	95	
	A2U20_02705	<i>higA2</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U21_08915	93	
	A2U20_02710	<i>higB</i>	mRNA interferase toxin HigB	bi	A2U21_08910	96	
	A2U20_02805	<i>tdeA</i>	Putative toxin and drug export protein A	bi	A2U21_08785	97	
	A2U20_02840	<i>relE</i>	Addiction module toxin RelE	bi	A2U21_03035	59	
	A2U20_03250	<i>mazF</i>	Programmed cell death toxin MazF	uni	A2U21_06190	52	
	A2U20_03635	<i>hicA2</i>	Addiction module toxin HicA	bi	A2U21_02950	94	
	A2U20_04570	<i>higA3</i>	Antitoxin HigA_mRNA interferase antitoxin	uni	A2U21_00995	63	
	A2U20_04795	<i>pezT</i>	Antitoxin/toxin system zeta toxin	uni	A2U21_03130	53	
	A2U20_04825	<i>cdtC</i>	Cytolethal distending toxin subunit CdtC	bi	A2U21_06570	99	
	A2U20_04830	<i>cdtB</i>	Cytolethal distending toxin subunit CdtB	uni	A2U21_06565	95	
	A2U20_04835	<i>cdtA</i>	Cytolethal distending toxin subunit CdtA	bi	A2U21_06560	99	

(Continued)

Table 3. (Continued)

Group	D74 locus_tag	D74 Name	Product/ Function	Hit <sup>a</sup>	Nagasaki locus_tag	% Identity <sup>b</sup>
	A2U20_04880	<i>higB2</i>	mRNA interferase toxin HigB	uni	A2U21_05325	55
	A2U20_04885	<i>higA4</i>	Antitoxin HigA_mRNA interferase antitoxin	uni	A2U21_08915	56
	A2U20_05030	<i>higA5</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U21_06130	98
	A2U20_05035	<i>higB3</i>	mRNA interferase toxin HigB	bi	A2U21_06135	99
	A2U20_05120	<i>cdtC2</i>	Cytolethal distending toxin subunit CdtC	uni	A2U21_06570	98
	A2U20_05125	<i>cdtB2</i>	Cytolethal distending toxin subunit CdtB	bi	A2U21_06565	96
	A2U20_05130	<i>cdtA2</i>	Cytolethal distending toxin subunit CdtA	uni	A2U21_06560	99
	A2U20_05385	<i>higA6</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U21_06310	99
	A2U20_05390	<i>higB4</i>	mRNA interferase toxin HigB	bi	A2U21_06305	100
	A2U20_05455	<i>higB5</i>	mRNA interferase toxin HigB	bi	A2U21_06250	99
	A2U20_05460	<i>higA7</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U21_06245	100
	A2U20_05480	<i>cdiA</i>	Contact-dependent growth inhibition (CDI) toxin	-		
	A2U20_05555	<i>chpS</i>	Antitoxin ChpS	bi	A2U21_06185	99
	A2U20_05585	<i>tabA</i>	Toxin-antitoxin biofilm protein TabA	bi	A2U21_06150	100
	A2U20_05690	<i>hicA3</i>	Addiction module toxin HicA	bi	A2U21_04705	62
	A2U20_06620	<i>higA8</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U21_05330	93
	A2U20_06625	<i>higB6</i>	mRNA interferase toxin HigB	bi	A2U21_05325	99
	A2U20_06830	<i>cdiA2</i>	Contact-dependent growth inhibition (CDI) toxin	-		
	A2U20_06850	<i>cdiA3</i>	Contact-dependent growth inhibition (CDI) toxin	-		
	A2U20_06865	<i>cdiA4</i>	Contact-dependent growth inhibition (CDI) toxin	-		
	A2U20_06875	<i>cdiA5</i>	Contact-dependent growth inhibition (CDI) toxin	-		
	A2U20_07065	<i>pezT2</i>	Antitoxin/toxin system zeta toxin	bi	A2U21_03130	75
	A2U20_07375	<i>ltxB</i>	Leukotoxin export ATP-binding protein LtxB	uni	A2U21_07020	52
	A2U20_08285	<i>cdiA6</i>	Contact-dependent growth inhibition (CDI) toxin	-		
	A2U20_08295	<i>cdiA7</i>	Contact-dependent growth inhibition (CDI) toxin	-		
	A2U20_08515	<i>pasI</i>	Persistence and stress-resistance antitoxin PasI	bi	A2U21_04275	98
	A2U20_10085	<i>hicB</i>	Antitoxin HicB	bi	A2U21_00970	100
	A2U20_10110	<i>higA9</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U21_00995	98
	A2U20_10115	<i>higB7</i>	mRNA interferase toxin HigB	bi	A2U21_01000	99
	A2U20_10565	<i>hipA2</i>	Serine/threonine-protein kinase toxin HipA	bi	A2U21_00565	99
Other	A2U20_00515	<i>nanH</i>	Sialidase	bi	A2U21_11540	99
	A2U20_01700	<i>espP</i>	Putative extracellular serine protease	bi	A2U21_09655	75
	A2U20_01715	<i>espP2</i>	Putative extracellular serine protease	bi	A2U21_09640	78
	A2U20_03355	<i>vacJ</i>	Putative VacJ lipoprotein	bi	A2U21_08285	97
	A2U20_05150	<i>sirA</i>	Regulator of disulfide bond formation	bi	A2U21_06540	99
	A2U20_05155	<i>sirB</i>	Invasion protein expression up-regulator SirB	bi	A2U21_06535	100
	A2U20_07890	<i>sodA</i>	Superoxide dismutase	bi	A2U21_03770	100
	A2U20_08325	<i>sodC</i>	Superoxide dismutase	bi	A2U21_04050	92

<sup>a</sup> The Hit column contains a '-' (no hit), 'uni' or 'bi' RAST server results from a one-to-one BLASTP comparison of the protein coding sequences in the Nagasaki genome using the D74 genome as the reference. "bi" represents a bidirectional best hit in which the reverse hit from the Nagasaki comparison genome to the D74 reference genome was also the best hit. "uni" indicates a uni-directional hit in which the reverse hit from the comparison genome to the reference genome was not also the best hit. "-" indicates no hit or match was found.

<sup>b</sup> Global pairwise nucleotide percent sequence identity.

<https://doi.org/10.1371/journal.pone.0205700.t003>

Corresponding orthologs for *pilT* and *bmaC* were not found in Nagasaki. Gene *leuC* (A2U21\_06460) in Nagasaki was the uni-directional best match to *pilT* (A2U20\_03770) in D74 with 56% global nucleotide sequence identity and gene *aidA2* (A2U21\_04065) in

Nagasaki was the uni-directional best match to *bmaC* (A2U20\_08910) in D74 with 51% global nucleotide sequence identity (Table 3). Eighteen CDSs encoding predicted adhesins were identified in Nagasaki including five outer membrane genes (*ompA*, *ompP1*, *ompP2*, *ompP5*, and *ompD15*), one *fimD* fimbrial usher gene (chromosomally adjacent to a predicted pseudogene *fimB*), two pertactin family virulence factor genes *aidA* and *aidA2*, and two type IV pili (Table 4). *In silico* analysis of the region in D74 containing the two *fimD* genes compared to the *fimD* gene in Nagasaki indicated that the D74 *fimD2* (A2U20\_08925) aligns with the 5' end of the Nagasaki *fimD* (A2U21\_03995) and the D74 *fimD* (A2U20\_08920) aligns with the 3' end of the Nagasaki *fimD* gene. This suggests that the two *fimD* genes in D74 could have potentially arisen from a frameshift within a single gene. A 51% sequence identity was observed between corresponding orthologs *esiB* (A2U20\_01865) in D74 and *esiB2* (A2U21\_08055) in Nagasaki (Table 3 and Table 4).

Eight CDSs encoding predicted hemolysins were identified in D74 including two (locus tags A2U20\_06885 and A2U20\_07380) unique to strain D74 (Table 3). D74 harbors two putative Serralyisin B genes, *prtB* and *prtB2*, and a putative Serralyisin C gene *prtC*. In contrast, only putative Serralyisin B gene *prtB* was identified in strain Nagasaki (Table 4). A notable size difference appears to exist between the proteins encoded by both *prtB*, *prtB2*, and *prtC* in D74, 1,954, 1,703, and 1,911 amino acids respectively, compared to the putative Serralyisin B 910 amino acid protein encoded by *prtB* in Nagasaki. Nine CDSs encoding predicted hemolysins were identified in Nagasaki, five of which were identified as unique to Nagasaki. These include *hpmA* (A2U21\_11150) and locus-tags A2U21\_00465, A2U21\_11140, and A2U210\_11145. Nagasaki harbors two putative hemolysin transporter genes *shlB1* and *shlB2* (chromosomally located next to each other), while D74 contains only *shlB1* gene (Tables 3 and 4). A notable size difference was also observed for the *shlB* gene in D74 encoding a putative 582 amino acid protein compared to the putative 329 amino acid protein encoded by the *shlB* in Nagasaki.

Two CDSs encoding genes predicted to function in secretion were identified in D74, A2U20\_09675 and *esiB*, a putative secretory immunoglobulin A-binding encoding gene, while three were identified in Nagasaki. These include A2U21\_02375 and two putative secretory immunoglobulin A-binding encoding genes *esiB* and *esiB2*, chromosomally located next to each other.

Thirty-nine CDSs encoding predicted toxins were identified in D74 and forty-six CDSs encoding predicted toxins were identified in Nagasaki. A noticeable difference between the CDSs encoding predicted toxins identified in D74 and Nagasaki was the seven *cdiA* genes encoding putative contact-dependent growth inhibition toxin A harbored only in strain D74 (Table 3). These *cdiA* genes harbored by D74 are discussed in more detail below.

Both strains harbored a large number of toxin-antitoxin (TA) systems. TA systems are small genetic elements comprised of two components, a stable protein toxin and its more labile antagonistic antitoxin, which can be a protein or non-coding RNA [52]. TA systems were originally identified as plasmid-borne loci, which functioned to promote plasmid maintenance by killing daughter cells that lacked the TA encoded plasmid [52]. TA loci were subsequently discovered in numerous bacterial and archaeal chromosomes and provide several functions, such as stabilization of genomic regions, anti-addiction against similar plasmid-borne toxins, defense against phage infection, biofilm formation, control of the stress response, and bacterial persistence [52–54]. Six types of TA systems (types I to VI) have been described to date based on the type (either RNA or protein) and mode of action of the antitoxin [53]. The Type II TA system is highly abundant among prokaryotes and has been extensively studied [52, 54–56]. In type II TA systems, both toxin and antitoxin are small proteins encoded by genes in a bicistronic operon [52, 54–56]. The antitoxin blocks the toxicity of the toxin by forming a complex with it [52, 54–56]. Both D74 and Nagasaki contain several Type II TA families including

**Table 4. Predicted virulence-associated genes identified in *H. parasuis* Nagasaki.**

Group	Nagasaki locus_tag	Nagasaki Name	Product/ Function	Hit <sup>a</sup>	D74 locus_tag	% Identity <sup>b</sup>	
Adhesin	A2U21_02720	<i>aidA</i>	Type V secretory pathway, adhesin AidA	-			
	A2U21_03995	<i>fimD</i>	Putative F17-like fimbrial usher	bi	A2U20_08925	99	
	A2U21_04065	<i>aidA2</i>	Type V secretory pathway, adhesin AidA	bi	A2U20_08340	72	
	A2U21_06820	<i>ompA</i>	Outer membrane protein A precursor	bi	A2U20_04855	98	
	A2U21_07040	<i>ompD15</i>	Surface antigen (D15), outer membrane protein	bi	A2U20_04655	99	
	A2U21_07315	<i>ompP2</i>	Outer membrane protein P2 precursor	bi	A2U20_04380	85	
	A2U21_07950	<i>ompP1</i>	Outer membrane protein precursor P1	bi	A2U20_03710	90	
	A2U21_08150	<i>ompP5</i>	putative outer membrane protein P5	bi	A2U20_03490	92	
	A2U21_08195	<i>pilQ</i>	Type IV pilus biogenesis protein PilQ	bi	A2U20_03445	94	
	A2U21_08210	<i>pilN</i>	Type IV pilus biogenesis protein PilN	bi	A2U20_03430	98	
	A2U21_08215	<i>pilM</i>	Type IV pilus biogenesis protein PilM	bi	A2U20_03425	98	
	A2U21_08410	<i>pulJ</i>	Type II secretory pathway, component PulJ	bi	A2U20_03220	96	
	A2U21_08415	<i>pulG</i>	Type II secretory pathway, pseudopilin PulG	bi	A2U20_03215	97	
	A2U21_08925	<i>pilW</i>	Type IV pilus biogenesis/stability protein PilW	bi	A2U20_02695	100	
	A2U21_10845	<i>pilD</i>	Tfp pilus assembly pathway, fimbrial leader peptidase	bi	A2U20_07225	94	
	A2U21_10850	<i>pilC</i>	Type IV fimbrial assembly protein PilC	bi	A2U20_07220	99	
	A2U21_10855	<i>pilB</i>	Type IV fimbrial assembly ATPase PilB	bi	A2U20_07215	98	
	A2U21_10860	<i>pilA</i>	Type IV pilin PilA	bi	A2U20_07210	87	
	Hemolysin	A2U21_00465		Putative hemagglutinin/hemolysin-related protein	-		
		A2U21_00470	<i>prtB</i>	Serralsin B, hemolysin-type calcium-binding region	bi	A2U20_10970	69
A2U21_02620		<i>ahpA</i>	Hemolysin regulation protein AhpA	bi	A2U20_09890	99	
A2U21_04100		<i>osmY</i>	Osmotically-inducible protein OsmY_putative hemolysin	bi	A2U20_08375	100	
A2U21_11130		<i>shlB1</i>	Hemolysin transporter protein ShlB	-			
A2U21_11135		<i>shlB2</i>	Hemolysin transporter protein ShlB	bi	A2U20_08300	53	
A2U21_11140			putative hemolysin	-			
A2U21_11145			Putative hemolysin	-			
	A2U21_11150	<i>hpmA</i>	Hemolysin	-			
Secretion	A2U21_02375		Periplasmic/secreted protein	bi	A2U20_09675	93	
	A2U21_08050	<i>esiB</i>	Putative secretory immunoglobulin A-binding protein	-			
	A2U21_08055	<i>esiB2</i>	Putative secretory immunoglobulin A-binding protein	bi	A2U20_01865	51	
Toxin	A2U21_00565	<i>hipA</i>	Serine/threonine-protein kinase toxin HipA	bi	A2U20_10565	99	
	A2U21_00620	<i>ebgC</i>	EbgC protein_Toxin-antitoxin biofilm protein TabA	bi	A2U20_00255	98	
	A2U21_00970	<i>hicB</i>	Antitoxin HicB	bi	A2U20_10085	100	
	A2U21_00995	<i>higA</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U20_10110	98	
	A2U21_01000	<i>higB</i>	mRNA interferase toxin HigB	bi	A2U20_10115	99	
	A2U21_02950	<i>hicA</i>	Addiction module toxin HicA	bi	A2U20_03635	94	
	A2U21_03035	<i>relE</i>	Addiction module toxin RelE	bi	A2U20_02840	59	
	A2U21_03130	<i>pezT</i>	Antitoxin/toxin system zeta toxin	bi	A2U20_07065	75	
	A2U21_03940		Toxin-antitoxin system, antitoxin component	-			
	A2U21_04275	<i>pasI</i>	Persistence and stress-resistance antitoxin PasI	bi	A2U20_08515	98	
	A2U21_04650	<i>higA2</i>	Antitoxin HigA_mRNA interferase antitoxin	uni	A2U20_01495	54	
	A2U21_04705	<i>hicA2</i>	Addiction module toxin HicA	bi	A2U20_05690	62	
	A2U21_04735	<i>hicA3</i>	Addiction module toxin HicA	uni	A2U20_03635	94	
	A2U21_04850	<i>relE</i>	mRNA interferase toxin RelE	-			
	A2U21_05325	<i>higB2</i>	mRNA interferase toxin HigB	bi	A2U20_06625	99	
A2U21_05330	<i>higA3</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U20_06620	93		
A2U21_05725	<i>relE2</i>	Addiction module toxin RelE	-				

(Continued)

Table 4. (Continued)

Group	Nagasaki locus_tag	Nagasaki Name	Product/ Function	Hit <sup>a</sup>	D74 locus_tag	% Identity <sup>b</sup>
	A2U21_05930	<i>hicB2</i>	Antitoxin HicB	-		
	A2U21_06130	<i>higA4</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U20_05030	98
	A2U21_06135	<i>higB3</i>	mRNA interferase toxin HigB	bi	A2U20_05035	99
	A2U21_06150	<i>tabA</i>	Putative Toxin-antitoxin biofilm protein TabA	bi	A2U20_05585	100
	A2U21_06185	<i>chpS</i>	Antitoxin	bi	A2U20_05555	99
	A2U21_06190	<i>pemK</i>	Programmed cell death toxin PemK	bi	A2U20_05550	97
	A2U21_06245	<i>higA5</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U20_05460	100
	A2U21_06250	<i>higB4</i>	mRNA interferase toxin HigB	bi	A2U20_05455	99
	A2U21_06305	<i>higB5</i>	mRNA interferase toxin HigB	bi	A2U20_05390	100
	A2U21_06310	<i>higA6</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U20_05385	99
	A2U21_06560	<i>cdtA</i>	Toxin	bi	A2U20_04835	99
	A2U21_06565	<i>cdtB</i>	Cytolethal distending toxin subunit CdtB	bi	A2U20_05125	96
	A2U21_06570	<i>cdtC</i>	Toxin	bi	A2U20_04825	99
	A2U21_06840	<i>cdtA2</i>	Toxin	uni	A2U20_04835	99
	A2U21_06845	<i>cdtB2</i>	Cytolethal distending toxin subunit CdtB	uni	A2U20_04830	94
	A2U21_06850	<i>cdtC2</i>	Toxin	uni	A2U20_04825	94
	A2U21_08015	<i>relE3</i>	Addiction module toxin RelE	-		
	A2U21_08020	<i>stbD</i>	RelB/StbD replicon stabilization protein; antitoxin to RelE/StbE	-		
	A2U21_08785	<i>tdeA</i>	Putative membrane toxin/ drug export protein A	bi	A2U20_02805	97
	A2U21_08910	<i>higB6</i>	mRNA interferase toxin HigB	bi	A2U20_02710	96
	A2U21_08915	<i>higA7</i>	Antitoxin HigA_mRNA interferase antitoxin	bi	A2U20_02705	93
	A2U21_09500	<i>hipA2</i>	Toxin HipA	bi	A2U20_02115	95
	A2U21_09900	<i>vapD</i>	Virulence-associated protein D; endoribonuclease	-		
	A2U21_10000	<i>vapC</i>	VapC toxin family PIN domain ribonuclease	bi	A2U20_01850	78
Other	A2U21_03770	<i>sodA</i>	Superoxide dismutase	bi	A2U20_07890	100
	A2U21_04050	<i>sodC</i>	Superoxide dismutase	bi	A2U20_08325	92
	A2U21_06535	<i>sirB</i>	Invasion protein expression up-regulator SirB	bi	A2U20_05155	100
	A2U21_06540	<i>sirA</i>	Regulator of disulfide bond formation	bi	A2U20_05150	99
	A2U21_08285	<i>vacJ</i>	putative VacJ lipoprotein	bi	A2U20_03355	97
	A2U21_09640	<i>espP</i>	Putative serine protease	bi	A2U20_01715	78
	A2U21_09655	<i>espP2</i>	Putative serine protease	bi	A2U20_01700	75
	A2U21_11540	<i>nanH</i>	Sialidase	bi	A2U20_00515	99

<sup>a</sup> The Hit column contains a '-' (no hit), 'uni' or 'bi' RAST server results from a one-to-one BLASTP comparison of the protein coding sequences in the Nagasaki genome using the D74 genome as the reference. "bi" represents a bidirectional best hit in which the reverse hit from the Nagasaki comparison genome to the D74 reference genome was also the best hit. "uni" indicates a uni-directional hit in which the reverse hit from the comparison genome to the reference genome was not also the best hit. "-" indicates no hit or match was found.

<sup>b</sup> Global pairwise nucleotide percent sequence identity.

<https://doi.org/10.1371/journal.pone.0205700.t004>

*relBE*, *mazEF*, *vapBC*, and *higBA*, which was the most abundant with 10 putative *higBA* loci identified in D74 and 6 putative *higBA* loci identified in Nagasaki, with a varying degree of similarity (Table 3 and Table 4).

Eight orthologs of other virulence-associated proteins were identified in both strains (Table 3 and Table 4). A 75% sequence identity was observed between corresponding orthologs *espP* (A2U20\_01700) in D74 and *espP2* (A2U21\_09655) in Nagasaki and a 78% sequence identity was observed between corresponding orthologs *espP2* (A2U20\_01715) in D74 and *espP* (A2U21\_09640) in Nagasaki (Table 3 and Table 4). Additionally a notable size difference

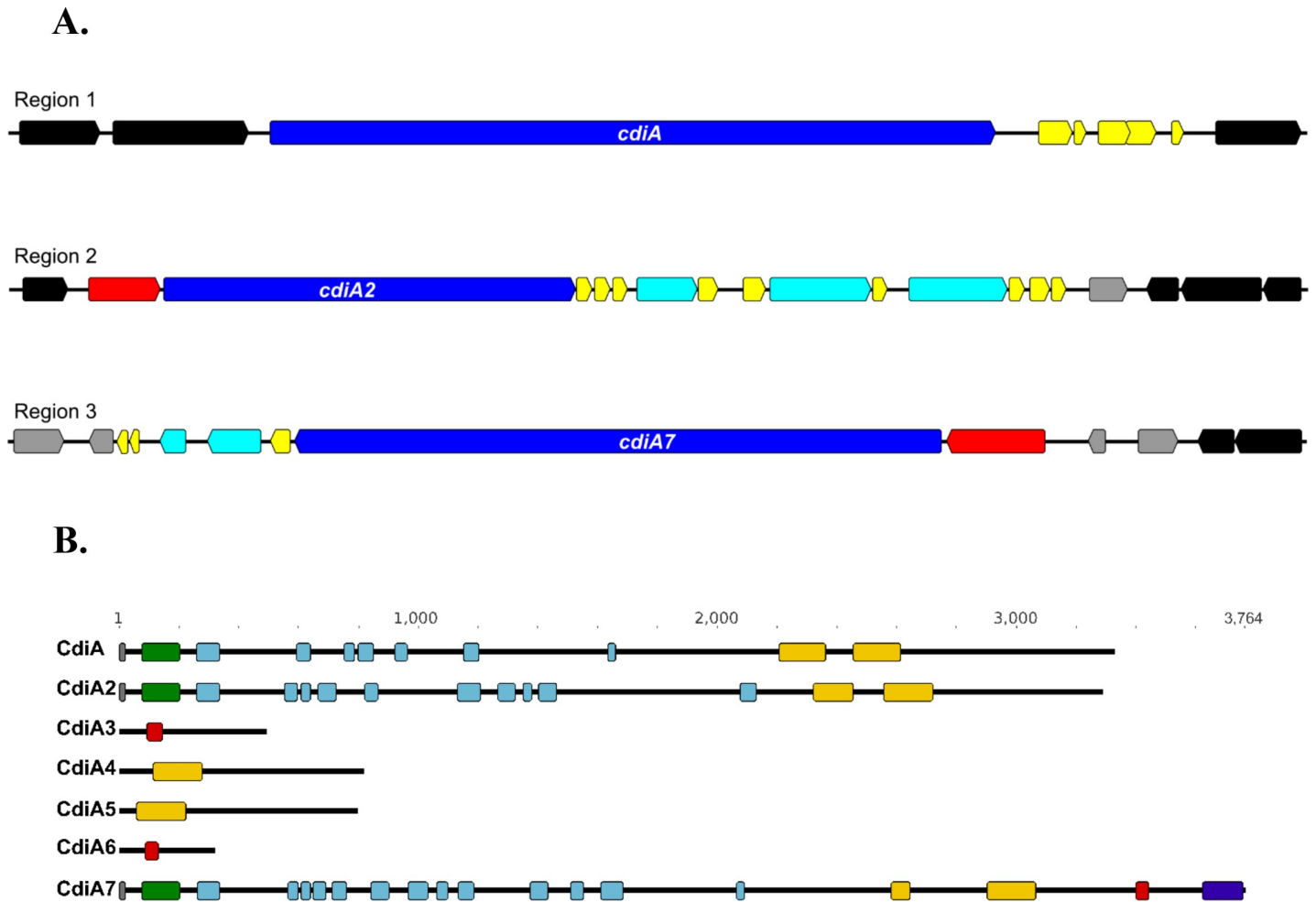
was also observed for the putative proteins encoded by *espP* (A2U20\_01700), 1,071 amino acids, and *espP2* (A2U20\_01715), 985 amino acids, in D74, compared to the putative proteins encoded by *espP* (A2U21\_09640), 781 amino acids, and *espP2* (A2U21\_09655), 772 amino acids, in Nagasaki.

### ***cdiA* genes encoding putative contact-dependent growth inhibition proteins identified in *H. parasuis* D74**

Contact-dependent growth inhibition (CDI) is a process used by Gram-negative bacteria to deliver diverse growth inhibiting nuclease toxins into the cytoplasm of neighboring cells upon cell-cell contact [57–61]. CDI is mediated by the CdiB and CdiA proteins, which are members of the TpsB and TpsA two-partner secretion (TPS) group of proteins [61–63]. CdiB facilitates secretion of the CdiA “exoprotein” onto the cell surface [61, 64]. CdiA then binds to specific outer-membrane receptors on susceptible bacteria and transfers its C-terminal toxin domain (CdiA-CT) into the target cell [58–61]. CDI+ bacteria also produce small immunity proteins (CdiI) that protect them from toxin delivered by neighboring cells of closely related species or sibling cells by binding to the CdiA-CT and neutralizing its toxin activity [58–61].

The putative *cdiA* genes in strain D74 are located in three regions with a *cdiBAI* organization with “orphan” *cdiA* genes located downstream, which is similar to the organization found in *E. coli* strains and other gamma-proteobacteria [61, 65]. Region one includes *cdiA* (A2U20\_05480) and five upstream potential *cdiI* candidate genes (locus tags A2U20\_05485–A2U20\_05505) (Fig 3A). The second region encompasses *fhaC* (A2U20\_06825) located upstream of *cdiA2* (A2U20\_06830), which, consistent with other TpsB proteins, shares substantial homology with CdiB (Fig 3A) [66]. Nine potential *cdiI* candidate genes (locus tags A2U20\_06835–A2U20\_06845, A2U20\_06855–A2U20\_06860, A2U20\_06870, and A2U20\_06880–A2U20\_06890) and *cdiA3*, *cdiA4*, and *cdiA5* (locus tags A2U20\_06850, A2U20\_06865, A2U20\_06875) are located downstream of *cdiA2*. Genes *cdiA3*, *cdiA4*, and *cdiA5* were identified as “orphan” *cdiA* genes based on their shared characteristics such as smaller size compared to larger full-length *cdiA* genes and resemble the 3'-ends of larger *cdiA* genes. (Fig 3A) [65]. The third region contains the *cdiA7* (A2U20\_08295) gene with the *shlB* gene (A2U20\_08300) gene located upstream (Fig 3). Similar to FhaC and other TpsB proteins, ShlB shares substantial homology with CdiB [66]. A potential *cdiI* candidate (A2U20\_08290), followed by two potential “orphan” *cdiA* genes, *cdiA6* (A2U20\_08285) and a predicted pseudogene *fha* (A2U20\_08280), followed by two potential *cdiI* candidate genes (locus tags A2U20\_08270 and A2U20\_08265) are located downstream of *cdiA7* (Fig 3A).

CdiA proteins share a number of characteristics in common with the TpsA family protein FHA [61, 67]. CdiA proteins typically contain an N-terminal region homologous to FHA containing the TPS domain required for interaction with the TpsB partner, CdiB or FhaC respectively, and haemagglutinin repeats that are predicted to form a  $\beta$ -helical structure [61, 67]. In addition, most *cdiA* homologues encode the VENN peptide motif, which delineates the beginning of C-terminal toxin domain as well as the conserved and variable regions [61, 67]. The predicted proteins encoded by the *cdiA* genes in D74 were evaluated for the presence of these domains. Domains identified in CdiA include a ESPR signal peptide domain (PF13018) required for export and multiple haemagglutination activity domains, specifically one Haemagg\_act domain (PF05860), seven Fil\_haemagg domains (PF05594), two Fil\_haemagg\_2 domains (PF13332) (Fig 3B). CdiA2 is similar to CdiA and contains a ESPR signal peptide domain (PF13018) required for export and multiple haemagglutination activity domains, including one Haemagg\_act domain (PF05860), ten Fil\_haemagg domains (PF05594), two Fil\_haemagg\_2 domains (PF13332) (Fig 3B). CdiA3 and CdiA6 are similar to each other and



**Fig 3. *cdiA* genes encoding putative contact-dependent growth inhibition proteins identified in *H. parasuis* D74.** (A) Organization of *cdiA* loci in D74. The three genomic regions containing putative *cdiA* genes are depicted. Gene function was assigned based on the results of BLASTX searches. Each arrow represents a gene within the locus; direction of the arrow indicates orientation within the closed genome sequence. Dark blue arrows represent the *cdiA* genes, red arrows represent *cdiB* homologues, light blue arrows represent potential "orphan" *cdiA* genes, yellow arrows represent putative *cdiI* candidates, and grey arrows represent genes predicted to have functions relating to horizontal gene transmission, black arrows represent genes whose predicted function is unrelated to contact-dependent inhibition. Genomic regions are not shown to scale. (B) Domain architecture of the predicted CdiA proteins. Domain content of the CdiA proteins was determined using a pfam database search. Grey boxes represent ESPR signal peptide domains (PF13018), green boxes represent Haemagg\_act domains (PF05860), light blue boxes represent Fil\_haemagg domains (PF05594), orange boxes represent Fil\_haemagg\_2 domains (PF13332), red boxes represent PT-VENN domains (PF04829), and purple boxes represent EndoU\_bacteria domains (PF14436).

<https://doi.org/10.1371/journal.pone.0205700.g003>

contain a pre-toxin domain with VENN motif that marks the beginning of the C-terminal toxin domain similar to other previously reported CdiA proteins (Fig 3B). CdiA4 and CdiA5 are also similar to each other and both contain a Fil\_haemagg\_2 domain (PF13332) (Fig 3B). CdiA7 is the largest CdiA protein in D74 and contains a ESPR signal peptide domain (PF13018) required for export and multiple haemagglutination activity domains, including one Haemagg\_act domain (PF05860), thirteen Fil\_haemagg domains (PF05594), two Fil\_haemagg\_2 domains (PF13332), a pre-toxin VENN domain (PF04829), and a EndoU\_bacterial nuclease domain (PF14436) (Fig 3B). Overall, *cdiA3*, *cdiA4*, *cdiA5*, and *cdiA6* are similar to previously reported "orphan" *cdiA* genes as they encode much smaller proteins that lack a conserved export signal and other functional domains. The region containing the orphan *cdiA* genes was further evaluated and no additional *cdi*-related protein domains were found beyond



those depicted in Fig 3B for CdiA3, CdiA4, CdiA5, and CdiA6. This indicates that the orphan *cdiA* genes are not the result of frameshift or indel mutations causing disruption of a larger intact *cdiA* gene(s). This genomic organization of a *cdi* locus containing one or more orphan *cdiA* genes downstream of full-length *cdiA* is common in many species of bacteria [65]. In contrast, CdiA, CdiA2, and CdiA7 are larger and contain most of the functional domains harbored by previously characterized CdiA-CT proteins; however, CdiA7 shares more similarity to other CdiA-CT proteins in that it additionally contains a PT-VENN motif and a recognizable nuclease domain at C-terminus.

### ***vtaA* family of trimeric autotransporter genes identified in *H. parasuis* strains D74 and Nagasaki**

Pina et al. first identified thirteen proteins encoded by the *vtaA* family of trimeric autotransporter genes in strain Nagasaki based on the occurrence of a C-terminal YadA anchor domain, which defines this family of proteins [15]. That study also identified 17 homologues harbored by other *H. parasuis* strains with relatively conserved sequence within the passenger domain among *vtaA* homologues from pathogenic isolates and a high degree of divergence among non-virulent isolates [15]. The authors subsequently named these genes *vtaA* or virulence-associated trimeric autotransporter genes and classified the proteins encoded by *vtaA* genes into three groups based on sequence comparison of the C-terminal YadA anchor domain, with groups 1 and 2 being strongly associated with virulent *H. parasuis* isolates [15]. Previous studies have demonstrated that VtaA proteins are involved in virulence as well as being immunogenic, are produced during an infection, and are capable of conferring protection [68–70]. Recently, a comparison between the D74 and Nagasaki draft genomes identified only three *vtaA* genes harbored by D74 compared to thirteen harbored by Nagasaki [34]. Unfortunately, many *vtaA* genes identified within each strain were incomplete in the draft sequences, preventing a reliable one-to-one assignment of the *vtaA*-like ORFs to specific *vtaA* genes and subsequent evaluation of the predicted protein structure.

To ensure identification of all potential *vtaA* genes, all protein coding sequences in Nagasaki and D74 were searched for the occurrence of a YadA anchor domain and no additional YadA anchor domain containing proteins were identified. The 13 *vtaA* genes identified in Nagasaki have been named according to their location along the chromosome and are listed in Table 5 along with their respective name and group originally assigned by Pina et al. [15]. Similarly, the 3 *vtaA* genes identified in D74 have been named according to their location along the chromosome and are also listed in Table 5. The predicted protein structure of all VtaA proteins for both Nagasaki and D74 were evaluated for known domain content (Fig 4). All Nagasaki VtaA proteins contain an N-terminal extended signal peptide or ESPR domain (PF13018) for Type V secretion, followed by 1–4 YadA head domains (PF05658), and 3–5 YadA anchor domains (PF03895). This region containing the head and stalk domains is referred to as the passenger domain region [71]. Following the passenger domain region, all Nagasaki VtaA proteins contain 2–8 collagen triple helix repeat domains (PF01391) followed by the C-terminal YadA anchor domain (PF03895) (Fig 4A). In contrast, none of the D74 VtaA proteins contain a collagen triple helix repeat domain (PF01391) and the predicted size for both VtaA\_D1 (4054 AA) and VtaA\_D2 (6778 AA) proteins is substantially larger than any predicted Nagasaki VtaA protein (Fig 4B). Additionally, VtaA\_D3 differs from VtaA\_D1 and VtaA\_D2 proteins. VtaA\_D3 contains a tryptophan-ring motif domain or TAA-Trp-ring (PF15401) and does not contain an N-terminal ESPR domain (PF13018) (Fig 4B). The absence of the ESPR domain suggests that VtaA\_D3 may not be exported across the inner membrane. When we compared our VtaA predicted protein sequences to those reported by Pina et al. [15] two

Table 5. *vtaA* genes identified in *H. parasuis* Nagasaki.

Nagasaki locus_tag	Nagasaki Name	Name assigned by Pina et al. [15]	Group assigned by Pina et al. [15]	Hit <sup>a</sup>	D74 locus_tag	D74 Name	% Identity <sup>b</sup>
A2U21_00055	vtaA_N1	vtaA1	1	-			0
A2U21_00360	vtaA_N2	vtaA10	2	uni	A2U20_04585	vtaA_D1	49.73
A2U21_00905	vtaA_N3	vtaA4	1	-			0
A2U21_03315	vtaA_N4	vtaA8 <sup>c</sup>	1	bi	A2U20_05995	vtaA_D3	36.84
A2U21_04400	vtaA_N5	vtaA9	1	uni	A2U20_04585	vtaA_D1	30.74
A2U21_05305	vtaA_N6	vtaA6	1	-			0
A2U21_05400	vtaA_N7	vtaA2	1	-			0
A2U21_05640	vtaA_N8	vtaA11	2	uni	A2U20_04585	vtaA_D1	50.94
A2U21_06175	vtaA_N9	vtaA3 <sup>d</sup>	1	-			0
A2U21_06350	vtaA_N10	vtaA12	1	bi	A2U20_04585	vtaA_D1	94.09
A2U21_07110	vtaA_N11	vtaA13	1	uni	A2U20_04585	vtaA_D1	59.87
A2U21_08125	vtaA_N12	vtaA5	3	-			0
A2U21_09355	vtaA_N13	vtaA7	3	-			0

<sup>a</sup> The Hit column contains a '-' (no hit), 'uni' or 'bi' RAST server results from a one-to-one BLASTP comparison of the protein coding sequences in the D74 genome using the Nagasaki genome as the reference. "bi" represents a bidirectional best hit in which the reverse hit from the D74 comparison genome to the Nagasaki reference genome was also the best hit. "uni" represents uni-directional indicates a hit in which the reverse hit from the comparison genome to the reference genome was not also the best hit. "-" indicates no hit or match was found.

<sup>b</sup> Local percent sequence identity.

<sup>c</sup> 18 amino acid insertion compared to the Pina et al. [15] sequence.

<sup>d</sup> Single amino acid difference at position 1191.

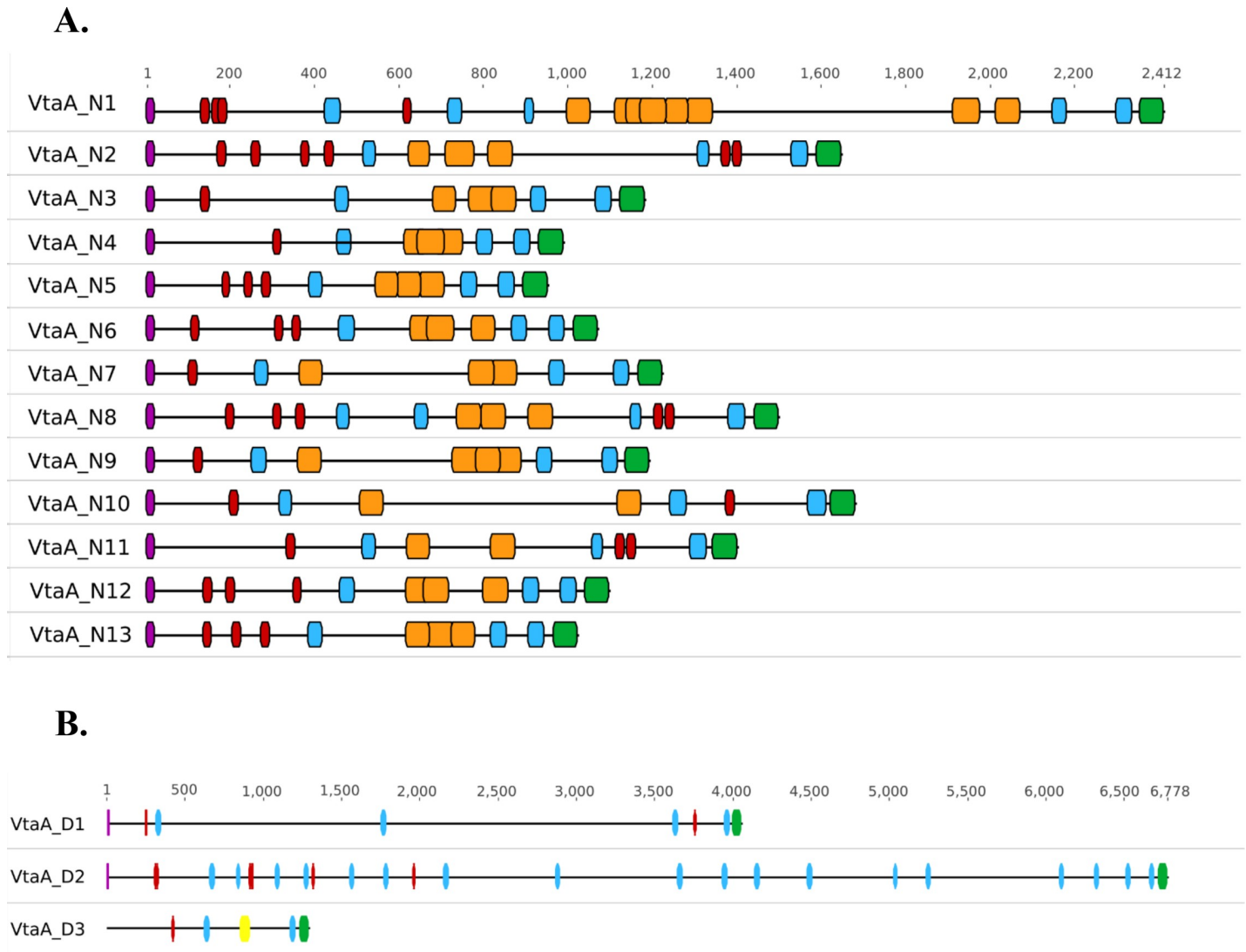
<https://doi.org/10.1371/journal.pone.0205700.t005>

differences were identified and the other 11 out of 13 sequences were found to be 100% identical. The two differences identified were an amino acid change in VtaA\_N9 (S1191G) relative to the previously reported sequence and an 18 amino acid insertion (AGPTGPGQGPAGPTGQ DP) after amino acid 737 of VtaA8 that is not present in our VtaA\_N4 sequence. The absence or presence of the insertion is located within the third collagen repeat domain and does not affect the presence or absence of this domain nor the total number of collagen repeat domains predicted for the two proteins.

When the genome regions in Nagasaki and D74 were compared, the sequence and gene content up and downstream of the *vtaA* gene locations in Nagasaki were similar to the reciprocal locations in D74. No evidence of extensive genome rearrangement at these locations was observed. A different gene or genes were identified in D74 at the same location of a reciprocal *vtaA* gene in Nagasaki, which could have arisen from small insertion events. Further evaluation of the *vtaA* genes in both strains suggested that they are not in an operon configuration. Pina et al. identified an active promoter upstream of *vtaA\_N3* [15]. *In silico* analysis indicated that the upstream region of all 13 Nagasaki *vtaA* genes contained highly similar promoter sequences to the *vtaA\_N3* promoter. A highly similar promoter sequence to the *vtaA\_N3* was additionally observed upstream of *vtaA\_D2* in D74. This sequence conservation implies similar expression and/or regulation mechanisms among these genes.

### Methylation motifs and RM-systems in *H. parasuis* strains D74 and Nagasaki

In bacteria, the most common post-replicative modification of DNA is methylation by methyltransferase (MTase) enzymes resulting in three types of epigenetic markers: N6-methyladenine (m6A), N4-methylcytosine (m4C) and 5-methylcytosine (m5C) [47, 72]. DNA methylation



**Fig 4. Domain architecture of the predicted VtaA proteins in *H. parasuis* strains D74 and Nagasaki.** (A) Domain architecture of the predicted VtaA proteins from Nagasaki. Assigned gene name designations are shown at left. Schematic depictions of the 13 Nagasaki VtaA proteins is shown. The domains were identified by a pfam database search. ESPR signal peptides (PF13018) are shown in purple, YadA head domains (PF05658) in red, YadA stalk domains (PF05662) in blue, YadA anchor domains (PF03895) in green, collagen domains (PF1391) in orange, TAA-Trp-ring domains (PF15401) in yellow. (B) Domain architecture of the predicted VtaA proteins from D74. Assigned gene name designations are shown at left. Schematic depictions of the pfam domains, colored as in Panel A, for the 3 VtaA proteins from D74.

<https://doi.org/10.1371/journal.pone.0205700.g004>

serves several key roles in bacterial processes, including mismatch repair, the timing of DNA replication, conferring protection against bacteriophages, and regulating gene expression. [73–78]. Analysis of the SMRT DNA sequencing kinetics was used to identify total base modifications in the genomes of *H. parasuis* D74 and Nagasaki, and the modified sequence motifs for each strain are summarized in Tables 6 and 7.

A total of 15 sequence motifs were identified in strain D74, and N6-methyladenine (m6A) was the most prevalent type of modification detected (Table 6). Focusing on strain Nagasaki, a total of 14 recognition sites for methylation or sequence motifs were identified and, similar to D74, N6-methyladenine (m6A) was the most prevalent type of modification detected (Table 7). This analysis revealed a surprising degree of diversity in motifs observed between

**Table 6. Methylation motifs detected in *H. parasuis* D74.**

Motif <sup>a</sup>	Modification Type	# Detected	# in Genome	% Detected	Mean Modification QV	Mean Motif Coverage	Partner Motif
<b>A</b> GCNNNNNGCT	m6A	767	772	99.4%	535.6	376.2	AGCNNNNNGCT
GT <b>A</b> NNNNNNNTGG	m6A	817	825	99.0%	436.6	380.3	CCANNNNNNTAC
CC <b>A</b> NNNNNNNTAC	m6A	813	825	98.5%	421.3	381.9	GTANNNNNNTGG
<b>A</b> AGCTT	m6A	623	630	98.9%	442.4	384.0	AAGCTT
<b>G</b> ATC	m6A	15,534	15,718	98.8%	410.7	381.5	GATC
GT <b>A</b> HNNNNNNCTTG	m6A	217	220	98.6%	420.1	378.9	
CA <b>A</b> GNNNNNNGNTAC	m6A	55	57	96.5%	397.1	348.8	
<b>A</b> GCNNNNGGATC	m6A	47	55	85.5%	346.8	365.8	
GC <b>A</b> GGVNNNDG	m6A	316	659	48.0%	99.1	382.2	
VA <b>A</b> AGCTCKD	m6A	207	447	46.3%	131.9	391.1	
<b>A</b> HBYAGYAD	m6A	662	2,632	25.2%	108.6	376.7	
DD <b>T</b> GTNDNDG	modified_base <sup>b</sup>	1,575	8,346	18.9%	52.3	356.8	
<b>T</b> NNNNNNH	modified_base <sup>b</sup>	179,073	1,197,696	15.0%	52.7	367.5	
DT <b>N</b> VVNDG	modified_base <sup>b</sup>	9,920	78,834	12.6%	48.6	366.4	
<b>A</b> GNNNNNH	m6A	15,854	200,853	7.9%	112.6	371.8	

<sup>a</sup>Bold underlined bases indicate methylated base in motif sequence.

<sup>b</sup>Base modification not identified or recognized by software.

<https://doi.org/10.1371/journal.pone.0205700.t006>

these closely related strains given that the methylation motif 5' -GA<sup>m6</sup>TC-3' was the only motif shared or observed in both D74 and Nagasaki (Table 6 and Table 7).

The genomes of *H. parasuis* Nagasaki and D74 were assessed using the Restriction Enzyme Database REBASE ([www.rebase.neb.com](http://www.rebase.neb.com)) [47] for determination of putative MTases involved with each motif and for comparisons with known modification systems. A total of 26 genes associated with restriction-modification systems were identified in *H. parasuis* D74, including

**Table 7. Methylation motifs detected in *H. parasuis* Nagasaki.**

Motif <sup>a</sup>	Modification Type	# Detected	# in Genome	% Detected	Mean Modification QV	Mean Motif Coverage	Partner Motif
<b>A</b> GNNNNNCTT	m6A	1,322	1,324	99.8%	719.7	571.6	AAGNNNNCTT
<b>A</b> CNNNNNTGG	m6A	1,089	1,092	99.7%	676.2	592.7	CCANNNNNGTT
CC <b>A</b> NNNNNGTT	m6A	1,088	1,092	99.6%	700.0	622.7	AACNNNNNTGG
GT <b>A</b> NNNNNNCTTG	m6A	217	218	99.5%	641.8	610.1	CAAGNNNNNNNTAC
CA <b>A</b> GNNNNNNTAC	m6A	215	218	98.6%	671.3	582.9	GTANNNNNNNCTTG
<b>G</b> ATC	m6A	14,603	14,778	98.8%	602.6	617.8	GATC
<b>A</b> GGNNNNCCT	m6A	454	460	98.7%	688.5	590.4	AGGNNNNCCT
<b>A</b> AGVNNNNCTT	m6A	303	1,013	29.9%	130.6	608.1	
<b>R</b> AHDBAGYA	m6A	721	2,881	25.0%	107.0	582.3	
<b>T</b> NNNNNNH	modified_base <sup>b</sup>	181,179	1,123,132	16.1%	58.7	579.9	
DD <b>T</b> NVVNDG	modified_base <sup>b</sup>	9,439	61,834	15.3%	54.8	582.8	
<b>T</b> SNKNGG	modified_base <sup>b</sup>	7,871	64,478	12.2%	52.9	587.1	
<b>A</b> GDNNNNH	m6A	11,510	134,477	8.6%	178.5	591.9	
<b>B</b> AGCVNNNH	m6A	1,133	24,217	4.7%	140.2	629.3	

<sup>a</sup>Bold underlined bases indicate methylated base in motif sequence.

<sup>b</sup>Base modification not identified or recognized by software.

<https://doi.org/10.1371/journal.pone.0205700.t007>

**Table 8. Putative *H. parasuis* D74 restriction modification systems.**

Type <sup>a</sup>	Gene <sup>b</sup>	D74 locus_tag	D74 Name	Predicted Recognition Sequence	REBASE Name
I	M	A2U20_05240	hsdM		M.Hpa74ORFFP
I	S	A2U20_05245	hsdS		S.Hpa74ORFFP
I	R	A2U20_05260	hsdR2		Hpa74ORFFP
I	R	A2U20_10200	hsdR3		Hpa74ORFIP
I	S	A2U20_10215	hsdS2		S1.Hpa74ORFIP
I	S	A2U20_10220	hypothetical protein CDS		S2.Hpa74ORFIP
I	M	A2U20_10230	hsdM2		M.Hpa74ORFIP
I	R	A2U20_10460 <sup>c</sup>	hsdR4 <sup>c</sup>		Hpa74ORFJP
I	S	A2U20_11315	hsdS3	GTANNNNNNCTTG	S.Hpa74III
I	R	A2U20_11320	hsdR5	GTANNNNNNCTTG	Hpa74III
I	M	A2U20_11330	hsdM3	GTANNNNNNCTTG	M.Hpa74III
II	M	A2U20_00375	hindIIIM	AAGCTT	M.Hpa74I
II	R	A2U20_00380 <sup>c</sup>	hindIIIR <sup>c</sup>	AAGCTT	Hpa74IP
II	RM	A2U20_00575	bcgIA		Hpa74ORFBP
II	S	A2U20_00580	bcgIB		S.Hpa74ORFBP
II	S	A2U20_00800	bcgIB		S1.Hpa74ORFDP
II	S	A2U20_00805 <sup>c</sup>	bcgIB2 <sup>c</sup>		S2.Hpa74ORFDP
II	M	A2U20_03405	dam	GATC	Hpa74II
II	RM	A2U20_07465	hypothetical protein CDS		Hpa74ORFGP
II	R	A2U20_07700	restriction endonuclease CDS	GGCC	Hpa74ORFHP
II	M	A2U20_07705	haeIIIM	GGCC	M.Hpa74ORFHP
II	M	A2U20_10010	yhdJ		M.Hpa74ORFOP
II	M	A2U20_10475	modification methylase CDS		M.Hpa74ORFLP
II	R	A2U20_10480	HNH endonuclease CDS		Hpa74ORFLP
II	S	A2U20_10485	bcgIB3		S.Hpa74ORFMP
II	RM	A2U20_10490	bcgIA2		Hpa74ORFMP

<sup>a</sup>Systems were designated Type I, II, or III based on REBASE analyses.

<sup>b</sup>Gene designations of methylase (M), restriction (R), fused restriction-modification (RM), or specificity (S), along with predicted recognition sequence and REBASE name, were determined using REBASE analysis.

<sup>c</sup>Predicted pseudogene.

<https://doi.org/10.1371/journal.pone.0205700.t008>

11 genes associated with Type 1 restriction-modification (RM) systems and 15 genes associated with Type II RM systems (Table 8). Genes associated with Type III RM systems were not identified in D74 (Table 8). REBASE predicted three recognition sequences corresponding to a specific motif detected by the SMRT sequencing analysis. REBASE analysis indicated that the putative Type I RM enzymes S.Hpa74III, Hpa74III, and M.Hpa74III were predicted to be responsible for the 5' -GTA<sup>m6</sup>NNNNNNNCTTG-3' modification (Table 8). The putative Type II RM enzymes M.Hpa74I and M.Hpa74IP were indicated by REBASE to be responsible for the motif 5'-A<sup>m6</sup>A GCTT-3' modification (Table 8). The putative Type II RM enzyme *dam* or Hpa74II was predicted to be responsible for the 5'-GA<sup>m6</sup>TC-3' modification for D74 (Table 8). The remaining two motifs detected by the SMRT sequencing analysis did not correspond with a REBASE predicted recognition sequence. Two putative Type II RM enzymes Hpa74ORFHP and M.Hpa74ORFHP were predicted by REBASE to recognize the sequence motif 5'-GGC C-3, which was not a motif detected by the SMRT sequencing analysis. Three putative RM enzyme genes identified in D74 by REBASE analysis are predicted pseudogenes. These include S2.Hpa74ORFDP, Hpa74ORFJP, and M.Hpa74IP, associated with the Type II RM system

**Table 9. Putative *H. parasuis* Nagasaki restriction modification systems.**

Type <sup>a</sup>	Gene <sup>b</sup>	Nagasaki locus_tag	Nagasaki Name	Predicted Recognition Sequence <sup>b</sup>	REBASE Name <sup>b</sup>
I	M	A2U21_00095	hsdM	GTANNNNNNNCTTG	M.HpaNNII
I	R	A2U21_00110	hsdR	GTANNNNNNNCTTG	HpaNNIIP
I	S	A2U21_00115	hsdS	GTANNNNNNNCTTG	S.HpaNNII
I	R	A2U21_01085	hsdR2		HpaNNORFEP
I	S	A2U21_01100	hsdS2		S1.HpaNNORFEP
I	S	A2U21_01105	hypothetical protein CDS		S2.HpaNNORFEP
I	M	A2U21_01115	type I restriction-modification system subunit M CDS		M.HpaNNORFEP
I	R	A2U21_04210	hsdR4		HpaNNORFIP
I	S	A2U21_04220	hsdS3		S1.HpaNNORFEP
I	S	A2U21_04225	hsdS4		S2.HpaNNORFEP
I	M	A2U21_04230	hsdM2		M.HpaNNORFEP
I	R	A2U21_06440	hsdR5		HpaNNORFJP
I	S	A2U21_06445	hsdS5		S.HpaNNORFJP
I	M	A2U21_06450	hsdM3		M.HpaNNORFJP
II	S	A2U21_00855	bcgIB		S1.HpaNNORFDP
II	S	A2U21_00860	bcgIB2		S2.HpaNNORFDP
II	RM	A2U21_00865	bcgIA		HpaNNORFDP
II	M	A2U21_03345	hypothetical protein CDS		M.HpaNNORFGP
II	M	A2U21_03850	hhalM	GCGC	M.HpaNNORFHP
II	R	A2U21_03855	type II RM endonuclease	GCGC	HpaNNORFHP
II	M	A2U21_06885	hpaIIM	CCGG	M.HpaNNORFAP
II	M	A2U21_08235	dam	GATC	M.HpaNNI
II	M	A2U21_09040	bspRIM	CGCG	M.HpaNNORFLP
II	R	A2U21_09045	hypothetical protein CDS	CGCG	HpaNNORFLP
II	M	A2U21_09985	restriction endonuclease subunit M CDS		M1.HpaNNORFMP
II	M	A2U21_09990	restriction endonuclease CDS		M2.HpaNNORFMP
II	RM	A2U21_09995	restriction endonuclease CDS		HpaNNORFMP
II	M	A2U21_10160	hypothetical protein CDS		M.HpaNNORFNP
II	M	A2U21_10520	hypothetical protein CDS	BA	M.HpaNNORFOP
III	M	A2U21_01380	restriction endonuclease subunit M CDS	GGAG	M1.HpaNNORFFP
III	M	A2U21_01385	bamHIM	GGAG	M2.HpaNNORFFP
III	R	A2U21_01390	type III restriction endonuclease subunit R CDS		HpaNNORFFP
III	M	A2U21_11010	site-specific DNA-methyltransferase CDS		M.HpaNNORFPP
III	R	A2U21_11015	restriction endonuclease CDS		HpaNNORFPP

<sup>a</sup>Systems were designated Type I, II, or III based on REBASE analyses.

<sup>b</sup>Gene designations of methylase (M), restriction (R), fused restriction-modification (RM), or specificity (S), along with predicted recognition sequence and REBASE name, were determined using REBASE analysis.

<https://doi.org/10.1371/journal.pone.0205700.t009>

indicated by REBSE to be responsible for the motif 5' -A<sup>m6</sup>A GCTT-3' modification (Table 8). In contrast, none of the putative RM enzymes genes identified in Nagasaki by REBASE analysis are predicted pseudogenes.

Focusing on *H. parasuis* Nagasaki, 34 genes associated with restriction-modification systems were identified, including 14 genes associated with Type 1, 15 genes associated with Type II, and 5 genes associated with Type III RM systems (Table 9). Only two of the REBASE predicted recognition sequences corresponded to a specific motif detected by the SMRT sequencing analysis. REBASE analysis indicated that the putative Type I RM enzymes M.HpaNNII,

HpaNNIIP, and S.HapNNII were predicted to be responsible for the 5' -GTA<sup>m6</sup>NNNNNNN CTTG-3' modification (Table 9). The putative Type II RM enzyme *dam* or M.HpaNNI was predicted to be responsible for the 5'-GA<sup>m6</sup>TC-3' modification (Table 9). The remaining five motifs identified by the SMRT sequencing analysis represent yet unknown recognition sequences.

Expression of MTases can undergo phase variation by slipped-strand mispairing due to the presence of simple sequence repeats (SSRs), such as homopolymeric tracts [79–83]. All of the putative RM genes identified in D74 (Table 8) and Nagasaki (Table 9) were searched for the presence of SSRs within the coding region and in the region encompassing 150 bp upstream of the putative start codon. No SSRs were observed in the upstream region for five RM genes identified in D74 (A2U20\_00805c, A2U20\_10480, A2U20\_10215, A2U20\_11315, and A2U20\_10200), while six homopolymeric tracts of consisting of five or more bases were observed in the upstream region of *hsdR2* (A2U20\_05260) (S6 Table). SSRs were observed within the coding region of all of the D74 RM genes and the numbers of SSRs ranged from two (A2U20\_00580 and A2U20\_11330) to 38 (A2U20\_07465) (S6 Table). No SSRs were observed in the upstream region for six RM genes identified in Nagasaki (A2U21\_00115, A2U21\_01085, A2U21\_04210, A2U21\_03345, A2U21\_03850, A2U21\_06885), while six homopolymeric tracts of consisting of five or more bases were observed in the upstream region of A2U21\_09985 (S7 Table). SSRs were observed within the coding region of all of the Nagasaki RM genes and the numbers of SSRs ranged from two (A2U21\_10520) to 31 (A2U21\_04210 and A2U21\_00110) (S7 Table). While further studies are warranted, the occurrence of these homopolymeric tracts within these regions indicates the potential of these genes to undergo phase variation by slip strand mispairing.

## Conclusions

This report provides the closed whole-genome sequence annotation and genome-wide methylation patterns for the *H. parasuis* non-virulent D74 strain and for the highly virulent Nagasaki strain. This collective information will enable reliable one-to-one assignment of specific genes of interest and subsequent evaluation of predicted protein structures. Highlights of the information gained from this study include the sequence and annotation of a plasmid harbored by strain D74 that shares a high degree of similarity to other plasmids harbored by members of the *Pasteurellaceae* family, which could prove useful in future allelic replacement and/or functional genomic studies. Evaluation of the virulence-associated genes contained within the genomes of D74 and Nagasaki led to the discovery of a large number of TA systems, primarily Type II TA families, within both genomes. Five predicted hemolysins were identified as unique to Nagasaki and seven putative contact-dependent growth inhibition toxin proteins were identified only in strain D74. Assessment of all potential *vtaA* genes revealed thirteen present in the Nagasaki genome and three in the D74 genome. Subsequent evaluation of the predicted protein structure revealed that none of the D74 VtaA proteins contain a collagen triple helix repeat domain and a much larger predicted amino acid size for two D74 VtaA proteins compared to any predicted Nagasaki VtaA protein. Fifteen methylation sequence motifs were identified in D74 and fourteen methylation sequence motifs were identified in Nagasaki using SMRT sequencing analysis. Only one of the methylation sequence motif was observed in both strains highlighting the diversity between D74 and Nagasaki. Subsequent analysis also revealed diversity in the restriction-modification systems harbored by D74 and Nagasaki. Our hope is that the assembly and annotation of these genomes, coupled with the comparative genomic analyses reported in this study, will aid in the identification of genetic elements that underlie and influence phenotypic differences between these isolates. Together, this

information can support future research and the development of vaccines with improved efficacy towards *H. parasuis* in swine to decrease the prevalence and disease burden caused by this pathogen.

## Supporting information

**S1 Table. D74 CDS list.**

(XLSX)

**S2 Table. Nagasaki CDS list.**

(XLSX)

**S3 Table. AMR MIC data.**

(XLSX)

**S4 Table. D74 capsule genes.**

(DOCX)

**S5 Table. Nagasaki capsule genes.**

(DOCX)

**S6 Table. D74 RM gene SSRs.**

(XLSX)

**S7 Table. Nagasaki RM gene SSRs.**

(XLSX)

**S1 File. D74 annotations.**

(GB)

**S2 File Nagasaki annotations.**

(GB)

**S3 File pD74 annotations.**

(GB)

## Acknowledgments

We thank Iowa State University Veterinary Diagnostic Laboratory for assistance in producing the phenotypic antibiotic resistance data and Yale Center for Genomic Analysis for its assistance in producing the PacBio sequence data. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

## Author Contributions

**Conceptualization:** Tracy L. Nicholson, Brian W. Brunelle.

**Data curation:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Formal analysis:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Funding acquisition:** Tracy L. Nicholson.



**Investigation:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Methodology:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Resources:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Software:** Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Supervision:** Tracy L. Nicholson.

**Validation:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Writing – original draft:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

**Writing – review & editing:** Tracy L. Nicholson, Brian W. Brunelle, Darrell O. Bayles, David P. Alt, Sarah M. Shore.

## References

1. Biberstein EL, White DC. A proposal for the establishment of two new *Haemophilus* species. *J Med Microbiol.* 1969; 2(1):75–8. Epub 1969/02/01. <https://doi.org/10.1099/00222615-2-1-75> PMID: 5821848.
2. Aragon V, Segales J, Oliveira S. Glasser's Disease. In: Zimmerman JJ, Krieger LA, Ramirez A, S K.J., Stevenson GW, editors. *Diseases of Swine.* 10 ed. West Sussex, UK: Wiley-Blackwell; 2012. p. 760–9.
3. Costa-Hurtado M, Aragon V. Advances in the quest for virulence factors of *Haemophilus parasuis*. *Vet J.* 2013; 198(3):571–6. Epub 2013/10/03. <https://doi.org/10.1016/j.tvjl.2013.08.027> PMID: 24084037.
4. Zhang B, Tang C, Liao M, Yue H. Update on the pathogenesis of *Haemophilus parasuis* infection and virulence factors. *Vet Microbiol.* 2014; 168(1):1–7. Epub 2013/08/27. <https://doi.org/10.1016/j.vetmic.2013.07.027> PMID: 23972951.
5. Little TW. *Haemophilus* infection in pigs. *Vet Rec.* 1970; 87(14):399–402. Epub 1970/10/03. PMID: 5531588.
6. Riley MG, Russell EG, Callinan RB. *Haemophilus parasuis* infection in swine. *J Am Vet Med Assoc.* 1977; 171(7):649–51. Epub 1977/10/01. PMID: 914694.
7. Brockmeier SL. Prior infection with *Bordetella bronchiseptica* increases nasal colonization by *Haemophilus parasuis* in swine. *Vet Microbiol.* 2004; 99(1):75–8. Epub 2004/03/17. <https://doi.org/10.1016/j.vetmic.2003.08.013> PMID: 15019114.
8. Kielstein P, Rapp-Gabrielson VJ. Designation of 15 serovars of *Haemophilus parasuis* on the basis of immunodiffusion using heat-stable antigen extracts. *J Clin Microbiol.* 1992; 30(4):862–5. Epub 1992/04/01. PMID: 1572971; PubMed Central PMCID: PMCPMC265175.
9. Tadjine M, Mittal KR, Bourdon S, Gottschalk M. Development of a new serological test for serotyping *Haemophilus parasuis* isolates and determination of their prevalence in North America. *J Clin Microbiol.* 2004; 42(2):839–40. Epub 2004/02/10. <https://doi.org/10.1128/JCM.42.2.839-840.2004> PMID: 14766867; PubMed Central PMCID: PMCPMC344452.
10. Howell KJ, Weinert LA, Luan SL, Peters SE, Chaudhuri RR, Harris D, et al. Gene content and diversity of the loci encoding biosynthesis of capsular polysaccharides of the 15 serovar reference strains of *Haemophilus parasuis*. *J Bacteriol.* 2013; 195(18):4264–73. Epub 2013/07/23. <https://doi.org/10.1128/JB.00471-13> PMID: 23873912; PubMed Central PMCID: PMCPMC3754760.
11. Howell KJ, Peters SE, Wang J, Hernandez-Garcia J, Weinert LA, Luan SL, et al. Development of a Multiplex PCR Assay for Rapid Molecular Serotyping of *Haemophilus parasuis*. *J Clin Microbiol.* 2015; 53(12):3812–21. Epub 2015/10/02. <https://doi.org/10.1128/JCM.01991-15> PMID: 26424843; PubMed Central PMCID: PMCPMC4652097.
12. Jia A, Zhou R, Fan H, Yang K, Zhang J, Xu Y, et al. Development of Serotype-Specific PCR Assays for Typing of *Haemophilus parasuis* Isolates Circulating in Southern China. *J Clin Microbiol.* 2017; 55(11):3249–57. Epub 2017/09/08. <https://doi.org/10.1128/JCM.00688-17> PMID: 28878007; PubMed Central PMCID: PMCPMC5654909.

13. Costa-Hurtado M, Ballester M, Galofre-Mila N, Darji A, Aragon V. VtaA8 and VtaA9 from *Haemophilus parasuis* delay phagocytosis by alveolar macrophages. *Vet Res.* 2012; 43:57. Epub 2012/07/31. <https://doi.org/10.1186/1297-9716-43-57> PMID: 22839779; PubMed Central PMCID: PMCPMC3462726.
14. Martinez-Moliner V, Soler-Llorens P, Molerés J, Garmendia J, Aragon V. Distribution of genes involved in sialic acid utilization in strains of *Haemophilus parasuis*. *Microbiology.* 2012; 158(Pt 8):2117–24. Epub 2012/05/23. <https://doi.org/10.1099/mic.0.056994-0> PMID: 22609756.
15. Pina S, Olvera A, Barcelo A, Bensaid A. Trimeric autotransporters of *Haemophilus parasuis*: generation of an extensive passenger domain repertoire specific for pathogenic strains. *J Bacteriol.* 2009; 191(2):576–87. Epub 2008/11/18. <https://doi.org/10.1128/JB.00703-08> PMID: 19011035; PubMed Central PMCID: PMCPMC2620822.
16. Ruiz A, Oliveira S, Torremorell M, Pijoan C. Outer membrane proteins and DNA profiles in strains of *Haemophilus parasuis* recovered from systemic and respiratory sites. *J Clin Microbiol.* 2001; 39(5):1757–62. Epub 2001/04/28. <https://doi.org/10.1128/JCM.39.5.1757-1762.2001> PMID: 11325986; PubMed Central PMCID: PMCPMC88021.
17. Wang X, Xu X, Wu Y, Li L, Cao R, Cai X, et al. Polysaccharide biosynthesis protein CapD is a novel pathogenicity-associated determinant of *Haemophilus parasuis* involved in serum-resistance ability. *Vet Microbiol.* 2013; 164(1–2):184–9. Epub 2013/02/26. <https://doi.org/10.1016/j.vetmic.2013.01.037> PMID: 23434184.
18. Zhang B, Xu C, Liao M. Outer membrane protein P2 of the *Haemophilus parasuis* SC096 strain contributes to adherence to porcine alveolar macrophages cells. *Vet Microbiol.* 2012; 158(1–2):226–7. Epub 2012/02/22. <https://doi.org/10.1016/j.vetmic.2012.01.023> PMID: 22342495.
19. Zhang B, Xu C, Zhou S, Feng S, Zhang L, He Y, et al. Comparative proteomic analysis of a *Haemophilus parasuis* SC096 mutant deficient in the outer membrane protein P5. *Microb Pathog.* 2012; 52(2):117–24. Epub 2011/12/17. <https://doi.org/10.1016/j.micpath.2011.11.002> PMID: 22172379.
20. Frandoloso R, Martinez-Martinez S, Gutierrez-Martin CB, Rodriguez-Ferri EF. *Haemophilus parasuis* serovar 5 Nagasaki strain adheres and invades PK-15 cells. *Vet Microbiol.* 2012; 154(3–4):347–52. Epub 2011/08/16. <https://doi.org/10.1016/j.vetmic.2011.07.022> PMID: 21839589.
21. He L, Dai K, Wen X, Ding L, Cao S, Huang X, et al. QseC Mediates Osmotic Stress Resistance and Biofilm Formation in *Haemophilus parasuis*. *Front Microbiol.* 2018; 9:212. Epub 2018/03/01. <https://doi.org/10.3389/fmicb.2018.00212> PMID: 29487590; PubMed Central PMCID: PMCPMC5816903.
22. Wen Y, Wen Y, Wen X, Cao S, Huang X, Wu R, et al. OxyR of *Haemophilus parasuis* is a global transcriptional regulator important in oxidative stress resistance and growth. *Gene.* 2018; 643:107–16. Epub 2017/12/13. <https://doi.org/10.1016/j.gene.2017.12.010> PMID: 29229515.
23. Chen L, Wu D, Cai X, Guo F, Blackall PJ, Xu X, et al. Electrotransformation of *Haemophilus parasuis* with in vitro modified DNA based on a novel shuttle vector. *Vet Microbiol.* 2012; 155(2–4):310–6. Epub 2011/09/20. <https://doi.org/10.1016/j.vetmic.2011.08.020> PMID: 21925811.
24. Blanco I, Galina-Pantoja L, Oliveira S, Pijoan C, Sanchez C, Canals A. Comparison between *Haemophilus parasuis* infection in colostrums-deprived and sow-reared piglets. *Vet Microbiol.* 2004; 103(1–2):21–7. Epub 2004/09/24. <https://doi.org/10.1016/j.vetmic.2004.06.011> PMID: 15381262.
25. Oliveira S, Galina L, Blanco I, Canals A, Pijoan C. Naturally-farrowed, artificially-reared pigs as an alternative model for experimental infection by *Haemophilus parasuis*. *Can J Vet Res.* 2003; 67(2):146–50. Epub 2003/05/23. PMID: 12760482; PubMed Central PMCID: PMCPMC227044.
26. Luan SL, Chaudhuri RR, Peters SE, Mayho M, Weinert LA, Crowther SA, et al. Generation of a Tn5 transposon library in *Haemophilus parasuis* and analysis by transposon-directed insertion-site sequencing (TraDIS). *Vet Microbiol.* 2013; 166(3–4):558–66. Epub 2013/08/10. <https://doi.org/10.1016/j.vetmic.2013.07.008> PMID: 23928120.
27. Liu H, Xue Q, Zeng Q, Zhao Z. *Haemophilus parasuis* vaccines. *Vet Immunol Immunopathol.* 2016; 180:53–8. Epub 2016/10/04. <https://doi.org/10.1016/j.vetimm.2016.09.002> PMID: 27692096.
28. United States DoA. Swine 2012 Part I: Baseline Reference of Swine Health and Management in the United States. Fort Collins, CO: 2015.
29. United States DoA. Swine 2012 Part II: Reference of Swine Health and Health Management in the United States. Fort Collins, CO: 2015.
30. White M. Pig health—Glässers disease. <http://www.nadis.org.uk/bulletins/glaessers-disease.aspx>: 2010.
31. El Garch F, Kroemer S, Galland D, Morrissey I, Woehrle F. Survey of susceptibility to marbofloxacin in bacteria isolated from diseased pigs in Europe. *Vet Rec.* 2017; 180(24):591. Epub 2017/03/30. <https://doi.org/10.1136/vr.103954> PMID: 28348142.
32. Dayao D, Gibson JS, Blackall PJ, Turni C. Antimicrobial resistance genes in *Actinobacillus pleuropneumoniae*, *Haemophilus parasuis* and *Pasteurella multocida* isolated from Australian pigs. *Aust Vet J.* 2016; 94(7):227–31. Epub 2016/06/29. <https://doi.org/10.1111/avj.12458> PMID: 27349882.

33. Molerés J, Santos-Lopez A, Lazaro I, Labairu J, Prat C, Ardanuy C, et al. Novel bla<sub>ROB</sub>-1-bearing plasmid conferring resistance to beta-lactams in *Haemophilus parasuis* isolates from healthy weaning pigs. *Appl Environ Microbiol*. 2015; 81(9):3255–67. Epub 2015/03/10. <https://doi.org/10.1128/AEM.03865-14> PMID: 25747001; PubMed Central PMCID: PMC4393459.
34. Brockmeier SL, Register KB, Kuehn JS, Nicholson TL, Loving CL, Bayles DO, et al. Virulence and draft genome sequence overview of multiple strains of the swine pathogen *Haemophilus parasuis*. *PLoS One*. 2014; 9(8):e103787. Epub 2014/08/20. <https://doi.org/10.1371/journal.pone.0103787> PMID: 25137096; PubMed Central PMCID: PMC4138102.
35. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017; 27(5):722–36. Epub 2017/03/17. <https://doi.org/10.1101/gr.215087.116> PMID: 28298431; PubMed Central PMCID: PMC45411767.
36. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014; 9(11):e112963. Epub 2014/11/20. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509; PubMed Central PMCID: PMC4237348.
37. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 30(14):2068–9. Epub 2014/03/20. <https://doi.org/10.1093/bioinformatics/btu153> PMID: 24642063.
38. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008; 9:75. Epub 2008/02/12. <https://doi.org/10.1186/1471-2164-9-75> PMID: 18261238; PubMed Central PMCID: PMC2265698.
39. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. ACT: the Artemis Comparison Tool. *Bioinformatics*. 2005; 21(16):3422–3. Epub 2005/06/25. <https://doi.org/10.1093/bioinformatics/bti553> PMID: 15976072.
40. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 2010; 5(6):e11147. Epub 2010/07/02. <https://doi.org/10.1371/journal.pone.0011147> PMID: 20593022; PubMed Central PMCID: PMC2892488.
41. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999; 27(2):573–80. Epub 1998/12/24. PMID: 9862982; PubMed Central PMCID: PMC148217.
42. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Res*. 2006; 34(Database issue):D32–6. <https://doi.org/10.1093/nar/gkj014> PMID: 16381877; PubMed Central PMCID: PMC1347377.
43. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res*. 2007; 35(Web Server issue):W52–7. Epub 2007/06/01. <https://doi.org/10.1093/nar/gkm360> PMID: 17537822; PubMed Central PMCID: PMC1933234.
44. den Dunnen JT, Antonarakis SE. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. *Hum Mutat*. 2000; 15(1):7–12. Epub 1999/12/29. [https://doi.org/10.1002/\(SICI\)1098-1004\(200001\)15:1<7::AID-HUMU4>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1098-1004(200001)15:1<7::AID-HUMU4>3.0.CO;2-N) PMID: 10612815.
45. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016; 44(D1):D279–85. Epub 2015/12/18. <https://doi.org/10.1093/nar/gkv1344> PMID: 26673716; PubMed Central PMCID: PMC4702930.
46. 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(6):461–5. Epub 2010/05/11. <https://doi.org/10.1038/nmeth.1459> PMID: 20453866; PubMed Central PMCID: PMC2879396.
47. Roberts RJ, Vincze T, Posfai J, Macelis D. REBASE: restriction enzymes and methyltransferases. *Nucleic Acids Res*. 2003; 31(1):418–20. Epub 2003/01/10. PMID: 12520038; PubMed Central PMCID: PMC165516.
48. 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(W1):W16–21. Epub 2016/05/05. <https://doi.org/10.1093/nar/gkw387> PMID: 27141966; PubMed Central PMCID: PMC4987931.
49. Lancashire JF, Terry TD, Blackall PJ, Jennings MP. Plasmid-encoded Tet B tetracycline resistance in *Haemophilus parasuis*. *Antimicrob Agents Chemother*. 2005; 49(5):1927–31. Epub 2005/04/28. <https://doi.org/10.1128/AAC.49.5.1927-1931.2005> PMID: 15855515; PubMed Central PMCID: PMC1087615.
50. Kehrenberg C, Meunier D, Targant H, Cloeckaert A, Schwarz S, Madec JY. Plasmid-mediated florfenicol resistance in *Pasteurella trehalosi*. *J Antimicrob Chemother*. 2006; 58(1):13–7. Epub 2006/05/04. <https://doi.org/10.1093/jac/dkl174> PMID: 16670108.
51. Johnson TJ, Fernandez-Alarcon C, Bojesen AM, Nolan LK, Trampel DW, Seemann T. Complete genome sequence of *Gallibacterium anatis* strain UMN179, isolated from a laying hen with peritonitis. *J*

- Bacteriol. 2011; 193(14):3676–7. Epub 2011/05/24. <https://doi.org/10.1128/JB.05177-11> PMID: 21602325; PubMed Central PMCID: PMC3133324.
52. Yamaguchi Y, Park JH, Inouye M. Toxin-antitoxin systems in bacteria and archaea. *Annu Rev Genet.* 2011; 45:61–79. Epub 2011/11/09. <https://doi.org/10.1146/annurev-genet-110410-132412> PMID: 22060041.
  53. Kedzierska B, Hayes F. Emerging Roles of Toxin-Antitoxin Modules in Bacterial Pathogenesis. *Molecules.* 2016; 21(6). Epub 2016/06/21. <https://doi.org/10.3390/molecules21060790> PMID: 27322231.
  54. Hayes F, Van Melder L. Toxins-antitoxins: diversity, evolution and function. *Crit Rev Biochem Mol Biol.* 2011; 46(5):386–408. Epub 2011/08/09. <https://doi.org/10.3109/10409238.2011.600437> PMID: 21819231.
  55. Wang X, Wood TK. Toxin-antitoxin systems influence biofilm and persister cell formation and the general stress response. *Appl Environ Microbiol.* 2011; 77(16):5577–83. Epub 2011/06/21. <https://doi.org/10.1128/AEM.05068-11> PMID: 21685157; PubMed Central PMCID: PMC3165247.
  56. Yamaguchi Y, Inouye M. Regulation of growth and death in *Escherichia coli* by toxin-antitoxin systems. *Nat Rev Microbiol.* 2011; 9(11):779–90. Epub 2011/09/20. <https://doi.org/10.1038/nrmicro2651> PMID: 21927020.
  57. Aoki SK, Pamma R, Hernday AD, Bickham JE, Braaten BA, Low DA. Contact-dependent inhibition of growth in *Escherichia coli*. *Science.* 2005; 309(5738):1245–8. Epub 2005/08/20. <https://doi.org/10.1126/science.1115109> PMID: 16109881.
  58. Hayes CS, Koskiniemi S, Ruhe ZC, Poole SJ, Low DA. Mechanisms and biological roles of contact-dependent growth inhibition systems. *Cold Spring Harb Perspect Med.* 2014; 4(2). Epub 2014/02/05. <https://doi.org/10.1101/cshperspect.a010025> PMID: 24492845; PubMed Central PMCID: PMC3904093.
  59. Willett JL, Ruhe ZC, Goulding CW, Low DA, Hayes CS. Contact-Dependent Growth Inhibition (CDI) and CdiB/CdiA Two-Partner Secretion Proteins. *J Mol Biol.* 2015; 427(23):3754–65. Epub 2015/09/22. <https://doi.org/10.1016/j.jmb.2015.09.010> PMID: 26388411; PubMed Central PMCID: PMC4658273.
  60. Garcia EC. Contact-dependent interbacterial toxins deliver a message. *Curr Opin Microbiol.* 2017; 42:40–6. Epub 2017/10/28. <https://doi.org/10.1016/j.mib.2017.09.011> PMID: 29078204.
  61. Aoki SK, Diner EJ, de Roodenbeke CT, Burgess BR, Poole SJ, Braaten BA, et al. A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature.* 2010; 468(7322):439–42. Epub 2010/11/19. <https://doi.org/10.1038/nature09490> PMID: 21085179; PubMed Central PMCID: PMC3058911.
  62. Mazar J, Cotter PA. New insight into the molecular mechanisms of two-partner secretion. *Trends Microbiol.* 2007; 15(11):508–15. Epub 2007/11/09. <https://doi.org/10.1016/j.tim.2007.10.005> PMID: 17988872.
  63. Biochemistry Tommassen J. Getting into and through the outer membrane. *Science.* 2007; 317(5840):903–4. Epub 2007/08/19. <https://doi.org/10.1126/science.1146518> PubMed PMID: 17702930.
  64. Aoki SK, Malinverni JC, Jacoby K, Thomas B, Pamma R, Trinh BN, et al. Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Mol Microbiol.* 2008; 70(2):323–40. Epub 2008/09/03. <https://doi.org/10.1111/j.1365-2958.2008.06404.x> PMID: 18761695; PubMed Central PMCID: PMC2579741.
  65. Poole SJ, Diner EJ, Aoki SK, Braaten BA, t’Kint de Roodenbeke C, Low DA, et al. Identification of functional toxin/immunity genes linked to contact-dependent growth inhibition (CDI) and rearrangement hot-spot (Rhs) systems. *PLoS Genet.* 2011; 7(8):e1002217. Epub 2011/08/11. <https://doi.org/10.1371/journal.pgen.1002217> PMID: 21829394; PubMed Central PMCID: PMC3150448.
  66. Clantin B, Delattre AS, Rucktooa P, Saint N, Meli AC, Loch C, et al. Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science.* 2007; 317(5840):957–61. Epub 2007/08/19. <https://doi.org/10.1126/science.1143860> PMID: 17702945.
  67. Ruhe ZC, Low DA, Hayes CS. Bacterial contact-dependent growth inhibition. *Trends Microbiol.* 2013; 21(5):230–7. Epub 2013/03/12. <https://doi.org/10.1016/j.tim.2013.02.003> PMID: 23473845; PubMed Central PMCID: PMC3648609.
  68. Olvera A, Pina S, Perez-Simo M, Oliveira S, Bensaid A. Virulence-associated trimeric autotransporters of *Haemophilus parasuis* are antigenic proteins expressed in vivo. *Vet Res.* 2010; 41(3):26. Epub 2009/12/10. <https://doi.org/10.1051/vetres/2009074> PMID: 19995512; PubMed Central PMCID: PMC2820231.
  69. Olvera A, Pina S, Perez-Simo M, Aragon V, Segales J, Bensaid A. Immunogenicity and protection against *Haemophilus parasuis* infection after vaccination with recombinant virulence associated trimeric autotransporters (VtaA). *Vaccine.* 2011; 29(15):2797–802. Epub 2011/02/16. <https://doi.org/10.1016/j.vaccine.2011.01.105> PMID: 21320547.

70. Olvera A, Martinez-Moliner V, Pina-Pedrero S, Perez-Simo M, Galofre-Mila N, Costa-Hurtado M, et al. Serum cross-reaction among virulence-associated trimeric autotransporters (VtaA) of *Haemophilus parasuis*. *Vet Microbiol.* 2013; 164(3–4):387–91. Epub 2013/03/21. <https://doi.org/10.1016/j.vetmic.2013.02.022> PMID: 23510658.
71. Linke D, Riess T, Autenrieth IB, Lupas A, Kempf VA. Trimeric autotransporter adhesins: variable structure, common function. *Trends Microbiol.* 2006; 14(6):264–70. Epub 2006/05/09. <https://doi.org/10.1016/j.tim.2006.04.005> PMID: 16678419.
72. Roberts RJ, Belfort M, Bestor T, Bhagwat AS, Bickle TA, Bitinaite J, et al. A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 2003; 31(7):1805–12. Epub 2003/03/26. PMID: 12654995; PubMed Central PMCID: PMCPMC152790.
73. Casadesus J, Low D. Epigenetic gene regulation in the bacterial world. *Microbiol Mol Biol Rev.* 2006; 70(3):830–56. Epub 2006/09/09. <https://doi.org/10.1128/MMBR.00016-06> PMID: 16959970; PubMed Central PMCID: PMCPMC1594586.
74. Collier J, McAdams HH, Shapiro L. A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc Natl Acad Sci U S A.* 2007; 104(43):17111–6. Epub 2007/10/19. <https://doi.org/10.1073/pnas.0708112104> PMID: 17942674; PubMed Central PMCID: PMCPMC2040471.
75. Heithoff DM, Sinsheimer RL, Low DA, Mahan MJ. An essential role for DNA adenine methylation in bacterial virulence. *Science.* 1999; 284(5416):967–70. Epub 1999/05/13. PMID: 10320378.
76. Marinus MG, Casadesus J. Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev.* 2009; 33(3):488–503. Epub 2009/01/30. <https://doi.org/10.1111/j.1574-6976.2008.00159.x> PMID: 19175412; PubMed Central PMCID: PMCPMC2941194.
77. Stephens C, Reisenauer A, Wright R, Shapiro L. A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proc Natl Acad Sci U S A.* 1996; 93(3):1210–4. Epub 1996/02/06. PMID: 8577742; PubMed Central PMCID: PMCPMC40058.
78. van der Woude M, Braaten B, Low D. Epigenetic phase variation of the pap operon in *Escherichia coli*. *Trends Microbiol.* 1996; 4(1):5–9. Epub 1996/01/01. PMID: 8824788.
79. Atack JM, Srikhanta YN, Fox KL, Jurcisek JA, Brockman KL, Clark TA, et al. A biphasic epigenetic switch controls immunoevasion, virulence and niche adaptation in non-typeable *Haemophilus influenzae*. *Nat Commun.* 2015; 6:7828. Epub 2015/07/29. <https://doi.org/10.1038/ncomms8828> PMID: 26215614; PubMed Central PMCID: PMCPMC4525171.
80. Atack JM, Tan A, Bakaletz LO, Jennings MP, Seib KL. Phasevarions of Bacterial Pathogens: Methylomics Sheds New Light on Old Enemies. *Trends Microbiol.* 2018; 26(8):715–26. Epub 2018/02/18. <https://doi.org/10.1016/j.tim.2018.01.008> PMID: 29452952; PubMed Central PMCID: PMCPMC6054543.
81. Fox KL, Dowideit SJ, Erwin AL, Srikhanta YN, Smith AL, Jennings MP. *Haemophilus influenzae* phasevarions have evolved from type III DNA restriction systems into epigenetic regulators of gene expression. *Nucleic Acids Res.* 2007; 35(15):5242–52. Epub 2007/08/07. <https://doi.org/10.1093/nar/gkm571> PMID: 17675301; PubMed Central PMCID: PMCPMC1976455.
82. Srikhanta YN, Fox KL, Jennings MP. The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes. *Nat Rev Microbiol.* 2010; 8(3):196–206. Epub 2010/02/09. <https://doi.org/10.1038/nrmicro2283> PMID: 20140025.
83. Srikhanta YN, Maguire TL, Stacey KJ, Grimmond SM, Jennings MP. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci U S A.* 2005; 102(15):5547–51. Epub 2005/04/02. <https://doi.org/10.1073/pnas.0501169102> PMID: 15802471; PubMed Central PMCID: PMCPMC556257.