PLOS ONE

Pan-genome survey of the fish pathogen *Yersinia ruckeri* links accessory- and amplified genes to virulence

Andreas Riborg[1,2]*, Snorre Gulla[1], Eve Zeyl Fiskebeck[1], David Ryder[3], David W. Verner-Jeffreys[3], Duncan J. Colquhoun[1,4], Timothy J. Welch[5]

1 Norwegian Veterinary Institute, Ås, Norway, **2** Vaxxinova Norway AS, Bergen, Norway, **3** Centre for Environment, Fisheries and Aquaculture Science (CEFAS), Weymouth, Dorset, United Kingdom, **4** University of Bergen, Bergen, Norway, **5** National Centre for Cool and Coldwater Aquaculture, USDA-ARS, Leetown, WV, United States of America

* andreasriborg@gmail.com

Abstract

While both virulent and putatively avirulent *Yersinia ruckeri* strains exist in aquaculture environments, the relationship between the distribution of virulence-associated factors and *de facto* pathogenicity in fish remains poorly understood. Pan-genome analysis of 18 complete genomes, representing established virulent and putatively avirulent lineages of *Y. ruckeri*, revealed the presence of a number of accessory genetic determinants. Further investigation of 68 draft genome assemblies revealed that the distribution of certain putative virulence factors correlated well with virulence and host-specificity. The inverse-autotransporter invasin locus *yrIlm* was, however, the only gene present in all virulent strains, while absent in lineages regarded as avirulent. Strains known to be associated with significant mortalities in salmonid aquaculture display a combination of serotype O1-LPS and *yrIlm*, with the well-documented highly virulent lineages, represented by MLVA clonal complexes 1 and 2, displaying duplication of the *yrIlm* locus. Duplication of the *yrIlm* locus was further found to have evolved over time in clonal complex 1, where some modern, highly virulent isolates display up to three copies.

Introduction

Yersinia ruckeri is the causative agent of enteric redmouth disease (ERM), also known as yersiniosis in salmonids and other cold-water species of fish. While the disease has been primarily associated with farmed rainbow trout internationally, it also affects farmed Atlantic salmon in Norway, Finland, Australia, Chile and in the UK [1–5]. It has further been isolated from a wide range of other fish species, both farmed [6–8] and wild [9–11]. In Norway, *Y. ruckeri* strains appear to fall into at least three categories regarding their ability to cause disease in farmed salmon. All serious yersiniosis outbreaks since the late 1980s have been caused by a single genetic lineage of highly virulent serotype O1, known as multiple-locus variable number of tandem repeats analysis (MLVA) clonal complex 1 (CC1) [12]. Another MLVA clonal complex

(CC3) belonging to serotype O2 is considered only mildly pathogenic, being associated primarily with sporadic detections or low-mortality outbreaks [12]. A genetically diverse third category of isolates, predominantly sharing the same serotype (O1) as virulent CC1, are also found in Norwegian aquaculture environments and healthy fish. Despite commonly being present, these heterogeneous non-CC1, non-CC3 isolates are presumed avirulent as they have never been associated with disease [13].

Differences in *Y. ruckeri* virulence are also recognised internationally, with serotype O1 generally associated with high virulence and serotype O2 isolates considered to be of low or moderate virulence [10, 14, 15]. Serotype O1 (CC5) isolates dominate amongst disease cases in Atlantic salmon in Australia, while outbreaks of yersinosis in rainbow trout in the USA, UK and mainland Europe are almost exclusively caused by the serotype O1 CC2 lineage [7, 12]. Reports of avirulent isolates from Australia [7] and the UK [16] may indicate an international situation similar to that in Norway, with disease outbreaks dominated by highly virulent clones against a background of essentially avirulent, environmental strains.

Differences in virulence between *Y. ruckeri* lineages should be reflected in their genomes. Identification of key genetic differences should thus broaden our understanding of this widespread and economically important aquaculture pathogen. A considerable degree of host-specificity is apparent amongst various *Y. ruckeri* lineages [12], but cross-species virulence has been demonstrated, albeit at reduced efficiency [17], suggesting the presence of common virulence mechanisms. To gain a better understanding of the genetic elements necessary for the virulent phenotype observed in certain strains, a diverse collection of *Y. ruckeri* strains was whole-genome sequenced and analysed together with publicly available genomes. As mobile elements and putative virulence factors may contain highly repetitive sequences, a hybrid Illumina and MinION nanopore long-read sequencing approach was utilized for selected representative strains to obtain complete circular and highly accurate genome assemblies.

Results & discussion

Core-gene phylogenetic reconstruction

The maximum-likelihood phylogeny (Fig 1) inferred from 2388 core genes shared across the 86 *Y. ruckeri* genomes investigated verifies the existence of multiple discrete lineages separated according to host species and/or geography [7], and accurately reflects the MLVA-based population structure described previously [12]. Disregarding the distantly related YRB lineage, the tree bifurcates into two lineages (A and B), one of which (lineage A) contains all of the well-documented virulent isolates. This includes CC2 found globally in rainbow trout [12], CC5 from Australian salmon [7] as well as those of non-O1 serotypes such as strains BigCreek74 and SC09. Lineage A also contains all disease-associated Norwegian isolates, belonging respectively to the currently dominating CC1 and to CC10 from the late 1980s [1, 18], as well as the mildly virulent CC3 (serotype O2). Two notable exceptions to the association between virulence and affiliation to lineage A are CC7 and CC9, which in Norway are associated with biofilms in salmon rearing facilities and egg-fluid of apparently healthy salmon. Challenge trials with NVI-11076 (CC7) have demonstrated its avirulent nature in Atlantic salmon (manuscript in preparation). Isolates belonging to CC9 from Australia and New Zealand have been described as of low virulence, with New Zealand isolates being associated with screening of healthy fish or sporadic minor losses in Chinook salmon for which vaccination was not considered necessary [7]. The relatively close phylogenetic relationship between the well-documented virulent sub-lineages CC2 and CC5, and the putatively avirulent CC7 and CC9 (Fig 1), could facilitate identification of virulence-discriminatory genetic determinants in this part of the tree.
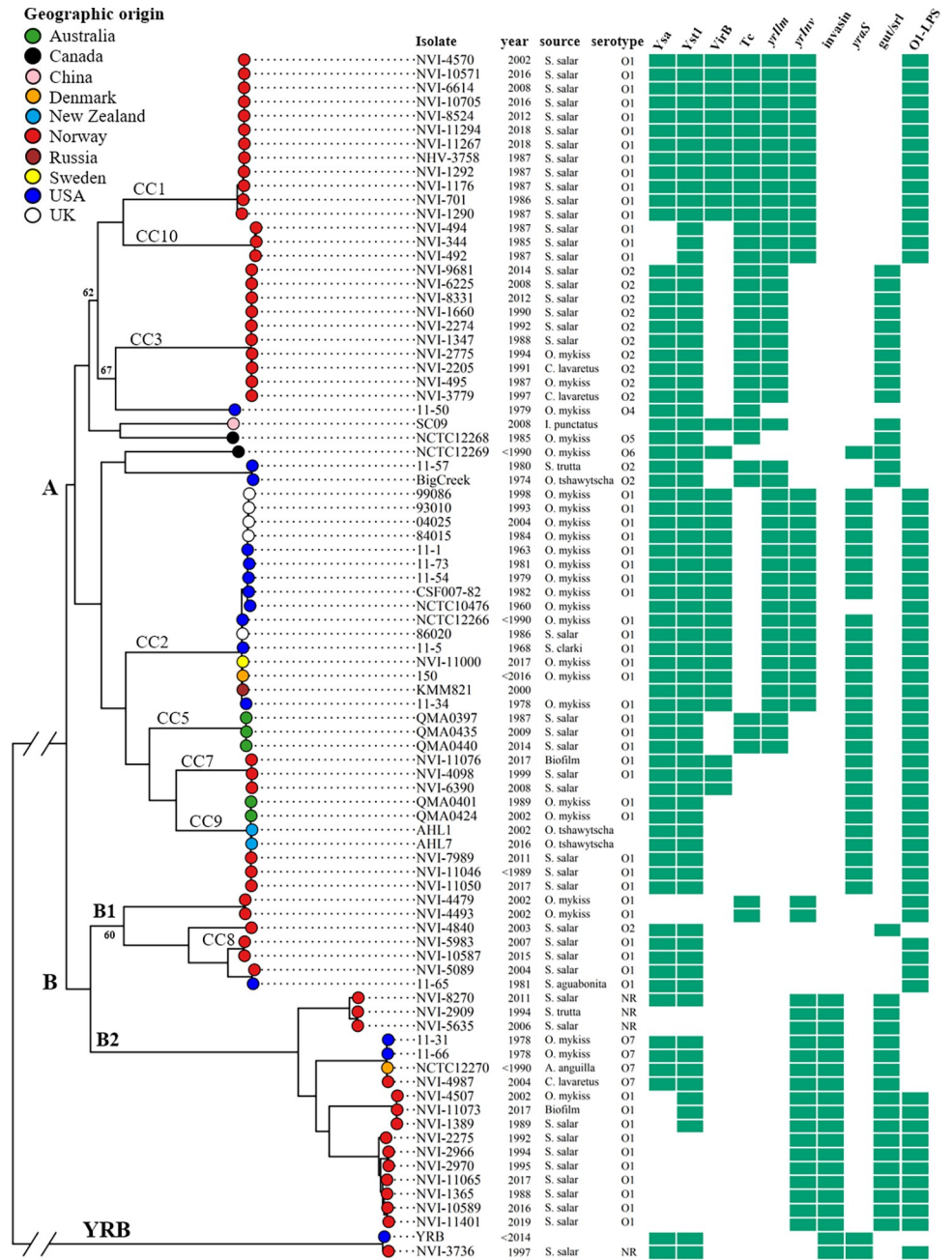
Fig 1. Maximum likelihood tree based on 2,387 genes, 49,389 core gene SNPs (16,762 SNPs were identified with the YRB lineage excluded; S2 Fig), with presence/absence data for specific systems or genes. MLVA clonal complexes (CC) and lineages are indicated on branches. Bootstrap values as percentages of 200 replicates are indicated if < 75 percent. Branch lengths between the YRB lineage and remaining genomes are truncated (see S1 Fig). Presence of genetic determinants indicated by green shading for Ysa Type 3 secretion system, Yst1 Type 2 secretion system, virB/virD4 Type 4 secretion system (VirB), Toxin complex (Tc), the putative invasins *yrIlm*, *yrInv* and undescribed invasin WP_042527435, sorbitol utilization genes (*gut/srl*), the alkyl sulphatase associated with SDS degradation (*yarS*), and the serotype O1-LPS synthesis cluster. Partial deletions of Ysa, Yst1 and the O1-LPS cluster are shown as absent if genes regarded as critical for function are not present. Serotype NR indicates no reaction with available antisera (O1, O2, O5).

https://doi.org/10.1371/journal.pone.0285257.g001

Juxtaposed to lineage A lies lineage B (further subdivided into B1 and B2), which contains isolates mostly originating from biofilm, health screening of clinically healthy farmed fish (CC8) and wild fish. In Norway, isolates from these sub-lineages are considered avirulent, as they have never been associated with disease outbreaks despite being widely present in aquaculture environments. While various serotypes are found within lineage B, most belong to serotype O1.

The strain YRB forms a distinct distal lineage together with a single isolate (NVI-3736) from wild salmon in Norway. While relatively distantly related to the remaining isolates studied (S1 Fig), both genomes display the same distinguishing genetic characteristics of the species in terms of genome size and gene content. The pangenome is considerably affected by including these two genomes, increasing by 507 total genes while 129 genes are lost from the core (**S2 Fig**), but status as core or accessory for the virulence-related genetic determinants discussed later was not affected.

Pan-genome analysis

Based on the phylogenetic reconstruction described above, representative isolates of putatively virulent and avirulent sub-lineages/CCs were selected for Nanopore sequencing to produce complete hybrid (Nanopore/Illumina) genome assemblies. Publicly available complete genome sequences (QMA0440, BigCreek74, SC09, CFS007, KMM821 and YRB) were also included (Table 1). As a large number of diverse plasmids was identified among these assemblies, most of which were exclusive to a single genome, plasmids were excluded from the pangenome analysis and investigated separately. With plasmids excluded, most of the genetic diversity amongst the complete assemblies could be related to a limited number of major deletions and a large number of diverse mobile elements. Accessory mobile elements ranging from single insertion sequence (IS) elements and small transposons (Tn), to large integrative and conjugative elements (ICE), as well as prophages and various cryptic phage-like elements, were identified. A large portion of the accessory genome resides within a few specific regions of the genome (Fig 2).

A deletion encompassing 20 genes within the O-antigen biosynthesis cluster in *Y. ruckeri* has been previously shown responsible for the O2 serotype [7]. Our dataset indicates that this deletion is a common feature of all non-O1 serotypes, including isolates that are unreactive to antisera from defined serotypes, implying that all serotypes derive from a common serotype O1 ancestor. BLAST searches based on ORFs from the LPS O1-antigen cluster revealed this phenomenon in all 86 genome assemblies studied here, with the exception of NVI-3736 in the YRB lineage. The O1-antigen cluster in this isolate appears to be intact but does not result in a visible reaction with O1-specific antisera. With the sole exception of NVI-3736, these finding are in agreement with a previous study that reported PCR targeting this region as being positive exclusively for serotype O1 strains [23]. The serotype O1 LPS-antigen cluster contains genes essential for synthesis of nonulosonic acid, a sialic acid component of the O-polysaccharide repeat of O1 *Y. ruckeri* [23]. Sialic acid-containing LPS is associated with host mimicry and evasion of the innate immune response in vertebrate hosts [24], and its presence may thus explain the dominance of serotype O1 amongst highly virulent strains. Additionally, serotype O1 LPS has been shown to be critically important for resistance to serum killing in rainbow trout [25] and this specific O1 antigen structure may possibly confer serum resistance.

The Yst1 type II (T2SS) and Ysa type III (T3SS) secretion systems are situated between core genes *mutS* and *fumA* [8]. This locus may contain accessory transposons, and is partially or completely lost in several sub-lineages (Fig 3). With the exception of CC10, all serotype O1 sequences have a putative biofilm adhesin polymer cluster associated with this locus, similar to

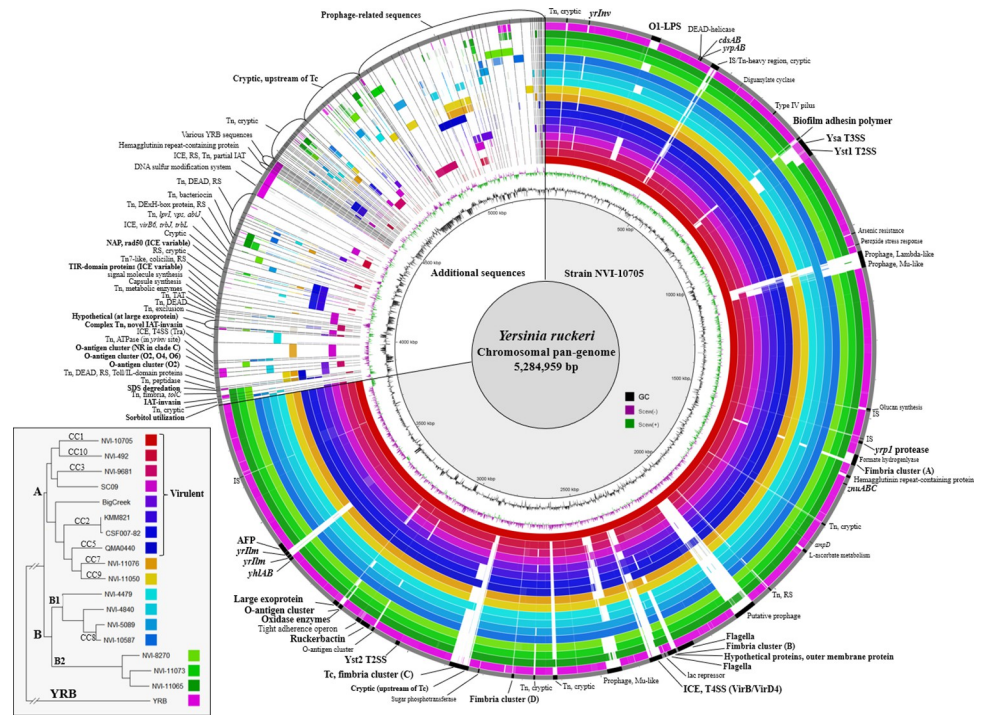**Table 1. Complete genome assemblies and plasmids used for pan-genome analyses.**

Strain	Lineage		Chromosome		Plasmids			Reference
	Main	Sub- / CC	accession	Size	Name	accession	Size	
NVI-10705	A	1	CP099805	3,809,437	pYR4	CP099806	115,589	Current
					pYR5	CP099807	5,191	
NVI-492	A	10	CP099813	3,654,750	pYR5	CP099814	5,286	Current
NVI-9681	A	3	CP099811	3,764,054	pYR7	CP099812	121,362	Current
SC09	A	s	CP025800	3,799,871	pLT	CP025802	57,905	[8]
					pWKY	CP025801	73,051	
BigCreek74	A	s	CP011078	3,699,725	-			Unpublished
KMM821	A	2	CP071802	3,773,395	pYR2	CP071803	16,925	Unpublished
					pYR3	CP071804	103,906	
CFS007-82	A	2	LN681231	3,799,036	pYR2	LN681229	16,923	[19]
					pYR3	LN681230	103,917	
QMA0440	A	5	CP017236	3,856,634	-			[7]
NVI-11076	A	7	CP099808	3,763,098	-			Current
NVI-11050	A	9	CP099815	3,853,871	-			Current
NVI-4479	B1	s	CP098710	3,660,364	-			Current
NVI-4840	B1	s	CP098703	3,657,071	pYR8	CP098705	6,876	Current
					pYR9	CP098704	78,774	
NVI-5089	B1	s	CP098701	3,724,773	pYR10	CP098702	89,273	Current
NVI-10587	B1	8	CP099809	3,700,651	pYR11	CP099810	4,985	Current
NVI-8270	B2	s	CP098694	3,655,667	-			Current
NVI-11073	B2	s	CP098722	3,585,796	-			Current
NVI-11065	B2	s	CP098723	3,703,894	-			Current
YRB	YRB	s	CP009539	3,605,216	-			[20]
NVI-11294†	A	1	CP098716	3,808,409	pYR4	CP098717	115,590	Current
					pYR5	CP098718	5,191	
NVI-11267†	A	1	CP098719	3,808,455	pYR4	CP098720	115,590	Current
					pYR5	CP098721	5,191	
NVI-10571†	A	1	CP098724	3,783,355	pYR4	CP098725	76,483	Current
					pYR5	CP098726	5,191	
NVI-8524†	A	1	CP098691	3,808,529	pYR4	CP098692	115,590	Current
					pYR5	CP098693	5,191	
NVI-6614†	A	1	CP098697	3,783,289	pYR4	CP098699	80,438	Current
					pYR5	CP098700	5,191	
					pYR6	CP098698	83,000	
NVI-4570†	A	1	CP098706	3,773,829	pYR4	CP098707	80,847	Current
					pYR5	CP098709	5,191	
					pYR6	CP098708	71,441	
NVI-1292†	A	1	CP098711	3,765,691	pYR4	CP098712	80,847	Current
					pYR5	CP098713	5,191	
NVH_3758†	A	1	CP023184	3,766,700	pYR4	CP032236	80,843	[21, 22]
NVI-1176†	A	1	CP098714	3,829,170	pYR4	CP098715	80,847	Current
NVI-701†	A	1	CP098695	3,758,144	pYR4	CP098696	80,847	Current

s: no MLVA CC defined

†: Not included in Fig 2. These CC1 assemblies vary only in complete- or partial prophage sequences (all of which are represented in Fig 2), plasmids, and *yrIlm-* copy number (Fig 5).

The significantly larger pYR4 variant (115,590bp) is due to a Mu-like prophage. The reduced size of pYR4 in NVI-10571 is due to loss of Tn-related sequences. KMM821 plasmids were identified as pYR2 and pYR3 by BLAST. pYR11 in NVI-10587 is nearly identical to *Aeromonas veronii* plasmid pWP3-W19-ESBL-03_3 (NZ_AP022041).

https://doi.org/10.1371/journal.pone.0285257.t001

Fig 2. Maximum likelihood core-gene phylogenetic tree (bottom left) and circular representation of genomic comparisons generated with the BLAST Ring Image Generator (BRIG) from complete *Y. ruckeri* genome assemblies. Plasmid sequences were removed from the data prior to analysis. Lineages and MLVA clonal complexes (CC) are indicated on the phylogenetic tree tree branches, with strains regarded as virulent labelled as such. The branch length of YRB is truncated (see S1 Fig). Coloured rings represent genomic sequences used as BLAST queries, corresponding to the colour labelling on tree nodes. Ordering top-bottom on the tree and inner-outer on the circular representation is identical (virulent strains in innermost rings). Solid colour in each respective ring represents ≥95% nucleotide identity to the pan-genome reference. Cut-off identity for BLAST was 85%. The innermost ring (black line with nucleotide distances indicated) represents the pan-genome reference sequence, consisting of the chromosome of strain NVI-10705 for the first 3,818,566 bp, followed by additional *Y. ruckeri* chromosomal sequences with each segment separated by a 800 bp artificial gap with black fill, and ordered by commonality or general association. Prominent features within the NVI-10705 chromosome are labelled by black fill on the outermost ring (grey). Annular graphs towards the centre represent GC percent (black) and GC skew (purple and green). All chromosomal sequences from additional complete genomes belonging to CC1 (see Table 1) are represented, and vary from NVI-10705 only in prophage-related sequences (not shown) and plasmids (Table 1). Hemagglutinin repeat-containing protein is present in all sequences (n = 86) but divergent in YRB (82% nucleotide identity). Abbreviations used: Restriction system (RS), toxin-antitoxin (TAT), insertion sequence (IS), transposon (Tn), integrative and conjugative elements (ICE), Toxin Complex (Tc), inverse autotransporter (IAT), no reaction (NR), DEAD-box helicase (DEAD), Anti-feeding prophage (AFP).

pgaABCD in *E. coli* which synthesizes and exports biofilm-enhancing linear homopolymers [26]. Several non-O1 sequences alternatively display a putative O-antigen modification cluster at this site. BLAST searches against the complete dataset of all 86 assemblies link this O-antigen cluster with serotype O2, suggesting that gain of this O-antigen cluster along with loss of the O1-antigen cluster may be required for the O2 serotype. Another putative O-antigen cluster of similar size and genetic composition is located some 300 Kbp downstream (between *ushA* and *copA*) and also linked to the O2 serotype, but occurs also in serotypes O4 and O6. The pattern of distribution of these accessory features between serotype O2 strains and the distantly related strain YRB, is a strong indication of horizontal gene transfer involving fairly large genomic segments. Other putative O-antigen modification clusters are notably also present at various locations within the accessory genome and may be involved in producing other serotypes and/ or subtypes of serotype O1, as described in some serotyping schemes [27].
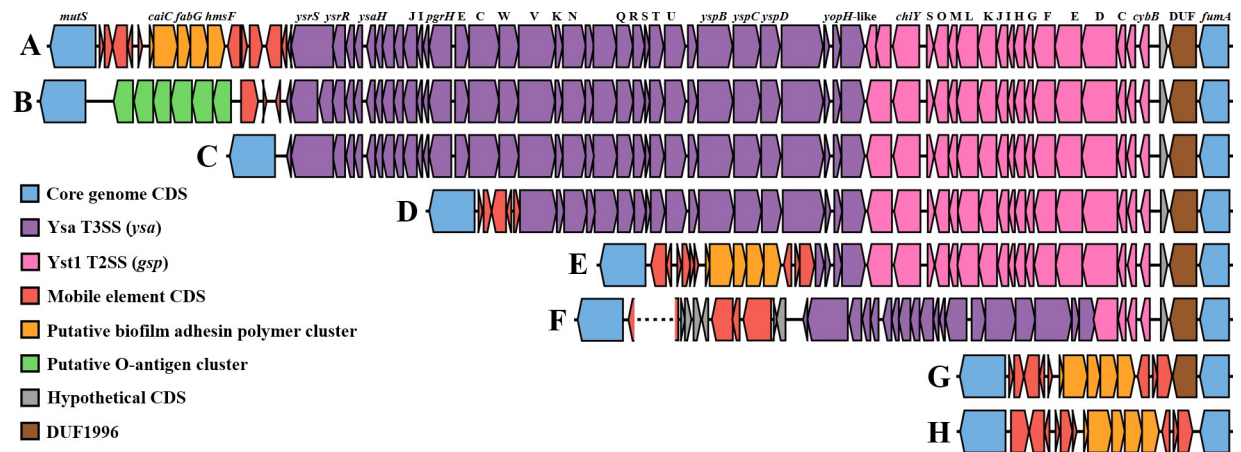
Fig 3. Genetic organization of the Ysa-Yts1 locus, between *mutS* and *fumA* in the *Y. ruckeri* core genome. Each gene is colored according to association with a system or other specific role as indicated by the panel on the left. Nomenclature for Ysa (*ysa*) and Yst1 (*gsp*) is according to Venecia and Young [64] and Iwobi et al. [63]. The putative O-antigen cluster consists of protein sequences (green) WP_045844464–67 and WP_042527525. Genotypes, labelled A to H, represent the following sequences: (A) NVI-10705 and typical configuration of serotype O1 sequences that possess intact Ysa and Yst1; (B) NVI-9681 and typical configuration of serotype O2 with a O-antigen cluster in green; (C) SC09; (D) NVI-492 with partial Ysa deletion characteristic of CC10; (E) NVI-11073 with Ysa deletion; (F) NVI-5635 with partial Ysa and Yst1 deletion and a putative ICE inserted between *mutS* and the partial Ysa; (G) NVI-4479 with complete deletion of Ysa and Yst1; (H) NVI-11065 with complete deletion of Ysa and Yst1 including DUF 1996. The dotted line indicates a contig split (incomplete genome sequence).

https://doi.org/10.1371/journal.pone.0285257.g003

Another highly variable region may contain toxin complex (Tc) genes and a fimbria cluster, and in most strains various cryptic and prophage-related elements. These cryptic genes are similar in putative function but highly variable in sequence between strains and comprise ~10% of the accessory genome. Yet another highly variable region is found upstream of a putative large exoprotein (WP_045844312), which resembles a contact-dependent growth inhibition operon, and contains accessory O-antigen modification genes and oxidase enzymes in CC1 and some other serotype O1 strains. This region may also contain transposons carrying a putative peptidase or hypothetical genes.

DEAD-box helicases and restriction systems are common Tn payloads in *Y. ruckeri* [7]. Most strains have acquired several accessory restriction systems, and at least one DEAD-box helicase in addition to that present in the core genome. Restriction systems are also common within prophages and cryptic phage-like elements. While prophages make up a substantial portion of the accessory genome, no known virulence-related prophages were identified, nor does the presence of any prophage seem to align with the virulence trait. The T6SS-like anti-feeding prophage (AFP), which has been previously related to virulence [28], is present in all assemblies.

The presence/absence of genes involved in sorbitol-utilization (Sorbitol-specific membrane transport and utilization, *gut*/*srl*; [7]) and sodium dodecyl sulfate (SDS) degradation (*yraS* alkyl sulphatase, WP_080717331) varies between lineages and correlates with the phenotypes of CC1 (negative for both traits), CC2 (degrades SDS) and CC3 (utilizes sorbitol). Sorbitol utilization has previously been considered a trait specific to serotype O2 [29, 30], although rare instances of sorbitol-fermenting serotype O1 strains have been reported [2, 31]. In our dataset, the sorbitol utilization genes are present in most isolates that display deletion of the LPS O1-antigen cluster, i.e. non-O1 serotypes, but they are also present throughout lineage B2 which include several serotype O1 isolates (Fig 1). The ability to degrade SDS has been previously associated with virulence, as a 'heat-sensitive factor' [32, 33], but this has since been refuted [34]. Here, we find the presence of the alkyl sulphatase involved in primary

degradation of SDS to correlate well with core-gene based phylogeny, but not with virulence or serotype. Disruption of *lacZ* was identified in isolates NVI-4840 (lineage B1) and NVI-8270 (lineage B2) due to insertion of, respectively, a transposase and transposase-related sequence. These isolates were found to be *ortho*-nitrophenyl-β-galactoside (ONPG) negative (API20E, not shown). ONPG negative *Y. ruckeri* strains have, to our knowledge, only been reported once previously [35]. The scarcity of reports of ONPG-negativity, and the sorbitol-fermenting serotype O1 phenotype, suggests that strains belonging to lineage B have been poorly represented in previous studies. Genes absent in the YRB lineage include the lac repressor, the formate hydrogenlyase cluster, and genes involved in L-ascorbate metabolism, although impact on phenotype is not known.

Of the wide variety of pilin/fimbriae gene clusters found both chromosomally- and in plasmids, many of which probably overlap in function, four were considered to be of particular interest. One (A) is present in 8 of the 16 complete genome sequences, but its presence does not correlate well with phylogeny or virulence. Another (B), located within the core-genome flagellar region, was reported by Barnes et al. [7] to be exclusively missing in 'serotype O1 strains from the USA' i.e. CC2. Here, we found these genes also absent in CC7, and lineages B and YRB. Some adjacent hypothetical proteins, including a predicted outer membrane LPS/ capsule polymerase, have nearly identical distributions, but are also present in CC2. The third (C) fimbria cluster is associated with the Tc cluster, while the fourth (D) is present in all complete sequences, albeit with a duplicated pilin subunit gene in CC1.

A large putative ICE containing a conjugative plasmid-like (Tra) type IV secretion system (T4SS) is present in SC09 and in the putatively avirulent NVI-11076 (CC7). Another ICE, although with a virB/virD4-type T4SS, was found to be present in CC1 and CC2, as well as in strains SC09 and NVI-11076. This virB/virD4 ICE has been implicated in virulence previously [36].

Four accessory inverse-autotransporter invasin-like genes (IAT-invasins) were identified, some of which have been previously linked to *Y. ruckeri* virulence [21, 37]. Of these four, individual strains carry up to two different IAT-invasins, situated in transposons or (apparently) seamlessly integrated into the genome at various genomic locations.

Accessory virulence-associated determinants

In addition to accessory features identified during pan-genome analyses involving complete assemblies, the full dataset of 86 genomes was screened for the presence of several specific virulence factors described previously (Table 2). Many of these were found to be ubiquitous, also within the YRB outgroup, and are thus unlikely responsible for the observed variation in virulence between strains.

Plasmids. Human-pathogenic *Yersinia* spp. are well known for plasmid-borne virulence [48], and the possibility of plasmid-mediated virulence in *Y. ruckeri* has been frequently discussed [22, 49–53]. A wide variety of plasmids were identified amongst the complete genome sequences studied here, ranging from small ~5 Kbp cryptic mobilizable plasmids, to large conjugative plasmids of up to ~120 Kbp carrying complete conjugative T4SS (Tra). While plasmids were identified in most sub-lineages/CCs considered virulent, they seem less common in the putatively avirulent ones, with several isolates displaying complete absence (Table 1). Additional plasmids not shown here notably exist amongst the non-complete genome assemblies. BLAST searches in all 86 genomes for plasmids identified within complete assemblies (BLASTn, replication protein) revealed them to be strictly sub-lineage/CC-specific with two exceptions, i.e. the small mobilizable cryptic plasmid pYR5 present in both CC1 and CC10, and the large conjugative plasmids pYR3 and pYR4, present in CC2 and CC1 respectively. Plasmids pYR3 and pYR4 may be considered as variants of the same plasmid as they share the

Table 2. Distribution of verified and putative virulence factors previously described or discovered here, across the 86 genome assemblies (BLASTn).

Name	Classification	Reference	Distribution
yrp1	Protease	[38]	Core
yrpAB	Peptidase	[34]	Core
yhlBA	Hemolysin	[39]	Core
Yst2	Type II secretion system	[8]	Core
Anti-feeding prophage	Type VI-like secretion system	[28]	Core
Ruckerbactin	Siderophore iron acquisition	[40]	Core
cdsAB	L-cysteine acquisition	[41]	Core
znuABC	Zinc acquisition	[42]	Core
barA-uvrY	Response regulator	[43]	Core
ompF	Outer membrane protein	[44]	Core
filamentous hemagglutinin	Outer membrane protein	[45]	Core
Yst1	Type II secretion system	[8]	Shell
Ysa	Type III secretion system	[46]	Shell
yrIlm	IAT-invasin	[21]	Shell
yrInv	IAT-invasin	[21]	Shell
Unnamed invasin	IAT-invasin	this paper	Shell
yraS	HSF/SDS	[32]	Shell
Biofilm polymer cluster	Putative secreted polymer	this paper	Shell
Tc	Toxin complex	[7]	Shell
VirB/VirD4	Type IV secretion system	[36]	Shell
Novel invasin	IAT-invasin	this paper	Cloud
STIR, *tcpA*	Putative secreted proteins	[36]	Cloud

Core indicates presence in 100%, shell in 15–95%, and cloud in less than 15% of genome assemblies (n = 86). Some features relevant for virulence for which no particular gene or system have been attributed, e.g. phospholipase activity and various pilin/fimbriae clusters, were not screened. Plasmids are not shown. The biotype 2 trait which has a casual association with virulence is caused by non-synonymous SNPs or small indels [47] and thus outside the scope of this study. Rare deletions in Yst2, the tight adherence operon and iron acquisition have been described previously [7] but were not observed amongst the sequences studied here.

same replicative and conjugative plasmid backbone [22]. A large pilin cluster and the conjugative T4SS present on pYR3 and pYR4 have been previously implicated in virulence [22, 53], which is further supported here by their unique presence in the highly virulent CC1 and CC2. The omnipresence of pYR3/pYR4 across the highly virulent and successful CC1 and CC2 does indicate some competitive advantage for this plasmid relating to virulence or survival in aquaculture environments. However, the variety of plasmids identified amongst other virulent strains, and complete absence of plasmids in the virulent CC5, strongly indicates that no specific plasmid or the presence of plasmids in general are essential for virulence in salmonid fish.

The toxin complex locus. Insecticidal Toxin Complexes, or simply Toxin Complexes (Tc), are high-molecular weight secreted protein complexes known to provide insecticidal properties [54], with activity against mammalian cells demonstrated for the Tc gene products of *Y. enterocolitica* and *Y. pestis* [55, 56]. Although the presence of Tc genes in *Y. ruckeri* has been mentioned previously [7, 57], no study has investigated them specifically. We found the presence of the Tc gene cluster, consisting of A, B and C toxin components TcA (WP_162486770), TcdB (WP_234049470) and RhsA (WP_096823580), to align well with pathogenicity towards Atlantic salmon (CC1, CC3 and CC10 in Norway, and CC5 in Australia). These genes are, however, absent in the rainbow trout-associated CC2, which notably constitutes the most studied *Y. ruckeri* sub-lineage to date, and this may be the reason for *Y. ruckeri* not being included in studies of *Yersinia* Tc genes [58, 59]. The Tc cluster has an

identical distribution as fimbria cluster C located immediately upstream. A role in virulence or host-specificity for the Tc genes and/or this associated fimbria cluster seems likely given the distribution of these genes. They cannot be crucial components for *Y. ruckeri* virulence in general, however, given their absence in the highly virulent and globally prominent CC2.

The Yts1-Ysa locus. The Yts1 T2SS and Ysa T3SS of *Y. ruckeri* correspond to the Yts1 and Ysa of *Y. enterocolitica* (BLAST), located within the 199 kb 'plasticity zone' of enteropathogenic biotype 1B strains [60]. While some evidence suggests a role in survival within host macrophages [61], the exact role of these secretion systems during infection is not well understood. Both systems are nonetheless required for full virulence of *Y. enterocolitica* in mammalian hosts following per-oral administration [62–64]. The Yst1-Ysa1 locus is present throughout the deepest branching *Y. ruckeri* lineages, indicating its presence in the founding lineage, although it appears to be lost in several sub-lineages within lineage B which exclusively contain putatively avirulent strains. The phylogenetic placement of strains that have suffered deletions in the Yst1-Ysa1 locus suggests several independent deletion events, as does the variation in genes deleted (Fig 3). While the Ysa T3SS is largely conserved in virulent strains, all investigated isolates of the virulent CC10 display a large deletion that includes the phosphorelay system, operator region, and a range of essential injectosome proteins, indicating that Ysa is not essential for virulence in Atlantic salmon. The loss of the Yts1 T2SS is, on the other hand, seemingly exclusive to certain putatively avirulent sub-lineages/CCs and may thus be required for virulence. However, the broad presence of both intact Ysa and Yts1 in several other putatively avirulent strains does not implicate these systems in playing a major standalone role in enabling virulence.

Chromosomal type IV secretion system. T4SSs are capable of transmitting DNA and proteins to neighboring cells through a conjugative pilus and are essential components of conjugative plasmids and ICE, although specialized variants adapted for natural competence or virulence exist, and T4SS are key virulence factors in a number of bacterial species [65–67].

The virB/virD4-type T4SS ICE present in CC1, CC2 and strains SC09 and NVI-11076, have been linked to virulence in *Y. ruckeri* strain SC09 by a recent series of publications. These studies focused on a set of effector genes located immediately downstream of the core ICE components, the products of which are presumably secreted into host cells by the ICE T4SS or a different secretion system during infection [36, 68]. The downstream genes mobilized by the ICE are however quite variable between strains [36], and in other *Y. ruckeri* strains/lineages generally have functions associated with DNA-repair and recombination (Fig 4). While the effector genes are present exclusively in strain SC09 (STIR, tcpA in Table 2), the presence of this virB/virD4-type T4SS ICE in the highly virulent CC1 and CC2, while uncommon in *Y. ruckeri* otherwise, may indicate some role in virulence. Besides secretion of effectors, some specialized T4SS contribute to virulence by the T4SS pilus conferring attachment to host cells [69], and even non-pilus T4SS surface structures may grant cell-to-cell adhesive capabilities [70]. As such, while this locus cannot be essential for virulence in salmonids, a potential role in *Y. ruckeri* virulence should not be ruled out.

Inverse autotransporter invasins. The inverse autotransporter (IAT) protein family, also known as type Ve-secreted proteins, are characterized by an N-terminal β-barrel domain embedded in the outer membrane of Gram-negative bacteria. This structure functions both as an anchor and facilitator for secretion of the C-terminal domain of the same protein, which becomes exposed on the surface of the cell [71]. These proteins are often referred to as adhesins or invasins as they have been shown to facilitate adhesion to and invasion of host cells or tissues [71–73]. Two IAT-invasins in *Y. ruckeri* have been studied in detail recently in a series of papers that describe *yrIlm* and *yrInv* and their contribution to biofilm formation and virulence [21, 37]. In our dataset, we identified four IAT-invasins, all consisting of an N-terminal
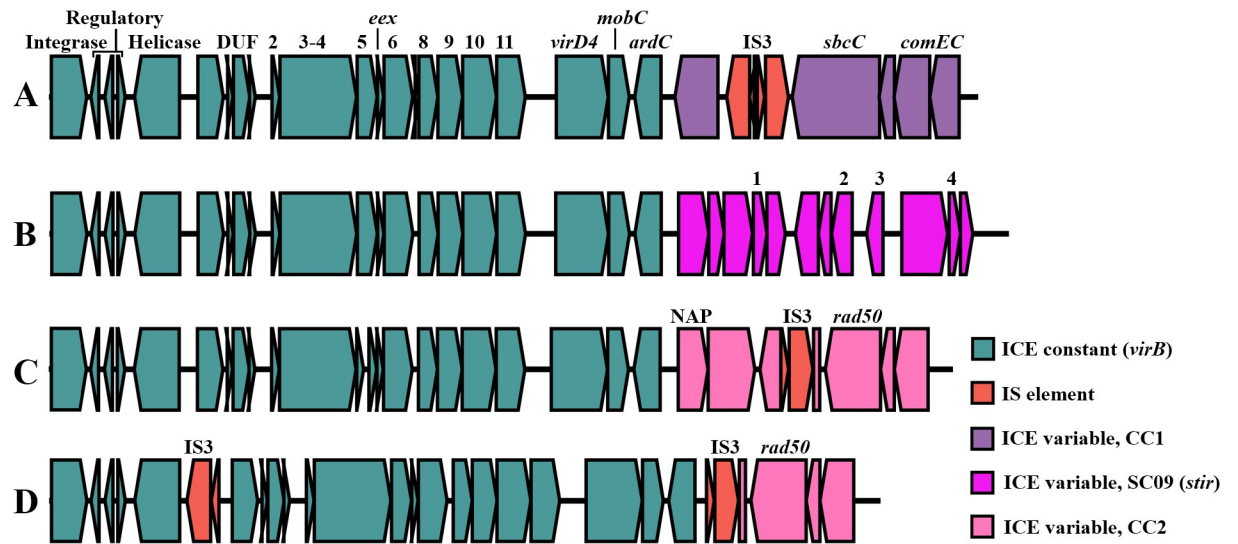
Fig 4. Genetic organization of the accessory VirB/VirD4 ICE and associated genes. Genes are coloured according to function or association with each variant. Genes indicated by number represent *virB2—virB11. virB3* and *virB4* are fused. Genotypes A to D represent the following sequences: (A) NVI-10705 and typical configuration in CC1 isolates with downstream genes exclusive to CC1; (B) SC09 with downstream genes exclusive to this strain; (C) CFS007-82 with typical configuration of CC2 isolates; (D) NVI-11076 (CC7) with some of the downstream genes present in CC2.

https://doi.org/10.1371/journal.pone.0285257.g004

inverse-autotransporter domain, an array of bacterial immunoglobulin-like domains (BIG), and a C-terminal lectin-like domain.

yrIlm contains up to 22 nearly identical BIG-domains, is situated between the virulence factors *yhlA* hemolysin and AFP [21], within a composite transposon flanked by IS256 transposase direct repeats (Fig 5A). Strikingly, the distribution of *yrIlm* seems to correlate fully with virulence, being present in all sub-lineages well documented as causing yersiniosis in fish, while being absent from all those considered avirulent (Fig 1). Its presence in the virulent Australian CC5 (Atlantic salmon), and absence in CC7 and CC9, may explain the observed difference in virulence between these relatively closely related sub-lineages.

The paralogous and notably shorter *yrInv* is also present in many virulent strains, although occurring throughout lineage B2 while absent in the virulent CC5, and can therefore not be considered essential for virulence. Furthermore, an IAT-invasin (WP_042527435) gene is widespread amongst *Yersinia* species at >60% AA identity, identical protein domain-organization and chromosomal location (between *aroP* and *ampE*), to our knowledge previously undescribed, is present in *Y. ruckeri* lineages B2 and YRB and thus appears to be negatively correlated with pathogenicity in salmonids. Finally, a novel IAT-invasin (locus tag ND012_09315 in CP098710) represents a paralogue of *yrInv* and *yrIlm*, with size and protein domain-organization similar to *yrInv*, although found exclusively in two nearly identical isolates within lineage B (NVI-4479, NVI-4493) and is thus not of significance for *Y. ruckeri* virulence.

yrIlm gene amplification. Gene amplification involving complex transposons is a mechanism utilized by bacteria to rapidly increase expression levels of a beneficial protein under extreme selection pressure, and it is often associated with increased antibiotic resistance following amplification of resistance cassettes [74]. Amplification of *yrIlm* is present in both of the complete genome sequences for CC2 (Fig 5). The temporal (1982/2000) and spatial (USA/Russia) separation of these strains indicates that *yrIlm* duplication may be a characteristic feature of CC2. In contrast, early isolates (late 1980s) of the exclusively Norwegian CC1, contain
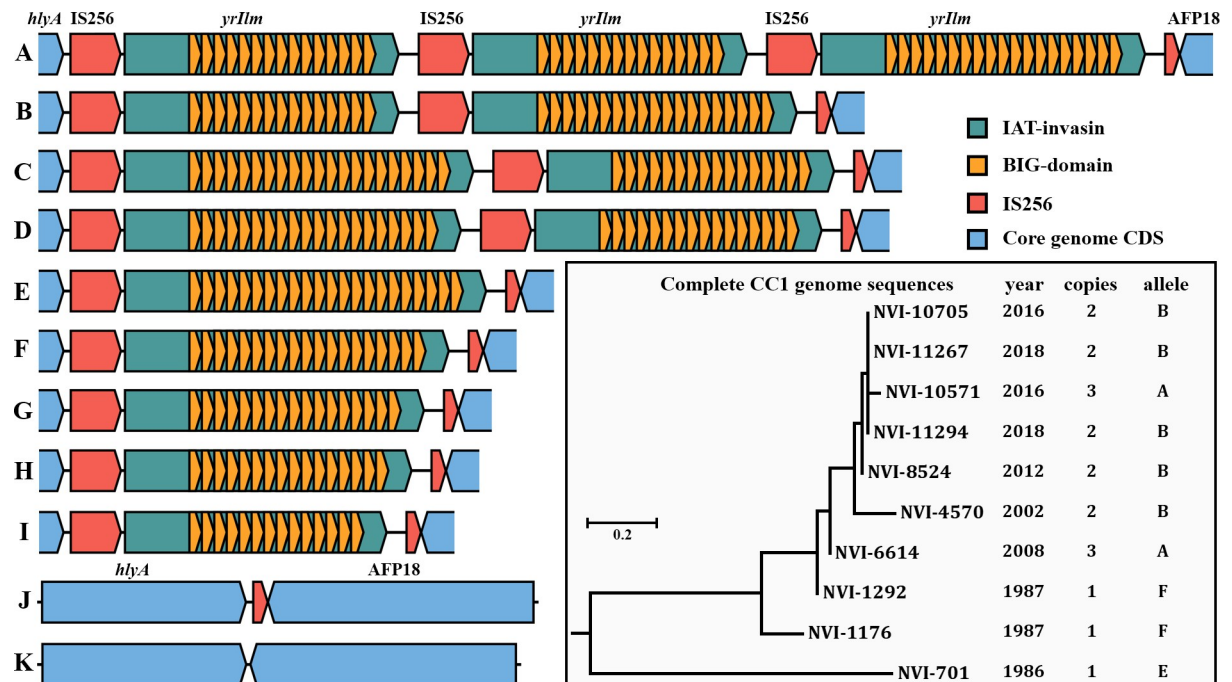
Fig 5. Genetic organization of the *yrIlm* locus, between *yhlA* and AFP18 in the *Y. ruckeri* core genome and maximum likelihood core gene phylogeny of complete CC1 genomes indicating year of isolation, *yrIlm* copy number and variant (bottom right). The complete sequence of strain NVH-3758 (CP023184; not included in this tree), isolated in Norway in 1987, contains a single copy of *yrIlm* [21]. Locus variants, labelled A to K, represent the following complete genomes: (A) NVI-6614 and NVI-10571 (CC1); (B) most modern CC1; (C) CFS007-82 (CC2); (D) KMM821 (CC2); (E) NVI-701 (CC1) and NVI-9681 (CC3); (F) NVI-1176 and NVI-1292 (CC1), and QMA0440 (CC5); (G) BigCreek74; (H) NVI-492 (CC10); (I) SC09; (J) empty site with partial IS256 in lineage B2; (K) empty site with no IS256 in lineages B1 and YRB, and MLVA clonal complexes CC7 and CC9.

https://doi.org/10.1371/journal.pone.0285257.g005

only a single *yrIlm* copy, whereas all later isolates are found to harbour two or three copies, suggesting gradual gene amplification over the last few decades (Fig 5B).

Not only does amplification of *yrIlm* likely increase expression levels, the additional allele displays differing numbers of BIG-repeats, likely affecting the length of the protein extending from the cell surface which may in turn affect adherence to host cells or other surfaces. However, as most available *Y. ruckeri* genome sequences are based on short read sequencing technologies that are inherently unsuitable for accurately sequencing such complex repeat structures, amplification of *yrIlm* may be more widespread than revealed here.

While the presence of flanking direct repeats may facilitate amplification, these also make the element prone to loss by homologous recombination in the absence of a positive selection pressure [74]. Loss of *yrIlm* by homologous recombination seems to have occurred in some lineages in which a sequence corresponding to a single partial IS256 remains at this locus (Fig 5). The complete absence of IS256 sequence in lineages B1 and YRB, as well as in CC7 and CC9, indicates loss by a different mechanism, or that the element was never present in these lineages.

Conclusions

A relatively small number of highly virulent strains of *Y. ruckeri* are capable of causing serious disease outbreaks in salmonid aquaculture. Our findings point to the presence of a single gene, the IAT-invasin *yrIlm*, as a major contributor to virulence in Atlantic salmon and rainbow trout pathogenic strains, as this gene is present in all virulent sub-lineages and is amplified in

the highly virulent CC1 and CC2. Amplification of *yrIlm* may contribute towards explaining the increasing impact of yersiniosis in Norwegian aquaculture from the mid-2000s and onwards [12, 75]. In virulent strains of non-O1 serotypes, where *yrIlm* is present, the lack of serotype O1 LPS may be responsible for the oft-cited lower virulence of such strains. While a multitude of other virulence-associated accessory features are also present in the pan-genome, some are more common in virulent lineages and likely provide some general or host-specific benefits. However, they seem to be less critical as they are generally not omnipresent amongst virulent strains. Further analysis is needed to verify the role of YrIlm and the importance of *yrIlm* copy number in virulence. In addition, further work is needed to define the role of the other accessory genetic determinants that may or may not contribute to variance in virulence and host specialization.

Materials and methods

Strains & culture

Detailed information on the *Y. ruckeri* isolates sequenced is provided in S1 Table. Isolates were cultured at 22˚C on 5% bovine blood agar (BA) or in tryptic soy broth (TSB). Sorbitol and ONPG phenotypes were assessed with API20e (BioMérieux, Marcy-l'Étoile, France) according to the manufacturers recommendations, although incubated at 22˚C for 48 hours. SDS phenotype was assessed on TSA agar plates with 1% SDS prepared according to Furones, Gilpin & Munn [33], with colonies surrounded by crystalline deposits after 48h incubation at 22˚C interpreted as a positive result. MLVA typing was performed according to Gulla, Mohammad & Colquhoun [76] or data derived from previous work [12, 18]. Serotypes were derived from previously published data or typed using in-house polyclonal rabbit antisera for *Yersinia ruckeri* O1, O2 and O5 with the slide agglutination technique.

Preparation of sequencing templates

For Illumina sequencing, DNA templates were extracted from pure cultures on BA with the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations for Gram-negative bacteria. For nanopore sequencing, DNA templates were extracted from overnight culture in TSB with the Gentra Puregene Yeast/Bact. Kit (Qiagen) following the manufacturer's descriptions for Gram-negative bacteria. DNA templates were assessed for purity with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), quantified by Qubit dsDNA HS Assay Kit on a Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA), and checked for integrity by agarose gel electrophoresis.

Whole genome sequencing

Genome assemblies were downloaded from the National Center for Biotechnology Information (NCBI), derived from previous work [18], or generated from 300bp paired end sequencing libraries prepared with NexteraFlex (Illumina, San Diego, CA, USA) and sequenced with the MiSeq platform (Illumina). Trimmomatic v0.38 [77] was used for trimming adapter sequences, cropping low-quality nucleotides from raw reads and filtering out raw reads prior to *de novo* assembly with SPAdes v3.13.0 or v3.9.0 [78] using the '—careful' option and otherwise default settings, or with Unicycler v0.4.8 [79] with default settings ('normal' mode) (see S1 Table).

Phylogenetic analysis

Annotation generated and downloaded genome assemblies was performed with Prokka v1.13 [80] with default settings. Core-gene alignments were generated using Roary v3.12.0 [81] with the MAFFT aligner, using a 95% identity cut-off, with a core definition of presence in 100% of isolates (option '-e—mafft -i 95 -cd 100'), and concatenated using snp-sites v2.4.1 [82]. Maximum likelihood trees were generated with MEGA v10.2 [83], bootstrap re-sampling (200 replicates) was used to assess branch support. Trees were visualized with MEGA (Figs 2, 5B and S1 Fig), or in R [84] with the ggtree package [85] (Fig 1).

Generation of complete genome assemblies

For generation of complete genome assemblies, to complement the already/publicly available complete assemblies, strains were selected based on core-gene phylogeny to cover the majority of relevant *Y. ruckeri* strains/lineages. Sequencing libraries for each strain selected for nanopore sequencing were prepared from 500 ng of genomic DNA with the Oxford Nanopore ligation sequencing kit (SQK-LSK109, Oxford Nanopore Technologies (ONT), Oxford, UK), following the protocol for Flongle as described by the manufacturer, with no fragmentation performed. External reagents used were the NEBNext Ultra II end repair/dA-tailing module (New England Biolabs, catalog #E7546) and AMPure XP paramagnetic beads (Beckman Coulter Life Sciences, CA, USA, catalog #A63880). Sequencing libraries were loaded onto Flongle flow cells (FLO-MIN106, ONT) and sequenced for 24 h with a MinION sequencer (ONT). Base calling was performed with Guppy v4.0.15 (ONT). Base called reads were filtered Nano-Filt v2.7.1 [86] by quality ('-q 8') and length (8Kbp, '-l 8000'). To ease hybrid assembly, base called reads were filtered by increasing length (up to 38Kbp) for some samples (S1 Table) until an uncompressed file size of approximately 500MB was obtained. Hybrid assemblies were produced with Unicycler with default settings by supplying trimmed Illumina reads and base called, filtered nanopore reads for each strain. As a control for representation of i.e. small plasmids, complete assemblies were compared to corresponding Illumina-based drafts by alignment of the draft contigs to the complete assembly by BLASTn.

To assess assembly quality of the highly repetitive *yrIlm*-region (between *yhlA* and AFP18 in the core genome), the region was, for the hybrid assemblies, aligned with nanopore-only assemblies produced with Flye version 2.9 ([87]; genome size set to 4Mbp and otherwise default options), as well as with five individual base called filtered nanopore reads spanning the entire *yrIlm*-region. In cases where Unicycler was in agreement with all individual nanopore reads, the Unicycler assembly was accepted. Otherwise, if the Flye assembly was in agreement with all individual nanopore reads, the Flye assembly was polished by mapping Illumina reads with Bowtie2 v2.3.4 [88] and the consensus of mapped reads was extracted with sam2-consensus v2 (developed by Edgardo M. Ortiz & Deise J. P. Gonçalves). The consensus was then used as a template to manually edit the Unicycler-produced hybrid assembly to account for missing repeats (mapped consensus sequence added) or excessive repeats (sequence deleted). Individual nanopore reads successfully aligning to both *yhlA* and AFP18 were always in agreement with each other in terms of the presence and number of duplicated *yrIlm* and internal 300bp BIG-domain repeats.

Secondary circular contigs were identified as plasmids by BLAST and/or identification of a plasmid replication protein. Plasmids were removed from assemblies prior to complete pan-genome analyses.

Pan-genome analysis

The complete genome of strain NVI-10705 (CC1) was selected as the base for building the pan-genome. Each additional genome was utilized as a reference sequence with the current pan-genome aligned as a query with BLAST Ring Image Generator (BRIG) v0.95 [89], using the option 'perc_identity 85' and otherwise default BLAST settings. Genetic features not already present in the pan-genome were characterized according to annotation (Prokka annotations and annotations of the assemblies downloaded from NCBI) or BLASTp, or by genetic context (i.e. new hypothetical genes that were intermixed with prophage genes were characterized as 'prophage-related') and added to the pan-genome.

In silico screening analysis

All genomes were assayed for novel and previously described putative virulence factors (including serotype- and metabolism-related) by BLASTn. Genes found to be split between different contigs in the Illumina-only sequences (e.g. *yrIlm*) were assayed using a truncated version as a query. For gene clusters (e.g. Tc genes and putative O-antigen clusters) and larger systems (e.g. secretion systems and O1-LPS), nucleotide sequences of three genes evenly spaced throughout the cluster were used as queries. In cases where these three searches did not agree (e.g. for Ysa and Yst1 in some instances), sequences were assessed individually by manual inspection to obtain presence/absence data for Fig 1.

Genetic maps (Figs 3–5) were visualized with DNAplotlib v1.0 [90] based on annotated sequence data. BIG-domains in IAT-invasins were identified by BLASTp.

Supporting information

S1 Table. Detailed information for sequenced isolates.
(XLSX)

S1 Fig. Core gene phylogeny, radial-style ML trees. The left tree is a subtree with lineage YRB excluded, with the attachment point of the YRB branch indicated by blue arrow. The right tree is the complete tree. See Fig 1 for bootstrap details and relevant metadata for each sequence.
(DOCX)

S2 Fig. Roary summary statistics. Summary statistics from Roary output respectively for complete genome assemblies only (top) and all assemblies (bottom), with (left) and without (right) YRB lineage genomes. The core gene alignment from 'All assemblies (n = 86)' was used to generate the trees in Fig 1 and S1 Fig.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Eve Zeyl Fiskebeck.

Investigation: Andreas Riborg.

Methodology: Andreas Riborg.

Writing – original draft: Andreas Riborg.

Writing – review & editing: Snorre Gulla, Eve Zeyl Fiskebeck, David Ryder, David W. Verner-Jeffreys, Duncan J. Colquhoun, Timothy J. Welch.

References

1. Sparboe O, Koren C, Håstein T, Poppe TT, Stenwig H. The first isolation of Yersinia ruckeri from farmed Norwegian salmon. Bull Eur Assoc Fish Pathol. 1986; 6, 41–42.

2. Rintamäki P, Valtonen ET, Frerichs GN. Occurrence of Yersinia ruckeri infection in farmed whitefish, Coregonus peled Gmelin and Coregonus muksun Pallas, and Atlantic salmon, Salmo salar L., in northern Finland. J Fish Dis. 1986; 9: 137–140. https://doi.org/10.1111/j.1365-2761.1986.tb00993.x

3. Costa AA, Leef MJ, Bridle AR, Carson J, Nowak B F. Effect of vaccination against yersiniosis on the relative percent survival, bactericidal and lysozyme response of Atlantic salmon, Salmo salar. Aquaculture. 2011; 315, 201–206. https://doi.org/10.1016/j.aquaculture.2011.02.031

4. Bastardo A, Bohle H, Ravelo C, Toranzo AE, Romalde JL. Serological and molecular heterogeneity among Yersinia ruckeri strains isolated from farmed Atlantic salmon Salmo salar in Chile. Dis Aquat Organ. 2011 Feb 22; 93(3):207–14. https://doi.org/10.3354/dao02296 PMID: 21516973

5. Wheeler RW, Davies RL, Dalsgaard I, Garcia J, Welch TJ, Wagley S et al. Yersinia ruckeri biotype 2 isolates from mainland Europe and the UK likely represent different clonal groups. Dis Aquat Organ. 2009 Mar 9; 84(1):25–33. https://doi.org/10.3354/dao02039 PMID: 19419004

6. Gudmundsdottir BK, Gudmundsdottir S, Gudmundsdottir S, Magnadottir B. Yersiniosis in Atlantic cod, Gadus morhua (L.), characterization of the infective strain and host reactions. J Fish Dis. 2014 Jun; 37 (6):511–9. https://doi.org/10.1111/jfd.12139 PMID: 23786306

7. Barnes AC, Delamare-Deboutteville J, Gudkovs N, Brosnahan C, Morrison R, Carson J. Whole genome analysis of *Yersinia ruckeri* isolated over 27 years in Australia and New Zealand reveals geographical endemism over multiple lineages and recent evolution under host selection. Microb Genom. 2016 Nov 30; 2(11):e000095. https://doi.org/10.1099/mgen.0.000095 PMID: 28348835

8. Liu T, Wang KY, Wang J, Chen DF, Huang XL, Ouyang P et al. Genome Sequence of the Fish Pathogen Yersinia ruckeri SC09 Provides Insights into Niche Adaptation and Pathogenic Mechanism. Int J Mol Sci. 2016; 17(4), 557. https://doi.org/10.3390/ijms17040557 PMID: 27089334

9. Willumsen B. Birds and wild fish as potential vectors of Yersinia ruckeri. J Fish Dis. 1989; 12, 275–277. https://doi.org/10.1111/j.1365-2761.1989.tb00313.x

10. Barnes AC. Enteric redmouth disease (ERM) (Yersinia ruckeri). In Woo P. T. K., & Bruno D. W (Eds.), Fish diseases and disorders. Volume 3: viral, bacterial and fungal infections. 2011. pp. 484–511. http://dx.doi.org/10.1079/9781845935542.0000

11. Wade J. Characterization of Yersinia ruckeri and enteric redmouth disease (ERM) to inform pathogen transfer risk assessments in British Columbia. DFO Canadian Science Advisory Secretariat Research Document 2019/022. 2019. available from: https://waves-vagues.dfo-mpo.gc.ca/Library/40853032.pdf

12. Gulla S, Barnes AC, Welch TJ, Romalde JL, Ryder D, Ormsby MJ et al. Multilocus Variable-Number Tandem-Repeat Analysis of Yersinia ruckeri Confirms the Existence of Host Specificity, Geographic Endemism, and Anthropogenic Dissemination of Virulent Clones. Appl Environ Microbiol. 2018 Aug 1; 84(16):e00730–18. https://doi.org/10.1128/AEM.00730-18 PMID: 29884756

13. Riborg A, Gulla S, Strand D, Wiik-Nielsen J, Rønneseth A, Welch TJ, et al. qPCR screening for Yersinia ruckeri clonal complex 1 against a background of putatively avirulent strains in Norwegian aquaculture. J Fish Dis. 2022 Aug; 45(8):1211–1224. https://doi.org/10.1111/jfd.13656 PMID: 35648597

14. Bullock GL, Stuckeye HM, Shotts JR. Enteric redmouth bacterium: comparison of isolates from different geographic areas. J Fish Dis. 1978; 1(4), 351–356. https://doi.org/10.1111/j.1365-2761.1978.tb00039.x

15. McCarthy DH, Johnson KA. A serotypic survey and cross-protection test of North American field isolates of Yersinia ruckeri. J Fish Dis. 1982; 5: 323–328 https://doi.org/10.1111/j.1365-2761.1982.tb00487.x

16. Verner-Jeffreys D, Haig SJ, Welch TJ, Pond MJ, Stone D, Davies RL, et al. Characterisation of a serotype O1 Yersinia ruckeri isolate from the Isle of Man: Further evidence that O antigen serotype is not a reliable indicator of virulence. Bull Eur Assoc Fish Pathol. 2011; 31, 86.

17. Haig SJ, Davies RL, Welch TJ, Reese RA, Verner-Jeffreys DW. Comparative susceptibility of Atlantic salmon and rainbow trout to Yersinia ruckeri: relationship to O antigen serotype and resistance to

serum killing. Vet Microbiol. 2011 Jan 10; 147(1–2):155–61. https://doi.org/10.1016/j.vetmic.2010.06. 022 PMID: 20691549

18. Riborg A, Colquhoun DJ, Gulla S. Biotyping reveals loss of motility in two distinct Yersinia ruckeri lineages exclusive to Norwegian aquaculture. J Fish Dis. 2022 May; 45(5):641–653. https://doi.org/10. 1111/jfd.13590 PMID: 35180320

19. Nelson MC, LaPatra SE, Welch TJ, Graf J. Complete Genome Sequence of Yersinia ruckeri Strain CSF007-82, Etiologic Agent of Red Mouth Disease in Salmonid Fish. Genome Announc. 2015 Jan 29; 3(1):e01491–14. https://doi.org/10.1128/genomeA.01491-14 PMID: 25635018

20. Johnson SL, Daligault HE, Davenport KW, Jaissle J, Frey KG, Ladner JT et al. Thirty-Two Complete Genome Assemblies of Nine Yersinia Species, Including Y. pestis, Y. pseudotuberculosis, and Y. enterocolitica. Genome Announc. 2015 Apr 30; 3(2):e00148–15. https://doi.org/10.1128/genomeA.00148-15 PMID: 25931590

21. Wrobel A, Ottoni C, Leo JC, Gulla S, Linke D. The repeat structure of two paralogous genes, Yersinia ruckeri invasin (yrInv) and a "Y. ruckeri invasin-like molecule", (yrIlm) sheds light on the evolution of adhesive capacities of a fish pathogen. J Struct Biol. 2018 Feb; 201(2):171–183. https://doi.org/10. 1016/j.jsb.2017.08.008 PMID: 28888816

22. Wrobel A, Ottoni C, Leo JC, Linke D. pYR4 From a Norwegian Isolate of *Yersinia ruckeri* Is a Putative Virulence Plasmid Encoding Both a Type IV Pilus and a Type IV Secretion System. Front Cell Infect Microbiol. 2018 Oct 30; 8:373. https://doi.org/10.3389/fcimb.2018.00373 PMID: 30460204

23. Welch TJ, LaPatra S. Yersinia ruckeri lipopolysaccharide is necessary and sufficient for eliciting a protective immune response in rainbow trout (Oncorhynchus mykiss, Walbaum). Fish Shellfish Immunol. 2016 Feb; 49:420–6. https://doi.org/10.1016/j.fsi.2015.12.037 PMID: 26766181

24. Severi E, Hood DW, Thomas GH. Sialic acid utilization by bacterial pathogens. Microbiology (Reading). 2007 Sep; 153(Pt 9):2817–2822. https://doi.org/10.1099/mic.0.2007/009480-0 PMID: 17768226

25. Welch TJ. Characterization of a novel Yersinia ruckeri serotype O1-specific bacteriophage with virulence-neutralizing activity. J Fish Dis. 2020 Feb; 43(2):285–293. https://doi.org/10.1111/jfd.13124 PMID: 31828808

26. Wang X, Preston JF 3rd, Romeo T. The pgaABCD locus of Escherichia coli promotes the synthesis of a polysaccharide adhesin required for biofilm formation. J Bacteriol. 2004 May; 186(9):2724–34. https:// doi.org/10.1128/JB.186.9.2724-2734.2004 PMID: 15090514

27. Romalde J. L., MagariÑos B., Barja J. L., & Toranzo A. E. (1993). Antigenic and Molecular Characterization of Yersinia ruckeri Proposal for a New Intraspecies Classification. Systematic and Applied Microbiology, 16(3), 411–419. https://doi.org/10.1016/S0723-2020(11)80274-2.

28. Jank T, Eckerle S, Steinemann M, Trillhaase C, Schimpl M, Wiese S et al. Tyrosine glycosylation of Rho by Yersinia toxin impairs blastomere cell behaviour in zebrafish embryos. Nat Commun. 2015 Jul 20; 6:7807. https://doi.org/10.1038/ncomms8807 PMID: 26190758

29. O'Leary PJ, Rohovec JS, Sanders JE, Fryer JL. Serotypes of Yersinia ruckeri and their immunogenic properties. Oregon State University Agricultural Experiment Station Technical paper no. 6235. 1982. Available from: https://repository.library.noaa.gov/view/noaa/10921

30. Cipriano RC, Pyle JB. Development of a culture medium for determination of sorbitol utilization among strains of Yersinia ruckeri. Microbios Letters 1985; 28:79–82.

31. Stevenson RMW, Airdrie DW. Serological variation among Yersinia ruckeri strains. J Fish Dis. 1984; 7 (4), 247–254. https://doi.org/10.1111/j.1365-2761.1984.tb00930.x

32. Furones MD, Gilpin MJ, Alderman DJ, Munn CB. Virulence of Yersinia ruckeri serotype I strains is associated with a heat sensitive factor (HSF) in cell extracts. FEMS Microbiol Lett. 1990 Jan 1; 54(1–3):339–43. https://doi.org/10.1016/0378-1097(90)90309-e PMID: 2323544

33. Furones MD, Gilpin ML, Munn CB. Culture media for the differentiation of isolates of Yersinia ruckeri, based on detection of a virulence factor. J Appl Bacteriol. 1993 Apr; 74(4):360–6. https://doi.org/10. 1111/j.1365-2672.1993.tb05139.x PMID: 8486541

34. Navais R, Méndez J, Cascales D, Reimundo P, Guijarro JA. The heat sensitive factor (HSF) of Yersinia ruckeri is produced by an alkyl sulphatase involved in sodium dodecyl sulphate (SDS) degradation but not in virulence. BMC Microbiol. 2014 Sep 30; 14:221. https://doi.org/10.1186/s12866-014-0221-7 PMID: 25266819

35. Stevenson RMW, Daly JG. Biochemical and Serological Characteristics of Ontario Isolates of Yersinia ruckeri. Can J Fish Aquat Sci. 1982; 39(6), 870–876. https://doi.org/10.1139/f82-118

36. Liu T, Li L, Wei W, Wang K, Yang Q, Wang E. Yersinia ruckeri strain SC09 disrupts proinflammatory activation via Toll/IL-1 receptor-containing protein STIR-3. Fish Shellfish Immunol. 2020 Apr; 99:424–434. https://doi.org/10.1016/j.fsi.2020.02.035 PMID: 32087278

37. Wrobel A, Saragliadis A, Pérez-Ortega J, Sittman C, Göttig S, Liskiewicz K et al. The inverse autotransporters of Yersinia ruckeri, YrInv and YrIlm, contribute to biofilm formation and virulence. Environ Microbiol. 2020 Jul; 22(7):2939–2955. https://doi.org/10.1111/1462-2920.15051 PMID: 32372498

38. Secades P, Guijarro JA. Purification and characterization of an extracellular protease from the fish pathogen Yersinia ruckeri and effect of culture conditions on production. Appl Environ Microbiol. 1999 Sep; 65(9):3969–75. https://doi.org/10.1128/AEM.65.9.3969-3975.1999 PMID: 10473403

39. Fernández L, Prieto M, Guijarro JA. The iron- and temperature-regulated haemolysin YhlA is a virulence factor of Yersinia ruckeri. Microbiology (Reading). 2007 Feb; 153(Pt 2):483–489. https://doi.org/10.1099/mic.0.29284-0 PMID: 17259619

40. Fernández L, Márquez I, Guijarro JA. Identification of specific in vivo-induced (ivi) genes in Yersinia ruckeri and analysis of ruckerbactin, a catecholate siderophore iron acquisition system. Appl Environ Microbiol. 2004 Sep; 70(9):5199–207. https://doi.org/10.1128/AEM.70.9.5199–5207.2004

41. Méndez J, Reimundo P, Pérez-Pascual D, Navais R, Gómez E, Guijarro JA. A novel cdsAB operon is involved in the uptake of L-cysteine and participates in the pathogenesis of Yersinia ruckeri. J Bacteriol. 2011 Feb; 193(4):944–51. https://doi.org/10.1128/JB.01058-10 PMID: 21169490

42. Dahiya I, Stevenson RM. The ZnuABC operon is important for Yersinia ruckeri infections of rainbow trout, Oncorhynchus mykiss (Walbaum). J Fish Dis. 2010 Apr; 33(4):331–40. https://doi.org/10.1111/j.1365-2761.2009.01125.x PMID: 20070462

43. Dahiya I, Stevenson RM. The UvrY response regulator of the BarA-UvrY two-component system contributes to Yersinia ruckeri infection of rainbow trout (Oncorhynchus mykiss). Arch Microbiol. 2010 Jul; 192(7):541–7. https://doi.org/10.1007/s00203-010-0582-8 PMID: 20480360

44. Wang E, Qin Z, Yu Z, Ai X, Wang K, Yang Q et al. Molecular Characterization, Phylogenetic, Expression, and Protective Immunity Analysis of OmpF, a Promising Candidate Immunogen Against *Yersinia ruckeri* Infection in Channel Catfish. Front Immunol. 2018 Sep 13; 9:2003. https://doi.org/10.3389/fimmu.2018.02003 PMID: 30271401

45. Ormsby MJ, Grahame E, Burchmore R, Davies RL. Comparative bioinformatic and proteomic approaches to evaluate the outer membrane proteome of the fish pathogen Yersinia ruckeri. J Proteomics. 2019 May 15; 199:135–147. https://doi.org/10.1016/j.jprot.2019.02.014 PMID: 30831250

46. Gunasena DK, Komrower JR, MacIntyre S. The fish pathogen Yersinia ruckeri possesses a TTS system. Adv Exp Med Biol. 2003; 529:105–7. https://doi.org/10.1007/0-306-48416-1_19 PMID: 12756737

47. Welch TJ, Verner-Jeffreys DW, Dalsgaard I, Wiklund T, Evenhuis JP, Cabrera JA et al. Independent emergence of Yersinia ruckeri biotype 2 in the United States and Europe. Appl Environ Microbiol. 2011 May; 77(10):3493–9. https://doi.org/10.1128/AEM.02997-10 PMID: 21441334

48. Cornelis GR, Boland A, Boyd AP, Geuijen C, Iriarte M, Neyt C et al. The virulence plasmid of Yersinia, an antihost genome. Microbiol Mol Biol Rev. 1998 Dec; 62(4):1315–52. https://doi.org/10.1128/MMBR.62.4.1315-1352.1998 PMID: 9841674

49. De Grandis SA, Stevenson RMW. Variations in plasmid profiles and growth characteristics of Yersinia ruckeri strains. FEMS Microbiol Immunol. 1982; 15(3):199–202. https://doi.org/10.1111/j.1574-6968.1982.tb00067.x

50. Stave J, Cook T, Roberson B. Chemiluminescent responses of striped bass, Morone saxatilis (Walbaum), phagocytes to strains of Yersinia ruckeri. J Fish Dis. 1987; 10, 1–10. https://doi.org/10.1111/j.1365-2761.1987.tb00712.x

51. Guilvout I, Quilici ML, Rabot S, Lesel R, Mazigh D. BamHI restriction endonuclease analysis of Yersinia ruckeri plasmids and their relatedness to the genus Yersinia 42- to 47-megadalton plasmid. Appl Environ Microbiol. 1988 Oct; 54(10):2594–7. https://doi.org/10.1128/aem.54.10.2594-2597.1988 PMID: 3202639

52. Garcia JA, Dominguez L, Larsen JL, Pedersen K. Ribotyping and plasmid profiling of Yersinia ruckeri. J Appl Microbiol. 1998 Dec; 85(6):949–55. https://doi.org/10.1111/j.1365-2672.1998.tb05258.x PMID: 9871314

53. Méndez J, Fernández L, Menéndez A, Reimundo P, Pérez-Pascual D, Navais R, et al. A chromosomally located traHIJKCLMN operon encoding a putative type IV secretion system is involved in the virulence of Yersinia ruckeri. Appl Environ Microbiol. 2009 Feb; 75(4):937–45. https://doi.org/10.1128/AEM.01377-08 PMID: 19088314

54. Pinheiro VB, Ellar DJ. Expression and insecticidal activity of Yersinia pseudotuberculosis and Photorhabdus luminescens toxin complex proteins. Cell Microbiol. 2007 Oct; 9(10):2372–80. https://doi.org/10.1111/j.1462-5822.2007.00966.x PMID: 17573906

55. Tennant SM, Skinner NA, Joe A, Robins-Browne RM. Homologues of insecticidal toxin complex genes in Yersinia enterocolitica biotype 1A and their contribution to virulence. Infect Immun. 2005 Oct; 73(10):6860–7. https://doi.org/10.1128/IAI.73.10.6860-6867.2005 PMID: 16177365

56. Hares MC, Hinchliffe SJ, Strong PCR, Eleftherianos I, Dowling AJ, Ffrench-Constant RH, et al. The Yersinia pseudotuberculosis and Yersinia pestis toxin complex is active against cultured mammalian cells. Microbiology (Reading). 2008 Nov; 154(Pt 11):3503–3517. https://doi.org/10.1099/mic.0.2008/018440-0 PMID: 18957603

57. Cascales D, Guijarro JA, García-Torrico AI, Méndez J. Comparative genome analysis reveals important genetic differences among serotype O1 and serotype O2 strains of Y. ruckeri and provides insights into host adaptation and virulence. Microbiologyopen. 2017 Aug; 6(4):e00460. https://doi.org/10.1002/mbo3.460 PMID: 28317294

58. Fuchs TM, Bresolin G, Marcinowski L, Schachtner J, Scherer S. Insecticidal genes of Yersinia spp.: taxonomical distribution, contribution to toxicity towards Manduca sexta and Galleria mellonella, and evolution. BMC Microbiol. 2008 Dec 8; 8:214. https://doi.org/10.1186/1471-2180-8-214 PMID: 19063735

59. Springer K, Sänger PA, Moritz C, Felsl A, Rattei T, Fuchs TM. Insecticidal Toxicity of *Yersinia frederiksenii* Involves the Novel Enterotoxin YacT. Front Cell Infect Microbiol. 2018 Nov 14; 8:392. https://doi.org/10.3389/fcimb.2018.00392 PMID: 30488025

60. Thomson NR, Howard S, Wren BW, Holden MT, Crossman L, Challis GL et al. The complete genome sequence and comparative genome analysis of the high pathogenicity Yersinia enterocolitica strain 8081. PLoS Genet. 2006 Dec 15; 2(12):e206. https://doi.org/10.1371/journal.pgen.0020206 PMID: 17173484

61. Bent ZW, Poorey K, Brazel DM, LaBauve AE, Sinha A, Curtis DJ et al. Transcriptomic Analysis of Yersinia enterocolitica Biovar 1B Infecting Murine Macrophages Reveals New Mechanisms of Extracellular and Intracellular Survival. Infect Immun. 2015 Jul; 83(7):2672–85. https://doi.org/10.1128/IAI.02922-14 PMID: 25895974

62. Haller JC, Carlson S, Pederson KJ, Pierson DE. A chromosomally encoded type III secretion pathway in Yersinia enterocolitica is important in virulence. Mol Microbiol. 2000 Jun; 36(6):1436–46. https://doi.org/10.1046/j.1365-2958.2000.01964.x PMID: 10931293

63. Iwobi A, Heesemann J, Garcia E, Igwe E, Noelting C, Rakin A. Novel virulence-associated type II secretion system unique to high-pathogenicity Yersinia enterocolitica. Infect Immun. 2003 Apr; 71(4):1872–9. https://doi.org/10.1128/IAI.71.4.1872-1879.2003 PMID: 12654803

64. Venecia K, Young GM. Environmental regulation and virulence attributes of the Ysa type III secretion system of Yersinia enterocolitica biovar 1B. Infect Immun. 2005 Sep; 73(9):5961–77. https://doi.org/10.1128/IAI.73.9.5961–5977.2005

65. Zhu J, Oger PM, Schrammeijer B, Hooykaas PJ, Farrand SK, Winans SC. The bases of crown gall tumorigenesis. J Bacteriol. 2000 Jul; 182(14):3885–95. https://doi.org/10.1128/JB.182.14.3885-3895.2000 PMID: 10869063

66. Cascales E, Christie PJ. The versatile bacterial type IV secretion systems. Nat Rev Microbiol. 2003 Nov; 1(2):137–49. https://doi.org/10.1038/nrmicro753 PMID: 15035043

67. Voth DE, Broederdorf LJ, Graham JG. Bacterial Type IV secretion systems: versatile virulence machines. Future Microbiol. 2012 Feb; 7(2):241–57. https://doi.org/10.2217/fmb.11.150 PMID: 22324993

68. Liu T, Wang E, Wei W, Wang K, Yang Q, Ai X. TcpA, a novel Yersinia ruckeri TIR-containing virulent protein mediates immune evasion by targeting MyD88 adaptors. Fish Shellfish Immunol. 2019 Nov; 94:58–65. https://doi.org/10.1016/j.fsi.2019.08.069 PMID: 31470137

69. Vayssier-Taussat M, Le Rhun D, Deng HK, Biville F, Cescau S, Danchin A et al. The Trw type IV secretion system of Bartonella mediates host-specific adhesion to erythrocytes. PLoS Pathog. 2010 Jun 10; 6(6):e1000946. https://doi.org/10.1371/journal.ppat.1000946 PMID: 20548954

70. González-Rivera C, Khara P, Awad D, Patel R, Li YG, Bogisch M, et al. Two pKM101-encoded proteins, the pilus-tip protein TraC and Pep, assemble on the Escherichia coli cell surface as adhesins required for efficient conjugative DNA transfer. Mol Microbiol. 2019 Jan; 111(1):96–117. https://doi.org/10.1111/mmi.14141 PMID: 30264928

71. Tsai JC, Yen MR, Castillo R, Leyton DL, Henderson IR, Saier MH Jr. The bacterial intimins and invasins: a large and novel family of secreted proteins. PLoS One. 2010 Dec 22; 5(12):e14403. https://doi.org/10.1371/journal.pone.0014403 PMID: 21203509

72. Leo JC, Oberhettinger P, Schütz M, Linke D. The inverse autotransporter family: intimin, invasin and related proteins. Int J Med Microbiol. 2015 Feb; 305(2):276–82. https://doi.org/10.1016/j.ijmm.2014.12.011 PMID: 25596886

73. Sadana P, Mönnich M, Unverzagt C, Scrima A. Structure of the Y. pseudotuberculosis adhesin InvasinE. Protein Sci. 2017 Jun; 26(6):1182–1195. https://doi.org/10.1002/pro.3171 PMID: 28370712

74. Sandegren L, Andersson DI. Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat Rev Microbiol. 2009 Aug; 7(8):578–88. https://doi.org/10.1038/nrmicro2174 PMID: 19609259

75. Gulla S, Olsen AB. Yersiniosis. In Sommerset I., Walde C. S., Bang Jensen B., Bornø B., Haukaas A., & Brun E. (Eds.), The health situation in Norwegian aquaculture 2019. 2020. pp. 80–82. Published by the Norwegian Veterinary Institute. Available from: https://www.vetinst.no/rapporter-og-publikasjoner/rapporter/2020/fish-health-report-2019

76. Gulla S, Mohammad SN, Colquhoun DJ. Multi-locus Variable-number Tandem-repeat Analysis of the Fish-pathogenic Bacterium Yersinia ruckeri by Multiplex PCR and Capillary Electrophoresis. J Vis Exp. 2019 Jun 17;(148). https://doi.org/10.3791/59455 PMID: 31259900

77. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1; 30(15):2114–20. https://doi.org/10.1093/bioinformatics/btu170 PMID: 24695404

78. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May; 19(5):455–77. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599

79. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol. 2017 Jun 8; 13(6):e1005595. https://doi.org/10.1371/journal.pcbi.1005595 PMID: 28594827

80. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014 Jul 15; 30(14):2068–9. https://doi.org/10.1093/bioinformatics/btu153 PMID: 24642063

81. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MT et al. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 2015 Nov 15; 31(22):3691–3. https://doi.org/10.1093/bioinformatics/btv421 PMID: 26198102

82. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. *SNP-sites*: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom. 2016 Apr 29; 2(4):e000056. https://doi.org/10.1099/mgen.0.000056 PMID: 28348851

83. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol. 2018 Jun 1; 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096 PMID: 29722887

84. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. ISBN 3-900051-07-0, http://www.R-project.org/

85. Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017; 8(1), 28–36. https://doi.org/10.1111/2041-210X.12628

86. De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 2018 Aug 1; 34(15):2666–2669. https://doi.org/10.1093/bioinformatics/bty149 PMID: 29547981

87. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019 May; 37(5):540–546. https://doi.org/10.1038/s41587-019-0072-8 PMID: 30936562

88. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4; 9 (4):357–9. https://doi.org/10.1038/nmeth.1923 PMID: 22388286

89. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011 Aug 8; 12:402. https://doi.org/10.1186/1471-2164-12-402 PMID: 21824423

90. Der BS, Glassey E, Bartley BA, Enghuus C, Goodman DB, Gordon DB et al. DNAplotlib: Programmable Visualization of Genetic Designs and Associated Data. ACS Synth Biol. 2017 Jul 21; 6(7):1115–1119. https://doi.org/10.1021/acssynbio.6b00252 PMID: 27744689