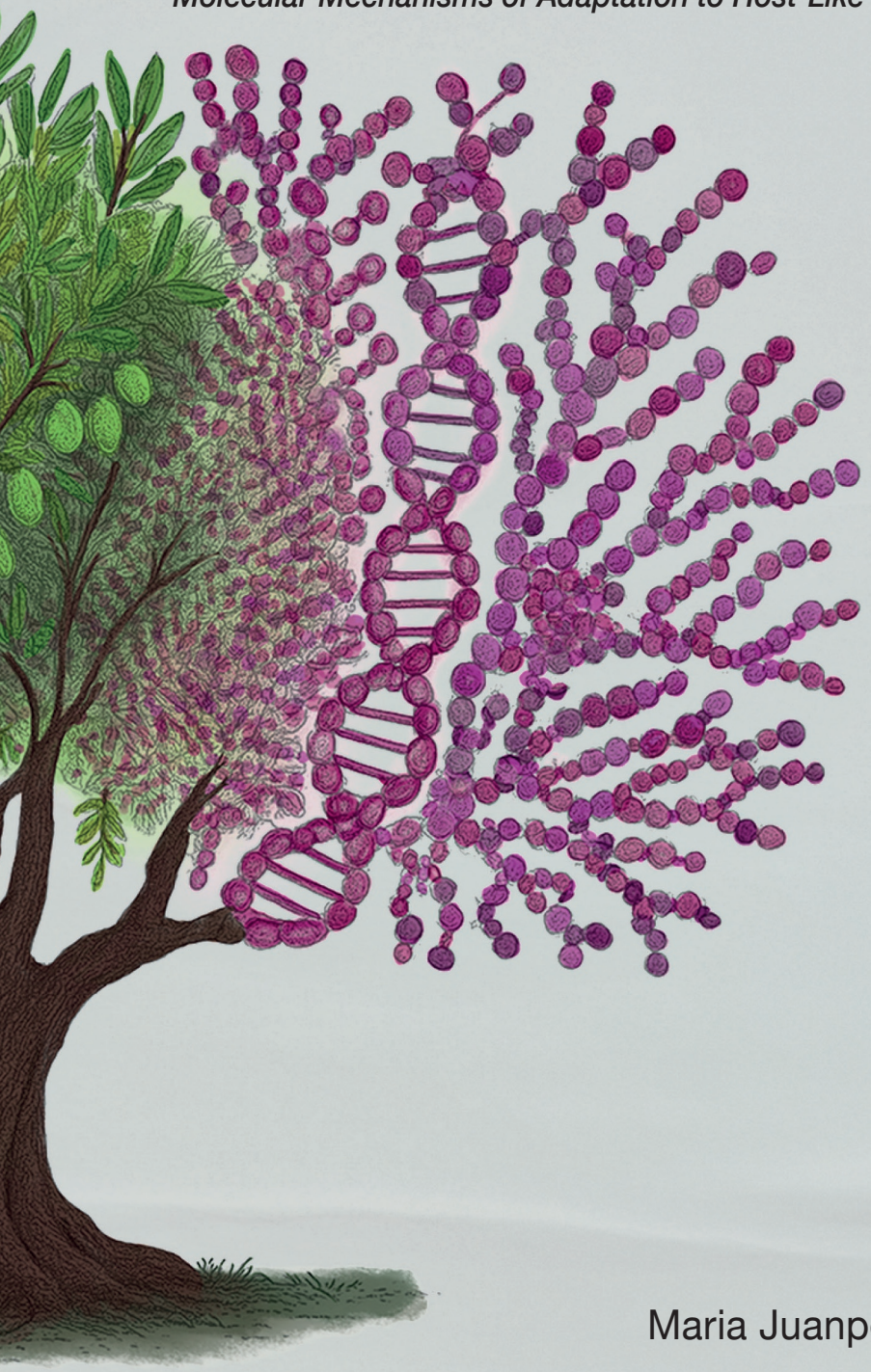
Finding the Achilles' Heel of *Streptococcus suis*

Molecular Mechanisms of Adaptation to Host-Like Environments



Maria Juanpere-Borràs

Propositions

1. Transposon libraries are a starting point rather than a final result.
   (this thesis)
2. Regulation of a virulence factor is more important than the virulence factor.
   (this thesis)
3. Freedom to define questions, formulate hypotheses, and choose methods is essential for becoming an independent researcher.
4. All fundamental research is of societal relevance.
5. In molecular cloning, protocols are unnecessary.
6. A person's understanding of a topic is best assessed by the ability to ask critical questions.
7. Gender quotas increase impostor syndrome in women.

Propositions belonging to the thesis, entitled

Finding the Achilles' Heel of *Streptococcus suis*. Molecular Mechanisms of Adaptation to Host-Like Environments.

Maria Juanpere-Borràs
Wageningen, 16 January 2026

Finding the Achilles' Heel of *Streptococcus suis*

Molecular Mechanisms of Adaptation to Host-Like Environments

Maria Juanpere-Borràs

Thesis committee

Promotors
Prof. Dr Jerry M. Wells
Professor of Host-Microbe Interactomics Group
Wageningen University & Research

Co-promotors
Dr Peter van Baarlen
Researcher, Host-Microbe Interactomics Group
Wageningen University & Research

Other members
Prof. Dr Hauke Smidt, Wageningen University & Research
Dr Manouk Vrieling, Wageningen Bioveterinary Research
Dr José Antonio Escudero García-Calderón, Universidad Complutense de Madrid, Spain
Dr Virginia Aragón Fernández, Institute of Agrifood Research and Technology, Barcelona, Spain

This research was conducted under the auspices of the Graduate School Wageningen Institute of Animal Science (WIAS).

Finding the Achilles' Heel of *Streptococcus suis*

Molecular Mechanisms of Adaptation to Host-Like Environments

Maria Juanpere-Borràs

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 16 January 2026
at 3.30 p.m. in the Omnia Auditorium.

Maria Juanpere-Borràs

Finding the Achilles' Heel of Streptococcus suis. Molecular Mechanisms of Adaptation to Host-Like Environments, 206 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2026)
With references, with summary in English

DOI: https://doi.org/10.18174/681597

“To the impostor in me”

Nothing in life is to be feared, it is only to be understood.

Marie Skłodowska-Curie

Table of contents

Chapter 1

General Introduction

The rapid rise of antimicrobial resistance (AMR) in recent decades has become one of the most pressing threats to global health [1]. Driven by widespread antibiotic use and the ability of bacteria to exchange resistance genes through horizontal transfer, multidrug-resistant pathogens are now increasingly common in both clinical settings and livestock production, representing a growing zoonotic threat [2–4]. One such example is *Streptococcus suis*, a Gram-positive, encapsulated bacterium responsible for severe diseases in pigs, such as meningitis and septicemia, and an emerging cause of human infections [4–6]. *S. suis* causes substantial economic losses in the swine industry and contributes to the global AMR problem through the need to treat and control outbreaks using metaphylactic antibiotic use [7,8]. Recent genomic studies have revealed high levels of AMR in *S. suis* and the emergence of zoonotic lineages with resistance to penicillin and ceftriaxone, which are standard treatments for *S. suis* infections in humans [9–11]. This thesis uses Tn-seq and molecular genetic analyses of sensory and epigenetic systems to uncover how *S.suis* adapts and causes disease, providing fundamental insights and identifying candidate targets, that can inform future strategies to control and treat this pathogen.

1. *S. suis*

S. suis is common member of the porcine tonsillar associated microbiota and has been detected in other parts of the upper respiratory tract, small intestine, and birth canal [12,13]. Large-scale genomic studies have shown that the *S. suis* population comprises both commensal ecotypes and pathogenic ecotypes, which have evolved from ancestral commensals, likely driven by the intensification of pig farming [4]. The first confirmed case of *S. suis* infection was reported in 1954, and the first documented human case occurred in Denmark in 1968 [6,14,15]. *S. suis* is now recognized as a major cause of adult meningitis and septicemia in Southeast Asia, with at least 913 meningitis cases reported between 1980 and 2015. In pigs, outbreaks of *S. suis* disease are most frequently reported in China and Vietnam, but cases are also widespread in Europe, North America, Australia, and New Zealand [14,16].

A study assessing the impact of *S. suis* in Germany, the Netherlands, and Spain reported that mortality during the nursery phase is usually below 5%. However, in untreated outbreaks, mortality can rise to 20–30%.[17]. Consequently, the use of antimicrobials in pig production is considered a cost-effective strategy to minimize economic losses [17].

1.1 *S. suis* pathogenesis

Piglets acquire the microbiota from the sow shortly after birth, and studies have shown that the composition and diversity of the microbiota before weaning influence susceptibility to post weaning *S. suis* disease [18].

S. suis is very common in the tonsillar microbiota and can reside deep within tonsillar crypts, and inside macrophages within the tonsil lymphoid tissue which may access lymphatic vessels and traffic to regional lymph nodes. From there, lymphatic drainage to the thoracic duct and into the subclavian vein provides access to the bloodstream [19]. *S. suis* can also be detected in the nasal cavity and in the intestinal tract of piglets, suggesting other possible routes of entry into the body [20,21]. In inflamed tissues, direct passage into fenestrated blood vessels is also possible. Its polysaccharide capsule (CPS) and other surface proteins enable survival in blood by evading complement activation and phagocytosis, facilitating systemic dissemination [22–24]. This systemic spread can lead to septicemia, characterized by fever and in severe cases, sudden death [22]. From the bloodstream, *S. suis* can disseminate to multiple organs, including the joints, heart, and lungs, where it may cause arthritis, endocarditis, or pneumonia. At later stages, it is also able to cross the blood–brain barrier via the brain microvasculature and the choroid plexus, leading to meningitis, the most frequently reported manifestation in both pigs and humans [23–25].

A large number of virulence factors have been described for *S. suis* but it remains unclear which are critical for pathogenesis due to the genetic diversity of the species and redundancy of function [21,23]. At least 34 genes have been described as contributing to epithelial cell adhesion, yet it remains unclear whether any are critical for successful colonization [20,23]. For example, Suilysin (Sly) the cholesterol dependent cytolysin is considered an important virulence factor and produced by many strains. However, some virulent isolates lacking Sly still cause disease in animals, showing that it is not strictly essential for pathogenesis [23,24]. At later stages of infection, enolase was identified as contributing to central nervous system (CNS) invasion by binding plasminogen and promoting its activation to plasmin, thereby facilitating the degradation of endothelial tight junctions and translocation across the blood–brain barrier. However, translocation *in vitro* can still occur, albeit at reduced levels, in the absence of plasminogen, suggesting alternative or compensatory mechanisms [25]. The virulence factors implicated in *S. suis* pathogenesis include cell surface proteins, toxins, enzymes, sensory systems, transcriptional regulators, transporters, epigenetic regulators, and immune evasion mechanisms [23,26]. Two traits emerge as major contributors to the success of *S. suis* as a pathogen: the regulation of CPS biosynthesis and the widespread acquisition of AMR, which enables *S. suis* to persist in piglets even after metaphylactic treatment.

1.2 The role of the CPS in *S. suis* pathogenesis

The CPS of *S. suis* is a complex structure composed of long chains of glucose, galactose, N-acetylglucosamine, rhamnose, and sialic acid [27,28]. In Gram-positive bacteria, these polysaccharide chains are covalently attached to the peptidoglycan wall [29]. While this general composition is conserved, subtle structural variations in the arrangement of these sugars and the presence or absence of side chains among different serotypes contribute to their significant antigenic diversity. Sialic acid, in particular, plays a crucial role in immune evasion by mimicking host glycoconjugates, thereby reducing complement activation and opsonization [28,30].

Capsule expression in *S. suis* is highly dynamic and adapts to distinct stages of infection. During colonization, capsule thickness may decrease, allowing surface adhesins to interact with epithelial cells and facilitate adherence and biofilm formation [30–32]. In contrast, during systemic dissemination, capsule production is upregulated, providing enhanced protection against antimicrobial peptides, phagocytosis, and complement-mediated killing, critical for bacterial survival in the bloodstream [31,33]. This capacity for dynamic regulation underscores the importance of capsule-associated regulatory systems in pathogenesis.

As the capsule is a key determinant of virulence in *S. suis*, its antigenic variability forms the basis for the serotyping system, which currently distinguishes 29 serotypes [23]. Serotype 2 is the most commonly associated with disease globally and is also the primary serotype identified in human infections, although serotypes 9, 1, and 7, predominate in certain regions, particularly in Europe [34,35]. Despite its utility, CPS-based classification does not fully reflect the genetic and phenotypic diversity of *S. suis*. Strains within the same serotype can vary in their pathogenic potential [36].

This limitation has led to the adoption of multilocus sequence typing (MLST) as a complementary approach to characterize *S. suis* populations with greater resolution. MLST analyzes the sequence profiles of several housekeeping genes to classify isolates into distinct sequence types (STs). As of 2016, more than 704 STs had been described, with ST1 being the most prevalent and strongly associated with invasive disease [4,34].

1.3 AMR in *S. suis*

The widespread use of antibiotics has contributed to the emergence and dissemination of AMR in *S. suis* strains. Large-scale genomic and phenotypic studies have shown that countries where antimicrobials were used as growth promoters for a longer period tend to have strains with higher MICs (Minimum Inhibitory Concentration), reflecting elevated resistance levels [7,8]. The correlation between extended antibiotic

use and resistance is further supported by studies on isolates from Spain, Denmark, and Sweden, where *S. suis* has shown notable resistance to macrolides such as erythromycin and tylosin, as well as to tetracyclines; the two most widely used antibiotic classes in pig production in these countries [9,10]. Resistance to β-lactams, the primary treatment for *S. suis* infections in both pigs and humans, remains generally low but has begun to emerge in certain regions, particularly in Asia and parts of Europe, where increases in MICs for penicillin, ampicillin, and ceftiofur have been reported [9–11]. Furthermore, in regions with extensive antibiotic use, a gradual increase in MICs over time, known as the MIC creep phenomenon, has been observed. Although antibiotic treatments remain effective, this gradual rise in resistance may eventually compromise their efficacy, highlighting the urgent need for alternative strategies [8].

Monitoring and predicting resistance in *S. suis* is challenging due to its genetic diversity, as different strains can employ distinct mechanisms against the same antibiotic, underscoring the importance of identifying reliable candidate markers that capture this diversity [8]. For example, macrolide resistance is frequently mediated by *erm(B)* or *mef(A/E)* genes, tetracycline resistance by *tet(O)* or *tet(M)*, and aminoglycoside resistance by genes such as *aph(3')-IIIa*; many of these determinants are located on integrative and conjugative elements (ICEs) or plasmids, promoting horizontal gene transfer between strains [37,38]. For example, serotype 2 strains tend to exhibit lower resistance to macrolides but higher resistance to tetracyclines compared to serotypes 7 and 9 [7,9,10]. Of note, because serotype 2 is the most prevalent in many regions, its lower macrolide resistance may contribute to the overall lower percentages of resistance reported for this antibiotic class [7,10].

2. Novel Antimicrobial Strategies

2. 1 The rise and decline of conventional antibiotics

The first antibiotic, salvarsan, was discovered in 1910, followed by sulfonamides and later penicillin. This marked the beginning of the golden age of antibiotics, which peaked in the 1950s with the discovery of numerous compounds, many of which were natural products produced by Actinomycetes [39]. These antibiotics, shaped by millions of years of microbial competition, were identified through natural product screening, and their mechanisms of action were often elucidated in subsequent studies. Over time, however, this natural reservoir has been largely exhausted [39]. In the past 20 years, the widespread use of antibiotics in medicine and livestock production has driven the emergence and spread of AMR strains, reducing the effectiveness of first-line treatments, including macrolides, tetracyclines, and

beta-lactams, and prompting the WHO to classify antimicrobial resistance as one of the top 10 global public health threats of the 21st century [1,40].

2.2 New approaches to antibiotic discovery

To counteract the antibiotic resistance crisis, a common strategy is the search for new natural compounds. However, this field has already been extensively explored, and discovering truly novel molecules has become increasingly difficult [40]. Alternatively, rational design focuses on developing synthetic molecules that target bacterial components to inhibit conditionally essential genes (CEG) or critical virulence factors[41].

A major challenge in rational antibiotic discovery is the lack of well-characterized targets. Designing compounds that effectively inhibit pathogens requires a deep understanding of bacterial biology [40]. Compared to other therapeutic areas, antibiotics have very few validated molecular targets, with most existing drugs focusing on cell wall biosynthesis, nucleic acid synthesis, or the protein synthesis machinery [39,42].

Progress for the identification of new targets has been limited by several challenges; i) although essential genes represent attractive candidates for novel antimicrobials, many have eukaryotic homologues that make them unsuitable for selective inhibition (ii) differences in bacterial growth between laboratory conditions and the host environment, and (iii) the complexity of regulatory networks, including sensory systems and epigenetic methylation pathways, which often exhibit high levels of interconnectivity and epistatic interactions [43–45].

Membrane and cell wall integrity are critical for bacterial survival during infection, and the CPS contributes to this bacterial protection but must be carefully regulated. While a robust capsule enhances resistance to antimicrobial peptides and other immune defenses, its downregulation at mucosal surfaces exposes adhesins that facilitate epithelial colonization and biofilm formation [30,31,33]. This balance underscores the importance of dynamic cell envelope regulation in pathogenesis.

This regulatory adaptability is rarely captured in standard gene essentiality screens, which are typically performed in nutrient-rich media and often underestimate the number of essential genes. In *E. coli*, initial studies identified 303 essential genes, but under nutrient-limited and stress conditions mimicking the host, an additional 258 essential genes were revealed [44]. Similarly, high-throughput approaches like transposon sequencing have demonstrated that disrupting single genes in pathways such as vitamin, nucleotide or amino acid biosynthesis is non-lethal in ideal laboratory

media but can be fatal under host-like stress conditions [46,47]. Even genes like *purA*, dispensable in rich media, have been reported among the few active during semi-dormancy in *S. aureus* [48], emphasizing how environmental context reshapes our perception of essentiality.

The inherent complexity of bacterial biochemical pathways presents further challenges. For instance, targeting biosynthetic pathways requires careful consideration of the point of disruption in relation to the desired effects and phenotype [44]. In *S. aureus*, deleting early enzymes in the wall teichoic acid (WTA) pathway, such as TarO, is not lethal but reduces virulence and increases sensitivity to β-lactam antibiotics. In contrast, deleting late enzymes like TarG is lethal because WTA intermediates accumulate and sequester undecaprenyl phosphate, a lipid carrier essential for peptidoglycan synthesis. However, this lethality can be by-passed by upstream loss-of-function mutations in early WTA genes [49,50]. These examples highlight the importance of considering not only individual gene functions but also their position within metabolic networks when evaluating drug targets. A similar interconnection exists in sensory systems, including two-component sensory networks, and epigenetic modifications such as DNA methylation, which rarely operate in linear isolated pathways but instead form highly dynamic and overlapping regulatory circuits [51–53].

3. Sensory systems

3.1 Canonical two-component systems

Sensory systems are common across bacterial species and play a central role in detecting and responding to diverse environmental stimuli, including changes in nutrient availability, pH, antimicrobial compounds, and physical stress [54]. In *S. suis*, several two-component systems have been implicated in stress tolerance and virulence, but the roles of many predicted sensory systems in adaptation to host environments and antimicrobial pressure remain poorly understood [55,56].

The most common sensory systems are two-component systems, which consist of a sensor histidine kinase (HK) and a cognate response regulator (RR). Histidine kinases typically span the membrane and contain two conserved domains: the DHp domain, which carries the conserved histidine residue that undergoes phosphorylation, and the catalytic domain, which supplies ATP for autophosphorylation. Upon sensing an external signal, the kinase autophosphorylates and transfers the phosphoryl group to a conserved aspartate on the response regulator. This modification induces structural changes that activate the regulator, commonly by exposing a DNA-binding

domain or promoting multimerization, ultimately leading to transcriptional activation or repression of target genes [57–59].

3.2 Other types of sensory systems

Another category of bacterial signaling pathways is three-component sensory systems. These systems resemble canonical two-component systems but include an additional auxiliary protein, often acting as the primary environmental sensor [60]. While HKs and RRs typically share conserved domain architecture and exhibit significant homology across species and with other signaling systems, the auxiliary sensory components are far more diverse [61,62]. This variability reflects their specialized roles in detecting distinct stimuli, and they generally lack enzymatic activity. Instead, auxiliary sensors regulate the kinase through direct protein-protein interactions. Upon sensing a specific stimulus, the auxiliary protein can either activate or inhibit the HK, thereby modulating downstream gene expression [60] .

A well-characterized example is the LiaFSR system, widely conserved among Firmicutes and strongly associated with membrane stress response (the name lia derives from lipid II-interfering antibiotics) [63]. Orthologues of LiaFSR are also present in *S. suis*, although their specific regulon and role in pathogenesis remain poorly defined. Under normal conditions, LiaF, which is the auxiliary transmembrane protein, inhibits LiaS (HK). Upon membrane stress, such as exposure to cell wall-targeting antibiotics. LiaF releases LiaS, which phosphorylates LiaR (RR), enabling to dimerize and activate genes involved in maintaining the cell envelope [64–67]. Some studies also describe a fourth component, LiaX, which is thought to modulate the system from the cell surface by interacting with antimicrobial peptides and releasing LiaF [59].

Due to the role of LiaFSR in protecting against antimicrobial peptides, several studies in *Streptococcus* Group A and B have investigated its contribution to virulence. In Group B *Streptococcus* (GBS), deletion of the response regulator *liaR* led to severely attenuated virulence in both sepsis and pneumonia mouse models [67]. In contrast, in Group A *Streptococcus* (GAS), *liaR* deletion increased virulence, while constitutive activation of the system, such as through *liaF* deletion, reduced virulence in an intramuscular infection model [68]. These findings illustrate how the impact of LiaFSR activity on pathogenesis can vary depending on the infection route and tissue niche being studied. As seen with CPS expression, what promotes virulence is not simply turning the system on or off, but the ability to tightly regulate its activity, allowing the bacterium to adapt dynamically to different host environments.

Other regulatory systems exist outside of the canonical phosphorylation cascades. In some cases, response regulators are activated through ligand-induced conformational

changes. A well-characterized example is the family of extracytoplasmic function sigma factors (ECFs), that typically respond to environmental and envelope stress in bacteria [69,70]. ECF sigma factors are regulated by membrane-anchored anti-sigma factors that undergo a conformational change in response to a specific signal, leading to the release and activation of the sigma factor. This mechanism has been extensively characterized for σ^E-RseA in *Escherichia coli* (ECF02 group) and σ^W-RsiW in *Bacillus subtilis* [71,72]. While ECF systems are common stress regulators in Gram-positive bacteria, any putative presence in *S. suis* remains speculative.

A recently described class of LytTR regulatory systems (LRS), first identified in *Streptococcus mutans*, has been predicted by in silico analyses to be broadly conserved among Gram-positive bacteria, although they remain largely uncharacterized in most species. [73] The two best-characterized examples, *hdrMR* and *brsMR*, function analogously to ECF systems. Each operon encodes a membrane-bound protein (HdrM or BrsM) and its cognate RR (HdrR or BrsR). Under non-inducing conditions, the membrane protein retains the response regulator in an inactive state, preventing gene activation. Upon an as-yet unidentified stimulus, the regulator is released and activates its own operon in a positive feedback loop, increasing cytoplasmic RR levels and promoting the expression of downstream target genes [74]. Notably, these two systems also exhibit cross-regulation, with each RR capable of activating the other operon. LRS systems have been implicated in key processes such as biofilm formation, natural competence, and bacteriocin production[75,76]. Genomic analysis of *S. suis* reveals the presence of two putative LytTR regulatory systems, neither of which have been experimentally characterized in this species to date. Their environmental stimuli, regulons, and roles in adaptation or virulence are directly addressed in this thesis.

4. DNA Methylation Systems: Phase Variation and Epigenetic Regulation

Phase variation is a reversible genetic mechanism that enables bacteria to generate phenotypic diversity within a clonal population [77]. This process often involves inverted repeats or homologous recombination, leading to rearrangement or switching of specific DNA sequences. As a result, genes can encode alternative protein variants or be turned ON and OFF, depending on the orientation or integrity of regulatory or coding regions. This mechanism helps promote genetic and phenotypic heterogeneity, increasing the chances that at least a subset of the population will be suited to survive under changing or stressful environmental conditions, thereby enhancing the adaptive potential of the bacteria [77,78].

DNA methylation systems, which are often components of bacterial restriction-modification (RM) systems, have emerged as a notable class of phase-variable genes playing roles in in both defense and regulation. These systems typically consist of DNA methyltransferases that modify specific nucleotide motifs, most commonly by adding a methyl group to adenine or cytosine residues (e.g., N^6-methyladenine, 5-methylcytosine[79]. Initially, DNA methylation was thought to function primarily as a protective mechanism against foreign genetic elements such as bacteriophages, by marking self-DNA and enabling the restriction of unmodified exogenous DNA. However, it has become increasingly clear that methylation can also serve as a form of epigenetic regulation, influencing transcription by modulating DNA accessibility, transcription factor binding, or promoter activity [77,80]. When methylation patterns vary across a bacterial population in a coordinated, heritable manner due to phase-variable methyltransferase expression, the result is a “phasevarion” (phase-variable regulon), a concept first described in *Haemophilus influenzae* and now recognized in multiple bacterial pathogens [53].

Bacterial DNA RM systems are classified into several types (Type I, II, III), each differing in subunit composition and mechanism of action. Phase variation has been identified in both Type I and Type III RM systems, with the latter often mediated by simple sequence repeats (SSRs) and the former by site-specific recombination between target recognition domains (TRDs) [81,82]. Here, we focus on Type I RM systems [80]. Type-I RM systems are composed of three subunits: a restriction endonuclease (HsdR), a methyltransferase (HsdM), and a specificity subunit (HsdS). The HsdS protein defines the target recognition sequence through two interchangeable target recognition domains (TRDs) [83,84].

A well-characterized example is the SpnD39III system in *Streptococcus pneumoniae*, which switches between six different *hsdS* alleles (SpnD39IIIA–F) through the action of the site-specific recombinase CreX. Each allele leads to a unique methylation profile and regulates distinct sets of genes involved in nutrient uptake, capsule biosynthesis, and stress response [81]. These shifts in gene expression have been shown to influence phenotypes such as capsule thickness, serum resistance, and virulence in infection models. A similar mechanism has been described in *Neisseria gonorrhoeae*, where recombination within the *hsdS* gene generates two different forms, a truncated and a full-length variant, resulting in distinct methylation patterns, although the downstream consequences on gene expression remain unknown [85].

In this thesis, we focus on two Type I methylation systems in *S. suis*, both composed of the canonical three subunits: HsdR, HsdM and HsdS. The first system (MS1) corresponds to a phase-variable MS containing two *hsdS* copies, allowing

recombination between TRDs and generating four distinct phase variants [86]. In contrast, the second system (MS2) does not undergo phase variation, but the *hsdS* gene exists in two different allelic forms: a truncated version, containing a single TRD, and a full-length version with two TRDs [87]. To our knowledge, the functional roles of these Type I systems in *S. suis* physiology, epigenetic regulation, and virulence have not been experimentally characterized prior to the work presented in this thesis. In order to obtain as much as possible unbiased insights into the roles of these Type I RM systems, we employed genome-wide approaches, including mutation screens and transcriptomics where possible.

5. Genome-Wide Screening Strategies

As mentioned above, many bacterial genes are traditionally studied under rich laboratory conditions, which do not reflect the complexity of the host environment. As a result, genes that appear dispensable *in vitro* may in fact be essential during infection. This limitation is further compounded in less-studied species, where a large fraction of genes remain uncharacterized. To systematically uncover gene functions and identify CEGs, especially those important for survival or virulence in specific environments, there is a need for high-throughput, genome-wide screening strategies [88,89].

5.1 Transposon libraries

One widely used approach is the generation of transposon mutant libraries combined with high-throughput sequencing, commonly referred to as transposon-insertion sequencing (TIS) methods. Several techniques fall under this category, including transposon insertion sequencing (Tn-seq) and transposon directed insertion-site sequencing (TraDIS) [89]. These methods rely on the random insertion of a transposon, typically Tn5 or *mariner* transposon, into the bacterial genome using a transposase enzyme. A transposon generally consists of an antibiotic resistance cassette flanked by inverted repeat sequences that facilitate insertion [90]. This process generates a highly diverse pool of mutants, each harbouring a transposon insertion at a different genomic location. The mutant library can then be subjected to any condition of interest, such as exposure to sub-lethal concentrations of antibiotics or host-related stress factors. After selection, the recovered mutant pool is processed and sequenced to determine the genomic locations of transposon insertions. Sequencing reads are mapped to the reference genome to determine the location and abundance of each insertion. By comparing insertion patterns between control and test conditions, genes that are depleted under selective pressure can be identified as conditionally essential, while enriched insertions may indicate a fitness advantage. Various statistical tools are available to support this analysis, accounting for insertion density, gene length, and biological variability [89,91,92].

5.2 Comparative genomics

Transposon libraries are not the only tool for identifying genes of interest. Advances in sequencing technologies, along with the decreasing cost and increasing speed of genome sequencing, have made it possible to analyze large collections of bacterial strains, including both pathogenic and non-pathogenic lineages [93]. These datasets can be assembled and annotated using automated bioinformatics pipelines that predict coding sequences and functional elements, offering powerful complementary strategies for functional gene discovery [94].

Comparative analyses across these genomes allow for the identification of conserved genes, including those of unknown function. Tools that cluster genes into orthologous groups enable more refined evolutionary comparisons, making it possible to detect functionally equivalent genes even when they have diverged significantly at the amino acid sequence level, something that simpler sequence alignments may miss. Essential genes often evolve more slowly, and their high level of conservation across diverse strains can serve as an indicator of their functional importance. Similarly, genes that are consistently found and more highly conserved in disease-associated strains, compared to non-pathogenic ones, may play roles in pathogenesis or virulence. Their conservation suggests they contribute to bacterial fitness in the host environment, making them compelling targets for further investigation [95,96].

6. Integrating Oxford Nanopore Technology into Functional Genomics Workflows

As seen in the case of comparative genomics, technological innovation is essential for scientific progress. However, progress depends not only on the development of new technologies, but also on their effective integration into existing methodologies. A good example of this is the incorporation of long-read Oxford Nanopore Technologies (ONT) into functional genomics [97–99].

ONT emerged as a breakthrough sequencing platform due to its ability to generate long sequencing reads, which offer major advantages over traditional short-read sequencing, particularly in resolving repetitive regions and assembling complete genomes [97–99]. Initial concerns about basecalling accuracy limited its early applications, but recent advances in sequencing chemistry and software have addressed these issues, expanding its utility across high-resolution genomic analyses [100,101]. Importantly, ONT is also relatively low-cost, portable, and operable at the bench, making it accessible to laboratories with limited infrastructure [100,102,103].

Yet, the value of such technologies goes beyond initial implementation. It is also important to understand and adapt them to specific research needs. Modifying protocols to reduce cost, increase flexibility, or simplify workflows allows scientists to tailor these tools to their research context. This also reduces reliance on third-party reagents or commercial products with hidden formulations, which are often more costly. Such efforts help broaden access to advanced molecular techniques and reduce dependence on external sequencing providers. Ultimately, these adaptations make high-throughput sequencing technologies more accessible and affordable.

7. Aims and outline of this thesis

The rise of AMR bacterial strains is a growing global health concern. As conventional antibiotics continue to lose efficacy, there is an urgent need to identify novel antimicrobial targets for the development of new therapies. Improving our understanding of the mechanisms underlying adaptation and pathogenesis in *S. suis*, an emerging zoonotic pathogen, is therefore increasingly important. This thesis aims to characterize genetic determinants, ranging from essential metabolic pathways to regulatory and conserved systems, that contribute to *S. suis* pathogenesis and may represent novel targets for therapeutic intervention.

The first aim of this thesis was to perform a genome-wide screen of *S. suis* genes under host-mimicking conditions. In **Chapter II**, we constructed a high-density transposon mutant library and screened it in active porcine serum (APS) and a cerebrospinal fluid (CSF) mimic. To enable the sequencing of the mutant pool entirely in-house, we implemented ONT for transposon insertion profiling. This Tn-seq approach allowed us to identify conditionally essential metabolic pathways specific to each environment, as well as candidate drug targets. Among the hits, we found the LiaFSR regulatory system to be conditionally essential in APS and highlighted several hypothetical proteins of potential interest.

To follow up on the findings from Chapter II, **Chapter III** focused on characterizing hypothetical proteins that appeared important for growth in APS. This study revealed two previously uncharacterized LRS systems in *S. suis*, both of which were found to engage in cross-regulation with LiaFSR, another sensory system identified in Chapter II. Together, these three systems form a regulatory network that shapes key virulence-associated traits, including competence, antimicrobial resistance, and biofilm formation. Their involvement in multiple host-adaptation processes also highlights them as promising candidate drug targets.

Comparative genomics of a large collection of *S. suis* strains identified two conserved DNA methylation systems in *S. suis* among pathogenic lineages: MS1, a phase-variable Type I RM system that is particularly conserved in pathogenic strains, and MS2, which exists in two forms (full-length and truncated). **Chapter IV** aimed to explore the regulatory roles of both systems, focusing on the epigenetic effects of the different MS1 and MS2 forms. Using a panel of 18 engineered mutant strains, we performed transcriptomic and phenotypic analyses that revealed how specific MS1–MS2 combinations modulate central metabolic pathways, including galactose metabolism and the incomplete TCA cycle, and influence growth dynamics in both rich and defined media. Notably, this work uncovered for the first time an epistatic interaction between the two methylation systems, demonstrating that DNA methylation not only regulates gene expression and bacterial fitness, but may also involve cross-system coordination regulation.

Chapter V focused on validating the implementation of ONT for transposon library sequencing. Given that 16S rRNA-based taxonomic classification is highly sensitive to sequencing errors, it served as an ideal benchmark for assessing ONT's accuracy. We compared short-read ONT sequencing (V3–V4 region) to the gold-standard Illumina platform and showed that, with current chemistry and basecalling, ONT performs comparably in terms of taxonomic resolution and diversity profiles, reinforcing the reliability of the ONT-based Tn-seq results described in Chapter II. In addition, we developed a simplified, cost-effective PCR-based barcoding protocol that eliminates the need for ligation, reducing both hands-on time and reagent costs. These advances contribute to the broader democratization of sequencing by making high-throughput methods more accessible, affordable, and adaptable to individual research needs.

Finally, **Chapter VI** summarizes the main findings and outcomes of this thesis, highlighting how they contribute to a broader understanding of *S. suis* biology and pathogenesis. It discusses the implications of newly identified regulatory systems, CEGs, and epigenetic modulators. Finally, chapter VI outlines future directions for both basic and applied research, emphasizing how integrative approaches combining functional genomics, comparative analysis, and methodological innovation can help advance the search for novel antimicrobials.

Chapter 2

Genome-Wide Identification of Conditionally Essential Genes Supporting *Streptococcus suis* Growth in Serum and Cerebrospinal Fluid

Maria Juanpere-Borras[1], Tiantong Zhao[1], Jos Boekhorst[1], Blanca Fernandez-Ciruelos[1], Rajrita Sanyal[1], Nissa Arifa[1, 2], Troy Wagenaar[1], Peter van Baarlen[1] and Jerry M. Wells[1].

[1] Host-Microbe Interactomics Group, Wageningen University and Research, Wageningen, Netherlands.
[2] Research Center for Genetic Engineering, National Research and Innovation Agency (BRIN), Bogor, Indonesia.

Abstract

Streptococcus suis is a major cause of sepsis and meningitis in pigs, and zoonosis through the emergence of disease-associated lineages. The ability of *S. suis* to adapt and survive in host environments, such as blood and cerebrospinal fluid (CSF), is important for pathogenesis. Here, we used Transposon insertion sequencing (Tn-Seq) coupled with Nanopore sequencing to identify conditionally essential genes for optimal growth of *S. suis* P1/7 in active porcine serum (APS) and CSF derived from choroid plexus organoids [104]. Through comparative fitness analyses we identified 33 conditionally essential genes (CEGs) supporting growth in APS and 25 CEGs in CSF, highlighting the importance of pathways involved in amino acid transport, nucleotide metabolism, and cell envelope integrity. Notably, the LiaFSR regulatory system and multiple ABC transporters were important for proliferation. We also identified several genes of unknown function as conditionally essential for growth, pointing to previously unrecognized genetic factors involved in *S. suis* adaptation during infection. These findings provide new insights into the genetic requirements for *S. suis* survival in APS and CSF and a deeper understanding of its ability to adapt to distinct physiological niches.

Introduction

S. suis is an encapsulated Gram-positive coccus that causes infections of the porcine respiratory tract as well as severe invasive infections such as septicemia, arthritis and meningitis [5,6]. Diseases caused by *S. suis* result in significant economic losses in the pork industry due to mortality, antimicrobial treatments and the use of autogenous vaccines [105]. *S. suis* is also a public health concern for persons working with diseased pigs or consuming raw pork products due to the high zoonotic potential of specific lineages [106–108]. Highest incidence of *S. suis* zoonotic disease occurs in Southeast Asia, with Vietnam having the highest number of cases [109]. Currently, 29 virulence-associated serotypes of *S. suis* have been reported, with serotype 2 being the most common cause of invasive disease in swine and human infections [106,107].

Asymptomatic carriage of *S. suis* in the upper respiratory, genital, and intestinal tracts of pigs is common[14]. *S. suis* exhibits relatively low invasiveness toward epithelial cells supporting the hypothesis that the primary route of entry is through the palatine tonsils [12,110]. Immunohistochemistry studies have shown that *S. suis* can enter deep into the tonsillar crypts, where the surface epithelium becomes a single cell layer thick, facilitating bacterial uptake and translocation [19]. One hypothesis is that *S. suis* is phagocytosed, but not killed, by specific subsets of tonsillar macrophages, allowing phagocytosed *S. suis* bacteria to replicate and travel through the efferent lymphatics to the bloodstream or directly enter the circulation via the blood vessels in the lymphoid tissue [111,112]. Once in the bloodstream, *S. suis* can cause sepsis, and meningitis upon crossing the blood-brain barrier or the blood-cerebrospinal fluid barrier [6,26]. The precise pathways and molecular mechanisms enabling *S. suis* to infect hosts remain poorly characterized [23]. This underscores the need for further studies to delineate the genetic and cellular factors underpinning *S. suis* pathogenesis [23].

Despite advances in understanding the pathogenesis of *S. suis*, the development of effective vaccines remains challenging. This difficulty arises from the extensive genetic diversity of *S. suis* [5], and the fact that similar virulence functions can be fullfilled by different genes across pathogenic lineages [4,20]. Virulence factors recognised to play important roles in *S. suis* pathogenesis are the cytotoxin suilysin, the capsular polysaccharide, and housekeeping genes enolase [25,113]. Surface-exposed enolase of *S. suis* hijacks the host plasminogen-plasmin proteolytic system to break down the host extracellular matrix and cleave epithelial tight junctions [25,113,114]. Other genes important in pathogenesis include those involved in stress resistance and metal homeostasis in host niches [26,115]. To colonise the host, survive and replicate during host infection, *S. suis* must compete with other microorganisms for scarce nutrients and adapt to changes in pH and oxygen levels [115]. Bacterial

sensory systems play a crucial role in detecting environmental changes, nutrient availability, and rapidly regulating gene expression to support replication, which is crucial for pathogenesis and transmission. Hence, a better knowledge and understanding of the genetics regulating metabolic adaptation of *S. suis* to different host niches might provide deeper insights into *S. suis* pathobiology and the development of new therapeutic strategies [116].

A substantial portion of the *S. suis* genome has hitherto remained uncharacterized, with numerous genes encoding hypothetical proteins with unknown functions, as described in NCBI databases (NC_012925). Consequently, critical genes for infection and pathogenesis may have remained undiscovered. To address this, unbiased genome-wide screening methods are necessary to identify annotated and hypothetical genes necessary for *S. suis* replication and survival in the host.

In this study, we used Tn-seq, a next-generation sequencing technique, to identify genes involved in *S. suis* growth and survival in APS and CSF extracted from the lumen of iPSC-derived choroid plexus (ChP) organoids [104]. CSF within the lumen of 30-40-day-old human ChP organoids closely resembles *in vivo* CSF [117], providing an alternative approach to extracting CSF from animals or humans. Significant differences in metabolite concentrations have been measured between CSF and serum [118]. The latter study, conducted on 58 healthy control individuals, showed that methionine, glutamic acid, and glycine were only detected in a small number of CSF samples but were present in nearly all serum samples. Additionally, inosine was exclusively detected in CSF samples, likely due to its critical roles in purine and energy metabolism and in neuroprotection, where it supports neural repair and reduces oxidative stress [118,119].

The Tn-seq procedure was recently applied in *S. suis* S10, to identify genes involved in pathogenesis using an *in vivo* model, where bacteria were inoculated intrathecally in pigs, and mutants were recovered from different body fluids [120]. In this study, a Tn-library was generated in *S. suis* P1/7, a zoonotic serotype 2 strain [23,115] which was cultured under both control and test conditions. By sequencing the flanking regions of transposon insertions and comparing their frequencies between conditions, we identified insertions that either enhanced or impaired bacterial fitness [121]. For cost-effective, rapid and accurate in-house sequencing of transposon insertions, we utilized ONT. To validate the functional role of specific genes, we performed bioassays using chemically defined media (CDM), APS and CSF to assay proliferation capacities of *S. suis* gene-specific deletion mutants.

Our high-throughput screening results revealed the critical role of specific metabolic pathways for *S. suis* P1/7 growth in APS and CSF. Nucleotide metabolism was essential for survival and proliferation in APS, while amino acid uptake was critical for proliferation in CSF. We also characterized a previously uncharacterized nucleotide ABC transporter in *S. suis*, essential for purine uptake. Furthermore, our analysis discovered that the LiaFSR three component system and several other uncharacterized genes potentially involved in *S. suis* pathogenesis. Notably, we demonstrated, for the first time, the feasibility of using CSF extracted from ChP organoids as a model system.

Materials and Methods

Bacterial strains and growth conditions

S. suis strain P1/7 (ATCC, BAA-853) was cultured in Todd-Hewitt medium (Thermo Fisher Scientific™, CM0189) supplemented with 0.2% of yeast extract (Thermo Fisher Scientific™, 212750)(THY) at 37 °C with 5% CO_2. High transformation-efficiency *Escherichia coli* strain Top10 (Thermo Fisher Scientific™, C404010) was cultured in LB medium (Merck, 1,102,850,500) at 37 °C with vigorous shaking. Chloramphenicol (Sigma-Aldrich, C0378-25G) was added to the media at a final concentration of 5 µg/mL for *S. suis* and 20 µg/mL for *E. coli*.

Growth measurements

Growth of *S. suis* wild-type (WT) and deletion mutants in THY was measured by absorbance (OD_{600}) every hour for 8 hours. Growth measurements in APS and CSF were performed hourly by making serial dilutions in PBS and plating on agar plates to obtain CFU/mL. For all strains, overnight cultures were first pelleted at 6000 rpm for 5 minutes, and pellets were resuspended in PBS. The appropriate volume of bacterial culture was inoculated into APS or CSF to achieve an OD_{600} of 0.015. For THY growth measurements, bacterial overnight cultures were directly inoculated without prior pelleting and resuspension in PBS.

Growth curves in CDM

The complete list of basic components and stock solutions used to prepare CDM is provided in supplementary materials Table S1. For the preparation of 2 mLof CDM, the following constituents were used: 450 µL of CDM buffer, 0.3 mg/mL amino acids, 50 mM of glucose, 0.01 mg/mL of pyruvate, 20 µL of vitamins, 20 µL of metal mixture, 4 µL of manganese, and 2 µL of choline chloride. Nucleobases were added at concentrations ranging from 5 to 100 mg/L. Subsequently, 200 µL of CDM was added to each well of a 96-well plate along with 2 µL of THY overnight cultures.

OD_{600} readings were obtained using a SpectraMax® M3 Multi-Mode Microplate Reader (Avantor, Radnor, PA, USA) at 37°C.

Transposon vector cloning

The materials and protocols for constructing the transposon library were provided by Tim Van Opijnen's lab (Tufts University School of Medicine, Boston, Massachusetts) and modified as described below [123]. *Magellan6* encodes spectinomycin resistance between two inverted repeat sequences [123]. Given that *S. suis* has intrinsic resistance to spectinomycin, we engineered the *Magellan6* plasmid such that the Himar1 *mariner* transposon carried a chloramphenicol resistance gene under the control of the P32 promoter. The plasmid containing the Himar1 *mariner* transposon, *Magellan6*, was purified using the Qiagen Miniprep kit (QIAprep Spin Miniprep Kit, 27106), following the manufacturer's recommendations. Linearization of the plasmid was performed using the restriction enzymes FastDigest *Sma*I and *Swa*I (Thermo Fisher Scientific™, 10324630 and 15390291). A DNA fragment containing the P32 promoter[124] and chloramphenicol acetyltransferase resistance gene, was amplified by PCR using primers P001 and P002 (Table S2), and the high-fidelity Q5 polymerase (NEB, M0493S), using plasmid pLABTarget as a template [125]. The DNA fragments were ligated together using Hifi Assembly (NEB, E2621S), to generate plasmid Meg6SS that now encoded Himar1 *mariner* transposon with the desired chloramphenicol resistance gene.

Generation of Tn-library for *S. suis* P1/7

Genomic DNA (gDNA) and plasmid DNA solutions were concentrated using a SpeedVac (Thermo Fisher Scientific™, RVT5105) until they reached a concentration of at least 350 ng/µL. A mixture of 36 µL of purified C9T transposase combined with 4 µg of gDNA and 4 µg of plasmid and incubated at 30 ^{o}C for 6 hours. DNA fragments containing the transposon were transformed into *S. suis* using the natural competence method described by Zaccaria et al. [126]. Briefly, 10 µL aliquots of treated DNA were combined with 5 µL of competence inducing peptide ComS and 100 µL of *S. suis* THY culture at OD_{600} values between 0.035 and 0.05. After 2 hours incubation, aliquots were pooled to a final volume of approximately 900 µL. A 1/10 dilution of 10 µL of the mixture was plated on THY chloramphenicol agar plates for CFU counting and the remaining 900 µl was centrifuged at 6000 rpm for 5 min. Then, 800 µL of supernatant was removed and the pellet resuspended in the remaining 100 µL before plating on THY agar plates containing 5 µg/mL chloramphenicol. This process was repeated four times, resulting in 4 libraries with colony counts of approximately 10K, 1.5K, 3.3K, and <1K, respectively.

Generation of choroid plexus organoids and extraction of cerebrospinal fluid

Six-well plates (Corning, 07-200-83) were pre-coated with Vitronectin (STEMCELL, 07180) in CellAdhere™ Dilution Buffer (STEMCELL, 07183) at room temperature for 1 hour. Human iPSC line EDi002-A (EBiSC™) was maintained on vitronectin-coated 6-well plates in mTeSR1 (STEMCELL, 85857). Medium was changed daily, and cells were passaged once per week. ChP organoids were generated using the STEMdiff™ Choroid Plexus Organoid Differentiation Kit (STEMCELL, 100-0824) and the STEMdiff™ Choroid Plexus Organoid Maturation Kit (STEMCELL, 100-0825), following previous protocols[117] with minor modifications. Briefly, iPSC were dissociated into single-cell suspensions using Accutase (STEMCELL, 07920). On day 1, 1 × 10^5 cells were seeded into a well of Corning® 96-well round-bottom ultra-low attachment microplate (Corning, 7007) in 100 µL of embryoid body (EB) Formation Medium and 10 µM Y-27632 (ROCK inhibitor; STEMCELL, 72302). On day 2 and day 4, fresh 100 µL of EB Formation Medium was added to each well. EBs with diameter ranging between 400 and 600 µm were typically observed on day 5 at which time EB Formation Medium was replaced with 200 µL/well of Induction Medium. On day 7, each EB was embedded in 15 µL of Matrigel® (Corning, 734-1101) dropwise on sheets of parafilm and incubated at 37 ˚C for 30 min to polymerize Matrigel® (16 EB per sheet of parafilm). The sheets of parafilm were each positioned above one well of a 6-well ultra-low adherent plate (STEMCELL, 100-0083) using sterile forceps. All 16 Matrigel® droplets were gently washed off sheets and put into one well using 3 mLof Expansion Medium. Each 6-well plate was shaken back and forth three times to ensure even distribution of EB and incubated at 37 ˚C for 3 days. On day 10 the Expansion Medium was carefully replaced with 3 mL/well of Choroid Plexus Differentiation Medium, and plates were placed on a platform rotator (Fisherbrand™, 15504080) in the incubator. On day 13, Choroid Plexus Differentiation Medium was refreshed. From day 15, Choroid Plexus Differentiation Medium was replaced with 3 mL/well Maturation Medium and renewed every 3 days. By day 30, ChP organoids epithelia resemble cyst-like structures filled with CSF-like fluid. ChP organoids between day 30 to day 40 were used to harvest CSF from the organoid lumen using a 0.30 × 12 mm BL/LB needle attached to a 1 mL syringe.

Transposon library screening

For screening in APS, 60 µL of *S. suis* Tn-library were inoculated into 10 mLpre-warmed THY broth or APS (Thermo Fisher Scientific™, 26250084), with two biological replicates for each condition and incubated for 5 hours at 37 °C with 5% CO_2. For CSF screening, 250 µL of Tn-library stock was thawed in 10 mL of THY medium and cultured for 3.5 hours until it reached a concentration of 5 × 10^8 CFU/mL, and then centrifuged and resuspended in 10 mLof PBS. Then 100 µL of the enriched Tn-library

in PBS (Fisher Scientific, 10769033) was inoculated in 10 mLof THY and 10 mLof CSF in triplicates and cultured at 37 °C with 5% CO_2 until cultures reached an OD_{600} of 0.6. Cultures were centrifuged at 4000 rpm for 10 min, the supernatant discarded, and the bacterial pellets stored at -20 degrees until further use. DNA was extracted from bacterial pellets using DNeasy PowerSoil kit (Qiagen, 47016) following the manufacturer's recommendations.

Sample preparation

After library growth under the different conditions (APS, CSF and THY), bacterial DNA was isolated and processed as described by Van Opijnen & Camilli [123]. Briefly, approximately 2 µg of genomic DNA was digested with the *MmeI* restriction enzyme (NEB, R0637L) and dephosphorylated using Quick CIP (calf intestinal alkaline phosphatase) (NEB, M0525S) to avoid re-ligation. The oligonucleotide primers incorporating the adapter sequences (P003 and P004, Table S2), were annealed by mixing equimolar concentrations of each primer in water in 1.5 mLEppendorf tubes, heating the tubes to 96°C for 3 minutes, and gradually cooling the tubes to room temperature. Phosphorylated adapters were ligated to the overhangs generated by digestion of bacterial DNA with *MmeI*, using T4 DNA ligase (NEB, M0202S). After adapters ligation, transposon-containing fragments of bacterial DNA were amplified by PCR. The forward primer annealed to the inverted repeat sequence that is present at each end of the transposon, while the reverse primer annealed to ligated adapter (P005 and P006, Table S2). The PCR reaction included 2 µL of the ligation mix and High-fidelity Q5 polymerase (NEB, M0493S) for 35 cycles. Total volumes of PCR reactions were loaded onto 0.8% agarose gel, and bands corresponding to 300 bp that were expected to contain transposon-enriched fragments were excised and purified using Invisorb fragment cleanup kit (Invitek molecular, 1020300300) (Figure 1).

Library sequencing with Nanopore

PCR amplicons were prepared and barcoded for sequencing following the manufacturer's instructions using the native barcoding kit (ONT, SQK-NBD112.24), the ligation sequencing kit (ONT, SQK-LSK112) kit and flow cells (ONT, FLO-MIN106D) on MinION Mk1c and Mk1b (ONT, Oxford, United Kingdom) sequencing devices; device outputs were handled using MinKNOW software (version 23.07.15). Sequencing ran until all reads were sequenced (approximately 24 hours). The resulting FAST5 files, containing raw nanopore data of each sample, were converted to FASTQ files via basecalling (Guppy version. 7.1.4).

Analysis of Tn-seq data

For nanopore reads sequence analysis (Figure 1), we filtered reads to include those within the expected size range of our PCR amplicon, roughly 150 to 180 base pairs

(bp). Subsequently, we conducted a two-step screening process for the 20 bp of transposon sequence upstream of the transposon insertion site so that only reads containing the forward Tn sequence (ACTTATCATCCAACCTGTTA) or the reverse Tn sequence (TAACAGGTTGGATGATAAGT) were retained for further analysis. For the filtered reads, only the 14 bp immediately adjacent to the respective forward or reverse Tn sequences were selected for alignment to the *S. suis* P1/7 reference genome using Bowtie2. Only reads with 100% identity to a single gene fragment were retained; reads that mapped to more than one gene in the genome were discarded. The number of reads per TA site was quantified, and the data was compiled into a Wiggle (WIG) file. TA sites with zero insertions (no matches) were also included in the WIG file during harmonization to ensure the data includes all TA sites per gene, for calculation and comparison during the fold-change analysis. The harmonized WIG files from each replicate were assessed for library quality using the Transit "tnseq_stats" module (version 3.2.7). This analysis provided metrics such as library saturation at TA sites, total mapped reads, and insertion count distributions. For CEGs analysis, the harmonized wig files from each replicate of both the control and test conditions were used as input into Transit (version 3.2.7), using the Transit "resampling test" module with default parameters. Genes that exhibited a log fold change >1 and achieved statistical significance (adj. p-value < 0.05) were compiled into a final list (Table 1). The set of filtered genes and encoded proteins were annotated by retrieving the corresponding gene descriptions and biological process annotations from NCBI and KEGG databases, respectively.

Gene deletion mutants

In frame deletion mutants were constructed using CRISPR/Cas9 based technology essentially as described by Gussak et al [127]. Briefly, mutants were generated by transforming *S. suis* P1/7 with a gRNA-Cas9 co-expression plasmid (pSStarget) and a linear DNA repair template. Upon transformation, Cas9 endonuclease was expressed and directed by the co-expressed guide RNA to specific target gene, where Cas9 induced a double-stranded cut that was repaired by the *S. suis* homologous recombination machinery using the introduced DNA repair template. Resulting colonies were screened for gene-specific deletion mutants and site-specific mutant strains were cured of the expression plasmids.

Three different 20 bp guide RNAs were designed for each target gene. Single-stranded oligonucleotides were designed using Benchling software, and synthesized by IDT Technologies. Each primer included a 4 bp overhang compatible with the overhangs of the linearized pSStarget plasmid (P007 to P048, Table S2). Equimolar concentrations of complementary primers for each guide were mixed with annealing buffer (10 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA) and annealed in a thermocycler

(Bio-Rad, Hercules, CA, USA) (5 min at 95°C followed by gradually cooling to 25°C at 1 °C/min).

The empty pSStarget plasmid was linearized using the BsaI enzyme (NEB, R3733S) and purified with the Invisorb Fragment CleanUp kit (Invitek, 1020300300). The annealed guide and digested plasmid were ligated using T4 DNA ligase (NEB, M0202S) for 1 hour at room temperature and subsequently incubated overnight at 4 °C. The ligation mixture was transformed into chemically competent *E. coli* Top10 cells and plated on LB agar plates containing 10 µg/mL of chloramphenicol, followed by overnight incubation at 37 °C. Colonies were screened for correct plasmid constructs using primers P077 and P078 (Table S2), and plasmids were extracted from positive colonies using the Qiagen Miniprep kit (QIAprep Spin Miniprep Kit, 27106).

The repair template was constructed by amplifying approximately 1000 bp upstream and downstream of the target gene using primers (P049 to P076, Table S2). These primers were designed to include approximately the first and last 30 bp of the target gene, ensuring that the entire gene was not deleted to avoid potential polar effects on downstream genes. PCR products were subsequently ligated through Splicing by Overlap Extension (SOE) PCR using external primers, HA1_fwd_X and HA2_rev_X (Table S1). After each PCR step, the PCR products were purified using the Invisorb Fragment CleanUp kit (Invitek, 1020300300).

The gRNA-Cas9 co-expression plasmids and repair templates were transformed into *S. suis* using the natural competence method described by Zaccaria et al[126]. Briefly, 250 µL of an overnight culture of *S. suis* P1/7 WT were inoculated into 10 mLof THY broth and grown until the OD_{600} reached between 0.035 and 0.058. A 100 µL aliquot of *S. suis* P1/7 culture was combined with 5 µL of competence-inducing peptide ComS, 200-500 ng of plasmid DNA, and 1 µg of repair template. The mixture was incubated at 37 °C with 5% CO_2 for 2 hours and plated on THY agar plates containing 5 µg/mL of chloramphenicol. Colonies with the correct gene deletion were identified using PCR and external primers HA1_fwd_X and HA2_rev_X (Table S2), and these colonies were cured from plasmids by performing two consecutive overnight passages in THY medium without chloramphenicol.

In silico amino acid sequence alignments and protein structure predictions

Amino acid sequences were obtained from the *S. suis* P1/7 genome annotation in NCBI (nucleotide ID NC_012925) for SSU_RS04755 and from Uniprot for PnrA (A0A0H2UPF3) and TmpC (P29724). Multiple sequence alignment of the amino acid

sequences of SSU_RS04755, PnrA and TmpC were performed using the Clustal Omega algorithm integrated within Jalview (version 2.11.3.3) software, using default parameters. Structure prediction of SSU_RS04755 was performed using the SWISS-MODEL server (https://swissmodel.expasy.org). Structure alignment of the predicted protein encoded by SSU_RS04755 from *S. suis* and PnrA was performed using PyMOL (version 3.0.3).

Identification of Promoter Regions with Putative LiaR Binding Boxes

Identification of promoter regions potentially containing a LiaR of *Bacillus subtilis* as input[128] binding box was performed using the FIMO tool from the MEME suite (https://meme-suite.org/meme/tools/fimo)[129]. The resulting list of genes was manually curated to retain only sequences with p-value < 0.05 and located within 200 bp upstream of a gene. This analysis identified 48 genes with a putative LiaR binding box in their promoter regions (Table S3).

Analysis of Differential Gene Expression by qPCR

S. suis P1/7 and *S. suis* Δ*liaR* mutant were grown to exponential phase (approximately OD_{600} 0.3) in THY broth at 37 °C with 0.5% CO_2. A 10 mLaliquot was pelleted by centrifugation at 4000 rpm for 10 minutes, the supernatant was discarded, and the pellet snap-frozen in liquid nitrogen before being stored at -80 °C overnight. RNA was extracted from the pelleted cells using the RNeasy Mini Kit (Qiagen, 74104) following the manufacturer's instructions with specific modifications. Pellets were resuspended in 700 µL of RLT buffer containing 0.1% β-mercaptoethanol and transferred to lysing matrix B 2 mLtubes (MP Biomedicals, 6911100). Bacterial cells were lysed using a FastPrep-24™ 5G bead beating grinder and lysis system (MP Biomedicals, Solon, OH, USA) with settings; 4.0 m/sec, All-MetalQuickprep adapter, 40 seconds. The tubes were centrifuged for 1 minute at 10,000 rpm, and the supernatant was transferred to a clean Eppendorf tube. Subsequent steps were carried out using the manufacturer's protocol. Final RNA concentrations were measured using the Qubit RNA Broad Range Kit (Thermo Fisher Scientific™, Q10211) and a Qubit 4 fluorometer (Thermo Fisher Scientific™, Waltham, MA, USA). The Quantitect Reverse Transcriptase Kit (Qiagen, 205311) was used for DNA deletion and cDNA synthesis, with 500 ng of RNA as input. For differential gene expression analysis, 96-well white PCR plates (Bio-Rad, ML9651) and GoTaq qPCR Master Mix (Promega, A6002) were used in a CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA, USA). Primers for each gene are listed in Table S1 (from P079 to P086). Gene expression was calculated using the 2-ΔΔCt method relative to reference gene *gyrA*. Independent experiments were performed in triplicate with three biological replicates each.

Results

Generation of *in vitro* Tn-libraries in *S. suis* P1/7

The transposon library was constructed using a modification of the Tn-Seq protocol optimized for *Streptococcus pneumoniae* described by Tim van Opijnen and colleagues [121], and adapted for *S. suis* P1/7 (see Methods). Using this optimized protocol, we obtained four libraries that were combined to obtain a single library. We evaluated library quality using the control libraries generated in THY medium, which served as the reference condition for subsequent essentiality comparisons. Across five replicates, saturation ranged from 44% to 67%, with total mapped read counts between 1.2 and 2.8 million. The distribution of insertion counts was highly skewed with heavy-tailed profiles, consistent with expectations for Himar1 libraries [130]. In total, 91% of coding sequences contained at least one insertion, indicating broad representation of the genome. These results indicate that the THY control libraries provided sufficient coverage to support reliable identification of conditionally essential genes.

Development of a nanopore sequencing protocol for sequencing of the Tn-libraries

After culturing the Tn-library in test (APS or CSF) and control (laboratory culture medium THY) media, DNA was extracted and processed as outlined in the Methods section to obtain a PCR amplicon suitable for sequencing. We opted for nanopore amplicon sequencing to have an in-house method for amplicon sequencing and analysis that could be optimized when appropriate. As it was the first time nanopore technology had been used for this purpose, we designed an in-house pipeline for processing and analyzing the sequenced nanopore reads (Figure 1), which is available at (https://github.com/MariaJuanpereBorras/Nanopore-TnSeq-Pipeline). For each sample, i.e. each Tn-library aliquot cultured in test or control media, nanopore sequencing generated between 0.7 and 5.4 million reads, which were filtered to remove fragments larger or smaller than the expected length of the amplicon (around 180 bp), resulting in an average discard rate of 5%. In the subsequent step, reads were selected for the presence of the transposon sequence (underlined region of P005, Table S2); 64% of the total reads were retained during this phase. After this step, reads were trimmed to retain only the 14 base pairs immediately adjacent to the transposon end sequence. When these 14 bp inserts were mapped to the *S. suis* P1/7 genome (NCBI accession no. NC_012925), 55% of the *de novo* sequenced reads had a perfect match and 51% had a unique perfect match; the remaining 4% were discarded because they did not map to a unique genomic location. After running the harmonized wig files in Transit (version 3.2.7) [131] a final list detailing the increased or decreased log2 fold change (logFC) for each gene in the test condition compared

to the control was generated. Filtering the data on negative logFC (< -1) and adjusted p-value (< 0,05), we obtained a final list of 33 conditionally essential genes (CEGs) for *S. suis* P1/7 grown in APS and 25 CEGs in CSF, both compared to growth in THY control medium (Table 1).

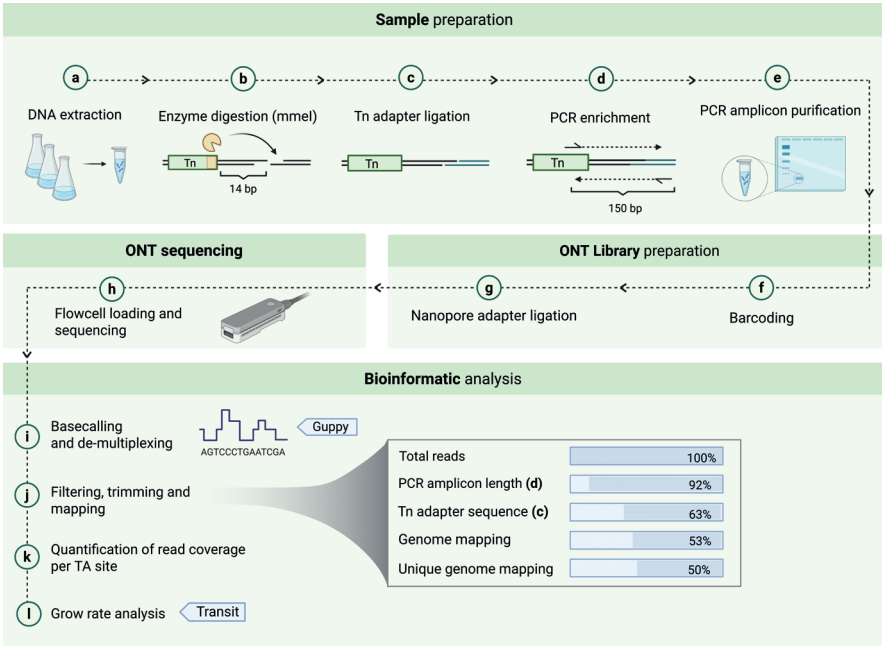


Figure 1. Workflow for Tn-Seq sample processing, sequencing, and bioinformatic analysis. (a-e) Sample preparation begins with DNA extraction, followed by enzyme digestion with MmeI, Tn adapter ligation, PCR enrichment, and amplicon purification. (f-h) ONT Library sequencing. (i-l) in-silico processing of sequencing output and data analysis. (Steps j and k were specifically developed for this study).

Table 1. List of conditionally essential genes, identified by Tn-seq, that were conditionally essential for *S. suis* optimal growth in APS and CSF compared to THY (adj. p-vale < 0.05; Log_2FC < -1).

Genes identified by tn-seq in APS					
Locus tag	**Name**	**Product**	**Main biological process**	**Operons**	**Log_2FC**
SSU_RS08850	-	adenylosuccinate synthase	Purine metabolism		**-3,19**
SSU_RS00280	*purB*	adenylosuccinate lyase	Purine metabolism		**-2,84**
SSU_RS06600	-	ABC transporter permease	Transporter	□	**-2,28**
SSU_RS02595	*ppc*	phosphoenolpyruvate carboxylase	Carbon metabolism		**-1,86**
SSU_RS05280	*ptsP*	phosphoenolpyruvate--protein phosphotransferase	Transporter		**-1,85**
SSU_RS06605	*trpX*	tryptophan ABC transporter substrate-binding protein	Transporter	□	**-1,84**
SSU_RS04185	*guaA*	glutamine-hydrolyzing GMP synthase	Purine metabolism		**-1,77**
SSU_RS06595	-	ABC transporter ATP-binding protein	Transporter	□	**-1,73**
SSU_RS06820	-	bifunctional oligoribonuclease/PAP phosphatase NrnA	Sulfur metabolism		**-1,65**
SSU_RS09155	-	hypothetical protein	NI		**-1,65**
SSU_RS09655	-	LacI family DNA-binding transcriptional regulator	Transcription factor		**-1,58**
SSU_RS06385	-	acyl-ACP thioesterase domain-containing protein	NI		**-1,56**
SSU_RS06740	-	hypothetical protein	NI		**-1,53**
SSU_RS01575	-	Fur family transcriptional regulator	Transcriptional factor		**-1,43**
SSU_RS04755	-	BMP family protein	Transporter	•	**-1,34**
SSU_RS07155	-	YtxH domain-containing protein	NI		**-1,34**
SSU_RS05100	-	mannonate dehydratase	Carbohydrate metabolism		**-1,29**
SSU_RS02715	-	amino acid ABC transporter substrate-binding protein	Transporter	◊	**-1,26**
SSU_RS02080	*liaF*	cell wall-active antibiotics response protein LiaF	Signal transduction		**-1,21**
SSU_RS04655	*pta*	phosphate acetyltransferase	Carbon metabolism		**-1,18**
SSU_RS07195	-	hypothetical protein	NI		**-1,18**

SSU_RS06140	*tpx*	thiol peroxidase	Oxidoreductases		**-1,17**
SSU_RS09860	*guaB*	IMP dehydrogenase	Purine metabolism		**-1,17**
SSU_RS04750	-	ABC transporter ATP-binding protein	Transporter	•	**-1,15**
SSU_RS04745	-	ABC transporter permease	Transporter	•	**-1,14**
SSU_RS02750	-	DegV family protein	NI		**-1,11**
SSU_RS05325	-	LacI family DNA-binding transcriptional regulator	Transcriptional factor		**-1,11**
SSU_RS06410	-	arsenate reductase	Oxidoreductases		**-1,11**
SSU_RS06645	-	lactonase family protein	Carbon metabolism		**-1,1**
SSU_RS05710		glucosaminidase domain-containing protein			**-1,05**
SSU_RS04420	*pyrE*	orotate phosphoribosyltransferase	Pyrimidine metabolism		**-1,02**
SSU_RS04510		ABC transporter substrate-binding protein/permease	Transporter	◊	**-1,02**
SSU_RS06155	*tehB*	SAM-dependent methyltransferase TehB	Transferases		**-1,01**

Genes identidied by tn-seq in CSF

Locus tag	Name	Product	Main biological process	Operons	Log_2FC
SSU_RS02635	-	hypothetical protein	NI		**-3.22**
SSU_RS06605	*trpX*	tryptophan ABC transporter substrate-binding protein	Transporter	□	**-2.97**
SSU_RS06600	-	ABC transporter permease	Transporter	□	**-2.88**
SSU_RS01560	-	hypothetical protein	NI		**-2.78**
SSU_RS06595	-	ABC transporter ATP-binding protein	Transporter	□	**-2.63**
SSU_RS00910	pgk	Phosphoglycerate kinase	Carbohydrate metabolism		**-2.61**
SSU_RS02500	-	hypothetical protein	NI		**-2.61**
SSU_RS07925	glyQ	Glycine--tRNA ligase alpha subunit	Translation		**-2.55**
SSU_RS04515	artM	Arginine transport ATP-binding protein ArtM	Transporter		**-2.54**
SSU_RS06355	mscL	Large-conductance mechanosensitive channel	Signaling and cellular processes		**-2.41**
SSU_RS03420	srlR	Glucitol operon repressor	Transcriptional factor		**-2.17**

Table 1. Continued.

Genes identidied by tn-seq in CSF					
Locus tag	**Name**	**Product**	**Main biological process**	**Operons**	**Log_2FC**
SSU_RS04835	*pstB3_1*	Phosphate import ATP-binding protein PstB	Transporter		**-2.13**
SSU_RS08105	*fabH*	3-oxoacyl-[acyl-carrier-protein] synthase 3	Lipid metabolism		**-1.85**
SSU_RS07155	-	hypothetical protein	NI		**-1.43**
SSU_RS06420	*rpiA*	Ribose-5-phosphate isomerase A	Carbohydrate metabolism		**-1.42**
SSU_RS04600	*lacR_2*	Lactose phosphotransferase system repressor	Transcriptional factor		**-1.38**
SSU_RS02705	*yecS_2*	L-cystine transport system permease protein YecS	Transporter	◊	**-1.3**
SSU_RS08470	*ilvH*	Putative acetolactate synthase small subunit	Carbohydrate metabolism		**-1.3**
SSU_RS02715	*fliY*	L-cystine-binding protein FliY	Transporter	◊	**-1.14**
SSU_RS03250	-	Phosphoglycerate dehydrogenase	Energy/Amino acid metabolism		**-1.1**
SSU_RS02710	*glnQ_2*	Glutamine transport ATP-binding protein GlnQ	Transporter	◊	**-1.08**
SSU_RS07955	*metQ*	putative D-methionine-binding lipoprotein MetQ	Transporter		**-1.03**
SSU_RS07855	-	hypothetical protein	NI		**-1.02**
SSU_RS03205	-	hypothetical protein	NI		**-1.01**
SSU_RS05450	-	hypothetical protein	NI		**-1**

Genes are ordered from lowest to highest FC; genes belonging to the same operon are marked with the same symbol in the 'Operons' column. Non identified genes are defined as NI.

Genes identified by Tn-seq exhibiting fitness impact in *S. suis* growth in APS compared to THY

The transposon library was cultivated for approximately four hours in THY or APS and Tn-seq was used to identify CEG that support growth of *S. suis* in APS. Out of the 33 CEG (Table 1), 7 (21%) did not have annotations in NCBI or the KEGG database. Among the annotated genes, 10 (30%) were predicted to be involved in metabolism, 5 (15%) in nucleotide metabolism, and 8 genes (24%) in substrate transport. Thus 'metabolism' and 'substrate transport' were the most abundant annotations of CEGs linked to growth of *S. suis* in APS (Figure 2). Furthermore, 3 genes were annotated as transcription factors, 2 as oxidoreductases, 1 as participating in signal transduction, and 1 as a methyltransferase.

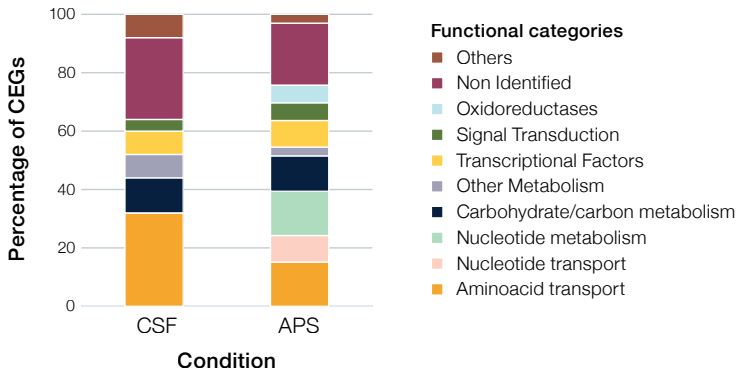


Figure 2. Stacked bar chart showing the percentage of CEGs (logFC < –1, adjusted p-value < 0.05) classified into functional categories for CSF and APS samples.

Of the 10 genes annotated as substrate transporters, 3 genes were predicted to belong to a tryptophan ABC transporter, 3 genes to an amino acid ABC transporter, and 1 gene (gene symbol *ptsP*) encoding a component of a sugar transporter. The remaining 3 genes were part of a single operon with unknown function. In this operon, a predicted substrate-binding lipoprotein (SSU_RS04755) is highly conserved across *S. suis* strains, including strains from serotype 2 and 9 [132]. This protein is also part of the *S. suis* secretome and has been studied as a potential vaccine candidate targeting *S. suis* [132,133]. An amino acid sequence homology search using NCBI blastp tool revealed a high identity (> 60%) with ABC transporters described as nucleoside transporters from *Streptococcus pneumoniae* (SPD_0739) and *Streptococcus agalactiae* (GBS0942) [134,135].

Genes identified by Tn-seq contributing to *S. suis* growth in CSF compared to THY

To explore whether nucleotide metabolism and transport pathways also play a critical role in *S. suis* growth in CSF we performed a Tn-seq screen comparing growth in CSF and THY. In contrast to APS, no genes involved in nucleotide metabolism or transport were conditionally essential for *S. suis* growth in CSF. Among the CEGs supporting growth in CSF, 8 genes (32%) were annotated with roles in amino acid transport, 7 genes (28%) as hypothetical proteins, 3 genes in carbohydrate metabolism, and 2 genes as transcriptional regulators (Table 1, Figure 2).

We found 8 CEGs that supported growth in both CSF and APS. The three genes that are predicted to form a tryptophan ABC transporter were identified as conditionally essential in both media (Table 2). Additionally, SSU_RS07195, annotated as hypothetical protein, and SSU_RS02715, annotated as cystine-binding lipoprotein, were also identified as conditionally essential for growth in CSF and APS. Transposon insertions in *liaF*, a membrane protein, and the hypothetical protein SSU_RS07195, were associated with increased proliferation during growth in CSF compared to THY, whereas insertions in the same genes (*liaF* and SSU_RS07195), were associated with reduced proliferation in APS compared to THY. Lastly, transposon insertions in the phosphate importer *pstB3_1* were associated with increased proliferation in APS but reduced proliferation in CSF.

Validating Tn-Seq results using In-Frame Deletion Mutants

To validate the Tn-seq results, we selected 5 CEGs identified in APS (*purA*, SSU_RS04755, *liaF*, SSU_RS09155, and SSU_RS07155; Table 1) to make in-frame deletions in *S. suis* strain P1/7. Given that genes related to nucleotide metabolism and transport were the most represented among the CEGs (Figure 2), we hypothesized that nucleotide availability was a limiting factor for *S. suis* growth in serum. We chose to generate a *purA* (SSU_RS08850) deletion mutant strain because (i) *purA* was annotated as playing a role in nucleotide metabolism, (ii) showed the highest fold change difference (-3.19) in the list of CEGs, and (iii) was reported as a CEG for *S. suis* survival in pigs in previous Tn-seq studies [120]. SSU_RS04755 was selected because it resides within a predicted nucleoside ABC transporter operon. Additionally, we selected the predicted *liaF* (SSU_RS02080) gene, annotated as participating in signal transduction, because *liaF* showed differential requirement for optimal growth in APS and CSF (see above). This suggests a differential role for *liaF* in growth under these two culture conditions. In several Gram-positive bacteria, *liaF* is part of a three-component system, along with a histidine kinase (*liaS*) and a transcription regulator (*liaR*)[136]. *LiaFSR* is involved in virulence and antibiotic resistance in streptococci [137,138], but it has not been studied in *S. suis*. We were unable to obtain a viable *liaF*

Table 2. Shared genes identified by Tn-seq analysis of S. suis P1/7 in APS and CSF media (adj. p-vale < 0.05; $Log_2FC < -1$; $Log_2FC > 1$).

Common genes identified by Tn-seq in APS and CSF					
Locus tag	**Name**	**Product**	**Operons**	**Log_2FC APS**	**Log_2FC CSF**
SSU_RS06600	-	ABC transporter permease	□	**-2.28**	**-2.88**
SSU_RS06605	*trpX*	tryptophan ABC transporter substrate-binding protein	□	**-1.84**	**-2.97**
SSU_RS06595	-	ABC transporter ATP-binding protein	□	**-1.73**	**-2.63**
SSU_RS07155	-	Hypothetical protein		**-1.34**	**-1.43**
SSU_RS02715	*fliY*	L-cystine-binding protein FliY		**-1.26**	**-1.14**
SSU_RS07195	-	Hypothetical protein		**-1.18**	**1.56**
SSU_RS02080	*liaF*	cell wall-active antibiotics response protein LiaF		**-1.21**	**1.3**
SSU_RS04835	*pstB3_1*	Phosphate import ATP-binding protein PstB 3		**2.17**	**-2.13**

Genes belonging to the same operon are marked with the same symbol in the 'Operons' column. Genes with decreased fitness (negative Log_2FC) are highlighted in orange, and genes with increased fitness (positive Log_2FC) are highlighted in green.

deletion mutant. Downstream of *liaF* are *liaS* (SSU_RS02085) and *liaR* (SSU_RS02090), which have overlapping reading frames and are controlled by the same promoter. Thus, we hypothesized that transposon insertions in *liaF* would disrupt the transcription of *liaS* and *liaR* and therefore generated gene deletion mutants for *liaS* and *liaR* as a proxy for *liaF* gene deletion.

From the group of unknown genes of CEGs for growth in APS, we selected SSU_RS09155, which had highest fold change (Table 1) and SSU_RS07155, which is predicted to encode a protein with unknown function and demonstrated a negative fold change in both APS and CSF. From the CEGs list in CSF, we also selected SSU_RS02635, encoding a protein with unknown function, and a fold change of -3,22 in CSF.

Growth rates of specific gene-deletion mutants in THY and APS were assessed by OD_{600} measurements and CFU counting each hour during an 8-hour period (Figure 3). In THY medium, the mutant strains had the same growth rate as the wild type across all phases of the growth curve, except for the *liaR* mutant, which had a one-$\log_{10}$ reduced CFU count in the stationary phase (Figure 3). In APS, the Δ*purA* mutant exhibited growth attenuation compared to the WT, reaching a maximum concentration of approximately 10^7 CFUs/mL, whereas the WT reached a concentration of nearly 10^9 CFUs/mL. This growth defect correlated with the Tn-seq data, where *purA* had a $\log_2$FC difference of -3.19, indicating that functional PurA contributes to growth of *S. suis* P1/7 in APS. Growth assays in APS showed that Δ*liaR,* Δ*liaS* and ΔSSU_RS09155 gene deletion-mutants exhibited a growth delay during exponential phase and reached stationary phase with lower bacterial cell densities compared to the wild type, confirming the APS Tn-seq results (Figure 3). The *liaR* mutant did not exhibit any difference in growth in CSF compared to the WT parental strain, whereas in CSF, the Tn-seq results revealed a positive logFC for *liaR.* The ΔSSU_RS04755 mutant exhibited a delayed growth rate compared to the WT during the first 3 hours of growth in APS, although the mutant ultimately reached the same bacterial densities at stationary phase as WT. Mutants ΔSSU_RS07155 and ΔSSU_RS02635 did not show significantly altered growth curves in any of the tested conditions.

LiaR Activates Transcription of the Hypothetical Protein SSU_RS07195

Deletion of the gene encoding transcriptional regulator LiaR led to lower growth in APS. We hypothesized that LiaR may regulate a gene(s) critical for growth in APS. If our hypothesis were correct, we expected that some of the genes regulated by LiaR might feature in our Tn-seq results. To explore this, we used FIMO (Find Individual Motif Occurrences, part of the MEME online software suite, see Methods) to search for potential LiaR binding sites across the *S. suis* genome (Table S3). FIMO predicted

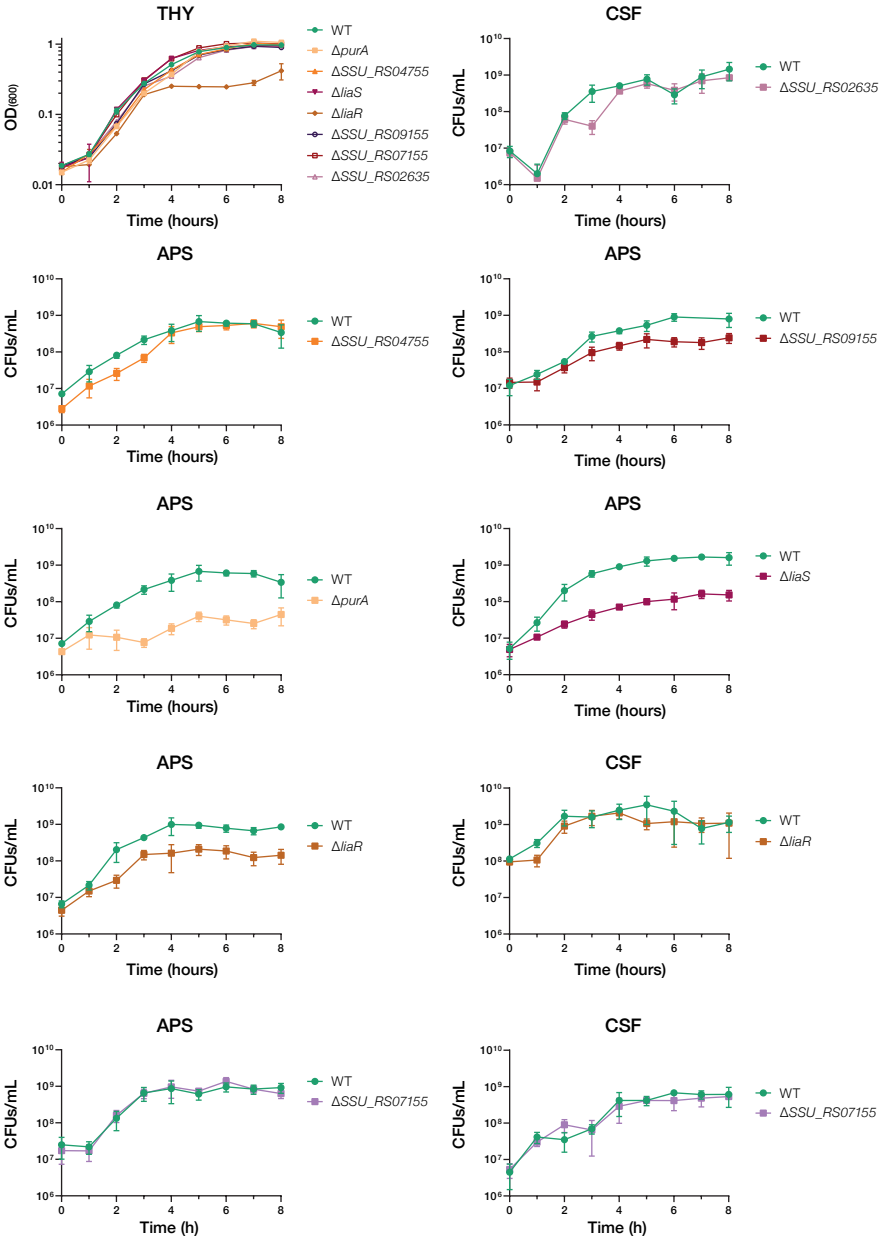


Figure 3. Growth curves in THY, APS and CSF. Data shown are means ± standard deviation of at least two biological replicates and two technical replicates for CFUs counting. CFUs/mL and OD600 axis are in Log10.

the presence of a putative LiaR binding site in the promoter of one of the genes encoding a hypothetical protein, SSU_RS07195, one of the genes identified as CEG in the APS Tn-seq results. To investigate possible functions of the hypothetical protein encoded by SSU_RS07195, we examined genomic position and neighbourhood of SSU_RS07195 in *S. suis* P1/7 in the NCBI database, and found the gene to be partially overlapping with gene SSU_RS07190, possibly part of a single operon together with a third gene SSU_RS10200 encoding an IS630 family transposase (Figure 4). The second gene SSU_RS07190 was predicted to encode a hypothetical protein containing a phage-shock protein (PspC) domain (Figure 4A). Both genes appear to be controlled by the same promoter, as the downstream gene lacks an independent promoter region, and the 5' end of SSU_RS07190 overlaps with the end of SSU_RS07195. Both *liaF* and SSU_RS07195 are CEGs supporting growth of *S. suis* P1/7 in aps with a negative $\log_2$FC (-1.21 and -1.18, respectively; Table 1). To validate these predictions, we performed qPCR to determine whether expression of SSU_RS07195 was altered in the Δ*liaR* gene deletion mutant using *liaF* and *spx* genes, as controls because they are known to be regulated by LiaR in other species [68,136], as references for comparison. Indeed, qPCR results showed that the expression of SSU_RS07195 was significantly reduced in *liaR* gene deletion mutant compared to WT, suggesting that LiaR positively regulates expression of SSU_RS07195 (Fig. 4). *liaF* and *spx* also showed reduced expression in *liaR* gene deletion mutant. Notably, the amount of downregulation of SSU_RS07195, -4.5 $\log_2$FC (Figure 4), was similar in magnitude to that observed for *liaF* and *spx*, -4.6 and -3.09 $\log_2$FC respectively (Figure 4).

In silico analysis of *S. suis* SSU_RS04755

Gene SSU_RS04755 has been predicted to encode a basic membrane family protein (BMP), a transmembrane component of specific ABC transporters. Growth curve experiments did not reveal a significant reduction in the growth rate of our ΔSSU_RS04755 mutant. The operon genes with SSU_RS04755 were predicted to encode an ATP-binding protein and a permease which are typical components of an ABC transporter. Both predicted ATP-binding protein and permease encoding genes appeared as CEGs in our Tn-seq results (Table 1). The protein sequence associated with SSU_RS04755 shares significant identity with lipoproteins that are components of nucleoside ABC transporters in Gram-positive bacteria (see above), suggesting that SSU_RS04755 and its operon genes may encode an ABC transporter complex involved in nucleoside transport. To investigate possible functional equivalence of SSU_RS04755 operon genes with ABC transporters, we performed amino acid sequence alignments of SSU_RS04755 with PnrA and TmpC, which are homologous nucleoside binding lipoproteins from *S. pneumoniae* and *Treponema pallidum* respectively [139,140]. Using the crystal structure of PnrA binding to adenosine [139], we

confirmed that all but one (T70) of the specific amino acids involved in adenosine binding by PnrA were conserved in SSU_RS04755 (Figure 5C). To verify if the structural positions of these amino acids were conserved, we performed a structural prediction (see methods) of the predicted *S. suis* lipoprotein SSU_RS04755 and aligned it to the structure of PnrA (ID 6y9U in Protein Data Bank (PDB)). We confirmed that all amino acids involved in the adenosine binding are located at the same position in both protein structures. The non-conserved amino acid T70 which is I49 in *S. suis* might bind adenosine, as it has a carboxyl group (COOH) in the same position as the corresponding residue in PnrA that is involved in adenosine binding (Figure 5A and B).

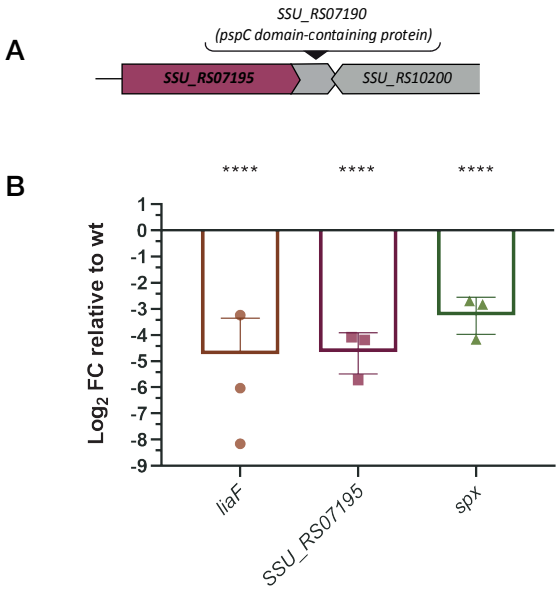


Figure 4. (A) Schematic representation of the SSU_RS07195 operon. (B) Expression of *liaF*, SSU_RS07195, and *spx* in the LiaR mutant relative to WT. Error bars represent standard deviation across biological replicates. Statistical analysis was performed using a one-way ANOVA (****, $p < 0.0001$).

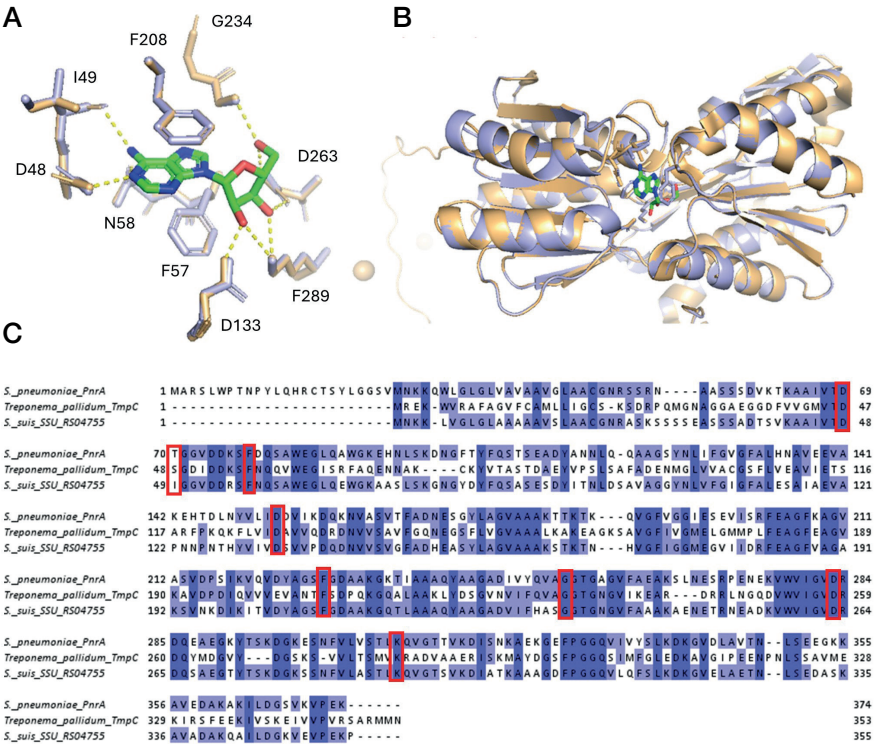


Figure 5. Structural and sequence analysis of protein homologs featuring *S. suis* lipoprotein SSU_RS04755 (blue) and S. pneumoniae PnrA (PDB ID: 6Y9U, Orange). (A) Zoomed-in view of the binding pocket with hydrogen bond interactions (dashed yellow lines). (B) Ribbon representation of the structural alignment of both proteins, highlighting active site interactions with a ligand (green) and key residues (sticks). (C) Multiple sequence alignment of homologous proteins from S. pneumoniae, T. pallidum, and S. suis, with conserved residues highlighted in blue and functionally significant residues marked with red boxes.

Growth in CDM Reveals Nucleobase Uptake Deficiencies of ΔSSU_RS04755 and Growth reduction of Δ*purA* Mutant

To further investigate our hypothesis that SSU_RS04755, SSU_RS04750, and SSU_RS04745 are part of an ABC transporter for nucleosides, we grew the ΔSSU_RS04755 mutant in CDM containing varying concentrations of purines and pyrimidines. When adenine was added to the CDM (Figure 6A), the ΔSSU_RS04755 mutant showed reduced growth compared to the WT, reaching an OD_{600} of approximately 0.4 versus 0.6 for the WT. Interestingly, the growth of both WT and ΔSSU_RS04755 with adenine supplementation was similar to their growth without

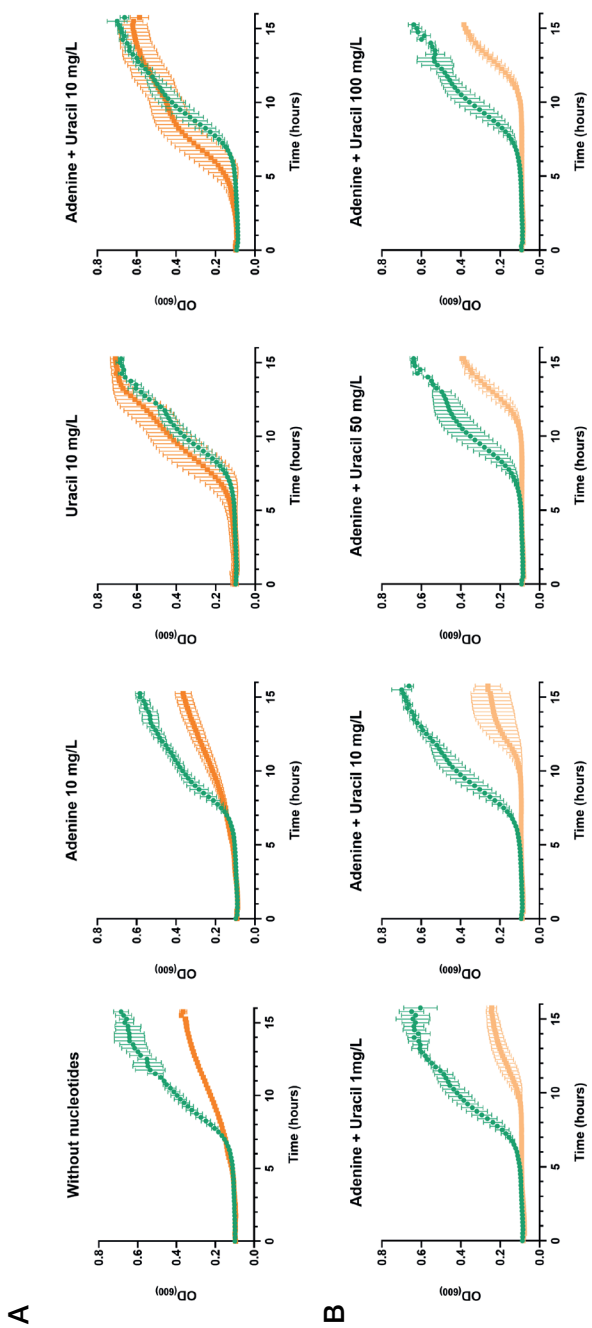


Figure 6. Growth curves of WT (green), *purA* (light orange) and ΔSSU_RS04755 (dark orange) in CDM. (A) WT and ΔSSU_RS04755 growth with different availability of adenine and uracil. (B) WT and Δ*purA* growth with increasing concentrations of adenine and uracil. The experiments were performed with at least 3 biological replicates of each strain. Data shown are means ± standard deviation.

nucleobases, suggesting a defect in the ΔSSU_RS04755 mutant's ability to uptake adenine. We observed that WT grew the same with and without addition of nucleobases, which might suggest that during preculture in THY prior to CDM growth assays, WT bacteria stored nucleobases intracellularly. However, following the same preculturing step in THY, growth of the corresponding ΔSSU_RS04755 deletion mutant in CDM, with adenine as sole source of nucleobase, was substantially lower, in agreement with the hypothesis that SSU_RS04755 encodes a protein that is part of an ABC transporter importing nucleobases.

A second CEG, *purA,* putatively encodes an adenylosuccinate synthase enzyme involved in the conversion of IMP (inosine monophosphate) to AMP (adenosine monophosphate). We hypothesized that the growth deficiency of the *purA* mutant in APS might have been due to its inability to synthesize adenine. We assessed growth of the *purA* mutant in CDM with increasing concentrations of nucleobases to compensate for the lack of de novo synthesized purines in the deletion mutant. Surprisingly, the addition of nucleobases could not recover the WT phenotype: the Δ*purA* deletion mutant exhibited a pronounced decrease in growth rate compared to the WT, reaching a maximum OD_{600} of approximately 0.4, whereas the WT OD_{600} reached up to 0.7. This phenotype was consistent regardless of the concentration of nucleobases added to the CDM (Figure 6B). In conclusion, The ΔSSU_RS04755 mutant showed impaired growth in CDM with adenine as the sole nucleobase, while the Δ*purA* mutant displayed reduced growth that was not restored by nucleobase supplementation.

Discussion

In this study, we constructed a Tn-library for *S. suis* strain P1/7 and employed Tn-seq to screen for CEGs that support growth of *S. suis* bacteria in two host body fluids, APS and CSF. APS was selected to simulate bloodstream conditions, as it supports more reproducible *S. suis* growth than whole blood, which likely restricts growth through neutrophil activation and antimicrobial peptide release [141]. Additionally, we chose to perform Tn-seq in CSF since virulent *S. suis* strains like P1/7 are able to grow in CSF. Using APS and CSF allowed controlled library screening while minimizing stochastic selection of mutants common in *in vivo* models [88]. To our knowledge, this is the first study to apply an in-house Nanopore sequencing approach to Tn-seq in *S. suis.* For downstream analysis, we designed a Tn-Seq data analysis pipeline to analyze nanopore sequencing output and generate a list of CEGs. The quality of the library was high, with saturation of potential TA insertion sites ranging from 44% to 67% and an average coverage of 91% across coding regions, well above the ≥35%

TA-site coverage threshold reported to yield reliable essentiality predictions in *Streptococcus*[142].

The Tn-seq results revealed 33 CEGs important for *S. suis* growth in APS and 25 CEGs in CSF. According to the functional annotations of the crucial genes mediating growth in APS, the main biological processes regulated by these genes were *de novo* biosynthesis of nucleotides and ABC transport systems. Genes conditionally essential for growth in CSF were primarily involved in amino acid transport. Five genes were common to both lists; three of these were part of a tryptophan ABC transporter operon, which has been previously reported as conditionally essential in Tn-seq studies on *S. suis*, corroborating reports that *S. suis* cannot synthesize tryptophan [120,143]. In APS, a substantial number of genes were annotated to be involved in nucleotide metabolism and nucleotide transport. Arenas and colleagues and Dresen and colleagues studied genes supporting *S. suis* growth in blood and CSF using a porcine *in vivo* Tn-library approach, and found that genes involved in *de novo* nucleotide biosynthesis, such as *purA, purB*, and *guaA*, were conditionally essential in both fluids[120,143]. The genes *purA* and *guaB* were also reported as CEGs for growth of Group A *Streptococcus* (GAS) in human blood [144], suggesting that genes related to nucleotide metabolism and transport are important for survival of streptococci in APS. Our Tn-seq screening in CSF did not identify CEGs annotated with functions in nucleotide biosynthesis. However, it revealed an enrichment of 10 genes involved in protein synthesis, 8 of which were annotated with functions in amino acid transport systems. This finding is consistent with reports showing that CSF contains lower amino acid levels than serum [118], which may necessitate the upregulation of amino acid transport systems for growth. Notably, these same studies identified inosine, a purine nucleoside that plays an essential role in purine nucleotide biosynthesis, as one of the metabolites found exclusively in CSF [118]. This may explain the absence of CEGs annotated with functions in nucleotide transport and biosynthesis pathways in the Tn-seq results from CSF, as inosine in CSF could act as a readily available precursor for purine nucleotide synthesis.

From our list of CEG in APS, five genes, i.e. *purA*, SSU_RS04755, SSU_RS04755, *liaF*, SSU_RS09155 and SSU_RS07155 were selected for verification of the library results by generating gene-specific in-frame deletion mutants. From our list of CEG in CSF, genes SSU_RS02635 and SSU_RS07155 with negative FC and *liaR* with positive FC were selected for verification. The choice of these genes was based on: (i) enrichment of certain biological processes according to CEG annotations; (ii) genes identified in Tn-library based studies in different streptococci, i.e. *purA* and *liaF*; (iii) uncharacterized genes in *S. suis* of possible interest, for example, those located in operons with gene of interest; and (iv) presence of a CEG in APS and CSF.

We failed to obtain viable *liaF* deletion mutants, but could obtain deletion mutants of operon genes *liaS* and *liaR*. Deletion mutants of the genes *purA, liaR, liaS,* and SSU_RS09155 had lower growth rates in APS and reached lower CFU/mL counts in stationary phase compared to the WT strain, thus validating the Tn-seq results. The *liaR* deletion mutant showed no growth deficiency in CSF compared to the WT, thus confirming the results obtained for this gene in CSF. Gene deletion mutants for the genes SSU_RS04755, SSU_RS02635, and SSU_RS07155 did not exhibit any significant reduction in growth rate under any of the conditions tested. This could be due to competition among mutants during library growth or a polar effect of the transposon insertion on downstream genes.

Gene SSU_RS04755 encodes a lipoprotein that is part of a nucleotide ABC transporter. Notably, all three genes in the SSU_RS04745–55 operon were identified as CEGs in APS, underscoring the importance of this nucleotide ABC transporter for survival in serum-like conditions, highlighting the importance of this candidate nucleotide ABC transporter to support *S. suis* growth in APS. Our growth experiments in CDM demonstrated that the SSU_RS04755 gene deletion mutant showed significant growth reduction when purines were the sole source of nucleobases, suggesting its essential role in purine uptake. Taken together, these results support the *in silico* predictions that the SSU_RS04745-55 operon encodes an ABC transporter for purine uptake.

In *S. pneumoniae* a nucleoside transporter with 66% homology to the *S. suis* SSU_RS04745-55 candidate purine transporter was shown to be important for growth in serum. Additionally, *S. pneumoniae* Δspd_0739 mutant, which encodes a lipoprotein that is part of a nucleotide transport system, exhibited reduced virulence compared to wild-type[145]. The spd_0739 lipoprotein component of the transporter was proposed as a vaccine candidate due to its high conservation among strains (>98%), its expression both *in vitro* and during *in vivo* infection, and its low homology (<11%) with human lipoproteins [145]. Additionally, in *Borrelia burgdorferi*, two purine uptake transporters have been reported as important for *in vivo* infection in mice, and mutants lacking these purine transporters showed no impaired growth in rich media *in vitro* [146] as we observed in our *in vitro* results. These results highlight the importance of purine transport and acquisition for in-host survival and pathogenesis.

We identified the genes *purA, purB*, *guaA, guaB* and *pyrE* involved in *de novo* nucleotide biosynthesis as CEGs supporting *S. suis* growth in APS. Previous studies had shown that disruption in the purine biosynthesis pathways may reduce bacterial colonization and intracellular growth, increase susceptibility to oxidative stress, and lower bacterial proliferation in human serum and blood [147–150]. Additionally,

genome-wide screens in *Staphylococcus aureus* demonstrated that *pur* and *pyr* gene deletion mutants were attenuated in various infection models[151–154]. The importance of the nucleotide biosynthesis pathway is not surprising, as it is (i) essential for the synthesis of nucleic acids; (ii) mediates production of signalling molecules like cyclic AMP and GMP, and (iii) generates energy storage molecules such as ATP and GTP [48]. Our *S. suis* P1/7 *purA* mutant showed significant growth defects in CDM and was not rescued by the addition of nucleobases suggesting that in limited-nutrient media, nucleobases uptake could not rescue viability-lowering effects of *purA* deletion. It thus appears that nucleobase uptake alone is insufficient to meet the purine demands of the *S. suis* P1/7Δ*purA* mutant. However, while several studies in other bacterial genera such as *Escherichia*, *Salmonella*, and *Bacillus* have demonstrated successful growth restoration of *pur* and *pyr* mutants following the addition of nucleobases to the media [150,155], no studies have been found regarding growth restoration in *Streptococcus*. Given that *de novo* purine biosynthesis involves an extensive list of *pur* genes, it appears that mutations at different points in the pathway may produce distinct phenotypic effects such that compensatory mutations or pathway rerouting is usually not restoring viability nor growth recovery. In studies where growth recovery had turned out possible, deleted genes were typically at the beginning of the pathway, such as *purF* and *purE* [150,155]. In contrast, in *S. suis* P1/7, *purA* is located toward the end of the pathway (KEGG. *Purine metabolism – Pathway ID: ssi00230)* and may therefore result in more pronounced phenotypic changes (Figure 7). The significance of *purA* in pathobiology is further highlighted by a study in clinical *S. aureus*, where it was found that *purA* was induced when *S. aureus* entered into a semi-dormant state in the human body upon exposure to acidic pH or neutrophils as part of the immune response to *S. aureus* infection [156].

Our study revealed the important regulatory role of the LiaFSR three-component system in mediating *S. suis* growth in APS, with a particular focus on the regulatory function of LiaR. Our data suggest that LiaR acts as a positive regulator of at least one CEG involved in supporting *S. suis* growth in APS, highlighting its role in stress adaptation and environmental sensing. Of note, *liaF* and gene SSU_RS07195 encoding a hypothetical protein displayed negative FC values during growth in APS, but positive FC values during growth in CSF, implying that LiaFSR activation is advantageous in APS, while its repression may be beneficial or neutral in CSF. This key role of LiaFSR sensory system to adapt to different host niches and nutrient availability has been described previously by Sanson and colleagues [68] who reported that in *Streptococcus pyogenes*, a mutant not producing LiaR showed reduced expression of virulence genes including *spxA2*, which is involved in oxidative stress response, in an *ex vivo* human blood model. We found that our *S. suis* P1/7 *liaR* gene deletion mutant exhibited reduced growth during culture in APS, while the deletion of

S. suis P1/7 *liaR* did not reduce bacterial growth during culture in CSF; instead, our Tn-seq data suggested that deletion of *liaR* was associated with increased bacterial growth in CSF. These results highlight the metabolic flexibility that disease-associated bacteria can be capable of in order to rapidly alter gene expression to adapt to different host niches. These results also reinforce the notion that some candidate antibacterial targets are only expressed in specific host niches, so that treatments including vaccinations should take such niche-specific antibacterial target expression into account when treating bacterial infections.

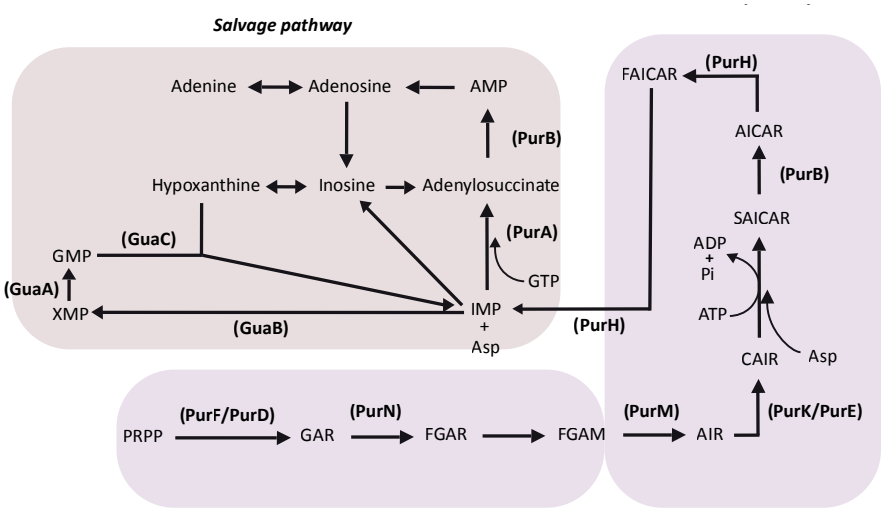


Figure 7. Schematic representation of purine metabolism in *S. suis*, showing the de novo and salvage pathways. Enzymes are indicated by gene names, and arrows denote the direction of metabolic conversions.

Acknowledgments

We thank Tim Van Opijnen's lab (Tufts University) for the transposon library protocol and plasmids, and Alex Gussak (Wageningen University & Research) for the CRISPR-Cas9 protocol for *S. suis* mutants design.

Data Availability

All sequencing data generated in this study (FAST5 and FASTQ files) have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB89170. The raw result files generated by the Tn-Seq analysis, and data supporting all figures in the manuscript are available from the 4TU.ResearchData

repository under DOI: 10.4121/ace99cb5-0414-4a91-9aa8-72c0d8c47381 (https://data.4tu.nl/datasets/ace99cb5-0414-4a91-9aa8-72c0d8c47381/1). All data are available under a Creative Commons CC-BY 4.0

Chapter 3

Cross-Talk Between LiaFSR and Two Novel LytTR Regulatory Systems Modulates Host-Adaptation Traits in *Streptococcus suis* P1/7

Maria Juanpere-Borras[1], Troy Wagenaar[1], Amy Raphaela[1], Peter van Baarlen[1] and Jerry M. Wells[1].

[1] *Host-Microbe Interactomics Group, Wageningen University and Research, Wageningen, Netherlands.*

Abstract

To survive within the host, *Streptococcus suis* must tightly regulate cell envelope integrity and surface structures such as the capsular polysaccharide (CPS), to resist host defensins and complement damage to the cell envelope. Here, we identify and characterize two previously unknown LytTR membrane-associated two component sensory systems in *S. suis*, named HdrMR and BrsMR, which are broadly conserved in Gram-positive bacteria but only partly characterized in *Streptococcus mutans.* Overexpression of the regulators HdrR and BrsR in *S. suis* reduces growth, viability and biofilm formation, but enhances natural competence. The LytTR sensory systems comprise a membrane-associated protein that binds a cognate transcriptional regulator via a cytosolic domain in the cytoplasm. Using CRISPR-Cas9 deletion mutants, transcriptomics, and structural protein modeling, we demonstrated a cross-regulatory mechanism between both LytTR regulatory systems and LiaFSR, a different three component cell envelop-stress sensing regulatory system comprising LiaF a membrane associated sensor, LiaS a histidine-kinase and LiaR a transcriptional regulator. Deletion of *liaR* or *liaS* sensitizes *S. suis* to cell wall-targeting antimicrobial peptides. Unphosphorylated LiaR was shown to repress expression of *hdrMR* and *brsMR* and HdrR downregulates *liaFSR*, indicating bidirectional control. Together, these findings reveal the interconnected role of the three regulatory systems in coordinating host-adaptative traits in *S. suis* and provide new potential targets for therapeutic intervention.

Introduction

S. suis is an encapsulated Gram-positive bacteria recognized as a major swine pathogen and emerging zoonotic agent[6,157,158]. It causes respiratory disease in pigs, as well as arthritis, meningitis, and septicemia[5,6]. *S. suis* infection in humans is predominantly linked to occupational exposure or to consumption of undercooked pork products and typically manifests as meningitis and septicemia[109,115,159]. *S. suis* is also an increasing concern regarding due to its contribution to antimicrobial resistance [8], which limits treatment options and contributes to recurrent outbreaks [159–161]. The absence of effective and cross-protective vaccines makes infection control challenging and underscores the need to identify new targets for vaccine development or other new prevention strategies[162].

One of the best-characterized virulence factors of *S. suis* is its CPS, which helps the bacterium evade phagocytosis by innate immune cells, membrane damage due to antimicrobial peptides and complement pathway-mediated killing [23]. The CPS is essential for bacterial survival in blood, and loss of capsule correlates with markedly increased leukocyte phagocytosis. Nevertheless, unencapsulated strains have been isolated from infective endocarditis, and downregulation of CPS has been shown to facilitate adherence to host cells [23]. This indicates that tight regulation of capsule expression is required for successful adaptation during different stages of infection, as CPS alone is not sufficient to cause disease. CPS expression must be finely balanced with that of other virulence factors, such as pili and adhesins so that the bacteria can evade immune defenses while still exposing the surface structures necessary for host interaction [31,33] .

The CPS of *S. suis* is composed of repeating units of glucose, galactose, N-acetyl-glucosamine, rhamnose, and sialic acid [27,28]. In Gram-positive bacteria, CPS chains are covalently linked to the peptidoglycan layer[29]. Peptidoglycan precursors are assembled in the cytoplasm, transferred to undecaprenyl phosphate to form Lipid I and Lipid II, and then flipped across the membrane as part of the lipid II cycle for incorporation into the cell wall and capsule anchoring [29,163]. Membrane disruption by antimicrobial peptides, antibiotics, or complement can impair lipid carrier recycling and cell envelope integrity, indirectly reducing proper capsule assembly [164,165]. To counteract cell envelope stress, Gram-positive bacteria including streptococci rely on dedicated sensory systems that monitor envelope integrity and activate protective responses [166]. A well-characterized example is the LiaFSR three-component system, originally named for its role in protecting against Lipid II-interacting antibiotics. LiaFSR detects membrane stress and triggers transcriptional changes that enhance resistance to membrane-active agents [63]. The LiaFSR system consists of the sensor

kinase LiaS, the response regulator LiaR, and the membrane-associated inhibitor LiaF. Genetic evidence indicates that LiaF maintains the system inactive under non-stress conditions, most likely through direct interaction with LiaS. Upon membrane damage, LiaF releases LiaS, which phosphorylates LiaR. Phosphorylation of LiaR promotes dimerization, enhancing its DNA-binding activity [64–67]. LiaR upregulates its own promoter, creating a positive feedback loop that amplifies the envelope stress response[68,128,137] .

In a previous Tn-seq study (Chapter II) we showed that LiaR is conditionally essential for *S. suis* optimal growth in activated porcine serum (APS), which retains an active complement system[167]. In that study, a gene annotated as a LiaR homolog, previously uncharacterized in *S. suis*, showed significantly lower growth in APS. Targeted deletion of *liaR*, along with its cognate histidine kinase *liaS*, confirmed their importance for proliferation in APS, as both mutants exhibited markedly reduced growth under these conditions.

In the same Tn-seq study, we also identified a previously uncharacterized gene (SSU1817/SSU_RS09155) encoding a hypothetical transmembrane protein that contributed to growth in APS [167]. Follow-up experiments using a targeted SSU1817 deletion mutant confirmed that loss of this gene reduced bacterial growth in APS. Here we show that SSU1817 is part of a two-gene operon homologous to the *hdrMR* sensory system originally characterized in *S. mutans* [75,76] . In *S. mutans*, the HdrRM system consists of a membrane-anchored protein (HdrM) and a LytTR-family transcriptional regulator (HdrR). Under homeostatic conditions, HdrM sequesters HdrR at the membrane, preventing it from activating transcription. When the HdrM–HdrR interaction is disrupted, either by *hdrM* deletion or by environmental cues, HdrR is released and activates transcription of its own operon, thereby establishing a positive feedback loop (Figure 3C) [74]. In addition to its autoregulatory loop, HdrR induces expression of the adjacent *brsRM* operon, which encodes a membrane protein and response regulator that are structurally analogous to HdrM and HdrR, respectively [168]. While both systems share similar architecture, BrsRM is thought to complement or extend the regulatory functions of HdrRM, suggesting a layered regulatory module that coordinates competence and stress adaptation[76,169]. The environmental signals that trigger or resolve activation of these two LytTR systems remain unknown[169], raising the possibility that they operate in coordination with other regulatory networks to regulate cellular homeostasis.

In this study, we reveal a cross-regulatory mechanism linking three sensory systems; *liaFSR*, *hdrMR*, and *brsMR*, in *S. suis* P1/7. We demonstrate that unphosphorylated LiaR negatively regulates the expression of the *hdrMR* and *brsMR* operons. Notably,

our transcriptomic data also reveal that HdrR overexpression reduces *liaR* expression, indicating that *hdrMR* activity also influences the LiaFSR system. Our results indicate that these three sensory systems must be tightly controlled, as their dysregulation leads to changes in growth, viability, competence, and biofilm formation, which can compromise the adaptation of *S. suis* to different host environments. To further assess the contribution of LiaR and LiaS to membrane stability, we exposed WT and *liaR/liaS* deletion mutants to LL-37, a human antimicrobial peptide that binds the capsule and disrupts membrane integrity[164]. Together, these findings identify a regulatory network that shapes *S. suis* adaptation during host interaction.

Methods

Bacterial strains and growth conditions

S. suis strain P1/7 (ATCC, BAA-853) was cultured in Todd-Hewitt medium (Thermo Fisher Scientific™, CM0189) supplemented with 0,2% (w/v) of yeast extract (Thermo Fisher Scientific™, 212750)(THY) at 37 °C with 5% CO_2. High transformation-efficiency *Escherichia coli* strain Top10 (Thermo Fisher Scientific™, C404010) was cultured in LB medium (Merck, 1,102,850,500) at 37 °C with vigorous shaking. Chloramphenicol (Sigma-Aldrich, C0378-25G) was added to the media at a final concentration of 5 µg/mL for *S. suis* and 20 µg/mL for *E. coli*. CDM (2 mL) was prepared by combining 450 µL CDM buffer with amino acids (0.3 mg/mL final), glucose (50 mM final), pyruvate (0.01 mg/mL final), vitamins (20 µL of stock), metal mixture (20 µL of stock), manganese (4 µL of stock), choline chloride (2 µL of stock), and nucleobases to a concentration of 50 mg/mL. The full list of components, stock concentrations, and sources is provided in Table S1 (Chapter II).

Statistical analysis

Statistical analyses were performed in Python (version 3.8.18) using pandas 2.0.3, numpy 1.24.3, scipy 1.10.1, statsmodels 0.14.0, and scikit-posthocs 0.7.0. For each experimental variable, data were first assessed for normality within each comparison group using the Shapiro–Wilk test and for homogeneity of variances across groups using Levene's test (Brown–Forsythe variant, center = median). If all groups passed normality ($p > 0.05$) and homogeneity ($p > 0.05$), a one-way ANOVA was performed followed by Tukey's honestly significant difference (HSD) test for pairwise comparisons. If either normality or variance homogeneity assumptions were not met, a Kruskal–Wallis test was applied followed by Dunn's post hoc test with Bonferroni adjustment. In all cases, p-values < 0.05 were considered statistically significant. Statistical significance levels in figures are reported as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****).

In sillico protein structure analysis

Amino acid sequences of selected genes from *S. suis* strain P1/7 and *S. mutans* UA159 were retrieved from the NCBI database. The *S. suis* sequences included HdrM (SSU1817), LiaF (SSU0386), and BrsM (SSU1906); the *S. mutans* sequences included HdrM (SMU_1855) and BrsM (SMU_2081). Predicted three-dimensional structures were generated using the SWISS-MODEL server (https://swissmodel.expasy.org). Structural similarity between proteins was assessed using TM-align (TM-align: A protein structure alignment algorithm using TM-score rotation matrix) [170], based on TM-scores and RMSD values. Structural models and alignments were visualized using PyMOL (version 3.0.3).

Conservation Analysis of LytTR Sensory Systems Across *S. suis* Genomes

Orthologous groups across 60 *S. suis* genomes were identified using OrthoFinder [171]. The genome sequences analyzed, along with their classification into pathogenic and non-pathogenic clades, were obtained from Murray *et al.* [15]. Orthogroups corresponding to the *hdrMR* operon (SSU1817, SSU1818) were selected and extracted using a custom Python script. For each genome, gene clusters containing all *hdrMR* orthogroup genes plus surrounding genomic context were saved as GenBank files. Alignments and level of DNA sequence identity of the *hdrMR* operon across all 60 *S. suis* genomes were generated using Clinker[172]. Due to the large size of these alignments, only representative subsets of 20 genomes (10 pathogenic and 10 non-pathogenic) were used for visualization in the manuscript. The same subset was then analyzed for the *brsMR* operon (SSU1906, SSU1907). Another custom script extracted the *brsMR* orthogroup genes along with neighboring context genes, generating GenBank files for Clinker visualization of gene arrangement and conservation.

Gene deletion mutants

In-frame deletion mutants were constructed using CRISPR/Cas9 based technology essentially as described by Gussak et al.[127]. Briefly, mutants were generated by transforming *S. suis* P1/7 with a gRNA-Cas9 co-expression plasmid (pSStarget) and a linear DNA repair template [127]. Upon transformation, Cas9 endonuclease was expressed and directed by the co-expressed guide RNA to specific target gene, where Cas9 induced a double-stranded cut that was repaired by the *S. suis* homologous recombination machinery using the introduced DNA repair template. Resulting colonies were screened for gene-specific deletion mutants, and plasmid constructs were cured.

For CRISPR-Cas-mediated gene deletions, three different 20 bp guide RNAs were designed for each target gene. Single-stranded oligonucleotides were designed using Benchling (https://benchling.com), and synthesized by IDT Technologies. Each primer included a 4 bp overhang compatible with the overhangs of the linearized pSStarget plasmid (Table S1, from P003 to P016). Equimolar concentrations of complementary primers for each guide were mixed with annealing buffer (10 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA) and annealed in a thermocycler (Bio-Rad, Hercules, CA, USA) (5 min at 95 °C followed by gradual cooling to 25°C at 1 °C/min). The empty pSStarget plasmid was linearized using the BsaI enzyme (NEB, R3733S) and purified with the Invisorb Fragment CleanUp kit (Invitek, 1020300300). The annealed guide and digested plasmid were ligated using T4 DNA ligase (NEB, M0202S) for 1 hour at room temperature and subsequently incubated overnight at 4 °C. The ligation mixture was transformed into chemically competent *E. coli* Top10 cells and plated on LB agar plates containing 10 µg/mL of chloramphenicol, followed by overnight incubation at 37 °C. Colonies were screened for correct plasmid constructs using primers P001 and P002 (Table S1), and plasmids and guide were extracted from these colonies using the Qiagen Miniprep kit (QIAprep Spin Miniprep Kit, 27106).

The repair template was constructed by amplifying approximately 1000 bp upstream and downstream of the target gene using primers P017 to P028 (Table S1). These primers were designed to include approximately the first and last 30 bp of the target gene, ensuring that the entire gene was not deleted to avoid potential polar effects on downstream genes. PCR products were subsequently ligated through Splicing by Overlap Extension (SOE) PCR using external primers, HA1_fwd_X and HA2_rev_X (Table S1). After each PCR step, the PCR products were purified using the Invisorb Fragment CleanUp kit (Invitek, 1020300300).

The gRNA-Cas9 co-expression plasmids and repair templates were transformed into *S. suis* using the natural competence method described by Zaccaria et al. [126]. Briefly, 250 µL of an overnight culture of *S. suis* P1/7 WT were inoculated into 10 mLof THY broth and grown until the OD600 reached between 0.035 and 0.058. A 100 µL aliquot of *S. suis* P1/7 culture was combined with 5 µL of competence-inducing peptide ComS [126], 200–500 ng of plasmid DNA, and 1 µg of repair template. The mixture was incubated at 37 °C with 5% CO_2 for 2 hours and plated on THY agar plates containing 5 µg/mL of chloramphenicol. Colonies with the correct gene deletion were identified using PCR and external primers HA1_fwd_X and HA2_rev_X (Table S1), and these colonies were cured from plasmids by performing two consecutive overnight passages in THY medium without chloramphenicol.

In silico prediction of putative *liaR* and *hdrR* binding boxes

A custom Python script was used to extract promoter regions for all coding sequences in the *S. suis* P1/7 genome, using GFF3 annotations from NCBI GCF_000091905.1. For each gene, a consensus promoter region was defined as the 250 bp upstream of the annotated start codon combined with the first 50 bp of the coding sequence. The script produced both a FASTA file and a CSV file containing promoter sequences and corresponding gene identifiers.

These promoter sequences were then analyzed using the FIMO tool from the MEME suite (https://meme-suite.org/meme/tools/fimo) [129] to identify potential binding sites for LiaR and HdrR. The LiaR binding motif and the HdrR binding motif used as input were based on the consensus sequences reported by Jordan et al. [128] et al. and Zou et al. [169], respectively.

Gene expression analysis

S. suis P1/7 WT and mutant strains were grown to exponential phase (approximately OD_{600} 0.3) in THY broth at 37 °C with 5% CO_2. A 10 mLaliquot was pelleted by centrifugation at 4000 rpm for 10 minutes, supernatant discarded, and the pellet snap-frozen in liquid nitrogen before being stored at -80 °C overnight. RNA was extracted from the pelleted cells using the RNeasy Mini Kit (Qiagen, 74104) following the manufacturer's instructions with specific modifications. Pellets were resuspended in 700 µL of RLT buffer containing 0.1% β-mercaptoethanol and transferred to lysing matrix B 2 mLtubes (MP Biomedicals, 6911100). Bacterial cells were lysed using a FastPrep-24™ 5G bead beating grinder and lysis system (MP Biomedicals, Solon, OH, USA) with the following settings: 4.0 m/sec, All-MetalQuickprep adapter, 40 seconds. The tubes were centrifuged for 1 minute at 10,000 rpm, and the supernatant was transferred to a clean Eppendorf tube. Subsequent steps were carried out using the manufacturer's protocol.

Final RNA concentrations were measured using the Qubit RNA Broad Range Kit (Thermo Fisher Scientific™, Q10211) and a Qubit 4 fluorometer (Thermo Fisher Scientific™, Waltham, MA, USA). The Quantitect Reverse Transcriptase Kit (Qiagen, 205311) was used for genomic DNA degradation and cDNA synthesis by reverse transcription of mRNA, with 500 ng of RNA as input. For differential gene expression analysis, 96-well white PCR plates (Bio-Rad, ML9651) and GoTaq qPCR Master Mix (Promega, A6002) were used in a CFX96 Real-Time PCR System (Bio-Rad, Hercules, CA, USA). Primers for each gene are listed in Table S1 (from P029 to P042). Gene expression was calculated using the 2^-ΔΔCt method relative to the reference gene *gyrA*. Independent experiments were performed in triplicate with three biological replicates each.

RNA-seq Data analysis

Raw paired-end FASTQ files were aligned to the genome of *S. suis* P1/7 (RefSeq: GCF_000091905.1) using Bowtie2, producing SAM alignment files. SAM files were converted to sorted and indexed BAM files using SAMtools. Gene-level read counts were obtained using featureCounts (v2.0.1), allowing multi-mapped reads (-O), proportional assignment (-M --fraction), and paired-end counting (-p). Genes with fewer than 10 mapped reads were excluded from further analysis. Differential gene expression analysis comparing read counts of gene transcripts in Δ*liaR* and Δ*hdrM* mutant strains to corresponding counts in the wild-type was performed using DESeq2 in R. Normalized counts, $\log_2$ fold-changes, and adjusted p-values were exported, and significant changes were defined as those with an adjusted adjusted p-value < 0.05 and $|\log_2 \text{fold-change}| > 1$). Data were visualised in Python using Matplotlib. Volcano plots were created for each mutant, plotting $\log_2$ fold change against $-\log_{10}$ adjusted p-values.

KEGG pathway enrichment analysis of significantly differentially expressed genes (DEGs) was conducted using KEGG Orthology (KO) identifiers assigned by the KofamKOALA tool (KofamKOALA) [173], based on the *S. suis* P1/7 genome annotation file (GenBank accession AM946016). Enrichment analysis was performed in R using the clusterProfiler package, applying Fisher's exact test with KEGG pathway mappings obtained via the KEGGREST API. Significantly enriched pathways were identified at an adjusted p-value < 0.05. Pathway enrichment results, including fold enrichment and gene-level regulation, were visualized in Python with customized horizontal bar plots showing the percentage of DEGs within each KEGG pathway and the corresponding fold-enrichment values.

Time-killing assay with LL-37

S. suis WT, Δ*liaR* and Δ*liaS* mutant strains were grown to exponential phase (OD_{600} 0.2–0.3) by diluting 500 µL of overnight culture into 9.5 mLfresh THY medium and incubating at 37°C. Once the OD_{600} reached 0.2–0.3, cultures were adjusted to an OD_{600} of 0.04, centrifuged at 6000 rpm for 5 minutes, and resuspended in the same volume of CDM in 15 mLFalcon tubes. From these tubes, aliquots of 100 µL bacterial suspension were transferred into flat-bottom 96-well plates, and LL-37 antimicrobial peptide was added at final concentrations of 0 µM, 2 µM, 5 µM, or 10 µM, adjusting the final volume to 200 µL per well with CDM. Each strain was evaluated in two independent experiments, each comprising two biological replicates with two technical replicates per biological replicate. Plates were incubated at 37°C, and bacterial viability was measured hourly over the initial 3-hour period (including T0) and again at 24 hours. At each time point, samples were serially diluted in PBS, plated onto THY agar plates, and incubated overnight at 37°C for CFU count.

Growth curves

Growth of *S. suis* WT and mutant strains was monitored in either THY or APS. In THY, growth was assessed either by measuring optical density at 600 nm (OD_{600}) or by determining colony-forming units per milliliter (CFU/mL) through serial dilutions in PBS and plating on THY agar plates followed by counting colonies the next day. Growth in APS was evaluated using CFU/mL measurements. In all cases, samples were collected every hour over an 8-hour period. For all strains, overnight cultures were first pelleted at 6000 rpm for 5 minutes, and pellets were resuspended in PBS. The appropriate volume of bacterial culture was inoculated into APS to achieve an OD_{600} of 0.015. For THY growth measurements, bacterial overnight cultures were directly inoculated without prior pelleting and resuspension in PBS. Each strain was analyzed in biological triplicate with two technical replicates per biological replicate. Technical replicates were averaged, and the mean and standard deviation across biological replicates were calculated at each time point using Python.

OD_{600} and CFU Correlation

To determine the number of viable cells corresponding to an OD_{600} of 0.04 in THY, *S. suis* cultures were prepared by diluting 500 µL of overnight culture into 9.5 mLof fresh THY and incubating at 37 °C. Samples were taken at two different growth stages: early exponential phase (OD_{600} of 0.08–0.1) and mid-exponential phase (OD_{600} of ~0.2–0.3, reached after 2–3 hours of incubation). In both cases, cultures were adjusted to an OD_{600} of 0.04, serially diluted in PBS, and plated onto THY agar. Plates were incubated overnight at 37 °C, and colonies were counted the following day to calculate CFU/mL. Each strain was tested in three biological replicates, with two technical replicates per biological replicate and technical replicates were averaged.

Competence efficiency assay

S. suis competence was induced using the natural transformation method described by Zaccaria et al. [126]. Briefly, overnight cultures were diluted into 10 mLof THY medium (250 µL overnight culture) and incubated at 37°C for 45–75 minutes until reaching an OD_{600} of 0.035–0.058. At this density, 100 µL aliquots were inoculated with 1 µg of the 3 kb plasmid pNZ8048 (carrying a chloramphenicol resistance gene) and 5 µL of the competence-inducing peptide ComS. Plates were incubated for 2 hours at 37°C. After incubation, samples were serially diluted and plated in duplicate on THY agar and THY agar supplemented with choramphenicol (Cm). Plates were incubated overnight at 37°C, and colonies were counted the following day to calculate CFU/mL.

Each strain was tested in two independent experiments, with three biological replicates per strain and two technical replicates per biological replicate. Technical replicates were averaged, and competence efficiency was calculated as the number of transformants per million viable cells

Biofilm formation assay

S. suis overnight cultures were diluted to an OD_{600} of 0.2 in THY medium. Subsequently, 300 µL of the diluted cultures was supplemented with fibrinogen, glucose, and NaCl at final concentrations of 1%, 1%, and 0.5%, respectively. Triplicates of 100 µL of the resulting mixture were aliquoted into separate wells of a flat-bottom 96-well plate. Surrounding wells were filled with 200 µL of PBS to maintain humidity and prevent edge effects during incubation. Plates were incubated at 37°C in a 5% CO_2 atmosphere for 24 h, 48 h, or 5 days. At each timepoint, planktonic media was removed and OD_{600} was measured to confirm comparable bacterial growth across cultures (data not shown). Wells were washed three times with PBS, fixed with methanol for 10 minutes, and stained with 1% crystal violet (Sigma-Aldrich, V5265) for 10 minutes. After staining, wells were washed five additional times with PBS and air-dried for 2 hours in a fume hood. Before solubilization, images of the stained biofilms were captured using a light microscope (ZEISS Primovert) at 40x magnification. Bound crystal violet was then solubilized in 200 µL of 95% ethanol by gentle agitation for 10 minutes. Absorbance at 595 nm was measured using a SpectraMax microplate reader. Each experiment was performed at least twice, with three replicate wells per timepoint.

Results

Identification of Two Novel LytTR Regulatory Systems in *S. suis*

A genome-wide transposon mutagenesis approach previously identified conditionally essential candidate genes required for *S. suis* growth and survival in APS [167]. One of the candidate genes SSU_RS09155 (SSU1817), encodes a hypothetical protein containing a transmembrane domain found in LiaF a negative regulator of the LiaFSR membrane stress-sensing regulatory system. SSU1817 is positioned within a predicted operon together with SSU1818, annotated as a putative transcriptional regulator containing a LytTR DNA-binding domain (Figure 3C). The genes SSU1818 and SSU1817 are located next to each other on the minus strand, in the order 5′–SSU1818–SSU1817. No promoter was detected immediately upstream of SSU1817, suggesting that both genes are transcribed together as a single operon from the promoter located upstream of SSU1818.

We searched for biological and functional information on homologs of SSU1817 in the literature and relevant databases (see Methods). The search revealed that *S. suis* P1/7 SSU1817 encodes a protein that shares approximately 30% of amino acid sequence identity (data not shown) with transmembrane proteins from a LytTR-type regulatory system described as HdrMR in *S. mutans* [74,168]. Hereafter, we will refer to SSU1818 as *hdrR* and SSU1817 as *hdrM*. In *S. mutans,* this system consists of a transmembrane protein (SMU_1855, HdrM) encoded downstream of a LytTR domain-containing transcriptional regulator (SMU_1854, HdrR). Although this regulatory system was initially described in *S. mutans*, another study also identified homologous LytTR regulatory systems in other streptococci, including *S. pneumoniae* and *S. agalactiae*, as well as in more distantly related genera like *Lactobacillus* and *Bacillus*, showing that these systems are broadly distributed across diverse bacterial lineages [169].

Given the low amino acid sequence identity between *S. suis* HdrM and the *S. mutans* HdrM, we used the SWISS-MODEL server to predict the structure of both proteins and used TM-align to assess their similarity (see Methods). Structural alignment (Figure 1A) using TM-align yielded an RMSD of 3.49 Å across 157 aligned residues, with a TM-score of 0.569 (normalized to the reference structure). TM-scores between 0.5 and 1.0 generally indicate that two proteins share the same overall fold, suggesting that despite limited amino acid sequence similarity, the two transmembrane proteins may adopt comparable structural conformations.

To further investigate whether the *S. suis hdrM* and its neighbouring gene *hdrR* which putatively encoding a LytTR family regulator, are orthologous to the corresponding regulatory system described in *S. mutans*, we examined whether their regulatory mechanism is also conserved. A previously generated *S. suis* Δ*hdrM* mutant (Chapter II), and newly constructed a Δ*hdrR* mutant were analysed. The expression levels of *hdrM* and *hdrR* were quantified in these mutants relative to the WT strain using qPCR. Gene expression analysis showed that the Δ*hdrM* caused a strong upregulation of the operon, with a $\log_2$FC of about 8 (256-fold increase) compared to WT. In contrast, deletion of *hdrR* resulted in downregulation of the *hdrMR* operon.

In *S. mutans*, a second LytTR-type regulatory system called BrsMR is functionally linked to HdrMR. To explore whether *S. suis* possesses a comparable regulatory module to BrsMR, we conducted sequence similarity searches using the amino acid and nucleotide sequences of the *S. mutans brsM* and *brsR* genes against the *S. suis* P1/7 genome, but did not find any closely homologous sequences. As an alternative approach, we searched for similar binding motifs in the *S. suis* genome using the conserved HdrR binding site from *S. mutans* (ACCTCTTAG–12 bp–ACCACTTAA [75]) as a query input for FIMO (Find Individual Motif Occurrences; see Methods), screening

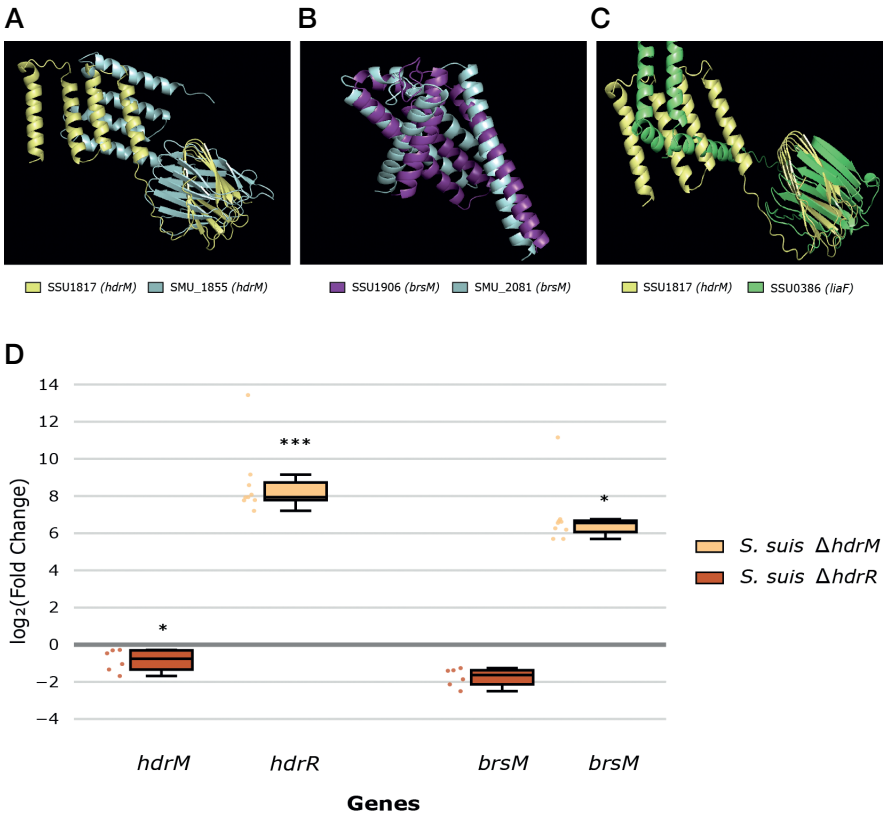


Figure 1. Structural alignment of predicted LytTR regulatory system components in *S. suis*. (A) Predicted structure of SSU1817 (putative HdrM in *S. suis*, yellow) aligned with SMU_1855 (HdrM from *S. mutans*, grey). (B) Predicted structure of SSU1906 (putative BrsM in *S. suis*, purple) aligned with SMU_2081 (BrsM from *S. mutans*, grey). (C) Predicted structure of SSU1817 (putative HdrM in *S. suis*, yellow) aligned with SSU0386 (LiaF from *S. suis*, green. (D) Relative expression of *brsM, hdrM* and *hdrR* in mutants of Δ*hdrM* (yellow) and Δ*hdrR* (red) measured by qPCR. Y-axis indicates fold change (FC) in expression altered genes relative to the WT strain. Data represent the mean of three biological replicates; error bars show standard deviation. Statistical significance levels in figures are reported as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****).

a custom dataset of *S. suis* promoter regions. This search identified 55 potential HdrR target genes (Table S1), including *hdrR* itself. Among these we found SSU1907 encodes a putative LytTR type transcriptional regulator (Figure 3C). Notably, SSU1906 is positioned immediately downstream of SSU1907, in a 5′–SSU1907_SSU1906

arrangement and lacks an independent promoter, suggesting that the two genes are co-transcribed (Figure 3C). Structural prediction of SSU1906 revealed high similarity to BrsM from *S. mutans*, with a TM-score of 0.837 and an RMSD of 1.97 Å over 140 aligned residues, indicating that both proteins likely share the same overall fold (Figure 1B).

We designated SSU1907 and SSU1906 as *brsR* and *brsM*, respectively, based on their genetic organization and protein fold similarity to the *S. mutans* BrsRM system. Gene expression analysis by qPCR in the *hdrM* and *hdrR* gene deletion mutants showed that deletion of *hdrM* led to a statistically significant increase in *brsR* and *brsM* expression, with transcript levels reaching up to 10-fold higher than in the WT strain. In contrast, deletion of *hdrR* resulted in expression levels comparable to the WT (Figure 1D).

Conservation of the *hdrMR* Sensory System and Its Genomic Context in *S. suis* Strains

To assess whether the *hdrMR* and *brsMR* regulatory systems are conserved among *S. suis* strains, particularly those associated with disease, we analyzed their presence, sequence conservation, and genomic organization in a set of genetically and phenotypically diverse *S. suis* strains. We first identified orthologous groups for the four genes (*hdrM*, *hdrR*, *brsM*, and *brsR*) and searched for homologs across 60 *S. suis* genomes, including 30 strains from the recently described pathogenic clades and commensal population[4]. For each gene, we aligned the coding sequence together with its genomic context, including up to three flanking genes upstream and three downstream.

In Figure 2, we show 20 representative examples selected from the full dataset of 60 *S. suis* genomes, including 10 from the pathogenic clade and 10 from the commensal clade. Our analysis reveals that the *hdrMR* operon is fully conserved at the nucleotide level in the pathogenic strains. Its surrounding genomic context is also largely conserved, with genes such as SSU1819, a major facilitator superfamily (MFS) transporter, and glutamate–tRNA ligase (*gltX)* consistently present. One pathogenic strain (*S. suis* 1506978) however, lacks the MFS transporter gene adjacent to the *hdrMR* operon. Although the *hdrMR* system is also present in all commensal strains, it shows reduced sequence identity, and its flanking genes are more variable or absent, consistent with the high diversity of non-disease associated S. suis strains. The *brsMR* system, on the other hand, is neither consistently present nor well conserved in strains from pathogenic or commensal clades, and its surrounding genomic organization is highly variable across the dataset (Figure 2).

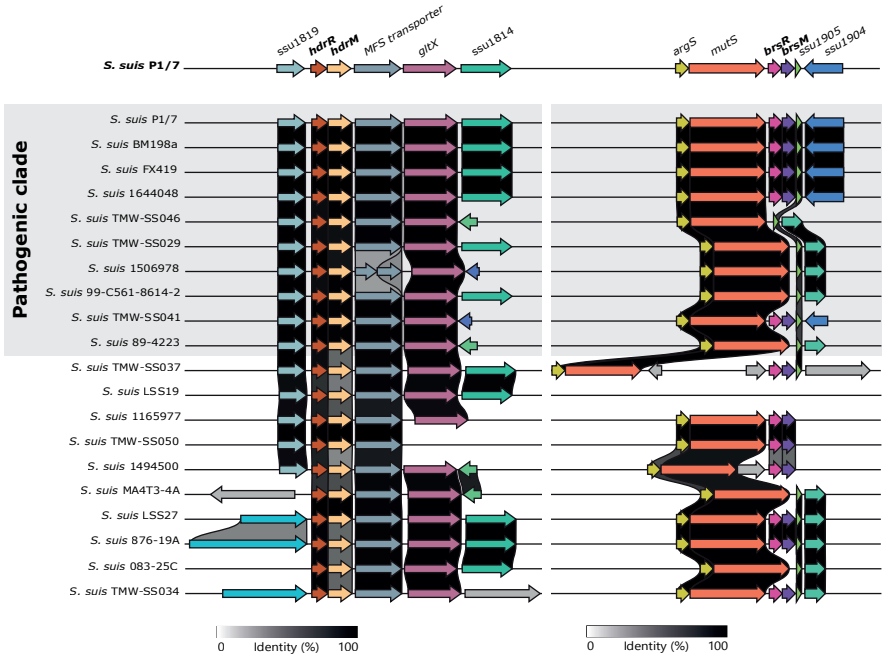


Figure 2. Conservation of the *hdrMR* and *brsMR* operons and their genomic organization across *S. suis* strains. Orthologs of *hdrMR* and *brsMR* were identified using OrthoFinder, and surrounding genomic regions were aligned and visualized with Clinker. Arrows represent individual genes; genes sharing the same color belong to the same orthogroup. For clarity, all gene clusters are displayed in the same orientation. Shading between genes indicates the percentage of amino acid identity. Strains from the pathogenic clade are highlighted with a grey background on the left.

Regulatory Crosstalk Between LytTR-Type and LiaFSR Sensory Systems in *S. suis*

To explore the potential regulon of HdrR, we re-examined the FIMO predictions of genes containing the putative HdrR binding motif and identified a gene encoding a LiaR homolog, the response regulator that is part of the well-characterized three-component LiaFSR system (Figure 3C, table S1). The LiaSFR regulatory system includes the genes *liaF* (a membrane-associated modulator), *liaS* (a histidine kinase), and *liaR* (a response regulator) [64,137,174]. According to NCBI conserved domain predictions, *hdrM* encodes a hypothetical transmembrane protein with a LiaF-like domain. To further examine this potential homology, we modelled the structures of a LiaF-like *S. suis* protein and HdrM and evaluated their similarity with TM-align.

The resulting TM-score was 0.46, with an aligned length of 133 residues and an RMSD of 4.11 Å, indicating moderate structural similarity. This structural similarity, combined with our previous transposon mutagenesis results identifying both *hdrM* and *liaR* as conditionally essential for growth in APS (Chapter II), supported the hypothesis that these two systems may be connected. We focused on the functional characterization of *hdrMR* because this operon is more widely conserved across *S. suis* strains than the predicted *brsMR* operon (Figure 2).

To investigate potential connections between regulatory roles of *liaSFR* and *hdrMR* in *S. suis* P1/7, we measured gene expression in *ΔliaR* and *ΔliaS* mutants [167]. We focused on genes previously identified as LiaR-regulated in *S. mutans*[137]: *liaF*, a PspC-domain-containing protein (hereafter referred to as PCDC), and *spxA* (SSU0386, SSU1424, and SSU0063) as well as *hdrR* (Figure 3A). In the *ΔliaR* mutant, expression of *liaF,* PCDC*,* and *spx* was significantly down-regulated compared to WT. Interestingly, expression of *hdrR* was strongly up-regulated, 10 to 80 fold higher than in the WT strain (Figure 3A). In contrast, the Δ*liaS* mutant, which cannot phosphorylate LiaR through its cognate histidine kinase, showed downregulation of *liaF,* PCDC*,* and *spx*, but no significant change in *hdrR* expression (Figure 3A).

To further extend these findings, we performed transcriptomic analysis on Δ*hdrM* and Δ*liaR* mutants. In the Δ*liaR* mutant, 1,178 genes showed LiaR-dependent differential expression (adjusted p value < 0.05 and absolute $\log_2$ fold change > 1). Within the 10 genes showing the greatest fold changes, we identified the complete *hdrMR* operon (SSU1818, SSU1817), and the *brsMR* system (SSU1907 and SSU1906), both strongly overexpressed (6 to 10 $\log_2$ FC) (Figure 3B, Table 1). The adjacent MFS transporter (SSU1816), located directly upstream of *hdrM* but not part of the operon, was also highly upregulated (Figure 3B, Table 1). This gene had previously been noted as part of the conserved genomic context in *S. suis* strains, particularly within the pathogenic clade (Figure 2). Beyond these sensory system components, we also identified a cluster of four co-regulated genes encoding enzymes of the tagatose-6-phosphate pathway for galactose catabolism, including *lacA* and *lacB*, and a gene of unknown function (SSU0352), all of which were significantly affected in the *ΔliaR* mutant (Table 1).

In contrast, the *ΔhdrM* mutant showed only 23 differentially expressed genes (adjusted p value < 0.05 and absolute $\log_2$ fold change > 0.5). Six genes were strongly upregulated, ($\log_2$ FC 3.8 to 9.37) and adjusted p-values < 0.05 (including the *hdrMR* operon, the *brsMR* operon, the upstream MFS transporter (SSU1816), and an additional hypothetical protein (SSU1546), (Figure 3B; Table 1). Although the overall number of changes were smaller than in the *ΔliaR* mutant, these sensory system-associated genes were similarly overexpressed in both mutants (Figure 3B, Table 1).

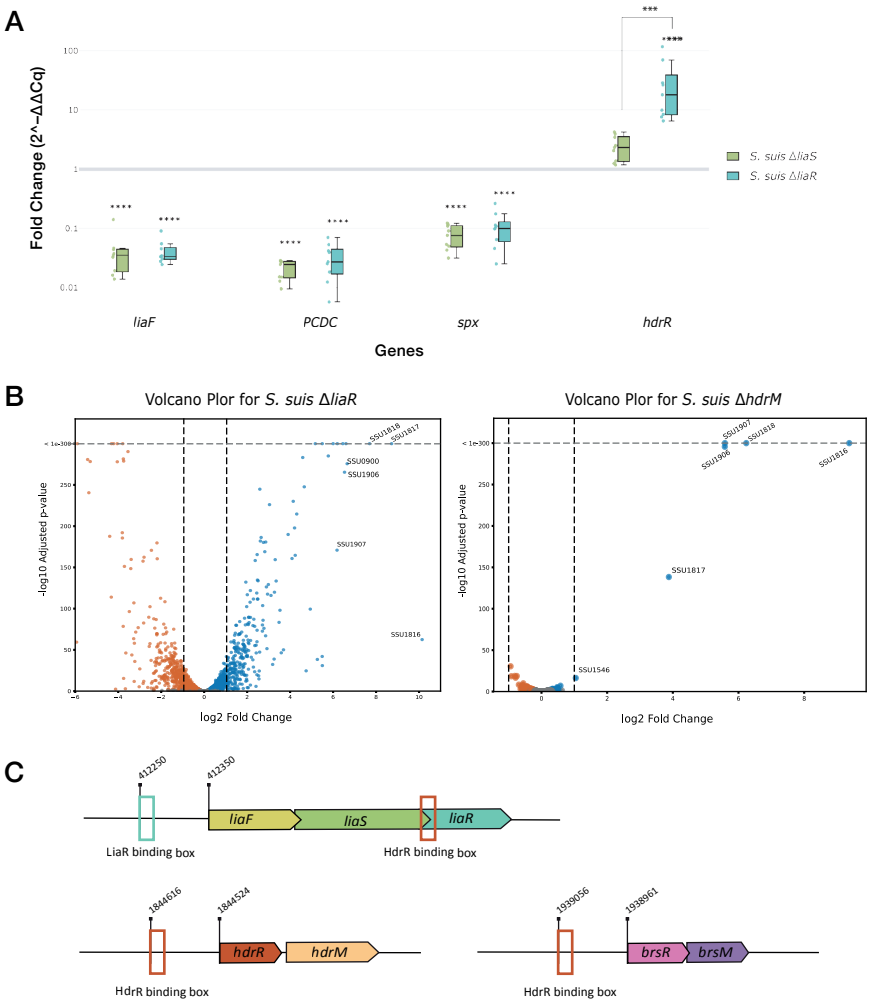


Figure 3. Regulatory interactions between the LiaFSR system and the LytTR-type regulators HdrR and BrsR in S. suis. (A) Expression of *liaF, liaH, spx,* and *hdrR* in Δ*liaS* and Δ*liaR* mutants, measured by qPCR. The Y-axis shows relative fold change (2^−ΔΔCq) on a logarithmic scale. Each dot represents a biological replicate; boxes indicate interquartile ranges, with lines marking the median. Statistical significance levels compared to the WT are reported as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****). (B) Volcano plots showing differential gene expression in the Δ*liaR* (left) and Δ*hdrM* (right) transcriptomes. Genes with $|\log_2 \text{fold change}| > 1$ and adjusted $p < 0.05$ are shown in color: upregulated genes in blue and downregulated genes in orange. In Δ*liaR*, genes with $\log_2$ fold change > 6 are labeled; in Δ*hdrM*, genes with $\log_2$ fold change > 1 are labeled. (C) Schematic representation of the *liaFSR, hdrMR,* and *brsMR* operons. Predicted LiaR binding sites are indicated in blue; predicted HdrR binding sites are indicated in orange.

Table 1. Differentially expressed genes in Δ*liaR* and Δ*hdrM* mutants.

***S. suis* ΔliaR**			
Gene id	**Gene name**	**Log_2FC**	**Adj. p-value**
SSU0352	AAA family ATPase	6.59	0
SSU0896	tagatose-6-phosphate kinase	6.00	0
SSU0898	galactose-6-phosphate isomerase subunit LacB	6.19	0
SSU0899	galactose-6-phosphate isomerase subunit LacA	6.45	0
SSU0900	galactose-6-phosphate isomerase subunit LacA	6.66	2.14E-276
SSU1906	***brsM***	6.53	4.22E-266
SSU1907	***brsR***	6.18	1.42E-171
SSU1816	**MFS transporter**	10.14	3.46E-63
SSU1817	***hdrM***	8.72	0
SSU1818	***hdrR***	7.69	0
SSU0386	*liaF*	-3.7	6.7E-152
SSU0387	*liaS*	-3.8	9.5E-193
SSU0388	*liaR*	-5.42	1.7E-281
***S. suis* ΔhdrM**			
Gene id	**Gene name**	**Log_2FC**	**Adj. p-value**
SSU1546	hypothetical protein	1.044	4.30E-17
SSU1906	***brsM***	5.58	2.14E-296
SSU1907	***brsR***	5.58	0
SSU1816	**MFS transporter**	9.37	0
SSU1817	***hdrM***	3.87	4.70E-139
SSU1818	***hdrR***	6.23	0
SSU0386	*liaF*	-0.31	0.028
SSU0387	*liaS*	-0.44	3.4E-07
SSU0388	*liaR*	-0.41	4.34E-06

Genes shown were selected based on strong expression changes or relevance to the study. Listed are gene IDs, annotated or assigned names, log_2 fold change, and adjusted p-values. Gene annotations are based on NCBI database entries. liaFSR operon genes are included due to their relevance in this study.

In the Δ*hdrM* mutant, the *liaF*, *liaS*, and *liaR* genes were significantly downregulated (log_2FC −0.31 to −0.44; adjusted $p < 0.05$) (Table 1). These findings align with the binding motif analysis, which predicted a potential HdrR binding site at the start of the *liaR* coding sequence (Figure 3C, Table S1). Overall, these results, together with qPCR data, indicate that deletion of either *liaR* or *hdrM* causes strong overexpression of the *hdrMR* and *brsMR* operons, while reducing *liaFSR* expression.

LiaFSR Contributes to Resistance Against Cell Wall–Targeting Antimicrobial Peptides

Since the LiaFSR system has not been previously characterized in *S. suis*, but is known to play a key role in maintaining membrane integrity in other Gram-positive bacteria, we assessed the response of Δ*liaR* and Δ*liaS* deletion mutants to LL-37, a human-derived antimicrobial peptide (AMP) that disrupts bacterial membranes by binding to negatively charged envelope components and inducing membrane permeabilization [175]. Time-kill assays were performed in CDM medium using LL-37 at concentrations of 2.5, 5, and 10 µM (Figure 4).

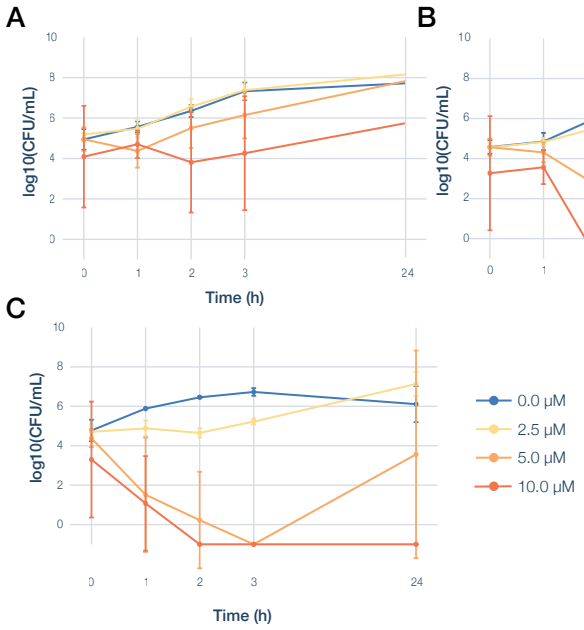


Figure 4. LL-37 time-kill assays for wild-type and LiaFSR mutants of *S. suis*. Bacterial survival was assessed over time in chemically defined medium (CDM) in the presence of increasing concentrations of the antimicrobial peptide LL-37. Viable counts (log_{10} CFU/mL) were determined at multiple time points over 24 hours. (A) WT strain. (B) Δ*liaS* mutant. (C) Δ*liaR* mutant. Each line represents the mean of biological replicates; error bars indicate standard deviation.

The results showed that both Δ*liaR* and Δ*liaS* mutants were more sensitive to LL-37 than the WT strain. While the WT was able to survive and grow at all tested concentrations, including after 24 hours of exposure (Figure 4A), both mutants showed a substantial reduction in CFU/mL at 5 and 10 µM during the first 3 hours. At 5 µM, only the Δ*liaR* mutant was able to partially recover after 24 hours, whereas Δ*liaS* remained highly sensitive throughout the experiment (Figure 4B and 4C).

hdrR and *brsR* Overexpression Compromises Viability, with Stronger Effects in *liaR*-Deficient Strains

To investigate interactions between the *hdrMR*, *brsMR*, and *liaFSR* sensory systems and their impact on bacterial physiology and fitness, we constructed a panel of combinatorial mutants. In addition to the original Δ*hdrM*, Δ*liaR*, Δ*liaS*, and Δ*hdrR* single mutants, we focused on the regulators *hdrR* and *brsR,* which were among the most strongly upregulated genes in the Δ*liaR* transcriptome. To investigate whether these regulators contribute to the Δ*liaR* phenotype, we generated Δ*liaR*Δ*hdrR* and Δ*infliaR*Δ*brsR* double mutants, as well as a Δ*liaR*Δ*hdrR*Δ*brsR* triple mutant. We also constructed a Δ*liaS*Δ*hdrM* double mutant to combine reduced *liaR* activity (due to lack of LiaR phosphorylation in the Δ*liaS* background) with *hdrR* overexpression resulting from *hdrM* deletion.

Previous growth analyses showed that both Δ*hdrM* and Δ*liaR* mutants had impaired growth in APS, but only the Δ*liaR* mutant showed a defect in THY medium. This strain grew normally until OD_{600} 0.2, but then entered a prolonged plateau phase, with OD_{600} values stabilising for several hours before resuming growth at approximately 8 h. To further assess the impact of these regulators, we measured the growth of all single and combinatorial mutants in THY (Figure 5A and 5B). Removing *hdrR* or *brsR* in the Δ*liaR* background, either alone or together, restored the growth profile to that of the WT (Figure 5B). The Δ*liaR*Δ*hdrR*, Δ*liaR*Δ*brsR,* and Δ*liaR*Δ*hdrR*Δ*brsR* strains all showed growth curves like the WT strain. The Δ*hdrM*Δ*liaS* mutant showed a similar pattern to Δ*liaR*, growing up to OD_{600} 0.2–0.3 before entering a phase with stable OD_{600} for several hours, then resuming growth around hour 7. The Δ*hdrM* and Δ*liaS* single mutants showed the same growth pattern as the WT strain.

Because Δ*hdrM* appeared to grow normally in THY (based on OD_{600}) but showed a defect in APS [167], we next measured CFU/mL distinguish whether changes in optical density reflected true differences in viability. We therefore repeated the growth assay in APS and THY performing CFU quantification in both media. In APS, Δ*hdrM* showed a reduction in CFU/mL beginning around hour 4, with viable counts dropping below the initial inoculum and increasing again around hour 7 (Figure 5A). This pattern resembled the behaviour of Δ*liaR* and Δ*liaS*Δ*hdrM* in THY, where OD_{600} stopped

increasing around hour 3–4 and began rising again at approximately hour 7–8 (Figure 5B). In contrast, in THY, Δ*hdrM* reached CFU levels comparable to the WT strain, confirming that the growth impairment of this mutant is specific to APS. We also measured the growth of Δ*hdrR* in APS, which showed no defect (Figure 5A).

These results raised the possibility that the apparent growth limitation seen in Δ*liaR* and Δ*liaS*Δ*hdrM* in THY might also reflect a transient loss of viability, as observed for the Δ*hdrM* mutant in APS, rather than a delayed lag or slow division rate. Since OD_{600} measurements alone cannot distinguish between these possibilities, we measured CFU counts for WT, Δ*liaR* and Δ*liaS*Δ*hdrM* Δ*liaR*Δ*hdrR*Δ*brsR* and at two key timepoints: early exponential phase (OD_{600} 0.08–0.1) and the onset of the growth defect (OD_{600} 0.2–0.3). All cultures were normalized to OD_{600} = 0.04 before plating (Figure 5C). At the first timepoint, CFU counts for Δ*liaR* were similar to WT, while Δ*liaS*Δ*hdrM* and Δ*liaR*Δ*hdrR*Δ*brsR* showed slightly lower values. By the later timepoint, CFU levels had dropped to ~10^6 CFU/mL in both Δ*liaR* and Δ*liaS*Δ*hdrM*, about tenfold lower than in WT or the triple mutant (Figure 5C). Since OD_{600} remained stable during this period, the data indicate that the reduced increase in optical density reflects loss of viability rather than a temporary growth arrest.

General Repression of Amino Acid Biosynthesis in Response to *hdrR* Overexpression

To better understand the molecular basis of the growth defect observed in the Δ*hdrM* mutant, particularly in the context of *liaR* deletion or reduced *liaR* activity (as in the Δ*liaS* background), we analyzed the transcriptomic data to identify pathways that might contribute to the loss of viability. KEGG pathway enrichment analysis was performed on differentially expressed genes (adjusted p value < 0.05) from the Δ*hdrM* mutant. Genes were annotated with KEGG orthologous (KO) terms, and enrichment was assessed using the clusterProfiler package and pathway mappings obtained via KEGGREST. The direction of regulation was determined from $\log_2$ fold change values. This approach enabled us to identify biological pathways significantly influenced by *hdrR* overexpression, providing insights into the physiological role of the *hdrMR* system.

Although the *hdrR* regulon showed only modest $\log_2$ fold changes at the level of individual genes, several KEGG pathways were significantly enriched, with enrichment scores reaching up to 60-fold (Figure 6). Notably, all enriched pathways were mainly composed of downregulated genes. The pathway with the highest representation was valine, leucine, and isoleucine biosynthesis. Additional enriched pathways included arginine, lysine, cysteine, methionine, and glycine, serine, and threonine metabolism, all central to amino acid biosynthesis. Vitamin B6 metabolism, a cofactor pathway involved in amino acid and D-alanine metabolism, showed the second-highest

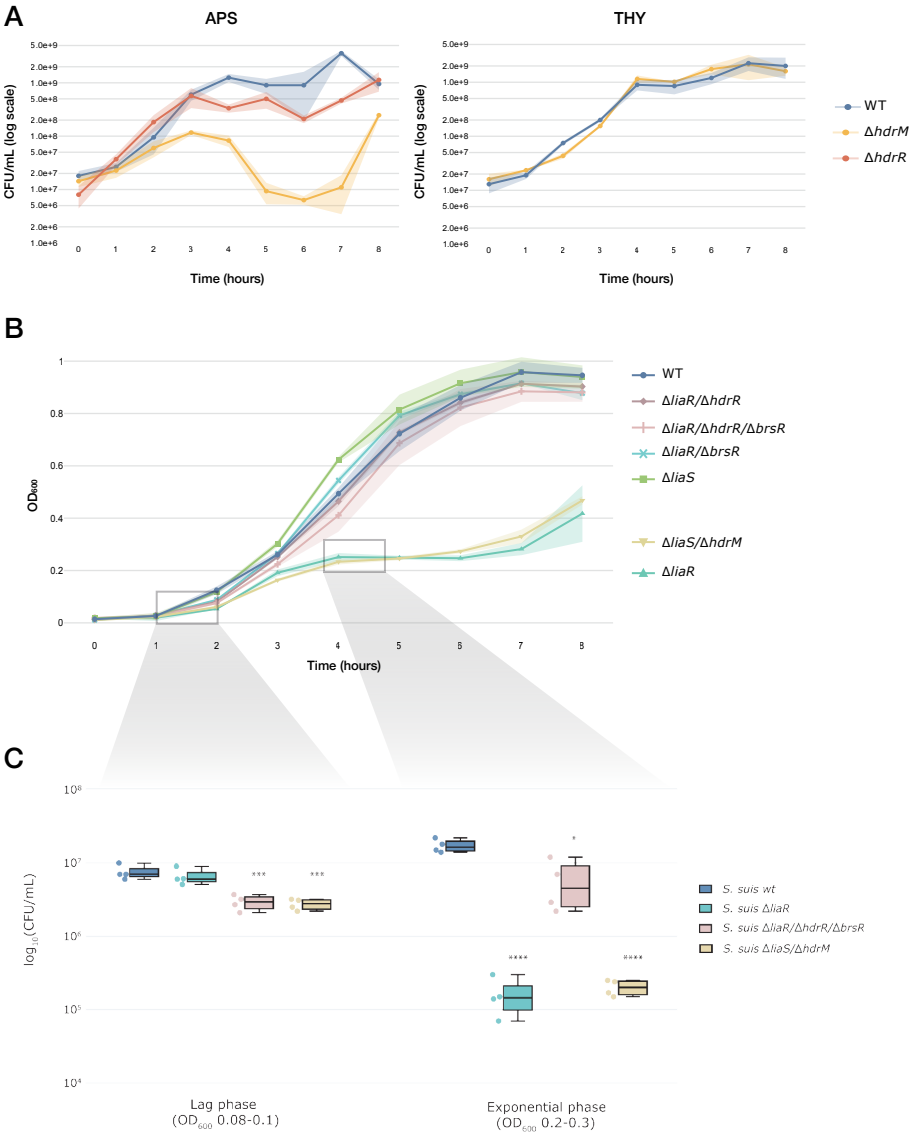


Figure 5. Growth and viability of *S. suis* LiaFSR and LytTR regulator mutants. (A) Time-course CFU/mL measurements in activated porcine serum (left) and THY medium (right) for wild type, Δ*hdrM*, and Δ*hdrR* strains. (B) Growth curves in THY medium (OD_{600}) for single and combinatorial mutants of the LiaFSR, HdrMR, and BrsMR systems. In both panels (A) and (B), shaded ribbons around the lines represent mean ± standard deviation (SD), visualizing variability across biological replicates. (C) CFU/mL quantification for selected strains at two defined time points along the growth curve: early exponential phase (OD_{600} ≈ 0.08–0.1) and

the onset of growth limitation ($OD_{600} \approx 0.2–0.3$). Samples were diluted to $OD_{600} = 0.04$ prior to plating to allow standardized comparison of viable cell counts. Statistical comparisons were performed separately for each time point using ANOVA followed by Tukey's HSD post hoc test when assumptions of normality were met, or Kruskal–Wallis with Dunn's post hoc test when they were not. Statistical significance levels in figures are reported as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****).

proportion of affected genes. We also observed enrichment of pantothenate and CoA biosynthesis, which provide essential cofactors for fatty acid and lipid metabolism. KEGG enrichment also highlighted the monobactam biosynthesis pathway, which includes enzymes involved in D-alanine metabolism and peptidoglycan biosynthesis, suggesting changes in cell-wall associated pathways (Figure 6).

Competence Activity Enhanced in Δ*liaS* and Δ*liaSΔhdrM* Mutants

Previous studies in *S. mutans* have shown that overexpression of *hdrR* is sufficient to activate the competence pathway, including strong upregulation of *comX* and *comY*, independently of the canonical ComCDE system [76]. In contrast, transcriptomic analysis of both Δ*hdrM* and Δ*liaR* mutants in *S. suis* revealed no significant changes in the canonical competence genes such as *comX*, *comYA*, or *comEC*.

Because competence induction in *S. suis* depends on exogenous ComS peptide[126], we tested whether HdrR influences transformation efficiency under inducing conditions. Transformation was assayed using the 3 kb plasmid pNZ8048 as donor DNA and efficiency was calculated as transformants per million viable cells (Figure 7). Both the Δ*liaS* and Δ*liaS*Δ*hdrM* mutants exhibited significantly higher transformation rates than the other strains, with more than a 10-fold increase compared to WT.

These results indicate that competence activation is enhanced when *hdrR* expression is increased in the context of reduced *liaR* activity. In the Δ*liaS* background, LiaR remains unphosphorylated and thus inactive, which leads to mild upregulation of *hdrR* (Figure 2). When *hdrM* is also deleted in this background, the repression of *hdrR* further boosts its expression, resulting in a marked increase in competence activity (Figure 7).

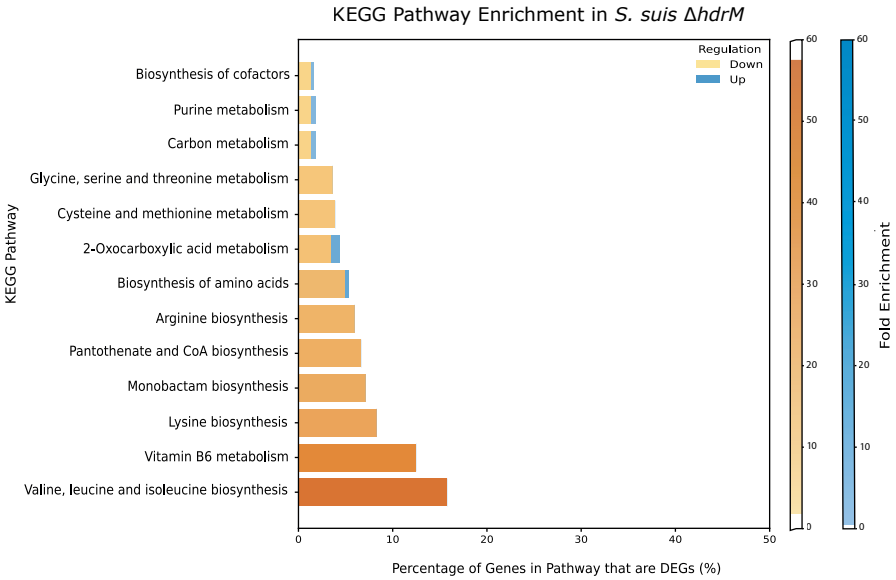


Figure 6. KEGG pathway enrichment analysis of differentially expressed genes in the Δ*hdrM* mutant. Pathways were identified using KO annotations and KEGG enrichment analysis. Bars indicate the percentage of differentially expressed genes (adjusted $p < 0.05$) mapped to each pathway. Regulation direction is indicated by color (blue = upregulated, orange = downregulated), while bar intensity reflects fold enrichment, with darker shades representing stronger enrichment.

Balanced Expression of *hdrR* and *brsR* Is Required for Biofilm Integrity

Biofilm formation has previously been shown to be disrupted by overexpression of *hdrR* in *S. mutans* [75]. To evaluate whether *hdrR*-linked regulation also influences biofilm development in *S. suis*, we assessed the ability of our full mutant panel to form biofilms under defined conditions. Static biofilm assays were optimised using glucose and NaCl as medium additives (see methods) and biofilm biomass was quantified by crystal violet staining after 24, 48, or 5 days. These timepoints were selected to capture both early biofilm formation and recovery capacity over time (Figure 8A).

At 24 hours, the Δ*hdrM*, Δ*liaR*, and Δ*liaR*Δ*hdrR*Δ*brsR* mutants showed significantly reduced biofilm formation compared to the wild type ($p < 0.01$ for all). The Δ*hdrM* mutant recovered by 48 hours, whereas Δ*liaR* only reached WT levels after 5 days. In contrast, the Δ*liaR*Δ*hdrR*Δ*brsR* triple mutant failed to form substantial biofilm at any timepoint. Interestingly, both the Δ*liaR*Δ*hdrR* and Δ*liaR*Δ*brsR* double mutants showed markedly improved biofilm formation relative to Δ*liaR* alone, reaching levels

comparable to wild type by 48 hours. This indicates that deletion of *hdrR* or *brsR* can compensate for the biofilm defect caused by loss of *liaR*, whereas deletion of both in the Δ*liaR* background (Δ*liaR*Δ*hdrR*Δ*brsR*), eliminated this compensatory effect.

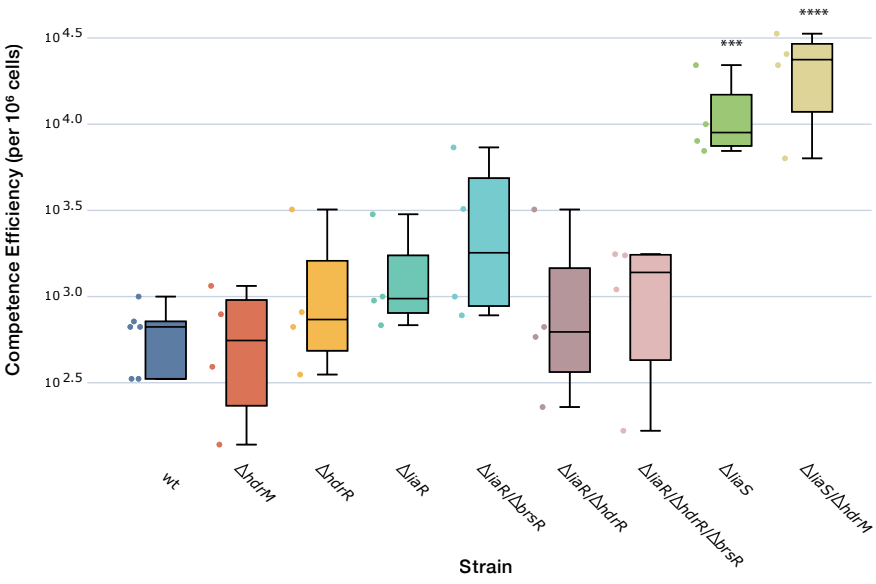


Figure 7. Competence efficiency in *S. suis* LiaFSR and LytTR regulator mutants. Competence efficiency was measured as the number of transformants per million viable cells. Each dot represents one biological replicate. Y-axis shows the $\log_{10}$-transformed efficiency values. Boxplots display the median, interquartile range, and all data points. Statistical significance was assessed by one-way ANOVA followed by Tukey's HSD test, or by Kruskal–Wallis and Dunn's post hoc test when normality assumptions were not met. Statistical significance levels in figures are reported as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****).

Microscopy images of stained biofilms after 48 hours (Figure 8B) support these findings, showing that WT cells formed a dense, uniform layer with extensive surface coverage. In contrast, the Δ*liaR* mutant and the Δ*liaR*Δ*hdrR*Δ*brsR* triple mutant showed severely impaired surface attachment. Δ*liaR* formed irregular, sparse clumps with limited cohesion, while the triple mutant displayed almost no visible biofilm layer. The Δ*liaR*Δ*hdrR* mutant developed larger, more structured clumps than Δ*liaR*, whereas Δ*liaR*Δ*brsR* produced small, dispersed aggregates with poor surface association. After 5 days of incubation (Figure 8C), the WT strain continued to display a thick, uniform biofilm layer. By this point, the Δ*liaR* mutant had partially recovered,

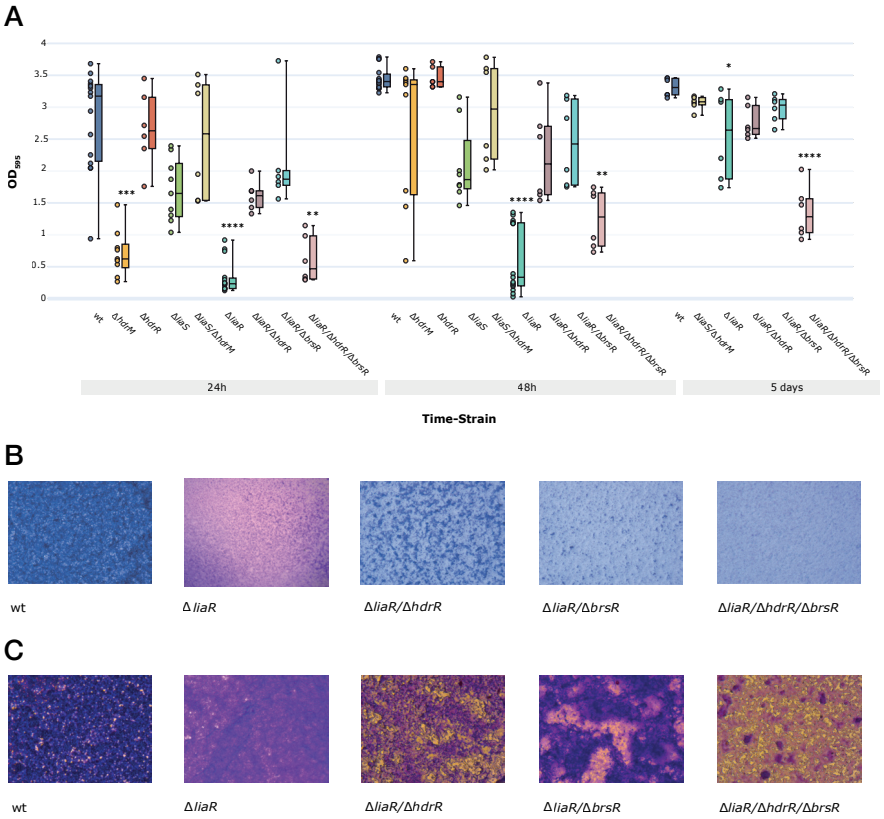


Figure 8. Biofilm formation in *S. suis* LiaFSR and LytTR regulator mutants. (A) Quantification of biofilm formation at 24 h, 48 h, and 5 days based on crystal violet staining. Absorbance at 595 nm (OD_{595}) is shown as boxplots with individual biological replicates overlaid as dots. Boxes represent the distribution of values; horizontal lines indicate the median. Statistical significance was assessed per timepoint using either one-way ANOVA with Tukey's HSD or Kruskal–Wallis with Dunn's test, depending on data normality (see Methods). Statistical significance levels in figures are reported as: $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***), and $p < 0.0001$ (****). (B) Microscopy images of stained biofilms after 48 hours. (C) Microscopy images of stained biofilms after 5 days. Images were taken at 40x magnification.

forming a more cohesive layer with fewer clumps. Both double mutants, Δ*liaR*Δ*hdrR* and Δ*liaR*Δ*brsR*, still exhibited clumped biofilm structures, though with increased coverage and larger aggregates compared to earlier timepoints. In contrast, the Δ*liaR*Δ*hdrR*Δ*brsR* triple mutant remained severely impaired, displaying only scattered, small clumps with minimal surface association.

Together, quantitative microscopy images correlate well with the OD_{595} measurements, reinforcing the observation that both complete absence and uncontrolled upregulation of LytTR regulatory systems disrupt biofilm development (Figure 8A-C). A balanced LytTR regulatory state is therefore critical for proper biofilm development in *S. suis*.

Discussion

S. suis is a zoonotic pathogen of growing concern due to the increasing prevalence of antimicrobial resistance and the absence of effective vaccines [8,162]. These challenges underscore the need to better understand the molecular mechanisms that support its survival within the host and ability to cause disease. Our study, identified and characterised the LiaFSR three component sensory system in *S. suis* P1/7 and two previously uncharacterised LytTR regulatory systems (HdrMR and BrsMR) first identified in *S. mutans* [33]. Together, these form a triad regulatory circuit that integrates competence and biofilm formation with envelop stress responses.

LiaFSR is a three-component sensory system composed of the membrane-anchored inhibitor LiaF, the histidine kinase LiaS, and the response regulator LiaR. LiaF contains a short transmembrane domain thought to monitor changes in membrane stability [66,176]. Upon sensing stress or membrane disruption, LiaF releases its inhibition of LiaS, which then phosphorylates LiaR [177]. Phosphorylated LiaR dimerizes and activates transcription of stress-response genes, initiating a protective response to restore envelope integrity [66]. In *S. suis* strain P1/7, deletion of either *liaS* or *liaR* led to a marked increase in susceptibility to LL-37, a human cathelicidin peptide, a cationic antimicrobial peptide [175]. After 3 hours of exposure, both Δ*liaS* and Δ*liaR* mutants showed a reduction in viability of approximately 1.5 to 2.5 $\log_{10}$ CFU/mL at 5.0 and 10.0 µM LL-37, respectively, compared to the WT strain (Figure 4). This phenotype correlated with the complete absence of LiaR in the Δ*liaR* mutant and a strong downregulation of *liaR* expression in the Δ*liaS* background, as confirmed by gene expression analysis (Figure 3A). These results support findings from other Gram-positive species, where disruption of LiaFSR increases susceptibility to agents that compromise membrane integrity [178–180]

The LiaFSR system protects the cell envelope, but its activity must be tightly regulated. The presence of a LiaR-binding site upstream of the operon allows phosphorylated LiaR to activate its own expression, creating a positive feedback loop [128]. Once this loop is engaged, LiaR levels can rapidly rise, potentially leading to dysregulation. Constitutive overexpression of LiaR in *S. pneumoniae* and *Streptococcus pyogenes,* has been shown to impair virulence and increase susceptibility to antibiotics such as

fluoroquinolones [68,181], underscoring the importance of tightly regulating LiaR activity. In this work, we identified two additional sensory systems, HdrMR and BrsRM, which to our knowledge have previously only been described in *S. mutans* [169]. These systems were identified based on structural predictions, Tm-align calculations revealed high similarity to the *S. mutans* HdrMR and BrsRM systems. Despite this structural similarity, the amino acid sequences in *S. suis* share less than 30% identity with those in *S. mutans*, suggesting that while the sequences have diverged, structural and likely functional conservation has been maintained, pointing to an important functional role for these systems in *S. suis*. This prompted our interest in these systems, with special attention to HdrMR, as comparative study across 60 *S. suis* strains revealed that the HdrMR system is highly conserved, with particularly high sequence identity among pathogenic strains. In *S. mutans*, each system comprises a membrane-bound inhibitor (HdrM or BrsM) and a cognate response regulator (HdrR or BrsR). The regulators appear to be held inactive by their associated membrane protein until an as-yet unidentified stimulus induces activation. This likely triggers the release of the regulator from its membrane-bound inhibitor, allowing it to activate its own operon via a positive feedback loop [73,74].

Our RNA-seq and gene expression analyses support this model: deletion of *hdrM* led to an increase in the expression of both *hdrMR* and *brsRM* (between 5,5 and 9,3 $\log_2$FC), indicating cross-regulation between the two systems. Furthermore, a conserved HdrR-binding motif was identified upstream of both operons, reinforcing the idea that HdrR directly regulates both loci.

As a result of this study, we identified LiaR as a strong inhibitor of the HdrMR and BrsRM sensory systems. Both our RNA-seq and targeted gene expression analyses indicate that these two systems, along with a predicted MFS transporter located downstream of the *hdrMR* operon, are among the 10 most highly upregulated genes in a Δ*liaR* mutant (between 6.5 and 10.2 $\log_2$FC). These findings point to a repressive role for LiaR on the expression of both LyTR systems.

Interestingly, although the Δ*liaR* mutant shows a ~70-fold upregulation of the *hdrMR* operon, deletion of *liaS*, which eliminates LiaR phosphorylation, results in a more moderate increase (~10-fold). This difference suggests that unphosphorylated LiaR, either by direct interaction or indirectly, is responsible for repressing *hdrMR* and *brsRM*. Supporting this, expression levels of genes known to be activated by phosphorylated LiaR, such as *spx* and the gene encoding the PspC-domain-containing protein [137,138], were similarly low in both Δ*liaR* and Δ*liaS* mutants, indicating that, in *S. suis* LiaR is not phosphorylated by alternative kinases in the absence of LiaS. Unphosphorylated LiaR cannot dimerize or activate the liaFSR operon (Figure 3A),

which likely explains the lower intracellular levels of LiaFSR observed in this background. The milder upregulation of *hdrMR* in the Δ*liaS* mutant is therefore consistent with reduced LiaR-mediated repression, due to decreased overall LiaR levels (Figure 9B). Interestingly, the data also revealed that HdrR may slightly downregulate *liaFSR*, as a putative HdrR binding motif is predicted at the start of the *liaR* gene, and RNA-seq data from the Δ*hdrM* mutant indicate a downregulation of the *liaFSR* operon, with $\log_2$ fold changes ranging from –0.31 to –0.44 (Table 1).

The structural similarity between HdrM and LiaF, both having a transmembrane domain composed of four α-helices and an intracellular β-sheet domain (Figure 1C), suggests that these systems might be triggered by similar, membrane-associated signals. The absence of an extracytoplasmic domain further supports the idea that signal detection occurs within or at the membrane. Interestingly, the two systems appear to be activated at different stages of the growth cycle. LiaFSR is known to be active during exponential phase, a period characterized by rapid cell division and continuous membrane synthesis. In contrast, HdrMR and BrsRM, might be activated under conditions of high cell density, such as during stationary phase [75,182]. This temporal separation suggests that the systems may be tuned to distinct physiological states, potentially coordinating membrane stress responses across different growth conditions.

Growth curve experiments highlight the critical role of *liaR* in supporting normal growth: the Δ*liaR* mutant was unable to grow beyond an OD_{600} of ~0.2. Cell viability assays confirmed that this defect coincides with entry into exponential phase, at which point cells begin to lyse. In contrast, deletion of *liaS* did not impair growth, indicating that the absence of activated LiaR alone is not responsible for the observed fitness defect. Remarkably, double deletion of *liaR* together with either *hdrR* or *brsR* fully restored both WT growth and cell viability. This demonstrates that the fitness loss is due to overexpression of the HdrMR/BrsMR systems, rather than the absence of LiaR itself, revealing an important inhibitory role for unphosphorylated LiaR that was previously unrecognized.

Accordingly, overexpression of LytTR systems appears to suppress *liaFSR* expression under high cell density conditions, when cell division ceases and peptidoglycan synthesis is no longer required. Upon entry into exponential phase, *liaFSR* is reactivated and subsequently downregulates LytTR system expression. This reciprocal regulatory mechanism may help coordinate cell envelope homeostasis in response to changing growth conditions (Figure 9). Interestingly, overexpression of *hdrR* alone, via deletion of *hdrM*, did not impair growth in THY. However, the Δ*hdrM* Δ*liaS* double mutant displayed the same severe growth defect and loss of viability as the Δ*liaR*

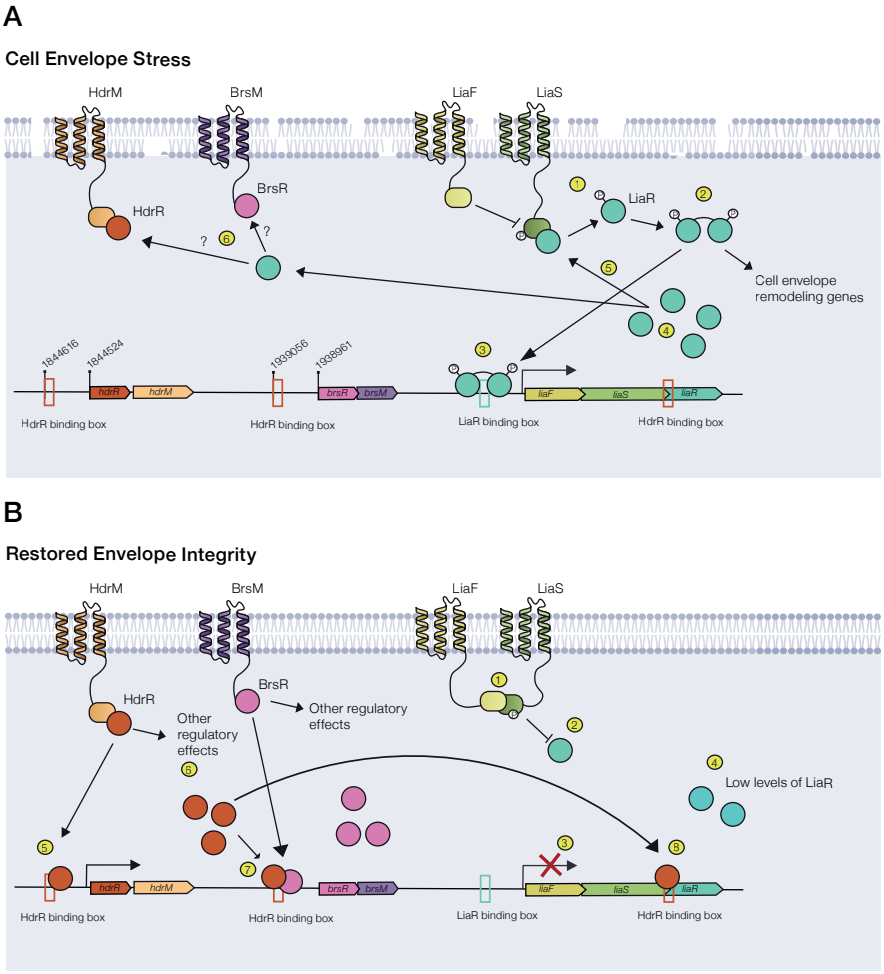


Figure 9. Proposed model of regulatory crosstalk between LiaFSR, HdrMR, and BrsMR systems in *S. suis*. (A) Under cell envelope stress, the sensor kinase LiaS phosphorylates LiaR, which activates expression of membrane protection genes and promotes its own expression through a positive feedback loop. High levels of unphosphorylated LiaR repress *hdrMR* and *brsMR* transcription, keeping HdrR and BrsR levels low. (B) When membrane integrity is restored, LiaR levels decrease, relieving repression of *hdrMR* and *brsMR*. HdrR and BrsR are derepressed and activate their own operons. HdrR also represses liaFSR, establishing a bidirectional regulatory circuit. This model integrates our findings and suggests that misregulation of this network impairs growth, competence, and biofilm formation. Numbers, refer to sequential processes occurring in the different physiological states relating to cell envelope stress or homeostasis.

mutant. In the absence of both LiaS and HdrM, intracellular levels of LiaR are reduced, and this repression is lost, allowing *hdrR* overexpression to drive the fitness defect. Notably, even though both the Δ*hdrM* and Δ*liaS* single mutants grew normally in THY but showed clear growth defects in APS (Figure 5A) [167]. This suggests that unphosphorylated LiaR, still present in the Δ*hdrM* and Δ*liaS* single mutant strains, may be tolerated under non-stress conditions but becomes insufficient to maintain homeostasis in serum. Consistent with this, transcriptomic analysis of the Δ*hdrM* mutant revealed downregulation of multiple amino acid biosynthesis pathways, which could impair peptidoglycan production and weaken the cell envelope. Together, these findings highlight the importance of precise LiaFSR-LytTR system regulation in supporting envelope integrity and fitness, particularly under host-like stress conditions.

Proper regulation of *liaFSR* and the LytTR systems is also important for several virulence-associated traits. Notably, competence was increased by approximately 10-fold in the Δ*liaS* and Δ*hdrM*Δ*liaS* mutants, the only two mutants showing both downregulation of *liaR* and upregulation of *hdrR*. The double mutant exhibited a slightly higher competence level than the Δ*liaS* single mutant, likely due to stronger *hdrR* upregulation, consistent with findings in *S. mutans*, where HdrMR contributes to competence development by activating late competence-related genes under high-density conditions [75,76].

This regulatory imbalance also impacted biofilm formation. Notably, deletion of *liaS*, which eliminates LiaR phosphorylation but retains unphosphorylated LiaR, had no measurable impact on biofilm formation, suggesting that it is the complete loss of LiaR-mediated repression, rather than the absence of its active form, that disrupts biofilm development. In *S. mutans*, HdrMR and LiaFSR systems have been implicated in biofilm formation, where their misregulation disrupts biofilm structure. This suggests that proper coordination of these systems is also important for biofilm development in *S. suis* [75,183]. Together, these findings suggest that LiaFSR and HdrMR form a mutually inhibitory regulatory circuit, in which each system constrains the activity of the other (Figure 9). This reciprocal repression likely functions as a homeostatic mechanism to prevent overactivation, allowing the cell to maintain membrane integrity in response to changing envelope-associated signals.

Additionally, our results reveal previously unrecognized roles for unphosphorylated LiaR in regulating key cellular processes. Unlike phosphorylated LiaR, which appears dispensable for these functions, the absence of unphosphorylated LiaR had a marked impact on growth, competence, and biofilm formation, highlighting its critical repressive role in maintaining regulatory balance. To elucidate the underlying mechanisms, DNA-binding assays and protein–protein interaction studies will be

required to determine how unphosphorylated LiaR represses LytTR regulatory systems. Understanding this regulatory interplay is particularly valuable given the strong influence these systems exert on virulence-related traits. Because of their central role in envelope stress sensing and virulence regulation, targeting LiaFSR or LytTR-type systems may represent a promising strategy for antimicrobial development, as previously suggested in *S. mutans* and other Gram-positive pathogens [178,179,181].

Data availability

All custom scripts, the complete OrthoFinder output, interactive HTML alignments, and any other analyzed data supporting the findings of this chapter will be deposited in a public repository upon publication. The RNA-seq data generated in this study have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB96309 and will be released upon publication.

Acknowledgements

I would like to sincerely thank Blanca Fernández-Ciruelos for her continuous advice and helpful discussions throughout this project. I am also grateful to Dr. Edwin J.A. Veldhuizen (Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, Utrecht University) for kindly providing the antimicrobial peptide LL-37 used in this study.

Supplementary Data

Table S1 Primers used in this study

Primer Nº	Primer name	Sequence
P001	Check_guide_pSStarget_fwd	cgcattgatttgagtcagctagg
P002	Check_guide_pSStarget_rev	tcggtgccacttttcaagttg
P003	Guide1_fwd_ hdrR	tgattacaggcgatgatgcctacg
P004	Guide1_rev_ hdrR	aaaccgtaggcatcatcgcctgta
P005	Guide2_fwd_ hdrR	tgatagtccttctcgggcaccagt
P006	Guide2_rev_ hdrR	aaacactggtgcccgagaaggact
P007	Guide3_fwd_ hdrR	tgatgttatccgaaccagccagct
P008	Guide3_rev_ hdrR	aaacagctggctggttcggataac
P009	Guide1_fwd_ brsR	tgatcgttagcagtcagtttgagg
P010	Guide1_rev_ brsR	aaaccctcaaactgactgctaacg
P011	Guide2_fwd_ brsR	tgatccttaacctgatacaagcgt
P012	Guide2_rev_ brsR	aaacacgcttgtatcaggttaagg
P013	Guide3_fwd_ brsR	aaacgctccccttttttaacagtt
P014	Guide3_rev_ brsR	tgataactgttaaaaaaggggagc
P015	Guide3_fwd_hdrM	tgatggttaatacaaaccagacag
P016	Guide3_rev_hdrM	aaacctgtctggtttgtattaacc
P017	HA1_fwd_hdrR	ctagggaagatttccgagaagg
P018	HA1_rev_ hdrR	ggttgaagttgttatccgaacccacaaggaagtccacgtgtc
P019	HA2_fwd_ hdrR	ggttcggataacaacttcaacc
P020	HA2_rev_ hdrR	ctgcccagtataaaatggaacg
P021	HA1_fwd_brsR	aacctcacacgttgctattaaacc
P022	HA1_rev_ brsR	atgaagaatggctcggtcacc
P023	HA2_fwd_ brsR	ggtgaccgagccattcttcatagcttcaatagtcaccaaatcttcc
P024	HA2_rev_ brsR	cttggttagtggtcaatcgacct
P025	HA1_fwd_hdrM	agcaacattccgacaaaaacg
P026	HA1_rev_ hdrM	acgggtgatatggcattttctac
P027	HA2_fwd_ hdrM	gtagaaaatgccatatcacccgtcagtcctataactgcagctcc
P028	HA2_rev_ hdrM	ctgcccagtataaaatggaac
P029	qPCR_fwd_gyrA	gcgtgggtcagcatatttacg
P030	qPCR_rev_gyrA	cgtcttcgtcgtttgacagg
P031	qPCR_fwd_liaF	ccacgaccactctcaagatagg

Table S1 Continued

Primer Nº	Primer name	Sequence
P032	qPCR_rev_liaF	gatactagttgctgaaagtgcaacc
P033	qPCR_fwd_spx	tgaagagttgagtgtgaacgag
P034	qPCR_rev_ spx	tgagcattttcaacattgcgc
P035	qPCR_fwd_PDCP	ccaagacatgacctgattgcg
P036	qPCR_rev_ PDCP	ggatattcgcttgacgacttgc
P037	qPCR_fwd_hdrR	ccctacaaaagtagatgggc
P038	qPCR_rev_ hdrR	tttttacaaggggagcagcg
P039	qPCR_fwd_hdrM	tattacccagaatcaccgcc
P040	qPCR_rev_ hdrM	gataagggacgcgatacagc
P041	qPCR_fwd_brsM	aggtgaccgagccattcttc
P042	qPCR_rev_ brsM	tcaggttaaggaggacttgcc

Chapter 4

Two Type I Restriction-Modification Systems interact to Modulate Gene Expression and Growth of Virulent Strains of *Streptococcus suis*

Maria Juanpere-Borras[1], Alex Gussak[1], Oshin Vellalara[1], Gemma Murray[2], Peter van Baarlen[1] and Jerry M. Wells[1].

[1] *Host-Microbe Interactomics Group, Wageningen University and Research, Wageningen, Netherlands*
[2] *Department of Genetics, Evolution and Environment, University College London, UK*

Abstract

Type I restriction-modification (RM) systems are traditionally known for protecting bacterial genomes against foreign DNA, but recent evidence suggests they also contribute to epigenetic regulation. Comparative genomics has shown that two co-existing Type I RM systems are present in *Streptococcus suis* sequence type 1 (ST1), the lineage regarded as the most virulent, globally distributed, and responsible for most reported human cases. To investigate these systems, we focused on strain P1/7, a reference isolate of ST1. *S. suis* P1/7 encodes two co-existing Type I RM systems. MS1, which is phase-variable with four *hsdS* alleles, and MS2, which exists in either a full-length or truncated form in different pathogenic lineages of the *S. suis* population. We generated a panel of 18 genetically defined mutants, including locked MS1 variants and deletions of *hsdS* or the entire RM loci, in combinations with MS2 full-length, truncated, or deleted variants of MS2. Notably, the MS1 B variant enabled a distinct transcriptional profile when paired with truncated MS2, characterized by upregulation of protein synthesis genes, and conferred a growth advantage over other MS1 phase variants. Deletion of the entire MS1 system in the presence of MS2 alleles resulted in the highest growth rates and downregulation of lactose PTS systems and the tagatose pathway, suggesting that MS1 constrains MS2. The lack of this inhibition could be problematic in host-like environments, where galactose and lactose are abundant sugar sources. These findings reveal a novel epistatic interaction between two RM systems and suggest that combinatorial methylation patterns may modulate gene expression and growth in *S. suis*. This work expands our understanding of RM systems as dynamic regulators of bacterial physiology and evolution, beyond their classical roles in restriction of methylated DNA.

Introduction

S. suis is an encapsulated, Gram-positive bacterium that primarily resides in the upper respiratory tract of pigs and piglets, mainly the tonsils, but may also be present in the digestive and reproductive tracts [6,33]. The species *S. suis* consists mostly of commensal lineages that asymptomatically colonize pigs and piglets, but also contains disease-associated lineages that cause severe invasive infections including respiratory disease, meningitis, arthritis, and septicaemia [6,14]. These infections significantly impact pig health and productivity and leading to substantial economic losses in the swine industry[105].

S. suis is also considered an emerging pathogen of humans, the highest numbers of cases in China and Southeast Asia [107,109]. Human infection is thought occur primarily through direct contact with pigs or pork products via skin wounds or through the consumption of raw or undercooked pork, a common dietary practice in parts of Southeast Asia [107,108].

Which of the numerous *S. suis* virulence factors involved in adhesion or invasion of host cells and avoidance of immune responses are critical for pathogenesis remains unclear, due to the high genetic diversity of *S. suis* strains and redundancy of gene function [23]. Genes encoding capsular polysaccharide biosynthesis and the cholesterol-binding pore-forming toxin suilysin are considered to play a major role in avoiding opsonization and phagocytosis [184,185]. Based on variation in capsular polysaccharide composition and antigenicity, 35 capsule serotypes of *S. suis* have been proposed [23,184], of which 6 have bene reassigned to other species [186,187]. Serotype 2 is most associated with invasive disease in humans and pigs and with zoonotic potential [106,186,188].

Genomic analyses of disease and non-disease associated isolates of *S. suis* have shown, among other findings, that disease-associated *S. suis* strains (DASS) have smaller genomes containing diverse mobile genetic elements and high recombination rates [189–191]. A more recent study on the genomic analyses of 3,071 isolates sampled from across Europe, North America, Asia and Australia, from 1960 to 2020, showed that most cases of porcine and human *S. suis* disease globally are caused by a small number of highly pathogenic lineages that have acquired specific pathogenicity-associated genomic islands through intra- and inter-species horizontal gene transfer [4]. Some DASS harbour gene clusters encoding phase-variable Type I RM systems, which classically function in DNA defense by restricting foreign DNA from lytic bacteriophages [192,193]. Beyond this defensive role, several studies have demonstrated that RM systems can also act as epigenetic regulators, influencing phenotypic variation, virulence, immune evasion, and adaptation to different host environments [83]. Such genetic and epigenetic plasticity is crucial for bacterial survival across diverse host niches and for evading host immune responses [194–196].

Functional type I RM systems consist of a pentameric complex composed of two HsdR restriction endonuclease proteins which recognizes and cleaves unmethylated DNA, two HsdM methyltransferase proteins that catalyze DNA methylation DNA nucleobases [197,198], and one HsdS specificity protein that contains two recognition domains determining the DNA recognition sequence [199]. Phase-variable Type I RM systems modulate DNA methylation patterns by recombining different configurations of the methylase specificity subunit HsdS. This recombination generates multiple epigenetic states, often referred to as phasevarions (phase-variable regulons), which result in reversible changes in gene expression without altering the underlying DNA sequence. Such mechanisms allow clonal bacterial populations to diversify phenotypically, enhancing their adaptability to varying (host) environments or selective pressures [200]. In *Streptococcus pneumoniae*, for example, distinct *hsdS* configurations generate unique methylation profiles that influence the expression of virulence-associated genes [81]. Type 1 RM systems have been found in *Streptococcus* species including *S. suis* as regulators of expression of various genes including genes that play roles in bacterial pathogenicity [81,201,202].

In this study, we investigated the activities of two Type I RM systems present in the genome of *S. suis* strain P1/7, an isolate that belongs to sequence type (ST) 1, the lineage considered the most virulent, globally distributed, and responsible for most reported human cases. The first methylation system, hereafter referred to as MS1, was discovered in a genetic lineage primarily composed of pathogenic ST1 strains isolated from human systemic infections [86] and contains the genes *hsdR*, *hsdM,* and two *hsdS* copies. In the P1/7 *hsd* operon, one *hsdS* gene is transcribed in the same direction as *hsdM* and *hsdR*, while the second *hsdS* copy is transcribed in the opposite direction but appears to be transcriptionally silent[86]. Both *hsdS* genes contain inverted repeat sequences, enabling the bacterium to undergo site-specific recombination and exchange of different *hsdS* genes, generating a total of four distinct *hsdS* variants[86](Figure 1A). Studies of analogous phase-variable Type I RM systems in *S. pneumoniae* have shown that phase locked variants having different *hsdS* configurations are associated with distinct epigenetic methylation profiles, changes in gene expression, and altered virulence phenotypes [81]. These findings suggest that the four MS1 phase variants in *S. suis* may similarly contribute to variability in gene regulation and traits relevant to host colonization and pathogenicity.

A second Type I RM system (MS2) is present in the genome of most *S. suis* ST1 strains (Figure 2), hereafter referred to as MS2. It consists of a single gene copy of *hsdR, hsdM* and *hsdS.* The *hsdS* gene locus exists in two distinct forms among *S. suis* strains: a full-length version that is commonly found in genomes of disease-associated lineages (ST1) and a truncated variant that contains only one of the target

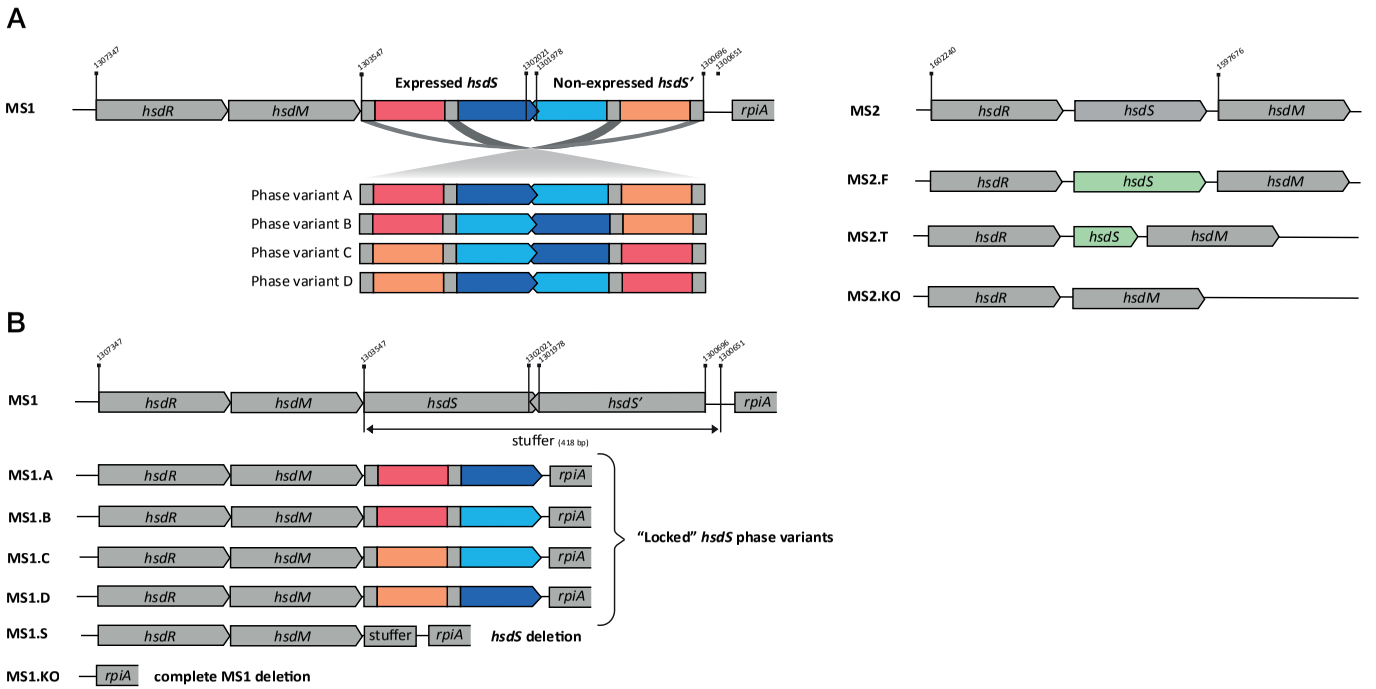


Figure 1. (A) Genomic organization of the MS1 locus in *S. suis* P1/7. The *hsdS* gene encodes two target recognition domains (TRDs). Four distinct TRDs are present in the genome, each represented in a different colour. Site-specific recombination between these TRDs generates four MS1 phase variants. Also shown are a mutant lacking the *hsdS* gene (ms1.s) and a mutant with a complete deletion of the entire MS1 locus (ms1.ko). (B) Genomic organization of the MS2 locus in *S. suis* P1/7. Shown are the full-length and truncated variants of MS2, as well as a mutant lacking the *hsdS* gene. In both panels, the loci are depicted in their chromosomal context in the *S. suis* P1/7 genome, with their corresponding genomic positions indicatedt.

recognition domains that recognizes a bipartite DNA sequence (Figure 1B)[87]. Methylation systems with truncated *hsdS* have also been found in other pathogenic bacteria, such as *Haemophilus parasuis where the* truncated *hsdS* contributes to phase variation and host immune evasion[86,87]. The consistent presence of the full-length *hsdS* variant in one of the most highly pathogenic and zoonotic *S. suis* lineages (ST1) suggests it may confer epigenetic regulatory traits that enhance *S. suis* virulence.

Bacteria often integrate multiple regulatory inputs to modulate gene expression[203], but crosstalk between co-existing DNA methylation systems has not been explored. This study investigates whether such interactions occur between the two Type I RM systems in *S. suis* P1/7, and which traits or genes are regulated by each RM system upon targeted deletion of specific RM genes.

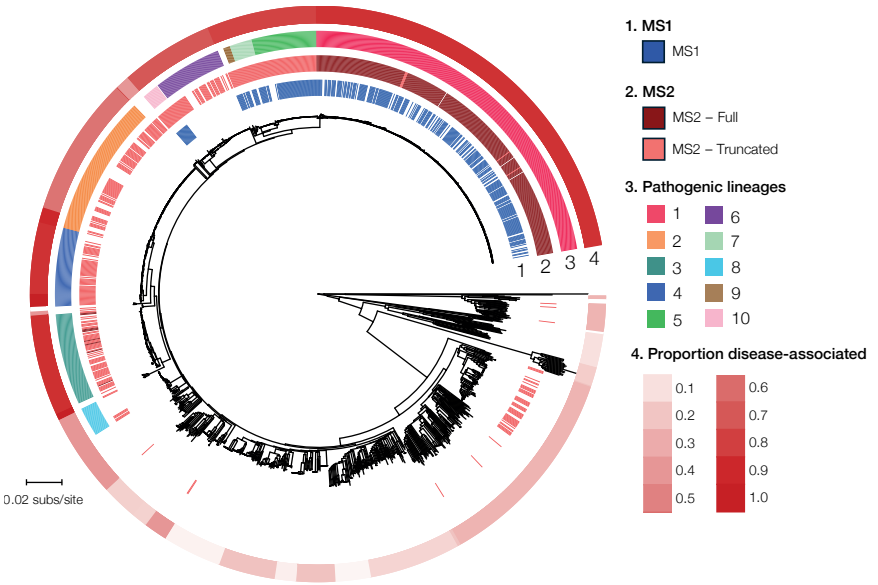
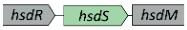
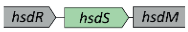
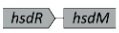
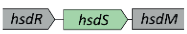
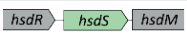
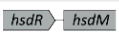
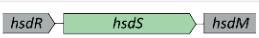
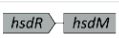


Figure 2. Phylogenetic distribution of MS1 and MS2 methylation systems across the *S. suis* population. The circular tree is annotated with four concentric rings: ring 1 shows the presence of MS1 (blue); ring 2 indicates MS2 full-length (dark red) and truncated (light red) variants; ring 3 highlights the major pathogenic lineages (coloured blocks); and ring 4 depicts the proportion of disease-associated isolates (heat scale).

To investigate the regulatory roles of the two Type I RM systems of *S. suis* P1/7, we generated a panel of 18 *S. suis* P1/7 mutant strains (Table 1), each carrying locked phase-variants of the *hdsS* gene in MS1 (MS1 A to MS1D) or a deletion of *hdsS* or a complete deletion of MS1 Figure 1A. This design enabled us to examine the specific regulatory effects of all MS1 *hsdS* phasevarions, in (i) the absence of MS2 (ii) the presence of MS2 carrying a full length *hsdS* gene and (iii) the presence of MS2 carrying a truncated *hdsS* (Figures 1A, 1B and Table 1). Our transcriptomic and phenotypic analyses revealed, differences in gene expression for mutants carrying truncated and full-length MS2 variants in absence of MS1, suggesting functional interaction between the two RM systems.

Table 1. Assigned names of mutants and corresponding schematic representations of the MS1 and MS2 methylation systems in *S. suis* P1/7.

Mutant name	Methylation System 1	Methylation System 2
ms1.a_ms2.t		
ms1.a_ms2.f		
ms1.a_ms2.ko		
ms1.b_ms2.t		
ms1.b_ms2.f		
ms1.b_ms2.ko		
ms1.c_ms2.t		
ms1.c_ms2.f		
ms1.c_ms2.ko		
ms1.d_ms2.t		
ms1.d_ms2.f		
ms1.d_ms2.ko		
ms1.s_ms2.t		
ms1.s_ms2.f		
ms1.s_ms2.ko		
ms1.ko_ms2.t		
ms1.ko_ms2.f		
ms1.ko_ms2.ko		

For MS1, the two TRD domains of hsdS are shown, with each of the four distinct TRDs represented in a different colour. The diagrams illustrate the genomic operon modifications in each mutant.

Materials and Methods

Identification of type I RM systems in a collection of *S. suis* genomes

We searched for these two MS systems in a previously described collection of draft genome assemblies of *S. suis* strains, consisting of isolates from Germany (133), Denmark (148), the Netherlands (145), Spain (124), the UK (440), China (197), Canada (197), and Vietnam (190) [4]. These isolates were largely from pigs (both clinical and non-clinical samples), but also included a small number of clinical isolates from humans and non-clinical isolates from wild boar.

We used the description of the MS1 system from Atack *et al.* (2018) [86]: *S. suis* P1/7 (SSU1271 - SSU1274; 1300696-1307346) and the MS2 system from Wileman *et al.* (2019) [87]: *S. suis* P1/7 (SSU1589; 1597720-1598925). We extracted these regions and searched for similar regions in the draft assemblies in our data set using *blastn*. Due to frequent rearrangements within the MS1 system, we searched for the four genes that constitute the system individually. We scored the MS1 system as present if each component gene was present with >80% nucleotide identity across >90% of the length, and the genes were all within 10kb of one another.

The MS2 system is sometimes present in a truncated form, where the first half of the gene is present and the second half absent. We scored the system as fully present if the gene was present with >80% nucleotide identity across >90% of the length. We scored it as truncated if the first half of the gene was present, but the second half absent (40-70% of the gene remained, with matches beginning within the first 150 bases of the full length gene, and the ending with first 150 bases of the midpoint of the gene).

Mutant design and construction

All mutants described in this work were generated using the CRISPR-Cas9 method based on the pSStarget plasmid following the protocol by Gussak et al. [127]. Briefly, *S. suis* was transformed using a target-specific pSStarget plasmid and a repair template that can be integrated in the genome via homologous recombination. Incorporation of the repair template in the genome removes the sequence targeted by Cas9, allowing the use of pSStarget for sequence-specific counter selection [127]. The repair templates comprised the sequence to be inserted flanked by 1000 bp regions upstream and downstream of the targeted locus to allow for homologous recombination by the P1/7 native recombination machinery. To make deletions, repair templates consisted of only the sequences flanking the locus targeted for deletion. After transformation, selected colonies were re-streaked on selective agar plates (THY with 5μg/mL chloramphenicol) and the genomic edit verified by colony PCR

and sequencing. Colonies with the verified edit were then grown in liquid culture without antibiotic to cure the plasmid.

Repair templates were constructed by joining two PCR amplicons using splicing by overlap-extension PCR. For the construction of repair templates with three fragments, Hifi assembly (NEB) was used to assemble a plasmid using pUC57-kan as backbone. After transformation into *E. coli* NEBturbo, the linear repair template was amplified from the assembled plasmids using primers annealing to the outer ends of the homologous flanks. The repair template encoding the truncated *hsdS* variant of MS2 was ordered as a synthetic DNA fragment (GeneArt, Thermo Fisher). The cloning of single guide RNA (sgRNA) into pSStarget was performed as described in the original publication [127]. All PCRs were performed using Q5 polymerase (NEB). All primers and sgRNAs used for the construction of these strains are listed in Table S1.

To generate the MS1 set of mutants (ms1.a, ms1.b, ms1.c, ms1.d and ms1.s), we first constructed a strain in which *hsdS* and *hsdS'* were replaced by a "stuffer" sequence - a short non-coding DNA fragment amplified from the commonly used pUC19 plasmid, resulting in ms1.s. This intermediate served as the recipient for subsequent construction of "locked" phase variants, each encoding one of the four possible *hsdS* alleles (ms1.a, ms1.b, ms1.c and ms1.d). Repair templates for these alleles were generated by amplifying the corresponding *hsdS* variant sequences from previously described plasmids (kindly provided by Dr. John Atack, Griffith University). To insert these variants (ms1.a, ms1.b, ms1.c and ms1.d) into the *hsdS* locus, we used a CRISPR-Cas9 targeting plasmid (pSStarget) designed to cut the stuffer sequence, followed by homologous recombination with the appropriate repair template. This approach was necessary because each *hsdS* variant contains the Cas9 target sequence used for deletion of the wild-type *hsdS*, precluding their direct use as templates during the initial knockout step.
A clean MS1 knockout strain (ms1.ko) was generated using a similar strategy, employing a repair template composed of the flanking regions of the entire MS1 locus and the same pSStarget used for ms1.s construction, resulting in deletion of all MS1 locus genes.
These six MS1 strains (ms1.a–d, ms1.s and ms1.ko) were then used as parental strains for MS2 system editing. For this, each strain was transformed with a second pSStarget plasmid specific to the second target recognition domain (TRD) of the MS2 *hsdS* gene, along with a repair template encoding either a truncated *hsdS* variant or a complete deletion of the *hsdS* gene.

Growth assays

The 18 mutant strains (Figure 1, and see above) were cultured overnight in Todd-Hewitt medium (THY; Thermo Fisher Scientific™, CM0189) supplemented with 0,2% (w/v) of

yeast extract (Thermo Fisher Scientific™, 212750). The optical density at 600 nm (OD_{600}) of the overnight cultures was measured to ensure all cultures reached an OD_{600} of approximately 0.8. Cultures were then pelleted by centrifugation at 4,000 rpm for 10 minutes, the supernatant was discarded, and the bacterial pellets were resuspended in phosphate-buffered saline (PBS) to the same volume as the discarded supernatant.
For growth assays, 2.5 µL of the resuspended bacterial cultures were inoculated into 197.5 µL of either fresh THY medium or a chemically defined medium (CDM), prepared according to the composition listed in Table S1 (Chapter II), using galactose (Sigma-Aldrich, G0750) instead of glucose as the carbon source. Negative controls consisting of uninoculated THY and CDM were included in all experiments. The OD_{600} was measured every 15 minutes using a SpectraMax® M3 Multi-Mode Microplate Reader (Avantor, Radnor, PA, USA) for at least 15 hours at 37°C without shaking. Growth of each strain was assessed in duplicate, and all experiments were performed independently three times.

RNA Extraction

The 18 mutant strains were grown to exponential phase (approximately OD_{600} = 0.3) in THY broth at 37°C with 5% CO_2. A 10 mLaliquot of each culture was pelleted by centrifugation at 4,000 rpm for 10 minutes, the supernatant was discarded, and the pellet was snap-frozen in liquid nitrogen before being stored at -80°C overnight.
RNA was extracted from the pelleted cells using the RNeasy Mini Kit (Qiagen) following the manufacturer's instructions with specific modifications. Bacterial pellets were resuspended in 700 µL of RLT buffer containing 0.1% β-mercapto-ethanol and transferred to lysing matrix B 2 mLtubes (MP Biomedicals). Bacterial lysis was performed using the FastPrep-24™ 5G bead beating grinder and lysis system with the following settings: 4.0 m/sec, All-Metal Quickprep adapter, 40 seconds. The tubes were centrifuged for 1 minute at 10,000 rpm, and the supernatant was transferred to a clean Eppendorf tube. To ensure complete DNA removal, an additional DNase treatment step was included using the RNase-Free DNase Set (Qiagen, Cat. No. 79254), following the manufacturer's protocol. Subsequent RNA purification steps were carried out according to the standard Qiagen protocol.
Final RNA concentrations were measured using the Qubit RNA Broad Range Kit (Invitrogen) and a Qubit 4 fluorometer (Invitrogen). To assess RNA integrity, RNA integrity numbers (RIN) were determined for 15 randomly selected samples from each RNA extraction experiment using a Bioanalyzer (Agilent). Each mutant strain was tested in three biological replicates, with RNA extracted independently in three separate experiments.

Data analysis

Growth curves were generated from raw OD_{600} data, preprocessed by converting time measurements to hours and filtering the dataset to include the first 8 hours for THY and 10 hours for CDM. Growth curves were plotted using the Python (v. 3.8.18) visualization library Matplotlib (v. 3.6.2), with error bars representing standard deviation across replicates. The area under the curve (AUC) for bacterial growth was calculated to assess differences in growth dynamics across the 18 mutant strains. Growth data were preprocessed by converting time measurements to minutes and filtering the dataset to include time points up to 8 hours for THY and up to 10 hours for CDM. The AUC was computed using the trapezoidal rule, AUC values were determined for each replicate and averaged. The following statistical analyses were performed to assess differences in AUC values among strains. Data normality was evaluated using the Shapiro-Wilk test. If the data were sampled from a normal distribution, one-way ANOVA was applied, followed by Tukey's post hoc test for multiple comparisons. If normality assumptions were not met, a Kruskal-Wallis test was used, followed by Dunn's multiple comparisons test. A significance threshold of $p<0.05$ was applied in all analyses. Statistical analyses were conducted using Python (v. 3.8.18) using pandas v.2.0.3, numpy v.1.24.3, scipy v.1.10.1, statsmodels v.0.14.0, and scikit-posthocs v.0.7.0.

RNA sequencing data was processed as follows. Raw paired-end FASTQ files were first aligned to the *S. suis* P1/7 genome (RefSeq: GCF_000091905.1) using Bowtie2, generating SAM alignment files for each sample. The SAM files were then converted to BAM format using samtools version 1.13 with default options to optimize storage and processing efficiency. BAM files were subsequently sorted and indexed to facilitate downstream analysis. Gene-level read counts were obtained using featureCounts, specifying the option to allow for multi-mapping reads (-O), the -M --fraction options to distribute multi-mapped reads proportionally, and the option to count reads from paired-end sequencing (-p). A minimum threshold of 10 mapped reads per gene was applied to filter out low-abundance transcripts before further analysis. The resulting count matrices were merged across all samples for further statistical evaluation. Differential gene expression analysis was performed using DESeq2 (v. 1.46.0) in R 4.4.2, correcting for replicate variation. The final output included normalized gene expression values, log2 fold changes, and adjusted p-values (Benjamini-Hochberg FDR correction) for all comparisons.
Partial constrained redundancy analysis (RDA) was performed using the Canoco software (v. 5.15). Gene expression data were first normalized using Reads Per Kilobase of transcript per Million mapped reads (RPKM) to adjust for gene length and sequencing depth. To further reduce the influence of gene-specific baseline expression levels, expression values for each gene were then normalized to their average expression across all samples.

The RDA was constrained by strain genotype, with biological replicate variability statistically controlled. All heatmaps were generated using R v.4.4.2 to construct the expression matrices, Python v.3.10 for visualization and design, using the libraries pandas v.1.5.3, numpy v.1.23.5, and plotly v.5.13.0. Adobe Illustrator v.29.5.1 was used for final figure assembly.

Results

Conservation of two Type I RM system in *S. suis*

The first Type I RM system we investigated is Methylation System 1, hereafter referred to as MS1, exhibits phase variation via site specific recombination events between two *hsdS* genes resulting in four distinct *hsdS* specificity variants (Figure 1, Table 1). The second Type 1 RM system is Methylation System 2, hereafter referred to as MS2, is a non-phase-variable methylation system present in most pathogenic strains. However, the *hsdS* gene in MS2 is only full-length (ms2.f) in the most virulent lineage, which is associated with invasive disease in both pigs and humans. In other pathogenic strains the *hsdS* gene contains one target recognition domain (TRD), hereafter referred to as truncated *hsdS* (ms2.t) (Figure 1). Since both TRD domains are required to position the methylase dimer accurately on DNA, the truncated MS2 likely confers altered or limited methylation specificity [86].

To investigate the genomic distribution of these two MS systems, we analyzed their conservation across a published global *S. suis* phylogeny from disease and non-disease contexts [4] . As shown in Figure 2, MS1 is conserved in several highly pathogenic lineages, including lineages ST1, ST16, ST17 and ST87. It is absent in highly diverse commensal population of *S. suis* (Figure 2). The frequency of disease-associated isolates is visualized in Figure 2 as a shaded heatmap (red = 100%, white = 0%). These patterns suggest MS1 may have been acquired by one or more pathogenic lineages, via horizontal gene transfer and could potentially provide a selective advantage for within-host growth or transmission. The second MS system, MS2 (SSU_1588–1590), carries a full-length *hsdS* allele (ms2.f) only in lineage 1, the most virulent lineage associated with invasive disease in both pigs and humans (Figure 2). The truncated version of MS2 is common in other pathogenic lineages and occasionally present in commensal and divergent lineages (ms2.t). This raises the question of whether MS1 and MS2 interact functionally in strains where both are present, (i.e. in ST1) and whether this contributes to the high virulence of this lineage. To explore potential regulatory interactions between MS1 and MS2, we constructed a panel of 18 *S. suis* P1/7 mutant strains (Table 1), each with defined combinations of locked MS1 phase variants (A–D), MS1 deletions (Δ*hsdS* or complete MS1 knockout),

and MS2 variants (full-length, truncated, or deleted *hsdS*) (Figure 1, Table1). This strain panel enabled us to assess the individual and combined regulatory impacts of MS1 and MS2 on global gene expression and bacterial growth under defined conditions.

In summary, MS1 is a phase variable type I RM system conserved in disease-associated lineages while MS2 is a non-phase-variable Type I system present in most pathogenic strains. The full-length form of the *hsdS* gene in MS2 is uniquely present in the highly virulent zoonotic lineage ST1 (Figure 2).

Global transcriptomic effects of MS1 and MS2 combinations

To evaluate the transcriptomic effects of MS1 and MS2 we constructed 18 distinct genetic variants of MS1 and MS2 in *S. suis* that occur in nature (Table 1). We performed pairwise differential gene expression (DGE) analyses across all strains, yielding a total of 153 comparisons. To visualize the global structure of transcriptomic variation across all strains, we filtered the differentially expressed genes (DEGs) based on an adjusted p-value < 0.05 and a $\log_2$FC > 1 or < -1 and generated a matrix showing the number of DEGs for each pairwise comparison (Figure 3). Pairwise comparisons of the four locked MS1 phase variants (MS1 A, MS1 B, MS1 C, and MS1 D), each encoding a distinct *hsdS* variant, in the context of an intact MS2 system revealed limited transcriptomic differences, with few DEGs detected (from 0 to 18 DEGs per comparison). These findings suggest that although the MS1 system is phase-variable, its individual *hsdS* variants exert relatively subtle effects on the transcriptome.

Notably, the transcriptome of the MS1 deletion strains (ms1.ko_ms2.f, ms1.ko_ms2.t) diverged markedly from the other strains (with 13–91 DEGs), including ms1.ko_ms2.ko, suggesting that the presence of MS2 is required for this transcriptional change. Interestingly, truncated MS2-*hsdS* (ms2.t) caused strong increases in the number of DEGs in the background of MS1 B variant but not the MS1 A, C, or D variants (Figure 3). In contrast, no significant transcriptomic differences were observed between ms1.s (deletion of *hsdS*) and the locked MS1 variants. These results indicate that complete deletion of the MS1 system (ms1.ko) rather than partial disruption (ms1.s) is required to reveal the full extent of MS2 activity, in both truncated and full-length forms.

To confirm that the biggest transcriptome changes cluster around ms1.ko_ms2.t, ms1.ko_ms2.f and ms1.b_ms2.t we assessed how much the specific combinations of MS1 and MS2 configuration contributed to the global transcriptomic variation using RDA (Figure 4). The resulting RDA plot confirmed that most strain combinations clustered tightly, reflecting broadly similar transcriptomic profiles, whereas four strains (ms1.b_ms2.t, ms1.ko_ms2.t, ms1.ko_ms2.f and ms1.ko_ms2.ko) separated clearly

from the main cluster. Therefore, removing MS1 entirely (ms1.ko), having a truncated or full length MS2 (ms2.t or MS2.f) or the MS1.b variant together with ms2.t explain most of the variation across the dataset, suggesting that MS1 normally constrains the effects of transcription on MS2. Furthermore, the combination of ms1.ko_ms2.f produced the most distinct transcriptome changes and deletion of only MS1 *hsdS* is not sufficient to reveal the strong effects of MS2 (see ms1.s_ms2.t and ms1.s_ms2.f). Together these results show an epistatic interaction with the transcriptomic landscape being shaped by MS1-MS2 interactions, not the independent effect of one system alone, with the strongest effects seen in ms1.ko (especially with ms2.f), followed by ms1.ko with ms2.t, and to a lesser extent ms1.b with ms2.t and ms1.ko with ms2.ko.

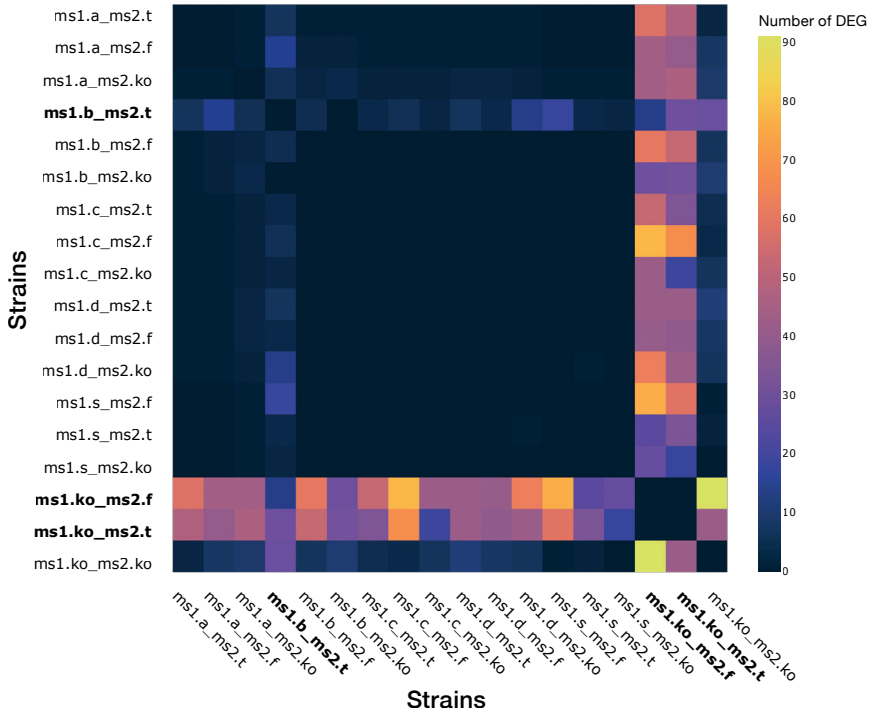


Figure 3. Heatmap of DEGs Across *S. suis* Methylation System Mutant Pairs (adj. p-value < 0.05, |logFC| > 1, excluding MS1 and MS2 genes.

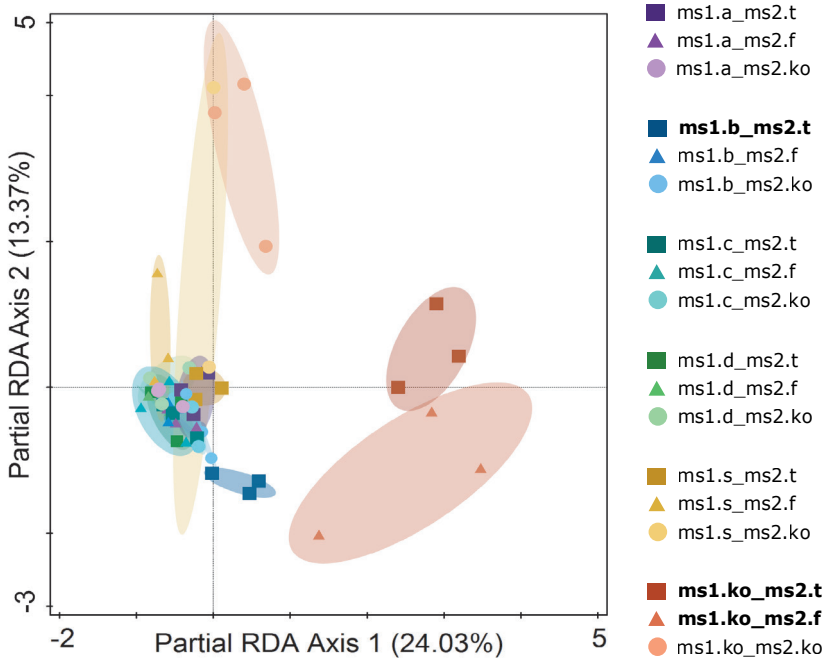


Figure 4. Partially constrained redundancy analysis (RDA) of transcriptomic profiles across all 18 *S. suis* P1/7 MS1-MS2 locked mutant strains. Most strains clustered tightly reflecting broadly similar transcriptome profiles, whereas four outlier combinations (ms1.ko_ms2.f, ms1.k_ms2.t, ms1.b_ms2.t, and ms1.ko_ms2.ko) separated from the main cluster. The strongest transcriptomic shifts were observed in ms1.ko_ms2.f, followed by ms1.ko_ms2.t, and to a lesser extent ms1.b_ms2.t and ms1.ko_ms2.ko. The RDA was constrained by genotype, controlling for replicate variability. Each point represents a biological replicate, and colours/symbols correspond to specific MS1–MS2 genetic combinations. Shaded ellipses represent 95% confidence intervals around each genotype. The first two RDA axes explain 24.03% and 13.37% of the variance, respectively.

Key mutants drive global changes in carbohydrate and amino acid metabolism

As a result of the findings shown in figures 2 and 3, we identified three key mutants that have pronounced transcriptional changes compared to the rest of mutants: ms1.ko_ms2.t, ms1.ko_ms2.f, and ms1.b_ms2.t. To identify the DEGs (log2FC >1 or < -1 and adj. p-value < 0,05) that explain the separate clustering of the key mutants (Figure 3). Each key mutant was compared to the other 17 in pairwise comparisons, selecting only genes that were differentially expressed in at least 8 of the 17 comparisons (Figure 5).

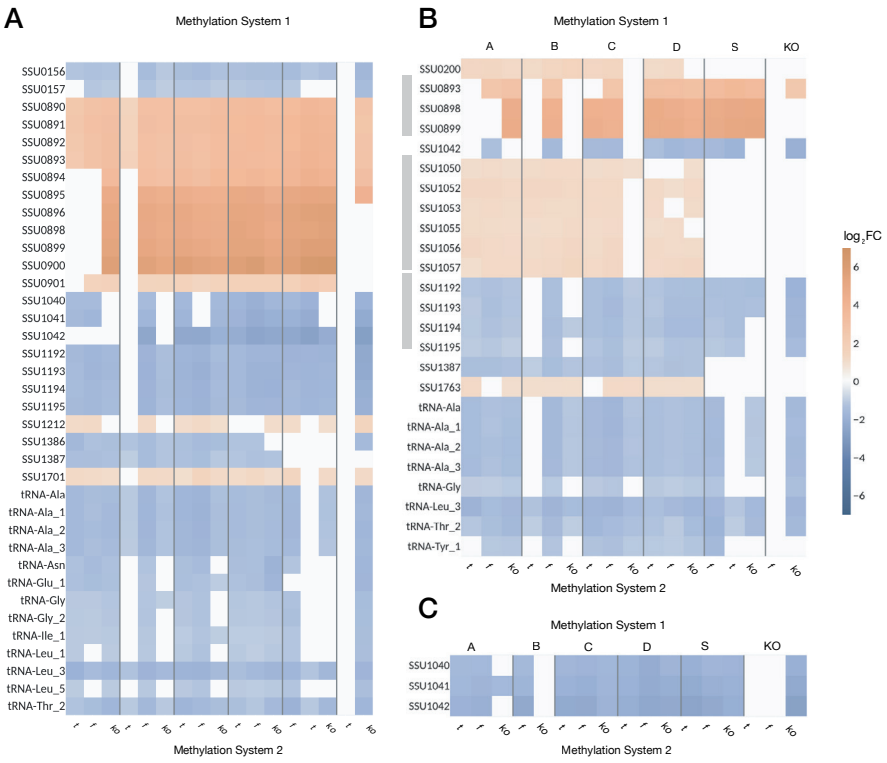


Figure 5. Differential gene expression heatmaps. Each panel shows the $\log_2$ fold change (logFC) in gene expression for one mutant strain compared against all other strains. (A) ms1.ko_ms2.f, (B) ms1.ko_ms2.t, and (C) ms1.b_ms2.t. Heatmaps display genes that were significantly differentially expressed (adjusted $p < 0.05$ and $\log_2FC > 1$ or < -1). Expression values are scaled by log2FC; shades of orange indicate upregulation and shades of blue indicate downregulation. Genes belonging to the same operon are indicated by grey sidebar markers.

For ms1.ko_ms2.f this analysis revealed a cluster of 10 genes in the *S. suis* genome (SSU0890–SSU0901) that exhibited consistent overexpression in across the majority of the pairwise comparisons during culture in THY (Figure 5A). These 10 genes showed a $\log_2FC$ up to 6.55 in most of the mutants. The SSU0890-SSU0901 gene cluster encodes a set of proteins essential for carbohydrate transport and metabolism in *S. suis* (Table 2). Annotation from the NCBI and KEGG databases indicated that the proteins encoded within the cluster include two subunits of the lactose phosphotransferase system (PTS) (SSU0892 and SSU0893), seven enzymes involved in catabolic pathways (SSU0890, SSU0891, SSU0894 – SSU0899) and a DeoR/GlpR family transcriptional regulator (SSU0901) that potentially regulates the other genes in

the cluster [204]. According to KEGG database, the 10 cluster genes are part of the galactose/tagatose metabolism pathway, dedicated to the uptake and degradation of lactose and galactose. Together, these findings indicate that the ms1.ko_ms2.f consistently down-regulates a lactose PTS operon and associated metabolic genes, suggesting that removal of MS1 reduces carbohydrate uptake and catabolism except in comparison to ms1b_ms2.t.

Table 2.

Gene ID	Name	Product
SSU0890		Aldose 1-epimerase family protein
SSU0891	*lacG*	6-phospho-beta-galactosidase
SSU0892	*lacE*	Lactose-specific PTS transporter subunit EIIC
SSU0893	*lacF*	PTS lactose/cellobiose transporter subunit IIA
SSU0894		PRD domain-containing protein
SSU0895	*lacD*	Tagatose-bisphosphate aldolase
SSU0896		Tagatose-6-phosphate kinase
SSU0898	*lacB*	Galactose-6-phosphate isomerase subunit LacB
SSU0899	*lacA*	Galactose-6-phosphate isomerase subunit LacA
SSU0901		DeoR/GlpR family DNA-binding transcription regulator
SSU1055		PTS system mannose/fructose/sorbose family transporter subunit IID
SSU1056		PTS mannose/fructose/sorbose/N-acetylgalactosamine transporter subunit IIC
SSU1057		PTS system mannose/fructose/N-acetylgalactosamine-transporter subunit IIB
SSU1192		Amino acid ABC transporter ATP-binding protein
SSU1193		Transporter substrate-binding domain-containing protein
SSU1194		Amino acid ABC transporter permease
SSU1195		Amino acid ABC transporter permease
SSU1040		NADP-dependent isocitrate dehydrogenase
SSU1041		Citrate synthase
SSU1042		Aconitate hydratase AcnA

In ms1.ko_ms2.t (Figure 5B), three genes in the above mentioned 10 gene cluster: SSU0893, SSU0898, and SSU0899 were consistently upregulated in most of the other mutants compared to the ms1.ko_ms2.t. These genes encode PTS lactose transporter subunit IIA, galactose-6-phosphate isomerase subunit LacB, and galactose-6-phosphate isomerase subunit LacA, respectively. Moreover, a different

gene cluster, SSU1055–SSU1057 showed significant (adj. p-value < 0,05) upregulation (logFC between 1 and 1.7) across all mutants carrying the *hsdS* variants of MS1 (ms1.a, ms1.b, ms1.c and ms1.d) but not in ms1.s variant, (lacking the *hsdS* of MS1) or in the ms1.ko mutants. The genes SSU1055–SSU1057 encode a carbohydrate transport system (Table 2) and three PTS components (SSU1055-SSU1057) which are predicted to transport mannose, fructose and sorbose (NCBI Gene annotation).

When analysing significantly (adj. p-value < 0,05) downregulated genes (logFC ≤ –1) in all strains relative to both ms1.ko_ms2.f and ms1.ko_ms2.t (Figure 5A and 5B), we found that the same set of genes consistently showed reduced expression. Notably, two gene clusters, SSU1192–SSU1195 and SSU1040–SSU1042, as well as several tRNA genes, were consistently downregulated compared to ms1.ko_ms2.f and ms1.ko_ms2.t.

Gene cluster SSU1192-SSU1195 encodes proteins involved in amino acid transport systems, while the genes SSU1040–SSU1042 encode the enzymes NADP-dependent isocitrate dehydrogenase, citrate synthase, and aconitate hydratase (AcnA). These three enzymes are key components of the incomplete tricarboxylic acid (TCA) cycle, functioning in a series of interconnected reactions that begin with oxaloacetate and culminate in the production of 2-oxoglutarate, a precursor for glutamate. Notably, SSU1040–SSU1042 is the only cluster significantly and consistently upregulated in all three key mutants (Figure 5).

Together, these results suggest that truncated or full-length variants of the *hsdS* gene in MS2 interact to reprogram metabolism and increase tRNAs and other genes related to protein synthesis and the TCA pathway in *S. suis* in the absence of MS1. In the ms1.b_ms2.t locked variant, only the TCA pathway is upregulated.

Growth phenotypes reflect transcriptomic differences in central metabolism and protein synthesis

Following the identification of key operons responsible for the differential gene expression between ms1.ko_ms2.f, ms1.ko_ms2.t and ms1.b_ms2.t compared to other strains, we sought to determine whether these transcriptional differences resulted in phenotypic variations. Because the genes were annotated with functions in growth and central metabolism, we conducted growth curve analyses for all 18 mutants in THY medium, the same medium used for the transcriptomic analyses. In addition, we performed growth analyses in CDM with galactose as the sole carbon source to assess whether growth fitness was depended on the available carbon source and to determine if the mutants could adapt to different nutrient contexts. To facilitate comparison of growth dynamics we calculated the AUC for each growth curve as a single measure of overall growth performance.

When cultured in THY medium, the key mutants ms1.ko_ms2.f and ms1.ko_ms2.t exhibited the fastest growth (Figure 6). This observation is consistent with our transcriptomic data showing the upregulation of genes involved in protein synthesis compared to the other mutants. Among the four MS1 locked variants, B exhibited the fastest growth. Variant B in combination with truncated ms2 (ms1.b_ms2.t) showed upregulation of the genes SSU1040-SSU1042, which encode proteins which are part of the incomplete TCA cycle (Figure 5C). The strain lacking the *hsdS* gene from MS1 (ms1.s) showed no differences in growth compared to the other MS1 variants. Among MS1 phase variants (A, B, C, D, and S), variant A showed the slowest growth. No major differences in growth were observed among the different MS2 backgrounds (t, f, and ko). However, the two key mutants ms1.ko_ms2.f and ms1. ko_ms2.t displayed significantly faster growth than ms1.ko_ms2.ko, highlighting the influence of MS2 on modulating growth in the absence of MS1.

We extended the phenotypic analysis by performing growth curve experiments in CDM with galactose as the sole carbon source (Figure 6). The key mutant ms1.ko_ ms2.t had the highest overall growth. In contrast key mutant ms1.ko_ms.2.f displayed similar growth performance as the ms1.ko_ms2.ko (Figure 6).

Interestingly, both MS1 variants B and D displayed more growth compared to variants A, C, and S, which in turn displayed very similar growth levels to each other (Figure 6). Only in CDM, MS1 variants A, C, S showed differences in overall growth depending on the MS2 genotype (ms2.ko, ms2.f, and ms2.t) (Figure 6). In these variants, ms2.t showed the least growth, followed by ms2.f and the most growth in ms2.ko (Figure 6).

Subtle transcriptomic differences among MS1 phase variants underlie growth variation

Next, we aimed to investigate if wether the subtle differences in gene expression among the MS 1 variant could account for the slightly lower overall growth of variant A and the higher overall growth of variant B. We performed a partial-constrained RDA focusing on differences between mutants excluding variation between replicates (Figure 7A). We studied MS1 phase variants a–d combined with either full-length (ms2.f) or truncated (ms2.t) MS2 variants.

The resulting RDA plot showed that (ms1.b_ms2.t) separated clearly from the other clusters. Genes contributing most to this separation were involved in anabolic processes such as in the TCA cycle and protein synthesis. Heatmap of gene expression confirmed the upregulation of genes associated with incomplete TCA cycle, ribosomal RNA operons (16S and 23S variants), and genes involved in amino acid transport in ms1b_ms2.t strain (Figure 7B).

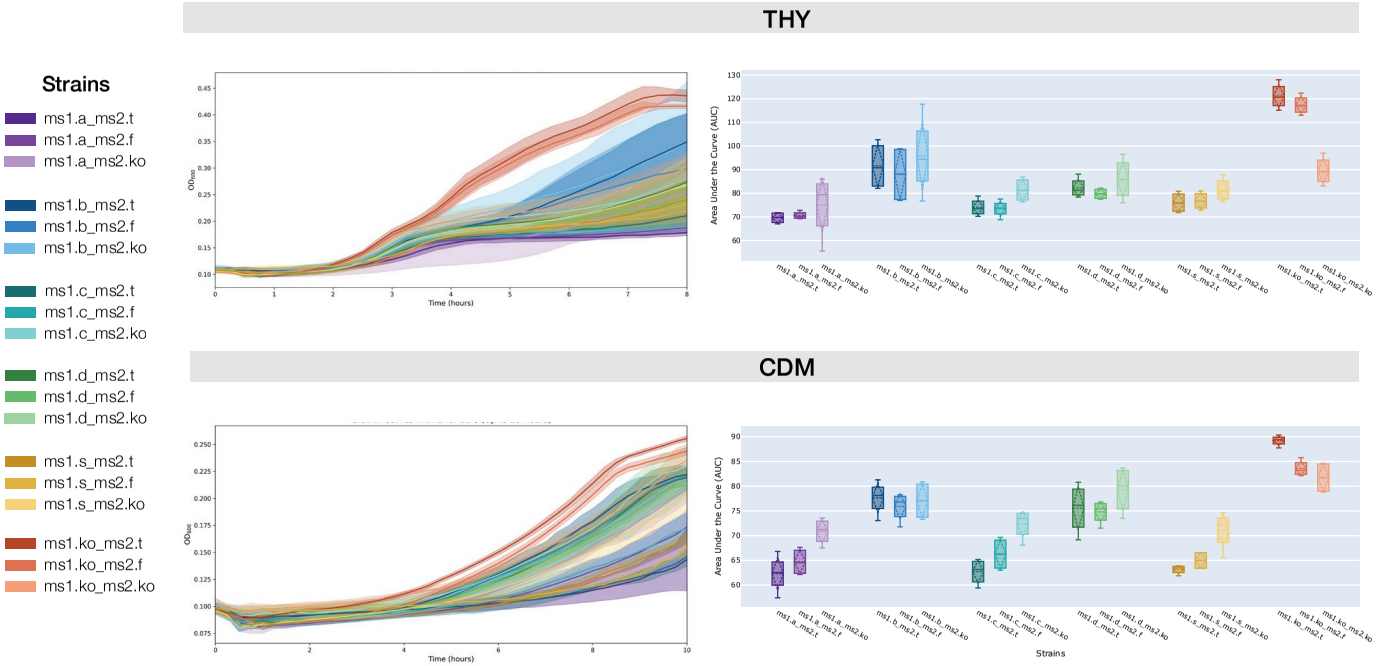


Figure 6. Growth analysis of *S. suis* RM system variants in THY and chemically defined medium (CDM with galactose). Left panels show growth curves over time, with mean ± SD (bold lines and shaded area). Right panels show area under the curve (AUC) distributions for 18 MS1/MS2 combinations. Color-coded boxplots represent MS1 phase variants A (purple), B (blue), C (green/blue), D (green), the ms1.S "stuffer" mutant (yellow), and MS1 ko (orange/brown). Across both media, MS1 knockout strains consistently grow fastest in all three MS2 backgrounds (wild-type, truncated, and knockout). MS1.B variants grow faster than other MS1 phase variants, but not as fast as the full MS1 knockout. MS1.A and ms1.S exhibit the slowest growth. Statistical differences among strains were tested using ANOVA or Kruskal-Wallis as appropriate (see Methods). Adjusted p-values for pairwise comparisons are provided in Supplementary Table S2 (THY) and Supplementary Table S3 (CDM-galactose).

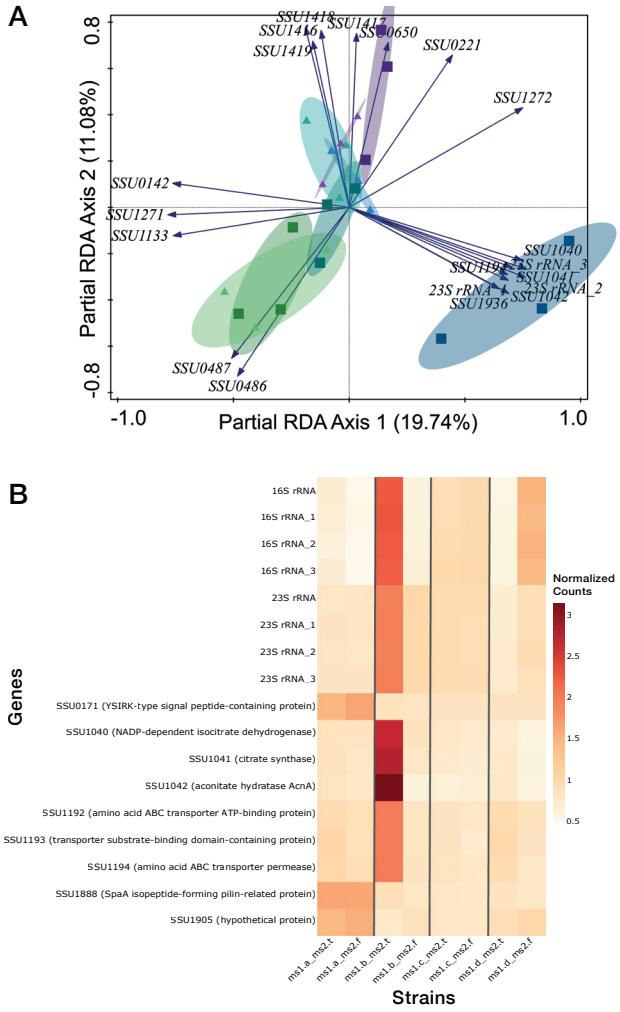


Figure 7. (A) Partial constrained redundancy analysis (RDA) of gene expression profiles from strains carrying MS1 phase variants a–d combined with either wild-type (ms2.f) or truncated (ms2.t) MS2 variant. The RDA was constrained by genotype, controlling for replicate variability. Each point represents a biological replicate, with ellipses indicating 95% confidence intervals for each genotype. Arrows represent genes that contribute most strongly to sample separation along the RDA axes. The first two axes explain 19.74% and 11.08% of the total constrained variance, respectively. (B) Heatmap of normalized gene expression across the same genotypes shown in panel A, showing normalized count values of genes that were significantly differentially expressed in at least one pairwise comparison (adjusted $p < 0.05$) with a $\log_2$FC greater than 1 or less than – 1. Normalized counts are scaled and color-coded, with deeper red indicating higher expression.

Gene expression of variants A, C and D were largely similar. Only variant A showed upregulation of pilin-like iso-peptide-forming protein gene SSU1888 and two hypothetical proteins of unknown function (SSU1905, and SSU0171). Within the gene cluster containing SSU1888, we identified two other genes SSU1889 and SSU1885 that were significantly upregulated (adj. p-value < 0.05), although not exceeding the heatmap threshold of $\log_2$FC >1 or <-1. SSU1889 encodes a SpaA isopeptide-forming pilin-related protein; SSU1885 encodes a protein containing a prealbumin-like fold domain.

Discussion

DNA methylation is a key regulatory mechanism influencing bacterial gene expression and adaptation [81,202]. In *S. suis*, two Type I MS have been identified, SSU1271-SSU1274 and SSU1588-SSU1590, here referred to as MS1 and MS2. Previous studies showed that MS play roles in pathogenicity [20,202], but the impact of MS systems on *S. suis* P1/7 physiology has not been previously reported.

MS1 possesses a genetic switch in the specificity gene (*hsdS*) where one copy of the *hsdS* gene is transcribed in the same direction as *hsdM* and *hsdR*, while the second copy is inverted and remains silent. This arrangement, coupled with the presence of inverted repeat sequences in both *hsdS* genes, enables site-specific recombination, allowing for the exchange of different *hsdS* genes. This genetic switch results in four distinct *hsdS* phase variants, which in *S. suis* are designated as a, b, c, and d [86]. MS2, on the other hand, is a methylation system that contains a specificity protein of which two variants have been identified in different *S. suis* isolates: a full-length and a truncated *hsdS* variant (Figure 1). The ms2 full-length *hsdS* is highly conserved in lineage 1 which is the most common cause of invasive disease in pigs and humans[36,87]. A truncated form of *hsdS* in MS2 is more commonly found in other pathogenic lineages (Figure 2) [87].

Our transcriptomic and phenotypic analysis of *S. suis* P1/7 mutant strains, containing various combinations of both Type I methylation systems (MS1 and MS2), reveals for the first time that two methylation systems interact epistatically to shape gene expression and bacterial physiology. Furthermore, to our knowledge, this is the first study exploring the possible interaction between two MS and impact of each system on gene expression and bacterial growth.

Previous studies on MS systems with different *hsdS* phase variants in *M. pulmonis* and *S. pneumoniae* have reported changes in these bacteria upon different phase

variants. In *M. pulmonis*, these changes included altered phage susceptibility, while in *S. pneumoniae*, *hsdS* variants mediated changes in the production of the polysaccharide capsule, nasopharyngeal colonization and evasion of phagocytosis, key processes involved in pathogenesis [205]. Since impact of different MS phase variants might be relevant to *S. suis* physiology, we used "locked" strains, in which recombination between the *hsdS* alleles was prevented by deleting the second inverted *hsdS* gene copy (Figure 1). These locked mutants ms1.a, ms1.b, ms1.c and ms1.d stably expressed a single *hsdS* variant and exhibit distinct effects on bacterial growth. Variant ms1.b confers the greatest growth advantage followed by *ms1.d*, *ms1.c*, and *ms1.a*. This hierarchy aligns with previous findings showing that *ms1.b* is the most prevalent variant in *S. suis* populations, whereas ms1.a is the least common [86]. The poor growth of ms1.a may stem from increased expression of pilus and cell wall anchoring genes which are metabolically costly and only upregulated in this specific phase variant (Table 2). However stochastic occurrence of the ms1.a variant *in vivo*, would lead to pilus production which might play a role in adhesion to host tissues, and biofilm formation, especially during specific stages of infection. Other examples of DNA methylation systems regulating the expression of energetically demanding surface structures such as flagella and fimbriae have been reported for *Salmonella enterica* serovar Typhimurium and *Escherichia coli* [206–208]. Therefore, the reason why variant A may be selectively maintained in pathogenic strains is due to its advantage for host interaction, despite the cost to growth potential.

Our study also examined why ms2 *hsdS* exists in both truncated and full-length variants (ms2.f and ms2.t). Atack et al. 2018 suggested that a truncated HsdS, which contains a single DNA binding-domain can dimerize to form a functional *hsdS* protein [86,199,209]. A full-length HdsS contains two DNA-binding domains in tandem. Interestingly the phenotype of truncated MS2 was strongly context-dependent, with marked transcriptomic changes only in the ms1.b background where genes related to protein synthesis were upregulated. It has been shown that the upregulation of protein synthesis genes has a near-linear relationship with bacterial growth[210,211], supporting the hypothesis that the transcriptional changes in ms1.b_ms2.t underlie the observed faster growth.

In support of an interaction between MS1 and MS2, mutants lacking MS1 while retaining MS2 (ms1.ko_ms2.f and ms1.ko_ms2.t) showed the highest number of DEG (18 to 95) exceeding the ones observed in other strains combinations. These transcription effects are lost when MS2 is deleted (ms1.ko_ms2.ko), indicating that there is an epistatic relationship between both systems. Most DEG in these mutants (ms1.ko_ms2.f and ms1.ko_ms2.t) were related protein synthesis correlating with their fast growth in THY. However, in galactose-CDM, ms1ko_ms2.f has similar growth rate

as the ms1.ko_ms2.ko. This could be explained by the fact that mutant ms1.ko_ms2.f has downregulated genes involved in galactose metabolism. Mutant ms1.ko_ms2.t maintained a growth advantage compared to the rest of mutants.

Downregulation of galactose genes might be detrimental for *S. suis* specially for survival in host environments[212], because galactose is an important component of the capsule [213]. The importance of galactose for capsule synthesis may explain why all strain from lineage 1, which have capsules that include galactose and a full ms2 *hsdS* also maintain MS1. Capsule production becomes more crucial in natural nutrient-limited environments, for example, niches where bacterial competition and host immune mechanisms induce stresses on bacterial growth [214,215].

The mutant ms1.s which was deleted of *hdsS* but not *hsdR* and *hsdM* showed no major growth or transcriptomic differences compared to the other mutants, showing that HsdM or HsdR can also function independently of *hsdS* in the tested conditions.

This study provides novel insights into the functional role of two different Type I MS in *S. suis* and their impact on gene expression and bacterial fitness in model strain P1/7. For MS1 four distinct phase variants of the *hsdS* gene have been observed in *S. suis* isolates although the frequency of switching is not known. Our results suggest , that each phase variant may benefit bacterial growth in different host niches or other environments. Additionally, we demonstrated that *S. suis* P1/7 strains carrying full-length and truncated versions of *hsdS* in MS2 exhibit differential gene expression and growth depending on the phase variant present in MS1. Deletion of *hdrR* or/and *hdrM* from MS1 in the presence of either truncated or full-length MS2 was associated with faster growth, likely due to a reduced metabolic burden.

In conclusion, we show that MS1 and MS2 form a regulatory network with epistatic interaction, influencing gene expression, metabolism and growth. In the context of evolution of pathogenic *S. suis* strains we find that the combination of MS1 and full-length MS2 is highly conserved in the most pathogenic lineage. Our results suggest that this is due to the importance of MS1 to control certain metabolic pathways, which are involved in uptake of galactose for capsule synthesis which is crucial for virulence. In summary, these two methylation systems contribute to the host adaption and the evolutionary trajectory of disease-associated lineages of *S. suis*.

Data availability

All custom scripts, and any other analyzed data supporting the findings of this chapter will be deposited in a public repository upon publication. The raw sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB96394 and will be released upon publication.

Supplementary Data

Table S1 Primers used in this study

Primer Nr	Primer name	Sequence
P620	US_MS1_KO_Fw	ATCATCACCACACCTAAAC
P621	US_MS1_KO_Rv	ctgaaagaggCCTACGTCCCTAAGTTAG
P622	DS_MS1_KO_Fw	gggacgtaggCCTCTTTCAGTATGAATAATAAC
P623	DS_MS1_KO_Rv	CTTGATGTTGACGTATTCTTC
P648	US_MS2_KO_Fw	GATTCCAAGTTCATCCAC
P649	US_MS2_KO_Rv	aggagggactGATTTCGAGTAAAAAATGAAACAAC
P650	DS_MS2_KO_Fw	actcgaaatcAGTCCCTCCTTTTCTACTC
P651	DS_MS2_KO_Rv	TTGAGAGAGTTGACGGGTG
P694	MS2 colonyPCR_Fw	cgatagaacatgaaggctagg
P695	MS2 colonyPCR_Rv	ttggttggtcaataggttgc
P778	pUC57_MS1_Fw	ctttgactgtataTCAGAATCCGAGTGACAG
P779	pUC57_MS1_Rv	tgaatatctttgagtgattcctttacggtgtc
P780	US_MS1_Fw	taaaggaatcactcaaagatattcaaagtgcag
P781	US_MS1_Rv	tatttctcctccaaaatttctaaaatttc
P782	stuffer_MS1.s_Fw	ttttagaaattttggaggagaaatatcatgtaactcgccttgatc
P783	stuffer_MS1.s_Rv	tttgctactaacttagggacgtaggtgagtaaacttggtctgac
P784	DS_MS1_Fw	cctacgtccctaagttag
P785	DS_MS1_Rv	ctcggattctgaTATACAGTCAAAGCTCTCG
P786	MS1.a_Fw	ttttagaaattttggaggagaaataATGACTCCTGAACAACTG
P787	MS1.a_Rv	tttgctactaacttagggacgtaggCTAAAGTAGATTTACTTTTTGAAAAAG
P788	MS1.b_Fw	ttttagaaattttggaggagaaataATGACTCCTGAACAACTG
P789	MS1.b_Rv	tttgctactaacttagggacgtaggTCAGTAATAAAGTTGGGC
P790	MS1.c_Fw	ttttagaaattttggaggagaaataATGACTCCTGAACAACTG
P791	MS1.c_Rv	tttgctactaacttagggacgtaggTCAGTAATAAAGTTGGGC
P792	MS1.d_Fw	ttttagaaattttggaggagaaataATGACTCCTGAACAACTG
P793	MS1.d_Rv	tttgctactaacttagggacgtaggCTAAAGTAGATTTACTTTTTGAAAAAG
P794	MS1 colonyPCR_Fw	tcagcaagaacgagctgtc
P795	MS1 colonyPCR_Rv	ggagatgggacaagagtgag
P813	MS1 repairtemp_Fw	cgaaagactagcgaagacg
P815	MS1 repairtemp_Rv	cctatcatctgactgctacg

sgRNA target	Locus Tag	Sequence
***hsdS* MS1**	SSU1271/1272	TTCTAACGATGAAGAGACCG
stuffer MS1.s		GTAGATAACTACGATACGGG
***hsdS* MS2**	SSU1589	ACTGTAAGATTATGGGGGAG

Table S2 Adjusted p-values for pairwise comparisons of growth (area under the curve, AUC) among S. suis RM system variants in THY medium

	ms1.a_ms2.ko	ms1.a_ms2.t	ms1.a_ms2.f	ms1.b_ms2.ko	ms1.b_ms2.t	ms1.b_ms2.f	ms1.c_ms2.ko	ms1.c_ms2.t
ms1.a_ms2.ko	1	1	1	1	1	1	1	1
ms1.a_ms2.t	1	1	1	0,3818	0,2422	1	1	1
ms1.a_ms2.f	1	1	1	0,5917	0,3818	1	1	1
ms1.b_ms2.ko	1	0,3818	0,5917	1	1	1	1	1
ms1.b_ms2.t	1	0,2422	0,3818	1	1	1	1	1
ms1.b_ms2.f	1	1	1	1	1	1	1	1
ms1.c_ms2.ko	1	1	1	1	1	1	1	1
ms1.c_ms2.t	1	1	1	1	1	1	1	1
ms1.c_ms2.f	1	1	1	1	1	1	1	1
ms1.d_ms2.ko	1	1	1	1	1	1	1	1
ms1.d_ms2.t	1	1	1	1	1	1	1	1
ms1.d_ms2.f	1	1	1	1	1	1	1	1
ms1.ko_ms2.ko	1	0,2566	0,4036	1	1	1	1	1
ms1.ko_ms2.t	0,9494	0,0032	0,0057	1	1	1	1	0,0870
ms1.ko_ms2.f	1	0,0053	0,0095	1	1	1	1	0,1338
ms1.s_ms2.ko	1	1	1	1	1	1	1	1
ms1.s_ms2.t	1	1	1	1	1	1	1	1
ms1.s_ms2.f	1	1	1	1	1	1	1	1

ms1.c_ms2.f	ms1.d_ms2.ko	ms1.d_ms2.t	ms1.d_ms2.f	ms1.ko_ms2.ko	ms1.ko_ms2.t	ms1.ko_ms2.f	ms1.s_ms2.ko	ms1.s_ms2.t	ms1.s_ms2.f
1	1	1	1	1	0,9494	1	1	1	1
1	1	1	1	0,2566	0,0032	0,0053	1	1	1
1	1	1	1	0,4036	0,0057	0,0095	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	0,0870	0,1338	1	1	1
1	1	1	1	1	0,0523	0,0817	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1
0,0523	1	1	1	1	1	1	1	0,3412	0,5917
0,0817	1	1	1	1	1	1	1	0,5030	0,8563
1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	0,3412	0,5030	1	1	1
1	1	1	1	1	0,5917	0,8563	1	1	1

Table S3 Adjusted p-values for pairwise comparisons of growth (area under the curve, AUC) among S. suis RM system variants in CDM-galactose

Strain	ms1.a_ms2.ko	ms1.a_ms2.t	ms1.a_ms2.f	ms1.b_ms2.ko	ms1.b_ms2.t	ms1.b_ms2.f	ms1.c_ms2.ko	ms1.c_ms2.t
ms1.a_ms2.ko	1,0000	0,0705	0,5063	0,5077	0,3506	0,8219	1,0000	0,1051
ms1.a_ms2.t	0,0705	1,0000	0,9999	0,0000	0,0000	0,0001	0,0124	1,0000
ms1.a_ms2.f	0,5063	0,9999	1,0000	0,0006	0,0003	0,0029	0,1659	1,0000
ms1.b_ms2.ko	0,5077	0,0000	0,0006	1,0000	1,0000	1,0000	0,8889	0,0000
ms1.b_ms2.t	0,3506	0,0000	0,0003	1,0000	1,0000	1,0000	0,7656	0,0000
ms1.b_ms2.f	0,8219	0,0001	0,0029	1,0000	1,0000	1,0000	0,9910	0,0002
ms1.c_ms2.ko	1,0000	0,0124	0,1659	0,8889	0,7656	0,9910	1,0000	0,0200
ms1.c_ms2.t	0,1051	1,0000	1,0000	0,0000	0,0000	0,0002	0,0200	1,0000
ms1.c_ms2.f	0,8979	0,9718	1,0000	0,0049	0,0023	0,0212	0,5233	0,9897
ms1.d_ms2.ko	0,0822	0,0000	0,0000	1,0000	1,0000	0,9936	0,3150	0,0000
ms1.d_ms2.t	0,8893	0,0002	0,0046	1,0000	1,0000	1,0000	0,9971	0,0003
ms1.d_ms2.f	0,9732	0,0005	0,0119	1,0000	0,9992	1,0000	0,9999	0,0008
ms1.ko_ms2.ko	0,0046	0,0000	0,0000	0,8883	0,9610	0,6060	0,0296	0,0000
ms1.ko_ms2.t	0,0000	0,0000	0,0000	0,0008	0,0017	0,0001	0,0000	0,0000
ms1.ko_ms2.f	0,0004	0,0000	0,0000	0,4153	0,5801	0,1684	0,0030	0,0000
ms1.s_ms2.ko	1,0000	0,0565	0,4478	0,5677	0,4039	0,8646	1,0000	0,0853
ms1.s_ms2.t	0,1605	1,0000	1,0000	0,0001	0,0000	0,0003	0,0338	1,0000
ms1.s_ms2.f	0,0141	1,0000	0,9808	0,0000	0,0000	0,0000	0,0020	1,0000

ms1.c_ms2.f	ms1.d_ms2.ko	ms1.d_ms2.t	ms1.d_ms2.f	ms1.ko_ms2.ko	ms1.ko_ms2.t	ms1.ko_ms2.f	ms1.s_ms2.ko	ms1.s_ms2.t	ms1.s_ms2.f
0,8979	0,0822	0,8893	0,9732	0,0046	0,0000	0,0004	1,0000	0,1605	0,0141
0,9718	0,0000	0,0002	0,0005	0,0000	0,0000	0,0000	0,0565	1,0000	1,0000
1,0000	0,0000	0,0046	0,0119	0,0000	0,0000	0,0000	0,4478	1,0000	0,9808
0,0049	1,0000	1,0000	1,0000	0,8883	0,0008	0,4153	0,5677	0,0001	0,0000
0,0023	1,0000	1,0000	0,9992	0,9610	0,0017	0,5801	0,4039	0,0000	0,0000
0,0212	0,9936	1,0000	1,0000	0,6060	0,0001	0,1684	0,8646	0,0003	0,0000
0,5233	0,3150	0,9971	0,9999	0,0296	0,0000	0,0030	1,0000	0,0338	0,0020
0,9897	0,0000	0,0003	0,0008	0,0000	0,0000	0,0000	0,0853	1,0000	1,0000
1,0000	0,0002	0,0317	0,0721	0,0000	0,0000	0,0000	0,8612	0,9976	0,7411
0,0002	1,0000	0,9827	0,9168	0,9999	0,0147	0,9439	0,1015	0,0000	0,0000
0,0317	0,9827	1,0000	1,0000	0,5060	0,0001	0,1220	0,9209	0,0006	0,0000
0,0721	0,9168	1,0000	1,0000	0,3103	0,0000	0,0569	0,9840	0,0016	0,0001
0,0000	0,9999	0,5060	0,3103	1,0000	0,1970	1,0000	0,0060	0,0000	0,0000
0,0000	0,0147	0,0001	0,0000	0,1970	1,0000	0,6571	0,0000	0,0000	0,0000
0,0000	0,9439	0,1220	0,0569	1,0000	0,6571	1,0000	0,0005	0,0000	0,0000
0,8612	0,1015	0,9209	0,9840	0,0060	0,0000	0,0005	1,0000	0,1323	0,0109
0,9976	0,0000	0,0006	0,0016	0,0000	0,0000	0,0000	0,1323	1,0000	1,0000
0,7411	0,0000	0,0000	0,0001	0,0000	0,0000	0,0000	0,0109	1,0000	1,0000

Chapter 5

Cost-Effective Nanopore Workflow for 16S rRNA Microbiome Profiling Validated in a Zebrafish Gut Model

Maria Juanpere-Borràs[1*], Jos Boekhorst[1], Peter van Baarlen[1], Sylvia Brugman[1], Adrià López Nadal[1*]

[1]Host-Microbe Interactomics Group, Wageningen University and Research, Wageningen, Netherlands.

Abstract

Advancements in long-read sequencing technologies, such as those developed by Oxford Nanopore Technologies (ONT), have opened new possibilities for 16S rRNA gene profiling of microbial communities by enabling full-length sequencing. In this study, we optimized and validated a cost-effective, in-house ONT workflow for zebrafish gut microbiome analysis, and compared its performance to the gold-standard Illumina NovaSeq platform. Using a PCR-based barcoding strategy, we successfully sequenced both the V3–V4 and V1–V9 regions of the 16S rRNA gene, achieving high read accuracy and balanced demultiplexing across samples. Our results demonstrate that ONT short-read (V3–V4) sequencing using the latest V14 chemistry and Dorado basecaller achieves resolution comparable to that of Illumina, supporting amplicon sequence variant (ASV) classification. Rarefaction analysis confirmed that the sequencing depth (~240,000 reads/sample) was sufficient to capture microbial diversity in the zebrafish gut, a complex and taxonomically rich environment. Furthermore, taxonomic classifications and diversity metrics were consistent between platforms, supporting the reliability of ONT for microbiota profiling. While full-length (V1–V9) ONT sequencing yielded higher taxonomic richness through Operational Taxonomic Units (OTUs), it did not significantly improve species-level resolution, likely due to current database and bioinformatics limitations. Our custom library preparation protocol simplified multiplexing, reduced costs, and improved scalability without compromising accuracy. These findings support ONT as a viable, scalable, and flexible alternative to Illumina for in-house microbiome research.

Introduction

Technological advancements in multi-omics approaches have significantly expanded our understanding of the microbiome, particularly in identifying its composition and elucidating its role in disease onset and development [216,217]. The bacterial composition of the microbiome is commonly determined through partial sequencing of the 16S ribosomal RNA gene, known as 16S profiling. The 16S rRNA gene is present in all bacteria and along its length (i.e., approximately 1500 bp) contains 9 variable regions (V1-V9) (Figure 1), the sequence of which inform taxonomic origin [218–220]. Traditionally, the platform chosen for 16S rRNA sequencing is the Illumina sequencing platform, which provides the basecalling accuracy needed for correct taxonomic profiling of the microbiome [220,221]. However, one major limitation of Illumina sequencing is its inability to produce reads longer than 500 base pairs, potentially limiting taxonomic resolution since these regions might not capture the necessary genetic variation to delineate between all bacterial species [219,220,222].

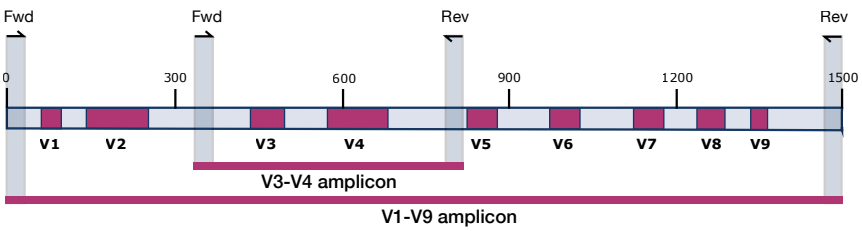


Figure 1. Schematic representation of the full-length 16S rRNA gene, highlighting conserved (grey) and variable (pink) regions (V1–V9). This visualization illustrates the structure of the gene and the relative coverage of the two amplicon strategies used in this study. The approximate locations of primer pairs targeting the V3–V4 and V1–V9 regions are indicated. Numbers along the gene correspond to nucleotide positions (in base pairs). Positions of each variable region were abstracted from study [248].

The recent development of next-generation long-read sequencing technologies, such as PacBio and nanopore sequencing as developed by ONT, offers the opportunity to sequence full-length 16S rRNA genes, potentially improving taxonomic classification [219,221,222]. ONT technologies, in particular, are amenable to in-house sequencing and offer convenience through a wide range of customizable library preparation kits and sequencing protocols. Additionally, the small and portable nature of ONT sequencing devices allows sequencing to be conducted virtually everywhere, from laboratory to directly in the field. This provides greater flexibility and

a reduction in both time and cost of sequencing, compared to outsourcing sequencing to an external company, although sequencing in field conditions carries a greater contamination risk [100,102,103]. The biggest drawback of ONT has traditionally been its lower basecalling accuracy, which is crucial for taxonomic classification. This limitation has been commonly addressed by combining short-read, high-accuracy technologies like Illumina with long-read, lower-accuracy technologies like ONT [97–99]. However, the latest advances in ONT chemistry and the Dorado basecaller model have raised sequencing accuracy to over 99% [100,101]. However, even a 1% error rate may introduce substantial sequencing noise for taxonomic classifications based on the full-length 16S rRNA gene, where fine-scale resolution is critical.

Microbiota profiling requires high basecalling accuracy, because even small sequence errors can result in misclassifications. Due to this sensitivity, we consider 16S rRNA gene profiling an ideal test case for evaluating the robustness of an in-house ONT sequencing platform. Numerous studies have already compared ONT and Illumina platforms for 16S rRNA microbiome profiling, highlighting their respective strengths and limitations. However, some studies have utilized earlier versions of ONT chemistry (e.g., V12), which have since undergone significant advancements, yet, comparative studies using the latest V14 chemistry remain limited [223–225]. Previous studies have not taken advantage of advancements in basecalling software, such as the transition from Guppy to the more accurate and powerful Dorado [226,227]. Additionally, while published comparisons focus on ONT's full-length 16S rRNA versus Illumina's short-read sequencing, a critical limitation often arises: full-length 16S rRNA reads can accumulate too many sequencing errors constraining analyses to OTUs rather than Amplicon Sequence Variants (ASVs) [228,229].

In this paper, we describe an ONT sequencing workflow to reliably and reproducibly assess changes in microbiota composition in zebrafish a widely used experimental animal model. We used a real biological case study to assess whether both sequencing technologies could detect the same changes in microbiota composition. In previous studies, Brugman and co-workers observed statistically significant and biologically relevant shifts in microbiota composition following the deletion of adaptive immune cells in Rag1-deficient zebrafish [230]. As T cells contribute to immune responses, including the expression of *cxcl8a*, the role of *cxcl8a* in shaping the gut microbiome was investigated [231]. To this end, we conducted 16S rRNA profiling of the zebrafish gut microbiota in both wild-type and a novel CRISPR-cas9-generated *cxcl8a* deletion mutant .

The results of our study demonstrates that ONT using the latest chemistry and basecalling tools, achieves sufficient accuracy for ASVs classification of the V3-V4 amplicon reads. The results obtained with the latest Nanopore chemistry (V14), combined with the Dorado basecalling model, matched the same taxonomic classification obtained using Illumina NovaSeq. Additionally, we optimized a customized library preparation protocol for ONT, significantly reducing the steps and time required for sample preparation. We also designed a bioinformatics pipeline for downstream analysis of sequencing reads for 16S rRNA profiling. Furthermore, we used both short (V3-V4) and long (V1-V9) reads to determine whether sequencing the full 16S rRNA gene provides new insights or deeper taxonomic classification compared to traditional short-read sequencing. To validate our new protocols, we used experimental data to assess the impact of Cxcl8a on shaping the microbiota composition of the zebrafish gut.

Materials and Methods

Ethics Statement

The present study was approved by the Dutch Committee on Animal Welfare (2019.W-0045.001) and the Animal Welfare Body (IvD) of the Wageningen University (Netherlands). Furthermore, we followed the standard biosecurity and institutional safety procedures at Wageningen University and Research (The Netherlands).

Animals

To generate *cxcl8a* deletion mutant in zebrafish, CRISPR-cas9 mediated gene-editing was performed using a guide RNA targeting the first exon of zebrafish *cxcl8a* (ENSDARG00000104795; GRCz11) sgRNA: The single guide used was AGGAATGAGCTTGAGAGGTC. The full sequence of the oligo [ccgctagc-taatacgactcactata-AGGAATGAGCTTGAGAGGTC–GTTTTAGAGCTAGAAATAGCAAG] and common reverse primer [AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTA GCCTTATTTTAA CTTGCTATTTCTAGCTCTAAAAC]. One cell stage embryos were injected with 150 ng/ul sgRNA, 2 ng/ul cas9 enzyme (IDT AltR s.p. Cas9 Nuclease V3 #1081058), and 0.3 µL phenol red. Analysis of the offspring of these F0 yielded heterozygous fish with different mutations in the germline. *cxcl8a* +/- F1 offspring (n=5 with different mutations) were selected and outcrossed to AB wildtype fish to yield F2 *cxcl8a* +/- zebrafish. The founder and offspring with a germline transmission of a 141 bp mutation in *cxcl8a* leading to a premature stop was selected and outcrossed to the double immune cell reporter transgenic Tg (mpeg1:mCherry/mpx:eGFPi^{114}) to obtain cxcl8a +/+ Tg (mpeg1:mCherry/mpx:eGFPi^{114}) and *cxcl8a* -/- (mpeg1:mCherry/mpx:eGFPi^{114}) zebrafish harbouring GFP+ neutrophils and mCherry+ macrophages.

Adult *cxcl8a* +/+ Tg (mpeg1:mCherry/mpx:eGFPi[114]) and *cxcl8a* -/- (mpeg1:mCherry/mpx:eGFPi[114]) zebrafish were housed and fed as previously described [232]. Embryos were obtained by natural spawning and raised with E2 media (0.10 mM NaCl in demineralized water, pH 7.3) in petri dishes at 28°C (12/12-hour light/dark cycle) [233]. Embryos were visually inspected daily and visibly infected embryos were removed using a MS-222 (tricaine mesylate: tricaine/E3 (8.4% v/v 24 mM tricaine) (Sigma-Aldrich, DL, United States of America) solution.

Experimental design

Fish were raised as previously described in López Nadal et al.[234] until 21 days-post fertilization (dpf). WT *cxcl8a* +/+ and mutant *cxcl8a* -/- fish were raised in separate tanks in triplicate. Fish were randomly sampled from each of three tanks.

DNA extraction

Guts were extracted from juvenile zebrafish, rinsed in sterile PBS, snap frozen in liquid nitrogen and preserved at -80°C for total gut DNA extraction. DNA extractions were performed as previously described[234,235] for each single gut. In brief, 100 µL lysis buffer (100mM NaCl, 10mM Tris pH 8, 15 mM EDTA 0.5% SDS) and 7µL proteinase K (QIAGEN, 19131) were added to the samples. Samples were incubated at 56°C overnight until fully dissolved. Then, 35 µL of 6M NaCl were added and samples were incubated under continuous shacking. After centrifugation, supernatant was transferred to a sterile 1.5 mLEppendorf tube. Next, 270 µL 100% (cold) ethanol was added and samples were kept in –80 C for one hour. Samples were centrifuged for 5 minutes at full speed, and the pellet washed with 70% ethanol. After centrifugation, the supernatant was removed, and pellets were air dried before adding 20 µL RNase/DNase free water. DNA concentrations were measured by Nanodrop and Qubit 2.0 fluorometer using Qubit™ dsDNA Quantification Assay Kits (ThermoFisher Scientific, Q32853). The same samples were used for illumina and Nanopore sequencing.

Illumina Sequencing

Next Generation Sequencing (NGS) Illumina NovaSeq 6000 technology with the following forward (FW) and reversed (RV) primers yielded 250 bp paired-end reads of the V3-V4 hypervariable region of the 16S rRNA gene: FW (CCTACGGGNGGCWG-CAG) and RV (GACTACHVGGGTATCTAATCC).

16S rRNA amplification and barcoding

PCR amplification reactions were performed using LongAmpl polymerase (NEB, M0323S) according to the manufacturer's instructions, with 100ng of template DNA and a final reaction volume of 50 µL. Primers targeting the V3–V4 region of the 16S rRNA gene were provided by Baseclear B.V., allowing us to perform PCR amplification

in our own lab using the same primers as those used by the company (Figure 1). For V1-V9 targeting region we used previously published primers from Y. Matsuo et, al.[236] (Figure 1, Table 1). Unique barcode sequences were added to the 5' end of the forward primers to ensure accurate sample identification during downstream data processing. Primer sequences can be found in Table 1. The barcode sequences were extracted from the protocol of the Native Barcoding Kit 24 (ONT, SQK-NBD112.24) from Nanopore technologies. To obtain V3-V4 amplicons, thermal cycling conditions were as follows; 30 cycles, annealing temperature of 54°C and extension time of 1 min. To obtain V1-V9 amplicons, the protocol was adjusted to an annealing temperature of 53°C and an extension time of 2 min. In both protocols, 5 µL of PCR product were loaded onto a 0.8% agarose gel to confirm correct fragment amplification (data not shown). The remaining 45 µL of PCR product were purified using the Invisorb Fragment CleanUp kit (Invitek molecular, 1020300300) following the manufacturer's instructions. Final DNA concentrations were measured using Qubit 2.0 fluorometer, and Qubit™ dsDNA Quantification Assay Kits (ThermoFisher Scientific, Q32853).

Library preparation and sequencing

Sequencing was organized into batches of around 15 samples, each identified by barcodes ranging from 1 to 15 (see Table 1). Each batch of samples were pooled together in equimolar concentrations to a final concentration of 200 fmol. The library preparation was conducted using the Ligation Sequencing Kit V14 (ONT, SQK-LSK114) following the manufacturer's instructions. Subsequently, the prepared library was loaded onto R10.4.1 flow cells (ONT, FLO-MIN114) using either the MinION Mk1c and Mk1b (ONT, Oxford, United Kingdom) sequencing devices, until most of the reads were sequenced. The total number of sequenced reads typically ranges from 14 to 19 million per flow cell, with the sequencing process usually taking between 24 and 48 hours to complete, depending on the amplicon length.

Sequencing data processing and downstream analysis

Raw Nanopore sequencing output was collected in .pod5 files and subsequently subjected to basecalling; results were stored in .fastq files using Dorado 0.5.3. Reads were demultiplexed using custom script. The sequencing reads were initially filtered by length, retaining those with lengths ranging from 300 to 700 base pairs. Subsequently, reads passing the filtering step were processed to identify the presence of barcode sequences and forward primer sequence within the first 100 base pairs of the reads or, alternatively, within the last 100 base pairs of the reads, considering reverse complement sequences of forward primer plus barcode. Our analysis pipeline transformed reverse complement sequences into corresponding 5'->3' orientations before classification to ensure uniform orientation of all reads

Table 1. Primer sequences used in this study, with barcode sequences in italics.

Name	Sequence
V3-V4 Bc1 FWD	*CACAAAGACACCGACAACTTTCTT*CCTACGGGNGGCWGCAG
V3-V4 Bc2 FWD	*ACAGACGACTACAAACGGAATCGA*CCTACGGGNGGCWGCAG
V3-V4 Bc3 FWD	*CCTGGTAACTGGGACACAAGACTC*CCTACGGGNGGCWGCAG
V3-V4 Bc4 FWD	*TAGGGAAACACGATAGAATCCGAA*CCTACGGGNGGCWGCAG
V3-V4 Bc5 FWD	*AAGGTTACACAAACCCTGGACAAG*CCTACGGGNGGCWGCAG
V3-V4 Bc6 FWD	*GACTACTTTCTGCCTTTGCGAGAA*CCTACGGGNGGCWGCAG
V3-V4 Bc7 FWD	*AAGGATTCATTCCCACGGTAACAC*CCTACGGGNGGCWGCAG
V3-V4 Bc8 FWD	*ACGTAACTTGGTTTGTTCCCTGAA*CCTACGGGNGGCWGCAG
V3-V4 Bc9 FWD	*AACCAAGACTCGCTGTGCCTAGTT*CCTACGGGNGGCWGCAG
V3-V4 Bc10 FWD	*GAGAGGACAAAGGTTTCAACGCTT*CCTACGGGNGGCWGCAG
V3-V4 Bc11 FWD	*TCCATTCCCTCCGATAGATGAAAC*CCTACGGGNGGCWGCAG
V3-V4 Bc12 FWD	*TCCGATTCTGCTTCTTTCTACCTGC*CTACGGGNGGCWGCAG
V3-V4 Bc13 FWD	*AGAACGACTTCCATACTCGTGTGA*CCTACGGGNGGCWGCAG
V3-V4 Bc14 FWD	*AACGAGTCTCTTGGGACCCATAGA*CCTACGGGNGGCWGCAG
V3-V4 Bc15 FWD	*AGGTCTACCTCGCTAACACCACTG*CCTACGGGNGGCWGCAG
V3-V4 REV	GACTACHVGGGTATCTAATCC
V1-V9 Bc1 FWD	*CACAAAGACACCGACAACTTTCTT*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc2 FWD	*ACAGACGACTACAAACGGAATCGA*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc3 FWD	*CCTGGTAACTGGGACACAAGACTC*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc4 FWD	*TAGGGAAACACGATAGAATCCGAA*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc5 FWD	*AAGGTTACACAAACCCTGGACAAG*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc6 FWD	*GACTACTTTCTGCCTTTGCGAGAA*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc7 FWD	*AAGGATTCATTCCCACGGTAACAC*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc8 FWD	*ACGTAACTTGGTTTGTTCCCTGAA*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc9 FWD	*AACCAAGACTCGCTGTGCCTAGTT*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc10 FWD	*GAGAGGACAAAGGTTTCAACGCTT*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc11 FWD	*TCCATTCCCTCCGATAGATGAAAC*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc12 FWD	*TCCGATTCTGCTTCTTTCTACCTG*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc13 FWD	*AGAACGACTTCCATACTCGTGTGA*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc14 FWD	*AACGAGTCTCTTGGGACCCATAGA*AGRGTTYGATYMTGGCTCAG
V1-V9 Bc15 FWD	*AGGTCTACCTCGCTAACACCACTG*AGRGTTYGATYMTGGCTCAG
V1-V9 REV	CGGYTACCTTGTTACGACTT

passing filtering. Reads lacking the specified sequence (barcode plus forward primer) were classified as unmatched. The same process was applied to the reverse primer in the last 100 bp of reads that remained after forward primer matching. The demultiplexed reads were then processed through the QIIME pipeline using default parameter settings for ASV calling with DADA2. For the V1–V9 region, the ONT-AmpSeq workflow was used with the default parameters as recommended by the authors [237], followed by taxonomic classification.

Statistical analyses

Rarefaction curves for all samples reached a plateau, indicating that sequencing depth was sufficient to capture most of the microbial diversity (data not shown). To ensure fair comparison across samples, we rarefied the data to 50,000 reads per sample, the lowest sequencing depth that still retained a comprehensive representation of taxonomic diversity across all samples. Redundancy analysis (RDA) and principal component analysis (PCA) were performed in Canoco v5.15 [238] using analysis type “constrained” or “unconstrained”, respectively. Response variables were log-transformed with the formula log(10,000*relative_ abundance + 1). RDA p-values were determined through permutation testing (n = 500 permutations). Boxplots were generated using Prism v.9.0.0 (GraphPad Software, San Diego, California USA). Cytoscape v3.9.1 [239] as used to visualize the diet-specific co-occurrence of ASVs based on their relative abundances. Additional data handling and format conversions were done in Python 3.8.8 (https://www.python.org/).

Results

16S rRNA Sequencing Workflow Overview

We set out to design and validate a workflow to efficiently amplify and sequence specific region of the 16S rRNA gene using ONT (see Methods). This workflow (Figure 2) allows users to customize several parameters, including primer selection, amplicon size, barcode sequences, and the number of samples to pool together. Additionally, PCR conditions, such as annealing temperature and extension times, can be adjusted based on users-defined criteria.

The workflow is divided into three main sections (Figure 2). The first section involves a customized library preparation protocol adapted to the experimental design. This preparatory step was performed prior to initiating the ONT library preparation protocol. The second section includes steps using the Nanopore sequencing kit in accordance with the manufacturer’s instructions. The third and final section involves the bioinformatic analysis of the raw data, processed to facilitate ASVs calling and taxonomic classification.

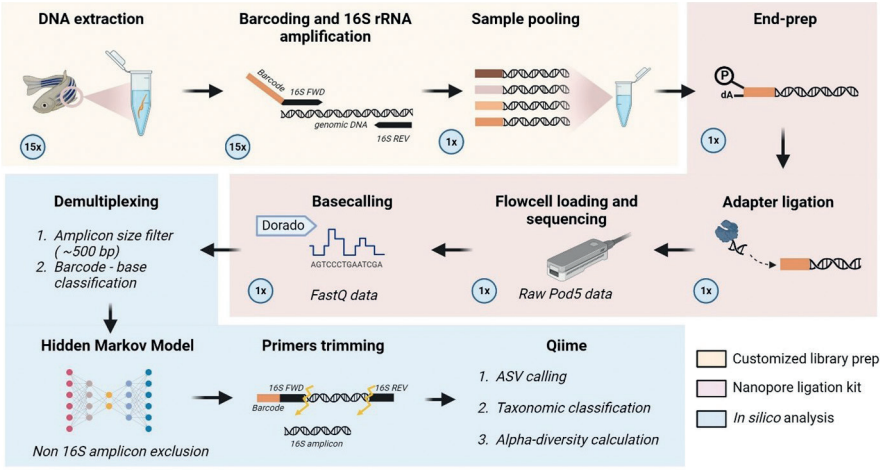


Figure 2. ONT 16S rRNA sequencing workflow.

The customized library preparation (Figure 2) begins with the DNA extraction from the biological samples. The choice of DNA extraction kit is independent of the subsequent workflow steps. For PCR amplification, users can select forward (FWD) and reverse (REV) primers targeting the variable regions for the 16s rRNA gene. For this study, we performed PCR amplification targeting the V3-V4 and V1-V9 regions (Figure 1). Each forward primer included a barcode sequence as a single DNA oligo overhang at the 5' end. Consequently, a unique DNA sequence including a standard forward primer, but a unique barcode was used for each sample. Up to this point, each sample had been processed independently. Once samples had been tagged with a unique barcodes, equimolar concentrations of each sample PCR amplicon were pooled together into a single tube to reach up to the final concentration specified in the Nanopore ligation kit protocol. For our study, after a small trial to ensure sufficient read coverage and given the high microbiome diversity of the zebrafish gut, we decided to pool 15 samples together.

From this step onward, the protocol provided with the Nanopore ligation kit was implemented (Figure 2), which consists of two main steps: (i) amplicon end-preparation and (ii) adapter ligation. After the adapter ligation step, the resulting DNA fragments were loaded onto the flow cell and sequenced until all PCR amplicons had been processed, and no further reads were generated. Raw sequencing data were converted to .fastq files for downstream analysis.

For bioinformatic downstream analysis (Figure 2), we first filtered all generated reads by amplicon size: to obtain reads with sizes corresponding to the rRNA gene region, of interest reads were filtered to 350-550 bp, while to obtain reads with sizes corresponding to the V1-V9 region, reads were filtered to 1300-1600 bp. Next, we classified reads based on their barcode sequences, retaining only those reads containing DNA sequences that matched our designed barcodes at 100% identity. This process resulted in 15 separate .fastq files, each .fastq file containing reads with a unique barcode, and an additional file for reads with no barcode or reads with less than 100% similarity (mismatches) to our barcodes. All reads passing these filtering steps were then processed using Hidden Markov Models to discard reads that did not resemble 16S rRNA. Finally, reads were trimmed to remove any sequences outside the primer regions, including the primer sequences themselves, retaining only the 16S rRNA fragments bordered by the respective primers. After this step, filtered and trimmed reads were parsed to the Qiime pipeline for ASVs or ONT-Ampseq for OTU calling, taxonomic classification, and alpha- and beta-diversity calculation for the samples.

Data Output Comparison: Nanopore vs. Illumina

Due to the limitation of 15 barcodes and a total of 38 samples, sequencing was carried out in 3 rounds to accommodate all samples. Each round was performed separately for both the V3–V4 and V1–V9 amplicon regions, resulting in a total of 6 sequencing runs. This yielded approximately 35 million reads for V3-V4 region and 41 million reads for V1-V9 region.

To evaluate whether an excessive number of reads were discarded during any of the bioinformatic filtering steps, and to assess the specificity of PCR as well as the efficiency of the barcoding system (i.e., ensuring that no barcode interfered with 16S rRNA amplification due to non-specific binding elsewhere in the genome), we obtained detailed reports of reads per sample and total reads throughout the processing pipeline (Figure 3, Table S1).

From these reports, we found that the read count distribution was more consistent in the V3-V4 samples, whereas the V1-V9 dataset exhibited greater variability (Figure 3), suggesting a less uniform equimolar input of the barcoded samples in V1-V9 compared to V3-V4. The removal of non-16S rRNA amplicons (i.e., amplicons including zebrafish DNA, and DNA from other microorganisms) had minimal impact on the total read count in both datasets, with 87.68% and 96.82% of average reads per sample retained for the V3-V4 and V1-V9 primer sets, respectively. Subsequent barcode and primer trimming steps further refined the datasets, retaining an average of 72.79% of reads per sample in the V3-V4 dataset and 76.88% in the V1-V9 dataset.

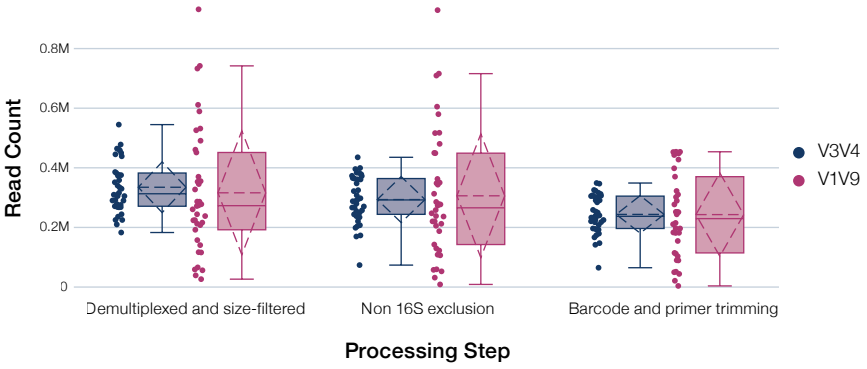


Figure 3. Distribution of read counts per sample across key bioinformatic processing steps for V3–V4 and V1–V9 16S rRNA gene amplicon datasets. Box plots depict the number of reads retained at each of three sequential steps. For each step, the blue and red boxes represent the V3–V4 and V1–V9 datasets, respectively. Boxes display the interquartile range (IQR), with the median indicated by a horizontal line within the box. Individual sample values are plotted as points to illustrate variability.

A total average of 243,716 reads per sample in V3-V4 and 243,047 reads per sample in V1-V9, were retained for ASV or OTU assignment and taxonomic classification (Figure 3). Illumina sequencing yielded an average of 6.2 million reads per sample, provided directly by the sequencing company.

Alpha Diversity Analysis of Gut Microbiome in Mutant and Wild-Type Zebrafish Using V3-V4 Amplicon Sequencing by Illumina and Nanopore sequencing platforms

Gut samples (n=19/group) from *cxcl8a* +/+ and *cxcl8a* -/- fish (21dpf) were used to determine prokaryotic community composition. All samples were sequenced using both amplicon sequencing of V3-V4 16S rRNA gene region using Illumina (white background) and ONT (grey background) sequencing platforms (Figure 4). The number of observed ASVs indexes (richness indicator) was comparable between *cxcl8a* +/+ and *cxcl8a* -/- zebrafish; Illumina sequencing identified a slight but significantly higher number of ASVs than ONT (Figure 4A). Nonetheless, the increased number of ASVs did not reflect an increase in phylogenetic diversity (Figure 4B). Other alpha-diversity metrics such as Evenness or Shannon (diversity indicators) were comparable between genotypes and sequencing platforms (Figure 4C, 4D).

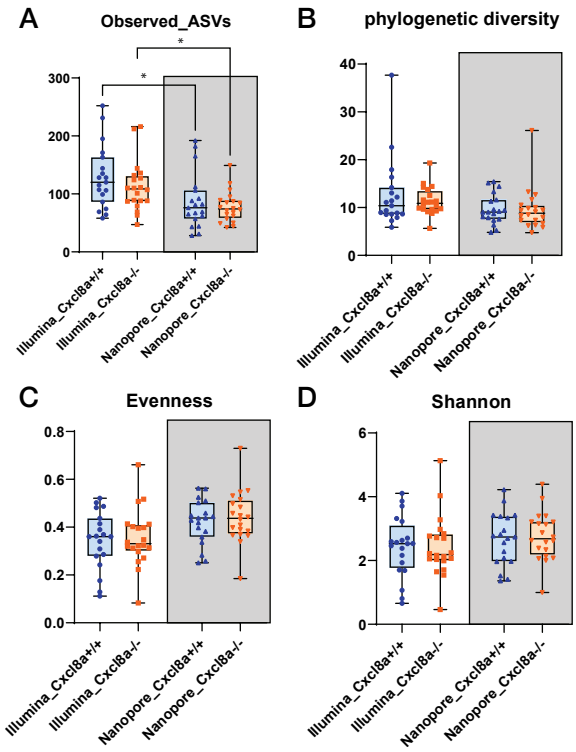


Figure 4. Alpha-diversity metrics of *cxcl8a*+/+ and *cxcl8a*-/- zebrafish gut samples sequenced using Illumina (white background) and Nanopore (grey background) sequencing platforms.

Comparable Beta Diversity in Zebrafish Gut Microbiomes Across Sequencing Platforms

To visualise the differences in the taxonomy and diversity of microbiota composition of the zebrafish gut resulting from short-read Illumina or long-read sequencing ONT platforms, we compared beta-diversity by PCA and RDA. Illumina NovaSeq 6000 and ONT sequencing explained roughly the same amount of variation in microbiota composition on the first two PCA axes, with PC1 explaining 27.88% and 25.57% of microbiota composition variation respectively, and PC2 explaining 16.8% vs 17.62% of variation in microbiota composition, respectively (Figure 5A). In the results of both sequencing platforms, samples similarly clustered in two separate groups depending on their genotype (*cxcl8a* +/+ and *cxcl8a* -/-). When comparing the variation in taxonomic composition of microbiota measured by Illumina and ONT sequencing platforms by RDA (p=0.002), the same trends were found as uncovered by PCA.

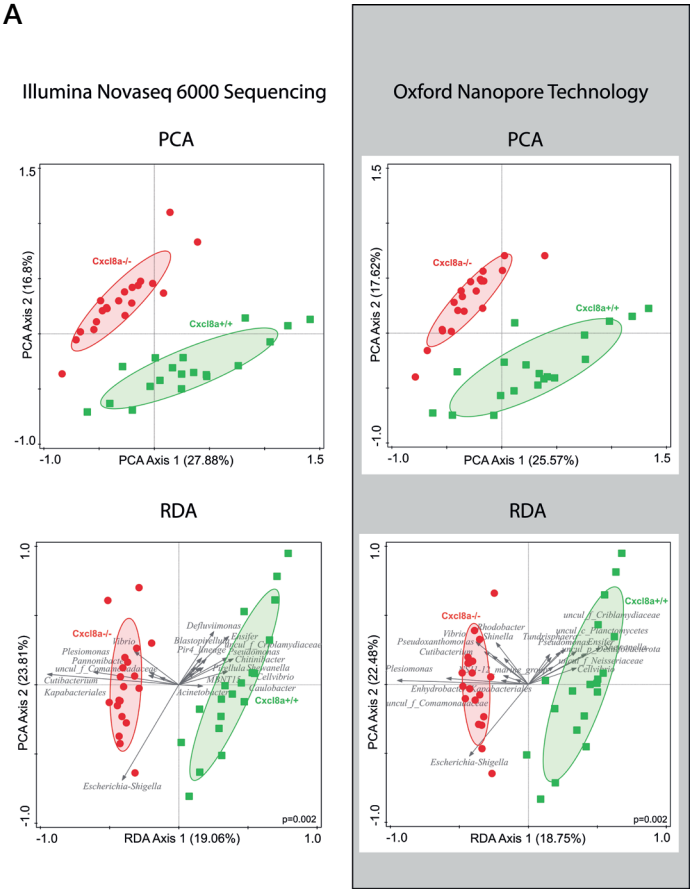


Figure 5. (A) Beta diversity PCA (Principal Component Analysis) and RDA (Redundancy Analysis) of *cxcl8a*+/+ and *cxcl8a*-/- zebrafish gut samples. (B) Beta diversity PCA (Principal Component Analysis) and RDA (Redundancy Analysis) of the two different sequencing technologies used in our study (Illumina and Nanopore).

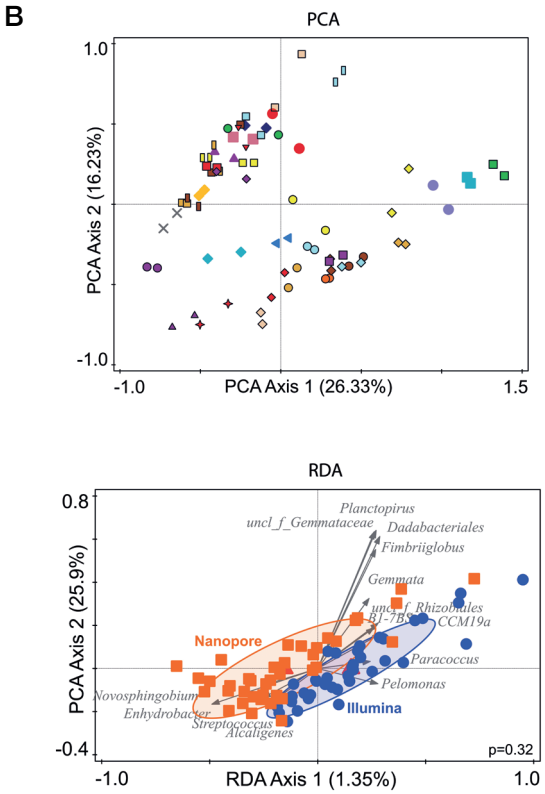


Figure 5. Continued.

PCA and RDA analyses identified the same genera as the most substantially different in relative abundance between WT and *cxcl8a* deletion mutant fish, as measured by both Illumina and ONT sequencing platforms.

To further explore whether the Illumina and ONT sequencing platforms capture different patterns in microbiota composition, we performed PCA and RDA analyses across all wild-type and *cxcl8a* mutant samples (Figure 5B). Unlike Figure 5A, which presents RDA results separately for each sequencing platform, Figure 5B integrates both Illumina and ONT data into a single PCA and RDA analysis. Notably, the total variation explained by PC1 and PC2 was comparable to the platform-specific analyses, suggesting overall consistency between sequencing methods. To visualize overlap between the two platforms at the sample level, we marked each sample with the same symbol across both platforms. Samples sequenced by both Illumina and

ONT consistently clustered closely together, indicating that the microbiota profiles generated by each platform were highly similar. Additionally, when all samples were included in a combined RDA, sequencing platform did not significantly explain the variation in microbial composition (p = 0.32), further supporting the agreement between platforms across genotypes.

Finally, we wanted to assess whether the relative abundances of the six most informative species were found in a similar manner in both sequencing techniques. We selected 6 interesting ASVs and compared their abundances in the two different sequence analysis datasets (Illumina white background) and Nanopore (grey background) (Figure 6). For the selected six species, we found minimal variations in the respective relative abundances from samples obtained from either WT or *cxcl8a* deletion mutant fish, measured either by Illumina or ONT sequencing platform. Therefore, we conclude that both sequencing platforms provide comparable results at the level of 16S rRNA gene diversity.

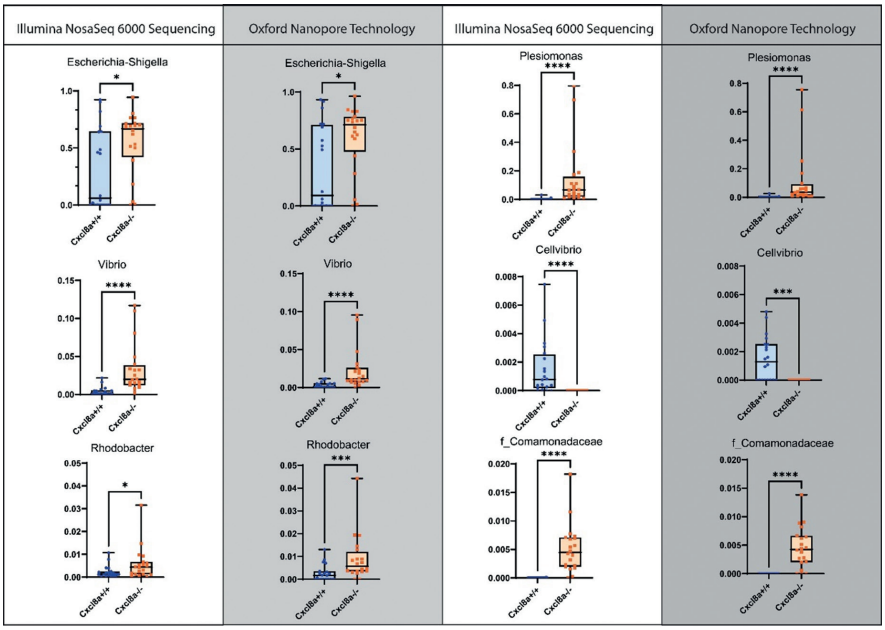


Figure 6. Boxplots of relative abundance of 6 interesting species in our *cxcl8a*+/+ and *cxcl8a*-/- genotypes sequenced using Illumina (white background) and Nanopore (Grey Background).

Evaluation of Taxonomic Diversity and Resolution by 16S rRNA Region and Sequencing Platform

Following taxonomic classification using OTUs for V1-V9 and ASVs for V3-V4, we observed that the number of ASVs detected at each taxonomic level, from phylum to species, was consistent across Illumina and Nanopore sequencing platforms, suggesting that both platforms identify largely overlapping ASVs within the V3-V4 region (Figure 7).

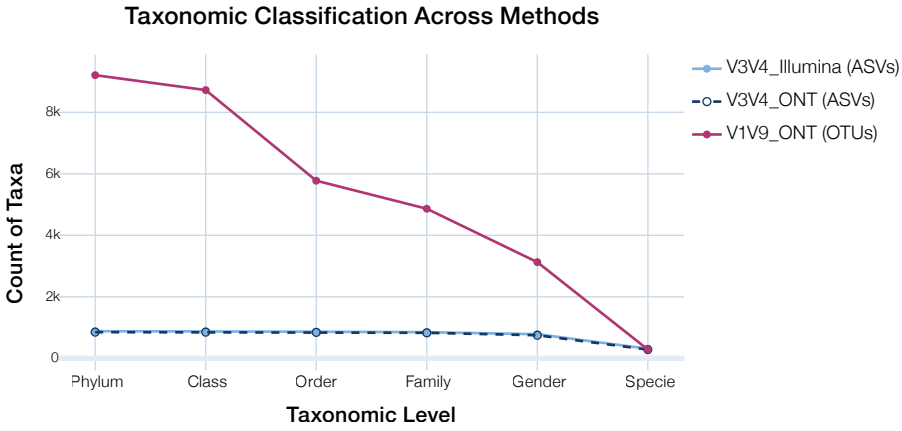


Figure 7. Taxonomic diversity of microbiota composition of zebrafish gut samples as measured by Illumina and ONT sequencing platforms. The number of identified taxa is shown at different taxonomic levels for V3-V4 (ASVs) using Illumina and ONT, and V1-V9 (OTUs) using ONT

In contrast, using 16S rRNA V1-V9 region for taxonomic classification resulted in a substantial increase in the number of OTUs detected at each taxonomic level compared to the number of ASVs identified in V3-V4 region, significantly improving taxonomic resolution (Figure 7). Despite the higher taxonomic resolution observed with the full-length V1–V9 region, from phylum to genus, the number of taxa successfully assigned decreased with deeper classification levels. By the species level, the number of taxa assigned using V1–V9 was comparable to that obtained with the shorter V3–V4 region, although notable differences in specific taxa remained (Figure 7, 8). Specifically, at the phylum level, taxonomic classification based on V3-V4 sequencing across all fish samples, primarily identified two dominant taxa: Proteobacteria, which accounted for 94.3% of the classified sequences, and Bacteroidota, representing 2.5%. Taxonomic classification based on V1-V9 sequencing identified a broader range of taxa, including Proteobacteria, Actinobacteriota, Bacteroidota, Firmicutes, and Planctomycetota (Figure 8). At the genus

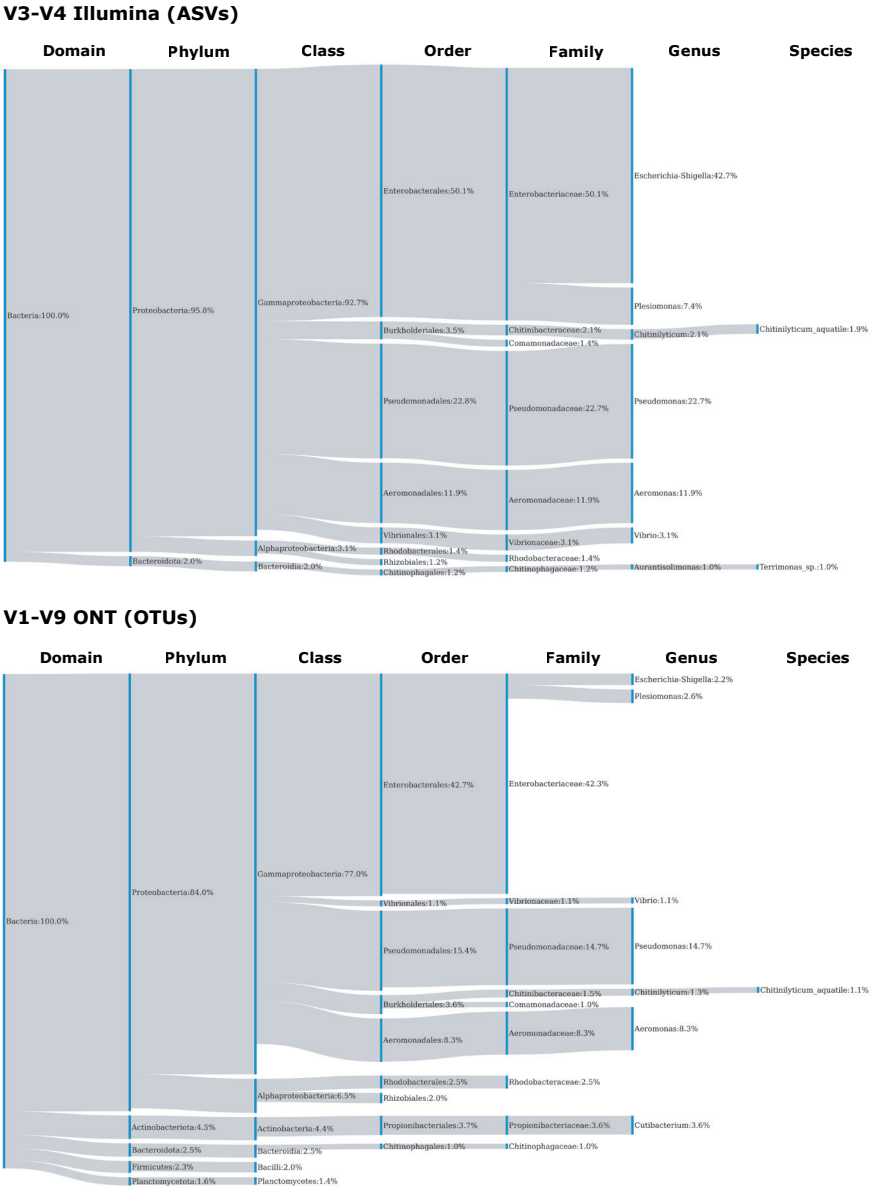


Figure 8. Sankey plots representing the taxonomy classification of microbial communities using (A) ASV-based classification from V3V4 regions and (B) OTU-based classification from nanopore sequencing of the V1V9 region. The width of the flows corresponds to the relative abundance of taxa, expressed as percentages, at each taxonomic level.

level, there were notable differences in the relative abundances of taxa between the two amplicon regions (Figure 7). For example, the genus *Cutibacterium* had a relative abundance of 3.6% in the V1-V9 analysis but was absent from the V3-V4 taxonomic assignment (Figure 8). Additionally, a substantial difference was observed in the relative abundance of Escherichia-Shigella, which comprised 42% of the assigned taxa in the V3-V4 analysis, while only 2.2% of the taxa were assigned to this genus in the V1-V9 analysis (Figure 8).

Nanopore Sequencing Demonstrates Lower Per-Sample Cost Compared to Illumina
The cost of microbiome composition studies is largely determined by the number of samples and the sequencing depth required. Illumina sequencing is usually outsourced to sequencing service providers, whereas nanopore sequencing can be done in-house. It is therefore of interest to evaluate cost of ONT sequencing compared to Illumina sequencing. Here we present a comparison of the cost associated with this study, from obtaining a purified DNA sample to the generation of sequencing data.

Nanopore library preparation and sequencing consist of two main steps: the customized library prep and ONT sequencing kit (Figure 2). For the customized library prep DNA samples were amplified using a forward primer with a barcode overhang at the 5' end, followed by purification with a PCR clean-up kit (Figure 2; see Methods). The cost of these steps is highly dependent on the choice of reagents, and there is a wide range of reagents with their associated cost available. In our study, the cost of attaching specific barcodes to PCR amplicons was 1.84 euros per sample. Further costs are associated with sample clean-up and sample preparations for sequencing. In our optimized procedure, samples are pooled together during sample preparation (Figure 2), which reduces the amount of reagents needed for the *end-prep* and *adapter ligation* steps of the Nanopore kit, since these steps were carried out in a single reaction (Figure 2). The total cost was €606.09 per sequencing run (table 3). As we pooled 15 samples (representing DNA from 15 gut samples) together, the cost per input sample was approximately €40.4 (Table 2).

Table 2 Library Preparation and Sequencing Cost Comparison for ONT V14 Ligation Kit. Cost estimates exclude consumables and working hours.

	Product	Product name	Price/unit (€)	Reactions/ unit	Price/reaction (€)	Reactions	ONT (€)
Barcoding PCR	Polymerase	LongAmp Hot Start Taq	199	500	0.4	15	6
	PCR clean-up kit	InviSorb® Fragment CleanUp	359	250	1.44	15	21.6
Sequencing	ONT kit	Ligation sequencing kit V14 (SQK-LSK114)	570	6	95	1	95
	3rd party reagents	Ultra II End-prep enzyme mix	287	24	11.9	1	11.9
		Quick T4 DNA ligase	1459	100	14.59	1	14.59
	Flowcell	FLO-MIN114	475	1	475	1	457
Total price							606.09
Price per sample							**40.4**

Discussion

In this study, we designed a PCR-based barcoding protocol for ONT 16S rRNA sequencing. We then demonstrated that the taxonomic classifications obtained from the V3-V4 region were highly consistent to those generated by the current gold standard, Illumina NovaSeq. Previous studies utilizing ONT for 16S rRNA sequencing have faced challenges in ASVs identification (i.e. taxonomic classification based on sequence variations at the single-nucleotide level) due to high error rates [240] and the reliance on ligation-based barcoding, which is costly, labour-intensive, and prone to biases [241,242]. Our approach achieved sufficient accuracy for taxonomic classification at the ASV level while significantly reducing per-sample sequencing costs and processing time. By integrating our optimized ONT workflow, we addressed these limitations and demonstrated its effectiveness in a real biological case study, assessing the gut microbiota of *cxcl8a+/+* and *cxcl8a-/-* zebrafish. Furthermore, we extended our analysis to compare full-length 16S rRNA V1-V9 sequencing with short-read V3-V4 sequencing, revealing an increased number of OTUs in long-read sequencing, albeit with some discrepancies in taxonomic assignments. This highlights the trade-offs between improved richness and potential classification challenges due to limitations in current 16S rRNA reference databases and the lack of optimized bioinformatics pipelines for full-length 16S rRNA gene analysis.

A major challenge in ONT-based taxonomic classification is the reliance on alignment-based bioinformatics pipelines, such as EPI2ME and EMU [243,244]. These methods do not perform *de novo* clustering, limiting the identification of OTUs or ASVs, which are essential for high-resolution microbiome analysis [240,243,244]. Our results demonstrate that ONT short-read sequencing can achieve ASVs classification with an accuracy comparable to Illumina when processed with an optimized bioinformatics pipeline. Using the latest ONT chemistry (V14) and the Dorado basecalling model, we obtained taxonomic classifications nearly identical to those generated by the current golden standard Illumina NovaSeq sequencing platform, confirming that the ONT sequencing platform is capable of resolving complex microbial community composition at high resolution. Although Illumina sequencing detected more ASVs, these additional variants did not increase taxonomic assignments, as alpha and beta diversity metrics showed strong concordance between platforms. Beta diversity analyses further demonstrated a strong overlap between ONT and Illumina-derived sequences. The microbiome composition of *cxcl8a* wild-type and gene deletion zebrafish mutant showed clearly visible differences in clustering and statistically significantly different relative abundances of specific genera. These differences were consistently detected across both sequencing platforms, reinforcing the reliability of the ONT sequencing platform for microbiome profiling next to the gold-standard Illumina platform.

Recent advancements, such as the ONT-AmpSeq pipeline, have improved taxonomic resolution by enabling OTU classification of full-length 16S rRNA sequences [237], raising interest in the potential benefits of long-read sequencing for microbiome profiling. To evaluate the taxonomic resolution of full-length 16S rRNA sequencing, we applied OTU-based classification via the ONT-AmpSeq workflow to ONT V1-V9 long sequencing reads and compared them to the ASV-based classification of V3-V4 reads from both ONT and Illumina. While full-length 16S rRNA gene sequencing detected a greater number of OTUs, this did not translate into improved taxonomic resolution compared to short-amplicon V3–V4, as the number of taxa assigned at the species level did not increase with long-read sequencing. This observation is consistent with the results of previous studies [245,246]. We detected some discrepancies, such as differences in *Escherichia-Shigella* abundance (42.7% in V3–V4 versus 2.2% in V1–V9). Our results suggest that the V3-V4 region may lack sufficient variability to detect taxonomically meaningful sequence variants within the *Enterobacteriaceae* family, potentially leading to an overestimation of *Escherichia-Shigella* due to extensive sequence similarity between this taxon with other related genera [245,247]. This reinforces previous findings that taxonomic resolution is highly dependent on the selected variable region, as different bacterial genera exhibit varying levels of conservation across the 16S rRNA gene. Therefore, relying on a single variable region, such as V3-V4, may not always provide accurate bacterial taxonomic classification [245,246,248]. Full-length 16S rRNA sequencing has the potential to address discrepancies in microbiota composition that arise from the use of different variable regions across studies. However, improvements are still needed in reference databases containing full-length 16S rRNA genes, as well as in the algorithms used to distinguish true biological variation from sequencing errors. Distinguishing intraspecies genetic variation remains a major challenge, which is further compounded by the need for algorithms capable of processing the substantially longer reads (1500 bp) produced by long-read sequencing technologies [246].

We developed a simplified and cost-effective PCR-based barcoding strategy that eliminates the need for ligation-based barcoding, significantly streamlining the workflow. Unlike ONT's standard barcoding kits, which require individual sample processing that may come with biases [241,242,249] our approach directly incorporates barcodes during an initial PCR amplification step, which enables simultaneous preparation of multiple samples. This study was conducted with up to 15 pooled samples; other studies in our lab have successfully pooled up to 98 samples (data not shown), further demonstrating the scalability and robustness of our strategy. Previous efforts to enhance ONT barcoding efficiency have made important progress, but usually still rely on ligation-based strategies or requirement for additional amplification steps, which can increase workload and introduce potential biases

[250,251]. Our method, using V14 ONT kits with novel chemistry, simplifies multiplexing while maintaining high sequencing accuracy, as evidenced by the balanced read distribution and successful demultiplexing of all samples. The consistent average read counts per sample obtained from both V3–V4 and V1–V9 ONT sequencing (approximately 242,000 and 244,700 reads per sample, respectively) indicate that our PCR-based barcoding strategy is effective regardless of amplicon length and the primers used. Rarefaction analysis further confirmed that a sequencing depth of around 50,000 reads per sample is sufficient to capture the full taxonomic diversity, demonstrating that our approach not only yields sample sequencing depth but is also well-suited for studying complex microbial communities, such as those found in the zebrafish gut. A pronounced decrease in read numbers was observed during primer trimming and filtering, particularly in samples with initially high read counts, which brought the final read distribution closer to the dataset's average. This suggests that a fraction of the excess reads in these samples likely represented amplicons with primer sequence errors, lacking primer regions, or chimeric sequences containing more than one primer sequence. In any of these cases, the reads were effectively discarded during the filtering process. Additionally, by allowing flexible primer selection, our approach overcomes the constraint of predefined 16S rRNA primer set kits, offering greater adaptability for microbiome studies [241,242].

Cost analysis revealed a substantial reduction in per-sample sequencing expenses associated with the ONT compared to Illumina sequencing platform. In our study, we employed a conservative pooling strategy of 15 samples per flow cell and still achieved a competitive per-sample cost compared to those typically charged by commercial sequencing providers. Subsequent studies in our lab, using the same strategy, have successfully pooled up to 98 samples per flow cell (data not shown), suggesting that further cost savings could be achieved. A key advantage of ONT sequencing is its portability, enabling in-house sequencing without the need for extensive and expensive laboratory infrastructure including highly skilled technicians.

Our study highlights the robustness of ONT short-read sequencing for microbiome profiling while maintaining comparable taxonomic classification accuracy as the Illumina sequencing platform, demonstrating the feasibility of ONT sequencing platforms as a cost-effective and scalable alternative. ONT platform portability provides workflow flexibility, making it a practical choice for in-house sequencing. By developing and implementing an optimized PCR-based barcoding strategy, we streamlined typical 16S rRNA gene sequencing workflows, reducing processing time and costs while enabling the simultaneous preparation of multiple samples. This approach enhanced the scalability of ONT sequencing, making it more amenable to large-scale microbiome studies. While full-length sequencing holds promise for

capturing greater taxonomic diversity, classification accuracy remained constrained by database limitations and lack of optimized algorithms. By optimizing ONT workflows, we provide a sequencing approach that balances cost, accuracy, and scalability, offering valuable insights into the future of nanopore-based microbiome research and taxonomic classifications.

Data Availability

All analyzed data supporting the findings of this chapter will be deposited in a public repository upon publication. The raw sequencing data have been deposited in the European Nucleotide Archive (ENA) under accession number PRJEB96324 and will be released upon publication. The pipeline developed for Nanopore-based 16S rRNA analysis will be made publicly available on GitHub upon publication.

Acknowledgments

We would like to thank prof. dr. Michiel Kleerebezem and prof. dr. Jerry Wells for their valuable feedback and careful review of the manuscript. Their insights contributed to the clarity and overall quality of the final work

Chapter 6

General Discussion

General Discussion

The rapid emergence of antimicrobial resistance (AMR) bacterial strains is a growing global health concern, and *S. suis* is no exception [1]. As both a major swine pathogen and an emerging zoonotic agent, *S. suis* represents a classic One Health challenge. In livestock production, metaphylactic use of antibiotics are prevent the spread *S. suis* invasive disease in fattening pens, and in some regions continue to be applied prophylactically or even for growth promotion[7,8]. Such practices create a strong selective pressure that promotes the development of AMR bacteria [9–11]. Infections caused by these strains can then be transmitted to humans directly or contribute resistance genes to other pathogens through horizontal gene transfer, with potential consequences for human health and environmental microbial communities [38]. Meanwhile, the efficacy of conventional antibiotics continues to decline, reinforcing the need for novel therapeutic strategies [9–11]. Against this backdrop, this thesis combines genome-wide functional genomics, mechanistic regulatory network analysis, and investigation of phase-variable Type I restriction–modification (RM) systems as epigenetic regulators of gene expression at the molecular level. By dissecting the bacterial mechanisms required for survival and adaptation in host niches, it identifies regulatory mechanisms and highlights potential targets for the rational design of new antibiotics. This integrated approach provides not only candidate targets but also an understanding of why such targets are critical for *S. suis* survival under host-like conditions.

One common strategy to identify new drug targets in bacteria is genome-wide screening for conditionally essential genes (CEG), defined as genes whose inactivation leads to growth impairment in laboratory media. A widely used approach for identifying these CEG is the use of transposon mutant library (Tn-library) screens (e.g. Tn-seq, TraDIS), performed under laboratory conditions. While such studies yield valuable fundamental knowledge of bacterial physiology, they may not reveal the genes required for survival *in vivo*. We addressed this limitation by performing Tn-seq in laboratory culture medium and physiologically relevant body fluids of the host such as active porcine serum (APS) and cerebrospinal fluid (CSF). By avoiding the use of animal models which have ethical considerations and technical drawbacks in terms of bottlenecks during infection we could screen in a biological fluids which are close to infection conditions. This strategy provided the foundation for the Tn-seq study in Chapter II and directly led to the discovery of key sensory systems explored in Chapter III.

1. Challenges in constructing Tn-libraries in non-model organisms or strains

An important practical consideration in this work was the difficulty of constructing high-coverage Tn-libraries in a non-model organism such as *S. suis*. In well-characterized bacteria, a broad toolkit of genetic resources enables efficient library generation. A commonly used strategy involves cloning the transposon and its transposase into a plasmid with a temperature-sensitive replication system, which is then introduced into the bacterial host. Expression of the transposase from an inducible promoter triggers insertion of the transposon into the genome. The plasmid carries a temperature sensitive replicon and is subsequently cured by elevating the temperature to block replication [252–255].

In *S. suis*, the limited availability of suitable plasmids with a temperature-sensitive origin of replication and inducible promoters for controlled expression of the transposase constrained our design. We therefore opted to perform transposon insertion into purified genomic DNA in vitro, followed by transformation of *S. suis* P1/7 via natural competence [256]. The success of this approach was made possible by the prior development of a natural competence induction method using synthetic ComS peptide [126].

In addition to *S. suis* P1/7, we attempted to construct libraries in multidrug-resistant (MDR) field isolates, but transformation efficiency was too low to obtain high-coverage libraries. Different ComS alleles pair with matching ComR regulators, making the system incompatible across species[257] . Given the high genetic diversity of *S. suis*, this specificity may also vary between strains, which could contribute to the low efficiency observed in MDR isolates [258–260]. Furthermore, differences in transformation efficiency between strains have been shown in other streptococci to result from low *comR* expression, suggesting that similar mechanisms might also operate in *S. suis* [261]. Developing such a system would have required extensive tool-building, including new plasmids, inducible promoters, and selection markers. This made it unfeasible within the timeline of the project, restricting our functional genomics analysis to transformable strains and highlights a technical bottleneck for expanding genome-wide mutagenesis to diverse *S. suis* lineages.

This experience illustrates a wider challenge in bacterial genetics: many veterinary and zoonotic pathogens remain technically difficult to manipulate compared to model species, which means our understanding of their biology is shaped as much by available tools as by scientific questions. Our development of such tools and techniques therefore makes an important contribution to the field.

2. Considerations for the Design and Interpretation of Tn-library Screens

One of the first and most critical considerations when designing a transposon screen is the selection of the environment in which the library will be tested, as **population bottlenecks** can drastically reduce library diversity. Bottlenecks occur when only a small fraction of the input mutant pool survives to establish infection, not because of the disrupted gene's role, but due to stochastic events such as random loss during host entry or early immune clearance. These effects are most severe in the initial stages of infection, where host barrier defenses, including mucosal surfaces, phagocytic clearance, or restricted tissue access, can limit the number of bacteria that successfully initiate colonization [89,262]. To circumvent these issues while still applying infection-relevant selective pressure, we designed our screens in host biological fluids that replicate specific stages of *S. suis* pathogenesis. APS models the bloodstream environment encountered during systemic spread, including presence of active complement factors, and CSF derived from choroid plexus organoids represents the infection niche after bacteria cross either the blood-brain barrier (BBB) or blood cerebral spinal fluid barrier (BCSFB) (Chapter II). Using these controlled host-derived environments avoided the strong bottlenecks of *in vivo* infection models while still providing physiologically relevant stressors. As a result, the Tn-seq screen identified 33 and 25 CEGs that were important for optimal *S. suis* growth in APS and CSF, respectively.

However, while transposon-based approaches offer a powerful, genome-wide method to study gene function, they also present several limitations that must be considered when interpreting the results. Because mutants are grown as a **pooled population**, inter-mutant interactions can influence fitness measurements. For instance, defective mutants may be rescued by others through cross-complementation, or conversely, may be outcompeted despite being independently viable. This means that pooled-screen fitness estimates may overestimate or underestimate the true contribution of a given gene to bacterial survival. By constructing single, gene-specific in-frame deletion mutants, it is possible to determine whether genes identified as conditionally essential (CE) in the pooled screen also show a fitness defect when tested in pure culture. From the list of genes identified in our screen, six were selected for targeted validation. Of these, three deletions (SSU_RS04755, SSU_RS02635, and SSU_RS07155) did not reproduce the fitness defect observed in the transposon library and displayed the same growth rate as the wild-type (WT) strain. This illustrates both the strength and the limitation of pooled approaches: they allow broad discovery, but careful single-gene validation remains necessary to disentangle artefacts from genuine biological requirements.

Another challenge arises from the **operonic organization** typical of bacteria. Insertions near the start of an operon can cause polar effects, disrupting the expression of downstream genes and complicating data interpretation. This was the case with the *liaFSR* operon. Our Tn-seq results suggested that disruption of *liaF* led to reduced fitness. However, we were unable to obtain a clean in-frame deletion mutant of *liaF*, whereas deletion mutants for *liaR* and *liaS* were successfully constructed and showed reduced fitness when grown in APS. This suggests that the observed fitness defect in the Tn-seq data may not reflect the effect of *liaF* disruption alone, but rather reflect polar effects of the inserted transposon, leading to downregulation of *liaR* and *liaS*. Such cases emphasize the need to interpret Tn-seq results in the genetic context of operons rather than at the level of individual genes, a point that becomes especially important in bacteria with compact, operon-rich genomes.

Another critical factor in Tn-seq experimental design is the **sampling time point**. Genes required only during specific growth phases may be missed, or their importance overestimated, if the library is assessed too early or too late. This was the case for the nucleotide ABC transporter characterized in this study. Based on the Tn-seq results, the mutant appeared to have a significant fitness defect. However, follow-up growth analysis of the single in-frame deletion mutant showed that the phenotype was limited to a delay during the exponential phase. The mutant ultimately reached the same concentration as the WT strain. Since the library was assessed during mid-exponential phase, the reduced representation of this mutant likely reflects a temporary growth lag rather than a sustained fitness cost. This example underscores how experimental design choices, and particularly the sampling time point, can bias interpretation.

The overall **library saturation** depends on the genome size of the bacterium and the number of independent mutants obtained, and it strongly influences the resolution of the screen. If resolution is insufficient, small genes or those in sparsely covered regions may lack insertions purely by chance. In Tn-seq analysis, such genes can be incorrectly classified as essential, simply because no mutants were recovered rather than due to a genuine fitness requirement. To address this issue, several statistical approaches have been developed to account for and correct these effects [89]. This risk is lower when studying CEG, since their contribution is inferred from relative changes in mutant abundance between laboratory media and host-mimicking conditions, and thus requires that the mutant is already present in the input library [131]. In our study, saturation of potential TA insertion sites ranged from 44% to 67%, with an average coverage of 91% across coding regions. In *Streptococcus*, the high density of TA sites means that even libraries with moderate coverage encompass most genes, and libraries with TA-site saturation above 35% have been reported to

yield high-confidence essentiality predictions[142]. The 91% representation of coding sequences observed here further supports the reliability of our library for identifying CEGs. By reporting these metrics transparently, we ensure that our conclusions are viewed considering library quality, which is essential for reproducibility and for comparison across studies.

Beyond coverage, the accuracy of **functional interpretation** also depends on the reliability of genome annotations. In *S. suis*, a large proportion of genes are assigned predicted functions based solely on sequence homology, without experimental validation. The high genetic diversity among strains further adds to this challenge, making it difficult to confidently interpret Tn-seq results, especially when attempting to identify enriched pathways or broader physiological trends [258–260]. Databases such as KEGG infer gene functions based on conserved domains, but these remain predictions. Some proteins may have multiple roles or regulatory interactions that are not captured by standard annotations. This limitation stresses importance of treating pathway-level insights as hypotheses rather than definitive assignments, and underscores the importance of targeted functional validation. These concepts shaped the research strategy described in this thesis and when possible Tn-seq results were, validated using precise, single-gene in-frame deletion and functional studies.

3. Functional complexity as a key aspect of conditional essentiality

In the tested conditions of this thesis. The genes we identified and studied in more detail are functionally complex because they participate in complex metabolic and regulatory networks. For example *purA*, the two-component system genes *liaS/liaR*, and the hypothetical protein SSU1817 (*hdrM*). Our aim was to explore why these genes appeared as CE and to evaluate whether their roles in environmental adaptation and regulation could make them suitable drug targets (Chapter II and III). In parallel, we investigated two DNA methylation systems identified in a comparative genomic study, which highlighted their high conservation among *S. suis* disease-associated strains (Chapter IV).

3.1 The importance of purine biosynthesis pathway in environmental adaptation

In Chapter II, the Tn-seq results revealed distinct metabolic requirements of *S. suis* depending on the growth niche. In APS, genes involved in *de novo* purine biosynthesis (*purA, purB, guaA, guaB,* and *pyrE*) and purine transporters were CE. This aligns with previous studies linking purine biosynthesis to bacterial colonization, intracellular

survival, increased susceptibility to oxidative stress, and reduced proliferation in human serum and blood [120,143,144,147–150]. However, a striking finding was that none of these purine biosynthesis or transport genes appeared in the Tn-seq list for CSF. Instead, in CSF, a large proportion of the identified CEGs were associated with amino acid transport systems, consistent with the low amino acid content of this fluid. This is further supported by ^13C-isotopologue profiling, which showed strong labeling of multiple amino acids under CSF conditions, indicating active *de novo* amino acid biosynthesis in this medium [118,262,263].

To investigate why nucleotide biosynthesis is important in APS but not CSF, we constructed an in-frame deletion mutant of *purA*, which had the largest fold change difference in our screen (–3.19 $\log_2$FC). We initially hypothesized that the nucleotide concentration in APS might be lower than in CSF or THY but culturing the *purA* mutant in chemically defined medium (CDM) with varying concentrations of nucleobases failed to restore growth, regardless of nucleobase type or concentration. Similar observations have been reported in other bacteria, where disruption of purine biosynthesis genes leads to growth defects that cannot be rescued by supplementation, often due to broader metabolic imbalances[263]. A plausible explanation is the role of *purA* in multiple interconnected processes that are essential for bacterial physiology. For example, purine nucleotides are essential precursors for key signaling molecules such as cyclic AMP, cyclic GMP, and they are alsothe principle energy molecules ATP and GTP, and the alarmone (p)ppGpp, which coordinate stress responses and global transcriptional changes [48,264].

We propose that in nutrient-limited media such as APS and CDM, purine biosynthesis is strongly activated to sustain proliferation, reflecting both the energy requirements (ATP, GTP) and the broader demand for purine nucleotides as building blocks for nucleic acids and signaling molecules. This is supported by our Tn-seq data, which identified only late-stage enzymes in the purine biosynthetic pathway (*purA*, *purB*, *guaA*, *guaB*) as CE in APS, whereas early pathway enzymes such as PurF were not required. Bacteria possess salvage pathways to recycle nucleotides but this process requires the terminal enzymes of the de novo pathway, including *purA*, *purB*, *guaA*, and *guaB*, therefore disruption of these genes is associated with a strong fitness defect in APS [264]. Another possible explanation for the higher impact of late-pathway disruptions is that early regulatory enzymes, such as PurF, can adjust metabolic flux in response to nucleotide levels, whereas inactivation of downstream enzymes commits resources irreversibly to the pathway, causing intermediate accumulation, nucleotide imbalance, and metabolic stress [264,265]. THY is nutrient-rich, and the abundance of metabolic precursors likely reduces the need for *de novo* purine biosynthesis, with alternative routes sufficient to maintain nucleotide homeostasis.

In APS, the limited nutrient availability makes it more difficult for the cell to access alternative sources of AMP and ATP.

Although CSF is essential for neuronal homeostasis, its composition is relatively nutrient-poor compared to serum, with low concentrations of proteins, amino acids, and glucose [118]. From a bacterial perspective, CSF therefore represents a nutrient-limited environment, requiring *de novo* biosynthesis of amino acids and other metabolites to sustain growth[262]. Our data suggest that in CSF, *S. suis* metabolism channels aspartate preferentially into amino acid biosynthesis rather than purine biosynthesis. Aspartate serves as a precursor for both pathways (Figure 3), but given the scarcity of free amino acids in CSF, *de novo* amino acid synthesis likely takes precedence to sustain protein production and growth, whereas purine demands may be partly met through reduced energy requirements in this niche. Taken together these results suggest that CSF represents a less-energy-demanding environment than APS. Serum contains active complement components that damage bacterial membranes via the alternative and lectin pathways, creating high demands for ATP and nucleotide precursors to sustain membrane integrity. In contrast, CSF lacks this immunological pressure, making it comparatively less stressful for *S. suis*[31]. This interpretation is supported by the absence of Ppc (a central catabolic gene in *S. suis*) in the Tn-seq list (Chapter II, page 27 to 29) of CEGs in CSF. Because Ppc drives anaplerotic flux into the TCA cycle, its absence as a CEG indicates that high-flux through the TCA cycle is less critical in CSF, consistent with lower overall energy requirements. In addition, deletions in cell envelope integrity genes such as *hdrM*, *liaR*, or *liaS* (Chapters II and III) were highly detrimental in APS but had no measurable effect in CSF (Chapters II). Together, these findings link complement-mediated envelope stress in APS to increased energy and nucleotide demands, while CSF allows growth with reduced pressure on both the TCA cycle and purine metabolism [266].

Further evidence of the connection between membrane stability and purine metabolism comes from the work of Shaffer et al., who identified a strong functional association between CvpA, a membrane stability-associated protein, and *purF*, the key regulatory enzyme initiating the *de novo* purine pathway [155]. Similarly, deletion of the membrane sensor *hdrM* resulted in downregulation of several purine biosynthesis enzymes, including *purA* and *guaA* (Chapter III), reinforcing the connection between cell envelope stress and purine metabolism and explaining why disruption of this pathway is detrimental in APS but not in CSF.

To summarize, the main purine biosynthesis genes identified in our Tn-seq screen, *purA, purB, guaA*, and *guaB***,** are required not only for *de novo* synthesis but also for the salvage pathway. Supplementation with nucleobases did not restore growth,

indicating that the limitation is most likely a shortage of adenine nucleotides (ATP, ADP, AMP) rather than a simple lack of precursors. We hypothesize that this shortage does not impair growth in THY because alternative metabolic routes can supply ATP. In contrast, APS imposes both nutrient limitation and additional stresses on the cell, as reflected by the importance of membrane-associated sensory systems, making *de novo* purine biosynthesis indispensable. In CSF, metabolism shifts toward amino acid biosynthesis, which may divert aspartate away from the purine pathway and at the same time reduce the risk of toxic intermediate accumulation, thereby mitigating the fitness impact of *purA* deletion. In THY membrane related stress is absent and nutrients abundant, explaining why *purA* deletion has no detectable effect. Further work should measure intracellular ATP/ADP ratios and membrane potential under APS and CSF conditions to clarify how nucleotide biosynthesis in supports growth under APS conditions

3.2. Tn-seq identification of membrane-stress sensors and complex regulatory networks

In Chapter II, using Tn-seq we identified membrane-stress sensors and in Chapter III we characterized their complex regulatory systems through genetic engineering techniques. In Chapter IV we explored the role of variable Type-I RM systems in altering gene regulatory networks and metabolism demonstrating how epigenetic mechanisms contribute to adaptation across different host environments.

3.2.1 Regulatory network of membrane-stress sensory systems

Our Tn-seq data also highlighted several small, poorly annotated proteins that were CE under our tested conditions. One such example is hypothetical protein, *hdrM* (660 bp), which showed a significantly reduced fitness (log_2FC –1.65) in APS. This phenotype was confirmed using targeted deletion mutants, which showed reduced proliferation in APS. The *hdrM* gene was predicted to encode a transmembrane protein, located within a putative operon together with *hdrR*, which encodes a putative DNA-binding regulator, indicating a possible role in environmental sensing and signal transduction (Chapter II).

In Chapter III, through structural modeling, genomic context analysis, and transcriptional profiling, we identified *hdrM* as part of a previously undescribed LytTR family regulatory system in *S. suis* P1/7 designated HdrMR. This system is widely distributed across Gram-positive bacteria, yet its functions remain poorly understood [169]. The first functional characterization of this system was provided by Merritt et al. in *S. mutans,* who demonstrated its interconnection with a second regulatory system, BrsMR [74,169]. These systems were shown to regulate adaptive traits such as biofilm formation and competence[75,168].

Our work extended these findings to *S. suis*, confirming the presence and conservation of both LytTR regulatory systems across multiple *S. suis* strains and demonstrating their cross-regulation in *S. suis* P1/7. Moreover, we found that their functional relevance in *S. suis* P1/7 appears to be strongly influenced by their regulatory integration with the LiaFSR three-component system.

Although the LiaFSR system is well characterized in other Gram-positive bacteria, it had not previously been described in *S. suis* [64,66,68,137,174,267]. Our Tn-seq data revealed that components of this system contribute to bacterial fitness in APS (Chapter II), prompting us to investigate it further. Subsequent genomic, structural, and transcriptomic analyses confirmed the presence of a functional LiaFSR system in *S. suis* P1/7 and revealed that it plays a central role in modulating the activity of both HdrMR and BrsMR, thereby linking multiple regulatory pathways involved in host adaptation.

A key insight from chapter III was that predicting the tertiary structure of a protein, rather than just its primary sequence, is a valuable approach to identify its function. For example, the *hdrM* CEG gene identified by Tn-seq, shared less than 30% amino acid sequence identity with homologous genes in other *Streptococcus* species. However, structural model alignments of potentially homolgous genes in *S. mutans* revealed a high degree of conformational similarity. This was later confirmed by a series of functional assays.

Moreover, similar tertiary structure was observed between HdrM and LiaF, another CEG identified by Tn-seq in *S. suis*. This similarity suggested that HdrM and LiaF might be transmembrane proteins which respond to comparable stimuli, a hypothesis supported by our subsequent demonstration of cross-regulation between the HdrMR and LiaFSR systems. These results underscore the value of structural conservation as a predictor of functional, even when sequence identity is low.

LiaR, the response regulator of the LiaFSR system, is known for its role in conferring resistance to lipid II–interfering antibiotics [63]. In this well-established model, phosphorylation of LiaR by its cognate histidine kinase LiaS leads to dimerization and activation of downstream mechanisms that promote cell envelope integrity [64–67]. Surprisingly, we found that unphosphorylated LiaR also exerts regulatory activity. Deletion of *liaS* did not significantly affect the expression of the LytTR systems, whereas deletion of *liaR* led to significant upregulation, with *hdrM*, *hdrR*, *brsM*, and *brsR* being the top four most upregulated genes. For the genes known to be regulated by *liaR*, there were no differences in Δ*liaR* and Δ*liaS,* suggesting that LiaR's regulatory functions are not complemented by other histidine kinases.

These results highlight the functional versatility of LiaR and underscore the complexity of regulatory systems, which often operate through multilayered interactions rather than linear signaling pathways. Further studies are needed to determine whether LiaR exerts this regulatory effect through direct interaction with the *hdrMR* and *brsMR* promoters, binding with HdrR/BrsR regulators or through indirect mechanisms involving additional regulatory factors.

Finally, we highlight how this regulatory cross-talk directly impacts virulence-associated traits. In particular, the impaired growth observed in the *liaR* deletion mutant was largely attributable to the overexpression of the regulators *hdrR* and *brsR*. While deletion of *hdrM* alone did not affect growth, combining it with *liaS* deletion reproduced the phenotype, supporting a dual control model in which both the release of HdrR and a reduction in LiaR levels are required to trigger the response.

Based on these observations, and the close structural similarity between HdrM and LiaF, we propose that HdrM and LiaF function as membrane-bound sensors that modulate regulatory activity in response to membrane status. When membrane integrity is compromised, such as under antibiotic stress, HdrM may sequester HdrR, while LiaS phosphorylates LiaR to activate membrane-stabilizing responses. In contrast, under conditions of membrane stability, such as in stationary phase or during reduced growth, HdrM may release HdrR and LiaF may inhibit LiaS, thereby reducing LiaR phosphorylation and decreasing the intracellular pool of active LiaR (Figure 1). Increased *hdrR* expression under high cell density conditions has been observed previously[75], supporting its potential role in density-dependent regulatory responses. Transcriptomic analysis of the *hdrM* mutant also revealed downregulation of multiple amino acid biosynthetic pathways and purine biosynthesis, suggesting a shift toward reduced metabolic activity and energy demand.

Integrating membrane status into signal-responsive regulation benefits processes such as competence and biofilm formation. In our study, simultaneous upregulation of *hdrR* and downregulation of *liaR,* produced in both traits levels comparable to, or exceeding, those of the WT. This could be explained, for example, by transcriptomics studies in *S. suis* using competence induction peptide (CPS) have shown that, during the first 15 minutes of competence, there is a global downregulation of biosynthetic metabolism and cell envelope synthesis [268]. *hdrR* remains upregulated throughout this 15-minute period compared to non-competent bacteria, whereas *liaR* shows a distinct peak only at 15 minutes. This peak corresponds to the maximum transformation efficiency, after which both the efficiency and *hdrR* expression decrease, and by 45 minutes the expression of *hdrR* and *liaR* is like that of the control without CPS (Figure 2) [268]. Similarly, studies in *S. suis* and *Vibrio vulnificus* have

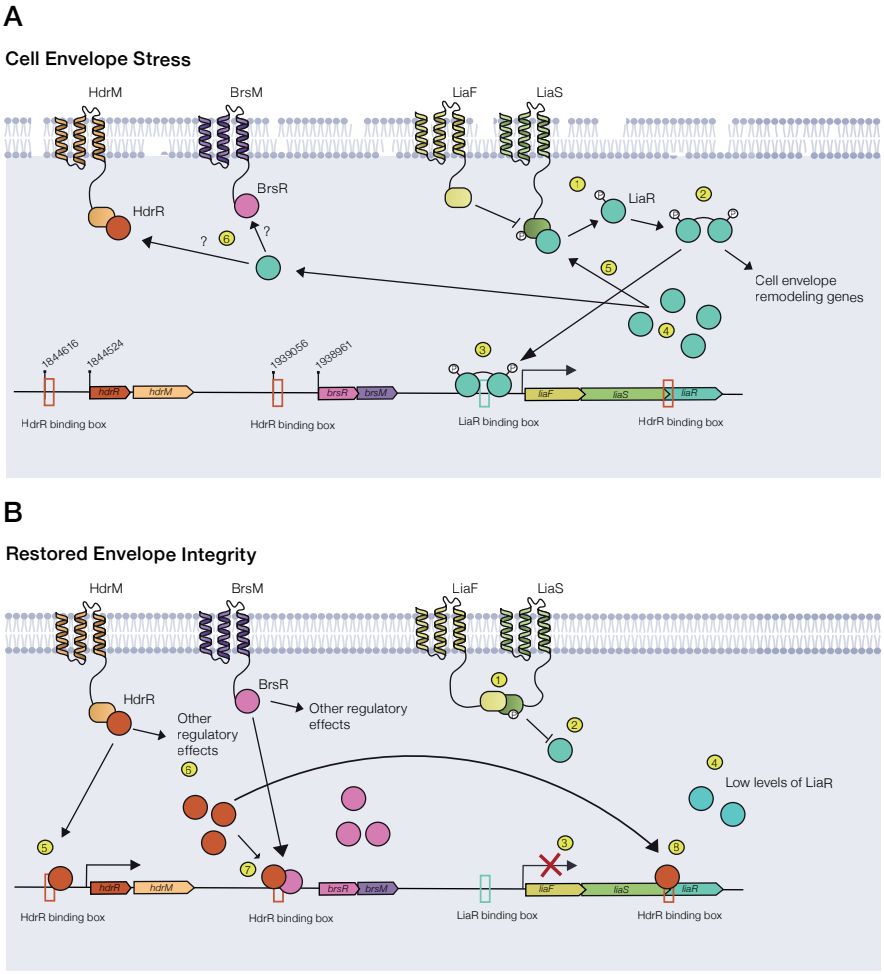


Figure 1. Proposed model of regulatory crosstalk between LiaFSR, HdrMR, and BrsMR systems in *S. suis*. (A) Under cell envelope stress, the sensor kinase LiaS phosphorylates LiaR, which activates expression of membrane protection genes and promotes its own expression through a positive feedback loop. High levels of unphosphorylated LiaR repress *hdrMR* and *brsMR* transcription, keeping HdrR and BrsR levels low. (B) When membrane integrity is restored, LiaR levels decrease, relieving repression of *hdrMR* and *brsMR*. HdrR and BrsR are derepressed and activate their own operons. HdrR also represses liaFSR, establishing a bidirectional regulatory circuit. This model integrates our findings and suggests that misregulation of this network impairs growth, competence, and biofilm formation. Numbers, refer to sequential processes occurring in the different physiological states relating to cell envelope stress or homeostasis. Figure derived from Chapter III.

examined the effect of the capsule on biofilm formation. In *V. vulnificus*, CPS-deficient mutants display increased cell surface hydrophobicity, which enhances attachment to surfaces and intercellular adhesion within biofilms, resulting in larger biofilms than those formed by the WT. Conversely, CPS overproduction reduces hydrophobicity, promoting cell dispersal and disrupting biofilm structure. The capsule appears to mask key adhesins or hydrophobic surface molecules involved in biofilm formation [30–32]. However, any compromise of envelope structure must be carefully controlled, as excessive destabilization could lead to osmotic lysis or increased susceptibility to immune defenses and antimicrobials. Consistent with this, our LL-37 experiments demonstrated that activated LiaR is critical for resistance to host-derived antimicrobial peptides.

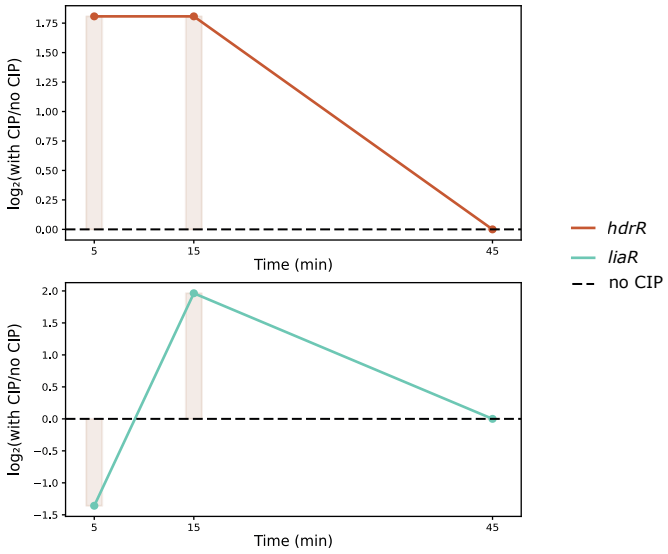


Figure 2. Data are from Zaccaria et al. [268], who reported fold change (FC) values with false discovery rate (FDR) correction. A value of 0 FC indicates that the difference was not statistically significant after FDR correction, whereas non-zero FC values correspond to significant differences. CIP / Control FCs at 15 min and 45 min were taken directly from the study. The 5 min values were inferred by combining the reported CIP / Control FC at 15 min with the fold change between 5 min and 15 min in the CIP-treated condition, also provided in the same dataset. This reconstruction assumes that expression in the control condition did not change significantly between 5 min and 15 min, as no data for this interval were available, while the reported fold change between 15 min and 45 min in the control condition was 0, indicating no significant change. The control baseline was therefore kept constant ($\log_2$FC = 0).

Our data suggest that the balance between LiaR and HdrR plays a key role in maintaining this equilibrium. Deletion of both *hdrR* and *brsR* completely abolished biofilm formation, underlining their importance in envelope-associated adaptations. Interestingly, a moderate reduction in LiaR levels, as observed in the *liaS* mutant, together with increased *hdrR* expression through *hdrM* deletion, appears to promote these traits, while complete loss of *liaR* causes broad regulatory disruption. This pattern likely reflects a physiological strategy in which the cell adjusts, but never fully abolishes, LiaR activity thereby safeguarding envelope integrity within a functional range.

The key insights from this chapter can be summarized into three main concepts: (i) the function of a protein is often better predicted by its conformational structure rather than its nucleotide or amino acid sequence; (ii) regulatory proteins often exhibit pleiotropy, influencing multiple, distinct biological processes simultaneously rather than acting in a single pathway; and (iii) core physiological traits, required for adaptation to host-like conditions, are typically regulated by multiple, overlapping systems, that provide robustness and buffer against misregulation.

3.2.2 Contribution of *S. suis* methylation systems shaping metabolism and virulence-adaptative traits

In Chapter IV, we studied two conserved Type I MS in *S. suis* strains (MS1 and MS2). MS1 *hsdS* contains inverted repeats that allow recombination between TDR domains resulting in 4 different phase variants of *hsdS*. Previous studies have shown that all 4 variants exist in culture, suggesting that the different phase variants may confer different phenotypes for adaptation in different environments [86]. Unlike MS1, MS2 specificity protein (*hsdS*) is not recombinogenic existing in two variants: truncated and full-length. In other bacteria like *S. pneumoniae* different phase variants alter gene expression through changes in the methylation pattern, in particular promoter sites[81]. MS2 is present in most disease-associated strains in a truncated form, except in ST1, where it occurs almost exclusively as a full-length variant. Interestingly, MS1 is predominantly found in lineage fre1 and, to a lesser extent, in lineage 5 and 10 (Chapter IV, Figure 2). In this chapter we investigated the effects of each variant using transcriptomic and phenotypic analyses, and our findings indicate epistatic interactions between MS1 and MS2 that shape metabolism and virulence-adaptative traits.

Locked MS1 phase variants revealed that each variant confers a distinct growth profile, with ms1.b providing the greatest growth advantage and ms1.a the poorest. This is in accordance with Atack J et al., who showed that ms1.b is the predominant variant when cultured in laboratory media [86]. These effects correlated with specific transcriptional changes, such as the upregulation of pilus and sortase dependent cell wall anchoring genes in ms1.a, which impose an energetic and envelope-remodelling

burden, reducing growth rate but likely providing benefits for host interaction *in vivo*. This observation is consistent with results of Chapter III, where energy intensive membrane remodelling was shown to reduced growth rate reinforcing the trade-off between proliferation and host adaptation. We also found that ms1.b effects on gene expression were dependent on the presence/absence of full length or truncated variant of the *hsdS* gene of MS2.

The main finding of this chapter is that MS1 modulates the regulatory effects of MS2, in both the full-length and truncated variants. In both cases, the most prominent transcriptional changes were the upregulation of protein synthesis genes and the downregulation of phosphotransferase systems (PTS). However, the PTS specificities differed substantially between variants. In the full-length MS2 background, the downregulated PTS systems targeted lactose and galactose, as well as the tagatose--6-phosphate pathway, whereas in the truncated MS2 background, the downregulated PTS systems mainly targeted mannose and fructose. Accordingly, growth in CDM with galactose as the sole carbon source, eliminates the growth advantage of the full length MS2 variant, whereas the truncated MS2 variant retained it.

Interestingly, in the full-length MS2 background, lactose/galactose PTS-tagatose operons (SSU0890-SSU0901) were also strongly downregulated, and this same operon was induced in the *liaR* mutant described in Chapter III. This operon encodes the transport and catabolism functions needed to import and convert lactose and galactose into glycolytic intermediates and ultimately into UDP-galactose, a direct building block for capsule polysaccharides. Its induction in the *liaR* mutant, which suffers from membrane instability, likely reflects a strategy to boost capsule biosynthesis as a protective response. Interestingly, in the *liaR* mutant, PTS genes were upregulated while incomplete TCA cycle genes were downregulated, whereas in the methylation system mutants the opposite pattern was observed, TCA cycle genes were upregulated while PTS genes were downregulated. We hypothesise that this inverse relationship arises from competition for phosphoenolpyruvate (PEP), a key metabolite needed for both PTS-mediated sugar uptake and anabolic reactions such as anaplerosis via PEP carboxylase (Figure 3). In model bacteria like *Escherichia coli* and *Enterococcus faecalis*, high PTS activity has been demonstrated to limit PEP availability for these alternative pathways, causing shifts in carbon flux and biomass allocation [269–271]. In *S. suis*, which has an incomplete TCA cycle (lacking α-ketoglutarate dehydrogenase and succinyl-CoA synthetase), it is plausible that a similar adjustment occurs. While direct evidence in *S. suis* or related bacteria is lacking, these patterns in well-characterised organisms make this a plausible and testable hypothesis.

Tight regulation of the PTS–tagatose pathway is therefore critical: its upregulation boosts capsule biosynthesis by increasing the supply of UDP-galactose precursors, but at the expense of diverting carbon and energy away from other biosynthetic pathways, ultimately limiting growth. Conversely, downregulation reduces capsule precursor availability, which could compromise capsule integrity, but frees metabolic resources to support higher growth rates.

The frequent co-occurrence of full-length MS2 with MS1 in pathogenic strains may reflect a need for MS1 to counterbalance the high metabolic activity driven by MS2. How this balance is maintained under nutrient-limited conditions remains unknown; in such environments, an inability to upregulate PTS systems could be detrimental, with consequences for nutrients uptake, capsule integrity and associated-virulence traits. This first description of the MS1–MS2 interaction highlights the importance of further experiments in nutrient-restricted media and with alternative carbon sources to understand how metabolism is regulated.

ST1 strains are the most strongly associated with disease and are also the main group carrying full-length MS2. In this form, MS2 downregulates PTS systems involved in lactose uptake and galactose catabolism. Since these sugars are important carbon sources during colonization of pig tissues, suppressing their utilization could severely impair colonization[212]. Our data suggest that MS1 exerts an inhibitory effect on MS2, which may explain why strains carrying full-length MS2 almost always also harbour MS1. The presence of MS1 could provide additional control over MS2 activity, allowing bacteria to utilize PTS systems and the tagatose pathway during tissue colonization while restricting them in other environments, such as blood, where glucose is abundant and rapid growth is advantageous. Consistent with this idea, Arenas *et al.* reported that the tagatose-6-phosphate and Leloir pathways, both involved in galactose metabolism, with the latter also supplying UDP-galactose as a precursor for capsule biosynthesis, were downregulated in *S. suis* isolated from heart tissue [118,272]. This regulatory interplay may also be linked to the strong association of ST1 with invasive disease.

Comparative analyses of *S. suis* strains carrying truncated MS2 but lacking MS1 could reveal how changes in genomic context influence adaptation. While our findings provide strong evidence for epistatic interactions between MS1 and MS2, the precise molecular basis of this relationship remains unresolved. Future work should combine methylome sequencing, with biochemical and genetic studies, to determine whether these effects arise from direct protein–protein interactions, DNA methylation–dependent transcriptional changes, or indirect network-level mechanisms. In summary, epigenetic interactions between MS1 and MS2 likely shaped the evolutionary emergence of the most virulent *S. suis* lineage.

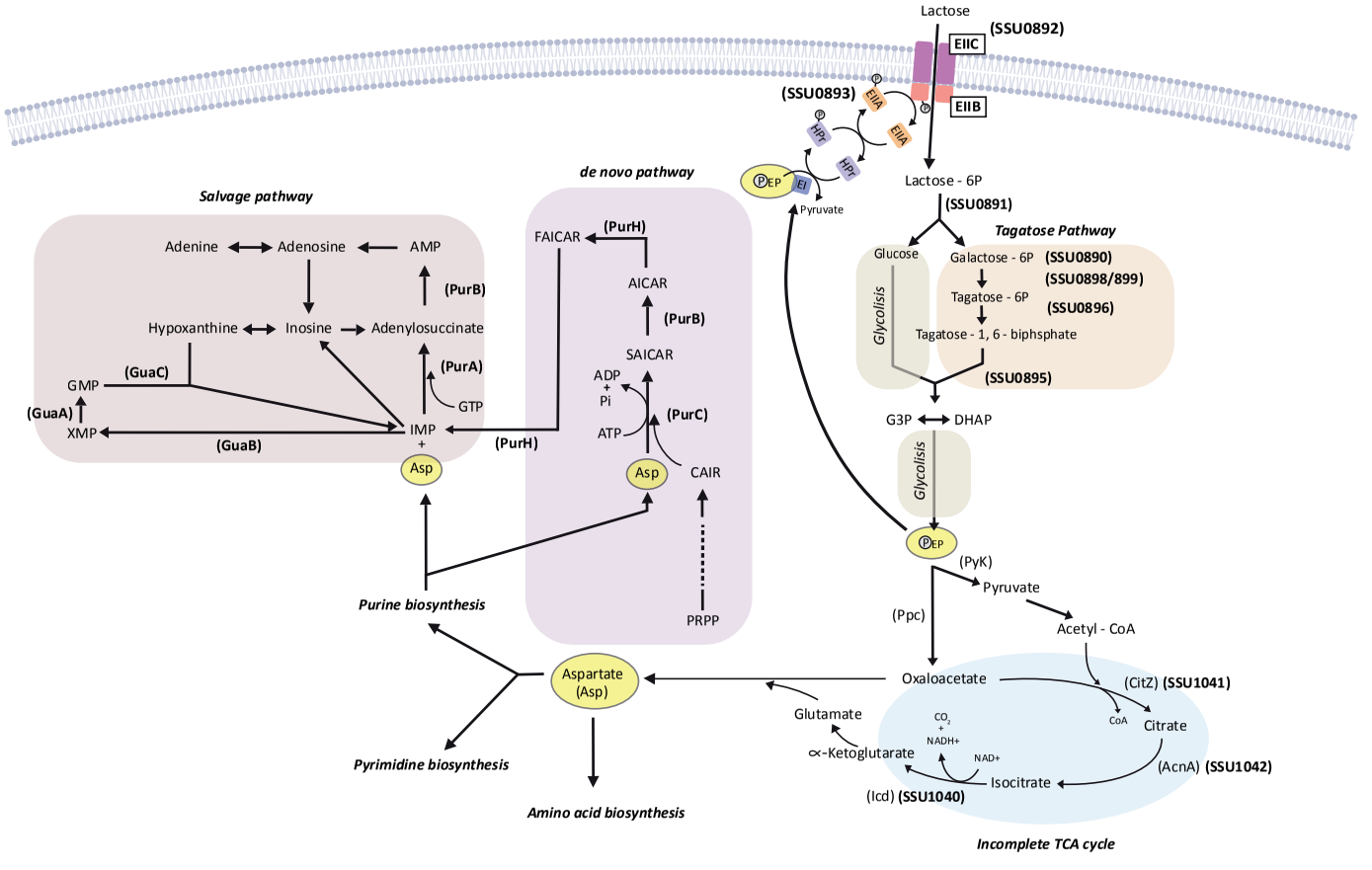


Figure 3. Schematic representation of PTS-mediated lactose metabolism through the tagatose-6-phosphate pathway, glycolysis, the incomplete TCA cycle, and aspartate-derived de novo and salvage nucleotide biosynthesis in *S. suis*. Gene identifiers correspond to enzymes analyzed in this chapters II, III and IV.

4. Potential new drug targets and future work

Defining a gene as conditionally essential can be misleading, as different strains may rely on different genes, or even distinct regulatory mechanisms, to perform the same function. A more robust approach is to identify conditionally essential functions, then determine which genes contribute to those functions and define the regulatory connections. This functional perspective allows us to better predict the relevance of potential targets across diverse strains and conditions.

Our data suggest that cross-regulation between the HdrMR and LiaFSR systems is one potential mechanism enabling dynamic control of cell envelope integrity. Disruption of either system alone, such as deletion of *liaR* or *hdrM*, reduces growth in APS and increases antimicrobial susceptibility, suggesting both could serve as therapeutic targets. However, during colonization processes such as biofilm formation, single deletions allowed partial recovery and compensation over time. In contrast, simultaneous deletion of *hdrR*, *liaR*, and *brsR* completely abolished biofilm formation. This inability to recover demonstrates that while *S. suis* can compensate for the loss of a single system, targeting multiple regulators at once removes this flexibility.

This compensatory capacity aligns with previous studies reporting that *liaR* deletion reduced virulence in sepsis models yet bacteria could still be recovered from tissues, while other studies observed no significant virulence difference between *liaR* mutants and WT in tissue infection models[67,68]. These findings reinforce that while regulators like *liaR* and *hdrM* are promising targets, their inhibition alone may not be sufficient. Instead, combining inhibition of these regulators with conventional antibiotics could increase killing before the bacteria can adapt regulatory network compensation occurs, thereby enhancing treatment efficacy. Targeting the histidine-kinase *liaS* would likely produce similar, though less pronounced, effects compared to *liaR*. Notably, the ATP-binding domains of histidine-kinases have already been explored as attractive candidates for the rational design of novel antibiotics [273] .

Our data further suggest that HdrM downregulates the purine biosynthesis pathway, which, based on our findings for *purA* in APS and other published studies, is likely to contribute to maintaining energy balance under stress conditions. This highlights purine metabolism as a but relatively underexplored antimicrobial target. Several enzymes in the *de novo* pathway have been proposed as antimicrobial targets, with PurK and PurE standing out because they are absent in humans [264,274,275]. Further studies are required to confirm their essentiality in *S. suis* under infection-relevant conditions and to evaluate their potential as safe and effective targets.

Similarly, MS1 *hsdM* merits further investigation as a potential target. In P1/7, deletion of MS1 markedly alters metabolism and reduces uptake of sugars, including those processed via the tagatose pathway, which are important for capsule synthesis. Targeting MS1 would therefore be most relevant in disease-associated strains carrying both full-length MS2 and MS1. To predict the broader applicability of such an approach, it is important to understand the effects of MS2 in strains lacking MS1, as the genetic context may strongly influence the outcome of MS1 inhibition.

As both HdrM and MS1 are conserved among disease-associated strains, their potential as therapeutic targets warrants further exploration. However, their precise effects on antimicrobial sensitivity, colonization, and *in vivo* fitness should be evaluated across multiple genetic backgrounds to assess the robustness of these targets. For translational progress, future studies must bridge the gap between controlled *in vitro* experiments and the heterogeneous, dynamic environments encountered during infection.

5. Incorporation and validation of Oxford Nanopore Technologies for Tn-library sequencing

We chose to incorporate Oxford Nanopore Technologies (ONT) sequencing into our transposon library workflow because it offered a practical and flexible in-house alternative to outsourcing Illumina sequencing (Chapter II). ONT devices allow rapid turnaround, independent experimental scheduling, and reduced logistical delays, which are particularly advantageous for iterative experiments like Tn-library screening.

We further wanted to validate that ONT could provide the sequence accuracy required for high-confidence mapping of transposon insertions. To do this, we used a complex and taxonomically diverse system, the zebrafish microbiota, as a benchmark. Sequencing 16S rRNA amplicons from zebrafish gut samples allowed us to compare ONT (R10.4.1 chemistry, Dorado basecalling) against Illumina and assess taxonomic resolution in a challenging community. The strong agreement in taxonomic composition and diversity metrics confirmed that ONT provides the accuracy needed for reliable transposon library analyses (Chapter V).

Building on this validation, we addressed another limitation of standard ONT workflows: the cost, time, and DNA requirements of the ligation-based barcoding step. In Chapter V, we describe the optimisation of a PCR-based barcoding strategy in which unique barcodes are incorporated directly during the 16S rRNA PCR amplification step. This strategy allows immediate pooling of amplicons after PCR

clean-up, reducing what could be up to 98 individual ligation reactions down to a single pooled ligation. In addition to saving time and reducing reagent use, this method significantly lowers the per-sample DNA requirement, an important consideration when working with low-yield extractions or large numbers of samples.

As part of this work, we also aimed to make ONT-based transposon library sequencing more accessible to other researchers. To this end, we will make the analysis script for Tn-library mapping publicly available, and the ONT 16S rRNA analysis pipeline and barcoding protocol will also be released openly. Providing these tools and methods to the scientific community is intended to facilitate adoption of ONT in similar applications and to reduce reliance on costly, inflexible commercial services.

By validating ONT accuracy in a complex microbiota system and optimizing its barcoding workflow, we have established a rapid, scalable, and cost-efficient sequencing pipeline that can be readily applied to Tn-library mapping and other high-throughput bacterial genomics applications.

Concluding remarks

By combining genome scale fitness mapping in host relevant fluids, mechanistic dissection of bacterial envelope surveillance systems, epigenetic control of metabolism, and practical sequencing innovations, this thesis advances our understanding of how *S. suis* adapts within the host. The findings reinforce that antibacterial strategies will benefit from targeting coordinated networks rather than single endpoints and from aligning interventions with the distinct demands of specific host niches. The methodological contributions lower the barrier to implementable, cost-conscious sequencing in laboratories that wish to perform similar studies. Together, these advances contribute to a framework for rational antimicrobial discovery that is grounded in the stressors bacteria face during infection.

References

1. World Health Organization. Antimicrobial resistance. *https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance* (2023).
2. Sun, D., Jeannot, K., Xiao, Y. & Knapp, C. W. Editorial: Horizontal gene transfer mediated bacterial antibiotic resistance. *Frontiers in Microbiology* vol. 10 Preprint at https://doi.org/10.3389/fmicb.2019.01933 (2019).
3. Napit, R. *et al.* Metagenomic analysis of human, animal, and environmental samples identifies potential emerging pathogens, profiles antibiotic resistance genes, and reveals horizontal gene transfer dynamics. *Sci Rep* 15, (2025).
4. Murray, G. G. R. *et al.* The emergence and diversification of a zoonotic pathogen from within the microbiota of intensively farmed pigs. *Proc Natl Acad Sci U S A* 120, (2023).
5. Weinert, L. A. *et al.* Genomic signatures of human and animal disease in the zoonotic pathogen Streptococcus suis. *Nat Commun* 6, (2015).
6. Feng, Y. *et al.* Streptococcus suis infection: An emerging/reemerging challenge of bacterial infectious diseases? *Virulence* vol. 5 477–497 Preprint at https://doi.org/10.4161/viru.28595 (2014).
7. Ma, L. *et al. Genomic Insight into the Antimicrobial Resistance of Streptococcus Suis-Six Countries, 2011-2019*. https://github.com/rrwick/.
8. Hadjirin, N. F. *et al.* Large-scale genomic analysis of antimicrobial resistance in the zoonotic pathogen Streptococcus suis. *BMC Biol* 19, (2021).
9. Aarestrup, F. M., Rasmussen, S. R., Artursson, K. & Jensen, N. E. *Trends in the Resistance to Antimicrobial Agents of Streptococcus Suis Isolates from Denmark and Sweden $*.
10. Vela, A. I. *et al.* Antimicrobial susceptibility of clinical strains of Streptococcus suis isolated from pigs in Spain. *Vet Microbiol* 105, 143–147 (2005).
11. Zhao, X. *et al.* Identification and characterization of Streptococcus suis strains isolated from eastern China Swine Farms, 2021–2023. *Sci Rep* 15, (2025).
12. Staats, J. J., Feder, I., Okwumabua, O. & Chengappa, M. M. *Streptococcus Suis: Past and Present*. *Veterinary Research Communications* vol. 21 (1997).
13. Fredriksen, S. *et al.* Environmental and maternal factors shaping tonsillar microbiota development in piglets. *BMC Microbiol* 22, (2022).
14. Wertheim, H. F. L., Nghia, H. D. T., Taylor, W. & Schultsz, C. Streptococcus suis: An emerging human pathogen. *Clinical Infectious Diseases* 48, 617–625 (2009).
15. Murray, G. G. R. *et al.* The emergence and diversification of a zoonotic pathogen from within the microbiota of intensively farmed pigs. *Proc Natl Acad Sci U S A* 120, (2023).
16. Yu H *et al.* Human Streptococcus suis outbreak, Sichuan, China. *Emerg Infect Dis* 12(6), 914–20 (2006).
17. Neila-Ibáñez, C. *et al.* Stochastic Assessment of the Economic Impact of Streptococcus suis-Associated Disease in German, Dutch and Spanish Swine Farms. *Front Vet Sci* 8, (2021).
18. Fredriksen, S. *et al.* Streptococcus suis infection on European farms is associated with an altered tonsil microbiome and resistome. *Microb Genom* 10, (2024).
19. Isabela Maria Fernandes de Oliveira. Exploiting the tonsil microbiota to prevent Streptococcus suis infections. (Wageningen University, 2023).
20. Roodsant, T. J., Van Der Putten, B. C. L., Tamminga, S. M., Schultsz, C. & Van Der Ark, K. C. H. Identification of Streptococcus suis putative zoonotic virulence factors: A systematic review and genomic meta-analysis. *Virulence* vol. 12 2787–2797 Preprint at https://doi.org/10.1080/21505594.2021.1985760 (2021).
21. Vötsch, D., Willenborg, M., Weldearegay, Y. B. & Valentin-Weigand, P. Streptococcus suis - The 'two faces' of a pathobiont in the porcine respiratory tract. *Frontiers in Microbiology* vol. 9 Preprint at https://doi.org/10.3389/fmicb.2018.00480 (2018).
22. Gottschalk M, S. M. Streptococcosis. In: Diseases of swine. *Wiley* 12, 934–950 (2019).
23. Segura, M., Fittipaldi, N., Calzas, C. & Gottschalk, M. Critical Streptococcus suis Virulence Factors: Are They All Really Critical? *Trends in Microbiology* vol. 25 585–599 Preprint at https://doi.org/10.1016/j.tim.2017.02.005 (2017).
24. Willemse, N. & Schultsz, C. Distribution of type I restriction-modification systems in streptococcus suis: An outlook. *Pathogens* 5, (2016).

25. Zhao, T. *et al.* Identification of plasminogen-binding sites in Streptococcus suis enolase that contribute to bacterial translocation across the blood-brain barrier. *Front Cell Infect Microbiol* 14, (2024).
26. Tram, G., Jennings, M. P., Blackall, P. J. & Atack, J. M. Streptococcus suis pathogenesis—A diverse array of virulence factors for a zoonotic lifestyle. in *Advances in Microbial Physiology* vol. 78 217–257 (Academic Press, 2021).
27. Smith, H. E., De Vries, R., Slot, R. V. t. & Smits, M. A. The cps locus Streptococcus suis serotype 2: Genetic determinant for the synthesis of sialic acid. *Microb Pathog* 29, 127–134 (2000).
28. Lecours, M. P. *et al.* Sialylation of Streptococcus suis serotype 2 is essential for capsule expression but is not responsible for the main capsular epitope. *Microbes Infect* 14, 941–950 (2012).
29. Larson, T. R. & Yother, J. Streptococcus pneumoniae capsular polysaccharide is linked to peptidoglycan via a direct glycosidic bond to β-D-N-acetylglucosamine. *Proc Natl Acad Sci U S A* 114, 5695–5700 (2017).
30. Gao, S. *et al.* Bacterial capsules: Occurrence, mechanism, and function. *npj Biofilms and Microbiomes* vol. 10 Preprint at https://doi.org/10.1038/s41522-024-00497-6 (2024).
31. Tanabe, S.-I. *et al.* Pleiotropic effects of polysaccharide capsule loss on selected biological properties of Streptococcus suis. *The Canadian Journal of Veterinary Research* 73, 65–70 (2008).
32. Lee, K. J., Kim, J. A., Hwang, W., Park, S. J. & Lee, K. H. Role of capsular polysaccharide (CPS) in biofilm formation and regulation of CPS production by quorum-sensing in Vibrio vulnificus. *Mol Microbiol* 90, 841–857 (2013).
33. Gottschalk, M. & Segura, M. The pathogenesis of the meningitis caused by Streptococcus suis: the unresolved questions. *Vet Microbiol* (2000).
34. Li, K. *et al.* Molecular characterization of Streptococcus suis isolates recovered from diseased pigs in Europe. *Vet Res* 55, 117 (2024).
35. Lun, Z.-R., Wang, Q.-P., Chen, X.-G., Li, A.-X. & Zhu, X.-Q. Streptococcus suis: an emerging zoonotic pathogen. https://doi.org/10.1016/S1473 (2007) doi:10.1016/S1473.
36. Scherrer, S. *et al.* Population structure, genetic diversity and pathotypes of Streptococcus suis isolated during the last 13 years from diseased pigs in Switzerland. *Vet Res* 51, (2020).
37. Wang, J. *et al.* Characterization of integrative and conjugative elements carrying antibiotic resistance genes of Streptococcus suis isolated in China. *Front Microbiol* 13, (2022).
38. Uruén, C., García, C., Fraile, L., Tommassen, J. & Arenas, J. How Streptococcus suis escapes antibiotic treatments. *Vet Res* 53, 91 (2022).
39. Hutchings, M., Truman, A. & Wilkinson, B. Antibiotics: past, present and future. *Curr Opin Microbiol* 51, 72–80 (2019).
40. Piddock, L. J. V., Malpani, R. & Hennessy, A. Challenges and Opportunities with Antibiotic Discovery and Exploratory Research. *ACS Infect Dis* 10, 2445–2447 (2024).
41. Johnson, B. K. & Abramovitch, R. B. Small Molecules That Sabotage Bacterial Virulence. *Trends Pharmacol Sci* 38, 339–362 (2017).
42. Abbas, A., Barkhouse, A., Hackenberger, D. & Wright, G. D. Antibiotic resistance: A key microbial survival mechanism that threatens public health. *Cell Host Microbe* 32, 837–851 (2024).
43. Farha, M. A., Tu, M. M. & Brown, E. D. Important challenges to finding new leads for new antibiotics. *Curr Opin Microbiol* 83, (2025).
44. Brown, E. D. & Wright, G. D. Antibacterial drug discovery in the resistance era. *Nature* 529, 336–343 (2016).
45. Juhas, M., Eberl, L. & Church, G. M. Essential genes as antimicrobial targets and cornerstones of synthetic biology. *Trends Biotechnol* 30, 601–607 (2012).
46. Block, A. M., Wiegert, P. C., Namugenyi, S. B. & Tischler, A. D. Transposon sequencing reveals metabolic pathways essential for Mycobacterium tuberculosis infection. *PLoS Pathog* 20, (2024).
47. Autret, N. & Charbit, A. Lessons from signature-tagged mutagenesis on the infectious mechanisms of pathogenic bacteria. *FEMS Microbiol Rev* 29, 703–717 (2005).
48. Goncheva, M. I., Chin, D. & Heinrichs, D. E. Nucleotide biosynthesis: the base of bacterial pathogenesis. *Trends Microbiol* 30, 793–804 (2022).
49. Lu, Y. *et al.* Modulation of MRSA virulence gene expression by the wall teichoic acid enzyme TarO. *Nat Commun* 14, (2023).

50. Wang, H. *et al.* Discovery of wall teichoic acid inhibitors as potential anti-MRSA β-lactam combination agents. *Chem Biol* 20, 272–284 (2013).
51. Fernandez-Ciruelos, B., Potmis, T., Solomin, V. & Wells, J. M. Cross-talk between QseBC and PmrAB two-component systems is crucial for regulation of motility and colistin resistance in Enteropathogenic Escherichia coli. *PLoS Pathog* 19, (2023).
52. Elsen, S., Simon, V. & Attrée, I. Cross-regulation and cross-talk of conserved and accessory two-component regulatory systems orchestrate Pseudomonas copper resistance. *PLoS Genet* 20, (2024).
53. Atack, J. M., Tan, A., Bakaletz, L. O., Jennings, M. P. & Seib, K. L. Phasevarions of Bacterial Pathogens: Methylomics Sheds New Light on Old Enemies. *Trends Microbiol* 26, 715–726 (2018).
54. Lee, J. W. Bacterial Regulatory Mechanisms for the Control of Cellular Processes: Simple Organisms' Complex Regulation. *Journal of Microbiology* 61, 273–276 (2023).
55. He, L. Y., Le, Y. J., Guo, Z., Li, S. & Yang, X. Y. The Role and Regulatory Network of the CiaRH Two-Component System in Streptococcal Species. *Front Microbiol* 12, (2021).
56. Zheng, C. *et al.* Role of two-component regulatory systems in the virulence of Streptococcus suis. *Microbiol Res* 214, 123–128 (2018).
57. Krell, T. *et al.* Bacterial sensor kinases: Diversity in the recognition of environmental signals. *Annu Rev Microbiol* 64, 539–559 (2010).
58. Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu Rev Biochem* (2000).
59. Tymoszewska, A., Szylińska, M. & Aleksandrzak-Piekarczyk, T. The LiaFSR-LiaX System Mediates Resistance of Enterococcus faecium to Peptide Antibiotics and to Aureocin A53- and Enterocin L50-Like Bacteriocins. *Microbiol Spectr* 11, (2023).
60. Buelow, D. R. & Raivio, T. L. Three (and more) component regulatory systems - Auxiliary regulators of bacterial histidine kinases. *Mol Microbiol* 75, 547–566 (2010).
61. Krell, T. *et al.* Bacterial sensor kinases: Diversity in the recognition of environmental signals. *Annu Rev Microbiol* 64, 539–559 (2010).
62. Capra, E. J. & Laub, M. T. Evolution of two-component signal transduction systems. *Annu Rev Microbiol* 66, 325–347 (2012).
63. Mascher, T. Intramembrane-sensing histidine kinases: A new family of cell envelope stress sensors in Firmicutes bacteria. *FEMS Microbiol Lett* 264, 133–144 (2006).
64. Fritsch, F. *et al.* The cell envelope stress response mediated by the LiaFSRLm three-component system of Listeria monocytogenes is controlled via the phosphatase activity of the bifunctional histidine kinase LiaSLm. *Microbiology (N Y)* 157, 373–386 (2011).
65. Bombelli, A. *et al.* Effects of the antimicrobial glabridin on membrane integrity and stress response activation in Listeria monocytogenes. *Food Research International* 175, (2024).
66. Lin, Y. *et al.* Exportal and the liafsr regulatory system coordinate the response to cell membrane stress in streptococcus pyogenes. *mBio* 11, 1–15 (2020).
67. Klinzing, D. C. *et al.* The two-component response regulator LiaR regulates cell wall stress responses, pili expression and virulence in group B Streptococcus. *Microbiology (United Kingdom)* 159, 1521–1534 (2013).
68. Sanson, M. A. *et al.* The LiaFSR Transcriptome Reveals an Interconnected Regulatory Network in Group A Streptococcus. *Infect Immun* 89, (2021).
69. Otero-Asman, J. R. *et al.* The extracytoplasmic function sigma factor σVreI is active during infection and contributes to phosphate starvation-induced virulence of Pseudomonas aeruginosa. *Sci Rep* 10, 3139 (2020).
70. Le Jeune, A. *et al.* The extracytoplasmic function sigma factor SigV plays a key role in the original model of lysozyme resistance and virulence of Enterococcus faecalis. *PLoS One* 5, (2010).
71. Mascher, T. Signaling diversity and evolution of extracytoplasmic function (ECF) σ factors. *Curr Opin Microbiol* 16, 148–155 (2013).
72. Grove, A. Extracytoplasmic Function Sigma Factors Governing Production of the Primary Siderophores in Pathogenic Burkholderia Species. *Front Microbiol* 13, (2022).

73. Zou, Z. *et al.* LytTR Regulatory Systems: A potential new class of prokaryotic sensory system. *PLoS Genet* 14, (2018).
74. Xie, Z. *et al.* Regulatory control of the Streptococcus mutans HdrRM LytTR Regulatory System functions via a membrane sequestration mechanism. *Mol Microbiol* 114, 681–693 (2020).
75. Merritt, J., Zheng, L., Shi, W. & Qi, F. Genetic characterization of the hdrRM operon: A novel high-cell-density-responsive regulator in Streptococcus mutans. *Microbiology (N Y)* 153, 2765–2773 (2007).
76. Okinaga, T., Niu, G., Xie, Z., Qi, F. & Merritt, J. The hdrRM operon of Streptococcus mutans encodes a novel regulatory system for coordinated competence development and bacteriocin production. *J Bacteriol* 192, 1844–1852 (2010).
77. Van Der Woude, M. W. & Bäumler, A. J. Phase and antigenic variation in bacteria. *Clin Microbiol Rev* 17, 581–611 (2004).
78. Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches - The ON and OFF of bacterial phase variation. *Mol Microbiol* 33, 919–932 (1999).
79. Sánchez-Romero, M. A., Cota, I. & Casadesús, J. DNA methylation in bacteria: From the methyl group to the methylome. *Curr Opin Microbiol* 25, 9–16 (2015).
80. Seib, K. L., Srikhanta, Y. N., Atack, J. M. & Jennings, M. P. Epigenetic Regulation of Virulence and Immunoevasion by Phase-Variable Restriction-Modification Systems in Bacterial Pathogens. *Annu Rev Microbiol* 15, 22 (2025).
81. Manso, A. S. *et al.* A random six-phase switch regulates pneumococcal virulence via global epigenetic changes. *Nat Commun* 5, (2014).
82. Srikhanta, Y. N., Maguire, T. L., Stacey, K. J., Grimmond, S. M. & Jennings, M. P. The phasevarion: A genetic system controlling coordinated, random switching of expression of multiple genes. *National Institutes of Health* (2005).
83. Phillips, Z. N., Husna, A. U., Jennings, M. P., Seib, K. L. & Atack, J. M. Phasevarions of bacterial pathogens-phase-variable epigenetic regulators evolving from restriction-modification systems. *Microbiology (United Kingdom)* 165, 917–928 (2019).
84. Seib, K. L., Jen, F. E. C., Scott, A. L., Tan, A. & Jennings, M. P. Phase variation of DNA methyltransferases and the regulation of virulence and immune evasion in the pathogenic Neisseria. *Pathog Dis* 75, (2017).
85. Adamczyk-Poplawska, M., Lower, M. & Piekarowicz, A. Deletion of one nucleotide within the homonucleotide tract present in the hsdS gene alters the DNA sequence specificity of type I restriction-modification system NgoAV. *J Bacteriol* 193, 6750–6759 (2011).
86. Atack, J. M. *et al.* Streptococcus suis contains multiple phase-variable methyltransferases that show a discrete lineage distribution. *Nucleic Acids Res* 46, 11466–11476 (2018).
87. Wileman, T. M. *et al.* Pathotyping the Zoonotic Pathogen Streptococcus suis: Novel genetic markers to differentiate invasive disease-associated isolates from non-disease-associated isolates from England and Wales. *J Clin Microbiol* 57, (2019).
88. Bourgeois, J. & Camilli, A. High-Throughput Mutant Screening via Transposon Sequencing. *Cold Spring Harb Protoc* 2023, 707–709 (2023).
89. Cain, A. K. *et al.* A decade of advances in transposon-insertion sequencing. *Nat Rev Genet* 21, (2020).
90. Barquist, L., Boinett, C. J. & Cain, A. K. Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol* 10, 1161–1169 (2013).
91. Basta, D. W. *et al.* Inducible transposon mutagenesis identifies bacterial fitness determinants during infection in mice. *Nat Microbiol* 10, 1171–1183 (2025).
92. Van Opijnen, T. & Camilli, A. *Transposon Insertion Sequencing: A New Tool for Systems-Level Analysis of Microorganisms*. (2013).
93. Quainoo, S. *et al.* Whole-genome sequencing of bacterial pathogens: The future of nosocomial outbreak analysis. *Clin Microbiol Rev* 30, 1015–1063 (2017).
94. Richardson, E. J. & Watson, M. The automatic annotation of bacterial genomes. *Brief Bioinform* 14, 1–12 (2013).
95. Luo, H., Gao, F. & Lin, Y. Evolutionary conservation analysis between the essential and nonessential genes in bacterial genomes. *Sci Rep* 5, (2015).

96. Fields, F. R., Lee, S. W. & McConnell, M. J. Using bacterial genomes and essential genes for the development of new antibiotics. *Biochem Pharmacol* 134, 74–86 (2017).
97. Wick, R. R., Judd, L. M. & Holt, K. E. Assembling the perfect bacterial genome using Oxford Nanopore and Illumina sequencing. *PLoS Comput Biol* 19, (2023).
98. Linde, J. *et al.* Comparison of Illumina and Oxford Nanopore Technology for genome analysis of Francisella tularensis, Bacillus anthracis, and Brucella suis. *BMC Genomics* 24, (2023).
99. Bird, M. T. *et al.* Use of Nanopore Sequencing to Characterise the Genomic Architecture of Mobile Genetic Elements Encoding blaCTX-M-15 in Escherichia coli Causing Travellers' Diarrhoea. *Front Microbiol* 13, (2022).
100. Wohlers, I., Garg, S. & Hehir-Kwa, J. Y. Editorial: Long-read sequencing—Pitfalls, benefits and success stories. *Frontiers in Genetics* vol. 13 Preprint at https://doi.org/10.3389/fgene.2022.1114542 (2023).
101. Zhang, T. *et al.* The newest Oxford Nanopore R10.4.1 full-length 16S rRNA sequencing enables the accurate resolution of species-level microbial community profiling. *Appl Environ Microbiol* 89, (2023).
102. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 17, (2016).
103. Pugh, J. The Current State of Nanopore Sequencing. *Methods in molecular biology (Clifton, N.J.)* vol. 2632 3–14 Preprint at https://doi.org/10.1007/978-1-0716-2996-3_1 (2023).
104. Zhao, T. *et al.* Choroid plexus organoids reveal mechanisms of Streptococcus suis translocation at the blood-cerebrospinal fluid barrier. *Fluids Barriers CNS* 22, 14 (2025).
105. Arenas Busto, J. *Streptococcus Suis: Methods and Protocols. Humana, New York, NY* (2024). doi:10.1007/978-1-0716-3898-9.
106. Segura, M. Streptococcus suis research: Progress and challenges. *Pathogens* 9, 1–8 (2020).
107. Takeuchi, D. *et al.* Impact of a food safety campaign on streptococcus suis infection in humans in Thailand. *American Journal of Tropical Medicine and Hygiene* 96, 1370–1377 (2017).
108. Ho, D. T. N. *et al.* Risk factors of Streptococcus suis infection in Vietnam. A case-control study. *PLoS One* 6, (2011).
109. Mai, N. T. H. *et al.* Streptococcus suis meningitis in adults in Vietnam. *Clinical Infectious Diseases* 46, 659–667 (2008).
110. Segura, M., Calzas, C., Grenier, D. & Gottschalk, M. Initial steps of the pathogenesis of the infection caused by Streptococcus suis: fighting against nonspecific defenses. *FEBS Lett* 590, 3772–3799 (2016).
111. Segura, M. & Gottschalk, M. Streptococcus suis interactions with the murine macrophage cell line J774: Adhesion and cytotoxicity. *Infect Immun* 70, 4312–4322 (2002).
112. Williams, A. E. & Blakemore, W. F. Pathogenesis of Meningitis Caused by Streptococcus suis Type 2. *J Infect Dis* 162, 474–481 (1990).
113. Tenenbaum, T., Seitz, M., Schroten, H. & Schwerk, C. Biological activities of suilysin: Role in Streptococcus suis pathogenesis. *Future Microbiol* 11, 941–954 (2016).
114. Roy, D. *et al. Role of the Capsular Polysaccharide as a Virulence Factor for Streptococcus Suis Serotype 14*. *Can J Vet Res* (2014).
115. Singh Chhatwal Editor, G. Host-Pathogen Interactions in Streptococcal Diseases. *Curr Top Microbiol Immunol* (2013).
116. Rohmer, L., Hocquet, D. & Miller, S. I. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends Microbiol* 19, 341–348 (2011).
117. Pellegrini, L. *et al.* Human CNS barrier-forming organoids with cerebrospinal fluid production. *Science (1979)* 369, (2020).
118. Otto, C. *et al.* Comprehensive analysis of the cerebrospinal fluid and serum metabolome in neurological diseases. *Journal of Neuroinflammation* 21, (2024).
119. Nascimento, F. P., Macedo-Júnior, S. J., Lapa-Costa, F. R., Cezar-dos-Santos, F. & Santos, A. R. S. Inosine as a Tool to Understand and Treat Central Nervous System Disorders: A Neglected Actor? *Front Neurosci* 15, (2021).
120. Arenas, J. *et al.* Identification of conditionally essential genes for Streptococcus suis infection in pigs. *Virulence* 11, 446–464 (2020).

121. Van Opijnen, T., Bodi, K. L. & Camilli, A. tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods* 6, 767 (2009).
122. ATCC. Streptococcus suis (BAA-853). *https://www.atcc.org/products/baa-853*.
123. Van Opijnen, T. & Camilli, A. *Genome-Wide Fitness and Genetic Interactions Determined by Tn-Seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms*. (2014).
124. G. Venema. Molecular Biology and Genetic Modification of Lactococci. *J Dairy Sci* 76(8), 2133–214 (1993).
125. Van Der Els, S., James, J. K., Kleerebezem, M. & Bron, P. A. Versatile Cas9-Driven Subpopulation Selection Toolbox for Lactococcus lactis. *Appl Environ Microbiol* (2018).
126. Zaccaria, E. *et al.* Control of competence for DNA transformation in Streptococcus suis by genetically transferable pherotypes. *PLoS One* 9, (2014).
127. Gussak, A., Ferrando, M. L., Schrama, M., van Baarlen, P. & Wells, J. M. Precision Genome Engineering in Streptococcus suis Based on a Broad-Host-Range Vector and CRISPR-Cas9 Technology. *ACS Synth Biol* 12, 2546–2560 (2023).
128. Jordan, S., Junker, A., Helmann, J. D. & Mascher, T. Regulation of LiaRS-dependent gene expression in Bacillus subtilis: Identification of inhibitor proteins, regulator binding sites, and target genes of a conserved cell envelope stress-sensing two-component system. *J Bacteriol* 188, 5153–5166 (2006).
129. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: Scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018 (2011).
130. Dejesus, M. A. & Ioerger, T. R. Normalization of transposon-mutant library sequencing datasets to improve identification of conditionally essential genes. in *Journal of Bioinformatics and Computational Biology* vol. 14 (World Scientific Publishing Co. Pte Ltd, 2016).
131. DeJesus, M. A., Ambadipudi, C., Baker, R., Sassetti, C. & Ioerger, T. R. TRANSIT - A Software Tool for Himar1 TnSeq Analysis. *PLoS Comput Biol* 11, (2015).
132. Gómez-Gascón, L. *et al.* Comparative immunosecretome analysis of prevalent Streptococcus suis serotypes. *Comp Immunol Microbiol Infect Dis* 57, 55–61 (2018).
133. Weiße, C. *et al.* Immunogenicity and protective efficacy of a Streptococcus suis vaccine composed of six conserved immunogens. *Vet Res* 52, 112 (2021).
134. Franza, T. *et al.* NAD+ pool depletion as a signal for the Rex regulon involved in Streptococcus agalactiae virulence. *PLoS Pathog* 17, (2021).
135. Jahn, K., Kohler, T. P., Hammerschmidt, S., Swiatek, L. S. & Wiebe, S. Platelets, Bacterial Adhesins and the Pneumococcus. *Cells* vol. 11 Preprint at https://doi.org/10.3390/cells11071121 (2022).
136. Wolf, D. *et al.* In-depth profiling of the LiaR response of bacillus subtilis. *J Bacteriol* 192, 4680–4693 (2010).
137. Suntharalingam, P., Senadheera, M. D., Mair, R. W., Levesque, C. M. & Cvitkovitch, D. G. The LiaFSR system regulates the cell envelope stress response in streptococcus mutans. *J Bacteriol* 191, 2973–2984 (2009).
138. Vega, L. A. *et al.* LiaR-dependent gene expression contributes to antimicrobial responses in group A *Streptococcus*. Preprint at https://doi.org/10.1101/2024.04.04.588141 (2024).
139. Abdullah, M. R. *et al.* Crystal Structure and Pathophysiological Role of the Pneumococcal Nucleoside-binding Protein PnrA. *J Mol Biol* 433, (2021).
140. Deka, R. K. *et al.* The PnrA (Tp0319; TmpC) lipoprotein represents a new family of bacterial purine nucleoside receptor encoded within an ATP-binding cassette (ABC)-like operon in Treponema pallidum. *Journal of Biological Chemistry* 281, 8072–8081 (2006).
141. Bleuzé, M., Lavoie, J. P., Bédard, C., Gottschalk, M. & Segura, M. Encapsulated Streptococcus suis impairs optimal neutrophil functions which are not rescued by priming with colony-stimulating factors. *PLoS One* 19, (2024).
142. Rosconi, F. *et al.* A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat Microbiol* 7, 1580–1592 (2022).
143. Dresen, M. *et al.* Streptococcus suis TrpX is part of a tryptophan uptake system, and its expression is regulated by a T-box regulatory element. *Sci Rep* 12, (2022).
144. Breton, Y. Le *et al.* Genome-wide identification of genes required for fitness of group a streptococcus in human blood. *Infect Immun* 81, 862–875 (2013).
145. Saxena, S., Khan, N., Dehinwal, R., Kumar, A. & Sehgal, D. Conserved surface accessible nucleoside ABC transporter component SP0845 is essential for pneumococcal virulence and confers protection in vivo. *PLoS One* 10, (2015).

146. Jain, S., Sutchu, S., Rosa, P. A., Byram, R. & Jewett, M. W. Borrelia burgdorferi harbors a transport system essential for purine salvage and mammalian infection. *Infect Immun* 80, 3086–3093 (2012).
147. Mantena, R. K. R. *et al.* Reactive oxygen species are the major antibacterials against Salmonella Typhimurium purine auxotrophs in the phagosome of RAW 264.7 cells. *Cell Microbiol* 10, 1058–1073 (2008).
148. Schauer, K. *et al.* Deciphering the intracellular metabolism of Listeria monocytogenes by mutant screening and modelling. *BMC Genomics* 11, (2010).
149. Li, L. *et al.* Role of purine biosynthesis in persistent methicillin-resistant staphylococcus aureus infection. *Journal of Infectious Diseases* 218, 1367–1377 (2018).
150. Samant, S. *et al.* Nucleotide Biosynthesis Is Critical for Growth of Bacteria in Human Blood. *PLoS Pathog* (2008).
151. Mei, J. M., Nourbakhsh, F., Ford, C. W. & Holden, D. W. Identification of Staphylococcus auerus virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. *Mol Microbiol* 26, 399–407 (1997).
152. Wilde, A. D. *et al.* Bacterial Hypoxic Responses Revealed as Critical Determinants of the Host-Pathogen Outcome by TnSeq Analysis of Staphylococcus aureus Invasive Infection. *PLoS Pathog* 11, (2015).
153. Lan, L., Cheng, A., Dunman, P. M., Missiakas, D. & He, C. Golden pigment production and virulence gene expression are affected by metabolisms in Staphylococcus aureus. *J Bacteriol* 192, 3068–3077 (2010).
154. Valentino, M. D. *et al.* Genes contributing to Staphylococcus aureus fitness in abscess- and infection-related ecologies. *mBio* 5, (2014).
155. Shaffer, C. L. *et al.* Purine biosynthesis metabolically constrains intracellular survival of uropathogenic Escherichia coli. *Infect Immun* 85, (2017).
156. Huemer, M. *et al.* Molecular reprogramming and phenotype switching in Staphylococcus aureus lead to high antibiotic persistence and affect therapy success. *Proc Natl Acad Sci U S A* https://doi.org/10.1073/pnas.2014920118 (2021).
157. Fittipaldi, N., Segura, M., Grenier, D. & Gottschalk, M. Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and zoonotic agent Streptococcus suis. *Future Microbiology* vol. 7 259–279 Preprint at https://doi.org/10.2217/fmb.11.149 (2012).
158. Haas, B. & Grenier, D. Understanding the virulence of Streptococcus suis: A veterinary, medical, and economic challenge. *Med Mal Infect* 48, 159–166 (2018).
159. Wang, C.-Z. *et al.* Antibiotic Resistance Patterns and Molecular Characterization of Streptococcus suis Isolates from Swine and Humans in China. *Microbiol Spectr* 11, (2023).
160. Yongkiettrakul, S. *et al.* Antimicrobial susceptibility of Streptococcus suis isolated from diseased pigs, asymptomatic pigs, and human patients in Thailand. *BMC Vet Res* 15, (2019).
161. Brizuela, J. *et al.* Streptococcus suis outbreak caused by an emerging zoonotic strain with acquired multi-drug resistance in Thailand. *Microb Genom* 9, (2023).
162. Segura, M. Streptococcus suis vaccines: Candidate antigens and progress. *Expert Rev Vaccines* 14, 1587–1608 (2015).
163. Foster, S. J. & Popham, D. L. *Structure and Synthesis of Cell Wall, Spore Cortex, Teichoic Acids, S-Layers, and Capsules*.
164. Bucki, R., Leszczyńska, K., Namiot, A. & Sokołowski, W. Cathelicidin LL-37: A multitask antimicrobial peptide. *Arch Immunol Ther Exp (Warsz)* 58, 15–25 (2010).
165. Jayaraman, A., Walachowski, S. & Bosmann, M. The complement system: A key player in the host response to infections. *Eur J Immunol* 54, (2024).
166. Jordan, S., Hutchings, M. I. & Mascher, T. Cell envelope stress response in Gram-positive bacteria. *FEMS Microbiology Reviews* vol. 32 107–146 Preprint at https://doi.org/10.1111/j.1574-6976.2007.00091.x (2008).
167. Juanpere-Borras, M. *et al.* Genome-Wide Identification of Conditionally Essential Genes Supporting *Streptococcus suis* Growth in Serum and Cerebrospinal Fluid. Preprint at https://doi.org/10.1101/2025.05.03.652005 (2025).
168. Xie, Z., Okinaga, T., Niu, G., Qi, F. & Merritt, J. Identification of a novel bacteriocin regulatory system in Streptococcus mutans. *Mol Microbiol* 78, 1431–1447 (2010).
169. Zou, Z. *et al.* LytTR Regulatory Systems: A potential new class of prokaryotic sensory system. *PLoS Genet* 14, (2018).

170. Zhang, Y. & Skolnick, J. TM-align: A protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33, 2302–2309 (2005).
171. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol* 20, (2019).
172. Gilchrist, C. L. M. & Chooi, Y. H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* 37, 2473–2475 (2021).
173. Aramaki, T. *et al.* KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36, 2251–2252 (2020).
174. Schrecke, K. *The LiaFSR Three-Component System of Bacillus Subtilis: Mechanism of Stimulus Perception and Signal Transduction*. (2013).
175. Alalwani, S. M. *et al.* The antimicrobial peptide LL-37 modulates the inflammatory and host defense response of human neutrophils. *Eur J Immunol* 40, 1118–1126 (2010).
176. Suntharalingam, P., Senadheera, M. D., Mair, R. W., Levesque, C. M. & Cvitkovitch, D. G. The LiaFSR system regulates the cell envelope stress response in streptococcus mutans. *J Bacteriol* 191, 2973–2984 (2009).
177. Jani, S. *et al.* Low phosphatase activity of LiaS and strong LiaR-DNA affinity explain the unusual LiaS to LiaR in vivo stoichiometry. *BMC Microbiol* 20, (2020).
178. Panesso, D. *et al.* Deletion of liaR reverses daptomycin resistance in Enterococcus faecium independent of the genetic background. *Antimicrob Agents Chemother* 59, 7327–7334 (2015).
179. Reyes, J. *et al.* A liaR deletion restores susceptibility to daptomycin and antimicrobial peptides in multidrug-resistant enterococcus faecalis. *Journal of Infectious Diseases* 211, 1317–1325 (2015).
180. Vega, L. A. *et al.* LiaR-dependent gene expression contributes to antimicrobial responses in group A *Streptococcus*. *Antimicrob Agents Chemother* 68, (2024).
181. Sewgoolam, B. *et al.* Genome-wide antibiotic-CRISPRi profiling identifies LiaR activation as a strategy to resensitize fluoroquinolone-resistant *Streptococcus pneumoniae*. Preprint at https://doi.org/10.1101/2024.10.30.621020 (2024).
182. Kesel, S., Mader, A., Höfler, C., Mascher, T. & Leisner, M. Immediate and Heterogeneous Response of the LiaFSR Two-Component System of Bacillus subtilis to the Peptide Antibiotic Bacitracin. *PLoS One* 8, (2013).
183. Jung, C. J. *et al.* PspC domain-containing protein (PCP) determines Streptococcus mutans biofilm formation through bacterial extracellular DNA release and platelet adhesion in experimental endocarditis. *PLoS Pathog* 17, (2021).
184. Okura, M. *et al.* Capsular polysaccharide switching in Streptococcus suis modulates host cell interactions and virulence. *Sci Rep* 11, (2021).
185. Houde, M., Gottschalk, M., Gagnon, F., Van Calsteren, M. R. & Segura, M. Streptococcus suis capsular polysaccharide inhibits phagocytosis through destabilization of lipid microdomains and prevents lactosylceramide-dependent recognition. *Infect Immun* 80, 506–517 (2012).
186. Liedel, C., Rieckmann, K. & Baums, C. G. A critical review on experimental Streptococcus suis infection in pigs with a focus on clinical monitoring and refinement strategies. *BMC Vet Res* 19, (2023).
187. Kerdsin, A. *et al.* Streptococcus suis serotyping by a new multiplex PCR. *J Med Microbiol* 63, 824–830 (2014).
188. Huong, V. T. L. *et al.* Epidemiology, clinical manifestations, and outcomes of streptococcus suis infection in humans. *Emerg Infect Dis* 20, 1105–1114 (2014).
189. Estrada, A. A., Gottschalk, M., Gebhart, C. J. & Marthaler, D. G. Comparative analysis of Streptococcus suis genomes identifies novel candidate virulence-associated genes in North American isolates. *Vet Res* 53, 23 (2022).
190. Baig, A. *et al.* Whole genome investigation of a divergent clade of the pathogen Streptococcus suis. *Front Microbiol* 6, (2015).
191. Murray, G. G. R. *et al.* Genome Reduction Is Associated with Bacterial Pathogenicity across Different Scales of Temporal and Ecological Divergence. *Mol Biol Evol* 38, 1570–1579 (2021).
192. Vasu, K., Nagamalleswari, E. & Nagaraja, V. Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. *Proc Natl Acad Sci U S A* 109, (2012).

193. Egido, J. E., Costa, A. R., Aparicio-Maldonado, C., Haas, P. J. & Brouns, S. J. J. Mechanisms and clinical importance of bacteriophage resistance. *FEMS Microbiology Reviews* vol. 46 Preprint at https://doi.org/10.1093/femsre/fuab048 (2022).
194. Gómez-Díaz, E., Jordà, M., Peinado, M. A. & Rivero, A. Epigenetics of Host-Pathogen Interactions: The Road Ahead and the Road Behind. *PLoS Pathogens* vol. 8 Preprint at https://doi.org/10.1371/journal.ppat.1003007 (2012).
195. Dekker, J. P. Within-Host Evolution of Bacterial Pathogens in Acute and Chronic Infection. *Annu. Rev. Pathol. Mech. Dis. 2024* 19, 2023 (2025).
196. Oliveira, P. H. & Fang, G. Conserved DNA Methyltransferases: A Window into Fundamental Mechanisms of Epigenetic Regulation in Bacteria. *Trends Microbiol* 29, 28–40 (2021).
197. Gao, Q. *et al.* Bacterial DNA methyltransferase: A key to the epigenetic world with lessons learned from proteobacteria. *Front Microbiol* 14, (2023).
198. Murray, N. E. Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle). *Microbiol Mol Biol Rev* 64, 412–434 (2000).
199. Cooper, L. P. *et al.* DNA target recognition domains in the Type i restriction and modification systems of Staphylococcus aureus. *Nucleic Acids Res* 45, 3395–3406 (2017).
200. Anton, B. P. & Roberts, R. J. Beyond Restriction Modification: Epigenomic Roles of DNA Methylation in Prokaryotes. *Annu Rev Microbiol* 57, 57 (2025).
201. Nye, T. M. *et al.* DNA methylation from a type I restriction modification system influences gene expression and virulence in streptococcus pyogenes. *PLoS Pathog* 15, (2019).
202. Roodsant, T. J. *et al.* The streptococcal phase-variable type i restriction modificationsystem SsuCC20p dictates the methylome of Streptococcus suis impacting the transcriptome and virulence in a zebrafish-larvae infection model. *mBio* 15, (2024).
203. Lozada-Chávez, I., Janga, S. C. & Collado-Vides, J. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res* 34, 3434–3445 (2006).
204. Liang, Z. *et al.* Glycerol metabolic repressor GlpR contributes to Streptococcus suis oxidative stress resistance and virulence. *Microbes Infect* 27, (2025).
205. Trzilova, D. & Tamayo, R. Site-Specific Recombination – How Simple DNA Inversions Produce Complex Phenotypic Heterogeneity in Bacterial Populations. *Trends in Genetics* vol. 37 59–72 Preprint at https://doi.org/10.1016/j.tig.2020.09.004 (2021).
206. Douglas M. Heithoff, Robert L. Sinsheimer, David A. Low & Michael J. Mahan. An Essential Role for DNAAdenine Methylation inBacterial Virulence. *Science (1979)* 284, 967–969 (1999).
207. Marinus, M. G. & Casadesus, J. Roles of DNA adenine methylation in host-pathogen interactions: Mismatch repair, transcriptional regulation, and more. in *FEMS Microbiology Reviews* vol. 33 488–503 (2009).
208. Van Der Woude, M., Braaten, B. & Low, D. *Epigenetic Phase Variation of the Pap Operon in Escherichia Coli*. (1996).
209. Atack, J. M., Guo, C., Yang, L., Zhou, Y. & Jennings, M. P. DNA sequence repeats identify numerous Type I restriction-modification systems that are potential epigenetic regulators controlling phase-variable regulons; phasevarions. *FASEB Journal* 34, 1038–1051 (2020).
210. Hu, X. P., Dourado, H., Schubert, P. & Lercher, M. J. The protein translation machinery is expressed for maximal efficiency in Escherichia coli. *Nat Commun* 11, (2020).
211. Bosdriesz, E., Molenaar, D., Teusink, B. & Bruggeman, F. J. How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization. *FEBS Journal* 282, 2029–2044 (2015).
212. Kareem, B. O. *et al.* Environmental and genetic regulation of Streptococcus pneumoniae galactose catabolic pathways. *Nat Commun* 15, (2024).
213. Gao, S. *et al.* AI-2 quorum sensing-induced galactose metabolism activation in Streptococcus suis enhances capsular polysaccharide-associated virulence. *Vet Res* 55, 80 (2024).
214. Kareem, B. O. *et al.* Environmental and genetic regulation of Streptococcus pneumoniae galactose catabolic pathways. *Nat Commun* 15, (2024).
215. Paixão, L. *et al.* Host glycan sugar-specific pathways in streptococcus pneumonia: Galactose as a key sugar in colonisation and infection. *PLoS One* 10, (2015).

216. Zhao, H. The human microbiome and genetic disease: towards the integration of metagenomic and multi-omics data. *Human Genetics* vol. 140 701–702 Preprint at https://doi.org/10.1007/s00439-021-02277-0 (2021).
217. Arıkan, M. & Muth, T. Integrated multi-omics analyses of microbial communities: a review of the current state and future directions. *Molecular Omics* vol. 19 607–623 Preprint at https://doi.org/10.1039/d3mo00089c (2023).
218. Bertolo, A., Valido, E. & Stoyanov, J. Optimized bacterial community characterization through full-length 16S rRNA gene sequencing utilizing MinION nanopore technology. *BMC Microbiol* 24, (2024).
219. Zhang, H. *et al.* Comparison of the full-length sequence and sub-regions of 16S rRNA gene for skin microbiome profiling. *mSystems* 9, (2024).
220. Abellan-Schneyder, I. *et al.* Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *mSphere* 6, (2021).
221. Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 10, (2019).
222. Hrovat, K., Dutilh, B. E., Medema, M. H. & Melkonian, C. Taxonomic resolution of different 16S rRNA variable regions varies strongly across plant-associated bacteria. *ISME Communications* 4, (2024).
223. Nygaard, A. B., Tunsjø, H. S., Meisal, R. & Charnock, C. A preliminary study on the potential of Nanopore MinION and Illumina MiSeq 16S rRNA gene sequencing to characterize building-dust microbiomes. *Sci Rep* 10, 3209 (2020).
224. Rozas, M., Brillet, F., Callewaert, C. & Paetzold, B. MinIONTM Nanopore Sequencing of Skin Microbiome 16S and 16S-23S rRNA Gene Amplicons. *Front Cell Infect Microbiol* 11, (2022).
225. Matsuo, Y. *et al.* Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinIONTM nanopore sequencing confers species-level resolution. *BMC Microbiol* 21, (2021).
226. Stevens, B., Creed, T., Reardon, C. & Manter, D. Comparison of Oxford Nanopore Technologies and Illumina MiSeq sequencing with mock communities and agricultural soil. https://doi.org/10.21203/rs.3.rs-1731798/v1 (2022) doi:10.21203/rs.3.rs-1731798/v1.
227. Heikema, A. P. *et al.* Comparison of illumina versus nanopore 16s rRNA gene sequencing of the human nasal microbiota. *Genes (Basel)* 11, 1–17 (2020).
228. Yeo, K. *et al.* A comparison between full-length 16S rRNA Oxford nanopore sequencing and Illumina V3-V4 16S rRNA sequencing in head and neck cancer tissues. *Arch Microbiol* 206, (2024).
229. Cha, T. *et al.* Gut microbiome profiling of neonates using Nanopore MinION and Illumina MiSeq sequencing. *Front Microbiol* 14, (2023).
230. Brugman, S. *et al.* T lymphocytes control microbial composition by regulating the abundance of Vibrio in the zebrafish gut. *Gut Microbes* 5, 737–747 (2014).
231. Brugman, S. *et al.* T Lymphocyte–Dependent and –Independent Regulation of Cxcl8 Expression in Zebrafish Intestines . *The Journal of Immunology* 192, 484–491 (2014).
232. López Nadal, A., Peggs, D., Wiegertjes, G. F. & Brugman, S. Exposure to Antibiotics Affects Saponin Immersion-Induced Immune Stimulation and Shift in Microbial Composition in Zebrafish Larvae. *Front Microbiol* 9, 1–16 (2018).
233. Westerfield, M. *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio Rerio).* (University of Oregon Press, Eugene, Eugene, 2007).
234. López Nadal, A. *et al.* Omics and imaging combinatorial approach reveals butyrate-induced inflammatory effects in the zebrafish gut. *Anim Microbiome* 5, (2023).
235. López Nadal, A., Peggs, D., Wiegertjes, G. F. & Brugman, S. Exposure to Antibiotics Affects Saponin Immersion-Induced Immune Stimulation and Shift in Microbial Composition in Zebrafish Larvae. *Front Microbiol* 9, (2018).
236. Matsuo, Y. *et al.* Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinIONTM nanopore sequencing confers species-level resolution. https://doi.org/10.1186/s12866-021-02094-5 (2021).
237. Schacksen, P. S., Østergaard, S. K., Eskildsen, M. H. & Nielsen, J. L. Complete pipeline for Oxford Nanopore Technology amplicon sequencing (ONT-AmpSeq): from pre-processing to creating an operational taxonomic unit table. *FEBS Open Bio* (2024).

238. ter Braak, C. J. F. , & S. P. Canoco reference manual and user's guide: software for ordination, version 5.0. *Microcomputer Power.* (2012).
239. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res* 13, 2498–2504 (2003).
240. Ciuffreda, L., Rodríguez-Pérez, H. & Flores, C. Nanopore sequencing and its application to the study of microbial communities. *Comput Struct Biotechnol J* 19, 1497–1511 (2021).
241. Alon, S. *et al.* Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Res* 21, 1506–1511 (2011).
242. Sorefan, K. *et al.* Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence* 3, (2012).
243. Oxford Nanopore Technologies. EPI2ME: An extensible product for applied bioinformatics. *https://nanoporetech.com/resource-centre/epi2me-an-extensible-product-for-applied-bioinformatics-what-is-new-and-where-are-we-going* (2024).
244. Curry, K. D. *et al.* Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nat Methods* 19, 845–853 (2022).
245. Johnson, J. S. *et al.* Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nat Commun* 10, (2019).
246. Zhang, H. *et al.* Comparison of the full-length sequence and sub-regions of 16S rRNA gene for skin microbiome profiling. *mSystems* 9, (2024).
247. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 69, 330–339 (2007).
248. Chakravorty, S., Helb, D., Burday, M., Connell, N. & Alland, D. A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 69, 330–339 (2007).
249. Oxford Nanopore Technologies. *Introduction to the Protocol Overview of the Protocol Ligation Sequencing Kit V14 Features*. https://nanoporetech.com/document/ligation-sequencing-amplicons-sqk-lsk114 (2025).
250. Dommann, J. *et al.* A novel barcoded nanopore sequencing workflow of high-quality, full-length bacterial 16S amplicons for taxonomic annotation of bacterial isolates and complex microbial communities. *mSystems* (2024).
251. Lin, X. *et al.* High accuracy meets high throughput for near full-length 16S ribosomal RNA amplicon sequencing on the Nanopore platform. *PNAS Nexus* 3, (2024).
252. Wilson, A. C., Perego, M. & Hoch, J. A. *New Transposon Delivery Plasmids for Insertional Mutagenesis in Bacillus Anthracis*. *J Microbiol Methods* vol. 71 (2007).
253. Basta, D. W. *et al.* Inducible transposon mutagenesis identifies bacterial fitness determinants during infection in mice. *Nat Microbiol* 10, 1171–1183 (2025).
254. Bhavana, V. H. *et al.* A group B Streptococcus indexed transposon mutant library to accelerate genetic research on an important perinatal pathogen . *Microbiol Spectr* 11, (2023).
255. Yousief, S. W., Abdelmalek, N. & Paglietti, B. Optimizing phage-based mutant recovery and minimizing heat effect in the construction of transposon libraries in Staphylococcus aureus. *Sci Rep* 14, (2024).
256. Van Opijnen, T. & Camilli, A. *Genome-Wide Fitness and Genetic Interactions Determined by Tn-Seq, a High-Throughput Massively Parallel Sequencing Method for Microorganisms*.
257. Shanker, E. *et al.* Pheromone Recognition and Selectivity by ComR Proteins among Streptococcus Species. *PLoS Pathog* 12, (2016).
258. Uruén, C. *et al.* Genomic and phenotypic analysis of invasive Streptococcus suis isolated in Spain reveals genetic diversification and associated virulence traits. *Vet Res* 55, 11 (2024).
259. Scherrer, S. *et al.* Population structure, genetic diversity and pathotypes of Streptococcus suis isolated during the last 13 years from diseased pigs in Switzerland. *Vet Res* 51, (2020).
260. Zhang, A. *et al.* Comparative genomic analysis of Streptococcus suis reveals significant genomic diversity among different serotypes. *BMC Genomics* 12, (2011).
261. Haustenne, L., Bastin, G., Hols, P. & Fontaine, L. Modeling of the ComRS signaling pathway reveals the limiting factors controlling competence in Streptococcus thermophilus. *Front Microbiol* 6, (2015).

262. Chao, M. C., Abel, S., Davis, B. M. & Waldor, M. K. The design and analysis of transposon insertion sequencing experiments. *Nat Rev Microbiol* 14, 119–128 (2016).
263. Willenborg, J. *et al.* Characterization of the pivotal carbon metabolism of streptococcus suis serotype 2 under ex Vivo and chemically defined in vitro conditions by isotopologue profiling. *Journal of Biological Chemistry* 290, 5840–5854 (2015).
264. Kofoed, E. M. *et al.* Discovery of GuaB inhibitors with efficacy against Acinetobacter baumannii infection. *mBio* 15, (2024).
265. Grove, A. The delicate balance of bacterial purine homeostasis. *Discover Bacteria* 2, (2025).
266. Bazurto, J. V., Dearth, S. P., Tague, E. D., Campagna, S. R. & Downs, D. M. Untargeted metabolomics confirms and extends the understanding of the impact of aminoimidazole carboxamide ribotide (AICAR) in the metabolic network of salmonella enterica. *Microbial Cell* 5, 74–87 (2018).
267. Kano, T., Ishikawa, K., Furuta, K. & Kaito, C. Knockout of adenylosuccinate synthase purA increases susceptibility to colistin in Escherichia coli. *FEMS Microbiol Lett* 371, (2024).
268. Tymoszewska, A., Szylińska, M. & Aleksandrzak-Piekarczyk, T. The LiaFSR-LiaX System Mediates Resistance of Enterococcus faecium to Peptide Antibiotics and to Aureocin A53- and Enterocin L50-Like Bacteriocins. *Microbiol Spectr* 11, (2023).
269. Zaccaria, E., Wells, J. M. & van Baarlen, P. Metabolic Context of the Competence-Induced Checkpoint for Cell Replication in Streptococcus suis. *PLoS One* 11, (2016).
270. Hallenbeck, M., Chua, M. & Collins, J. The role of the universal sugar transport system components PtsI (EI) and PtsH (HPr) in Enterococcus faecium. *FEMS Microbes* 5, (2024).
271. Meza, E., Becker, J., Bolivar, F., Gosset, G. & Wittmann, C. Consequences of phosphoenolpyruvate:-sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in Escherichia coli. *Microb Cell Fact* 11, (2012).
272. Sauer, U. & Eikmanns, B. J. The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS Microbiol Rev* 29, 765–794 (2005).
273. Arenas, J. *et al.* In vivo transcriptomes of Streptococcus suis reveal genes required for niche-specific adaptation and pathogenesis. *Virulence* 10, 334–351 (2019).
274. Fernandez-Ciruelos, B. *et al.* Repurposing Hsp90 inhibitors as antimicrobials targeting two-component systems identifies compounds leading to loss of bacterial membrane integrity. *Microbiol Spectr* 12, (2024).
275. Ouyang, L. *et al.* In vitro antibacterial and antibiofilm activities of isobavachalcone against Enterococcus faecalis clinical isolates from China. *BMC Microbiol* 25, (2025).
276. Jenkins, A. *et al.* Role of purine biosynthesis in bacillus anthracis pathogenesis and virulence. *Infect Immun* 79, 153–166 (2011).

Summary

Antimicrobial resistance is an urgent global challenge across the One Health framework. Widespread antimicrobial use in human and veterinary medicine selects for resistant bacteria and reduces the effectiveness of existing treatments. A rational approach to discovery of novel antibacterial treatments begins with the identification of tractable, critical bacterial targets, followed by the design of antimicrobial agents that specifically inhibit those targets. Achieving these challenges requires a deep understanding of how bacteria sense their surroundings, adapt their physiology, persist in host environments, and, in the case of bacterial pathogens that damage hosts, express virulence.

This thesis addresses these challenges using *Streptococcus suis* a Gram-positive encapsulated pathogen of pigs and a zoonotic agent. *S. suis* commonly inhabits the tonsils of pigs as a commensal and is also detected in the gastrointestinal and urogenital tracts. Under conditions that are still not fully understood, some lineages of *S. suis* shift from colonization to invasive disease, beginning in the upper respiratory tract, spreading into the bloodstream with systemic infection and arthritis, and crossing the blood brain barrier to cause meningitis and sudden death. The central aim of this work was to define molecular mechanisms that allow *S. suis* to adapt to distinct host environments, and to identify "Achilles heels" of disease-associated *S. suis* lineages that could benefit antimicrobial discovery.

The first experimental chapter (Chapter II), presents a genome wide screen using a transposon mutant library (Tn library) constructed *in vitro*. We established an experimental framework that replaced animal challenge models with host fluids known to support *S. suis* growth, namely activate porcine serum (APS) and cerebrospinal fluid (CSF) derived from porcine choroid plexus organoids. We used Oxford Nanopore sequencing technology to obtain saturated genetic inactivated-gene libraries.

This screen identified conditionally essential genes (CEGs) whose disruption reduced growth or proliferation of *S. suis* P1/7, a virulent zoonotic serotype 2 strain, in APS or CSF. To support growth in serum, we observed a strong dependency of *S. suis* P1/7 on genes participating in *de novo* purine biosynthesis and membrane stability functions; in CSF, genes participating in amino acid transport systems were relevant to support growth of *S. suis* P1/7. Targeted deletion of specific genes using CRISPR-Cas9 gene-specific deletion methods confirmed results obtained in our Tn-Seq experiments, namely that APS and CSF are markedly different in terms of carbohydrate and amino acid sources, positioning nucleotide metabolism and envelope homeostasis as candidate targets for antimicrobial development.

Given the importance of envelope integrity in APS, in Chapter III we dissected the contribution of regulatory systems that sense and respond to membrane stress. We discovered and characterized the *S. suis* P1/7 orthologous LiaFSR three component system, which was identified in the Tn-Seq library study reported in Chapter II, and demonstrated its central role in protecting cell envelope integrity from host-produced antibacterial molecules. We further identified two additional *S. suis* LytTR family regulatory systems, one of which was also identified by the APS Tn-library screen,and revealed a cross regulatory network between these two LytTR regulatory systems and LiaFSR regulatory system. Functional analyses showed that this cross-regulatory network provides strict coordinated control over membrane stability, competence, biofilm formation, and cell viability. These findings show that in *S. suis* P1/7, a regulatory network adjusts capsule and membrane properties to the differential conditions (in terms of carbohydrate and amino acid availability) encountered by *S. suis* in different host niches (blood serum and cerebrospinal fluid), supporting host-adaptive traits.

In Chapter IV we examined how epigenetic regulation contributes to regulation of gene expression of serotype 2 *S. suis* by focusing on two type I restriction modification (RM) systems conserved in disease-associated lineages. One system, MS1, carries a phase variable specificity subunit. The second, MS2, is found with either a full length or a truncated specificity subunit. We constructed a panel of eighteen strains that combined the possible MS1 and MS2 configurations, from which we profiled their transcriptomes and assessed their growth in different media.

This analysis showed, for the first time, that MS1 exerts an inhibitory influence on MS2. When full-length MS2 operates in absence (by gene-specific deletion) of MS1, *S. suis* P1/7 downregulate sugar uptake systems while upregulating biosynthetic pathways, and display faster growth in rich media. We propose that co-occurrence of full length MS2 with MS1 in strongly disease-associated strains of the ST1 sequence type reflects selection for tighter control of carbohydrate acquisition and central metabolism. Such control would potentially support efficient sugar scavenging during colonization and rapid proliferation during invasive infection.

Because the transposon screen of Chapter II relied on Oxford Nanopore sequencing, in Chapter V we evaluated the accuracy of current Nanopore technologies in an application that demands precise taxonomic resolution. We compared Nanopore and Illumina sequencing for 16S rRNA profiling of the zebrafish gut microbiome, including V3-V4 amplicons analyzed by both platforms and full length V1-V9 amplicons analyzed by Nanopore. The V3-V4 results were highly similar across platforms, supporting the use of Nanopore sequencing platforms for applications that require

accurate measurements of DNA base composition, including the sequencing of Tn-libraries. Full length V1-V9 profiling showed promise for enhanced resolution, while underscoring the need for continued updates of reference DNA sequence databases. To improve affordability and throughput, we implemented a PCR-barcoding strategy and developed in-house pipelines for demultiplexing and analysing Nanopore long-read sequences that had proven to be useful across multiple microbiota studies of diverse host-associated microbial taxa.

The general discussion in Chapter VI outlines practical limitations for constructing transposon libraries, which vary by organism and delivery method, and it highlights analytical constraints that are essential for correct interpretation of Tn-library screening results. In Tn-library downstream analyses, we identified a biological link between nucleotide metabolism and envelope homeostasis, emphasizing that NADH and ATP energy supply and precursor availability influence capsule and membrane integrity, which co-determine survival in APS. Moreover, by comparing the transcriptomic signatures associated with membrane regulatory HdrMR-LiaSFR systems and the type I RM systems, the chapter proposes an inverse relationship between the activity of phosphotransferase transporters and the activation of biosynthetic pathways (e.g., the incomplete TCA cycle) in *S. suis*. Our downstream analyses of gene expression data suggest that phosphoenolpyruvate (PEP) sits at a critical pivotal branch point that supports sugar import through the phospho-transferase system and feeds central carbon metabolism which is essential for disease-associated *S. suis* growth and survival. Careful control of flux through this PEP node appears to determine whether the cell prioritizes resource acquisition over anabolism and proliferation.

This work provides several entry points for antimicrobial discovery. LiaS and LiaR, together with the membrane protein HdrM, are central to envelope integrity surveillance and could be leveraged to sensitize bacteria to host defenses. Elements of the MS1 system that moderate MS2 activity may, once targeted by bioactives, offer a way to disrupt the metabolic balance required for successful infection and in-host growth and survival. Genes encoding enzymes in *de novo* purine biosynthesis, which proved crucial to sustain growth in serum, represent additional candidates to be targeted by novel antibiotics. While further validation is needed to demonstrate feasibility of pharmacological modulation, this thesis contributes to our understanding of regulatory and enzyme functions of serotype 2 *S. suis* that are required to colonise specific host niches and that exert core regulatory and, or, metabolic bottlenecks.

Resum

La resistència als antimicrobians és un repte global urgent en el marc *One Health*. L'ús extens d'antimicrobians en medicina humana i veterinària selecciona bacteris resistents i redueix l'eficàcia dels tractaments existents. Els antibiòtics tradicionals es van descobrir en gran part a partir de microorganismes com fongs i Actinomicets, i la seva activitat es va anar perfeccionant durant llargs períodes evolutius, el qual es incompatible amb les necessitats clíniques actuals. Un enfocament racional comença amb la identificació de dianes moleculars adequades, seguit del disseny d'agents antimicrobians que les inhibeixin específicament. Això exigeix una comprensió profunda de com els bacteris perceben l'entorn, adapten la seva fisiologia, persisteixen en ambients de l'hoste i expressen virulència. Aquest repte és complex perquè els bacteris presenten redundància funcional i, en algunes espècies, mostren una elevada diversitat genètica entre soques.

Aquesta tesi aborda aquests reptes utilitzant *Streptococcus suis,* un patogen encapsulat Gram positiu que afecta principalment els porcs, però que també actua com a agent zoonòtic. *S. suis* habita una espècie bacteriana commensal, en la microbiota de les amígdales dels porcs i també es detecta al tracte gastrointestinal i urogenital. En condicions que encara no es tenen clares, *S. suis* pot passar de la colonització a la infecció: comença al tracte respiratori superior, es dissemina a la sang amb infecció sistèmica i artritis, i travessa la barrera hematoencefàlica per causar meningitis i, de vegades, mort sobtada. L'objectiu central d'aquest treball va ser definir els mecanismes moleculars que permeten a *S. suis* adaptar-se a diferents ambients de l'hoste, i identificar vulnerabilitats que puguin beneficiar el descobriment d'antimicrobians.

El primer capítol experimental (Capítol II) presenta un cribratge a escala genòmica utilitzant una llibreria mutant de transposons construïda *in vitro*. Vam establir un marc experimental que substitueix els models d'infecció en animals per fluids rellevants de l'hoste, concretament sèrum porcí activat (SPA) i líquid cefaloraquidi (LCR) derivat d'organoides del plexe coroide, per imitar nínxols clau de l'hoste i reduir l'ús d'animals. També vam integrar la tecnologia de seqüenciació d'*Oxford Nanopore* per a la llibreria de transposons. La llibreria va assolir una cobertura elevada a les regions codificants, proporcionant una saturació suficient per a anàlisis posteriors.

Aquest cribratge va identificar gens essencials condicionals l'alteració dels quals reduïa proliferació en SPA o LCR. En sèrum, vam observar una forta dependència de la biosíntesi de purines i de funcions d'estabilitat de membrana; al LCR, predominaven els sistemes de transport d'aminoàcids. L'eliminació dirigida de gens seleccionats

mitjançant CRISPR-Cas9 va confirmar diverses prediccions i aquests resultats indiquen que els requeriments del sèrum i del líquid cefaloraquidi difereixen clarament, situant el metabolisme dels nucleòtids i la homeòstasi de la capsula com a dianes candidates per al desenvolupament d'antimicrobians.

Atesa la importància de la integritat de l'embolcall en SPA, en el Capítol III vam estudiar la contribució de sistemes reguladors que detecten i responen a l'estrès de membrana. Vam caracteritzar el sistema de tres components LiaFSR, identificat a l'estudi de la llibreria de transposons (Capítol II), i vam demostrar el seu paper central a l'hora de protegir la cèl·lula dels antimicrobians de l'hoste. A més, vam identificar dos nous sistemes reguladors de la família LytTR, un dels quals també va ser detectat al cribratge amb la llibreria de transposons en sèrum, i vam descobrir una xarxa de regulació creuada entre aquests sistemes LytTR i LiaFSR. Els anàlisis funcionals van mostrar que aquesta xarxa proporciona un control coordinat de l'estabilitat de la membrana, la competència, la formació de biofilms i la viabilitat cel·lular. Aquestes troballes mostren que una xarxa reguladora ajusta les propietats de la càpsula i la membrana a l'entorn de l'hoste, donant suport a trets d'adaptació crucials.

Al Capítol IV vam examinar com la regulació epigenètica contribueix a l'adaptació centrant-nos en dos sistemes de restricció i metilació de tipus I conservats en llinatges associats a patogènesi. Un d'ells, MS1, conté una subunitat d'especificitat amb variació de fase. El segon, MS2, es troba amb una subunitat d'especificitat de forma completa o truncada. Vam construir un panell de divuit soques que combinava les possibles configuracions d'MS1 i MS2 i en vam caracteritzar els fenotips i els transcriptomes.

Aquesta anàlisi va mostrar, per primera vegada, que MS1 exerceix una influència inhibidora sobre MS2. Quan MS2 opera sense aquest fre, especialment en la seva forma completa, les bacteris redueixen l'expressió de diversos sistemes de captació de sucres alhora que incrementen les vies biosintètiques, i mostren un ploriferacio mes rapida en medis rics. Proposem que la coocurrència freqüent del MS2 de forma completa amb MS1 en soques altament associades a patogènesi (ST1) reflecteix una selecció per a un control més estricte de la captació de carbohidrats i del metabolisme central. Aquest control podria afavorir una captació eficient de sucres durant la colonització i una proliferació ràpida durant la patogènesi.

Com que el cribratge de transposons del Capítol II es va basar en seqüenciació *d'Oxford Nanopore*, al Capítol V vam avaluar la precisió de les tecnologies *Nanopore* actuals en una aplicació que exigeix una resolució taxonòmica precisa. Vam comparar *Nanopore* i *Illumina* per a la seqüenciació de l'ARNr 16S del microbioma

intestinal del peix zebra, incloent amplicons V3-V4 analitzats per ambdues plataformes i amplicons V1-V9 de analitzats amb *Nanopore*. Els resultats V3-V4 van ser altament similars entre plataformes, cosa que sosté l'ús de *Nanopore* per a aplicacions que requereixen una quantificació precisa, inclosa la seqüenciació de llibreries de transposons. La seqüenciació V1-V9 va mostrar potencial per a una resolució millorada, tot subratllant la necessitat d'actualitzacions en les bases de dades de referència. Per millorar l'assequibilitat i el rendiment, vam implementar una estratègia senzilla de *barcoding* per PCR i vam desenvolupar *pipelines* per l'anàlisi i *demultiplexing* de dades que ja han demostrat utilitat en diversos estudis de microbiota.

La discussió general del Capítol VI descriu limitacions pràctiques per a la construcció de llibreries de transposons, que varien segons l'organisme i el mètode utilitzat, i ressalta consideracions analítiques per a una interpretació correcta dels resultats. La discussió també aborda la connexió entre el metabolisme dels nucleòtids i la homeòstasi de la càpsula polisacàrida, destacant que la disponibilitat d'energia i de precursors influeixen en la integritat de la càpsula i de la membrana, cosa que al seu torn condiciona la supervivència en SPA. A més, comparant els perfils transcriptòmics associats als sistemes reguladors de membrana i als sistemes de metilació, el capítol proposa una relació inversa entre l'activitat dels transportadors del sistema de fosfotransferència i l'activació de vies biosintètiques (per exemple, el cicle de l'àcid cítric incomplet) en *S. suis*. El fosfoenolpiruvat se situa en un punt de ramificació crític perquè dona suport a la importació de sucres a través del sistema de fosfotransferència i també alimenta el metabolisme central. Un control acurat del flux en aquest node sembla determinar si la cèl·lula prioritza la captació de recursos o l'anabolisme i la proliferació ràpids.

Aquest treball proporciona diversos punts d'entrada per al descobriment d'antimicrobians. LiaS i LiaR, juntament amb la proteïna de membrana HdrM, són centrals per a la regulació de la càpsula polisacàrida i es podrien utilitzar per sensibilitzar les bacteris a les defenses de l'hoste. Elements del sistema MS1 que moderen l'activitat de MS2 podrien oferir una via per alterar l'equilibri metabòlic necessari per a una infecció. Enzims de la biosíntesi de purines, que van resultar crucials en sèrum, representen candidats addicionals. Tot i que cal una validació addicional per demostrar la viabilitat de la modulació farmacològica i per definir els marges terapèutics, la tesi acota la cerca cap a funcions repetidament necessàries en condicions de l'hoste i en punts clau de la regulació i del metabolisme.

Authorship Statement

Chapter I General introduction.

The ideas and content of this chapter, as well as the writing, were prepared by me. My supervisors provided advice on the structure and reviewed and commented on the phrasing.

Chapter II.

I contributed to the conceptualization, formal analysis, investigation, methodology (except where explicitly attributed to others), supervision, validation, visualization, writing - original draft, and writing - review and editing. Tiantong Zhao developed and extracted CSF from organoids. Jos Boekhorst designed the Tn-Seq pipeline. Rajrita Sanyal constructed the SSU_RS04755 and *purA* mutants and performed the associated experiments, Nissa Arifa constructed the *hdrM* mutant and performed its associated experiments, Troy Wagenaar constructed the *liaS* mutant and performed its associated experiments, and Varsha Seshadri constructed the *liaR* mutant. All students carried out their work under my supervision. Blanca Fernández-Ciruelos contributed to conceptualization and provided technical support and advice. My supervisors contributed to the conceptualization, in particular regarding the use of transposon libraries for screening and the use of CSF, and also reviewed and commented on the manuscript.

Chapter III.

I contributed to the conceptualization, formal analysis, validation, methodology (except where explicitly attributed to others), visualization, writing - original draft, and writing - review and editing. Troy Wagenaar performed the gene expression analysis and biofilm formation assays, and Amy Raphaela performed the growth curves, competence assays, biofilm assays, and viability assays. Both students worked under my supervision. My supervisors reviewed and commented on the manuscript.

Chapter IV.

I contributed to the conceptualization, formal analysis, validation, methodology (except where explicitly attributed to others), visualization, writing - original draft, and writing -review and editing. Gemma Murray performed the genomic context analysis, prepared the corresponding figure, and wrote the section describing this analysis. Alex Gussak and Oshin Vellalara constructed the 18 deletion-mutants, and Alex wrote the corresponding methodological section. My supervisors defined the general research question and reviewed and commented on the manuscript.

Chapter V.

The idea of comparing Illumina and Nanopore sequencing was developed by me, and the idea of applying this comparison to the zebrafish gut microbiota was proposed by Adrià López-Nadal. Adrià generated the zebrafish mutants and performed the analysis and biological interpretation of the differences between wild-type and mutant microbiota. I performed the 16S rRNA PCR amplification of all amplicons, the sequencing process, the design and optimization of the PCR-barcoding strategy, the demultiplexing part of the pipeline, the cost analysis, and the taxonomic comparison between Illumina and Nanopore. Jos Boekhorst performed the remaining bioinformatic analyses. I wrote the original draft and performed the review and editing. All authors reviewed and commented on the manuscript.

Chapter IV General discussion.

The ideas and content of this chapter, as well as the writing, were prepared by me. My supervisors provided advice on the structure and reviewed and commented on the phrasing.

In all chapters of this thesis, I have used generative AI (ChatGPT, OpenAI) to provide language support, enhance the clarity of the text, and assist in coding. I am fully responsible for the final content, analyses, and interpretation of the research. References: OpenAI. (2023). ChatGPT (April 20 version) [Large language model]. https://chat.openai.com/. OpenAI. (2025). ChatGPT (GPT-5) [Large language model]. https://chat.openai.com/.

Acknowledgements

With the conviction that no words will ever be enough to express my gratitude to everyone I mention, I still hope you may catch a glimpse of what I truly feel. Those who know me well will recognize that these words come from the heart, not from writing what is usually expected in an acknowledgements section. Politically correct has never really been my thing, so what follows is an honest reflection on how each of you contributed to my personal and scientific development. Sorry in advance for the length of these acknowledgments. Fortunately, I have many people to be grateful to. People are thanked in the chronological order in which they appeared in my life, as this is how my path has been shaped.

Primer de tot, començaré per la meva família: petita, però ja se sap que al pot petit hi ha la bona confitura. Començo pels que ja no hi són: àvia **Àngels**, àvia **Paquita**, avi **Joan** i tiet **Josep Maria**. Gràcies per treballar tant, per no queixar-vos mai de res, pel gran exemple que sempre m'heu donat, no amb paraules sinó amb accions. Gràcies per la humilitat que m'heu transmès i per la gran fortalesa que sempre vau mostrar. Sé que ja no hi sou, però heu deixat un llegat d'exemple i d'aprenentatge que no marxarà mai. Gràcies per fer-me forta.

Gràcies, avi, per les estones regant l'hort, movent l'aigua de les patateres a les tomaqueres, per intentar amagar-te quan anaves a fumar, i per la força que la teva mirada sempre transmetia. Gràcies, àvia, per tot el que em vas cuidar de petita, per despertar-me cada dia i vestir-me amb paciencia. Per ensenyar-me a cuinar, per la paciència fent galetes, coques, rosquilles i orelletes, pels berenars de pa amb oli i sucre. Per ajudar-me amb els quaderns d'estiu i cuidar-me quan em posava malalta. Avi i àvia, el vostre «salero» s'ha fet gran, però no us oblida.

Gràcies, **pare** i **mare**, per treballar tant i poder-me donar una bona vida i, sobretot, una bona educació. Gràcies, mare, per la teva incansable insistència perquè fes els deures, estudiés amb temps i per les incomptables hores preguntant-me el temari dels exàmens. Per ser exigent amb les meves notes, per demanar-me esforç i dedicació. Ets un exemple de dona forta que ha tingut dues feines tota la vida: una jornada laboral de vuit hores i una jornada de vint-i-quatre cuidant la casa i les dues filles. És increïble com de bé ho has fet i com n'era de difícil. Sempre el nostre exemple més gran.

Gràcies, pare, per les incomptables hores jugant al tren de *lego* i llegint-me contes; gràcies per no haver-me llegit mai cap conte de princesa salvada per un príncep; gràcies per contes com "Sóc molt Maria", "El lleó i el ratolí" o per pel·lícules com

“Simbad” o “Spirit”. Però, sobretot, per ensenyar-me què és la passió per la teva feina, què és gaudir del que fas: anar a treballar feliç i tornar cansat però satisfet, que la teva feina et faci feliç. La millor herència que ens podíeu deixar és aquesta. Gràcies, pare i mare, per assegurar-vos que no ens faltés mai de res i per l’amor que ens heu donat. Mai us ho agrairé prou.

Gràcies, **germaneta**, per ser una font de caos meravellosa; per ensenyar-me a insistir amb força quan vols alguna cosa; per mostrar-me la singularitat de cadascú i, sobretot, per aquesta força que tens dins que mou muntanyes. Sempre al teu costat.

I de la família que et toca, a la família que esculls. He tingut una sort inexplicable de conèixer-vos.

Rosaura, gràcies per fer-me de germana gran, per cuidar sempre de mi i de la despistada que sóc; per tots els moments de riure, que han estat molts, des de la llar d’infants fins ara, i espero que fins al dia que siguem velletes assegudes a una cadira davant de casa, prenent el sol i comentant com pugen les noves generacions, jajaja. Gràcies per incentivar-me sempre a fer el que em faci feliç sense por. Gràcies per celebrar les coses bones amb mi i per fer-me una estirada d’orelles quan toca! Una mica més tard vaig tenir la sort de conèixer-te, **Aleix**, sempre una mica als núvols, cosa que ja ens va bé per equilibrar la balança. Gràcies per les llargues nits estudiant matemàtiques, física i història junts. Gràcies per fer-nos riure i ser l’alegria del grup. Gràcies per voler sempre el millor per a mi; per dir-me, aquell dia a Barcelona, prenent un cafè: «Sí, dona, agafa la feina, que d’aquí a quatre anys aquí tot seguirà igual»; i per recordar-me que això sempre seria casa meva. Gràcies, **Anna**, per escoltar-me sempre i per l’exemple de decisió i independència, per passar del que diguin els altres i fer allò que creus correcte.

Gràcies, Rosaura, Anna i **Núria**: exemple de dones fortes i decidides, que no aguanten tonteries, que no segueixen les normes socials si no els escauen i que saben què es mereixen. Sou inspiració i exemple i, juntament amb l’Aleix, vosaltres sou pau, sou casa i sou refugi. No sé què faria sense vosaltres. Agraïda és poc.

Gràcies, **Daphne**, pel suport que sempre m’has donat; gràcies per veure sempre el millor de cadascú i per ser la primera a oferir ajuda. Sovint has estat qui m’ha dit que sí, que ho podia aconseguir; una de les persones que més m’ha fet creure en mi i en les meves capacitats. Ets un exemple de persona lliure i ambiciosa que sempre he admirat. Sempre has estat un gran suport per a mi, i estic segura que continuarem celebrant molts més èxits juntes. I, per descomptat, gràcies, **Eva**, per ser la tercera

en discòrdia, per aportar aquesta part de bogeria i alegria que tant ens fa falta, pel teu somriure contagiós i pel teu bon cor. Seguirem acumulant liades memorables i aventures èpiques, totes tres.

També voldria agrair tot allò viscut, rigut i après junts al **David**, a la **Gemma**, a l'**Alba**, a la **Paula**, al **Sula**, a la **Mercè**, a la **Duna**, al **Pau** I a la recent incorporada al grup, però no per això menys important, **Àngels**. Gràcies per ser *els més uapos del BC*, per totes les aventures i per fer junts una família una mica disfuncional i sorollosa, però molt maca.

També considero que una gran part de l'educació rebuda i de qui som l'hem d'agrair als mestres. Especialment, recordo amb afecte i admiració la professora **Imma** Oliveres Prats, gràcies per ensenyar-nos a pensar de manera crítica, per preocupar-te per nosaltres més enllà del temari, per donar-nos il·lusió i exemple de força. Has estat un gran referent que recordo amb carinyo i una part important del meu creixement com a persona.

A una de les millors coses que m'emporto del meu pas per la universitat, la **Laura**. Gràcies per totes les nits en vela estudiant, per tots els riures i els plors. Gràcies per haver demostrat ser una gran amiga en els bons moments, però sobretot en els dolents; gràcies pel teu cor enorme i pel suport incondicional. M'encanta veure la gran persona i científica en què t'estàs convertint; per sort i casualitat hem agafat camins molt semblants, i no puc ser més afortunada d'haver viscut tot el camí del doctorat amb tu. Espero poder ser a prop teu per celebrar els nostres futurs èxits, que de segur seran molts. I, per descomptat, gràcies també al boig que venia en el pack. Gràcies, **Èric**, per aguantar-nos sense jutjar, per donar suport sense fer preguntes, per tenir sempre una broma a punt i un somriure per regalar. Moltes gràcies als dos per tot.

També volia agrair **al Franco**, que ha estat present tant en el meu pas per la universitat com en la bogeria de mudar-nos a Holanda per seguir amb la nostra carrera. Ha estat una sort per a mi que hàgim coincidit, gràcies a tu m'he sentit menys sola en aquest país, i espero poder continuar compartint amb tu uns quants anys més per aquí.

Finalment, m'agradaria agrair a una professora que, durant la universitat, em va fer descobrir la biologia molecular. Gràcies, **Susana**, per una manera d'ensenyar diferent i més pràctica; per no fer-nos memoritzar conceptes, sinó entendre'ls; per prendre't el temps d'avaluar els nostres coneixements i no només la nostra capacitat de memorització. La teva manera d'ensenyar, tant com la matèria que imparties, va tenir

un impacte decisiu en el meu interès i en el rumb de la meva carrera, i t'ho agraeixo molt.

I also want to thank the first person who told me I could do a PhD, **Jeroen**. Thank you for taking me on as your student, for teaching me so much about cloning, and above all, for your patience when I first arrived in the Netherlands and my English was shaky. You were a great teacher whom I remember with affection, and the first to believe I could become a good researcher. I'll always be grateful.

A la meva terapeuta, **Cristina**. Gràcies per acompanyar-me aquests quatre anys amb tanta professionalitat i cura. Amb tu he après a reconèixer la síndrome de la impostora i a mirar-me amb més amabilitat. Sincerament, dubto que hagués arribat fins aquí sense el teu suport; avui sóc una investigadora i una persona millor gràcies a tu.

Claudia, tú fuiste mi primera amiga en este país. Gracias a ti, empezar una vida aquí fue un poco menos complicado. Gracias por hacerme reír hasta que me duele la tripa. Gracias por tu autenticidad y tu honestidad desarmante. Me enseñaste que el arte y la ciencia son dos caras de la misma moneda y que mirarlas juntas las hace todavía más bellas. Te admiro y soy muy feliz de que coincidiéramos en este rincón del mundo.

Gracias también a mi querida amiga **Cristina**, a este remolino de fuerza y energía que te hace ser tan maravillosa. No escuches a quien no ve lo bonito que es estar tan llena de vida. Gracias por tu apoyo incondicional todo este tiempo, por recordarme que no hay que atender a esa vocecilla que nos hace dudar. Gracias por tu determinación, tu fuerza, tu entereza y tu sonrisa. Espero seguir teniendo largas charlas contigo durante mucho tiempo. Sigue siendo así de auténtica. Nunca dejes de ser tan auténtica.

From the people who kept me grounded to those with whom I shared benches, ideas, and deadlines. I would first like to thank my supervisors, **Jerry** and **Peter**. Thank you for giving me this opportunity and for trusting me from the very beginning, even when you hardly knew me. Jerry, thank you for listening, for the trust you gave me, and for the freedom to follow my curiosity, even though my stubbornness sometimes left you with little choice. Thank you for the patience there. You treated my mistakes as part of learning and you did so with kindness. Your support in key moments made a real difference, and your encouragement to follow what makes me happy mattered just as much, during the PhD and after. Thank you for continuing to support me, teach me, and encourage me beyond the PhD. Peter, thank you for your

support as my co-supervisor. Your feedback and your help connecting me with the right people made a difference, and so did the way you brought order when it mattered. Special thanks for the guidance on supervising students. You helped me understand that commitment looks different from student to student and that each deserves the same attention and care. Many thanks to both of you.

Dear **Michiel**, where do I start. Thank you for being the person I most feared presenting to at lab meetings, in the best possible way. Thank you for your sharp eye and your disarming honesty. Thank you for your patience with my constant interruptions to your office for questions. Thank you for the mentoring you offered me even though I was not formally your responsibility. Most of all, thank you for seeing me, for listening, for understanding what moves me, and for sharing your passion and curiosity for molecular biology and for fundamental science. And last but not least, thank you for accepting this loud Spanish girl exactly as she is, loud and a little crazy. I hope I can keep counting on your advice in the future, vale?

And of course, a special and very big thank you to **Jos**. Thank you for the incredible support you have been throughout my PhD. You taught me bioinformatics with astonishing patience, and you were often one of my toughest critics, which undoubtedly made me a better scientist. Just as important, you helped me believe I had a place here. You stood by me through the doubts and helped me feel more sure of myself and my research. In our group you are a mentor in the fullest sense, not only on the technical side but also in how you look after people. You check that the PhDs we are doing okay, that we do not overwork, and that we take holidays when we need them. I am still not sure which of us is more stubborn or speaks faster, but I am sure this PhD would have been much harder without your kindness and support. Thank you for looking after all of us. I hope I can keep learning from you.

To the PhD students who were there when I had just started, my first teachers and people I admire. Thank you, **Simen**, for introducing me to bioinformatics and for that endearing mix of radical skepticism with a touch of drama. More often than not, you were right, and it was a very educational wake-up to how science really works. Very big thanks to **Alex**, one of the kindest, nicest, and smartest people I have had the honour to meet. It is always a fascinating mystery how your mind works, and how two minutes of troubleshooting with you are always incredibly helpful. Your heart is generous and your mind is brilliant. Thank you for everything, Alexito. And of course, thanks to Alex I met the sweetest **Rick**. Thank you for your smile and for the very precious cloning tricks! **Isabela**, thank you for being such a sweet and kind presence. You are a role model of strength and determination, someone who knows her worth and goes for what she deserves. You inspire me, and I am sure you will keep shining

wherever you go. To **Tiantong**, working alongside you in the lab was an education in humility, precision and thoughtful work. You taught me that the quietest voices are often the ones most worth listening to, and I am sure yours has a great deal of important things to say. I look forward to hearing them. **Hedi**, thank you for the time you spent at HMI. You made our days more fun and pleasant. Your enchanting smile and your sharp ironic humour were always a joy. You are a model of a strong, independent woman who does not give up. Seeing someone like you grow and keep surpassing herself every day made me feel stronger and more capable. Thank you for the example you set.

From those who paved the way to those who walk beside me now. To the current HMI PhD students, my peers and colleagues. **Zizhuang**, thank you for blending in with us so naturally and for opening your heart, your culture and the secret menu of Eastern Express to us. Thank you for the hugs when I needed them, for sending me home while you stayed late and for the weekend jump scares when I thought I was alone. Thank you for being there and for caring. I could not ask for more. To my dear **Oshin**, where do I start? Professionally, you taught me that you can do anything. Molecular biology? Sure. Microbiology? Sure. Organoids? Sure. There is Oshin, unstoppable, always respectful, always kind, always smiling, and always standing up for herself tirelessly. Personally, you are an example of a woman who does not follow the rules that this society has so cruelly and unfairly imposed on her. You think critically and intelligently, you made your own path, and you followed your own rules. You are also a thoughtful and kind friend whom I admire and respect, and I will have the honour to keep learning from very soon. For all this, Oshin, an enormous thank you. **Mohak**, my crazy, adventurous drama queen. You are the person who drives me the most crazy and the one I have the most fun with. You are endlessly organised, and your mind is a beautiful whirlwind of feelings and creativity that makes you such a special person. I hope you always find the safety you need to be yourself with us. I learned more from you than you think, including a good dose of patience. Querido **Ibai**, siempre me cuesta leer qué pasa por esa cabecita, pero ese misterio también es parte de tu encanto. Gracias por ser tú, por regalarme almendras y fruta cuando se me agotan las energías, por abrazarme cuando estoy triste, por escucharme cuando estoy frustrada y por darme a Herme para pegar cuando estoy enfadada. Tus reflexiones en nuestros paseos hicieron más ligera esta fase final del doctorado, y ojalá muy pronto sea yo quien esté ahí para ti. Ver tanta ambición es inspirador, y me muero de ganas de ver lo lejos que llegas. Soy muy afortunada de haberte conocido y espero que tengas que aguantar a esta pesada muchísimo tiempo más. Maite eta estimu handiz.

To the rest of the HMI team and former members, **Aline, Sylvia, Adam, Bart, Nienke, Sabrina, Evelien, Maaike, Marten and Stijn** thank you for the everyday support at the bench, the ideas, the good questions and the many small favours that make everything possible. A very special thank you to **Joyce**, the quiet wheel that keeps everything turning, and also very special thanks to **Linda** and **Anda** for their steady encouragement and their warmth and kindness.

To my wonderful students **Nissa, Ariadna, Rajrita, Troy, Amy** and **Hamzah**. I was very lucky to have you. Thank you for genuinely caring and for sharing your passion and eagerness to learn with me. Truly, much of this thesis would not have been possible without you. Working with you taught me as much as I hope I taught you. I am sure you all have a bright future ahead of you. Thank you for the support and kindness that you brought to every step.

To the ESRs of our ITN network, **Sandra, Jennifer, Dina, Nader, Hettie, Pareena, Lawrence, Linda and Ayelen,** thank you for your support, your insights and the sense of community you built. The yearly gatherings, the conversations and the small boosts of encouragement meant more than you know. You are an exceptionally smart and kind group, and I am confident your future in science is bright.

A mi cántabro favorito, mi amigo con pajaritos en la cabeza. Gracias, **David**, por acompañarme, por poner orden en mi pequeño caos, y por nuestras largas conversaciones en las que me escuchas y me devuelves la calma, aunque a veces se te vaya la mirada detrás de un pájaro. Ojalá sea por mucho tiempo más. Y, sobre todo, gracias por ser parte de esta pequeña familia que hacemos entre todos aquí en Wageningen. To my favorite Irish guy, **Donovan**. Thank you for your energy, your sensitivity and the way you stand up for what matters. You fight for causes that count, and your conviction inspires me every time. Querido **Alberto**, qué gusto tenerte cerca. Tu buena energía lo ilumina todo. Ojalá te quedes mucho tiempo más y sigas siendo el alma de la fiesta.

Adri, quina sort que els nostres camins s'hagin creuat en aquest racó de món. M'has inspirat moltíssim i m'has ensenyat una manera de viure completament fora de les convencionalitats, però sobretot, gràcies per ensenyar-me què vol dir ser autèntic sense por i lliure sense remordiments. Que no se t'apagui mai aquesta llum tan lliure.

To my dear paranimphs. **Alvja**, thank you for taking care of me as a sister would. You arrived in my life like a whirlwind and somehow made me feel at home. Thank you for filling the house with the smell of coffee in the morning and for making sure I eat enough broccoli at night. Little by little we became family here. We look after each

other, and you gave me all this before I even knew I needed it. You are one of the biggest thank yous of this thesis. I admire you deeply as a professional, I look up to you, and I cannot wait to celebrate all the successes ahead together. Ti voglio bene, amica mia. **Blanqui**, mi otro enorme gracias de esta tesis. Me siento muy afortunada de haberte tenido tan cerca y de que me acogieras con tanta generosidad. Tu mente me fascina, imaginación, inteligencia y picardía a partes iguales. Gracias por cuidarme y por mantener mis pies en la tierra. Gracias por ser como una hermana mayor. Sabes el respeto que te tengo. Tu imaginación no tiene límites y me muero de ganas de ver por qué caminos te llevará esa cabecita maravillosa. Ojalá pueda seguir a tu lado mucho tiempo más.

A **Santi**, mi montaña de paz. Gracias por escuchar, con paciencia incansable, cada día, todas mis quejas, frustraciones, enfados y, puntualmente, alguna que otra alegría que me da el lab. Gracias por recordarme lo mundana que es la ciencia cuando, a veces, me deslumbra lo que intenta aparentar ser. Pero, sobre todo, gracias por amar a todas las mujeres que hay en mí, desde la más frágil hasta la más fuerte, desde la más equivocada hasta la más sabia. Gracias a tu maravillosa, numerosa y cálida familia por abrirme las puertas de su casa. Y gracias a ti por elegirme para ser la tuya. T'estimo.

Finally, thank you to **all the women in science and beyond** who paved the way for girls like me to pursue the career of their choice. To those I have met, to those I have read, to those silenced, and to those burned. Thank you for raising your voices. We will keep being loud for all the girls who still don't have a chance.

About the author

Maria was born in Maspujols, a small farming village near Catalunya's "Golden Coast." She attended the rural school of Rocabruna, where the logic of nature offered a welcome break from memorizing arbitrary language rules, its patterns felt predictable and finally made sense. One of the presents she cherished most in her childhood was a microscope, with which she spent hours observing tiny insects. At the age of ten, she decided to become a marine biologist, fascinated by the sea, a passion that gradually evolved into a love for microbiology. She moved to Barcelona to pursue a BSc in Microbiology at the Universitat Autònoma de Barcelona, where she discovered a different pace of life among the city's crowds and energy. During her studies, she became fascinated by genetic engineering in microorganisms, which sparked her desire to pursue research in molecular microbiology.

To follow this path and improve her nearly nonexistent English skills, she moved to Groningen for an Erasmus+ internship in the Molecular Microbiology group. There, she worked on CRISPR-based mutations in yeast to optimize maltose transport, an experience that solidified her decision to pursue a PhD. To reach that goal, she returned to Barcelona to complete an MSc in Multidisciplinary Research in Experimental Sciences. Her MSc internship, carried out in the Bacterial Infection and Antimicrobial Therapies group, focused on the role of the Type VI Secretion System (T6SS) in *Pseudomonas aeruginosa* during co-biofilm formation with *Staphylococcus aureus*. During this period, she applied for PhD positions and eventually joined Wageningen University, under the supervision of Jerry Wells, to study molecular mechanisms of adaptation in *Streptococcus suis*, a topic that perfectly aligned with her long-standing fascination for molecular microbiology.

After four years, many breakdowns, and the formation of a new chosen family, she decided to continue "complicating her life" by pursuing a career in academia. Her next step is a postdoctoral position at Wageningen University, where she will study intestinal microbiota in organoid models. She is excited to "zoom out" from the molecular level toward a more multidisciplinary perspective, continue working with her wonderful colleagues, and keep surviving, at least for now, in the challenging but rewarding path of academic science. In this thesis, the results of her PhD are presented, outcomes that, after much struggle, might almost appear effortless.

Blanca Fernández-Ciruelos

Overview of completed training activities

A. The Basic Package (3 ECTS)

WIAS Introduction Day (2021)
WGS Scientific Integrity course (2024)
WGS Ethics in Animal Sciences course (2025)
Introduction course on Personal Effectiveness (2021)

B. Disciplinary Competences (14 ECTS)

Research proposal (2021)
Secondment IRTA (2021)
Microbial Product Screening (2024)
Proteomics course (2022)
Training course on Tnseq (2022)
Organization of training course on Tnseq (2022)

C. Professional Competences (5 ECTS)

Research and Data Management (2021)
Scientific Communication and Data management (2021)
IPR and Project Leadership (2022)

D. Societal Relevance

E. Presentation Skills (4 ECTS)

Poster presentation, 7th Joint Microbiology & Infection Conference (DGHM & VAAM), Würzburg (Germany), 2024.
Oral presentation, Host-Microbiome Genetics Symposium, Wageningen (The Netherlands), 2023.
Oral presentation, Spanish Meeting of Molecular Microbiology, Santander (Spain), 2024
Oral presentation, Host-Microbe Genetics Meeting, Leiden (The Netherlands), 2024.

F. Teaching competences (6 ECTS)

Master student Nissa Arifa (2023)
Master student Troy Wagenaar (2024)
Master student Amy Rafaela (2025)

Colophon

The research described in this thesis was financially supported by the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 956154. Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Claudia Sahagún Fernández
Layout by Promotie In Zicht // www.promotie-inzicht.nl
Printed by ProefschriftMaken // www.proefschriftmaken.nl