The application of high-throughput sequencing to study the genome composition and transcriptional response of *Haemophilus influenzae*

by

Pavelas Sazinas

Thesis submitted for the degree of Doctor of Philosophy in Medical Sciences

University of Warwick, Warwick Medical School

September 2016

Table of contents

List of figures

List of tables

Acknowledgements

First and foremost, I would like to wholeheartedly thank my supervisor, Dr Esther Robinson, for always guiding me with diligence, patience and moral support. Needless to say, Esther made my PhD journey an enjoyable and exciting experience. I would also like to thank my co-supervisor, Professor Mark Pallen, for his expertise, advice and good humour.

I would like to extend my gratitude to Dr Andrew Millard for his help with bioinformatic analyses, Dr Meera Unnikrishnan for helping me to set up cell culture experiments, and Dr Emma Denham for the RNA-related advice. I want to thank Richard Brown for his help with Python programming and for sharing PhD struggles and successes from our first day as PhD students. I thank Emily Stoakes, Holly Hall, Dr Jacqueline Chan, Dr Nicholas Duggett, Ross Slater and the rest of the lab team for making these last three and a bit years a fun ride. That, of course, also includes Dr Alexia Hapeshi, Dr Josie McKeown and Dr Gemma Kay, who get an additional special thanks for proofreading my thesis. I am very grateful to the rest of the Microbiology and Infection group for their assistance in the project and for the never-boring coffee and lunch breaks.

My most sincere gratitude to my partner, David, for always being there through all my ups and downs. His endless encouragement, love and support during my PhD, and particularly whilst writing the thesis, was invaluable. I must also thank my friends, especially Jonas, Mantas and Rasa, for always being able to lift my spirits during challenging times.

I want to extend my utmost gratitude to my mother, Vera, for her love and reassurance, for always being there for me and believing in my abilities. Finally, I dedicate this thesis to the memory of my grandmother, Domicele Lukauskiene, who was my biggest supporter and who would be very proud to see my academic achievements.

Author's declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree in this or any other university.

The work presented (including data generated and data analysis) was carried out solely by myself.

Signed.....

Date.....

Abstract

Haemophilus influenzae is an important human pathogen, responsible for respiratory infections, such as otitis media, bronchitis and epiglottitis, as well as invasive disease. Despite being the first free-living organism to have its whole genome sequenced, there have been only a few published studies investigating its transcriptional profile using next-generation sequencing (NGS). The work presented in this thesis aimed to use NGS to improve the understanding of how *H. influenzae* behaves during natural infection and to identify novel RNA structures with potentially important roles in pathogenesis.

The whole transcriptome of *H. influenzae* during infection-relevant conditions was analysed using high-throughput RNA sequencing. For the first time, the transcriptional profile of *H. influenzae* during stationary phase and nutritional stress was determined on a whole-genome scale. Differential gene expression analysis of an invasive strain, R2866, and a laboratory strain, Rd KW20, revealed differences in their transcriptional response, particularly during oxidative stress and iron starvation. Importantly, a new systematic and robust bioinformatic tool, "toRNAdo", was developed to identify non-coding RNA elements from the bacterial transcriptomic data. It enabled discovery of a repertoire of novel putative intergenic and antisense non-coding RNAs in *H. influenzae.* In addition, the first fully sequenced genome of a free-living organism, the Rd KW20 strain of *H. influenzae*, was re-sequenced and re-annotated for the first time. This enabled identification of multiple nucleotide-level differences between original and re-sequenced genomes of Rd KW20.

The work presented here facilitates future characterisation of novel RNA elements, with potentially important regulatory roles in pathogenesis in *H. influenzae*, and has implications for defining a model bacterial strain. Importantly, the findings present significant insight into the pathogenic lifestyle of *H. influenzae*. They provide the basis for further work, where novel vaccine and antibiotic targets may get developed.

List of abbreviations

A: adenine

A: alanine

ACT: Artemis Comparison Tool

ANI: average nucleotide identity

ATP: adenosine triphosphate

BAM: binary alignment/map

BHI: brain-heart infusion

BLAST: basic local alignment search tool

BLNAR: beta-lactamase negative ampicillin resistant

BLPACR: beta-lactamase-positive amoxicillin-clavulanate-resistant

bp: base pair

BR: broad range

C: cysteine

C: cytosine

°C: degree Celsius

cDNA: complementary DNA

CFU: colony-forming unit

CO_2: carbon dioxide

COPD: chronic obstructive pulmonary disorder

CRISPR: clustered, regularly interspaced short palindromic repeat

CTP: cytidine triphosphate

D: aspartic acid

DAPI: 4',6-diamidino-2-phenylindole

DAVID: Database for Annotation, Visualization, and Integrated Discovery

DMEM: Dulbecco's Modified Eagle medium

DMSO: dimethyl sulphoxide

DNA: deoxyribonucleic acid

dNTP: deoxynucleotide

dRNA-Seq: differential RNA-Seq

E: glutamic acid

EDTA: ethylenediaminetetraacetate

F: phenylalanine

FBS: fetal bovine serum

FNR: fumarate-nitrate transcriptional regulator

Fur: Ferric Uptake Regulator

g: gram

g: gravitational force

GFF: general feature format

GO: gene ontology

GTP: guanosine triphosphate

G: guanine

H: histidine

Hib: *H. influenzae* type b

HS: high sensitivity

HTS: high-throughput sequencing

I: isoleucine

ICE: integrative and conjugative element

IMP: inosinic acid

IVS: intervening sequence

K: lysine

kb: kilobase

kcal: kilocalory

KEGG: Kyoto Encyclopaedia of Genes and Genomes

L: leucine

l: litre

LOS: lipooligosaccharide

LPS: lipopolysaccharide

M: methionine

M: molar

Mb: megabase

μg: microgram

mg: milligram

MIV: medium IV

mJ: millijoule

μl: microlitre

ml: millilitre

mM: millimolar

MOI: multiplicity of infection

mol: mole

MOPS: 3-(*N*-morpholino)propanesulfonic acid

mRNA: messenger RNA

N: asparagine

NAD: nicotinamide adenine dinucleotide

NCBI: National Centre for Biotechnology Information

ncRNA: non-coding RNA

ng: nanogram

NGS: next-generation sequencing

nm: nanometer

nM: nanomolar

NNC: normalised nucleotide coverage

NTHi: non-typeable *H. influenzae*

NTP: nucleoside triphosphate

OD: optical density

PBS: phosphate-buffered saline

PCR: polymerase chain reaction

pM: picomolar

PPIX: protoporphyrin IX

Q: glutamine

R: arginine

RefSeq: Reference Sequence

RNA: ribonucleic acid

RNA-Seq: RNA sequencing

ROS: reactive oxygen species

RPKM: reads per kilobase per million reads

rpm: revolutions per minute

rRNA: ribosomal RNA

RT: room temperature

RT-qPCR: real-time quantitative polymerase chain reaction

S: serine

SAM: sequence alignment/map

sBHI: supplemented brain-heart infusion

SDS: sodium dodecyl sulphate

SE: standard error

SNP: single nucleotide polymorphism

SOD: superoxide dismutase

sRNA: small RNA

SSR: simple sequence repeat

T: threonine

T: thymine

TAE: tris-acetate-ethylenediaminetetraacetate

TE: tris-ethylenediaminetetraacetate

TPM: Transcripts per Million

tRNA: transfer RNA

U: unit

UTP: uridine triphosphate

UTR: untranslated region

UV: ultraviolet

V: valine

V: volt

w/v: weight/volume

Y: tyrosine

Chapter 1: Introduction

Haemophilus influenzae was the first free-living organism to have its whole genome sequenced in 1995, signifying the start of the genomic era in bacteriology (Fleischmann et al., 1995). Since then, the advances in next-generation sequencing (NGS) technologies allowed researchers to study bacterial organisms on a whole-genome scale, including the work presented in this thesis. *H. influenzae* is an important human pathogen, with historical and present-day significance. This study aimed to investigate the behaviour of this organism during infection-relevant conditions using NGS. It provides the most comprehensive repertoire of non-coding RNAs (ncRNAs) in *H. influenzae* to-date as well as explores implications of using old annotated genomes of model organisms. The findings presented in this work provide new insight into the pathogenesis of *H. influenzae* and lay important groundwork for future work on potentially important RNA elements in this organism.

1.1 *Haemophilus influenzae*

1.1.1 *Haemophilus* genus

Haemophilus is a genus of Gram-negative, non-motile coccobacilli, belonging to the Pasteurellaceae family of the phylum Proteobacteria. *Haemophilus* species are known to cause diseases in a range of hosts as well as behave as a commensal (Norskov-Lauritsen, 2014). These species share the growth requirement for at least one of blood-associated growth factors: haemin (factor X) and nicotinamide adenine dinucleotide (NAD; factor V) (Norskov-Lauritsen, 2014). The most significant and extensively studied member of the *Haemophilus* genus is *H. influenzae*, capable of causing important infections in humans, as described below. A list of *Haemophilus* species specific to the human host, along with their clinical presentations, is presented in Table 1.1. In addition, there are *Haemophilus* species with different animal hosts, including *Haemophilus parasuis*, *Haemophilus paracuniculus* and *Haemophilus felis*, isolated from pigs,

rabbits and cats respectively (Targowski and Targowski, 1979, Hoefling, 1991, Inzana et al., 1992).

Table 1.1: List of *Haemophilus* species with human host specificity.

***Haemophilus* species**	**Clinical presentation**	**Reference**
H. influenzae	Respiratory tract diseases, meningitis, bacteraemia, endocarditis	(Makela et al., 1992, Brouqui and Raoult, 2001, Agrawal and Murphy, 2011)
H. influenzae biogroup *aegyptius*	Acute conjunctivitis, Brazilian purpuric fever	(Pittman and Davis, 1950, Tondella et al., 1995)
Haemophilus parainfluenzae	Endocarditis	(Brouqui and Raoult, 2001)
Haemophilus haemolyticus	Common commensal, rare invasive disease	(Anderson et al., 2012)
Haemophilus parahaemolyticus	Pharyngitis, endocarditis	(Pittman, 1953)
Haemophilus paraphrohaemolyticus	Rare clinical presentation	(Norskov-Lauritsen et al., 2012)
Haemophilus ducreyi	Chancroid	(DiCarlo et al., 1995)
Haemophilus pittmaniae	Respiratory disease	(Boucher et al., 2012)
Haemophilus sputorum	Rare clinical presentation	(Norskov-Lauritsen et al., 2012)

1.1.2 Overview of *H. influenzae*

H. influenzae is a small-celled (1 μm x 0.3 μm), facultatively anaerobic, fastidious organism that has an absolute growth requirement for factors X and V. It forms pale grey colonies on chocolate agar and grows optimally at 35-37 °C with 5% CO_2, mirroring its ability to colonise the human upper respiratory tract. It was initially described incorrectly as the causative agent of influenza, which explains the aetiology of the name of the species (Pfeiffer, 1892). Originally called *Bacillus influenzae*, it was not renamed *H. influenzae* until more than two decades later (Pfeiffer, 1892, Winslow et al., 1920). *H. influenzae* was first identified as the cause of meningitis in 1911, while the involvement of capsulated strains in clinical manifestation of meningitis was characterised 20 years later (Wollstein, 1911, Pittman, 1931).

Human is the only natural host of *H. influenzae*, where it is generally present as a commensal. However, *H. influenzae* can also act as a pathogen in susceptible individuals, who include infants, the elderly and immunocompromised adults (van Wessel et al., 2011). In addition, viral infections as well as diseases like chronic obstructive pulmonary disorder (COPD) and cystic fibrosis also predispose patients to infection with *H. influenzae* (Smith et al., 1976, Murphy et al., 2004, Cardines et al., 2012).

1.1.3 Classification of *H. influenzae*

The pathogenicity and type of diseases caused by *H. influenzae* are dependent on the presence or absence of a capsule. This is a polysaccharide structure present outside the cell wall and acts as an important virulence factor in *H. influenzae* (see section 1.1.7). All typeable *H. influenzae* strains are encapsulated and classified based on the capsule structure (serotypes a-f), with type b being the most historically important as a common aetiological agent of an invasive disease (Pittman, 1931, Makela et al., 1992). Non-typeable *H. influenzae* (NTHi) strains lack a capsule and are most frequently found as respiratory tract

commensals, but also have the potential to cause local and invasive disease (see section 1.1.3.2).

1.1.3.1 *H. influenzae* type b

H. influenzae type b (Hib) used to be the most common cause of bacterial meningitis in infants and young children, as well as an important causative agent of diseases like epiglottitis, septic arthritis, pneumonia and generalised bacteraemia, before the introduction of an effective capsular polysaccharide vaccine (Makela et al., 1992). The initiation of Hib vaccine in the developed world in the late 1980s and early 1990s has resulted in a significant decrease in Hib carriage and the prevalence of Hib infections (Hargreaves et al., 1996, Agrawal and Murphy, 2011). Routine Hib vaccination began in the UK in 1992, which resulted in a decrease in over 96% of confirmed cases of Hib infections in children by 1999 (Heath et al., 2000, Ladhani, 2012). While there was an increase of Hib disease cases in the following few years, an improved vaccination programme resulted in the number of documented cases of the invasive Hib disease being reduced to 30 in England and Wales in 2010 (Ladhani, 2012).

In addition to the near elimination of Hib disease, there has been a change in those in the population who are most at risk of it: adults with comorbidities, as opposed to infants and young children, are now most commonly presenting with Hib infection in England and Wales, though the prevalence is very low (Collins et al., 2013). The Hib vaccine has also been increasingly introduced in the developing world, with similar successful outcomes (Lee et al., 2008). Hib still remains an important pathogen in the countries where Hib vaccination was never established, contributing to subsequent debilitating sequelae and high mortality rates (Ahmed et al., 2013). Other serotypes of typeable *H. influenzae* as well as non-typeable strains can also cause invasive disease (Shuel et al., 2011, Golebiewska et al., 2016, Tsang et al., 2016).

1.1.3.2 NTHi

NTHi is a common commensal in humans with the carriage rate reaching over 60% in children and remaining high in adulthood (Farjo et al., 2004). The presence of a number of virulence factors allows it to cause disease in humans, though the exact mechanisms of the transition from commensal to a pathogenic state are still poorly understood (see section 1.1.7). NTHi is a frequent causative agent of otitis media in infants and children, especially after the introduction of pneumococcal vaccination (Casey and Pichichero, 2004, Agrawal and Murphy, 2011). Otitis media is a very common disease in young children and, whilst the mortality is low, the possible sequelae make it a disease with a significant socioeconomic burden (Monasta et al., 2012). NTHi is also associated with respiratory tract infections (bronchitis and pneumonia) in infants, children and adults, and is linked to exacerbations of COPD (King, 2012).

NTHi is capable of causing invasive disease; high levels of invasive NTHi disease were observed in prematurely born infants and older adults (>65 year old) (Dworkin et al., 2007, van Wessel et al., 2011). Small nosocomial outbreaks of virulent NTHi have also been reported, particularly among elderly patients (Yang et al., 2010, Andersson et al., 2015). It remains unclear whether the introduction of the Hib vaccine has lead to an increase in the number of NTHi cases as well as a shift towards NTHi causing more invasive disease, as several studies give conflicting evidence on this matter (O'Neill et al., 2003, McConnell et al., 2007, Kalies et al., 2009, Agrawal and Murphy, 2011). This suggests that a more careful monitoring of the potentially increasing clinical threat of NTHi must be undertaken.

1.1.4 Antibiotic resistance

A significant contribution to the clinical importance of *H. influenzae* is its resistance to various antibiotics, particularly to beta-lactams. They are broad-range antibiotics that possess a beta-lactam ring, which act by inhibiting bacterial cell wall synthesis. A large number of *H. influenzae* strains containing

ROB-1 and TEM beta-lactamases have been identified (Scriver et al., 1994). These are enzymes that hydrolyse the beta-lactam ring and thereby deactivate beta-lactam antibiotics. High rates of beta-lactamase negative ampicillin resistant (BLNAR) strains, with a different beta-lactam resistance mechanism, have also been reported in USA and Japan (Hasegawa et al., 2004, Shuel et al., 2011). BLNAR strains are resistant to beta-lactam antibiotics due to the modification of their penicillin-binding proteins rather than an enzymatic action. Some strains have been shown to possess both mechanisms of beta-lactam resistance and are termed beta-lactamase-positive amoxicillin-clavulanate-resistant (BLPACR) (Matic et al., 2003).

Resistance to quinolones, which act to inhibit DNA replication via binding to the DNA gyrase or topoisomerase IV, is rare in *H. influenzae* and yet highly resistant strains have been reported (Rodriguez-Martinez et al., 2006). Both intrinsic and mobile genetic element-encoded resistance to chloramphenicol, macrolides, and tetracyclines (that inhibit protein synthesis via different modes of action) has been described as well, although not to the same extent as to beta-lactam antibiotics (Tristram et al., 2007).

Multi-drug resistant *H. influenzae* are consistently identified among clinical isolates, posing significant threat to current treatment strategies (Reis et al., 2002, Pfeifer et al., 2013, Skaare et al., 2014). Mobile genetic elements, such as integrative and conjugative elements (ICE), have in particular contributed to the increased spread of antibiotic resistance in naturally competent *H. influenzae* (Campos et al., 2003, Mohd-Zain et al., 2004). The transfer of genetic information can also occur between *H. influenzae* and closely-related species, such as *Haemophilus haemolyticus* and *Haemophilus parainfluenzae*, emphasising the need for an effective control of the transfer of antibiotic resistance among bacterial populations (Scheifele et al., 1982, Takahata et al., 2007, Sondergaard et al., 2015).

1.1.5 Rd KW20 and R2866 strains of *H. influenzae*

As of March 2016, there are 123 whole-genome sequences of *H. influenzae* strains, including 16 complete (without gaps in the sequence) genomes, on the National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/genome/). These include strains of *H. influenzae* that have been extensively used for laboratory research in the past. Rd KW20 (henceforward referred to as Rd), one of the two strains of *H. influenzae* used in this study, was isolated in mid-1940s by culturing a type d strain of *H. influenzae* and selecting a mutant that had lost the capsule (Alexander and Leidy, 1951). Since then it has been propagated as a standard laboratory strain and was used as the representative strain of *H. influenzae* for whole-genome sequencing in 1995 (Fleischmann et al., 1995). Rd was also the source of the first type II restriction enzyme *Hin*dII - the work, for which Hamilton O. Smith was awarded the Nobel prize in 1978 (Smith and Wilcox, 1970, Linn, 1978). The other strain of *H. influenzae* used in this study, R2866, was isolated in 1996 from the blood of a 30-month old child with meningitis and has since been used as an exemplar of an invasive NTHi strain (Nizet et al., 1996).

1.1.5.1 Phenotypic differences between Rd and R2866

R2866, an invasive NTHi isolate, has significant phenotypic differences to Rd, a classic laboratory strain. The major difference is an increased resistance of R2866 to the human serum (Erwin et al., 2005). This has been attributed to the expression of a glycosyltransferase gene, *lgtC*, responsible for modulating the composition of lipooligosaccharide (LOS) (see section 1.1.7) (Erwin et al., 2006). The ability of R2866 to circumvent the complement cascade by delaying the deposition of the C3 component has also been linked to an increased survival in the human serum (Williams et al., 2001). In addition, different growth rates have been observed for Rd and R2866 grown at pH 6.8, 7.0 and 8.0 (Ishak et al., 2014). Comparing the transcriptional response to iron starvation has also revealed differences in the iron-regulated genes between these two strains (Whitby et al., 2009).

1.1.5.2 Genetic differences between Rd and R2866

The availability of whole-genome sequencing provides the basis for determining genetic evidence for phenotypic differences. Rd and R2866 differ in their genome size by 0.1 Mb, as inferred from the NCBI database (see Table 1.2). Consequently, these strains also have a different number of genome features, with Rd having 1,610 and R2866 having 1,796 protein-coding genes. Interestingly, despite having a smaller genome, Rd is found to possess a larger number of pseudogenes than R2866.

Table 1.2: Whole-genome sequence features of Rd and R2866 strains, as inferred from the NCBI database.

Genome feature	Rd	R2866
Genome size (Mb)	1.83	1.93
GC content (%)	38.2	38.1
Number of genes	1,765	1,900
Number of protein-coding genes	1,610	1,796
Number of pseudogenes	67	22
Number of ribosomal RNAs (rRNAs)	19	19
Number of transfer RNAs (tRNAs)	58	59

1.1.5.2.1 Core and accessory genomes of Rd and R2866

The accessory genome of a set of bacterial strains or species represents genes that are present in some members of that set, but not in others. This is in contrast to the core genome, which is a subset of genes present in all compared strains or species. Core and accessory genomes were previously calculated for 13 strains of *H. influenzae* in a pairwise manner (Hogg et al., 2007). The core genome of Rd and R2866 was found to contain 1,576 genes. The Rd strain was shown to possess extra 134 genes, while R2866 had extra 259 genes. Rd contains a urease locus, which in the R2866 strain is replaced by *mtrF* - a gene linked to antimicrobial resistance (Veal and Shafer, 2003, Erwin et al., 2005). R2866, unlike Rd, possesses a tryptophanase locus, which is responsible for indole production and has been used for biotyping of *H. influenzae*, due its association with an invasive phenotype (Kilian, 1976, Martin et al., 1998).

A general feature of any accessory genome is the abundance of mobile genetic elements (Kung et al., 2010, Jackson et al., 2011). This provides the bacterial species with a pool of available genes for the transfer between different strains, potentially conferring new advantages, such as antibiotic resistance (Elwell et al., 1975). The presence of a 53-kb ICE, homologous to ICE*Hin*1056, as well as bacteriophage HP2 has been described in the R2866 strain, but not in Rd (Williams et al., 2002, Hogg et al., 2007, Juhas et al., 2007).

1.1.6 *H. influenzae* as a model organism

Due to its relatively small genome, genetic manipulability and close phylogenetic distance to *Escherichia coli*, *H. influenzae* has been a popular model organism for various microbiology studies (Evers et al., 1998, Kolker et al., 2003). The availability of a complete genome sequence for the Rd strain from the mid-1990s made *H. influenzae* an even more attractive model for genomic studies (Fleischmann et al., 1995). However, the original isolation of the Rd strain in the mid-1940s and its subsequent propagation in different laboratories over several decades might have lead to its significant genetic divergence (Alexander and Leidy, 1951). Current descendants of the original Rd strain have

likely accumulated a number of mutations, leading to small but possibly relevant changes in the behaviour of this strain. It is possible that the Rd strain, originally sequenced in 1995, is genetically and phenotypically different to other Rd descendants present in other laboratories (Fleischmann et al., 1995). The phenotypic differences in the same strain from different laboratories have been documented previously in *E. coli*, highlighting the accumulation of variations within the same strain over time (Soupene et al., 2003).

1.1.7 Virulence of *H. influenzae*

Virulence factors are pathogen-produced products, which facilitate the establishment and progression of disease. Both typeable and non-typeable *H. influenzae* strains contain virulence factors that promote infection. The presence of a capsule, especially in type b strains, significantly contributes to the ability of typeable strains to invade and persist in the bloodstream, by increasing bacterial resistance to the action of complement (Moxon and Vaughn, 1981, Sutton et al., 1982). NTHi strains have also been shown to invade the bloodstream and cause invasive disease, meaning that they utilise different virulence factors to achieve this (O'Neill et al., 2003).

In order to cause invasive or localised disease, *H. influenzae* needs to be able to circumvent mucosal barriers, evade host immune response and successfully adhere to the host epithelium. For that purpose, *H. influenzae* produces different virulence factors, aiding in establishment and development of disease, which are summarised in Table 1.3. Novel virulence factors still continue to be discovered in *H. influenzae*, such as *msf*, a recently characterised gene associated with an intracellular survival in macrophages (Kress-Bennett et al., 2016).

Table 1.3: Summary of major virulence factors in *H. influenzae*.

Virulence factor	Role in *H. influenzae*	Reference
Polysaccharide capsule	Invasion, persistence in the bloodstream, competence evasion	(Moxon and Vaughn, 1981, Sutton et al., 1982)
LOS	Adhesion, invasion, immune evasion, damage to host	(Johnson and Inzana, 1986, Swords et al., 2000)
Pili (fimbriae)	Adherence to nasopharyngeal tissue and respiratory mucus	(Loeb et al., 1988, Kubiet et al., 2000)
High molecular weight adhesins HMW1 and HMW2	Adherence to host epithelium	(St Geme et al., 1993)
Protein E	Adherence to components of the extracellular matrix (ECM)	(Singh et al., 2010)
Proteins P2 and P5	Adherence to respiratory mucus	(Reddy et al., 1996)
Autotransporter Hia	Adherence to host epithelium	(Barenkamp and St Geme, 1996, St Geme and Cutter, 2000)
Autotransporter Hap	Adherence to host epithelium and components of the ECM	(St Geme et al., 1994, Fink et al., 2002)

H. influenzae was among the first bacteria where simple sequence repeats (SSRs), associated with phenotype switching of cell surface structures, were reported (Weiser et al., 1989). They are 1-8 base pair (bp) long hypermutable tandem repeats; regions containing them have been defined as simple sequence contingency loci (Bayliss et al., 2001). SSRs contribute to virulence through facilitating phase variation - a process of reversible phenotypic change. SSRs appear in open reading frames and promoter regions of important virulence-associated contigency loci in *H. influenzae*. These include several genes coding for pilin subunits, adhesins HMW1 and HMW2, several LOS-associated genes as well as a restriction-modification gene, *mod* (Weiser et al., 1989, van Ham et al., 1993, Dawid et al., 1999, De Bolle et al., 2000). The phase variation of the latter gene induces global changes in gene expression in *H. influenzae* (Srikhanta et al., 2005). Ability to switch between different phenotypes provides *H. influenzae* with a more effective niche colonisation and persistence, thus contributing to its pathogenesis.

1.1.8 Stress response in *H. influenzae*

H. influenzae is likely to encounter a variety of stresses during infection of the host, including nutrient limitation and host immune response. Bacterial adaptation to these stresses ensures successful colonisation and subsequent disease progression. When DNA damage is inflicted, bacteria respond by inducing their "SOS" regulon, which largely contains genes involved in DNA repair (Radman, 1975). The SOS response has been described in *H. influenzae* as well, mediated by the transcriptional repressor LexA (Sweetman et al., 2005). The expression of five genes *recA*, *recN*, *recX*, *ruvA* and *lexA*, with roles in DNA repair and recombination, was induced as part of this response in *H. influenzae* (Sweetman et al., 2005). Other important stress response pathways in *H. influenzae* are discussed below.

1.1.8.1 Oxidative stress

Reactive oxygen species (ROS) are generated as part of normal cell metabolism in living cells and cause damage to DNA, proteins and other biomolecules (Gonzalez-Flecha and Demple, 1995, Tamarit et al., 1998, O'Rourke et al., 2003). *H. influenzae* has evolved to deal with this oxidative stress in a variety of mechanisms, as it needs to combat ROS generated by the host immune system, its own aerobic respiration and co-pathogens, such as *Streptococcus pneumoniae* (Harrison et al., 2012). Major ROS molecules include superoxide, hydrogen peroxide and hydroxyl radicals. The most important components of oxidative stress response in *H. influenzae* are summarised in Figure 1.1 and are described in detail below.

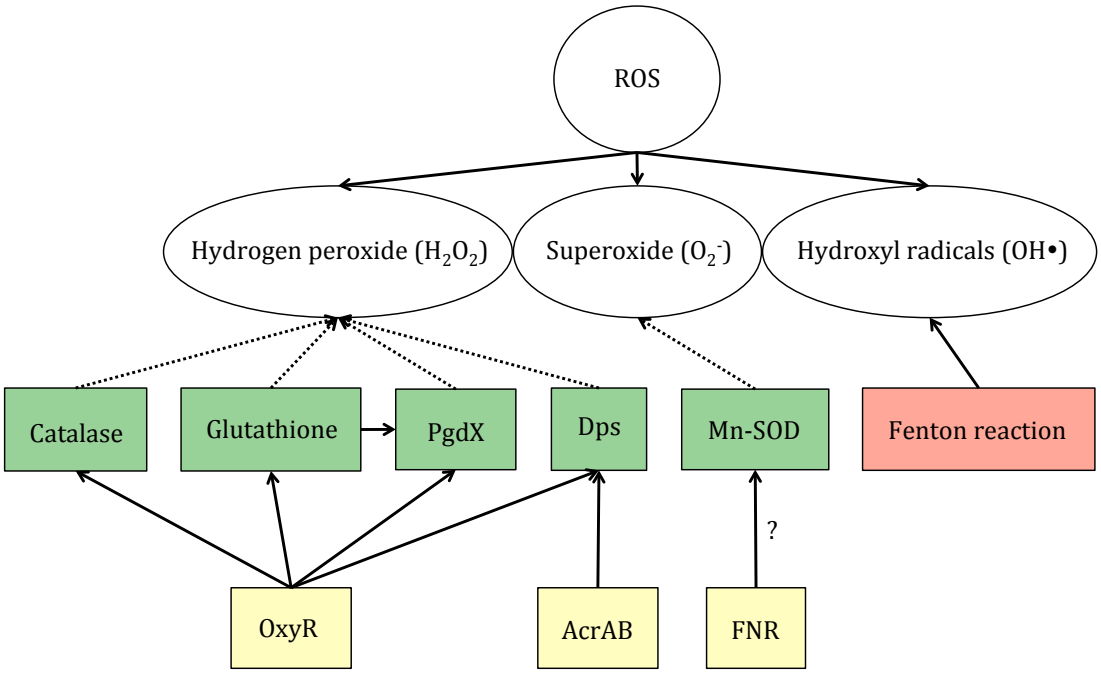

Figure 1.1: Major oxidative stress response components in *H. influenzae*. Solid and dotted arrows show positive and inhibitory effects respectively. In yellow are transcriptional regulators, while in green are oxidative stress effectors. The Fenton reaction, with a positive effect on production of hydroxyl radicals, is depicted in red.

1.1.8.1.1 Regulation of oxidative stress response

OxyR has been described as a major bacterial regulator of oxidative stress response (Zheng et al., 2001). In the *H. influenzae* strain 86-028NP, OxyR was shown to regulate the expression of 11 genes, most of which play a role in the defence against oxidative stress (Harrison et al., 2007). Mutation of the *oxyR* gene in *H. influenzae* resulted in increased sensitivity to hydrogen peroxide as well as decreased fitness and virulence in a rat model of *H. influenzae* infection (Whitby et al., 2012).

The two-component system ArcAB is another regulator that is involved in the response to oxidative stress in *H. influenzae*, particularly during anaerobic growth (Wong et al., 2007). It was shown to regulate the expression of the *dps* gene, encoding a ferritin-like protein, which protects against oxidative stress by sequestering free ferrous iron (see section 1.1.8.2) (Wong et al., 2007). Interestingly, *dps* is also part of the OxyR regulon in *H. influenzae* and is known to bind DNA protecting it from oxidative damage in *E. coli* (Martinez and Kolter, 1997, Harrison et al., 2007). The importance of *dps* in the defence against oxidative stress was further demonstrated by an increased sensitivity of *H. influenzae* to hydrogen peroxide in a *dps* knockout mutant (Harrison et al., 2015).

The ArcAB system has been found to be protective against the bactericidal activity of the human serum by regulating resistance to complement, which highlights its role in the survival of *H. influenzae* within the human host (De Souza-Hart et al., 2003, Wong et al., 2011). The fumarate-nitrate transcriptional regulator (FNR) has also been shown to play a role in defence against oxidative stress, with manganese-associated superoxide dismutase (SOD) likely being a part of its regulon (Kroll et al., 1993, Jiang et al., 2016a).

1.1.8.1.2 SOD

SOD is the main enzyme for protecting the cell against superoxide-related stress, which is an unavoidable consequence of aerobic respiration. Out of four known types of SODs, *H. influenzae* has only been shown to possess a functional manganese-associated SOD (Kroll et al., 1993). The gene *sodC,* coding for copper-zinc SOD, is present in some capsulated *H. influenzae*, although the enzyme itself is not functional (Kroll et al., 1991). Neither nickel nor iron SOD has been identified in *H. influenzae* to-date.

1.1.8.1.3 Catalase

Catalase is a key enzyme responsible for scavenging hydrogen peroxide from the cell. In *H. influenzae* it is encoded by the *hktE* gene, which is part of the OxyR regulon (Harrison et al., 2007). Catalase activity in *H. influenzae* was found to be lower during stationary growth phase when compared to exponential phase, which is contrary to catalase activity in *E. coli* (Loewen and Triggs, 1984, Bishai et al., 1994). It was also the most up-regulated gene in response to hydrogen peroxide in the 86-028NP strain (Harrison et al., 2007). It has been suggested that catalase is more effective when dealing with high concentrations of hydrogen peroxide, whereas other antioxidants, such as PgdX, are more efficient at scavenging low concentrations (Vergauwen et al., 2001, Vergauwen et al., 2003a, Pauwels et al., 2004).

1.1.8.1.4 PgdX and glutathione

PgdX is a glutathione-dependent peroxiredoxin-glutaredoxin and it was shown to play an important role in protecting *H. influenzae* against oxidative stress during stationary phase, when catalase is being down-regulated (Pauwels et al., 2004). Interestingly, a PgdX mutant was more resistant to hydrogen peroxide stress due to an increased activity of catalase, highlighting the evolved compensatory oxidative stress control mechanisms in *H. influenzae* (Pauwels et al., 2004, Harrison et al., 2015). Both PgdX and catalase are the two major

scavengers of hydrogen peroxide in *H. influenzae* as shown by the diminished resistance of double mutants to hydrogen peroxide (Vergauwen et al., 2006).

The activity of PgdX is reliant on the presence of glutathione, which was demonstrated by the sensitivity of a catalase mutant of *H. influenzae* to oxidative stress when grown in a medium lacking glutathione (Vergauwen et al., 2003a). Glutathione is a thiol molecule involved in the protection against hydrogen peroxide in both prokaryotes and eukaryotes. *H. influenzae* lacks the gene encoding glutathione, but possesses genes required for the import of extracellular glutathione and its subsequent metabolism (Vergauwen et al., 2003b). The presence of a large number of genes directly involved in the protection against oxidative stress in a relatively small *H. influenzae* genome emphasises the importance of these overlapping defence mechanisms for the organism.

1.1.8.2 Iron-mediated oxidative stress

H. influenzae has a growth requirement for haem, which is produced by ferrochelatase enzyme combining iron and protoporphyrin IX (PPIX) (Granick and Gilder, 1946). However, *H. influenzae* does not have enzymes for the biosynthesis of PPIX, meaning that the haem requirement must be satisfied by other means (White and Granick, 1963). For that purpose, *H. influenzae* has evolved different mechanisms to salvage extracellular haem, PPIX and iron molecules. Iron uptake must be carefully controlled by *H. influenzae*, as there is a clear link between oxidative stress and free intracellular ferrous iron (Touati, 2000). Ferrous iron interacts with hydrogen peroxide and generates free hydroxyl radicals through the Fenton reaction (see Figures 1.1; 1.2). Thus it is critical for bacteria like *H. influenzae* to carefully balance the necessary iron uptake against harmful oxidative effects.

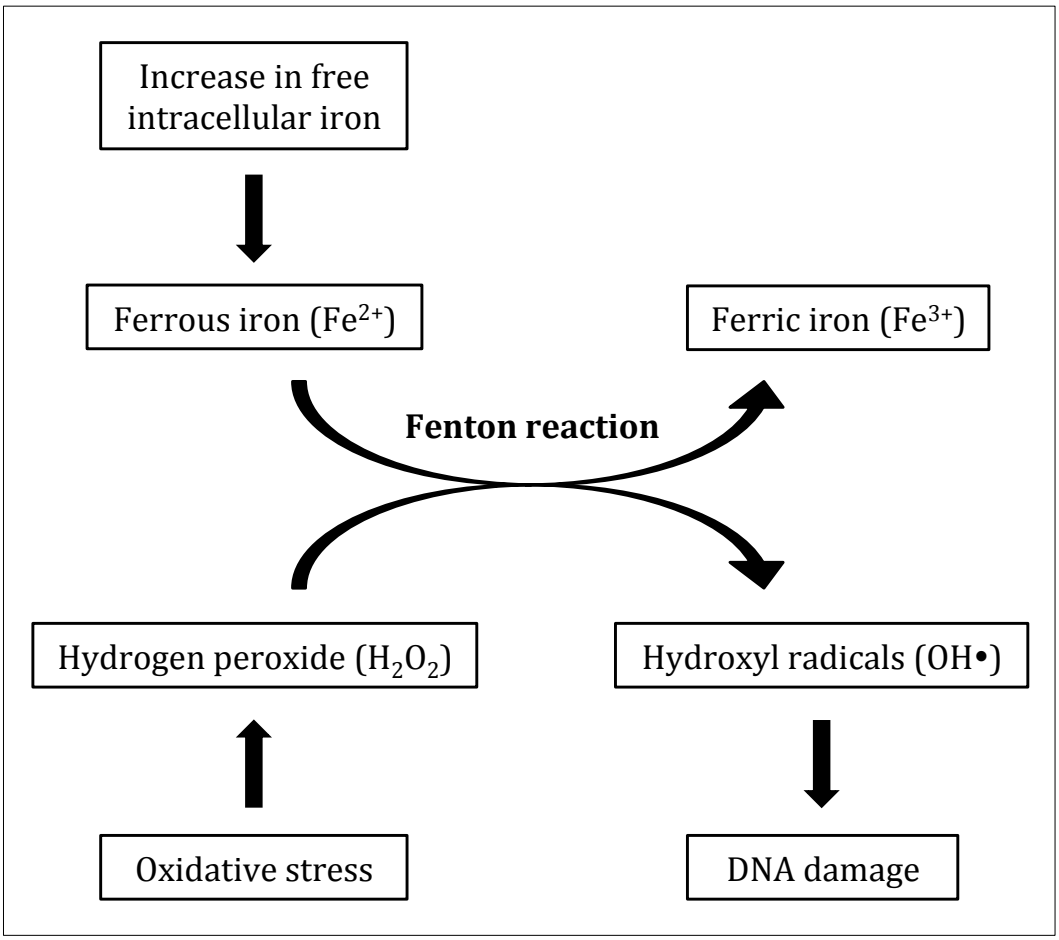

Figure 1.2: A schematic of the Fenton reaction, induced by oxidative stress and loss of iron homeostasis.

1.1.8.2.1 Regulation of iron homeostasis

The Ferric Uptake Regulator (Fur) is a transcriptional repressor of genes responsible for iron uptake into the cell. In *E. coli* it is regulated by OxyR, which highlights the close relationship between iron homeostasis and oxidative stress (Zheng et al., 1999). Despite not being part of OxyR regulon in *H. influenzae*, Fur still plays an important role in the defence against oxidative stress. Fur mutants were shown to have increased sensitivity to hydrogen peroxide (Harrison et al., 2015). This can be explained by the Fur-related loss of iron homeostasis, which leads to increased intracellular levels of ferrous iron and generation of more hydroxyl radicals through the Fenton reaction (see Figure 1.2).

Interestingly, several iron uptake genes were up-regulated upon hydrogen peroxide treatment in the 86-028NP strain (Harrison et al., 2007). This could possibly be due to the need to repair proteins containing iron-sulphur clusters damaged by oxidative stress, revealing the complexity of the interplay between iron homeostasis and oxidative stress defence (Flint et al., 1993). Ferritin and transferrin-like proteins are responsible for iron sequestration, and thus can protect the cell from accumulating free ferrous iron. Genes encoding such proteins have indeed been found to be up-regulated during oxidative stress in *H. influenzae* (Harrison et al., 2007). Double mutants of Fur and ferritin-like Dps in *H. influenzae* also showed very high sensitivity to hydrogen peroxide treatment, again emphasising a critical role that iron homeostasis plays during oxidative stress (Harrison et al., 2015).

1.1.8.3 Iron-starvation stress

H. influenzae is likely to reside within an iron-restricted environment during colonisation and infection of the human host, hence it needs to maximise its ability to acquire any available iron. *H. influenzae* has a range of overlapping mechanisms for iron uptake, the majority of which are regulated by Fur. The Fur regulon was described in the 86-028NP strain of *H. influenzae* and contains 73 proteins (Harrison et al., 2013). These include HitABC and HfeABCD iron transport systems, heme-binding proteins HxuABC, transferrin-binding proteins

TbpA and TbpB, a hemoglobin-haptoglobin binding protein HgpB, and a transmembrane protein TonB. The latter in particular plays a critical part in the iron uptake in *H. influenzae* as many of the surface-bound receptors, which are responsible for iron transport, function via a TonB-dependent mechanism (Postle, 1990, Jarosik et al., 1994, Jarosik et al., 1995). In addition, the expression of an *fhu* operon allows *H. influenzae* to acquire iron through the utilisation of extracellular siderophores, which are small iron-binding proteins secreted by other bacteria (Morton et al., 2010).

Several transcriptomic studies using microarrays have been performed to determine the response of *H. influenzae* to iron-starvation. In fact, a core iron-responsive regulon has been identified in five strains of *H. influenzae* - Rd, 10180, R2866, R2846 and 86-028NP (Whitby et al., 2006, Whitby et al., 2009, Whitby et al., 2013). The latter two strains were more sensitive to iron-starvation than others (Whitby et al., 2013). As expected, core regulon genes that were up-regulated during iron-starvation stress encoded the aforementioned proteins with important roles in iron acquisition: HitABC, HxuABC, TbpAB and TonB. A similar gene expression pattern was observed in *H. influenzae* in a chinchilla otitis media model, indicating an analogous iron-starved environment (Whitby et al., 2013). There were also a large number of iron-responsive non-core genes specific to each *H. influenzae* strain, suggesting that these strains possess different affinities and adaptation techniques to environments within the human host.

1.1.9 *H. influenzae* infection of human cells

H. influenzae is able to colonise and cause localised disease in a variety of niches within the human host, including the nasopharynx, middle ear and lung tissues. In addition, *H. influenzae* is capable of causing invasive disease after translocating across the epithelium and endothelium into the bloodstream. Cell culture models have been extensively used to study *H. influenzae* colonisation and infection processes. *H. influenzae* adheres to a number of different human cell types *in vitro*, including lung bronchial and alveolar epithelium,

oropharyngeal epithelium, conjunctival epithelium and brain microvascular endothelium (St Geme and Falkow, 1990, Holmes and Bakaletz, 1997, Daines et al., 2003, Morey et al., 2011, Singh et al., 2016).

H. influenzae has been shown to invade both epithelial and endothelial human cell types, a process mediated by the human cytoskeleton (St Geme and Falkow, 1990, Daines et al., 2003). Upon invasion, *H. influenzae* resides inside human cells in a non-replicative state within endosome-like compartments (Morey et al., 2011). The process of invasion likely has a significant contribution to persistence of *H. influenzae* during disease, aiding the evasion of host immune cells.

There are notable differences in pathogenesis between typeable *H. influenzae* and NTHi, with the former being a lot more effective at causing invasive disease. In typeable strains the capsule is required to avoid phagocytosis by macrophages and resist the action of complement (Sutton et al., 1982, Noel et al., 1992). It has been proposed that the level of encapsulation changes during colonisation process, with the absence of capsule linked to increased adhesion and invasion, due to enhanced interaction of surface-bound ligands with host receptors (St Geme and Falkow, 1991, St Geme and Cutter, 1996). This means that the capsule serves both as an asset and a hindrance to typeable *H. influenzae*, proving that a careful modulation of encapsulation is required during infection.

Due to complete absence of the capsule, NTHi strains are more efficient at adhesion and invasion than typeable strains. Despite NTHi primarily causing localised disease, its ability to adhere and invade human endothelial cells indicates its capacity to cause invasive disease through potential translocation into the bloodstream (Daines et al., 2004). The NTHi strain R2866 has been reported to transcytose human epithelium, i.e. traverse across the interior of the cell and emerge on the opposite side to its entry (VanWagoner et al., 2016). There is also evidence for *H. influenzae* undergoing paracytosis - a process by which bacteria pass between cells across the cell layer (van Schilfgaarde et al.,

1995). This shows that typeable and non-typeable *H. influenzae* strains all possess a variety of mechanisms to cause a diverse range of diseases in the human host. Interestingly, even the Rd strain, propagated in the laboratory for several decades, is still capable of adherence and invasion of both epithelial and endothelial cell types, suggesting its suitability as a model for *H. influenzae* virulence (Daines et al., 2003).

1.1.10 Gene expression in *H. influenzae*

Gene expression in prokaryotes is controlled through transcriptional and translational regulation pathways in response to environment (Nogueira and Springer, 2000, Balleza et al., 2009). Studying gene expression in bacteria can therefore shed light on their pathogenesis and potentially lead to the discovery of novel treatment strategies. A number of different methods have been used to explore gene expression in *H. influenzae.* Whole-genome microarrays were used to investigate genes involved in iron homeostasis (Whitby et al., 2009). Real-time quantitative polymerase chain reaction (RT-qPCR) was utilised to explore the differential gene expression of *H. influenzae* and *S. pneumoniae* in a co-culture as compared to monocultures (Cope et al., 2011). Northern blotting was used to study the effect of the presence of a *glpTQ* intergenic region on the expression of the flanking genes in *H. influenzae* (Song and Janson, 2003).

Since the advent of NGS, and RNA sequencing (RNA-Seq) in particular, there have been several studies to-date investigating the whole transcriptome of *H. influenzae* (see section 1.3.2) (Baddal et al., 2015). RNA-Seq was used to study the response of strains Rd and 86-028NP to the changes in oxygen levels in the growth medium (Jiang et al., 2016a, Jiang et al., 2016b). The transcriptional responses of *H. influenzae* to glucocorticosteroid treatment and intracellular nickel were also determined (Earl et al., 2015, Tikhomirova et al., 2015a). Whole transcriptomes of Rd and otitis media strains of *H. influenzae* were explored during co-culture with *S. pneumoniae* (Tikhomirova et al., 2015b). Finally, RNA-Seq was used to describe the transcriptional response of R2846 and a Hib strain

to increased pH levels in the growth medium (Ishak et al., 2014). Further applications of RNA-Seq to the study of the whole transcriptome of *H. influenzae*, including the identification of small regulatory RNAs across the whole genome, would allow a deeper insight into genes and the regulatory network involved in the pathogenesis of *H. influenzae* and its adaption to the changes in the environment.

1.2 Small RNAs in bacteria

The first regulatory RNAs were identified in prokaryotes over three decades ago, although the actual complexity and significance of this regulatory system in bacteria has only begun to be appreciated in recent years (Tomizawa et al., 1981, Mizuno et al., 1984). These short, non-coding, highly-structured RNA transcripts (~50-300 nucleotides), termed small RNAs (sRNAs), are increasingly shown to have important regulatory and housekeeping roles (Waters and Storz, 2009). The majority of sRNAs are non-coding, though some encode small proteins. For example, RNA III, which is a functional sRNA in *Staphylococcus aureus*, was also shown to possess a delta-hemolysin-encoding region (Benito et al., 2000). The first identified chromosomally-encoded sRNA was MicF in *E. coli*, which silences the expression of *ompF* - a porin-encoding gene (Mizuno et al., 1984). A large number of sRNAs have been identified since then, with research efforts stimulated by the advent of RNA-Seq (Mann et al., 2012).

1.2.1 Types of sRNAs

There are several types of regulatory sRNAs, classified based on the method of regulatory function. Small RNAs may associate with complementary nucleic acid sequences in a base-pairing manner either in *cis* or *trans* (see Figure 1.3). Antisense transcripts that target the messenger RNA (mRNA) of genes in the same position but on the opposite strand of DNA are called *cis*-sRNAs, whereas *trans*-sRNAs target the mRNA of genes that are elsewhere in the genome. *Cis*-

sRNAs have been well-studied in mobile genetic elements and were shown to require complete complementarity for effective functioning (Tomizawa et al., 1981). Only partial complementarity is required for *trans*-sRNA activity, allowing them to target multiple mRNAs (Kawamoto et al., 2006). Riboswitches are short transcripts normally present at the 5' untranslated region (UTR) of mRNA, which they regulate in response to the binding of small metabolites and changes in the secondary structure of the mRNA molecule (Mandal et al., 2003). Another type of sRNA is clustered, regularly interspaced short palindromic repeats (CRISPRs). They exhibit a protective, immunity-like role by targeting and degrading complementary bacteriophage DNA (Barrangou et al., 2007).

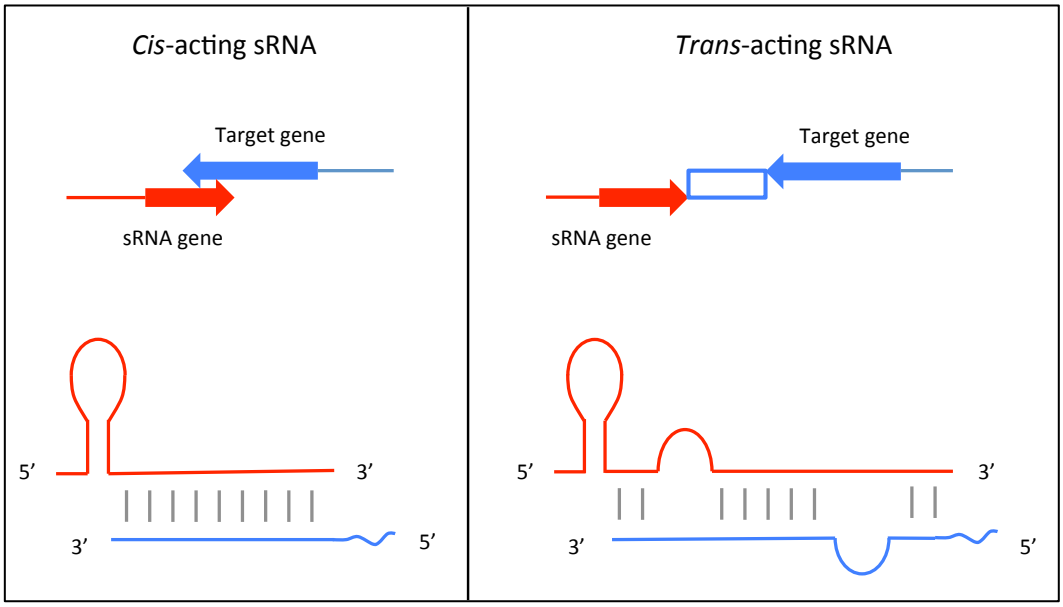

Figure 1.3: The mechanisms of action of *cis*- and *trans*-acting bacterial sRNAs. The *cis*-acting sRNA sequence is present on the opposite strand of the target gene and, once transcribed, exhibits complete complementarity. *Trans*-acting sRNAs are present elsewhere in the genome and exhibit partial complementarity to the target gene.

Binding of base-pairing sRNAs to the ribosomal binding site of mRNA directly interferes with the initiation of the translation process, as was originally proposed for the MicF sRNA in *E. coli* (Mizuno et al., 1984). In some cases, sRNAs positively affect mRNA translation by changing the secondary structure of the mRNA molecule and uncovering the ribosomal binding site, as was shown for RNA III in *S. aureus* (Morfeldt et al., 1995). Alternatively, base-pairing sRNAs can bind in the protein-coding region of the mRNA, leading to its degradation (Pfeiffer et al., 2009). Some sRNAs even associate with proteins and regulate their activity. A good example of this is 6S RNA. It regulates gene expression in *E. coli* via interaction with the sigma-70 holoenzyme of RNA polymerase (Wassarman and Storz, 2000). In addition, some sRNAs are able to sequester proteins through direct binding, as is the case with the CsrB sRNA, which inhibits the effects of the CsrA protein on the gene expression in *E. coli* (Liu et al., 1997).

1.2.2 Different roles of sRNAs

Many studies have been carried out on the roles of sRNAs in bacterial pathogenesis and stress responses. IsrM, an sRNA encoded on a pathogenicity island in *Salmonella enterica*, was identified as a virulence factor, as demonstrated by reduced intracellular replication inside macrophages by IsrM deletion mutants, along with their attenuated killing of mice (Gong et al., 2011). OxyS was one of the first sRNAs shown to be involved in the response to oxidative stress, by regulating the expression of as many as 40 genes in *E. coli* (Altuvia et al., 1997). Another sRNA involved in multiple gene targeting is RyhB, which is a regulator of iron homeostasis in *E. coli* (Masse and Gottesman, 2002). Its targets included genes coding for a superoxide dismutase and two iron-sequestering proteins. MicC is a good example of a single sRNA regulating responses to different stresses, namely changes in temperature and osmolarity, via interaction with just one target mRNA, encoding the OmpC porin in *E. coli* (Chen et al., 2004).

The RNA chaperone Hfq has been studied in many bacteria and was shown to be involved in a variety of processes including virulence and stress response, through facilitating interactions between sRNAs and their targets (Sittka et al., 2007, Kulesus et al., 2008). A recent study showed the importance of Hfq in *H. influenzae* in haem acquisition and the ability to cause disease in animal models (Hempel et al., 2013). Hfq associating with sRNAs in other bacteria suggests it may also do so in *H. influenzae*.

1.3 High-throughput sequencing

The sequencing of the first bacterial whole genome, the Rd strain of *H. influenzae*, was performed using Sanger methodology, which relies on the DNA chain termination approach using dideoxynucleotides (Sanger et al., 1977, Fleischmann et al., 1995). A decade later, the genomic field was revolutionised with the advent of high-throughput sequencing (HTS) technologies, which employ massively parallelised sequencing of millions of fragmented DNA molecules (Loman et al., 2012). These technologies provide a much faster and cheaper approach to sequencing compared to the conventional Sanger method, with completion of the sequencing within hours instead of months. In addition, the development of benchtop sequencers now makes HTS a feasible undertaking in a standard laboratory.

1.3.1 Illumina HTS platform

Illumina is the current leader among HTS platforms, which also include Life Technologies, Pacific Biosciences and Oxford Nanopore Technologies (Reuter et al., 2015). The Illumina platform relies on sequencing by synthesis, using modified deoxynucleotides (dNTP) as reversible terminators (Bentley et al., 2008). In this approach, millions of clonally amplified DNA fragments are attached to a flow cell surface and serve as templates for the synthesis of a complementary DNA (cDNA strand (see Figure 1.4). Polymerisation of cDNA is temporarily terminated after each addition of a fluorescently labelled dNTP.

Fluorescence imaging is used to identify the nucleotide and the polymerisation reaction is then resumed. The resulting short contiguous sequences are called reads and the number of reads per nucleotide is referred to as read depth.

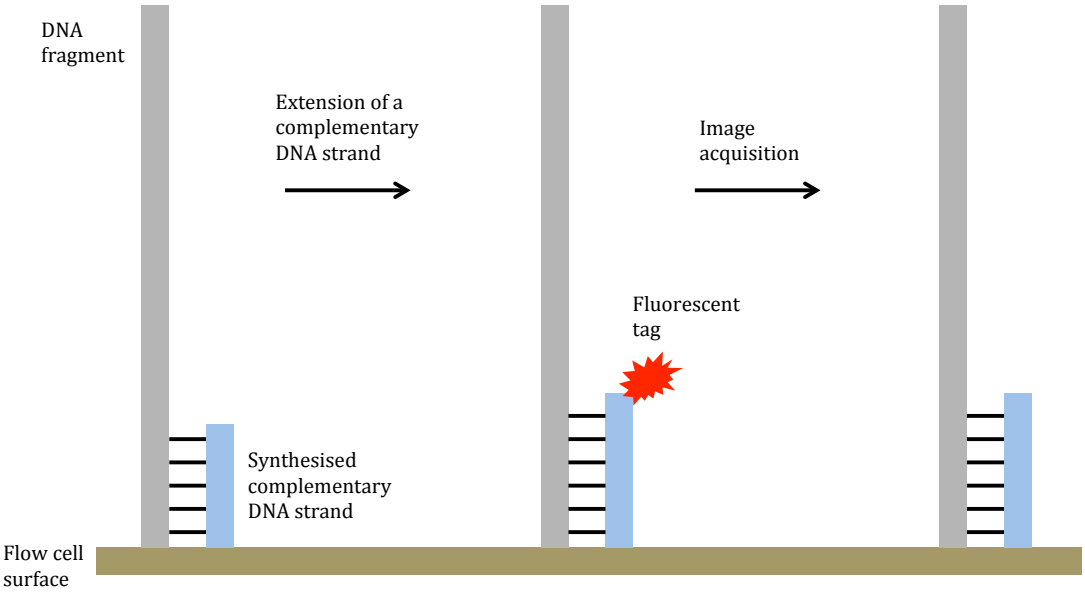

Figure 1.4: A simplified schematic of the Illumina "sequencing by synthesis" methodology.

The Illumina platform is recognised for low error rates, the ability to generate paired-end reads (by sequencing both ends of a DNA fragment) and its user-friendly workflow (Loman et al., 2012). The Illumina HiSeq™ series instruments are used for large-scale sequencing experiments, whereas the Illumina MiSeq™, MiniSeq™ and NextSeq™ are all established benchtop sequencers. The Illumina MiSeq™ currently allows for up to 50 million paired-end reads per single data output, with sequencing run times as short as 4 hours (http://www.illumina.com/systems.html). This makes it a particularly feasible tool for sequencing of prokaryotic genomes, which has already been successfully applied for research in microbiology (Eyre et al., 2012).

1.3.2 RNA-Seq

The study of bacterial gene expression provides the means of identifying genes potentially involved in important processes related to pathogenesis and persistence in the host. The advent of high-throughput RNA sequencing has contributed significantly to gene expression studies. Prior to RNA-Seq, microarrays used to be the technique of choice for the study of large scale gene expression (Hitzemann et al., 2013). They enable gene profiling by the complementary hybridisation of isolated RNA to DNA probes. In recent years there has been a shift in gene expression studies from microarrays towards RNA-Seq and it has been proposed that RNA-Seq should replace microarrays entirely (Hitzemann et al., 2013).

RNA-Seq offers significant advantages over microarrays, including single-nucleotide resolution, lower background noise and the ability to detect low-level and short transcripts. Importantly, no prior knowledge of the gene or reference sequence is required. This means that RNA-Seq can be used to discover novel transcripts, including ncRNAs (Mann et al., 2012). In addition, enrichment of primary transcripts from total RNA using differential RNA-Seq (dRNA-Seq) approach allows for the identification of transcriptional start sites (Sharma et al., 2010). RNA-Seq has already become the technique of choice for

simple differential gene expression studies as well as complex transcriptome analyses and has been successfully applied to study bacterial gene expression (Kumar et al., 2012, Spaniol et al., 2013, Baddal et al., 2015).

1.4 Research aims and objectives

Studying how *H. influenzae* behaves during natural infection is important for understanding the processes involved in its pathogenesis, which can potentially lead to the discovery of novel vaccine and antibiotic targets. Therefore, the main research aim of this project was to use the NGS technology in order to expand the current knowledge of the behaviour of *H. influenzae* under infection-relevant conditions. In addition, as ncRNAs have been largely understudied in *H. influenzae*, this study also aimed to utilise bioinformatic tools in order to identify these potentially important elements. The robust discovery of novel ncRNAs may provide a deeper insight into important regulatory processes during the host infection by *H. influenzae*.

The overall objectives of this study were:

- To identify any nucleotide-level variants between published and re-sequenced whole genomes of Rd and R2866 strains of *H. influenzae.*
- To carry out an in-depth analysis of the accessory genome of Rd and R2866 strains.
- To develop and optimise an *in vitro* invasion assay in order to establish the ability of Rd and R2866 strains from this laboratory to invade human epithelial cells.
- To characterise the transcriptional response of Rd and R2866 strains to infection-relevant conditions on a whole-genome scale, using the RNA-Seq method.
- To develop a robust bioinformatic tool for the discovery of novel putative ncRNA elements from bacterial RNA-Seq data.
- To identify and characterise a repertoire of putative intergenic and antisense ncRNAs in Rd and R2866 strains of *H. influenzae*.

Chapter 2: Materials and Methods

2.1 Media and solutions

Recipes for media and solutions used in this study are presented in Appendix A.

2.2 *H. influenzae*

2.2.1 Strains

Two strains of *H. influenzae* were used in this study - Rd and R2866. Rd is a non-encapsulated standard laboratory strain, first isolated in the mid-1940s (Alexander and Leidy, 1951). R2866 is an invasive non-typeable strain, isolated from the blood of a child with meningitis (Nizet et al., 1996).

2.2.2 Optical density

The optical density (OD) of bacterial broth cultures was measured on a spectrophotometer, using the wavelength of 600 nm. The medium without bacterial cultures was used as a blank.

2.2.3 Growth and storage conditions

H. influenzae was routinely grown on chocolate agar at 37 °C in a 5% CO_2 incubator, for 24-48 hours. A single bacterial colony was picked each time and inoculated in brain-heart infusion (BHI) medium (Oxoid, UK), supplemented with 15 μg/ml haemin (Sigma, UK) and 15 μg/ml NAD (Sigma, UK). Broth cultures were grown overnight in supplemented BHI (sBHI) in 50 ml Falcon tubes at 37 °C with shaking at 200 revolutions per minute (rpm). Overnight broth cultures were diluted to a starting OD_{600} of 0.05 in fresh sBHI. Growth was continued in 125 ml sterile glass flasks or 50 ml Falcon tubes at 37 °C with

shaking at 200 rpm. *H. influenzae* strain stocks were stored at -70 °C in 15% glycerol in sBHI.

2.2.4 Standard growth curve

Standard growth curves were set up with a starting OD_{600} of 0.05. Broth cultures were grown at 37 °C with shaking at 200 rpm and OD_{600} was measured every hour. Growth measurements were stopped when OD_{600} started decreasing.

2.2.5 Growth in medium IV

Broth cultures were grown overnight in sBHI medium, centrifuged at 4,000 rpm for 10 minutes, washed once in chemically-defined medium IV (MIV), centrifuged again and diluted in MIV to OD_{600} of 0.05. Standard growth curve measurements were then carried out.

2.2.6 Oxidative stress

Oxidative stress was induced in *H. influenzae* by growing broth cultures in sBHI until mid-exponential phase and treating them with hydrogen peroxide (Sigma, UK) (Wong et al., 2007). For an RNA sample acquisition, broth cultures at mid-exponential phase were treated with hydrogen peroxide for 10 minutes. Control cultures had an equal volume of sBHI medium added to them.

2.2.7 Iron-starvation stress

Iron-starvation stress was induced in *H. influenzae* by growing broth cultures until mid-exponential phase and treating them with 2,2-bipyridine (VWR International, UK), which is an iron-chelating agent (Harrison et al., 2013). For an RNA sample acquisition, broth cultures at mid-exponential phase were

treated with 2,2-bipyridine for 1 hour. Control cultures had an equal volume of 100% ethanol added to them.

2.3 DNA extraction

DNeasy® Blood & Tissue kit (QIAGEN, UK) was used to extract DNA from Rd and R2866 strains, according to manufacturer's instructions. 1 ml of an overnight broth culture was centrifuged at 5,000 g for 10 minutes and the pellet was resuspended in 180 μl of kit buffer ATL. 20 μl of Proteinase K was added, samples were vortexed and then placed in a thermomixer at 1,000 rpm, 56 °C for 40 minutes. After vortexing samples for 5 seconds, 200 μl of kit buffer AL and 200 μl of 100% ethanol were added. Samples were briefly vortexed, then pipetted onto a DNeasy® Mini spin column and centrifuged at 8,000 rpm for 1 minute. The flow-through was discarded and 500 μl of kit buffer AW1 added. Samples were then centrifuged for 3 minutes at 8,000 rpm, the flow-through discarded, and 500 μl of kit buffer AW2 added. Samples were further centrifuged at 13,500 rpm for 3 minutes to remove all ethanol from the membrane of the spin column. DNA was eluted twice in 100 μl of kit buffer AE by incubating the spin column at room temperature (RT) for 1 minute and centrifuging at 8000 rpm for 1 minute.

2.4 Quality control

2.4.1 Qubit® 2.0 Fluorometer assay

The Qubit® 2.0 Fluorometer assay kit (Life Technologies, UK) was routinely used to quantify DNA and RNA in studied samples, according to manufacturer's instructions. The reagent mix was prepared by diluting Qubit® reagent 1:200 in Qubit® buffer and briefly vortexing. 1-10 μl of RNA or DNA sample was combined with 190-199 μl of the reagent mix, vortexed briefly and incubated for 2 minutes before using the Qubit® Fluorometer. The RNA BR (broad range;

1-1,000 ng/μl) assay kit was used to quantify all RNA samples. The DNA HS (high sensitivity; 0.01-100 ng/μl) assay kit was used to measure concentrations of DNA and cDNA libraries in sequencing experiments. DNA BR (0.1-1,000 ng/μl) kit was used for all other DNA measurements. The fluorometer was calibrated each time using appropriate nucleic acid standards.

2.4.2 Agilent 2100 Bioanalyzer

The Agilent 2100 Bioanalyzer (Agilent, UK) instrument was used to check the quality of RNA (RNA 6000 Pico kit) and DNA (DNA High Sensitivity kit) samples, according to manufacturer's instructions. All kit reagents were equilibrated to RT for 30 minutes before use. Agilent Bioanalyzer chips contain micro-channels that connect sample and reagent wells on a single chip. Bioanalyzer electrodes drive electrophoresis of DNA or RNA molecules, which are separated by size. The data are presented in electropherogram images.

The RNA 6000 Pico gel was prepared by centrifuging 550 μl of the gel matrix in a spin column at 1,500 g for 10 minutes. The dye concentrate was vortexed for 10 seconds and 1 μl was mixed with 65 μl of filtered gel matrix. The gel-dye mix was centrifuged at 13,000 g for 10 minutes before use.

The High Sensitivity DNA dye concentrate was vortexed for 10 seconds and 15 μl was added to the supplied gel matrix. The gel-dye mix was vortexed for 10 seconds and centrifuged at 2,300 g for 10 minutes before use.

The chip was placed on the chip priming station and 9 μl of gel-dye mix was pipetted into an indicated well. The syringe plunger was pushed down for 30 seconds (RNA 6000 Pico kit) or 1 minute (High Sensitivity DNA kit) and then released. 9 μl of gel-dye mix was pipetted into remaining gel wells. When using the RNA 6000 Pico kit, 9 μl of the conditioning solution was pipetted into an indicated well and RNA samples were denatured at 70 °C for 2 minutes prior to loading on the chip. 1 μl of the DNA or RNA sample was combined with 5 μl of the marker in each sample well on the chip. 1 μl of the supplied ladder was

combined with 5 µl of the marker in the ladder well. The chip was vortexed at 2,400 rpm in an IKA vortex mixer for 1 minute and then run on the instrument within 5 minutes.

2.5 Polymerase chain reaction

Standard polymerase chain reaction (PCR) was set up using GoTaq® Green Master mix (Promega, UK). For a 50 µl reaction, the components used were 12.5 µl of Master mix, 1 µl of each PCR primer, 2-3 µl of sample, 7.5-8.5 µl of sterile water. A negative control was included in every PCR reaction. Specific PCR conditions are described in each relevant section.

2.5.1 Agarose gel electrophoresis

Agarose gels (0.8-2%) were prepared by dissolving agarose in Tris-Acetate-Ethylenediaminetetraacetate (TAE) buffer. SYBR® Green I nucleic acid stain (Thermo Fisher Scientific, UK) was diluted 1:10,000 in the gel. 10 µl of the PCR product and appropriate size marker were pipetted into wells on the gel. Gels were run in TAE buffer at 100 V for 45 minutes and visualised under UV light.

2.6 Whole-genome sequencing

2.6.1 Genomic DNA library preparation

The Nextera® XT kit (Illumina, UK) was used to prepare genomic DNA libraries for whole-genome sequencing, according to manufacturer's instructions. 1 ng of DNA in 5 µl volume was used as starting material. DNA samples were mixed with 10 µl of kit buffer TD and 5 µl of kit reagent ATM, centrifuged at 280 g for 1 minute and incubated at 55 °C for 5 minutes. 5 µl of kit buffer NT was added to stop the reaction and samples were vortexed at 280 g for 1 minute. After incubation at RT for 5 minutes, 15 µl of kit reagent NPM and 5 µl of each Index

adapter (i5 and i7) were added. Samples were centrifuged at 280 g for 1 minute and placed on a thermocycler with the following settings:

1. 72 °C for 3 minutes
2. 95 °C for 30 seconds
3. 95 °C for 10 seconds
4. 55 °C for 30 seconds
5. 72 °C for 30 seconds
6. Repeat steps 3-5 11 times
7. 72 °C for 5 minutes

DNA libraries were purified using Agencourt® AMPure® XP magnetic beads (Beckman Coulter, UK), which had been equilibrated to RT for at least 30 minutes. 25 μl of beads were mixed with each sample and incubated at RT for 5 minutes. Tubes were placed on a magnetic rack for 2 minutes and the supernatant was carefully removed. Beads were then washed twice with 200 μl of 80% ethanol for 30 seconds on the magnetic rack. Beads were left to air-dry for 15 minutes before being resuspended in 52.5 μl of Resuspension buffer. Samples were incubated at RT for 2 minutes and placed on the magnetic rack for 2 minutes. 50 μl of supernatant, containing the purified DNA library, was transferred to a new tube. The DNA concentration was measured with the Qubit® DNA HS assay kit (see section 2.4.1). DNA fragment size in each sample was determined using the Bioanalyzer High Sensitivity DNA kit (see section 2.4.2).

2.6.2 Whole-genome sequencing on the Illumina MiSeq™

The molar concentration of each DNA library sample was determined using the formula below:

Molar concentration (M) = Concentration (g/l) / (DNA fragment size (bp) * 650)

5 μl of the pooled library, diluted to 4 nM, was mixed with 5 μl of 0.2 M sodium hydroxide. Samples were briefly vortexed, centrifuged at 300 g for 1 minute and incubated at RT for 5 minutes. 990 μl of the ice-cold hybridisation buffer was added and the library was further diluted in the hybridisation buffer to 10-12 pM. 600 μl of denatured DNA library was loaded onto the MiSeq™ (Illumina, UK). The paired-end 2x250 bp kit was used for the whole-genome sequencing.

2.7 RNA extraction

Prior to any RNA-related work, all materials and designated RNA space were cleaned with 70% ethanol and sprayed with RNaseZap®.

2.7.1 Collection and preservation of bacterial cultures

Bacterial cells at 0.4-0.6 OD were taken for RNA extraction each time. The RNAprotect® reagent (QIAGEN, UK) was used to stabilise and preserve RNA samples for RNA extraction on a separate day. Bacteria were first grown in broth cultures (see sections 2.2.4-2.2.7) and mixed with 2x volume of RNAprotect®. Cultures in RNAprotect® were vortexed immediately for 5 seconds, incubated for 5 minutes at RT and centrifuged for 5 minutes at 5,000 g. The supernatant was then removed and pellets were stored at -70 °C.

For RNA extraction on the same day, bacterial broth cultures were centrifuged at 4,000 rpm for 10 minutes and pellets were resuspended in 1 ml of QIAzol® reagent from miRNeasy Mini kit (QIAGEN, UK). RNA extraction was then carried out using miRNeasy Mini kit per manufacturer's instructions (see section 2.7.2).

2.7.2 miRNeasy Mini kit

RNA extraction was performed with miRNeasy Mini kit (QIAGEN, UK), which enables purification of total RNA, including mRNA and small (<200 nucleotides) RNA transcripts. This method utilised phenol and guanidine thiocyanate (QIAzol® reagent) for sample lysis and a silica membrane-based procedure for RNA extraction.

Following the preservation in RNAprotect®, samples were treated with 100 µl of 1 mg/ml lysozyme in Tris-Ethylenediaminetetraacetate (TE) buffer (pH 8.0) and 20 µl of Proteinase K (QIAGEN, UK) to facilitate the breakdown of the bacterial cell wall. Samples were vortexed for 5 seconds and incubated at RT for 5 minutes, vortexing for 10 seconds every 2 minutes. 1 ml of QIAzol reagent was added and samples were vortexed for 3 minutes before being incubated at RT for 5 minutes. 200 µl of chloroform was added, tubes were shaken vigorously for 15 seconds and incubated at RT for 3 minutes. Samples were centrifuged at 12,000 g for 15 minutes at 4 °C and the resulting upper aqueous phase was transferred to RNase-free 2 ml tubes. 1.5x volume of 100% ethanol was added and the solution mixed by pipetting.

700 µl of the sample was added to the spin column, which was then centrifuged for 15 seconds at 10,000 rpm, with the flow-through discarded. This step was repeated for the remainder of the sample. 80 µl of DNase I (diluted 1:8 in RDD buffer) was pipetted directly onto the top of the silica membrane and left for 15 minutes at RT. 350 µl of Buffer RWT was added to the column and samples were centrifuged at 10,000 rpm for 15 seconds, with the flow-through discarded. DNase I treatment was repeated again, to ensure the efficient degradation of genomic DNA, which would otherwise interfere with sensitive downstream applications, RNA-Seq experiments in particular.

500 µl of RPE buffer was added to the spin column and samples were centrifuged for 15 seconds at 10,000 rpm, with the flow-through discarded. 500 µl of RPE buffer was added again and samples were centrifuged for 2 minutes at

10,000 rpm. The spin column was centrifuged for 1 minute at full speed to remove any residual ethanol. RNA was subsequently eluted with 20-40 μl RNase-free water by centrifuging at 10,000 rpm for 1 minute. The elution step was performed twice to increase the overall RNA yield. The total volume of eluted RNA was 40-60 μl and it was stored at -70 °C.

2.7.3 PCR to check for the genomic DNA contamination

RNA samples for RNA-Seq experiments were checked for genomic DNA contamination using standard PCR of the 16S rRNA gene. Genomic Rd DNA was used as a positive control. PCR primers for 16S rRNA were:
AGAGTTTGATCMTGGCTCAG (forward)
CGGTTACCTTGTTACGACTT (reverse).

PCR conditions were:
1. Initial denaturation at 94 °C for 3 minutes
2. Denaturation at 94 °C for 30 seconds
3. Annealing at 55 °C for 30 seconds
4. Extension at 68 °C for 1 minute
5. Repeat steps 2-4 29 times
6. Final extension at 68 °C for 5 minutes

Absence of detectable 16S rRNA PCR products was used to signify lack of genomic DNA contamination.

2.8 RNA-Seq

2.8.1 Depletion of rRNA

The Ribo-Zero™ rRNA Removal kit for bacteria (Illumina, UK) was used to remove most rRNA from total bacterial RNA, according to manufacturer's instructions. Ribo-Zero™ bead preparation was as follows. 225 μl of beads were equilibrated at RT for 30 minutes, placed on the magnetic rack for 1 minute and

the supernatant removed. Beads were washed twice with 225 μl of RNase-free water by placing them for 1 minute on the magnetic stand and removing the supernatant. Magnetic beads were then resuspended in 65 μl of Resuspension solution and 1 μl of RiboGuard RNase inhibitor (100 U/μl) was added.

1-5 μg of total extracted RNA was used and samples were topped up with RNase-free water to 28 μl. 4 μl of Ribo-Zero™ Reaction buffer and 8 μl of Ribo-Zero™ rRNA Removal solution were added to each sample. The mixture was incubated at 68 °C for 10 minutes and then cooled for 5 minutes at RT. The solution was mixed well with pre-prepared Ribo-Zero™ beads by repeated pipetting and vortexing for 10 seconds. Tubes were incubated for 5 minutes at RT and vortexed again at 50°C for 5 minutes. Tubes were then placed on the magnetic rack and the supernatant, containing rRNA-free RNA, was transferred to a new RNase-free tube.

Ethanol precipitation was used to purify RNA. RNA samples were topped up to 180 μl with RNase-free water. 18 μl of 3 M sodium acetate, 2 μl of glycogen (10 mg/ml) and 600 μl of ice-cold 100% ethanol were added to each sample. Tubes were vortexed and incubated at -20 °C for at least 1 hour before centrifuging at full speed for 30 minutes. RNA pellets were washed twice with 350 μl of ice-cold 70% ethanol and centrifuged at 14,000 rpm for 5 minutes each time. After air-drying the pellets, they were resuspended in 5.5 μl of RNase-free water. The efficiency of rRNA depletion was checked using the Bioanalyzer RNA 6000 Pico kit (see section 2.4.2).

2.8.2 Generation of the cDNA library

A modified TruSeq™ Stranded mRNA sample preparation protocol was used to prepare cDNA libraries for sequencing. All steps were carried out in sealable sterile 96-well plates.

5 µl of rRNA-depleted RNA was mixed with 13 µl of TruSeq™ Fragment, Prime, Finish mix and incubated at 94 °C for 8 minutes. The 96-well plate was briefly centrifuged at 280 g, incubated for 5 minutes at RT and 17 µl of the sample transferred to a new well on the same plate. 0.1x volume of Superscript II (Fisher, UK) was mixed with 0.9x volume of TruSeq™ First Strand Synthesis Act D. 8 µl of this mix was added to each sample and the plate was incubated at 25 °C for 10 minutes, 42 °C for 15 minutes and 70 °C for 15 minutes.

5 µl of TruSeq™ End Repair Control, diluted 1:50 in Resuspension buffer, and 20 µl of TruSeq™ Second Strand Marking Master mix were added to each sample and incubated at 16 °C for 1 hour. A purification step was performed by adding 90 µl of Agencourt® AMPure® XP magnetic beads, incubating the plate at RT for 15 minutes and then placing it on a magnetic rack for 5 minutes at RT. The supernatant was discarded and the beads were washed twice with 200 µl of 80% ethanol, by incubating for 30 seconds on the magnetic rack and discarding the supernatant. Beads were air-dried for 15 minutes and 17.5 µl of Resuspension buffer was added to each sample. The plate was incubated at RT for 2 minutes, placed on the magnetic rack for 5 minutes and 15 µl of the supernatant, containing double-stranded cDNA, was transferred to a new well.

2.5 µl of TruSeq™ A-Tailing Control, diluted 1:100 in Resuspension buffer, and 12.5 µl of TruSeq™ A-Tailing mix were added to each sample. The plate was incubated at 37 °C for 30 minutes and 70 °C for 5 minutes. 2.5 µl of TruSeq™ Ligation Control, diluted 1:100 in Resuspension buffer, 2.5 µl of TruSeq™ Ligation mix and 2.5 µl of Index adapter were added to each sample. The plate was centrifuged at 280 g for 1 minute and then incubated at 30 °C for 1 hour. 5 µl of TruSeq™ Stop Ligation buffer was added to stop the reaction.

The second purification step with 42 µl of Agencourt® AMPure® XP beads was carried out as described previously: beads were washed with 80% ethanol, resuspended in 52.5 µl of Resuspension buffer and 50 µl was transferred to a

new well. The third purification step with 50 μl of Agencourt® AMPure® XP beads was performed immediately after: beads were washed with 80% ethanol, resuspended in 22.5 μl of Resuspension buffer and 20 μl was transferred to a new well.

5 μl of PCR Primer Cocktail and 25 μl of PCR Master mix were added to each well and the following PCR reaction was set up:

1. Initial denaturation at 98 °C for 30 seconds
2. Denaturation at 98 °C for 10 seconds
3. Annealing at 60 °C for 30 seconds
4. Extension at 72 °C for 30 seconds
5. Repeat steps 2-4 14 more times
6. Final extension at 72 °C for 5 minutes

The final purification step was performed with 50 μl of Agencourt® AMPure® XP beads as described previously: beads were washed with 80% ethanol and resuspended in 32.5 μl of Resuspension buffer. 30 μl of Resuspension buffer, containing completed cDNA library, was then transferred to a new well. DNA concentration was measured using the Qubit® High Sensitivity DNA assay kit (see section 2.4.1). DNA fragment size of each library was determined using the Bioanalyzer High Sensitivity DNA kit (see section 2.4.2).

2.8.3 RNA-Seq on the Illumina MiSeq™

Libraries of cDNA were pooled and prepared for sequencing as previously described (see section 2.6.2). The paired-end 2x75 bp kit was used for RNA-Seq on the Illumina MiSeq™.

2.9 Northern blot

2.9.1 Primer design and PCR

Primers (Sigma, UK) for producing RNA probes were designed using Primer3 online tool (see Table 2.1) (Koressaar and Remm, 2007, Untergasser et al., 2012). The reverse primer for each probe had a T7 polymerase promoter sequence at the 5' end.

Table 2.1: List of primers used to design RNA probes for northern blotting.

T7 polymerase promoter sequence is depicted in bold.

Target	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
R2866_101	ccttagttggtttaggttgct	**ctaatacgactcactatagggaga**ctagataagcggcttttatatg
R2866_118	ggaagacaggattggtctc	**ctaatacgactcactatagggaga**gtggtggaactaagcagaatt

Standard GoTaq® (Promega, UK) PCR was set up as a 50 µl reaction, using *H. influenzae* strain R2866 genomic DNA as the template. The PCR conditions were:

1. Initial denaturation at 95 °C
2. Denaturation at 95 °C for 30 seconds
3. Annealing at 51 °C for 30 seconds
4. Extension at 72 °C for 15 seconds
5. Repeat steps 2-4 34 times
6. Final extension at 72 °C for 5 minutes

PCR products were checked on 2% agarose gel (see section 2.5.1).

2.9.2 PCR purification

PCR products were purified with the illustra™ GFX™ PCR DNA and Gel Band Purification kit (GE Healthcare, UK). 500 µl of kit Capture buffer type 3 was mixed with 40 µl of PCR product, transferred to a spin column, centrifuged at 16,000 g for 30 seconds and the flow-through discarded. The centrifugation step was repeated by adding 500 µl of kit Wash buffer type 1. The spin column was centrifuged at 16,000 g for 30 seconds to remove any residual ethanol. DNA was eluted in 17.5 µl of kit Elution buffer type 4 (10 mM Tris, pH 8.0) by centrifuging at 16,000 g for 1 minute. The elution step was repeated for a total of 35 µl of purified PCR product.

2.9.3 Generating the biotin-labelled RNA probe

The purified PCR product was used as a template for making the RNA probe. First, the nucleoside triphosphate (NTP) mix was prepared by mixing 2 µl of adenosine triphosphate (ATP), 2 µl of guanosine triphosphate (GTP), 2 µl of cytidine triphosphate (CTP), 0.5 µl of uridine triphosphate (UTP) (100 mM stock

each) and 15 µl of biotin-labelled UTP (10 mM stock). Subsequently, RNA transcription was set up by mixing 12 µl of the DNA template (previously purified PCR product), 2 µl of NTP mix, 4 µl of 5x transcription buffer, 2 µl of dithiothreitol (100 mM stock), 2 µl of T7 RNA polymerase and 1 µl of RNasin®. All reagents were purchased from Promega, UK. The reaction was incubated for 2 hours at 37 °C in a water bath. 2 µl of 0.2 M ethylenediaminetetraacetate (EDTA) was added to stop transcription.

RNA was precipitated by adding 4.5 µl of 4 M lithium chloride and 75 µl of 100% ethanol and incubating for 2 hours at -20 °C. The precipitated RNA probe was centrifuged for 10 minutes at 12,000 g and the RNA pellet was washed with 200 µl of 70% ethanol by centrifuging at 12,000 g for 10 minutes. The RNA pellet was dissolved in 100 µl of RNAse-free water and 1 µl of RNasin® was added. RNA probes were stored at -70 °C.

2.9.4 Denaturing RNA gel

A 2% denaturing RNA gel was prepared by dissolving agarose in MOPS (3-(*N*-morpholino)propanesulfonic acid) buffer and tempering at 56 °C for 30 minutes. 17.5 ml of 34% formaldehyde solution was added to 83 ml of the denaturing RNA gel and mixed thoroughly, before pouring into the gel cast.

10 µg of each RNA sample and biotin-labelled RNA ladder was run on the denaturing RNA gel. The volumes of RNA samples and RNA ladder were adjusted to 20 µl with RNAse-free water, mixed with 10 µl of RNA sample loading buffer, denatured at 70 °C for 5 minutes and immediately put on ice. 20 µl from each sample was carefully loaded into a well of the denaturing RNA gel. The gel was run for 2 hours and 15 minutes at 100 V in MOPS buffer to ensure sufficient separation of RNA size markers.

The gel was stained in a Diamond™ Nucleic Acid dye (Promega, UK), diluted

1:10,000 in MOPS buffer, in the dark on a rocking platform at 50 rpm for 30 minutes. The quality of RNA was determined by the presence of 16S and 23S rRNA bands, when viewed under UV light.

2.9.5 Capillary blotting

Capillary blotting apparatus was assembled by placing the lid of a medium-sized box at a 90° angle on top of the box filled with Transfer buffer. A long strip of standard filter paper was cut and soaked in Transfer buffer, before being placed on top of the lid, with the ends dipped in Transfer buffer. A 25 ml sterile pipette was used to roll across the filter paper to make it flat. The RNA gel was cut with a scalpel to the desired size and placed on top of the filter paper on the lid. A Hybond®-N+ nitrocellulose membrane (GE Healthcare, UK) was cut slightly larger than the gel and floated on the surface of RNase-free water for 5 minutes. The membrane was submerged in RNase-free water and then soaked for 5 minutes in Transfer buffer, before being placed on top of the gel. Another 25 ml sterile pipette was rolled across to exclude air bubbles. Two pieces of filter paper were cut and placed on top of the membrane, rolling across with a 25 ml sterile pipette each time. Strips of Parafilm M® were used to seal the area around the gel to prevent the bypassing of the capillary action. A large stack of paper tissues, with a heavy weight on top, was placed on the filter paper. This ensured that RNA would transfer from the gel onto the nitrocellulose membrane due to capillary action. The capillary blotting apparatus was incubated overnight in order to allow complete RNA transfer.

2.9.6 Hybridisation

Following an overnight incubation, the blotting apparatus was disassembled and the nitrocellulose membrane removed from the top of the gel. RNA was fixed to the membrane by UV crosslinking at 120,000 mJ for 2 minutes and then soaked for 10 minutes in Neutralisation solution. Subsequently, the membrane was placed inside a hybridisation tube, with the RNA side facing inwards. 20 ml

of pre-warmed pre-hybridisation solution was added to the hybridisation tube, which was then placed on rotation in a hybridisation oven for 1 hour at 58 °C (20 °C below the melting temperature of the probe). In the meantime, 900 ng - 1 μg of the RNA probe was mixed with 10 ml of pre-warmed pre-hybridisation solution, denatured at 65 °C for 15 minutes in a water bath and then immediately placed on ice. After a one-hour incubation at 58 °C, the pre-hybridisation solution was discarded from the hybridisation tube. 10 ml of the denatured RNA probe was added to the hybridisation tube with the nitrocellulose membrane, which was then placed on rotation in a hybridisation oven overnight at 58 °C.

2.9.7 Detection of biotin-labelled RNA

All wash and detection steps were carried out with the nitrocellulose membrane placed inside a black Incubation box (LI-COR, UK). The membrane was washed twice with 50 ml of Wash solution I for 5 minutes. It was then washed three times for 15 minutes with 50 ml of Wash solution II at 58 °C in a water bath. Subsequently, the membrane was washed with 30 ml of Wash buffer for 5 minutes on a rocking platform at 100 rpm and incubated in 15 ml of Odyssey® Blocking buffer (LI-COR, UK) for 30 minutes on a rocking platform at 50 rpm. 1.5 μl of IRDye® 800CW Streptavidin (LI-COR, UK) was added directly to the blot submerged in the Odyssey® blocking buffer. The membrane was left for a further 30 minutes at 50 rpm. Finally, the membrane was washed twice with 30 ml of Wash buffer for 30 minutes. Northern blots were visualized using IR740 Module lighting and LY800 filter, with a ten-minute exposure.

2.10 Cell culture

2.10.1 Maintenance

The human alveolar epithelial cell line A549 was kindly provided by Dr Meera Unnikrishnan. Cells were routinely grown in T75 flasks in a 5% CO_2 incubator at

37 °C. The cell culture medium used was Dulbecco's Modified Eagle medium (DMEM), supplemented with 10% fetal bovine serum (FBS) and 1 U/ml penicillin/streptomycin. A549 stocks were stored at -70 °C in the freezing medium: 50% DMEM, 40% FBS, 10% dimethyl sulphoxide (DMSO).

2.10.2 Passage

Cells were passaged upon reaching 90%-100% confluence. Briefly, cells were washed twice with warm phosphate-buffered saline (PBS) and detached from the surface of the flask by incubating in 1.5-2 ml of trypsin/EDTA for 4-5 minutes at 37 °C. 10 ml of warm cell culture medium was then added to the cells and they were centrifuged at 1,000 rpm for 5 minutes. The supernatant was removed and cells were resuspended in warm cell culture medium. If required, A549 cells were quantified by pipetting 10 μl of the cell suspension onto the haemocytometer and counting viable cells under the light microscope.

2.10.3 Infection with *H. influenzae*

A549 cells in the media without antibiotics were seeded at the concentration of $4x10^5$ cells/ml in 6-well or 24-well plates and grown overnight to a confluence of 80-90%. Bacterial cultures were grown to mid-exponential phase, centrifuged at 7,000 rpm for 7 minutes, washed in warm PBS and centrifuged again at 7,000 rpm for 7 minutes. Bacterial cells were resuspended in infection medium without antibiotics and were used to infect A549 cells at a set multiplicity of infection (MOI).

2.10.4 Invasion of A549 cells

H. influenzae invasion was studied by infecting A549 cells for a set amount of time and quantifying intracellular bacterial numbers. Optimisation of infection time is described in section 4.2.1.1. Extracellular bacteria were killed by treating the infected monolayer with infection medium containing 200 μg/ml

gentamicin for 1 hour. Cells were then washed several times with warm PBS and lysed either with 0.025% saponin (VWR International, UK) in PBS for 10 minutes at 37 °C or with 0.5% Triton™ X-100 in PBS for 10 minutes at 37 °C. Viable cell counts were performed as described below (see section 2.10.5).

2.10.5 Growth of *H. influenzae* in the infection medium

The effect of the FBS on the viability of *H. influenzae* in the infection medium was tested as follows. Bacterial broth cultures were grown in sBHI to mid-exponential phase, centrifuged at 7,000 rpm for 7 minutes, washed in warm PBS and centrifuged again at 7,000 rpm for 7 minutes. The supernatant was removed and bacteria were resuspended in warm infection medium without antibiotics. The infection medium was either without FBS or with 2%, 5% and 10% FBS. Bacterial cultures were diluted to OD_{600} of 0.1 and growth was measured at 1.5 and 3 hours. For viable cell counts, 100 µl of bacterial cultures were serially diluted, pipetted onto sBHI agar plates and distributed over the agar surface using sterile colony spreaders. Agar plates were kept in a 5% CO_2 incubator at 37 °C and single *H. influenzae* colonies were counted the following day.

2.10.6 Fluorescence staining of infected cells

For fluorescence staining, A549 cells were seeded at the concentration of $4x10^5$ cells/ml in a sterile 4-well Nunc™ Lab-Tek™ II Chamber Slide™ (Thermo Fisher Scientific, UK). 1 hour and 30 minutes before infection with *H. influenzae*, A549 cells were stained for 45 minutes with 5 µg/ml FM® 4-64 lipophilic membrane dye, diluted in the cell culture medium (Vida and Emr, 1995). Subsequently, cells were washed twice in warm PBS and grown for another 45 minutes, before proceeding with *H. influenzae* infection as described previously (see section 2.10.4). Cells were fixed with 1 ml of 4% paraformaldehyde (PFA) for 10 minutes at RT, washed twice in PBS and stained for 5 minutes with the nuclear counterstain, 4',6-diamidino-2-phenylindole (DAPI), diluted in PBS to a concentration of 250 ng/ml. The chamber and the gasket were removed from

the Chamber Slide™ and images of fluorescent cells were taken on a Leica DMi8 inverted fluorescence microscope at 100x magnification with oil immersion.

2.11 RNA extraction from infected human cells

The infected monolayer of A549 cells was directly treated with 1 ml of QIAzol® reagent and RNA was subsequently extracted using the miRNeasy Mini kit as described previously (see section 2.7.2).

2.11.1 Enrichment of bacterial RNA

Microb*Enrich*™ kit (Thermo Fisher Scientific, UK) was used to deplete eukaryotic RNA from a mixed sample containing bacterial and human RNA, according to manufacturer's instructions. Microb*Enrich*™ Oligo MagBeads were placed in an RNase-free tube on a magnetic rack for 3 minutes and the supernatant was removed. Beads were then washed once in RNase-free water and once in Microb*Enrich*™ Binding buffer by incubating them on the magnetic rack for 3 minutes and removing the supernatant. Beads were stored on ice until required.

27 μg of total RNA in 30 μl was combined with 300 μl of Microb*Enrich*™ Binding buffer and 12 μl of Microb*Enrich*™ Capture Oligo mix. RNA was then denatured at 70 °C for 10 minutes and the samples were incubated at 37 °C for 1 hour. They were then mixed with Microb*Enrich*™ Oligo MagBeads and incubated at 37 °C for 15 minutes. Tubes were placed on the magnetic rack for 3 minutes and the supernatant, containing enriched RNA, transferred to a new tube. The beads were resuspended in 100 μl of pre-warmed Microb*Enrich*™ Wash solution, incubated at 37 °C for 5 minutes and placed again on the magnetic rack for 3 minutes. The supernatant, containing any remaining enriched RNA, was pooled with the previously kept supernatant.

RNA was precipitated for 1 hour at -20 °C with 0.1x volume of 3 M sodium acetate, 4 μl glycogen (5 mg/ml) and 2.5x volume of ice-cold 100% ethanol. Samples were centrifuged at 13,000 rpm for 30 minutes and the supernatant discarded. RNA pellets were washed twice with 750 μl of 70% ethanol by centrifuging at 13,000 rpm for 5 minutes. Tubes were briefly centrifuged and any remaining supernatant was carefully removed. RNA pellets were air-dried for 5 minutes and resuspended in 50 μl of RNase-free water.

2.12 Bioinformatic data analysis

2.12.1 *H. influenzae* genome sequences

Whole-genome reference sequences of Rd and R2866 strains were available from the NCBI database (http://www.ncbi.nlm.nih.gov). Accession numbers were NC_000907 for Rd and CP002277 for R2866.

2.12.2 Whole-genome assembly

SPAdes software was used to assemble sequencing reads into joined contiguous sequences (contigs) (Bankevich et al., 2012). "Careful" mode was selected to reduce the number of mismatches as well as short insertions and deletions (indels). QUAST, included in SPAdes software, was used to assess the whole-genome assembly properties (Gurevich et al., 2013). Contigs were removed if they were shorter than 200 bp and the read coverage was lower than 10x. Mauve was used to align contigs to the appropriate reference genome sequence from the NCBI database (Darling et al., 2004). The Mauve Contig Mover module was subsequently used to reorder contigs based on Rd or R2866 reference genome (see section 2.12.1) (Rissman et al., 2009). Ordered contigs were concatenated into one complete sequence with the EMBOSS union online tool (http://www.bioinformatics.nl/cgi-bin/emboss/union). Qualimap was used to determine read coverage of each assembled genome after mapping sequencing

reads against the assembled genome (see section 2.12.5) (Garcia-Alcalde et al., 2012).

2.12.3 Whole-genome annotation

Prokka was used to annotate sequenced whole genomes of *H. influenzae* Rd and R2866 strains (Seemann, 2014). It was important to retain the original annotation of genome sequences of these strains. Hence, the makeblastdb (part of BLAST+ package) command-line tool was used to create a genus database from the reference genome sequences (Camacho et al., 2009). The genus database was used during Prokka annotation of sequenced genomes.

2.12.4 Sequence comparison and visualisation

Whole-genome and RNA-Seq data were visualised in the Artemis genome browser (Rutherford et al., 2000). The Artemis Comparison Tool (ACT) was used to compare the genomes of Rd and R2866 strains (Carver et al., 2005). For this purpose, comparison files were generated with an online tool WebACT (http://www.webact.org/WebACT/home) using the BLASTn algorithm with default parameters. The average nucleotide identity (ANI) was calculated using best hit and reciprocal best hit methods (http://enve-omics.ce.gatech.edu/ani/) (Goris et al., 2007).

2.12.5 Mapping and processing sequencing reads

Paired-end reads from RNA-Seq experiments were in the opposite orientation: the first read was reverse (3'-5') and the second read was forward (5'-3'). In order to visualize mapped RNA-Seq reads in Artemis, they needed to be of the same orientation. Hence, the first read was reverse complemented using the seqtk command-line tool, so that both reads were in the forward orientation. This was not required for whole-genome sequencing reads.

The reference genome was indexed using bowtie2-build command (Langmead and Salzberg, 2012). Sequencing reads were mapped to the reference genome using bowtie2 software (Langmead and Salzberg, 2012). Read alignment data was generated in SAM (sequence alignment/map) file format. SAMtools was used to convert alignment data to BAM (binary alignment/map) file format, which is a binary version of SAM file format (Li et al., 2009). SAMtools was subsequently used to sort and index BAM files.

2.12.6 Genome variant calling

The SAMtools command "mpileup" was used to generate a pileup format file from a sorted BAM file and a FASTA file of the reference genome (Li et al., 2009). This was used as input for VarScan2 software, which identifies single nucleotide polymorphisms (SNP) and indels present between two genome sequences (Koboldt et al., 2012). The minimum read coverage was set to 20. The minimum number of reads needed to support SNP or an indel was chosen as 15. The minimum quality for a bp was set to 30. The minimum allele frequency threshold was 0.9. Finally, the minimum allele frequency to be called a homozygote was set to 0.9.

2.12.7 Differential gene expression analysis

The R package DESeq2 uses a negative binomial distribution model to test for the differential expression in RNA-Seq data (Love et al., 2014). Sorted BAM and GFF (general feature format) files were used as input for the coverageBed tool, outputting a text file with read coverage information for every feature in the genome. These text files, one per biological replicate, were used as input for DESeq2. P-values were adjusted for a false discovery rate at 5% using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Data were further filtered by applying a standard cut-off of 2 for the fold change and 0.05 for adjusted p-value (Baddal et al., 2015).

2.12.8 Analysis of enriched functional groups

DAVID (Database for Annotation, Visualization, and Integrated Discovery) was used to identify gene ontology (GO) terms and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways that were enriched in lists of differentially expressed genes (Huang da et al., 2009a, Huang da et al., 2009b). Reference Sequence (RefSeq) protein identifiers for every gene from a list were used as input. KEGG pathway diagrams were generated using KEGG Mapper (http://www.kegg.jp/kegg/tool/map_pathway2.html).

2.12.9 TPM normalisation

For absolute expression analysis, RNA-Seq data was manually normalised using the Transcripts per Million (TPM) method (Wagner et al., 2012).

2.12.10 BLAST

All BLAST searches were performed online on the BLAST server (http://blast.ncbi.nlm.nih.gov) or using the BLAST+ package on the command line (Camacho et al., 2009). Homology search of ncRNAs was carried out using the E-value cut-off of 1e-05.

2.12.11 Identification of ncRNAs

Sorted BAM files and a GFF file, containing coordinates of the coding sequences, were used as input for coverageBed and genomeCoverageBed command-line tools, which are both part of the BEDTools suite (Quinlan and Hall, 2010). CoverageBed was used to produce read coverage information for each nucleotide that is present in every coding sequence in a genome. GenomeCoverageBed was used to produce read coverage information for each nucleotide in the genome: on both strands and for each strand separately. These files were used as input for a Python script, which was written in-house to

identify ncRNA sequences from RNA-Seq data. See Chapter 5 for a detailed description of the script.

2.12.12 RNA and protein family analysis

Protein domain and family analysis was carried out using the InterPro database (Mitchell et al., 2015). The Rfam database was used to identify homologues from known RNA families (Griffiths-Jones et al., 2003, Nawrocki et al., 2015).

2.12.13 RNA secondary structure and gene targets

Secondary RNA structure was predicted using the RNAfold web server (Hofacker and Stadler, 2006). Homologues of ncRNAs were identified using the GLASSgo online tool, using the "very high specificity" option (http://rna.informatik.uni-freiburg.de). Five homologous sequences were then used to predict potential gene targets using CopraRNA (Wright et al., 2013, Wright et al., 2014). Potential target sequences were analysed 75 bp around the start codon of each gene.

2.12.14 Figure generation and statistical analysis

Microsoft Excel was used to produce simple graphs of numeric data. False colour heatmaps were generated in R using the "heatmap.2" function of the "gplots" package. The Circos tool was used to visualize the genomic data in a circularized layout (Krzywinski et al., 2009). Venn diagrams were generated with the online tool Venny (http://bioinfogp.cnb.csic.es/tools/venny/). Image analysis was performed using the Fiji image processing package (Schindelin et al., 2012).

Chapter 3: Whole-genome analysis and comparison of Rd and R2866 strains of *H. influenzae*

3.1 Introduction

The accumulation of small genetic variations in descendants of the original bacterial strain can result in phenotypic differences between the same strains present in different laboratories, as described in section 1.1.6. The use of original whole genome sequences of *H. influenzae* strains to infer the behaviour of other descendants of the same strain could therefore introduce potential inaccuracies in generated data. Hence one of the major aims of this chapter was to re-sequence whole genomes of Rd and R2866 strains of *H. influenzae* and to explore any potential nucleotide-level variants between re-sequenced and original published whole genome sequences. This was also relevant for subsequent RNA-Seq experiments, where an accurate reference genome was required for mapping sequencing reads, in order to correctly deduce the transcriptional landscape of *H. influenzae* during infection-relevant conditions (see Chapter 4).

Genome annotation methods are being constantly improved, due to development of superior gene prediction algorithms as well as new published functional data (Stothard and Wishart, 2006). Therefore, it is advantageous to use the most recent genome annotations, to ensure up-to-date information about protein-coding sequences. Still, automated genome annotation tools also pose some caveats, including inconsistent annotation of homologous regions and abundance of hypothetical proteins (Richardson and Watson, 2013). In addition, sequencing errors could result in mis-annotation of a subset of genes. Therefore, whole-genome re-sequencing and re-annotation of Rd and R2866 strains in this study was used to infer any previous annotation issues. Furthermore, an up-to-date genome re-annotation also ensured inclusion of the

latest bacterial genome features, which was again relevant for subsequent differential gene expression analysis (see Chapter 4).

The availability of whole-genome sequences for different bacterial strains enables characterisation of their important genetic features, such as core and accessory genomes (see section 1.1.5.2.1). The accessory genome in particular can help to explain strain-specific phenotypes, as it may contain genes required for survival in certain environments. Thus the knowledge of differential presence of accessory genes is important in order to understand different behaviours of bacterial strains under a variety of growth conditions and stresses. The second major aim of this chapter was therefore to use whole-genome sequencing to characterise the accessory genome of Rd and R2866 strains and to use that to infer possible phenotypic differences between these strains. In addition, the description of the accessory genome was again relevant for subsequent RNA-Seq experiments, where accessory genes were likely to help to explain some of the differences in the response of Rd and R2866 to infection-relevant conditions (see Chapter 4).

3.2 Results

3.2.1 Whole-genome re-sequencing of Rd and R2866 strains

Assembly properties of re-sequenced whole genomes of Rd and R2866 strains are displayed in Table 3.1. N50 is a standard statistical measure of the size and quality of the assembly, where at least half of all nucleotides in the assembled genome belong to the contig of size N50 or larger. High N50 values are indicative of a good assembly quality, which was true for both Rd and R2866 strains (>100,000). Filtering assembled contigs by their length (>200 bp) and coverage (>10x) reduced the total number of contigs from 125 to 26 for Rd and from 147 to 29 for R2866. Filtered contigs were ordered based on the original published sequence and then concatenated into one contiguous sequence,

which was used for all subsequent analyses. Both strains had high genome coverage values, with 411x for the R2866 strain and 509x for the Rd strain. In addition, prior to calling nucleotide-level genome variants, reads of re-sequenced genomes were mapped to original whole genomes. Mapping properties are described in Table 3.1 as well.

Table 3.1: Whole-genome assembly and mapping properties for re-sequenced Rd and R2866 strains.

Assembly and mapping features	Rd	R2866
Total number of reads	3,876,150	3,334,598
Total number of contigs	125	147
Largest contig length (bp)	513,049	629,744
N50	128,059	362,305
Number of filtered contigs (>200 bp length; >10x coverage)	26	29
Genome size (bp)	1,798,888	1,909,348
GC content (%)	38.0	38.0
Genome coverage	509x	411x
Percentage of reads mapped to original reference genome (%)	99.37	98.79

3.2.2 Comparison of original and re-sequenced whole genomes of Rd and R2866

3.2.2.1 Identification of SNPs

SNPs are a form of nucleotide-level genetic variation, each one representing a difference in a single nucleotide between two DNA sequences. In this study, SNPs were identified between the original and re-sequenced Rd and R2866 whole genomes. The original genome was used as a reference for detecting SNPs, while the re-sequenced genome was a query.

While no SNPs were discovered for the R2866 strain, 122 SNPs were originally identified for Rd. However, 37 of these were called SNPs by the software due to the presence of an "N" nucleotide in the original published sequence, which meant that it could be any of the four nucleotides. Therefore, these were not true SNPs and were subsequently filtered out. In addition, there were ambiguous nucleotides present in the reference genome, representing any of the two possible nucleotides. If a query nucleotide was part of the possible pair of the ambiguous nucleotide in a reference, it was further filtered out. The final number of SNPs was 41, of which 14 were present in intergenic regions.

A list of 27 SNPs identified only in the coding sequences is displayed in Table 3.2. The majority of SNPs were non-synonymous, meaning that there was a change in the amino acid sequence of a protein. Only three SNPs were definitely synonymous (no changes in the amino acid sequence), while three more SNPs were undetermined, due to the ambiguity of the reference nucleotide. Most genes contained only one SNP, with the exception of HI0635, coding for a hemoglobin-binding protein, which contained three non-synonymous SNPs, as well as gene *oppA*, coding for an oligopeptide ABC transporter substrate-binding protein, with two non-synonymous SNPs. There were six SNPs present in genes coding for iron-associated proteins. SNPs were also located in a *mutS* gene and a *vacB* gene, which are associated with increased mutation rates and virulence respectively (Tobe et al., 1992, LeClerc et al., 1996).

Table 3.2: Strand-specific SNPs present in coding sequences of the Rd strain. The original (reference; Ref) and re-sequenced (query; Que) genomes were compared.

Coordinate (bp)	Ref	Que	Gene	Gene product	Amino acid substitution
29688	T	A	*lipB*	Lipoate-protein ligase B	V to D
120631	G/T	A	*hemR*	Haemin receptor	L to Q / R to Q
153123	A/G	T	*dnaQ*	DNA polymerase III subunit epsilon	C to F / Y to F
224090	C/T	A	*dam*	DNA adenine methylase	V to E / A to E
318285	C	G	*menD*	2-succinyl-5-enolpyruvyl-6-hydroxyl-3-cyclohexene-1-carboxylate synthase	L to V
597898	T	G	*fusA*	Elongation factor G	Y to D
673657	C	T	HI0634	tRNA-dihydrouridine synthase A	Y
677077	A	T	HI0635	Hemoglobin-binding protein	R to S
677154	A	G	HI0635	Hemoglobin-binding protein	I to V
677157	A	G	HI0635	Hemoglobin-binding protein	I to V
705310	C	G	*oapA*	Hemoglobin-binding protein	N to K
753312	T	G	*mutS*	DNA mismatch repair protein MutS	S to A
760386	T	C	HI0712	Hemoglobin-binding protein	N
765944	G	C	*nusG*	Transcription antitermination protein NusG	M to I
785858	T	A	HI0730	Organic solvent tolerance protein	V to D
911339	G/T	C	*vacB*	Virulence-associated protein	H / Q to H
932583	A/C	G	*rplU*	50S ribosomal protein L21	L to V / M to V
991151	G/T	A	*eno*	Phosphopyruvate hydratase	E / D to E
1072736	A	T	*ispH*	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	E to D
1073376	G/T	A	HI1008	Hypothetical protein	A to T / S to T
1135980	T	A	*hrpA*	ATP-dependent RNA helicase HrpA	V to E
1191409	A	G	*oppA*	Oligopeptide ABC transporter substrate-binding protein	T to A
1191430	A	G	*oppA*	Oligopeptide ABC transporter substrate-binding protein	T to A
1473587	C/T	A	*sbcB*	Exonuclease I	A to E / V to E
1474924	A/C	T	*phoR*	Phosphate regulon sensor protein PhoR	F / L to F
1495274	T	C	*fumC*	Fumarate hydratase	V to A
1725283	T	C	*nrdA*	Ribonucleotide-diphosphate reductase subunit alpha	N

3.2.2.2 Identification of indels

Indels, like SNPs, are a form of nucleotide-level genetic variation, where there is either a deletion or an insertion of any number of nucleotides in a reference DNA sequence when compared to a query DNA sequence. Indels were identified between the original and re-sequenced Rd and R2866 whole genomes. As with SNPs, the original genome was used as a reference for detecting indels, while the re-sequenced genome was a query.

A total of 226 indels were detected for the Rd strain. Of these, 93 were present in intergenic regions and the remaining 133 were located in genes. 61 indels were identified in 45 genes that were annotated as pseudogenes in the original published Rd genome sequence (see Table 3.3). The majority of these indels were present in the coding sequence in a way that they corrected frameshifts, which were most likely responsible for the annotation of the gene as a pseudogene. Only three indels, present in HI0247, HI1099a and HI1268 genes, did not have an obvious correction of the frame.

Table 3.3: Indels present in pseudogenes in the Rd strain. The original (reference; Ref) and re-sequenced (query; Que) genomes were compared.

Coordinate (bp)	Ref	Que	Pseudogene	Coordinate (bp)	Ref	Que	Pseudogene
108367	A	AG	HI0101	1195324	TG	T	HI1126.1
130905	AT	A	HI0116	1195695	C	CA	
166073	A	AT	HI0148.1	1249451	G	GC	HI1183
170592	C	CT	*dcuB*	1257711	TA	T	HI1191
265439	TC	T	HI0234	1257768	TG	T	
266933	G	GT	*perM*	1257775	TG	T	
277114	TAA	T	HI0247	1257781	TG	T	
313016	AG	A	HI0279m	1307904	T	TA	HI1235
354945	G	GC	HI0326	1330978	AG	A	HI1254
369921	G	GC	*napF*	1331102	C	CT	
370788	GC	G	*napA*	1331439	C	CT	
576723	GA	G	*devB*	1347623	C	CA	HI1268
607968	G	GC	HI0585	1377957	T	TG	*dgt*
654433	GC	G	HI0620.1	1476596	TC	T	*pstB*
787178	T	TAA	HI0732	1523755	GC	G	HI1434.2
787337	A	AT		1525290	GT	G	*thiI*
787392	A	AT		1541685	G	GA	HI1458m
892480	A	AG	HI0842	1569019	G	GC	*mor*
921808	AG	A	HI0869	1580044	G	GC	HI1506
922034	CA	C		1607054	GT	G	HI1534
929031	CA	C	*pepB*	1608081	G	GC	
1014632	AG	A	HI0956	1634585	TC	T	HI1565m
1034913	T	TA	HI0976	1637524	G	GC	HI1570
1082187	T	TG	HI1018	1648801	A	AT	HI1581
1127714	G	GT	HI1063	1662092	A	AG	*ftsK*
1128978	T	TC		1662266	AT	A	
1128980	T	TC		1662939	AC	A	
1144510	CG	C	*pnuC*	1687060	TG	T	HI1619
1144973	AC	A		1734148	G	GC	HI1666
1161063	C	CA	HI1099a	1789588	GC	G	HI1718
1195317	TG	T	HI1126.1				

The remaining 71 indels were detected in 51 coding sequences (see Table 3.4). The majority of these indels were located either towards the very start or the very end of a gene, therefore not greatly altering gene length. Six indels were present in the coding sequences that were split across two coding frames in the original genome, therefore correcting the frameshift. Seven indels were located in the middle of the coding sequences without any effect on the gene length. There were only three indels that were present in the middle of the coding sequences and resulted in a significantly truncated gene. The distribution of all indels and SNPs in the Rd genome is depicted in Figure 3.1.

Table 3.4: Indels identified in protein-coding genes in the Rd strain. The original (reference; Ref) and re-sequenced (query; Que) genomes were compared.

Coordinate (bp)	Ref	Que	Gene	Gene product	Comments
52066	C	CA	HI0050m	Integral membrane protein transporter	Start of gene
77924	A	AT	*ppnK*	Inorganic polyphosphate/ATP-NAD kinase	End of gene
104315	A	AG	*hitA*	Iron-utilisation periplasmic protein hFbpA	Corrects frameshift
142122	TG	T	*fbpC*	Ferric transporter ATP-binding protein	End of gene
142152	TA	T			
142169	TC	T			
142202	GC	G			
142225	TC	T			
142237	GC	G			
143334	G	GCA	*afuB*	Ferric transport system permease-like protein	End of gene
162571	G	GCA	HI0147	Hypothetical protein	End of gene
162605	G	GA			
201555	AG	A	*tatA*	Sec-independent protein secretion pathway component TatA	Start of gene
201574	AG	A			
391547	AG	A	HI0367	Hypothetical protein	End of gene
401967	AC	A	HI0380.2	tRNA-Lys	Start of gene
511410	CA	C	*aphA*	Acid phosphatase/phosphotransferase	Corrects frameshift
560060	GC	G	*ureH*	Urease accessory protein	End of gene
579334	A	AT	HI0559.1	Hypothetical protein	End of gene
588149	AG	A	*tex*	Transcription accessory protein	End of gene
608500	GC	G	*pepE*	Peptidase E	End of gene
620939	G	GT	*ccrB*	Camphor resistance protein CrcB	Possibly true
706625	TC	T	*oapA*	Hemoglobin-binding protein	Start of gene
721475	G	GT	HI0680	RarD protein	End of gene
850970	A	AG	*rpsM*	30S ribosomal protein S13	End of gene
926774	G	GA	HI0874	Hypothetical protein	End of gene
989395	A	AT	HI0930	Hypothetical protein	Start of gene
1036342	TG	T	*prmA*	Ribosomal protein L11 methyltransferase	Middle of gene, no effect
1036357	AT	A			
1036363	GA	G			
1068024	GC	G	HI1004	Peptidyl-propyl cis-trans isomerase	Corrects frameshift
1104427	C	CT	*ureF*	Urease accessory protein	Start of gene
1229388	TC	T	HI1159m	Thioredoxin domain-containing protein	Corrects frameshift
1231800	T	TCCGC	HI1162	Hypothetical protein	Start of gene

Coordinate (bp)	Ref	Que	Gene	Gene product	Comments
1244068	TG	T	HI1174	Opacity protein	Start of gene
1277863	A	AG	*lysS*	Lysyl-tRNA synthetase	Middle of gene; no effect
1277877	GA	G			
1290107	CA	C	*cmk*	Cytidylate kinase	End of gene
1290126	TC	T			
1300566	TC	T	*lpdA*	Dihydrolipoamide dehydrogenase	End of gene
1370921	A	AGC	*truB*	tRNA pseudouridine synthase B	End of gene
1375515	CG	C	HI1296	Nuclease	End of gene
1375546	A	AT			
1394146	TC	T	HI1317	Hypothetical protein	Start of gene
1394184	TC	T			
1449975	AT	A	HI1364	Transcriptional regulator	End of gene
1449988	TG	T			
1449998	GT	G			
1480065	G	GA	*pstS*	Phosphate ABC transporter substrate-binding protein	Corrects frameshift
1505885	G	GC	HI1410	Terminase large subunit-like protein	Start of gene
1509059	G	GC	HI1418	Hypothetical protein	End of gene
1527279	G	GCAC	*ispA*	Geranyltranstransferase	End of gene
1564858	C	CG	*muB*	DNA transposition protein	End of gene
1564863	G	GC			
1570070	T	TG	HI1493	Hypothetical protein	Possibly true
1571863	CG	C	HI1498.1	Hypothetical protein	End of gene
1587908	A	AG	HI1516m	Unannotated	Possibly true
1617702	CA	C	HI1546	Hypothetical protein	End of gene
1621283	G	GA	*bioD*	Dithiobiotin synthetase	End of gene
1647894	CA	C	*lpp*	15 kDa peptidoglycan-associated lipoprotein	End of gene
1647913	CA	C			
1647920	CA	C			
1675969	G	GT	*tyrS*	Tyrosyl-tRNA synthetase	End of gene
1689811	AT	A	HI1625	Hypothetical protein	End of gene
1691975	C	CA	HI1629	Hypothetical protein	End of gene
1691994	T	TA			
1697887	A	AT	*purR*	DNA-binding transcriptional repressor PurR	Middle of gene; no effect
1697907	TG	T			
1718873	A	AG	*tldD*	Hypothetical protein	Corrects frameshift
1738448	G	GC	HI1670	Solute/DNA competence effector	End of gene
1775364	ATC	A	HI1704	Hypothetical protein	End of gene

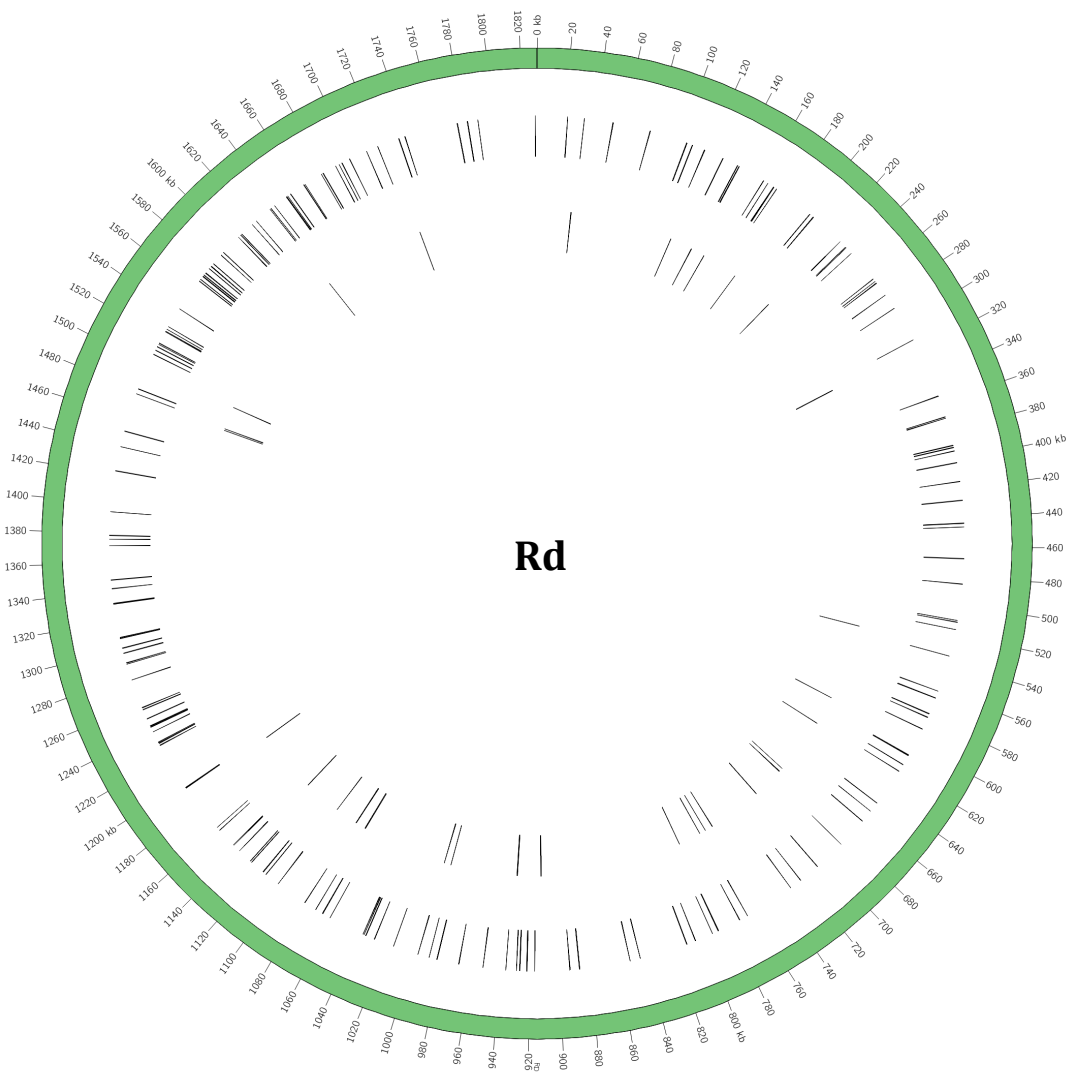

Figure 3.1: Positions of indels and SNPs in the Rd genome. The genomic locations of indels and SNPs are shown in two concentric circles. The outer circle shows the positions of all identified indels, while the inner circle shows the positions of all identified SNPs.

Only one indel was present in the R2866 strain at the 953,610 bp coordinate in the original genome. It was located in an intergenic region containing several tetranucleotide repeats. The indel itself constituted a single GCAA repeat, which was present in the original R2866 genome, but absent in the re-sequenced genome.

3.2.3 Comparison of Rd and R2866 genome sequences

In order to begin evaluating the genetic differences between Rd and R2866 strains, their re-sequenced genomes were aligned and visualised in ACT (see Figure 3.1). Several genome regions were clearly present in only one of the strains, forming their accessory genome (see section 3.2.3.1). ANI is another method to infer genetic relatedness between bacterial strains (Goris et al., 2007). In this study, ANI between Rd and R2866 strains was found to be 97.4%, which was the mean value from the two methods commonly used to calculate it (see Table 3.5).

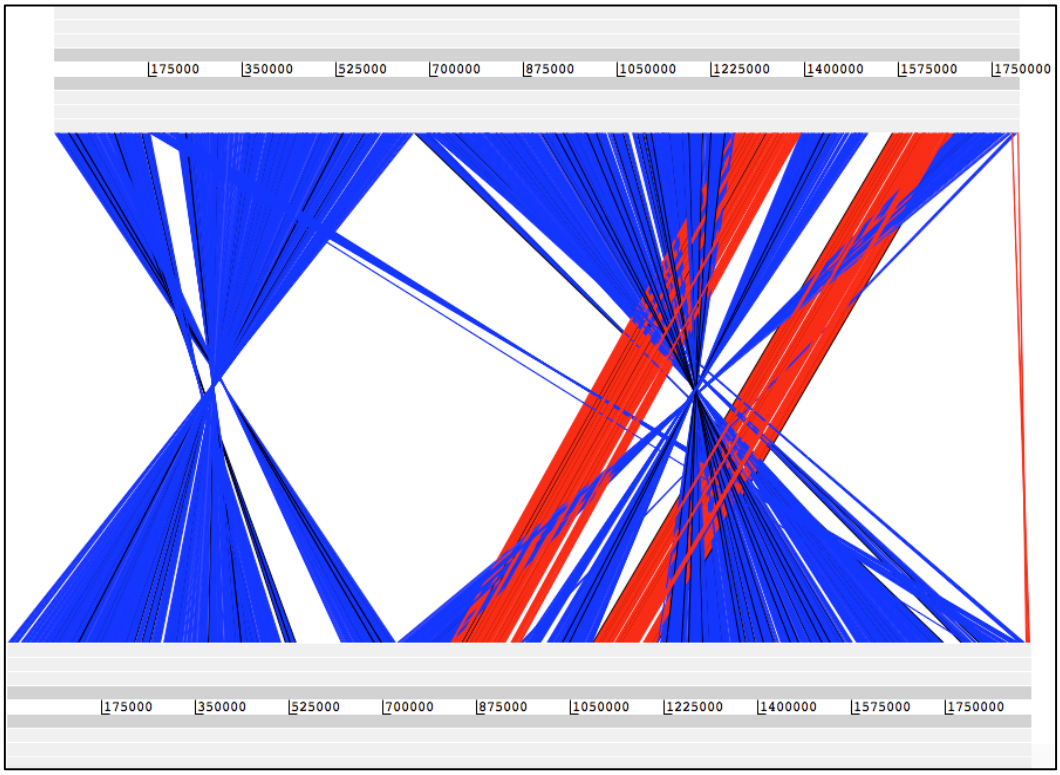

Figure 3.2: ACT view of the alignment of Rd and R2866 re-sequenced genomes. The Rd genome is shown at the top, while the R2866 genome is at the bottom. Blue and red lines show homologous sequences between the two strains.

Table 3.5: ANI between Rd and R2866 re-sequenced genomes. The best-hit value was determined by using the Rd genome as a query and the R2866 genome as a reference in a BLASTn comparison. Reciprocal best-hit method was a two-way BLASTn comparison, where Rd and R2866 genomes were used in turn both as a query and a reference.

ANI calculation method	ANI value (%)
Best-hit	97.3
Reciprocal best-hit	97.5

3.2.3.1 Rd and R2866 accessory genome

The accessory genome of Rd and R2866 strains of *H. influenzae* was determined manually in ACT, by identifying individual genes and gene clusters that were only present in one of the genomes. The total number of genes in the accessory genome of Rd and R2866 was 401. Of these, 134 genes were unique to the Rd strain and 267 were unique to the R2866 strain. Accessory genes, excluding those annotated as hypothetical proteins, are presented in Tables 3.6 and 3.7.

Table 3.6: Genes unique to the Rd strain when compared to R2866. Adjacent genes with the same annotation as well as major gene clusters were grouped together.

Gene name	Gene coordinates (bp)	Gene product
PROKKA_00088	94807 - 95238	PemK-like protein
PROKKA_00173	177022 - 177318	Chloride channel protein
PROKKA_00289	288703 - 289233	Transposase
PROKKA_00499	490791 - 491309	Ribosomal-protein-alanine N-acetyltransferase
hindIIR	575625 - 576401	Type II restriction endonuclease
hindIIM	576398 - 577954	Modification methylase
ureABCEFGH	607479 - 612940	Urease locus
PROKKA_00630	622672 - 623214	Transposase
PROKKA_00631	623339 - 623578	Putative membrane protein
PROKKA_00668	662731 - 664260	TRAP transporter, DctM subunit
arcC	664365 - 665297	Carbamate kinase
PROKKA_00727	716328 - 717242	Putative permease, DMT superfamily
PROKKA_00749	742135 - 742431	Bifunctional antitoxin/transcriptional repressor RelB
yafQ	742431 - 742739	mRNA interferase YafQ
ansB	782627 - 783676	L-asparaginase II
anmK	796479 - 797627	Anhydro-N-acetylmuramic acid kinase
murQ	797637 - 798548	N-acetylmuramic acid-6-phosphate etherase
tbp2_1	1009658 - 1010527	Transferrin-binding protein 2
PROKKA_01018	1010591 - 1011697	TPR repeat-containing protein precursor
PROKKA_01060	1059948 - 1060295	Putative cyclase
PROKKA_01067-01074	1068768 - 1076420	L-ascorbate utilisation locus
PROKKA_01075	1076614 - 1077402	Transcriptional regulator
ureF_2	1083828 - 1084796	Urease accessory protein
PROKKA_01083	1084793 - 1085707	Modification methylase
PROKKA_01302	1327053 - 1327589	Iron chelatin ABC transporter permease
PROKKA_01303	1327606 - 1328370	ABC transporter ATP-binding protein
yidK	1371852 - 1372142	Putative symporter YidK
hindIIIM	1464800 - 1465729	Modification methylase
hindIIIR	1465710 - 1466612	Type II restriction endonuclease
PROKKA_01423	1467153 - 1467575	Zeta toxin
holC	1467796 - 1468230	DNA polymerase III subunit chi
yhxB_2	1473006 - 1473515	Tail fiber protein/phosphomannomutase
PROKKA_01439	1480491 - 1481012	Phage regulatory protein, Rha family
PROKKA_01444	1482865 - 1483533	Putative phage-encoded protein
PROKKA_01445	1483894 - 1484193	Putative addiction module killer protein
PROKKA_01446	1484190 - 1484483	Putative addiction module antidote protein

Gene name	Gene coordinates (bp)	Gene product
PROKKA_01451	1486093 - 1487007	Integrase/recombinase
fiu	1519017 - 1519562	TonB-dependent receptor Fiu
fhuA	1520182 - 1521219	Ferric hydroxamate uptake
PROKKA_01496	1521278 - 1523047	ABC transporter ATP-binding protein
PROKKA_01498	1523438 - 1523785	Molybdate ABC transporter periplasmic molybdate-binding protein
PROKKA_01499-01501	1523795 - 1526604	*molABC* locus
modD	1526672 - 1527517	Molybdenum transport protein ModD
PROKKA_01503	1527527 - 1528126	ABC transporter ATP-binding protein
PROKKA_01504	1528128 - 1528367	Molybdate ABC transporter permease protein
PROKKA_01505-01550	1528715 - 1561534	Mu-like prophage, FluMu
PROKKA_01551	1562193 - 1563038	Adenine-specific DNA methylase
cysT	1563102 - 1563536	Sulfate transport system permease protein CysT
PROKKA_01553	1563592 - 1564329	Molybdate-binding periplasmic protein
oapA_3	1602540 - 1605476	Hemoglobin-binding protein
PROKKA_01591	1605545 - 1605961	Mu-like prophage FluMu G protein
PROKKA_01740-01741	1761540 - 1762106	Transposase
PROKKA_01752-01753	1773347 - 1776409	Hsf-like protein

Table 3.7: Genes unique to the R2866 strain when compared to Rd. Adjacent genes with the same annotation as well as major gene clusters were grouped together.

Gene name	Gene coordinates (bp)	Gene product
PROKKA_00017	17532 - 18215	GTPase Era
PROKKA_00019	20558 - 23287	Type I restriction enzyme EcoKI subunit R
PROKKA_00020	23624 - 23965	Transcriptional repressor DicA
PROKKA_00021	24335 - 24727	Putative transcriptional regulator
PROKKA_00022	24752 - 25429	Putative cation efflux protein
PROKKA_00050	52856 - 53302	CRISPR associated protein Cas2
mtrF	63686 - 65257	Antimicrobial resistance membrane protein MtrF
icsA	145675 - 147600	Outer membrane protein IcsA autotransporter precursor
PROKKA_00233	236619 - 237755	Putative transposase
rep1	285322 - 286140	Putative replicase protein
PROKKA_00287	286307 - 286459	DNA binding domain, excisionase family
PROKKA_00292	288551 - 288985	Putative transcriptional regulator
intA	289022 - 290233	Putative integrase
PROKKA_00389	390171 - 391112	Abortive infection bacteriophage resistance protein
ahpC	487938 - 488540	Peroxiredoxin
hsdR2	509752 - 512772	Putative type I restriction modification system, restriction enzyme component HsdR2
hsdS2	512873 - 514174	Putative type I restriction modification system, specificity component HsdS2
PROKKA_00509	514346 - 515809	Divergent AAA domain protein
hsdM2	515938 - 518310	Putative type I restriction modification system, methylase component HsdM2
PROKKA_00530-00568	539690 - 569387	Bacteriophage HP2
PROKKA_00570-00631	570441 - 623099	ICE, homologous to ICE*Hin*1056
PROKKA_00644	630888 - 632012	Putative peptidase
yjiG	632026 - 632496	Inner membrane protein YjiG
PROKKA_00646	632498 - 633130	Sporulation integral membrane protein YlbJ
hia	733204 - 736494	Adhesin Hia
PROKKA_00912	929171 - 929626	Putative 5'(3')-deoxyribonucleotidase
PROKKA_00913	929610 - 930320	Putative NAD-dependent protein deacetylase
doc	930369 - 931367	Death on curing protein
PROKKA_00916	931783 - 932607	Putative NAD-dependent protein deacetylase
PROKKA_00986-01004	993853 - 1004502	Putative bacteriophage
PROKKA_01006	1004954 - 1005919	P63C domain protein
PROKKA_01007	1006277 - 1006918	Putative prophage antirepressor protein
PROKKA_01078	1091522 - 1091761	DNA (cytosine-5-)-methyltransferase
PROKKA_01113	1128571 - 1129230	Putative TPR repeat protein

Gene name	Gene coordinates (bp)	Gene product
PROKKA_01115	1129418 - 1129864	Putative TPR repeat protein
PROKKA_01117	1130052 - 1130711	Putative TPR repeat protein
PROKKA_01119	1130899 - 1131345	Putative TPR repeat protein
hicAB; *hifABCDE*	1178066 - 1184978	Pilus assembly locus
PROKKA_01542	1577809 - 1578645	Putative lipooligosaccharide biosynthesis protein
losAB2	1579139 - 1580841	LOS biosynthesis *los2* locus
fpg2	1580844 - 1581656	Formamidopyrimidine-DNA-glycosylase 2
PROKKA_01546	1581705 - 1582430	Putative ABC transport system, periplasmic component
PROKKA_01547	1582440 - 1583180	Putative ABC transport system, ATPase component
PROKKA_01548-01549	1583177 - 1584189	Putative ABC transporter permease protein
PROKKA_01550	1584204 - 1584890	Iron-dicitrate transporter substrate-binding subunit
lex2AB	1680397 - 1681483	LOS biosynthesis *lex2* locus
PROKKA_01720-01767	1747866 - 1779744	Bacteriophage
tnaAB	1800781 - 1803518	Tryptophanase locus
mod2	1808282 - 1811182	Putative Type III restriction-modification system enzyme Mod
res2	1811184 - 1812830	Putative Type III restriction-modification system enzyme Res
PROKKA_01849	1865976 - 1867274	Archaeal ATPase
PROKKA_01851	1869130 - 1870020	Putative LysR-family transcriptional regulator
rimO	1870839 - 1872176	Ribosomal protein S12 methylthiotransferase

As expected, several mobile genetic elements were identified in the accessory genome of Rd and R2866 strains, some of which have been previously described in other studies (see Tables 3.6, 3.7). The 33-kb Mu-like prophage, FluMu, was present in Rd, but not in R2866 (Morgan et al., 2002). R2866 possessed the 30-kb bacteriophage, HP2, and another 11-kb putative bacteriophage. (Williams et al., 2002). As mentioned previously, the 53-kb ICE, homologous to ICE*Hin*1056, was present in the R2866 strain only (Juhas et al., 2007). The possession of this ICE has clinical significance for R2866, due to the presence of a *bla* gene, which encodes a beta-lactamase precursor and confers resistance to ampicillin. Both strains also contained a number of other genes that were annotated as part of mobile genetic elements, but did not seem to form specific gene clusters.

A number of pathogenicity-related genes were identified as unique to either Rd or R2866 strain. As described previously, a urease locus, *ureABCEFGH,* was only present in the Rd strain (see Table 3.6). This locus was thought to be associated with an increased survival of *H. influenzae* in the acidic environment of the human respiratory tract as well as circumventing lower pH inside intracellular vesicles (Murphy and Brauer, 2011). In addition, Rd possessed a gene encoding a putative transferrin-binding protein, with a possible role in the survival within iron-restricted environments inside the human host. A putative hemoglobin-binding protein, encoded by the *oapA_3* gene, was also present in Rd only. Gene *oapA* has been described in *H. influenzae* as playing an important role in the adhesion process (Weiser et al., 1995, Prasadarao et al., 1999).

Aside from the aforementioned ICE, responsible for the antibiotic resistance, a number of other pathogenicity-related accessory genes were unique to R2866 when compared to the Rd strain. As described previously, R2866 possessed a tryptophanase locus and the *mtrF* gene, their roles associated with virulence and antibiotic resistance respectively (see Table 3.7) (Kilian, 1976, Martin et al., 1998, Veal and Shafer, 2003, Erwin et al., 2005). Gene *icsA*, identified in R2866, encoded a protein homologous to a family of virulence-associated autotransporters (Davis et al., 2001). R2866, unlike Rd, also possessed two loci involved in the modification of LOS - *lex2AB* and *losAB2* (Deadman et al., 2009,

Hood et al., 2010). They are responsible for adding sugar moieties to LOS, thus contributing to its structural alteration between different *H. influenzae* strains, leading to variations in the fitness of this bacterium. Directly upstream of the *losAB2* locus was a putative LOS biosynthesis gene, *hgt1*, encoding a putative glycosyltransferase with a possible role in LOS modification (Hood et al., 2010).

A pilus-assembly locus was found to be present in R2866, but absent in Rd (see Table 3.7). Pili have been shown to promote the adhesion of *H. influenzae* to specific cell types (Gilsdorf et al., 1996). Gene *ahpC* encoding a peroxiredoxin was also unique to R2866 when compared to Rd. It is homologous to a *tsaA* gene in the 86-028NP strain of *H. influenzae*. Despite its homologue in *E. coli* having an important role in scavenging endogenous hydrogen peroxide, it was not responsive to the hydrogen peroxide treatment in the 86-028NP strain, suggesting an alternative function (Seaver and Imlay, 2001, Harrison et al., 2007). R2866 also possessed a gene encoding an important adhesin, Hia, which belongs to an autotransporter family and mediates the adhesion process of *H. influenzae* to several human cell types (Barenkamp and St Geme, 1996, St Geme and Cutter, 2000).

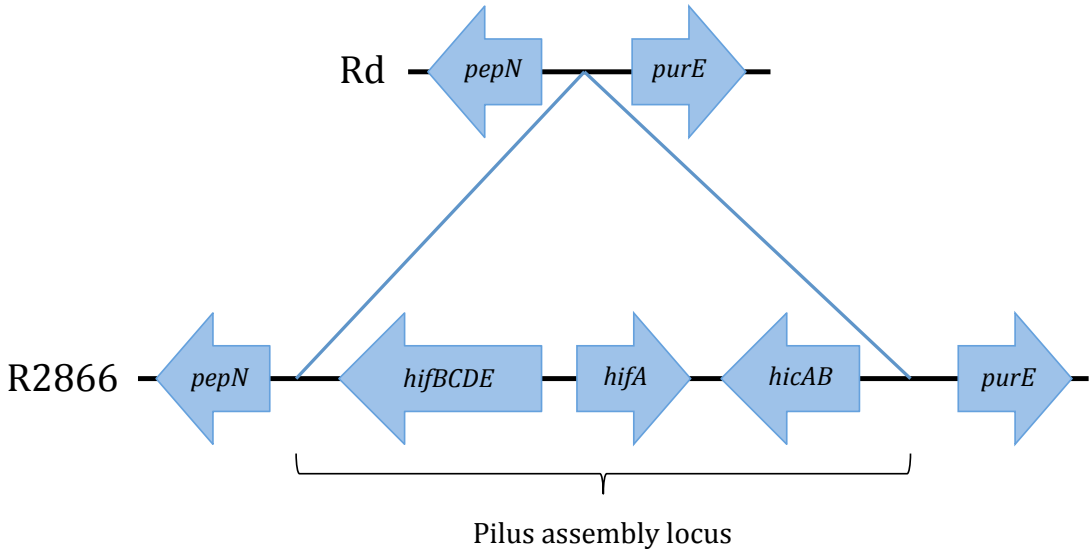

Figure 3.3: The accessory pilus assembly locus in R2866. Arrows denote the position and directionality of genes and gene loci.

Various metabolism-related genes and gene clusters were identified in the accessory genome of Rd and R2866 as well (see Tables 3.6, 3.7). A molybdate-uptake locus, *molABC,* was present in Rd and not in R2866 (Tirado-Lee et al., 2011). Rd also contained an L-ascorbate utilisation locus, comprised of seven genes and involved in an anaerobic catabolism of L-ascorbate as an alternative carbon source (see Figure 3.4) (Yew and Gerlt, 2002). Other metabolic genes present only in Rd were L-asparaginase, carbamate kinase as well as genes involved in the metabolism of anhydro-N-acetylmuramic acid. Interestingly, secretion of L-asparaginase II has been previously implicated in promoting virulence of *S. enterica* (Kullas et al., 2012). In addition to metabolism-associated genes, other noteworthy accessory features included a putative toxin-antitoxin system in Rd as well as several restriction-modification system genes present in both Rd and R2866 strains.

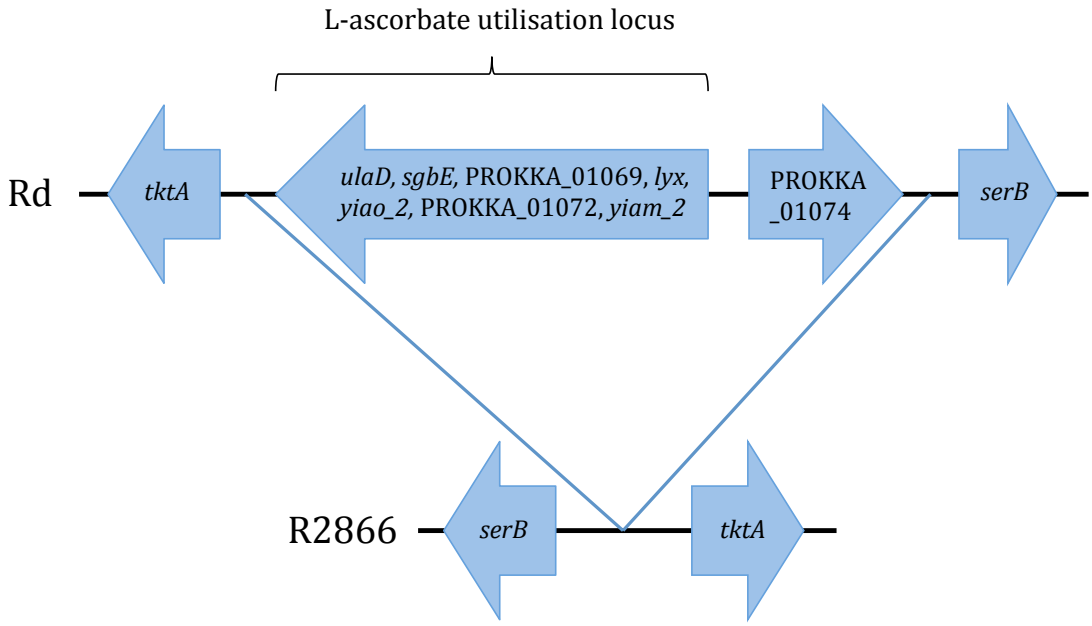

Figure 3.4: The accessory L-ascorbate utilisation locus in Rd. Arrows denote the position and directionality of genes and gene loci.

3.3 Discussion

The data presented in this chapter were used to describe potentially important differences in the genome composition of *H. influenzae* between and within two different strains. First, the availability of Rd and R2866 whole-genome sequences enabled detailed characterisation of the accessory genome of both strains. Second, re-sequencing of the Rd genome established the genetic variance between two descendants of the original Rd strain. These findings have implications for the use of the Rd strain as a model organism as well as setting the scene for the subsequent chapters, as discussed below.

The presence of multiple repetitive regions that are longer than raw Illumina sequencing reads (250 bp) creates gaps during the *de novo* assembly of a genome (Treangen and Salzberg, 2012). This results in a fragmented, contig-based draft genome sequence, as opposed to a complete whole genome, where gaps between contigs have been closed using additional techniques, such as PCR (Rogers et al., 2005). The inability to identify gaps between contigs as well as contig filtering steps leads to inevitable loss of some of the genome information. For this reason the re-sequenced genomes of both Rd and R2866 strains were shorter by 20-30 kb when compared to the original published genomes of the same strains. However, despite being draft versions of complete genomes, they still carry reliable genetic information about particular strains at a nucleotide-level, and thus can be used to infer changes between original and re-sequenced strains, as discussed later in this section.

The number of unique accessory genes identified in each strain was very similar to what was previously described for these two strains in a study by Hogg et al. (Hogg et al., 2007). Small differences can be attributed to the genome annotation and disparate statistical methods for identifying the accessory genome. R2866 was originally isolated from the blood of a sick patient (Nizet et al., 1996). Therefore, its possession of a larger number of virulence factors than Rd is not unexpected, as these factors likely contribute to an invasive phenotype. Pathogenicity-associated accessory genes in R2866 were well-

characterised and had distinct roles in infection-relevant processes, such as adhesion, modification of surface molecules, antibiotic resistance and the response to oxidative stress. Together, these pathogenicity-related factors establish R2866 as a more effective pathogen than Rd. The only characterised pathogenicity-associated gene cluster in Rd was the urease locus, whereas the rest had only putative roles in virulence.

All identified accessory metabolic gene clusters were only present in the Rd strain. This higher metabolic versatility may contribute to an increase in fitness of Rd during growth in different environments, though that still remains to be investigated. In addition, the identified accessory putative toxin-antitoxin system in Rd has homology to known toxin-antitoxin systems in *E. coli*, involved in the regulation of protein synthesis and bacterial growth (Christensen and Gerdes, 2003, Prysak et al., 2009). Several accessory genes, part of restriction-modification systems, were found in both Rd and R2866, representing diverse mechanisms of defence against foreign DNA as well as possible different roles in generating genetic diversity (Vasu and Nagaraja, 2013).

Re-sequencing of a reference genome has been previously applied to *Bacillus subtilis*, where a large number of small variations between two genomes was observed (Barbe et al., 2009). In addition, re-annotation of the re-sequenced genome of *B. subtilis* using up-to-date databases allowed them to identify novel putative functions for a number of genes. This highlights the benefit of re-sequencing reference genomes in an attempt to improve the annotation and correct previous sequencing errors. In this study, the whole genome of a model *H. influenzae* strain Rd was re-sequenced for the first time and compared to the original published sequence, resulting in identification of 41 SNPs and 226 indels.

The majority of SNPs identified in genes in Rd were non-synonymous, meaning that they could result in the loss of protein function (see Table 3.7). Six SNPs were found in genes coding for iron-associated proteins, with putative roles in survival within the human host. One SNP was also identified in a *vacB* gene

encoding a virulence-associated protein, though the ambiguity of the reference nucleotide means it could be a synonymous mutation. Since Rd has been propagated as a laboratory strain for several decades it is not surprising that the loss of function could occur in genes involved in the process of colonisation of its original niche and infection of the human host. These null mutations can actually be advantageous to bacteria, increasing their fitness during adaptation to new environments (Hottes et al., 2013). Of particular interest was a SNP identified in a *mutS* gene, which is part of a mismatch repair system in bacteria and has been known to be associated with increased mutation rates (LeClerc et al., 1996). However, the exact amino acid substitution in *mutS* identified in this study has not been described in the literature, making it hard to judge whether the SNP had any effect on the mutation rate of Rd.

While any of the identified SNPs in the Rd strain could be true mutations, it is also possible that some or even most of these SNPs originated due to errors in the original reference or re-sequenced genomes. The inherent errors present in the Illumina sequencing chemistry arise mainly due to inverted repeats and GGC sequences in the sequenced DNA, leading to SNP-associated errors (Nakamura et al., 2011). However, these errors are largely based on poor-quality nucleotides, which were filtered during SNP identification in this study. Therefore, it is unlikely to be the cause of SNP-related errors here. The original Sanger sequencing produced longer and more accurate reads than the Illumina sequencing chemistry (Shendure and Ji, 2008). However, since the original whole genome of Rd was the first to be sequenced over two decades ago, it can be expected that it would contain sequencing and assembly errors that can now be identified and corrected via re-sequencing of the same strain, coupled with existing computational methods (Fleischmann et al., 1995).

Nearly half of all identified indels in Rd were located in intergenic regions and, along with SNPs present in intergenic regions as well, could possibly contribute to changes in promoter sequences, which would affect the expression of downstream genes. Six indels were identified in genes that were split at a frameshift across two coding frames in the original genome. These indels

corrected the split in the gene by joining two coding frames. This clearly meant that the original frameshift was present as a sequencing error in the original genome. Most other indels were located at the very start or the very end of a gene, without a large effect on the gene length. Only three indels caused significant truncation of a gene, meaning that they could be either true mutations or a result of Illumina sequencing errors.

An interesting finding was the presence of a large number of indels in pseudogenes in the original published Rd genome. These pseudogenes contained one or more frameshifts, which resulted in significantly truncated coding sequences. Indels, identified in this study, corrected frameshifts in a way that pseudogenes were no longer truncated. This makes it likely that frameshifts were a result of previous sequencing errors and that these specific pseudogenes were in fact functioning genes. Correcting the number of pseudogenes in the original Rd sequence would reduce it from 67 to 22, which is also the number of pseudogenes present in the original published R2866 genome, further highlighting the falsehood of the majority of pseudogenes in the original Rd genome (see Table 1.2). Nevertheless, there were three pseudogenes in Rd, where indels did not correct the coding frame. Therefore, it is possible that these particular mutations were indeed true. Furthermore, some of the frameshifts could have originated as true mutations due to generation of a genomic library in *E. coli*, as part of the original sequencing protocol (Fleischmann et al., 1995).

The only genomic variant identified in R2866 was a tetranucleotide repeat in an intergenic region. This SSR has been previously described as being associated with the phase variation of genes involved in LOS biosynthesis (Jarosik and Hansen, 1994). The presence of this SSR in the intergenic region could signify phase-variable gene expression of an upstream gene, coding for a putative sigma factor. The overall lack of variations between R2866 original and re-sequenced genomes could be attributed to a more recent common ancestor of the two R2866 descendants. Additionally, the original R2866 genome was

sequenced in 2010 using HTS technologies, possibly contributing to a lower rate of sequencing and assembly errors than in the original published Rd genome.

While it is possible that some of SNPs and indels identified in this study are a result of Illumina sequencing errors, the correction of a large number of frameshifts within pseudogenes in the original Rd genome implies that the majority of indels and, possibly, SNPs originated due to errors in the original Rd sequence. This is further supported by some of the indels correcting the coding frames of genes in Rd. The fact that R2866 only had one variation, as opposed to 267 in Rd, signifies the improbability of Rd variations originating from Illumina sequencing errors. Despite the inability to reliably determine the exact source of most SNPs and indels in Rd, their presence still addresses a couple of important issues regarding the use of a model organism. First, it is necessary to be aware that the descendants of the same model strain could potentially have different genotypes and phenotypes. Second, using the old published genome sequence of a model strain could potentially introduce inaccuracies in the data analysis and experimental design, as demonstrated in this study by the presence of a large number of likely erroneous pseudogenes in the original Rd genome.

The data presented here lays important groundwork for subsequent chapters, where re-sequenced genomes of Rd and R2866 were used as references for building the transcriptional profile of these two strains. The knowledge of the accessory genome is also relevant for subsequent work, where the differential gene expression during several infection-relevant conditions will be examined in Rd and R2866.

Chapter 4: The response of *H. influenzae* during infection-relevant conditions

4.1 Introduction

4.1.1 Invasion of the host by *H. influenzae*

A recent dual RNA-Seq study by Baddal et al. explored the transcriptional response of the host and bacteria during general cell-association with *H. influenzae*, without distinguishing between adhesion and invasion processes. (Baddal et al., 2015). A more targeted approach, where either process is investigated separately, would reduce the potential transcriptional noise resulting from the mixture of both adherent and internalised *H. influenzae*. Therefore, the first major goal of this study was to optimise an invasion assay using an A549 human alveolar epithelial cell culture model, which could then be used to isolate RNA from both the host and intracellular bacteria simultaneously.

While both Rd and R2866 strains have previously been shown to infect A549 cells, the same strains from different laboratories could have varied phenotypes due to potential accumulated genetic differences (Daines et al., 2003, Soupene et al., 2003, Tsao et al., 2012). Small genetic variations were indeed observed for the Rd strain used in this study, when compared to another descendant of the original Rd strain (see Chapter 3). Therefore, the ability of Rd and R2866 strains, used in this laboratory, to infect A549 cells was also characterised. This then informed subsequent RNA-Seq experiments, where infection-relevant conditions were investigated.

4.1.1.1 Eukaryotic cell lysis methods

Triton™ X-100 and saponin are detergents that are used as standard lysis methods in bacterial invasion assays. Triton™ X-100 works by non-selectively permeabilising the lipid bilayer of eukaryotic cells (Koley and Bard, 2010).

Saponin creates pores in the eukaryotic plasma membrane by interacting with cholesterol molecules (Seeman et al., 1973). Sterile cold water can also be used to recover intracellular bacteria through the hypotonic lysis of eukaryotic cells. All three of these lysis methods have been previously successfully applied in bacterial invasion assays (Nair et al., 2000, Swords et al., 2000, Raffel et al., 2013). However, their potential bactericidal and bacteriostatic effects may vary between different bacterial species and need to be investigated for the specific bacteria used in the assay (Edwards and Massey, 2011).

4.1.1.2 Enrichment of bacterial RNA

Simultaneous transcriptome profiling of host-pathogen interactions poses some caveats. Bacterial RNA makes up a very small proportion of a mixed prokaryotic and eukaryotic RNA sample. In the study by Baddal et al. the mixed RNA sample only contained <1.5% of bacterial RNA (Baddal et al., 2015). Studying the transcriptome of a subset of infecting bacteria, i.e. differentiating between adhering and invading microorganisms, would reduce this ratio even more. Only high-end genome sequencers, like the Illumina HiSeq™, are suitable for such studies, since a high read depth needs to be achieved in order to recover any significant information of bacterial transcriptome (Humphrys et al., 2013, Westermann et al., 2016). Bacterial mRNA enrichment kits, such as Microb*Enrich*™ (Thermo Fisher Scientific, UK), are available in order to increase the ratio of bacterial RNA in the mixed sample at the expense of losing eukaryotic data. The Microb*Enrich*™ kit was used in this study for a proof-of-principle enrichment of bacterial RNA from a mixed host and *H. influenzae* RNA sample.

4.1.2 *H. influenzae* during infection-relevant conditions

Throughout the onset and course of both colonisation and infection of the human host, *H. influenzae* encounters a number of different stresses and environmental changes. Variations in the nutrient availability can lead to nutritional stress, forcing *H. influenzae* to scavenge any available nutrients in

order to survive. Nutrient limitation is also likely to lead to bacteria entering and persisting in stationary growth phase during natural infection (Kolter et al., 1993). Oxidative stress is another major obstacle for survival of *H. influenzae* in a colonised niche. It is primarily mediated by the activity of the human immune system as well as bacterial respiratory by-products and co-pathogens like *S. pneumoniae* (see section 1.1.8.1). To overcome this, *H. influenzae* has evolved various defence mechanisms, as discussed earlier (see section 1.1.8.1).

H. influenzae is likely to experience iron-limiting conditions during the colonisation and infection process and is therefore required to acquire iron and maintain iron homeostasis. High intracellular iron concentration can lead to oxidative stress through the Fenton reaction, whereas iron-starvation has a detrimental effect on bacterial growth (see sections 1.1.8.2; 1.1.8.3). It is important to study how *H. influenzae* responds to these stresses, as it may improve our understanding of its pathogenesis and help to identify novel strategies for disease prevention.

The second major aim of this chapter was to use the RNA-Seq technology to characterise the transcriptional response of Rd and R2866 strains during infection-relevant conditions. Differential gene expression of these strains during oxidative and iron-starvation stresses was examined in this study via treatment of bacterial cultures with appropriate agents. The nutrient-limiting MIV medium was used in this study to test nutritional stress response. This medium was originally developed to induce competence in *H. influenzae* during starvation conditions (Herriott et al., 1970). The final infection-relevant condition investigated in this study was the growth of *H. influenzae* during stationary phase in the standard rich medium, as compared to the growth at mid-exponential phase. The transcriptional response of a laboratory Rd strain was compared to an invasive R2866 strain, in order to shed light on what makes the latter a successful pathogen.

4.1.3 Important considerations for transcriptome profiling

4.1.3.1 Depletion of rRNA

In contrast to whole genome sequencing, where DNA is expected to be present at constant levels across the genome, the whole transcriptome contains a wide range of differentially expressed gene transcripts. In order to detect low-level transcripts, enough read depth needs to be acquired. 95-99% of total bacterial RNA consists of rRNA, hence removing rRNA transcripts before cDNA library preparation significantly increases the read depth for mRNA (Peano et al., 2013). The Ribo-Zero™ kit (Illumina, Inc.) has been reported to leave as little as <1% of rRNA, as compared to several other commercially available rRNA removal kits and methods (Giannoukos et al., 2012). It was used in this study to deplete rRNA in *H. influenzae* samples prepared for RNA-Seq experiments.

4.1.3.2 The use of replicates in transcriptome studies

The use of biological replicates in RNA-Seq experiments is important for downstream statistical inferences, such as differential gene expression analysis. Replicates help to identify outliers and improve the accuracy of data measurements. Importantly, the inclusion of replicates helps to account for the variation in gene expression between samples of the same origin. However, there is also inherent variation in RNA-Seq data itself. In addition to biological variation, between-sample variation also comes from a different number of sequencing reads per sample. Within-sample variation originates from longer genes having more reads mapped to them. To account for inherent RNA-Seq variation and to make gene expression comparable within and between samples, RNA-Seq data needs to be normalised first, as described below.

4.1.3.3 Normalisation methods

Reads per kilobase per million reads (RPKM) has been one of the most commonly used methods to normalise RNA-Seq data (Mortazavi et al., 2008). It

is calculated as the ratio of reads per gene length and the total number of reads. However, using the latter for normalisation creates a bias, as it does not represent the total number of transcripts (Wagner et al., 2012). Superior alternative normalisation methods include quartile, median and TPM (Dillies et al., 2013). The latter method is a modification of RPKM and has been used to normalise prokaryotic whole-transcriptome data (Wagner et al., 2012, Kroger et al., 2013). It is based on the total number of transcripts in RNA-Seq data, thus eliminating the bias present in the RPKM method. The TPM normalisation was used in this work to determine the absolute expression of genome features of *H. influenzae*.

A commonly used tool for the differential gene expression analysis of prokaryotic RNA-Seq data is DESeq2. (Love et al., 2014). It is implemented in R and is an advanced version of its predecessor DESeq, with added new features and improved statistical methodologies (Anders and Huber, 2010). DESeq2 models RNA-Seq data based on the negative binomial distribution and uses the aforementioned median normalisation method, which relies on calculating a median value of geometric mean ratios (Anders and Huber, 2010). Robust statistical methodologies and extensive use of the software in other bacterial transcriptomic studies make DESeq2 one of the best currently available tools for prokaryotic differential gene expression analysis (Bent et al., 2015, Jiang et al., 2016b). For these reasons, DESeq2 was used in this study to analyse differential gene expression in Rd and R2866 strains of *H. influenzae* during infection-relevant conditions.

4.2 Results

4.2.1 Development of a eukaryotic cell invasion assay

4.2.1.1 Optimisation of the invasion assay

Several aspects of the invasion assay were first optimised, including the MOI, eukaryotic cell lysis and infection time. The MOI represents an initial bacterial inoculum and is defined as the number of bacteria per single host cell. The standard MOI used in infection assays ranges from 1:1 to 100:1; a lower or higher MOI could potentially skew the data (Letourneau et al., 2011). MOI of 1:1, 10:1, 50:1 and 100:1 were tested for the Rd strain. A549 cells, infected with Rd at different MOI, were incubated for two hours prior to a one-hour gentamicin treatment. Triton™ X-100 was used to lyse infected A549 cells. There was a trend of a larger number of intracellular bacteria being recovered as the MOI increased, with 50:1 and 100:1 resulting in the highest numbers (see Figure 4.1).

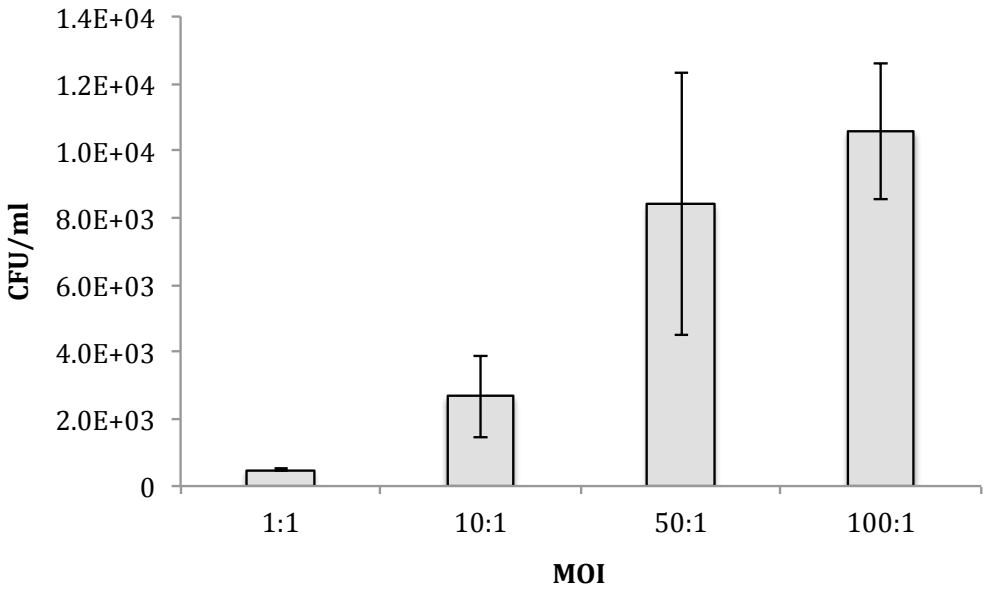

Figure 4.1: Optimisation of the MOI for the Rd strain. Results represent the mean CFU/ml (colony-forming units/ml) of recovered intracellular bacteria from three replicates. Error bars denote ± standard error (SE) of the mean.

The effect of Triton™ X-100, sterile cold water and saponin on the recovery of intracellular *H. influenzae* was tested after infecting A549 cells with the Rd strain for two hours at the MOI of 100:1 (see Figure 4.2). Saponin lysis of A549 cells resulted in the highest counts of intracellular bacteria, with over twice as much recovery as with Triton™ X-100. The recovery of invading bacteria with sterile cold water was only slightly higher than with Triton™ X-100. Saponin was therefore selected as a preferred eukaryotic cell lysis method for all subsequent invasion experiments.

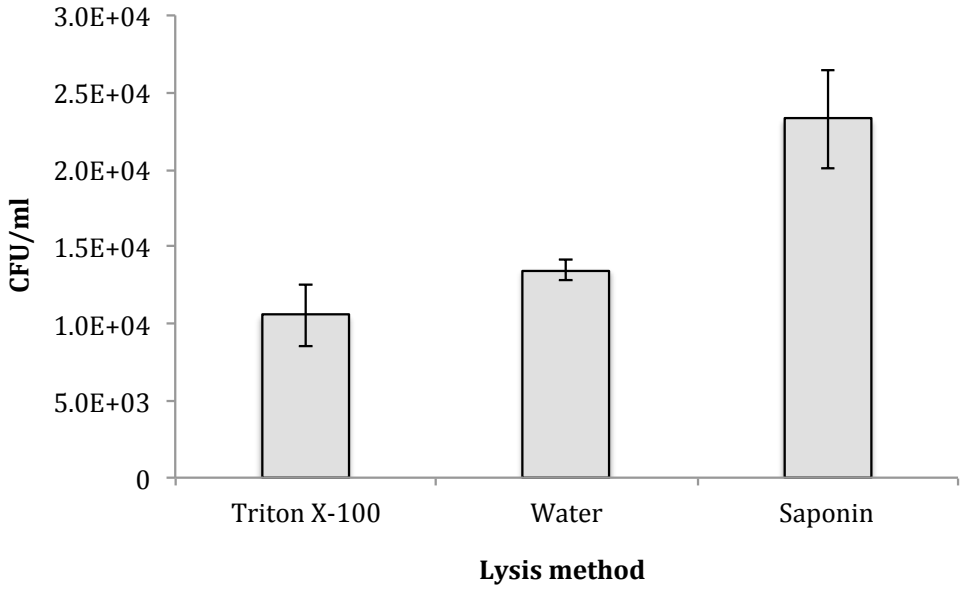

Figure 4.2: The effect of different eukaryotic cell lysis methods on the recovery of Rd strain. Results represent the mean CFU/ml of recovered intracellular bacteria from three replicates. Error bars denote ± SE of the mean.

In order to determine the optimal MOI for the R2866 strain, four different MOI were again tested - 1:1, 10:1, 50:1 and 100:1 (see Figure 4.3). A549 cells were incubated for two hours prior to the one-hour gentamicin treatment. Saponin was used as a eukaryotic cell lysis method, based on the previous optimisation. The same pattern of increasing intracellular bacterial numbers with a higher MOI was observed. Since one of the aims of this assay was to maximise the number of invading bacteria, the optimal MOI was chosen as 100:1. It was henceforth used in all infection experiments for both Rd and R2866.

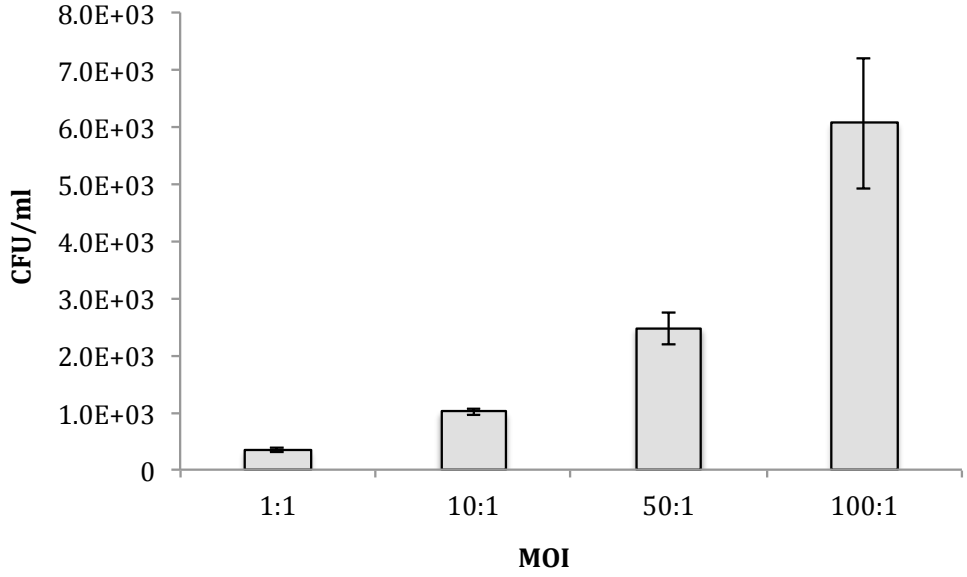

Figure 4.3: Optimisation of the MOI for the R2866 strain. Results represent the mean CFU/ml of recovered intracellular bacteria from three replicates. Error bars denote ± standard error (SE) of the mean.

The optimisation of infection time for Rd and R2866 strains was carried out next. The initial bacterial inoculum (at MOI of 100:1) was incubated with A549 cells for 2-5 hours, before proceeding with the gentamicin treatment in order to kill any remaining extracellular bacteria. For both Rd and R2866 strains, the number of recovered intracellular bacteria increased at 3 hours compared to 2 hours, but then decreased again at 4 hours (see Figure 4.4). However, there was another increase in the number of invading bacteria at 5 hours, particularly for the Rd strain. The optimal infection time was selected as 3 hours, due to a subsequent decrease in the number of intracellular bacteria at 4 hours. A larger number of recovered intracellular bacteria was observed at 3 hours for the Rd strain compared to R2866 with the same MOI of 100:1. The bacterial recovery at 3 hours was 8.47×10^4 CFU/ml for Rd and 4.97×10^3 CFU/ml for R2866.

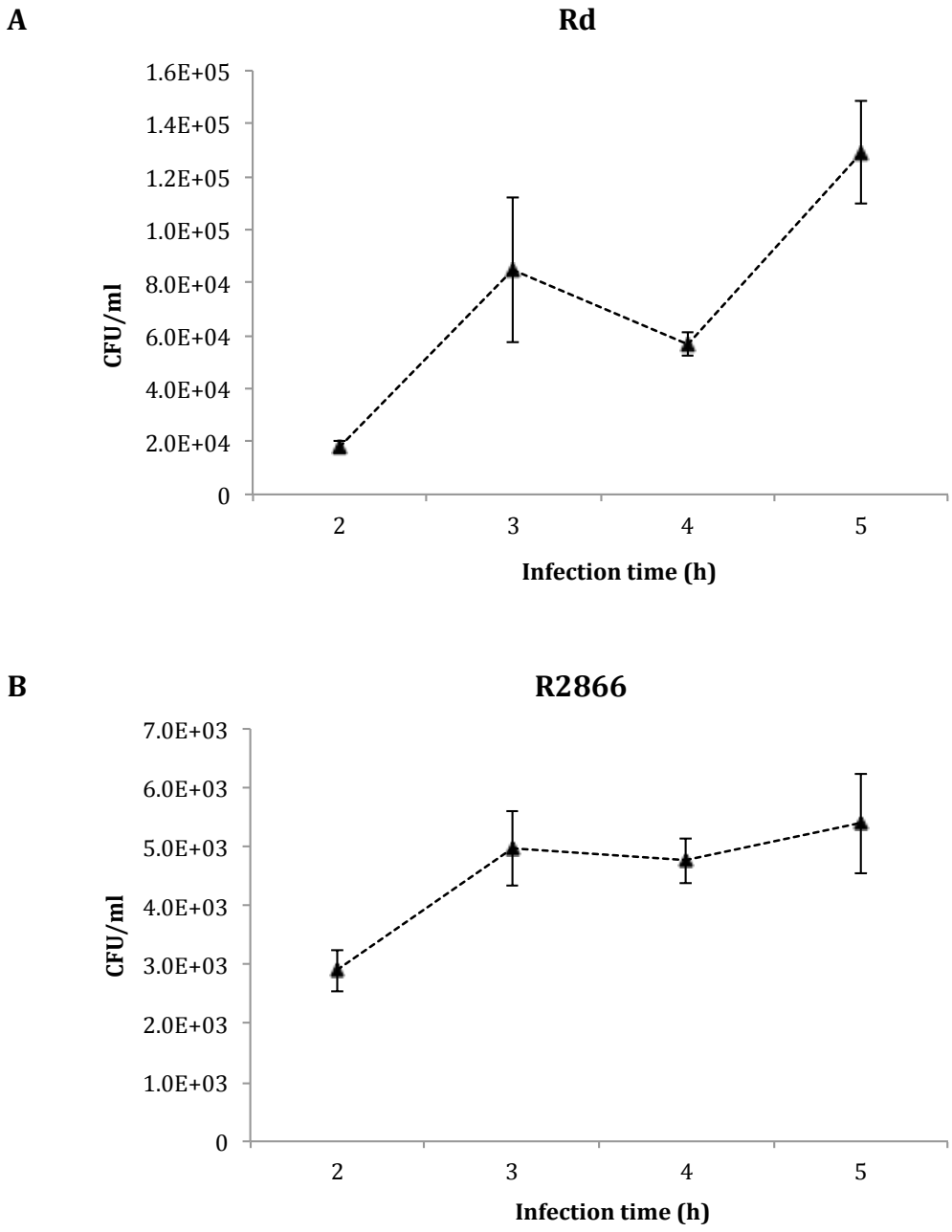

Figure 4.4: The optimisation of infection time. Results represent the mean CFU/ml of recovered intracellular bacteria from three replicates of Rd (A) and R2866 (B) strains. Error bars denote ± SE of the mean.

4.2.1.2 Effect of serum in the infection medium on the bacterial viability

In order to understand the effect of the infection medium, specifically the FBS component, on the viability of Rd and R2866, both strains were grown in DMEM either without FBS or with 2%, 5% and 10% FBS. Bacteria were quantified at 1.5 and 3 hours post-inoculation. As the FBS concentration in DMEM increased, the CFU/ml values rose for both Rd and R2866 (see Figure 4.5). When no FBS was present in DMEM, the viability of both Rd and R2866 decreased over time. Using 2% and 5% FBS resulted in an increase of Rd and R2866 numbers at 1.5 hours, compared to 0 hours. This was followed by a reduction in viability at 3 hours for both strains. The same growth pattern was observed for the R2866 strain at 10% FBS, though Rd continued to grow at 3 hours. The growth in 10% FBS (corresponding to the infection medium used in this study) resulted in the highest CFU/ml values for both Rd and R2866 strains.

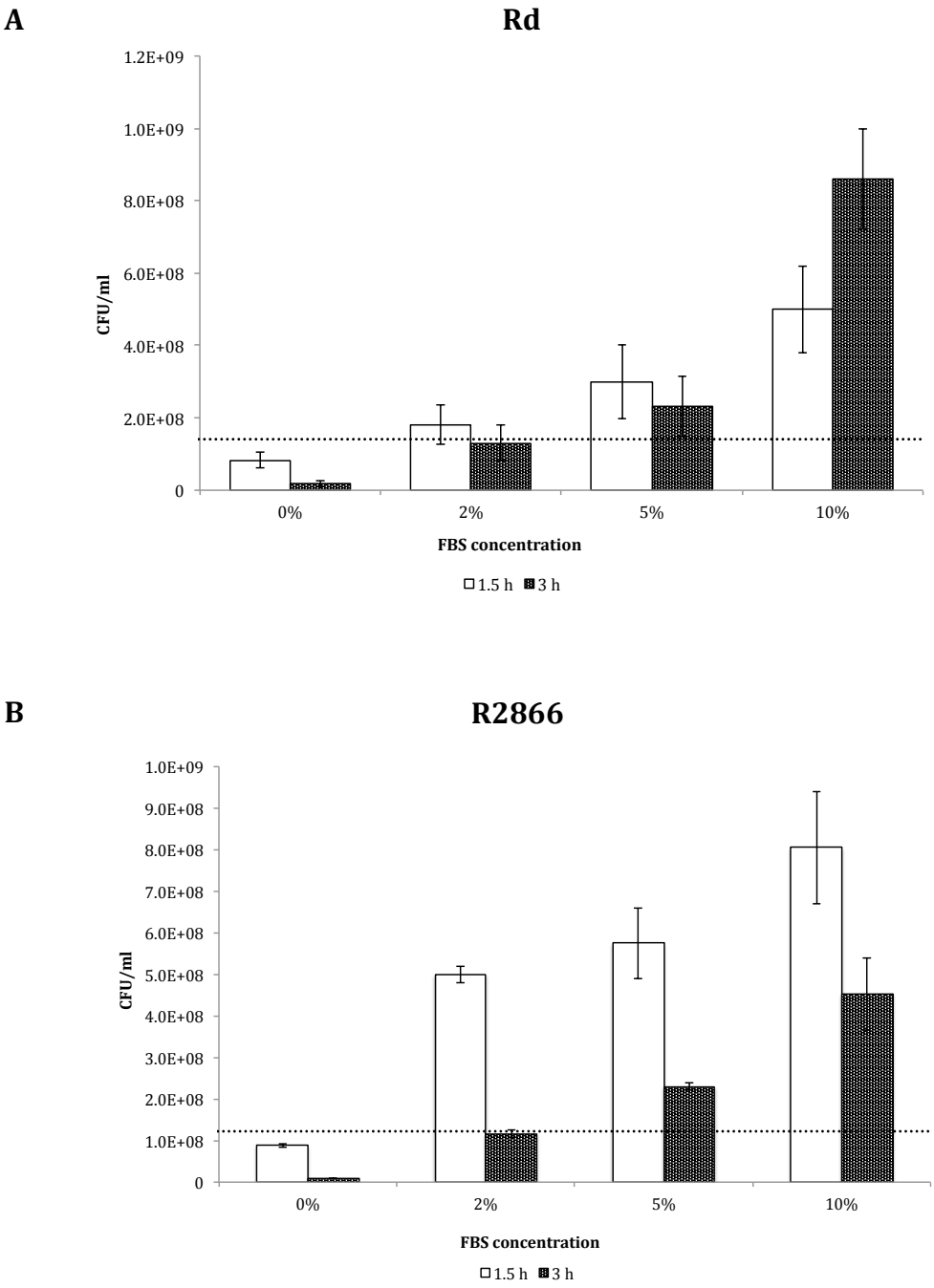

Figure 4.5: The effect of the FBS in the infection medium on the viability of *H. influenzae*. Results represent the mean CFU/ml from two replicates of Rd (A) and R2866 (B) strains. Error bars denote ± SE of the mean. A dashed horizontal line marks the CFU/ml value at 0 hours, as determined by a corresponding OD_{600} value.

4.2.1.3 Fluorescence microscopy of *H. influenzae* invasion of A549 cells

Fluorescence imaging of infected A549 cells was carried out in order to confirm the invasion of these cells by Rd and R2866 strains. The optimal MOI of 100:1 and infection time of 3 hours were used for the infection, prior to the one-hour gentamicin treatment. Following that, a lipophilic dye FM® 4-64 was used to stain the invading bacteria and eukaryotic vesicles, while cell nuclei were visualised using a standard DAPI stain (Vida and Emr, 1995). Non-infected A549 cells were used as control. Fluorescence imaging confirmed the presence of invading *H. influenzae* in A549 cells for both Rd and R2866 strains (see Figure 4.6). This was determined by observing the morphological differences between stained vesicles in a control sample and rod-shaped bacteria in infected cells.

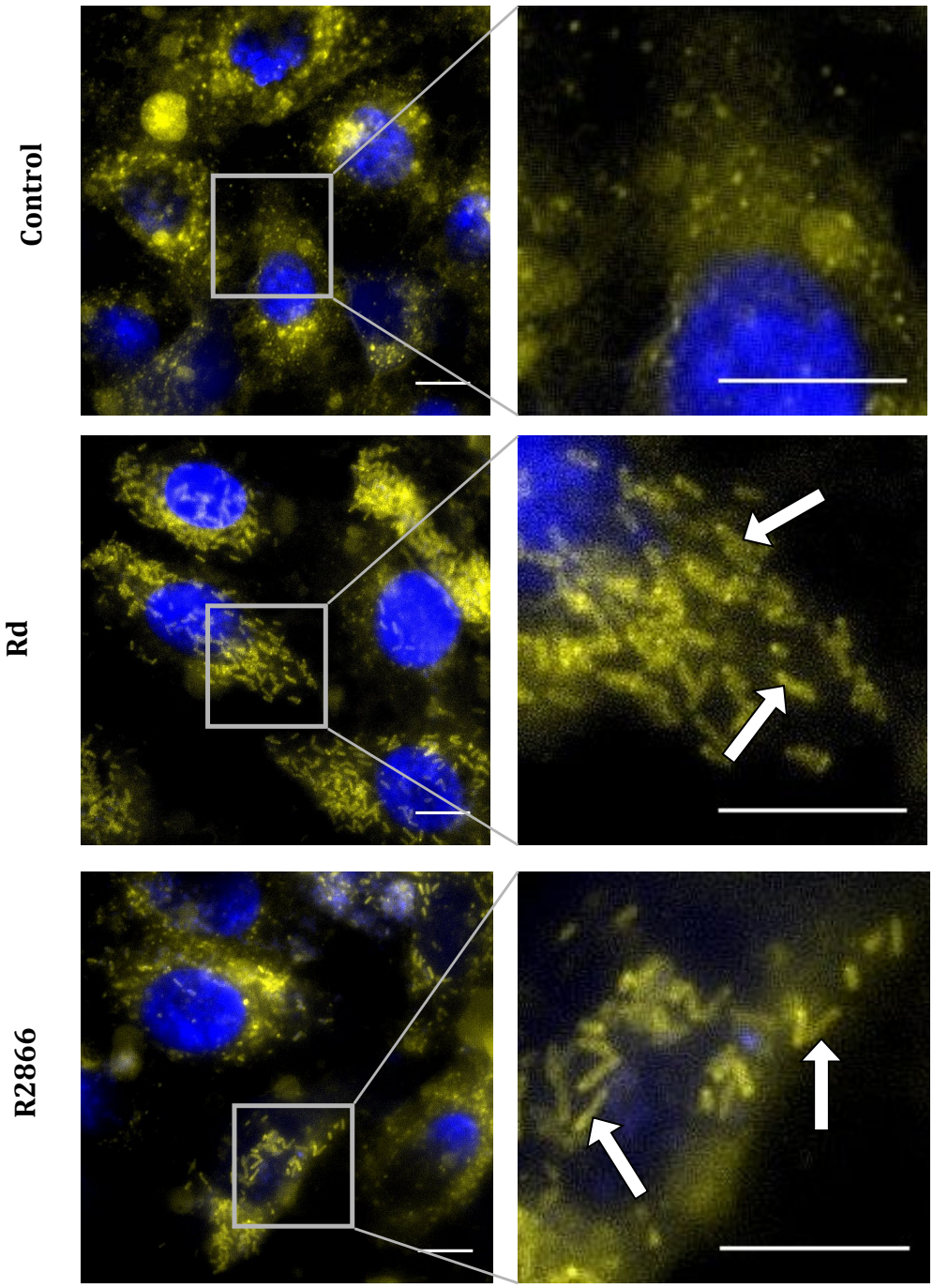

Figure 4.6: Fluorescence imaging of A549 cells infected with Rd and R2866 strains of *H. influenzae*. Mock represents non-infected cells. Images were taken on a Leica DMi8 inverted fluorescence microscope at the 100x magnification with oil immersion. Yellow (FM® 4-64 dye) - *H. influenzae* bacteria (arrows) and eukaryotic vesicles; blue (DAPI stain) - cell nuclei. Scale bar - 10 μm.

4.2.1.4 RNA extraction from infected cells

As proof of principle, RNA was extracted from A549 cells infected with the Rd strain using the optimised invasion assay conditions. The Bioanalyzer electropherogram images showed clear peaks of the human 18S and 28S rRNA as well as lower peaks of smaller fragment size, representing 5S rRNA, tRNA and other sRNAs (see Figure 4.7A). There was no trace of bacterial rRNA. Following the enrichment of the bacterial RNA with the Microb*Enrich*™ kit within the same mixed RNA sample, 18S and 28S rRNA peaks were absent, highlighting the efficiency of the removal of eukaryotic RNA (see Figure 4.7B). The largest peaks of the enriched RNA likely represented small eukaryotic RNAs (5S rRNA and tRNA). There were small peaks present of the higher fragment size as well, but it was not possible to interpret whether they represented the remaining human rRNA or indeed the enriched bacterial 16S or 23S rRNA, which would be expected to make up the majority of the bacterial RNA sample.

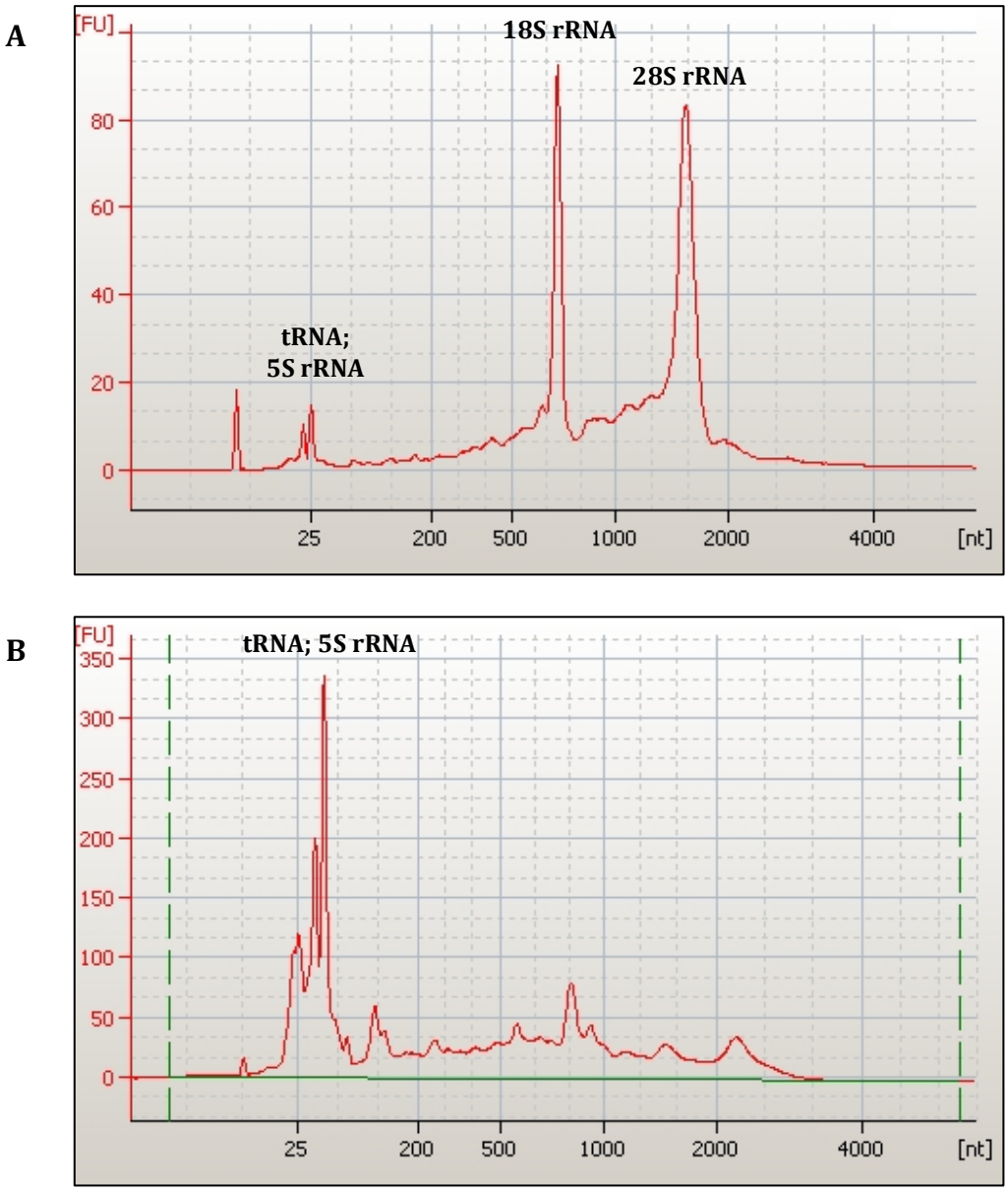

Figure 4.7: Quality control of the enrichment of bacterial RNA from a mixed host and *H. influenzae* RNA sample. Bioanalyzer electropherogram images depict traces of RNA following the total extraction of RNA from infected A549 cells (A) and subsequent enrichment of bacterial RNA (B). The first peak in each electropherogram image represents an RNA marker. The y-axis shows Bioanalyzer fluorescence units; the x-axis shows the RNA fragment size (bp).

4.2.2 RNA-Seq analysis of the transcriptional response of Rd and R2866 strains during infection-relevant conditions

4.2.2.1 RNA sample acquisition and preparation for RNA-Seq

4.2.2.1.1 Growth of Rd and R2866 strains in rich and nutrient-limiting media

Having established the ability of Rd and R2866 strains from this laboratory to infect human cells, the optimisation of infection-relevant conditions was carried out next. First, it was necessary to characterise the growth kinetics of Rd and R2866 in relevant media. That would then inform the acquisition of RNA samples at appropriate growth phases. Firstly, Rd and R2866 were grown in the rich sBHI medium for nine hours and a growth curve was plotted for each strain after measuring OD_{600} at one-hour intervals (see Figure 4.8). Rd and R2866 exhibited very similar growth, with mid-exponential phase being at OD_{600} of 0.3-0.5. Late stationary phase, when OD_{600} was no longer increasing, was at 9 hours for both strains. The transcriptional profile of *H. influenzae* would be compared between stationary and mid-exponential growth phases (see section 4.2.2.4.1). Therefore, bacteria were collected at these growth phases in the sBHI medium for the stationary phase RNA-Seq experiment.

The growth of Rd and R2866 in the nutrient-limiting MIV medium was also tested, in order to inform RNA sample acquisition for the nutritional-stress RNA-Seq experiment. As expected, Rd grew less well in MIV than in the sBHI medium, with mid-exponential phase being at an OD_{600} of 0.1-0.2 (see Figure 4.8). Rd reached late stationary phase in the MIV medium at 9.5 hours. The R2866 strain did not grow in MIV. For a nutritional stress RNA-Seq experiment, Rd broth cultures were collected at mid-exponential phase during growth in sBHI and MIV media. Nutritional stress was not tested for R2866, due to its inability to grow in the MIV medium.

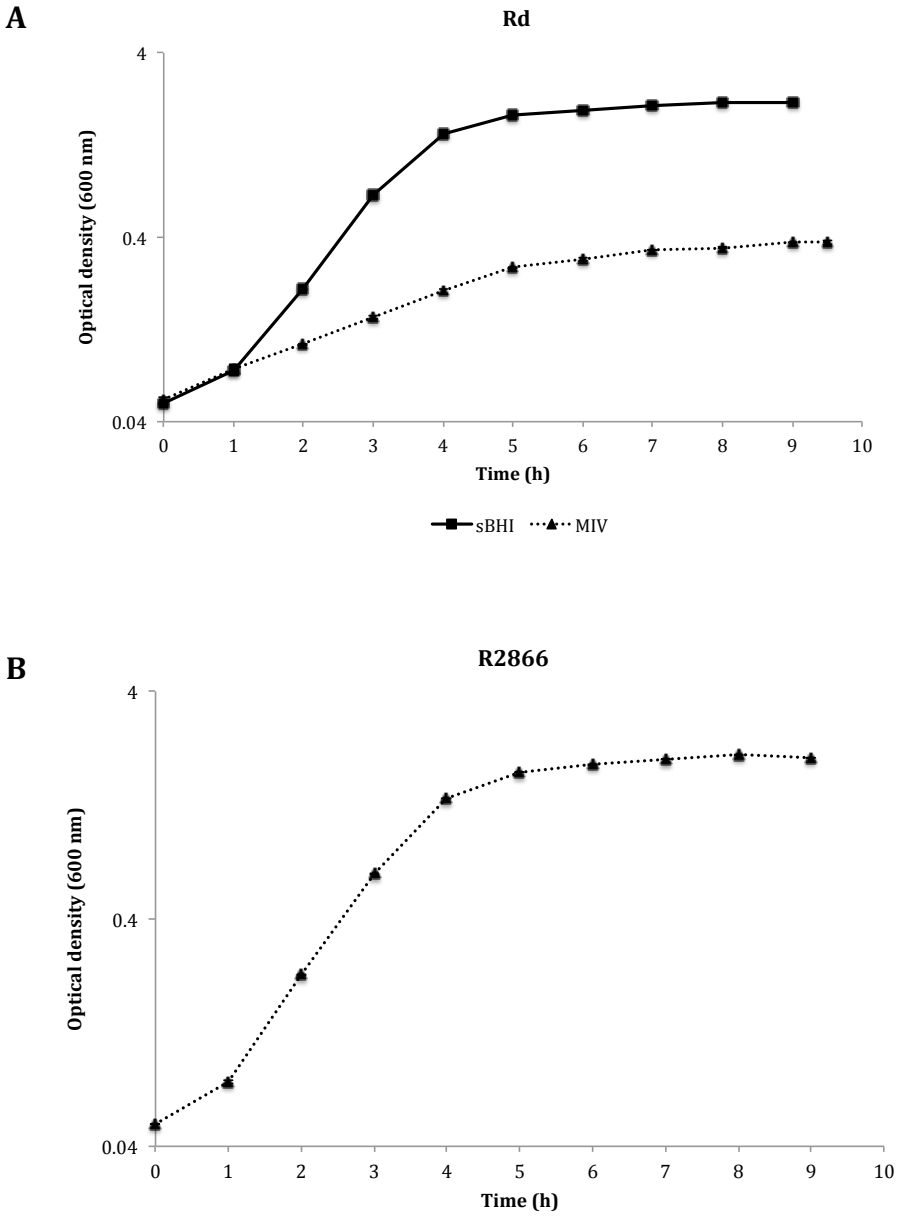

Figure 4.8: Standard growth curves for Rd and R2866 strains. Rd was grown in sBHI and MIV media (A); R2866 was grown in the sBHI medium (B). Results represent the mean OD_{600} from three replicates. Error bars denote ± SE of the mean. A logarithmic scale is used for the y-axis.

4.2.2.1.2 Optimisation of oxidative and iron-starvation stresses for Rd and R2866 strains

Hydrogen peroxide has been previously used as an agent for oxidative stress induction in *H. influenzae* (Wong et al., 2007). The effect of three different concentrations of hydrogen peroxide (1 mM, 5 mM and 15 mM) was tested on the growth of Rd and R2866 strains in this study. Hydrogen peroxide was added to sBHI broth cultures at mid-exponential phase and OD_{600} was measured at one-hour intervals. The R2866 strain was more sensitive to hydrogen peroxide treatment than Rd (see Figure 4.9). An appropriate concentration of hydrogen peroxide for the use in an oxidative stress RNA-Seq experiment was determined as 5 mM for both Rd and R2866 strains: it had an inhibitory effect on bacterial growth, yet did not result in a subsequent drop in OD_{600} as observed with the higher concentration of 15 mM.

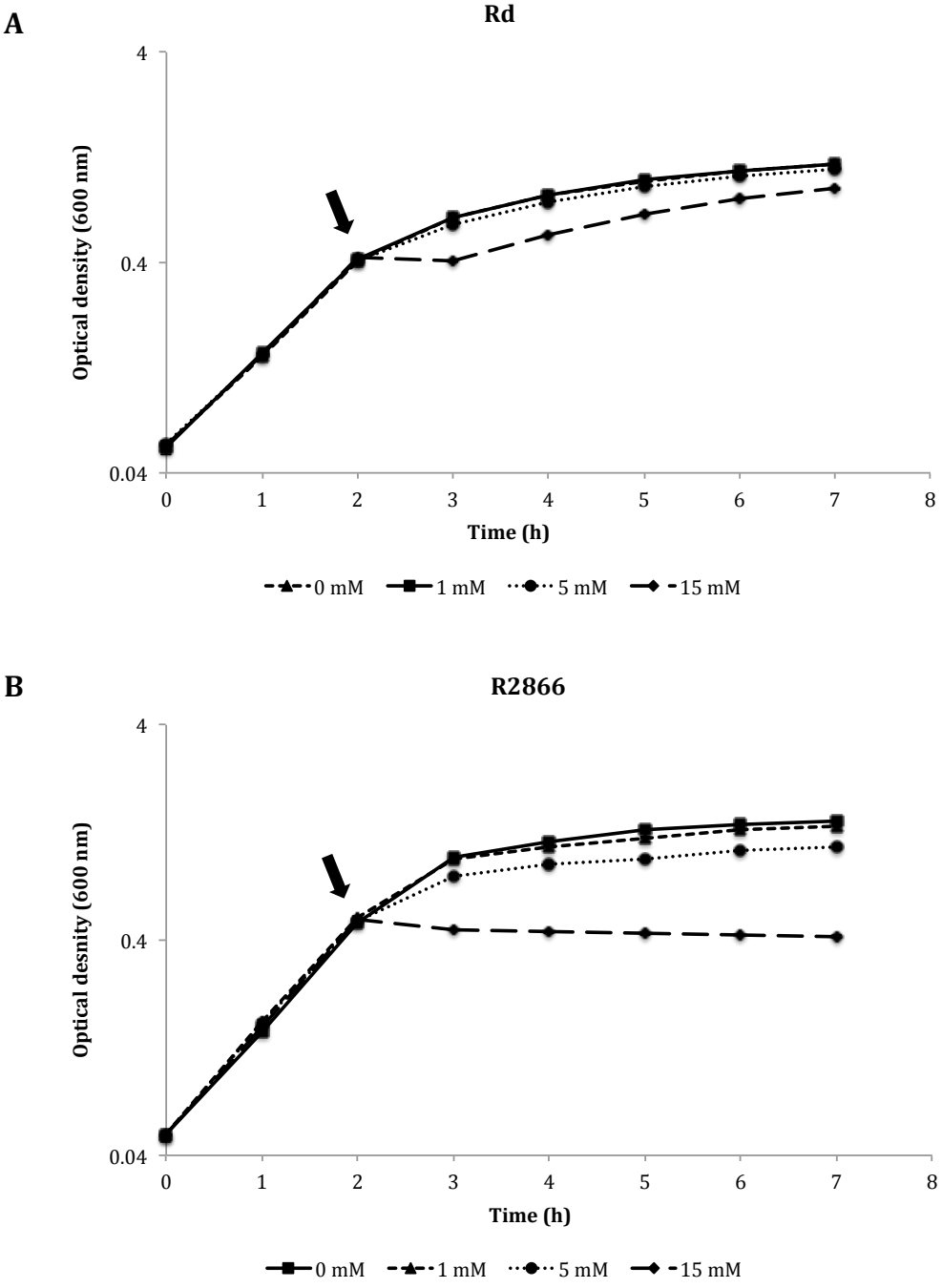

Figure 4.9: The growth response of Rd and R2866 strains to oxidative stress. Arrows denote addition of hydrogen peroxide to Rd (A) and R2866 (B) broth cultures. Results represent the mean OD_{600} from two replicates. Error bars denote ± SE of the mean. A logarithmic scale is used for the y-axis.

An iron chelator 2,2-bipyridine has been previously used to induce iron-starvation in *H. influenzae* (Harrison et al., 2013). The effect of three different concentrations of 2,2-bipyridine (1 mM, 2 mM and 5 mM) was tested on the growth of Rd and R2866. OD_{600} measurements were carried out hourly after addition of 2,2-bipyridine to sBHI broth cultures. As with oxidative stress, the R2866 strain was also more sensitive to iron-starvation than Rd (see Figure 4.10). It was also noted that the effect of 2,2-bipyridine on the growth was not noticeable in R2866 30 minutes after addition of the chelator, but was apparent after 1 hour. This was in addition to the liquid broth colour changing to dark red within 1 hour after treatment. An appropriate concentration of 2,2-bipyridine for use in an iron-starvation RNA-Seq experiment was deemed as 5 mM for both Rd and R2866 strains.

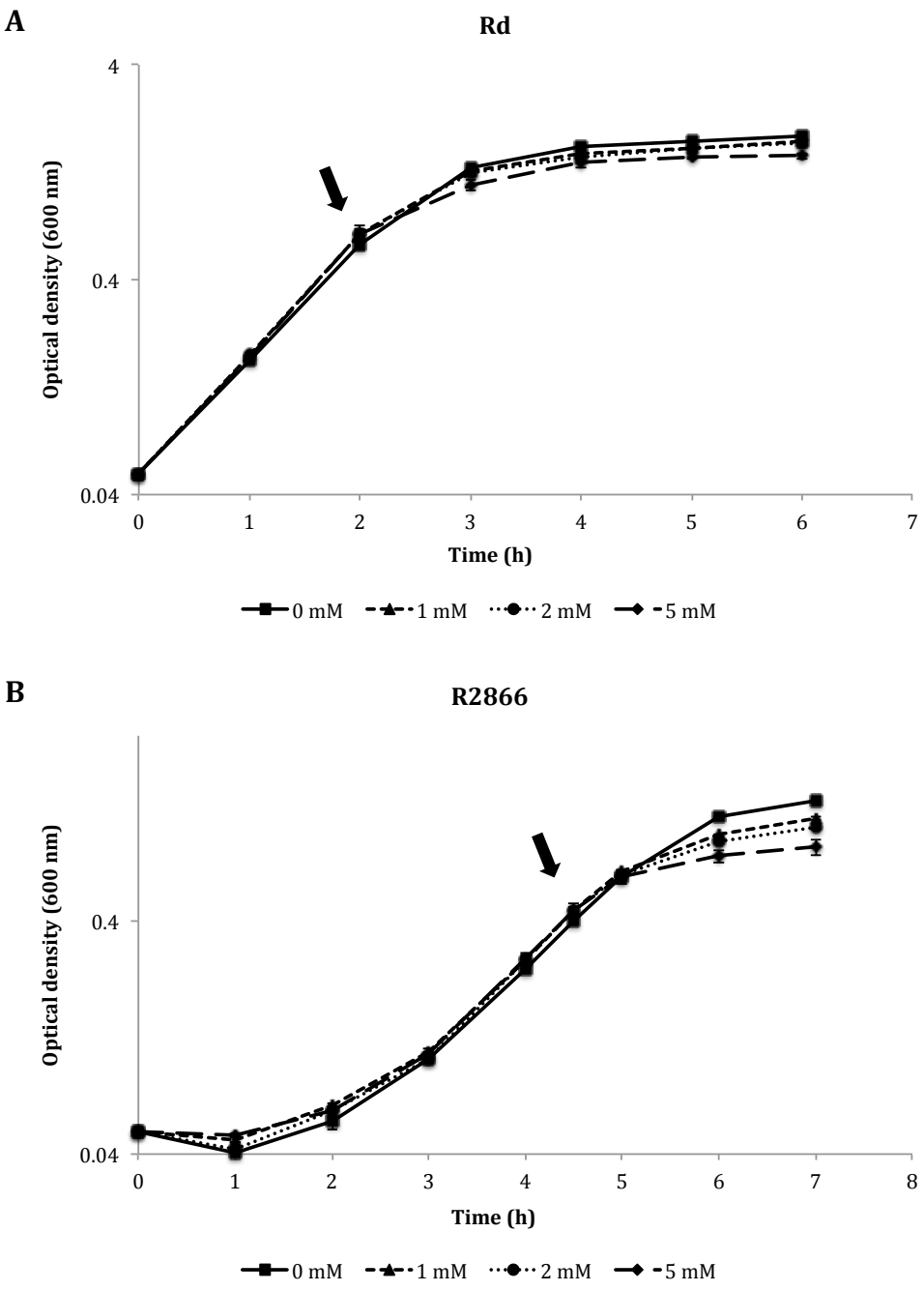

Figure 4.10: The growth response of Rd and R2866 strains to iron-starvation stress. Arrows denote addition of 2,2-bipyridine to Rd (A) and R2866 (B) broth cultures. Results represent the mean OD_{600} from two replicates. Error bars denote ± SE of the mean. A logarithmic scale is used for the y-axis.

4.2.2.2 Quality control of RNA extraction, rRNA depletion and cDNA library preparation

For each RNA-Seq experiment, Rd and R2866 strains were grown in triplicate and bacterial samples were collected at an appropriate growth phase or following a specific treatment, as described above. During RNA extraction and cDNA library preparation, several RNA and cDNA quality control checks were carried out using the Bioanalyzer. Figure 4.11 shows representative Bioanalyzer electropherogram images for Rd and R2866 after RNA extraction, rRNA depletion and cDNA library preparation. Following RNA extraction, Rd contained two clear 16S and 23S rRNA bands, while there were multiple RNA bands present in the R2866 strain. Depletion of rRNA was evident from the absence of clear rRNA bands and enrichment of mRNA. The cDNA Bioanalyzer electropherogram image represented amplified cDNA fragments and was used to quantify average cDNA fragment size, which was required for sequencing. RNA and cDNA samples used in all RNA-Seq experiments had equivalent Bioanalyzer electropherogram traces as shown in Figure 4.11. The cDNA libraries, prepared using TruSeq™ Stranded mRNA kit, were then sequenced on the Illumina MiSeq™ platform (see sections 2.8.2; 2.8.3).

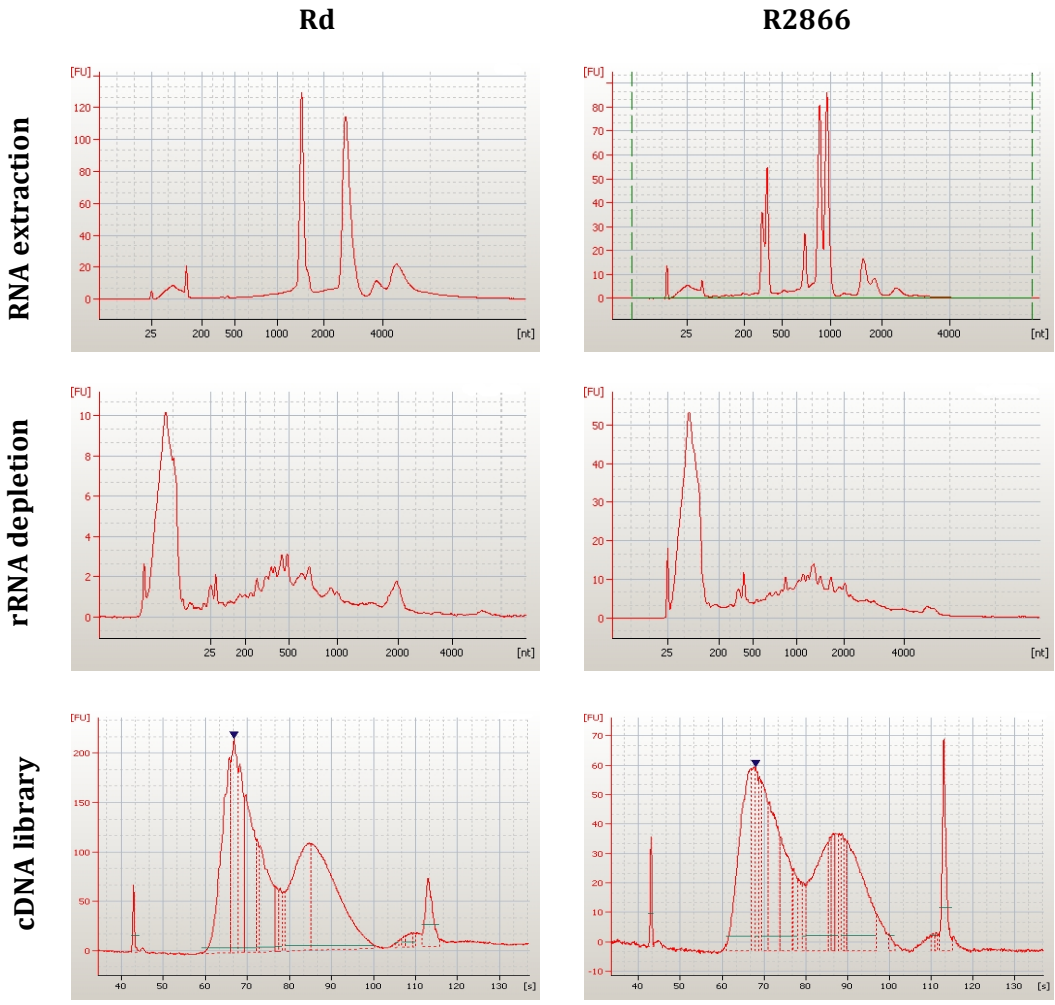

Figure 4.11: Bioanalyzer electropherogram images representative of RNA extraction, rRNA depletion and cDNA library preparation for Rd and R2866 strains. The triangular arrow points to a cDNA peak used to quantify the average fragment size of a cDNA library.

4.2.2.3 RNA-Seq data processing and quality control

Tables 4.1 and 4.2 show the number of RNA-Seq reads generated for every Rd and R2866 RNA-Seq sample as well as the percentage of RNA-Seq reads mapped to respective re-sequenced reference genomes. Each sample had over 3 million reads, with the maximum number of reads being over 30 million for two Rd samples. 99% or more of the generated RNA-Seq reads were mapped to reference genomes for all RNA-Seq samples. Next, the similarity between replicates from each RNA-Seq experiment was examined using the hierarchical clustering method. Replicates from the same condition in each RNA-Seq experiment indeed clustered together, apart from one R2866 replicate in the oxidative stress experiment (see Figures 4.12; 4.13).

Table 4.1: Generation and processing of RNA-Seq reads for the Rd strain.

RNA-Seq experiment	Sample group	Replicate	Total number of reads	Percentage of reads mapped to reference (%)
Stationary phase	Mid-exponential phase	1	8,513,820	99.73
		2	8,460,966	99.71
		3	8,054,466	99.71
	Stationary phase	1	9,577,404	99.67
		2	7,739,026	99.61
		3	6,922,290	99.69
Oxidative stress	- H_2O_2	1	9,865,746	99.28
		2	8,554,276	99.09
		3	9,956,172	99.25
	+ H_2O_2	1	9,351,424	99.26
		2	9,592,742	99.36
		3	8,010,936	99.31
Iron-starvation stress	- 2,2-bipyridine	1	8,465,004	99.69
		2	9,407,392	99.65
		3	8,395,522	99.66
	+ 2,2-bipyridine	1	11,748,542	99.62
		2	12,553,076	99.54
		3	9,969,430	99.41
Nutritional stress	sBHI medium	1	30,430,040	99.57
		2	13,007,846	99.72
		3	13,214,318	99.69
	MIV medium	1	33,644,884	99.60
		2	18,747,718	99.73
		3	15,066,398	99.70

Table 4.2: Generation and processing of RNA-Seq reads for the R2866 strain.

RNA-Seq experiment	Sample group	Replicate	Total number of reads	Percentage of reads mapped to reference (%)
Stationary phase	Mid-exponential phase	1	9,352,464	99.71
		2	10,588,010	99.69
		3	9,166,440	99.69
	Stationary phase	1	9,615,326	99.61
		2	9,387,932	99.64
		3	12,085,666	99.48
Oxidative stress	- H_2O_2	1	9,016,922	99.79
		2	8,098,468	99.05
		3	10,954,020	99.75
	+ H_2O_2	1	9,581,676	99.76
		2	5,566,528	99.70
		3	11,615,748	99.52
Iron-starvation stress	- 2,2-bipyridine	1	4,157,900	99.41
		2	4,822,644	99.45
		3	4,111,496	99.38
	+ 2,2-bipyridine	1	4,158,598	99.38
		2	4,318,270	99.00
		3	3,878,952	99.08

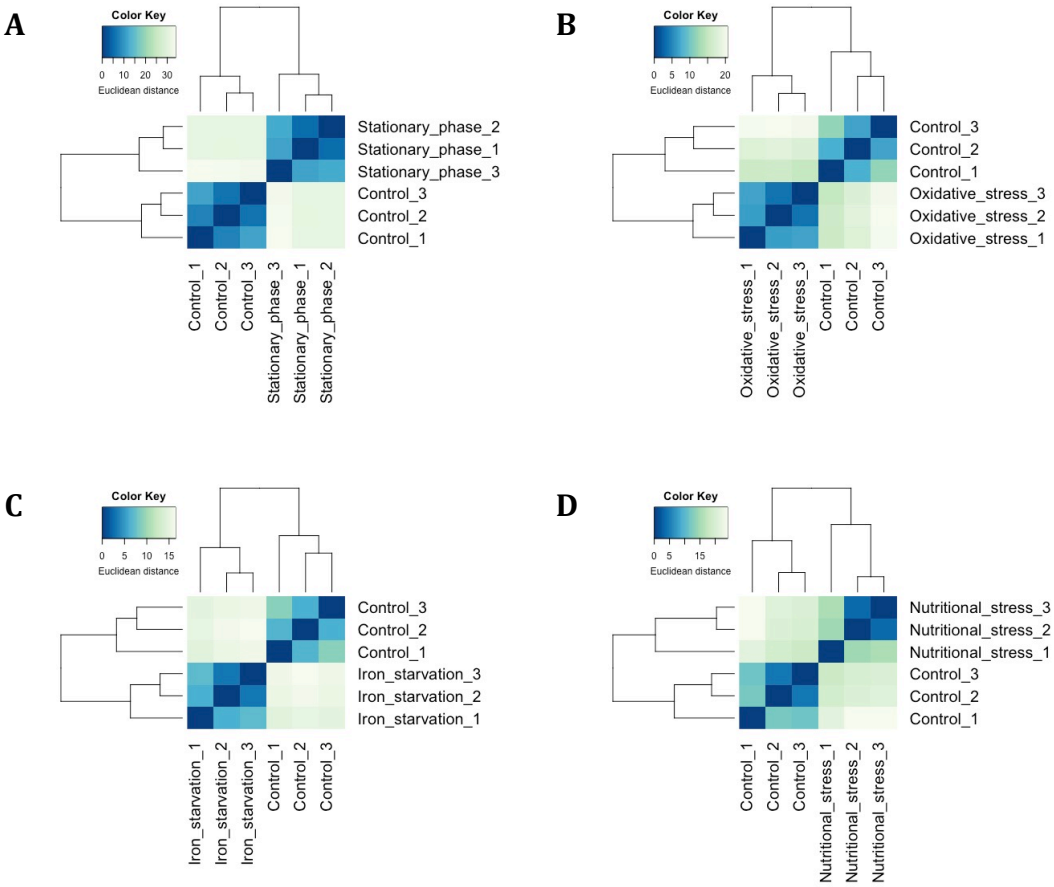

Figure 4.12: Quality control of Rd replicates used in the RNA-Seq experiment. Samples were clustered based on the euclidean distance calculated based on the $\log_2$-transformed gene expression data from each sample.

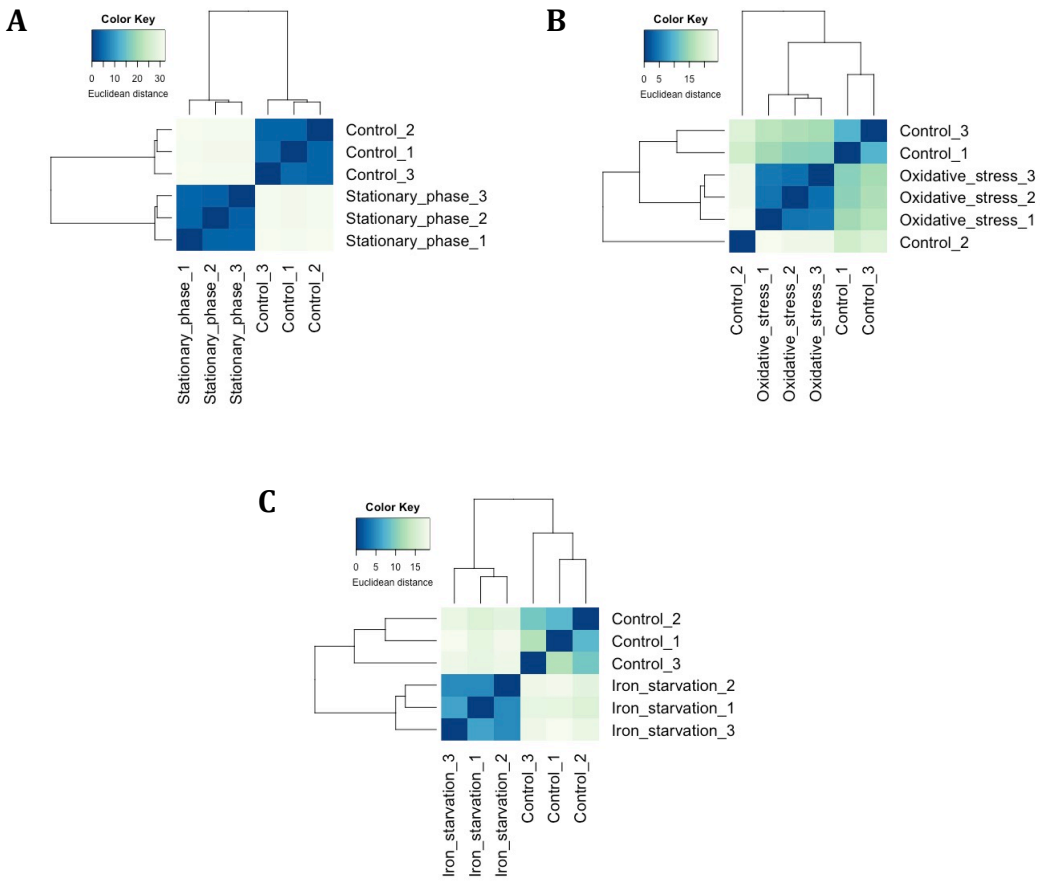

Figure 4.13: Quality control of R2866 replicates used in the RNA-Seq experiment. Samples were clustered based on the euclidean distance calculated based on the $\log_2$-transformed gene expression data from each sample.

4.2.2.4 Differential gene expression of *H. influenzae* during infection-relevant conditions

Following sequencing and mapping of RNA-Seq reads, differentially expressed genes were identified for each infection-relevant condition. As it is not possible to cover all genes in the dataset, the following sections will focus on important findings, with the emphasis on characterised and putative gene clusters and operons as well as genes with known and predicted roles in stress response and pathogenesis. Differentially expressed genes were also analysed for the enrichment of functional groups (GO terms and KEGG pathways), the most important of which will be presented in the following sections.

4.2.2.4.1 Transcriptional behaviour of Rd and R2866 strains during stationary growth phase

A total of 661 genes were differentially expressed in the Rd strain at stationary phase in the sBHI medium when compared to mid-exponential phase (see Appendix B). Of these, 324 were up-regulated and 337 were down-regulated. Similarly, 709 genes were differentially expressed in the R2866 strain at stationary phase, 356 of which were up-regulated and 353 were down-regulated. As described in detail below, there was induction of various metabolic pathways, oxidative stress response and iron acquisition. In particular, up-regulation of amino acid synthesis, coupled with reduction in protein biosynthesis and respiratory pathways, suggests that a classic stringent response was induced in both strains during stationary phase (Durfee et al., 2008).

4.2.2.4.1.1 Metabolic response

Multiple genes associated with metabolic pathways were up-regulated during stationary phase in both Rd and R2866. This was also evident from the majority of enriched functional groups being associated with metabolism in both strains (see Figures 4.14; 4.15). There was increase in expression of genes involved in pentose and hexose sugar metabolism, including the L-fucose operon, *fucAIKPU*, the D-ribose transport and utilisation locus, *rbsABCDKR* as well as a putative

galactose metabolism locus, *galKMT* (see Appendix B). Other up-regulated genes involved in sugar and amino acid metabolism were a putative glycogen biosynthesis and processing locus, *glgABCX,* as well as several genes with predicted roles in glycerol metabolism. The whole putative histidine biosynthesis operon was up-regulated in Rd, though only some members of this operon were induced in R2866 (see Figure 4.16).

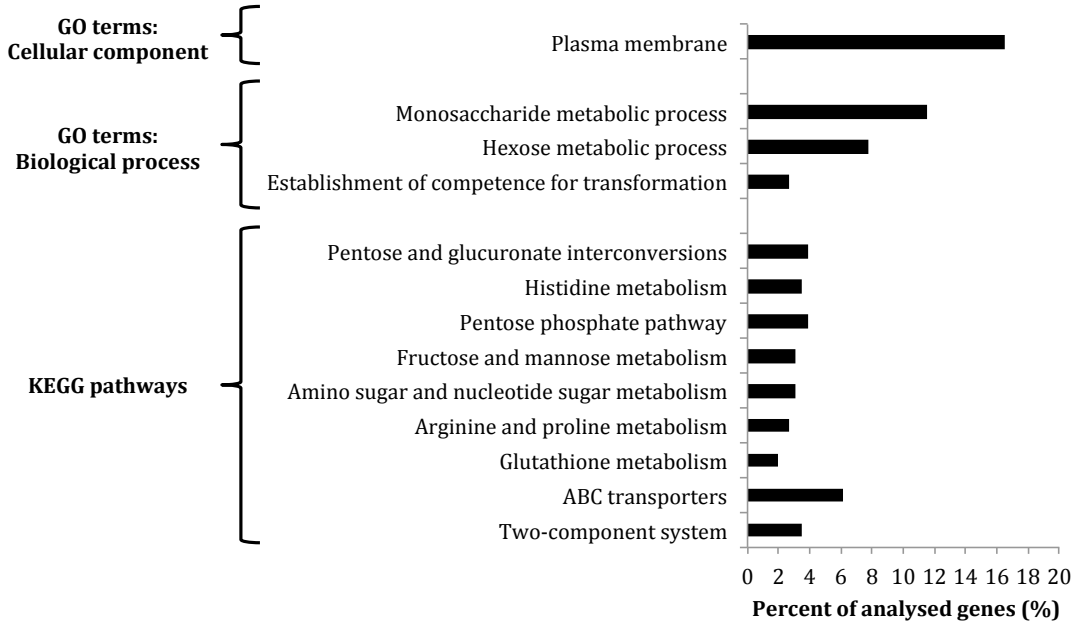

Figure 4.14: Functional group enrichment analysis of up-regulated genes in the Rd strain during growth at stationary phase, compared to growth at mid-exponential phase. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of up-regulated genes that belonged to each functional group.

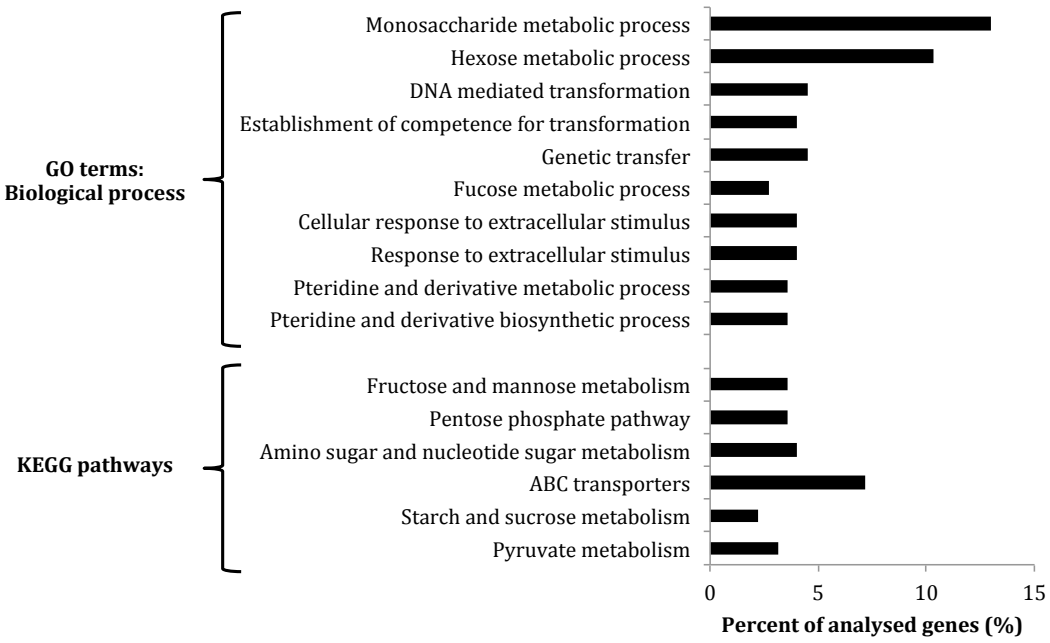

Figure 4.15: Functional group enrichment analysis of up-regulated genes in the R2866 strain during growth at stationary phase, compared to growth at mid-exponential phase. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of up-regulated genes that belonged to each functional group.

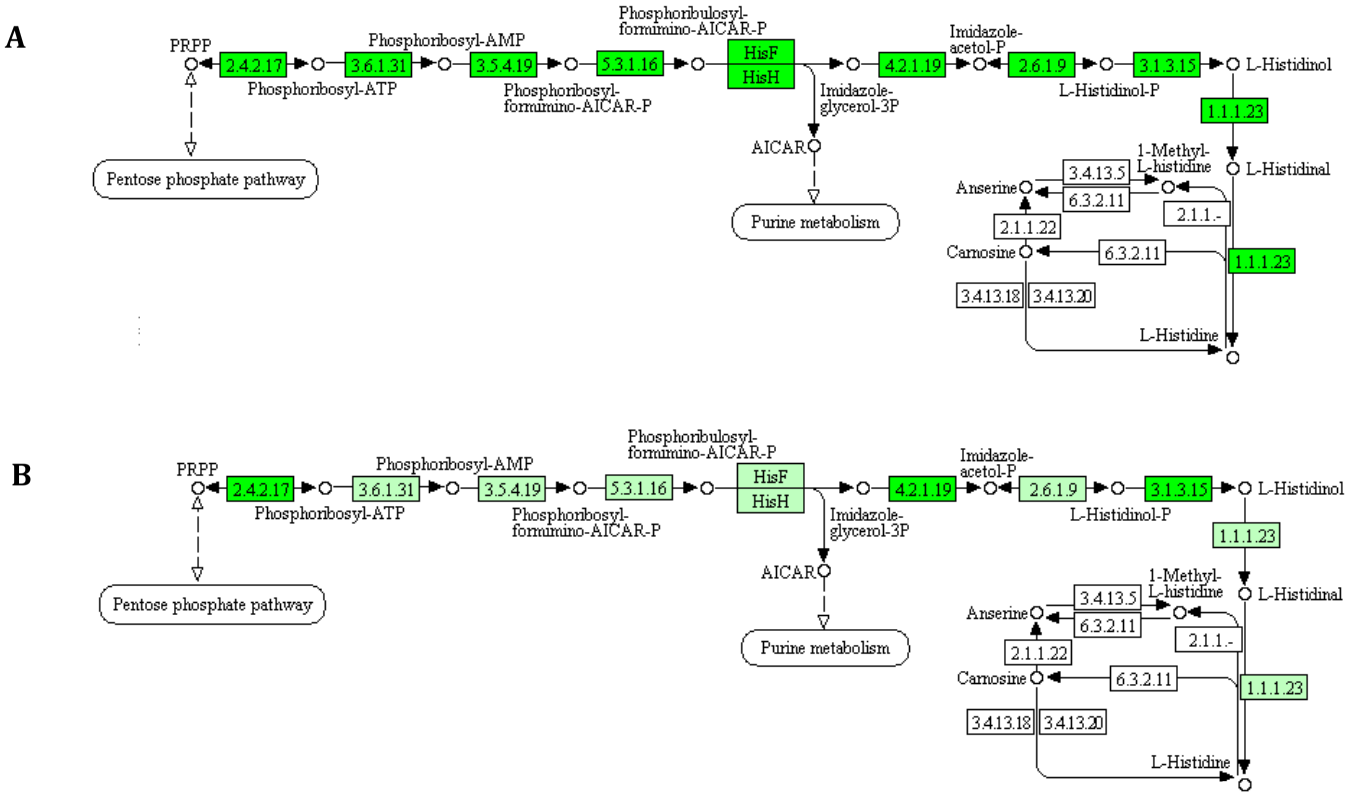

Figure 4.16: Induction of the histidine biosynthesis pathway in Rd and R2866 strains during stationary phase. KEGG diagrams depict genes up-regulated in the histidine biosynthesis pathway in Rd (A) and R2866 (B). Light green boxes represent genes present in *H. influenzae*, while darker green boxes show up-regulated genes.

Several amino acid and sugar transport genes were also induced, including *oppA* and *oppB* genes, encoding oligopeptide transporter proteins, and a transporter system, *afuABC* (see Appendix B). Despite being annotated as a ferric uptake system in Rd and R2866, the highly conserved bacterial gene locus *afuABC* was recently shown to be responsible for the transport of sugar-phosphates (Sit et al., 2015). The L-ascorbate utilisation locus, absent from R2866, was also induced in the Rd strain. In addition, there was up-regulation of the *atoABDE* operon, homologues of which encode proteins responsible for the catabolism of short-chain fatty acids in *E. coli* (Pauli and Overath, 1972).

A gene in R2866 with the highest level of up-regulation (over 123-fold) was another metabolic gene *tnaA*, encoding a tryptophanase enzyme (see Table 4.3). The *tnaB* gene, encoding a tryptophan permease protein, was the third most highly up-regulated gene in R2866. A gene *speF*, encoding an ornithine decarboxylase enzyme, was the second most highly up-regulated gene in the Rd strain (see Table 4.4). This enzyme is involved in the biosynthesis of putrescine from L-ornithine and is co-transcribed with a gene *potE*, encoding a putrescine-ornithine antiporter, which was also up-regulated in both strains (Kashiwagi et al., 1992) (see Appendix B). In contrast, genes from a putative spermidine and putrescine uptake locus, *potABCD*, were down-regulated in both (Furuchi et al., 1991).

Table 4.3: The ten most highly up-regulated and down-regulated genes in the R2866 strain during growth at stationary phase, compared to mid-exponential phase.

Gene	Product	Fold change	Fold change in Rd
Up-regulated genes			
tnaA	Tryptophanase	123.37	-
-	Alkylhydroperoxidase AhpD family core domain protein	56.49	28.79
tnaB	Tryptophan permease	32.04	-
afuA	Ferric transport system AfuABC; periplasmic-binding protein component	28.36	8.43
hxuC	Haem-hemopexin utilization protein C	27.20	-
hxuB	Haem-hemopexin utilization protein B	23.48	-
hxuA	Haem-hemopexin utilization protein A	20.61	2.14
yjiG	Inner membrane protein YjiG	20.30	-
-	Sporulation integral membrane protein YlbJ	17.38	-
-	Putative peptidase	14.64	-
Down-regulated genes			
rpL29	50S ribosomal protein L29	23.53	12.64
artP	Arginine ABC transporter; ATP-binding protein ArtP	23.51	82.19
rpL16	50S ribosomal protein L16	22.41	12.12
rpS17	30S ribosomal protein S17	20.09	10.40
rpS3	30S ribosomal protein S3	18.93	10.91
rpL7	50S ribosomal protein L7/L12	18.41	17.69
rpL22	50S ribosomal protein L22	17.49	11.06
artM	Arginine ABC transporter; permease protein ArtM	16.75	22.57
rpS19	30S ribosomal protein S19	16.64	10.44
potD	Spermidine/putrescine ABC transporter; periplasmic-binding protein	16.20	26.39

Table 4.4: The ten most highly up-regulated and down-regulated genes in the Rd strain during growth at stationary phase, compared to mid-exponential phase.

Gene	Product	Fold change	Fold change in R2866
Up-regulated genes			
-	Alkylhydroperoxidase AhpD family core domain protein	28.79	56.49
speF	Ornithine decarboxylase	22.04	7.69
-	RarD protein	19.86	4.00
fucI	L-fucose isomerase	18.19	9.14
dps	DNA protection during starvation protein	13.55	12.87
potE	Putrescine transporter	11.39	4.16
glpK	Glycerol kinase	10.59	11.21
fucR	L-fucose operon activator	10.27	5.17
-	DNA polymerase V subunit UmuD	9.75	-
glpF_1	Glycerol uptake facilitator protein	9.54	11.60
Down-regulated genes			
artP	Arginine transporter ATP-binding protein	82.19	23.51
artI	Arginine ABC transporter substrate-binding protein	40.84	13.75
potD_2	Spermidine/putrescine ABC transporter substrate-binding protein	26.39	16.20
deaD	ATP-dependent RNA helicase	23.54	8.57
artM	Arginine transporter permease subunit ArtM	22.57	16.75
cca	Multifunctional tRNA nucleotidyl transferase	18.02	14.88
rplL	50S ribosomal protein L7/L12	17.69	18.41
trpE	Anthranilate synthase component I	16.50	-
artQ	Arginine transporter permease subunit ArtQ	16.36	11.50
nrfA	Cytochrome c552	13.93	2.30

Among other metabolism-associated gene loci, down-regulated during stationary phase, was the *artIMPQ* operon, which is known to be involved in arginine uptake in *E. coli* (see Appendix B) (Wissenbach et al., 1995). Genes in this operon were all down-regulated at stationary phase in both Rd and R2866. Gene *artP,* encoding an arginine transporter ATP-binding protein, was the most highly down-regulated gene in the Rd strain (over 82-fold) (see Table 4.4).

4.2.2.4.1.2 Induction of competence

There was induction of competence-related genes during stationary phase in both strains. This was evident from up-regulation of the operon *comABCDEF*, which has a characterised role in competence and transformation induction in *H. influenzae* (see Appendix B) (Tomb et al., 1991). The expression of two genes, *rec-2* and *tfoX*, with characterised roles in DNA transformation, was also induced (Barouki and Smith, 1985, Zulty and Barcak, 1995). There was also induction of genes *radA* and *recN,* homologues of which are involved in DNA repair and general stress response in *E. coli* (Finch et al., 1985, Beam et al., 2002).

4.2.2.4.1.3 Oxidative stress response

Multiple genes involved in oxidative stress response were up-regulated in both strains during stationary phase. The most highly up-regulated gene (over 28-fold) in the Rd strain during stationary phase encoded an alkylhydroperoxidase family core domain protein, AhpD (see Table 4.4). It was also the second most highly up-regulated gene in R2866 (over 56-fold) (see Table 4.3). This gene was annotated as a hypothetical protein in the original Rd and R2866 published genome sequences. The amino acid sequence search in the InterPro database revealed this protein to contain a 50 amino acid long alkylhydroperoxidase AhpD core domain.

Other up-regulated genes in Rd and R2866 with known roles in oxidative stress response were a catalase gene, *hktE*, a peroxiredoxin gene, *pgdX*, and the *dps* gene (see Table 4.4) (see Appendix B) (see sections 1.1.8.1). The expression of

several other genes with putative roles in oxidative stress was induced either in Rd or R2866. Interestingly, a two-component system gene, *arcA*, with a characterised role in oxidative stress, was down-regulated in Rd only (Wong et al., 2007). In addition, a putative peroxiredoxin gene, *ahpC*, absent in Rd, was highly down-regulated (7.6-fold) in the R2866 strain.

4.2.2.4.1.4 Induction of iron acquisition

There were multiple up-regulated genes with putative and known roles in bacterial iron acquisition during stationary phase. This was especially evident in the R2866 strain, where four iron-related genes were among the ten most highly up-regulated genes (see Table 4.3). This included the *hxuABC* operon, which encodes haem utilisation proteins. The whole operon was up-regulated over 20-fold in R2866, though only the *hxuA* gene was up-regulated in Rd (see Appendix B).

The expression of *tbp1* and *tbp2* genes, encoding characterised transferrin-binding proteins, was highly induced (over 10-fold) in R2866, whereas only the *tbp1* gene was up-regulated in Rd (Gray-Owen and Schryvers, 1995). In contrast, there were two genes, *ftnA1* and *ftnA2*, coding for putative ferritin proteins, which were down-regulated in R2866. Putative metal transport-associated genes, *yfeA* and *yfeB,* were both induced in R2866, though only the latter was up-regulated in Rd (see Appendix B). There also appeared to be induction of iron-sulphur cluster formation during stationary phase in Rd and R2866. This was inferred from the up-regulation of genes *hscAB*, *fdx-1,* and *iscARSU*, which have putative roles in the formation of iron-sulphur clusters.

4.2.2.4.1.5 Reduction in protein biosynthesis and export

There was an apparent reduction in protein biosynthesis during stationary phase in Rd and R2866, as inferred from enriched functional groups (see Figures 4.17; 4.18). This was also evident from the down-regulation of a large number of ribosomal protein genes as well as genes coding for elongation and translation initiation factors and multiple tRNAs (see Figure 4.19) (see

Appendix B). The most highly down-regulated gene in R2866 was *rpL29*, encoding a 50S ribosomal protein L29 (see Table 4.3). Coupled with an increase in amino acid synthesis, this suggests that a classic stringent response was induced in *H. influenzae* during stationary phase. In addition, there was down-regulation of genes involved in protein export, including the Sec pathway in both strains and twin-arginine protein export genes in Rd only.

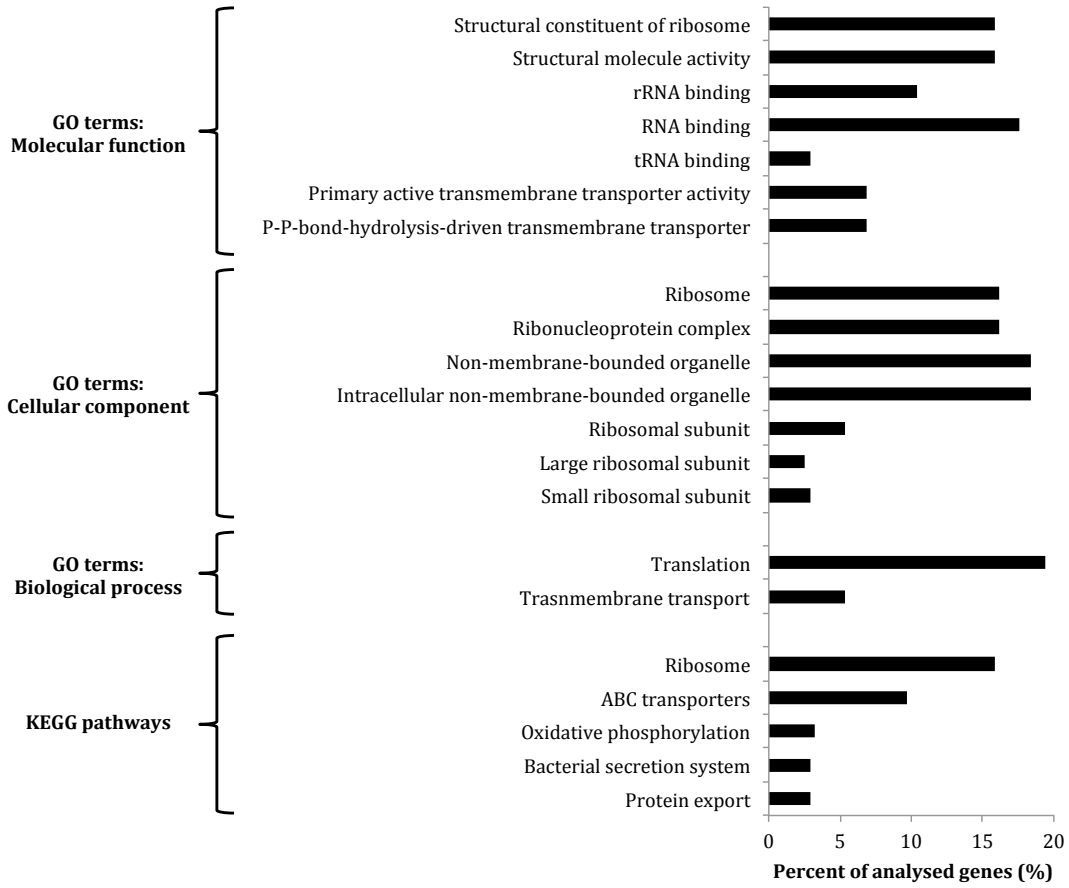

Figure 4.17: Functional group enrichment analysis of down-regulated genes in the Rd strain during growth at stationary phase, compared to growth at mid-exponential phase. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of down-regulated genes that belonged to each functional group.

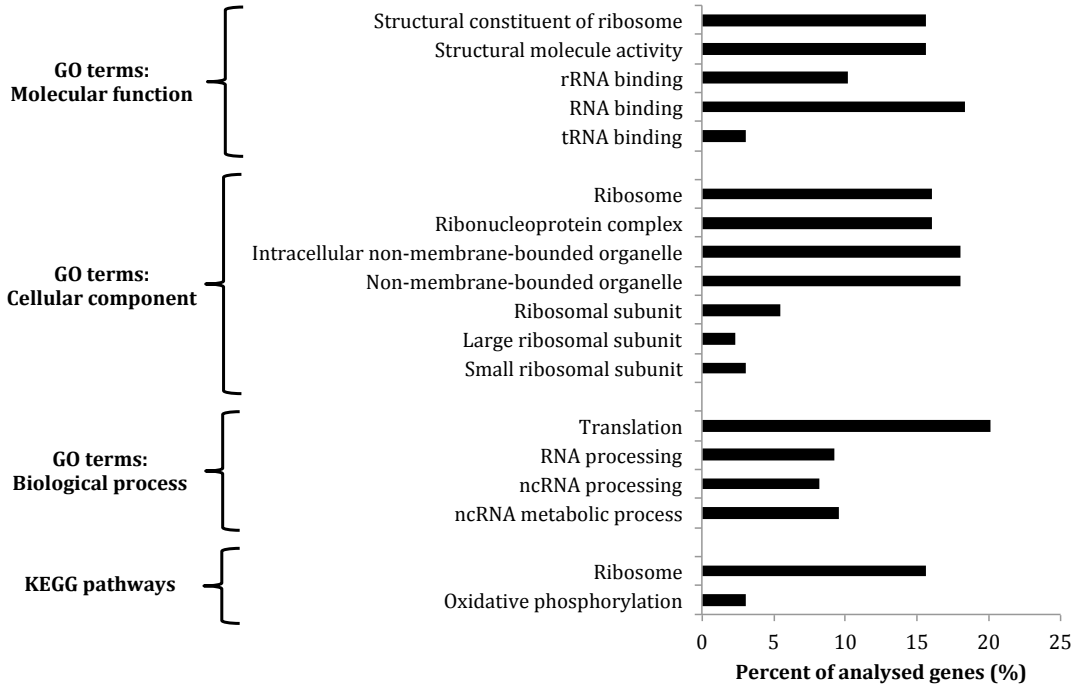

Figure 4.18: Functional group enrichment analysis of down-regulated genes in the R2866 strain during growth at stationary phase, compared to growth at mid-exponential phase. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of down-regulated genes that belonged to each functional group.

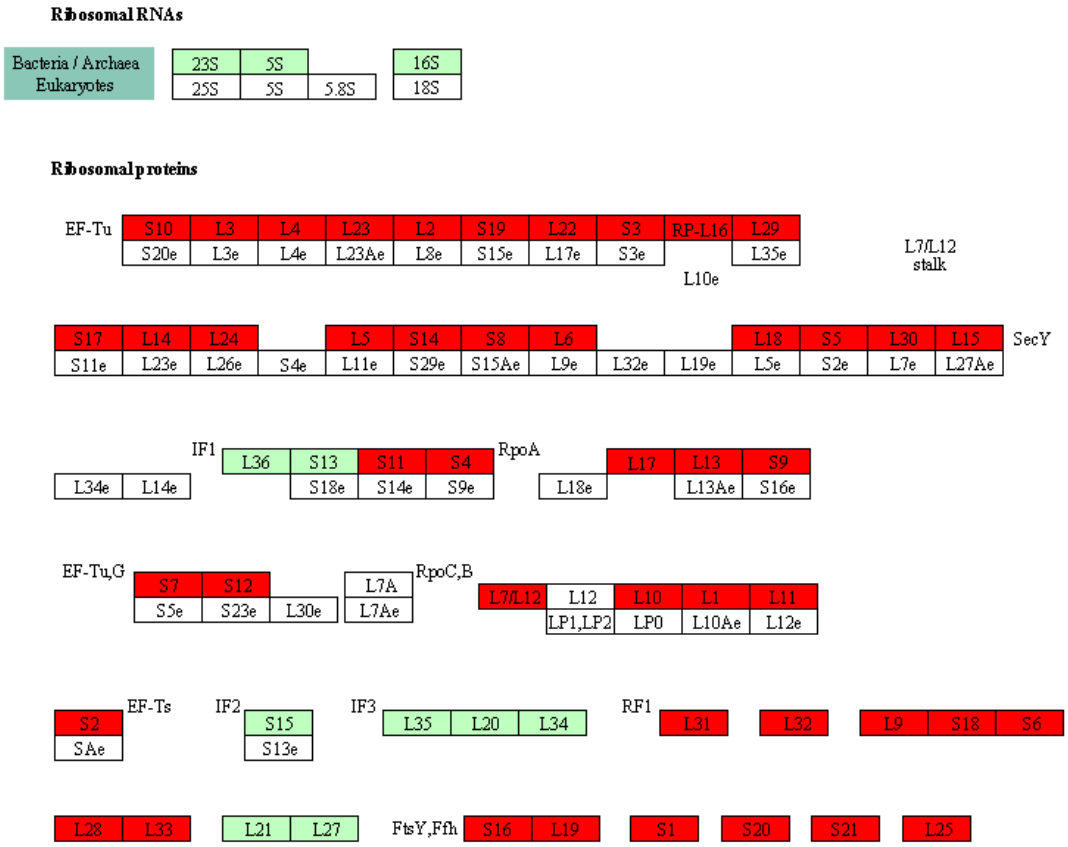

Figure 4.19: Down-regulation of ribosomal protein genes in *H. influenzae* during stationary phase. The KEGG diagram depicts down-regulated ribosomal protein genes in the R2866 strain. Light green boxes represent genes present in *H. influenzae*, while red boxes show down-regulated genes.

4.2.2.4.1.6 Reduction in bacterial respiration

Several gene loci with putative roles in bacterial respiration were down-regulated in Rd and R2866 during stationary phase. These included a putative ATP synthase operon, *atpABCDEFGHI*, a putative electron transport complex, as well as genes *cydA* and *cydB*, encoding predicted cytochrome oxidase subunits (see Appendix B). The *nqrABCDEF* operon, encoding subunits of a sodium-transport NADH-quinone reductase, was also down-regulated in Rd and R2866 (Hayashi et al., 1996).

4.2.2.4.1.7 Differential expression of mobile genetic elements and other notable genes

Multiple genes from mobile genetic elements were differentially expressed during stationary phase. Among these were up-regulated genes from the ICE and a putative prophage in R2866 as well as down-regulated genes from the FluMu prophage in Rd (see Appendix B). Among other notable differentially expressed genes was a putative heat-shock RNA polymerase sigma factor-32 gene, *rpoH*, up-regulated in Rd only, as well as putative chaperone genes, *groEL* and *groES,* up-regulated in both strains. Members of a type IV pilus gene cluster, *pilA*, *pilB* and *pilD*, along with genes from another major pilus gene locus, *hifABCDE*, were up-regulated in R2866 only (Mhlanga-Mutangadura et al., 1998, Bakaletz et al., 2005). Furthermore, the expression of two genes, *siaA* and *lsgB*, involved in the processing of LOS in *H. influenzae,* was induced in R2866 (Abu Kwaik et al., 1991, Jones et al., 2002). Finally, a characterised toxin-antitoxin locus, *vapBC1*, was down-regulated in R2866 only (Daines et al., 2007).

4.2.2.4.2 Transcriptional response of Rd and R2866 strains to oxidative stress

A total of 150 genes were differentially expressed in the Rd strain during oxidative stress compared to normal growth at mid-exponential phase (see Appendix B). Of these, 53 were up-regulated and 97 were down-regulated. For the R2866 strain, there were 96 genes that were differentially expressed during oxidative stress, when compared to normal growth at mid-exponential phase,

69 of which were up-regulated and 27 were down-regulated. As described in more detail below, there was induction of oxidative stress and SOS responses as well as differential regulation of metabolic and iron-uptake pathways.

4.2.2.4.2.1 Induction of oxidative stress response

There was a clear induction of oxidative stress response in both strains. The catalase gene, *hktE,* was the most highly up-regulated gene in both Rd and R2866 (see Tables 4.5; 4.6). Other up-regulated genes with well-characterised roles in oxidative stress were a glutathione-dependent peroxidase gene, *pgdX*, a 6-phosphogluconate dehydrogenase gene, *gnd,* as well as genes *pntA* and *pntB,* encoding proteins with homology to subunits of NAD(P) transhydrogenase in *E. coli* (see Appendix B) (Clarke and Bragg, 1985, Yoon et al., 1989, Pauwels et al., 2003). The last two genes were not up-regulated in the R2866 strain. The *dps* gene, encoding a ferritin-like protein with a characterised role in oxidative stress in *H. influenzae*, was up-regulated in both strains over 10-fold (see section 1.1.8.1). The oxidative stress response of Rd and R2866 was further supported by the enrichment of functional groups related to the response to stress (see Figures 4.20; 4.21).

Table 4.5: The ten most highly up-regulated and down-regulated genes in the Rd strain during oxidative stress, compared to normal growth.

Gene	Product	Fold change	Fold change in R2866
Up-regulated genes			
hktE	Catalase	76.94	111.22
acpD	Acyl carrier protein phosphodiesterase	26.07	11.36
-	DoxX	14.58	3.68
ilvC	Ketol-acid reductoisomerase	11.48	-
dps	DNA protection during starvation protein	10.31	13.68
-	DNA polymerase V subunit UmuD	10.31	13.26
-	Peroxiredoxin hybrid Prx5	7.56	8.30
ilvI	Acetolactate synthase 3 catalytic subunit	4.88	-
recN	DNA repair protein	4.68	8.78
ilvH	Acetolactate synthase 3 regulatory subunit	4.52	-
Down-regulated genes			
artP	Arginine transporter ATP-binding protein	21.55	-
dmsA_3	Anaerobic dimethyl sulfoxide reductase subunit A	11.89	-
artI	Arginine ABC transporter substrate-binding protein	11.01	-
artQ	Arginine transporter permease subunit ArtQ	9.12	-
dmsB	Anaerobic dimethyl sulfoxide reductase subunit B	7.98	-
nrfC	Nitrite reductase Fe-S protein	7.75	-
artM	Arginine transporter permease subunit ArtM	7.63	-
nrfB	Cytochrome c nitrite reductase pentaheme subunit	7.50	-
-	Twin-argninine leader-binding protein DmsD	7.24	-
nrfA	Cytochrome c552	6.89	-

Table 4.6: The ten most highly up-regulated and down-regulated genes in the R2866 strain during oxidative stress, compared to normal growth.

Gene	Product	Fold change	Fold change in Rd
Up-regulated genes			
hktE	Catalase	111.22	76.94
-	Hypothetical protein	26.61	2.57
-	Hypothetical protein	19.57	-
-	Putative NAD-dependent protein deacetylase	17.02	-
-	Hypothetical protein	15.63	-
dpsA	DPS ferritin-like protein	13.68	10.31
-	DNA polymerase V subunit UmuD	13.26	10.31
-	Putative 5'(3')-deoxyribonucleotidase	12.87	-
azoR	FMN-dependent NADH-azoreductase	11.36	26.07
-	Hypothetical protein	10.23	-
Down-regulated genes			
Hgd	2-(hydroxymethyl)glutarate dehydrogenase	7.82	3.12
-	Hypothetical protein	6.52	-
ygbM	Putative hydroxypyruvate isomerase YgbM	6.11	2.92
ygbL	Putative sugar aldolase/epimerase	5.98	2.91
-	Putative sugar epimerase	4.00	2.54
-	Putative permease	3.18	2.37
ftnA2	Ferritin protein A2	2.67	-
-	Hypothetical protein	2.57	-
galR	Galactose operon regulator	2.53	3.30
rpS18	30S ribosomal subunit protein S18	2.41	-

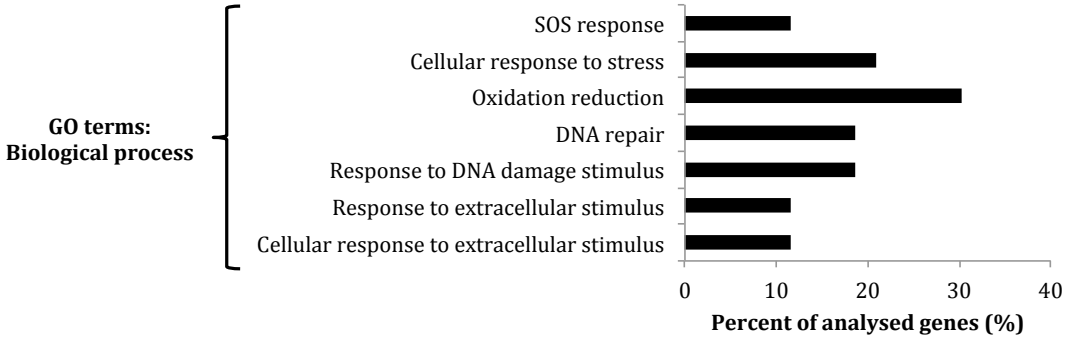

Figure 4.20: Functional group enrichment analysis of up-regulated genes in the Rd strain during oxidative stress, compared to normal growth. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of up-regulated genes that belonged to each functional group.

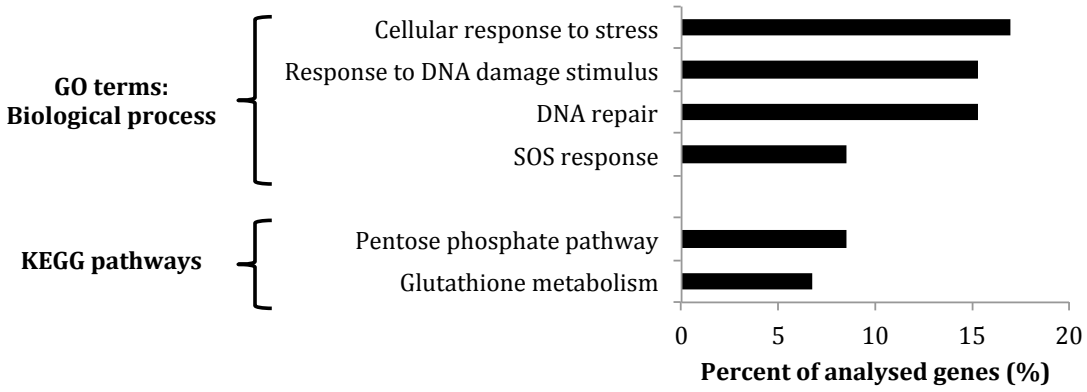

Figure 4.21: Functional group enrichment analysis of up-regulated genes in the R2866 strain during oxidative stress, compared to normal growth. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of up-regulated genes that belonged to each functional group.

Among other highly up-regulated genes with putative roles in oxidative stress response was a gene annotated as *azoR* in R2866 and *acpD* in Rd (see Appendix B). It encodes a protein homologous to an azoreductase enzyme in *E. coli*, with a characterised role in thiol-specific stress (Liu et al., 2009). In addition, there were other multiple up-regulated genes with putative roles in oxidative stress response. In contrast, a two-component system gene, *arcA*, with a role in oxidative stress in *H. influenzae*, was down-regulated in Rd, while a peroxiredoxin gene, *ahpC,* was down-regulated in R2866 (Wong et al., 2007).

4.2.2.4.2.2 SOS response

There were multiple up-regulated genes involved in DNA protection and repair, reflected by several enriched functional groups associated with DNA repair and response to DNA damage (see Figures 4.20; 4.21). Among these were genes *recA*, *recN*, *recX*, and *radA*, encoding proteins with characterised roles in DNA repair in *H. influenzae* and *E. coli* (see Appendix B) (Finch et al., 1985, Setlow et al., 1988, Beam et al., 2002, Pages et al., 2003). There was also up-regulation of a small putative gene locus, *ruvAB,* in both strains. It codes for proteins involved in Holliday junction processing in *E. coli*, with roles in DNA repair (Iwasaki et al., 1989). The *lexA* gene was up-regulated over 3-fold in both Rd and R2866. It encodes a homologue of the *E. coli* LexA protein, which is a major repressor of the SOS regulon (Little et al., 1981).

4.2.2.4.2.3 Induction of iron uptake

The expression of several iron-associated genes was induced in both Rd and R2866 during oxidative stress. These included several genes with predicted roles in iron-sulphur cluster formation and iron acquisition (see Appendix B). The expression of the haem utilisation locus, *hxuABC,* as well as *tbp1* and *tbp2* genes, coding for transferrin-binding proteins, was induced in R2866 only. Other iron-associated genes that were up-regulated in R2866 only were *hitA*, part of the iron acquisition operon, *hitABC*, and *hemR*, encoding a haemin receptor. In contrast, there was down-regulation of two putative ferritin-like

genes, *ftnA1* and *ftnA2,* in R2866 only. Several other putative iron-associated genes were down-regulated in Rd only.

4.2.2.4.2.4 Metabolic response during oxidative stress

There seemed to be an overall difference in gene down-regulation during oxidative stress between Rd and R2866 strains. The whole arginine-uptake gene locus, *artIMPQ,* was among the ten most highly down-regulated genes in Rd, while it was not down-regulated in R2866 at all (see Table 4.5). The most highly down-regulated gene in R2866 during oxidative stress was *Hgd*, encoding a putative 2-(hydroxymethyl)glutarate dehydrogenase enzyme (see Table 4.6). This gene was part of a six-gene locus, which was down-regulated over 3-fold in R2866 and to a lesser extent in Rd (see Appendix B). The locus encoded putative enzymes involved in the processing of carbohydrates.

Several other metabolism-associated gene clusters were differentially expressed during oxidative stress. Genes *ilvC*, *ilvI* and *ilvH*, with putative roles in the biosynthesis of leucine, isoleucine and valine amino acids, were highly up-regulated in the Rd strain only (Ricca et al., 1988). Among other notable genes highly down-regulated in Rd only were a putative DMSO reductase gene locus, *dmsABC,* and a nitrite reductase gene locus, *nrfABCD* (see Appendix B). In addition, there was down-regulation of several ribosomal protein genes in both strains.

4.2.2.4.3 Transcriptional response of Rd and R2866 strains to iron-starvation stress

A total of 175 genes were differentially expressed in the Rd strain during iron-starvation stress when compared to normal growth at mid-exponential phase (see Appendix B). Of these, 90 genes were up-regulated and 85 were down-regulated. 188 genes were differentially expressed in R2866 during iron-starvation stress when compared to normal growth at mid-exponential phase, of which 55 were up-regulated and 133 were down-regulated. As described

below, there was induction of iron acquisition, heat-shock response and LOS-associated pathways, while metabolic and transport pathways were reduced.

4.2.2.4.3.1 Induction of iron acquisition

There was a large number of iron transport-associated genes that were up-regulated in both strains, reflected by enriched functional groups related to metal and ion transport (see Figures 4.22; 4.23). However, this was more evident in the R2866 strain, where a larger number of such genes were induced. A gene encoding a putative TonB-dependent transport protein was the most highly up-regulated (35-fold) in R2866, but was only induced 3.5-fold in Rd (see Table 4.7) (see Appendix B). While the iron acquisition locus, *hitABC,* was up-regulated in both strains, the expression of the *hitA* gene in particular was induced 30.6-fold in the R2866 strain. A putative metal transporter gene locus, *yfeABCD,* and a haemin receptor gene, *hemR,* were up-regulated in both strains, though again to a higher extent in R2866.

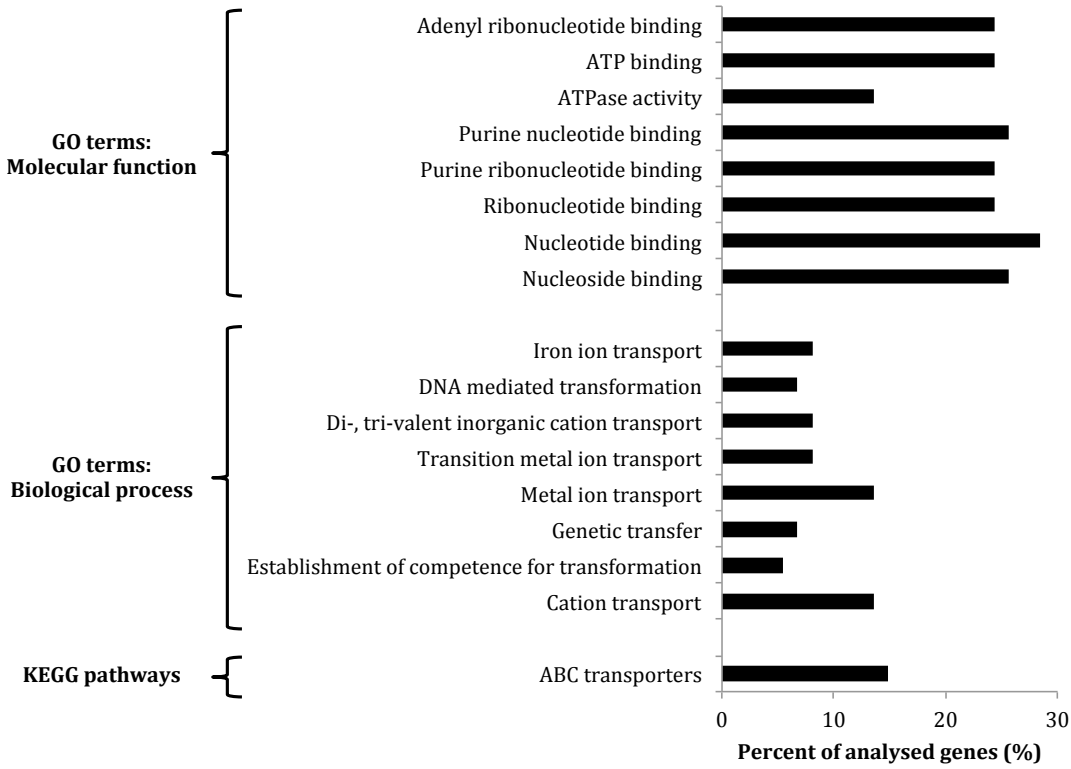

Figure 4.22: Functional group enrichment analysis of up-regulated genes in the Rd strain during iron-starvation stress, compared to normal growth. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of up-regulated genes that belonged to each functional group.

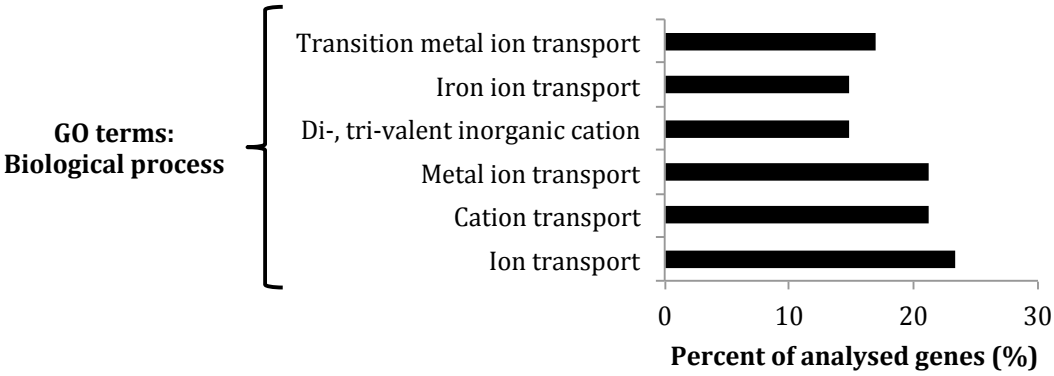

Figure 4.23: Functional group enrichment analysis of up-regulated genes in the R2866 strain during iron-starvation stress, compared to normal growth. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of up-regulated genes that belonged to each functional group.

Table 4.7: The ten most highly up-regulated and down-regulated genes in the R2866 strain during iron-starvation stress, compared to normal growth.

Gene	Product	Fold change	Fold change in Rd
Up-regulated genes			
-	Putative TonB-dependent transport protein	35.06	3.52
hxuC	Haem-hemopexin utilization protein C	30.77	-
hitA	Iron(III) ABC transporter periplasmic-binding protein	30.60	2.12
ompU1	Putative outer membrane protein OmpU1	23.72	2.87
hxuB	Haem-hemopexin utilization protein B	23.46	-
tbp2	Transferrin-binding protein 2	17.21	-
tbp1	Transferrin-binding protein 1	15.68	-
hxuA	Haem-hemopexin utilization protein A	15.01	-
copZ3_1	Copper chaperone protein	10.93	2.25
hitB	Iron(III) ABC transporter permease protein	10.83	2.13
Down-regulated genes			
ydjN	Putative transporter	10.68	17.82
metQ	DL-methionine transporter; periplasmic binding protein MetQ	9.13	9.96
metI	DL-methionine transporter; permease protein MetI	7.52	8.60
potD	Spermidine/putrescine ABC transporter; periplasmic-binding protein	7.09	3.01
metN	DL-methionine transporter; ATP binding protein MetN	7.01	7.72
artP	Arginine ABC transporter; ATP-binding protein ArtP	6.18	-
tcyA	L-cystine ABC transporter; periplasmic-binding protein TcyA	5.53	7.80
argH	Argininosuccinate lyase	5.24	-
fdnG_1	Formate dehydrogenase-N; major subunit	5.13	3.13
fdnI	Formate dehydrogenase-N; cytochrome B556(Fdn) gamma subunit; nitrate-inducible	4.66	2.75

Among other notable up-regulated iron-associated genes were the haem utilisation locus, *hxuABC,* and the *tonB* gene, known to play an important role in iron acquisition in *H. influenzae* (see Appendix B) (Jarosik et al., 1994, Jarosik et al., 1995). Other iron-related genes, induced in the R2866 strain only, were *tbp1* and *tbp2*, encoding transferrin-binding proteins, as well as *exbB* and *exbD*, coding for putative biopolymer transporter proteins (Morton et al., 1999). The latter two genes are homologous to and complement genes in *E. coli*, which encode TonB accessory proteins (Jarosik and Hansen, 1995). Several other putative metal transport and efflux pump genes were up-regulated during iron-starvation in both strains. Interestingly, some predicted metal-associated genes were down-regulated in both Rd and R2866, including a haem-binding lipoprotein gene, *hbpA* (annotated as *dppA* in Rd).

4.2.2.4.3.2 Induction of LOS-related genes

As with iron acquisition-associated gene clusters, the rest of the up-regulated genes also mostly differed between Rd and R2866. The most highly up-regulated gene in the Rd strain was *lptF*, encoding a putative lipopolysaccharide export system permease protein, which was not induced in R2866 at all (see Table 4.8). The second and third most highly up-regulated genes in Rd encoded putative proteins associated with LOS biosynthesis. Other LOS-related up-regulated genes in Rd included *lptA* and *lptC*, encoding putative lipopolysaccharide (LPS) export system proteins (see Appendix B). The latter was the only putative LOS-associated gene to be induced in the R2866 strain.

Table 4.8: The ten most highly up-regulated and down-regulated genes in the Rd strain during iron-starvation stress, compared to normal growth.

Gene	Product	Fold change	Fold change in R2866
Up-regulated genes			
lptF	Lipopolysaccharide export system permease protein LptF	22.71	-
-	Lipopolysaccharide biosynthesis protein	6.30	-
-	Lipopolysaccharide biosynthesis protein	5.50	-
-	Na(+)-translocating NADH-quinone reductase subunit E	4.95	-
ppc	Phosphoenolpyruvate carboxylase	4.92	-
merR2	Mercuric resistance operon regulatory protein	4.77	-
-	Branched chain amino acid ABC transporter substrate-binding protein	4.59	-
-	Hypothetical protein	4.59	-
-	Cobalt transport protein CbiM	4.55	-
-	Cobalt ABC transporter; permease protein CbiQ	4.37	-
Down-regulated genes			
-	Proton glutamate symport protein	17.82	10.68
hlpA	D-methionine-binding lipoprotein MetQ	9.96	9.13
pstA_1	Phosphate ABC transporter permease	8.60	7.52
-	Amino-acid ABC transporter ATP-binding protein	8.56	2.21
-	Amino acid ABC transporter substrate-binding protein	7.80	5.53
metN	DL-methionine transporter ATP-binding protein	7.72	7.01
-	Amino acid ABC transporter permease	7.04	2.41
dppA	Haem-binding lipoprotein	5.41	4.49
-	Long chain fatty acid CoA ligase	5.19	-
merT	Mercuric ion transport protein	4.69	3.67

4.2.2.4.3.3 Heat-shock response

Both Rd and R2866 switched on their heat-shock response during iron-starvation. This was apparent from the up-regulation of *dnaK*, *groES* and *groEL* genes, encoding chaperone proteins with roles in heat-shock response in *E. coli* (see Appendix B) (Fayet et al., 1989, Liberek et al., 1992). Another up-regulated gene locus with a putative role in heat-shock response was the ATP-dependent protease locus, *hslUV* (Kanemori et al., 1997).

4.2.2.4.3.4 Regulation of metabolism and transport

Many metabolic and transport-associated gene clusters were down-regulated in both strains during iron-starvation. The most highly down-regulated gene (17.8-fold) in the Rd strain encoded a putative proton glutamate symport protein (see Table 4.8). It was annotated as a putative transporter gene *ydjN* in the R2866 strain and was the most highly down-regulated gene (10.7-fold) in that strain as well (see Table 4.7). The *ydjN* gene was previously shown to be responsible for uptake of cystine in *E. coli* (Chonoles Imlay et al., 2015). In agreement with this, a putative L-cystine uptake locus, *tcyABC,* was down-regulated in both strains (see Appendix B).

Among other highly down-regulated gene loci, associated with metabolism and transport, was a putative DL-methionine ABC transporter locus, *metINQ* (see Appendix B) (Merlin et al., 2002). Genes *metI* and *metQ* were annotated as *ptsa_1* and *hlpA* respectively in the Rd strain. There was also down-regulation of putative formate dehydrogenase and nitrite reductase loci, as well as genes *ilvA*, *ilvC* and *ilvD,* with predicted roles in the biosynthesis of isoleucine and valine amino acids.

Among metabolic genes down-regulated in Rd only were members of the histidine operon, *hisABCDFGHIE*, putative tryptophan synthase subunit genes, *trpA* and *trpB,* as well as the *ilvHI* locus, encoding putative acetolactate synthase subunits (see Appendix B). Other metabolic and respiratory genes down-regulated in R2866 only were the putative arginine uptake operon, *artIMPQ*, the

tryptophanase locus, *tnaAB*, the putative fumarate reductase operon, *frdABCD* and genes from a putative ATP synthase locus, *atpABCDEFGHI*. There were also multiple down-regulated ribosomal protein genes in the R2866 strain, reflected by a large number of enriched functional groups being associated with ribosomal subunits and the translation process (see Figure 4.24). Among other notable genes down-regulated only in R2866 was a putative peroxiredoxin gene, *ahpC.*

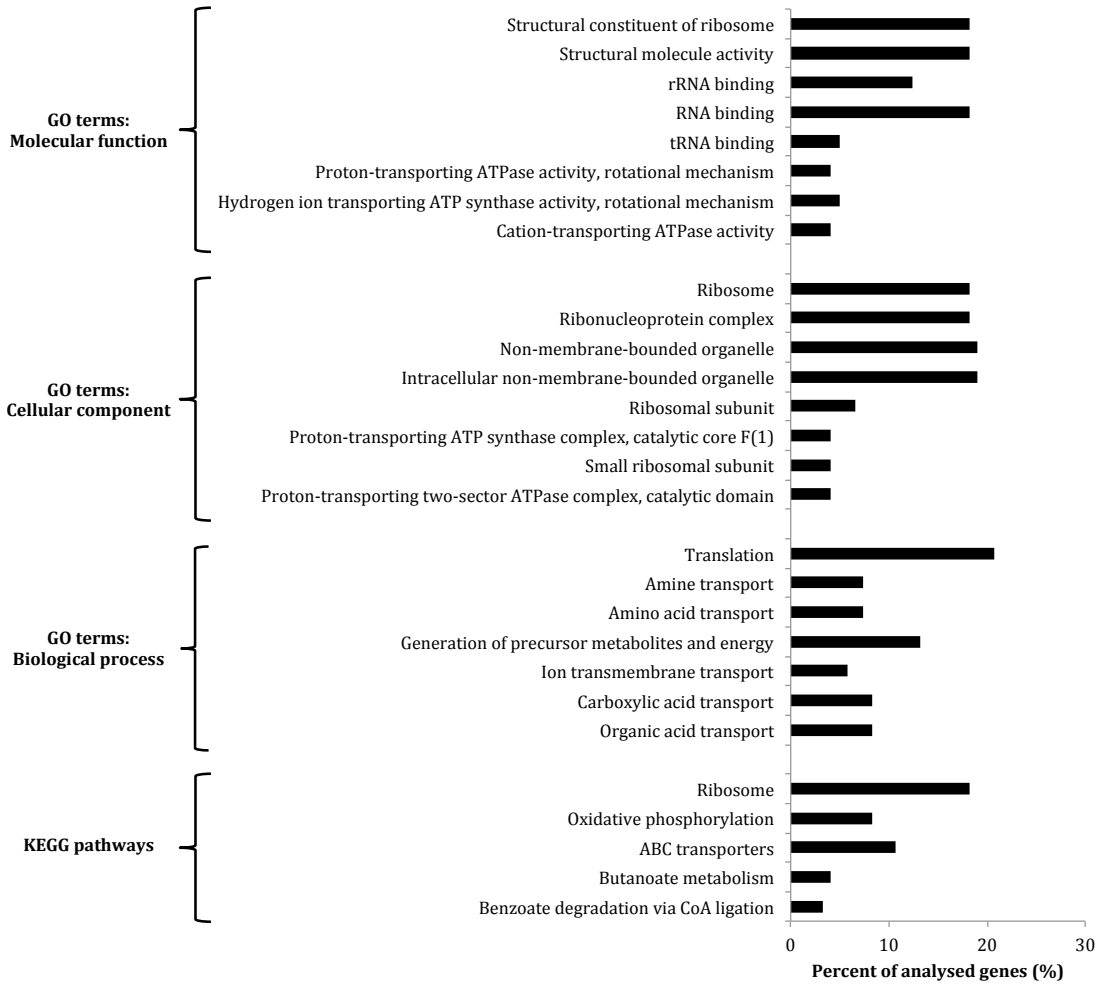

Figure 4.24: Functional group enrichment analysis of down-regulated genes in the R2866 strain during iron-starvation stress, compared to normal growth. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of down-regulated genes that belonged to each functional group.

4.2.2.4.4 Transcriptional response of the Rd strain to nutritional stress

A total of 231 genes were differentially expressed in the Rd strain during growth in the chemically-defined MIV medium, compared to growth in the rich sBHI medium (see Appendix B). Of these, 135 genes were up-regulated and 96 were down-regulated. As described in detail below, multiple metabolic and iron-associated pathways, along with competence and oxidative stress response, were up-regulated, while purine biosynthesis and transport pathways were reduced.

4.2.2.4.4.1 Metabolic and iron-associated response

There was up-regulation of multiple metabolic pathways as well as iron transport-associated gene clusters, reflected by the enriched functional groups (see Figure 4.25). Genes *sdaA* and *sdaC*, encoding putative L-serine deaminase and serine transporter proteins respectively, were among the most highly up-regulated genes (see Table 4.9). In addition, there was up-regulation of genes with putative roles in glycogen processing, glycerol utilisation, sugar-phosphate transport and oligopeptide transport (see Appendix B). Putative down-regulated metabolic gene clusters included formate dehydrogenase and arginine uptake loci. Among induced iron transport-related genes were the iron-acquisition complex, *hitABC*, members of a putative cytochrome c synthesis and haem transport locus, *ccmABCDEFGH,* as well as a gene encoding an iron-associated protein, TonB.

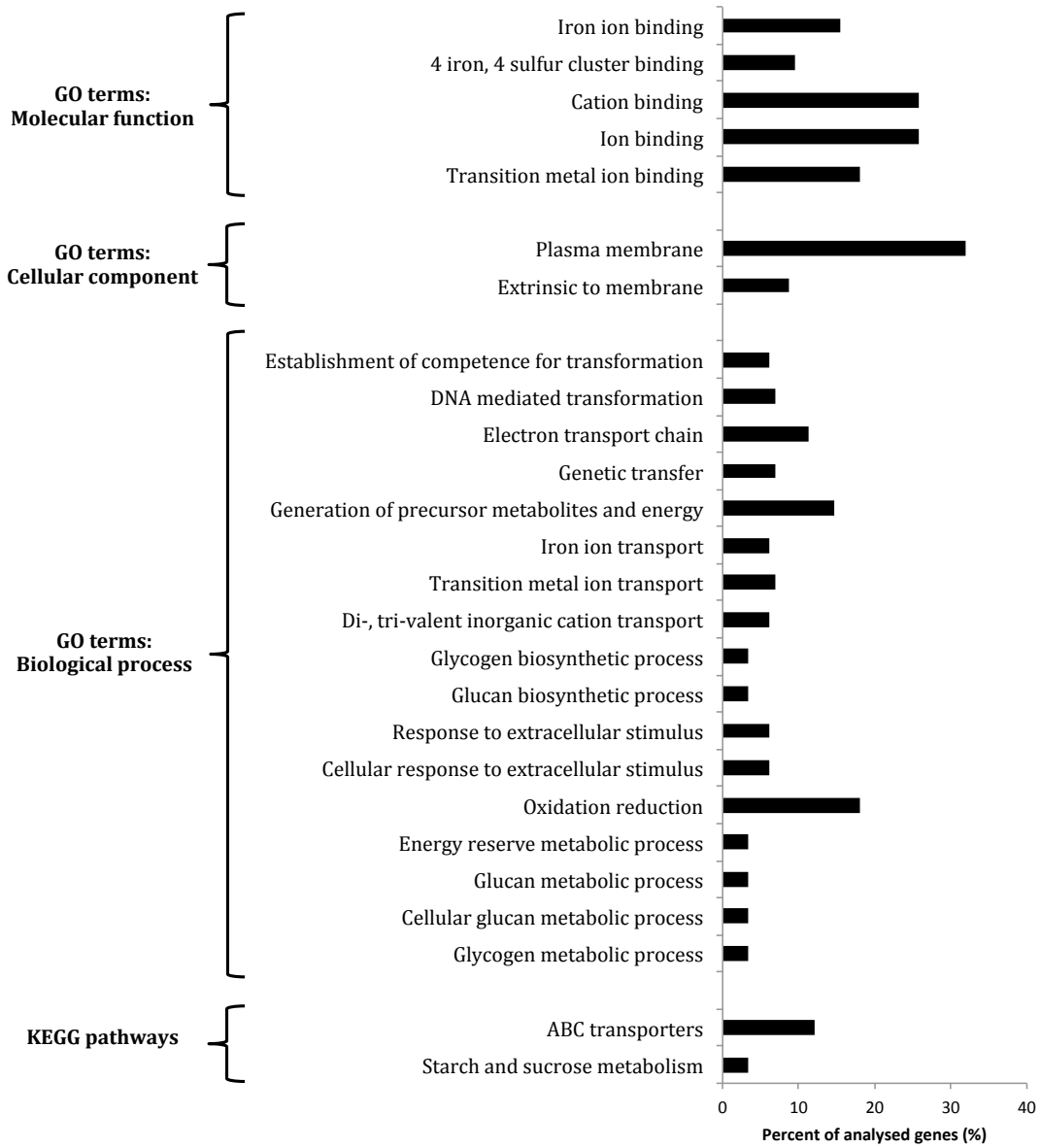

Figure 4.25: Functional group enrichment analysis of up-regulated genes in the Rd strain during nutritional stress, compared to growth in the rich medium. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of up-regulated genes that belonged to each functional group.

Table 4.9: The ten most highly up-regulated and down-regulated genes in the Rd strain during nutritional stress, compared to growth in the rich medium.

Gene	Product	Fold change
Up-regulated genes		
dps	DNA protection during starvation protein	11.82
sdaA	L-serine deaminase	10.55
sdaC	Serine transporter	10.22
gntP_1	Gluconate permease	7.84
garK	Glycerate 2-kinase	7.54
-	Aldolase	7.21
dprA	DNA processing chain A	7.16
-	Hypothetical protein	6.85
-	3-hydroxyisobutyrate dehydrogenase	6.57
ygbM	Putative hydroxypyruvate isomerase YgbM	6.51
Down-regulated genes		
yjcD	Putative permease YjcD	66.56
purH	Bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase	59.93
purD	Phosphoribosylamine-glycine ligase	35.82
purM	Phosphoribosylaminoimidazole synthetase	35.77
purE	Phosphoribosylaminoimidazole carboxylase catalytic subunit	33.62
purN	Phosphoribosylglycinamide formyltransferase	22.74
purK	Phosphoribosylaminoimidazole carboxylase ATPase subunit	19.39
mtr	Tryptophan-specific transport protein	17.58
lctP	L-lactate permease	11.89
-	Short chain dehydrogenase/reductase	10.51

4.2.2.4.4.2 Competence-related response

There was a clear induction of competence-related genes during nutritional stress in Rd. This was apparent from enriched functional groups as well as the competence-associated operon, *comABCDEF,* being up-regulated in Rd (see Figure 4.25) (see Appendix B). Among other genes with predicted roles in competence was *dprA*, encoding a DNA processing chain A, which was previously characterised as having a role in DNA transformation in *H. influenzae* (Karudapuram et al., 1995). Interestingly, there was down-regulation of the *rec-2* gene, encoding a protein involved in DNA transformation, and a putative SOS regulon repressor, *lexA.*

4.2.2.4.4.3 Reduction in purine biosynthesis and transport

There was down-regulation of purine biosynthetic and metabolic pathways during nutritional stress, which was particularly evident from the enriched functional groups (see Figures 4.26). The most highly down-regulated gene (66.6-fold) was *yjcD*, encoding a putative permease (see Table 4.9). The homologue of this gene in *E. coli* was previously shown to be involved in purine transport (Kozmin et al., 2013, Papakostas et al., 2013). Consistent with that, the whole *pur* regulon, with a putative role in purine biosynthesis, was also highly down-regulated in Rd (Zhang et al., 2008). The down-regulation of the pathway leading to generation of the purine precursor, inosinic acid (IMP), is depicted in Figure 4.27.

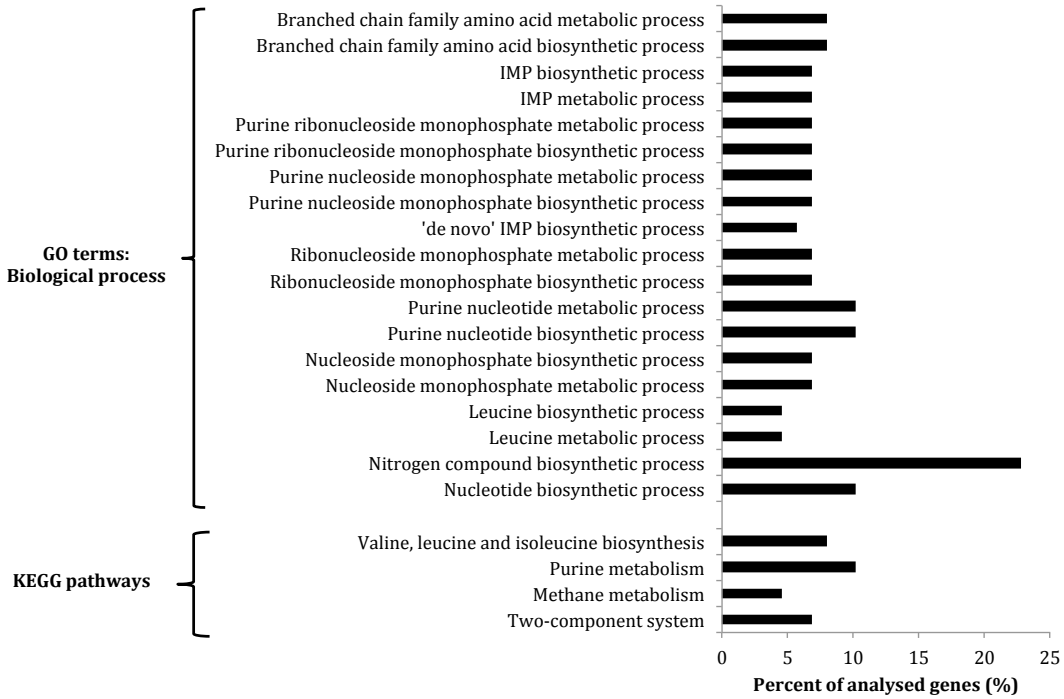

Figure 4.26: Functional group enrichment analysis of down-regulated genes in the Rd strain during nutritional stress, compared to growth in the rich medium. Groups were ordered from top to bottom based on the decreasing adjusted p-value. The x-axis displays the percentage of down-regulated genes that belonged to each functional group.

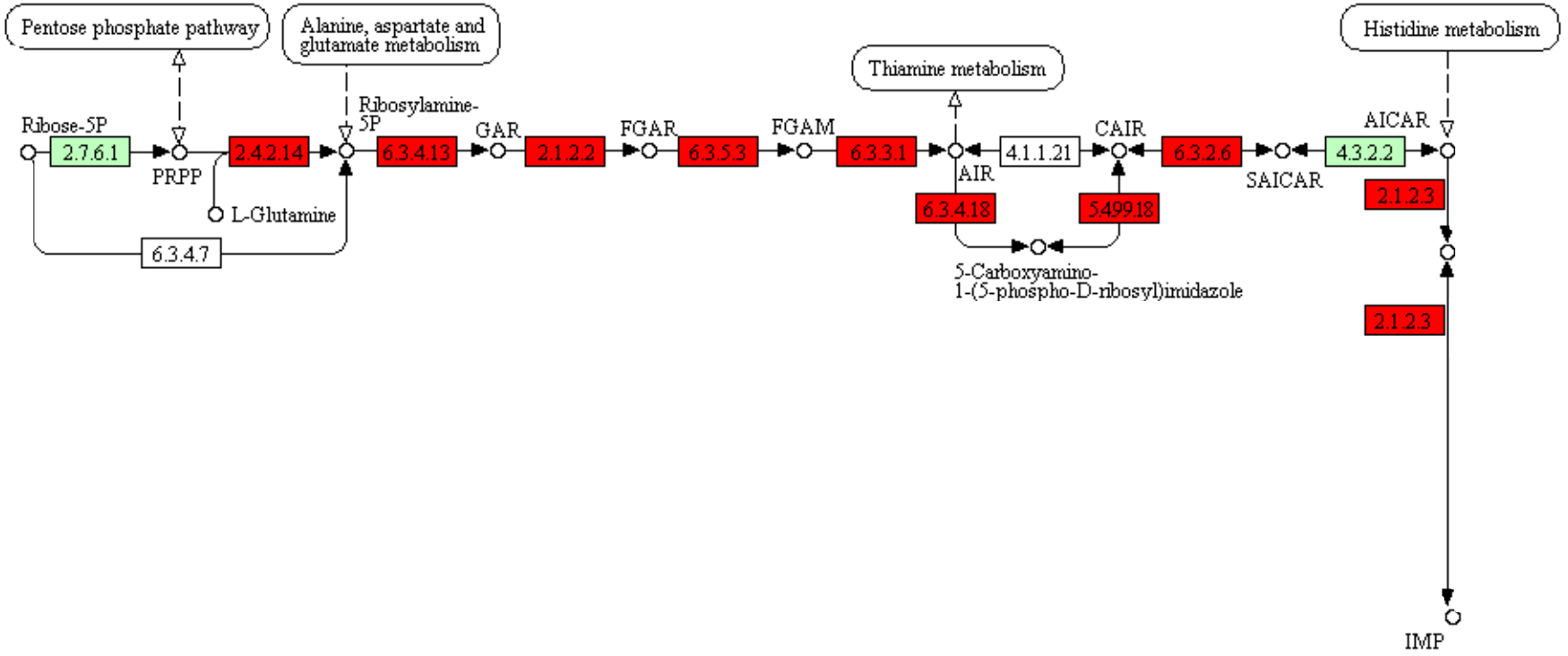

Figure 4.27: Down-regulation of the purine biosynthesis pathway in Rd during nutritional stress. The KEGG diagram depicts down-regulated genes in the purine precursor (IMP) biosynthetic pathway. Light green boxes represent genes present in *H. influenzae*, while red boxes show down-regulated genes.

4.2.2.4.4.4 Induction of oxidative stress response during nutritional stress

There was induction of several genes involved in the oxidative stress response during nutritional stress. The most highly up-regulated gene was *dps*, with a well-characterised role in oxidative stress response in *H. influenzae* (see Table 4.9) (see section 1.1.8.1). A peroxiredoxin gene, *pgdX*, with a protective role in protection against oxidative stress, was also up-regulated (see Appendix B) (see section 1.1.8.1). Several bacterial respiration genes were induced as well, including DMSO, nitrate and nitrite reductase operons, as well as a putative electron transport locus (Stewart and Bledsoe, 2005). Interestingly, the *arcA* gene, involved in oxidative stress defence in *H. influenzae*, was down-regulated during nutritional stress (Wong et al., 2007).

4.2.2.4.5 Comparison of transcriptional response of Rd and R2866 across different conditions

There were several genes in both Rd and R2866 strains that were up-regulated or down-regulated in more than one RNA-Seq condition (see Figure 4.28) (see Appendix B). Over half of up-regulated genes in oxidative stress were also induced during stationary phase in both strains. A glutamate dehydrogenase gene, *gdhA,* was the only up-regulated gene in all four conditions in Rd. Among up-regulated genes in Rd, common to three conditions and associated with oxidative stress, were *pgdX* and *dps*, while the *arcA* gene was down-regulated. It was also noteworthy that several iron acquisition-associated genes were up-regulated in all three conditions in R2866, including the whole haem utilisation locus, *hxuABC*, the iron acquisition gene, *hitA*, the haem receptor gene, *hemR*, as well as genes *tbp1* and *tbp2*, encoding transferrin-binding proteins.

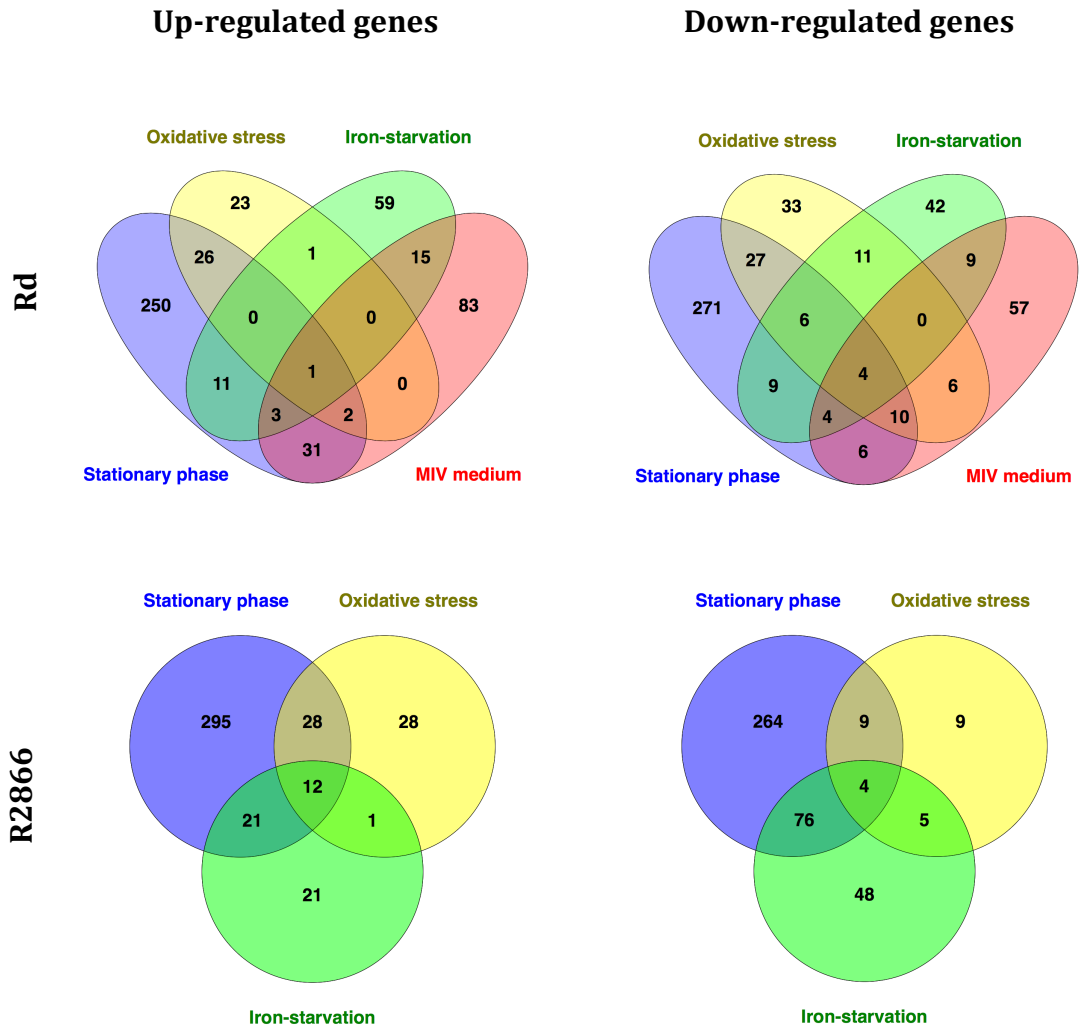

Figure 4.28: Venn diagrams of common differentially expressed genes across different RNA-Seq experiments in Rd and R2866 strains.

4.2.2.5 The whole transcriptome across different conditions in Rd and R2866 strains

The absolute expression of whole genomes across different conditions was compared in both Rd and R2866 strains. The DESeq2 normalisation method relies on the geometric mean of gene expression across all samples that are being compared between two RNA-Seq conditions. The normalisation data is thus specific to each differential gene expression analysis and is not comparable across several different RNA-Seq experiments. Therefore, the expression data for each gene in each condition was normalised using the TPM method. Gene expression during mid-exponential phase was the most similar to nutritional stress and oxidative stress in the Rd strain, and to oxidative stress in the R2866 strain (see Figures 4.29; 4.30). Gene expression during stationary phase was the least similar to all other conditions in both strains. The most highly expressed gene across all conditions in both strains was the *ssrA* gene, encoding a transfer-messenger RNA.

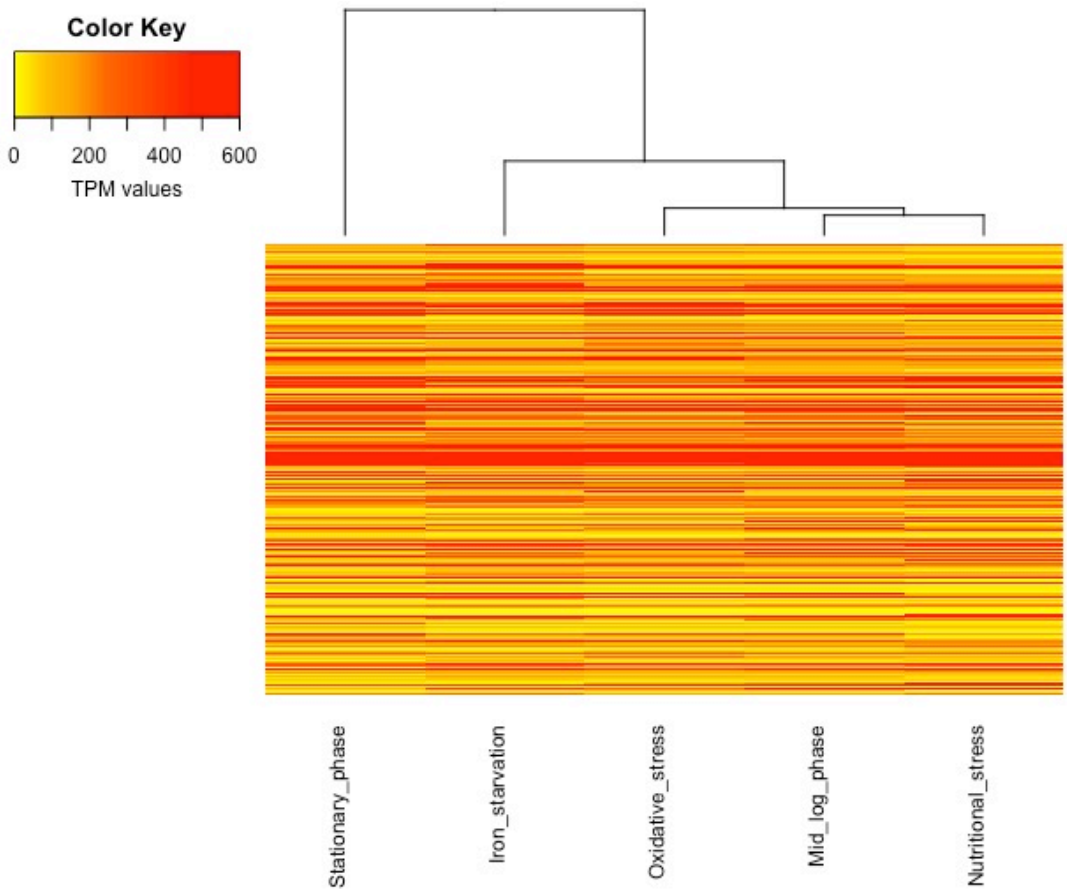

Figure 4.29: Comparison of the whole genome expression across five different conditions used in RNA-Seq experiments in the Rd strain. RNA-Seq conditions were clustered based on the euclidean distance between TPM-normalised expression of the whole genome.

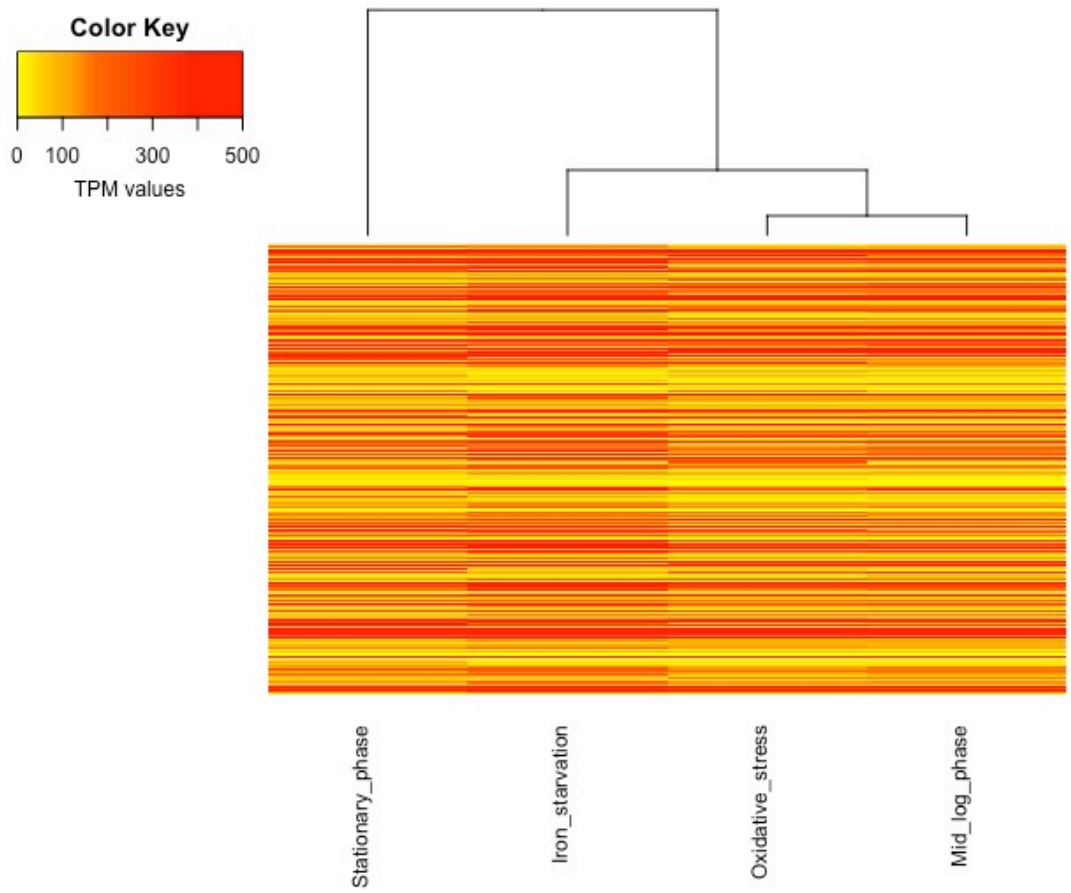

Figure 4.30: Comparison of the whole genome expression across four different conditions used in RNA-Seq experiments in the R2866 strain. RNA-Seq conditions were clustered based on the euclidean distance between TPM-normalised expression of the whole genome.

4.3 Discussion

This is the first time that HTS has been applied to study whole transcriptomes of Rd and R2866 strains of *H. influenzae* during infection-relevant conditions. These included growth during stationary phase, growth in nutrient-limiting medium as well as oxidative and iron-starvation stresses. Differentially expressed genes were identified in each of the conditions in comparison to growth at mid-exponential phase. In addition, an invasion assay in a human cell culture model was optimised for Rd and R2866 strains. The number of recovered intracellular bacteria was maximised for each strain to enable robust transcriptomic analyses of *H. influenzae* infection in future studies. The data presented here also established that Rd and R2866 strains of *H. influenzae* used in this laboratory were able to cause infection in the human A549 cell culture model. This directly supports the study of the transcriptional response of these strains during infection-relevant conditions.

4.3.1 Development of the eukaryotic invasion assay

As expected, using 100 bacteria per single host cell in the optimised invasion assay resulted in the highest intracellular numbers for both Rd and R2866. 100:1 is at the upper limit of accepted MOI and is not likely to be representative of a natural infection (Letourneau et al., 2011). An even higher MOI could potentially induce cellular damage and would make the infection model even more different from the natural infection. However, maximising the number of invading bacteria is important for future host-pathogen transcriptomic studies, where sufficient amounts of bacterial RNA will need to be extracted. In addition, bacterial invasion tends to be a heterogeneous process, meaning that not all eukaryotic cells are infected equally, with some cells potentially remaining completely non-invaded (Gerceker et al., 2000). A higher MOI would therefore ensure a lower transcriptional noise, as the possibility for a eukaryotic cell to remain non-infected would be reduced. This will be particularly important for future dual RNA-Seq experiments of both *H. influenzae* and the human host.

The optimal infection time was selected as 3 hours, due to the following decrease in the number of intracellular bacteria at 4 hours. The subsequent increase in the number of invading *H. influenzae* at 5 hours is curious. These fluctuations in intracellular numbers could possibly be explained by the complex dynamics of a bacterial infection, where killed intracellular bacteria could be replaced by new invading *H. influenzae*.

The experiment with *H. influenzae* grown in DMEM with varying concentrations of FBS suggests that *H. influenzae* requires FBS in order to grow in DMEM. The initial growth and subsequent death in the presence of FBS in the medium indicates FBS sustaining the growth of *H. influenzae* over a limited time period. At 3 hours post-inoculation, CFU/ml values for both Rd and R2866 were higher than at the start of the experiment when grown in the presence of 10% FBS (same as the infection medium). Therefore, in the invasion assay, the starting MOI actually increases over 3 hours and is higher for the Rd strain compared to R2866 at the time of the gentamicin treatment. This could explain why a higher number of intracellular bacteria was observed for Rd compared to an invasive R2866 strain in this study.

4.3.1.1 Enrichment of bacterial RNA in a mixed bacterial-host sample

RNA isolation from host cells infected with *H. influenzae* was carried out in this study. Since the proportion of bacterial RNA in a mixed sample is expected to be very low, the read depth for sufficient detection of bacterial transcription would only be achievable using high-end sequencers like the HiSeq™. However, due to financial constraints, the inability to use high-end sequencing instruments meant that the method needed to be further optimised by enriching bacterial RNA in a mixed sample. While the enrichment resulted in the significant depletion of 18S and 28S rRNA, it was not possible to observe bacterial rRNA bands in the Bioanalyzer electropherogram image, suggesting that the majority of RNA in the sample was still eukaryotic RNA. Nevertheless, the enrichment ensured that the proportion of bacterial RNA in the sample was now higher than

in the original sample. This could now potentially allow sequencing of the bacterial RNA in the enriched sample at the appropriate read depth using a benchtop instrument like the MiSeq™. The optimised invasion assay could also be used directly for a dual RNA-Seq experiment to study gene expression of both *H. influenzae* and the human host. Potential heterogeneity in the number of invaded and non-invaded eukaryotic cells could be overcome by utilising a recently described method, where non-infected cells are differentiated via fluorescence-activated cell sorting technology (Westermann et al., 2016).

4.3.2 The transcriptional response of Rd and R2866 strains during infection-relevant conditions

4.3.2.1 R2866 is more sensitive to oxidative stress and iron-starvation than Rd

It was interesting to observe that an invasive R2866 strain was more sensitive to both iron-starvation and oxidative stresses than a standard laboratory Rd strain, based on the growth kinetics. While R2866 may be more susceptible to environmental stresses examined in this study, it was shown to be highly resistant to human serum through complement evasion, which is the most likely explanation for its invasive phenotype (Williams et al., 2001).

The optimised concentration of hydrogen peroxide for the induction of oxidative stress in this study was 5 mM. The R2866 strain had a greater growth defect in reaction to this treatment than Rd, highlighting the variability in stress response between different strains. Concentrations of up to 20 mM of hydrogen peroxide did not have an effect on the viability of the wild type 86-028NP strain of *H. influenzae*, when grown in the MIV medium (Juneau et al., 2015). In contrast, the same strain grown in the rich sBHI medium was sensitive to as low as 0.5 mM of hydrogen peroxide, revealing the significant effect that the growth medium seems to have on the sensitivity of *H. influenzae* strains to hydrogen peroxide treatment. (Harrison et al., 2007). The millimolar concentrations of hydrogen peroxide tested in this study were similar to the amounts produced

by a co-pathogen *S. pneumoniae* as well as active neutrophil cells (Duane et al., 1993, Pericone et al., 2003).

As with the hydrogen peroxide treatment, 5 mM of 2,2-bipyridine was selected as the optimal concentration for iron chelation in this study. A much lower concentration of 0.5 mM was previously used for iron chelation in the 86-028NP strain, though the effect on the bacterial growth was not described in that study (Harrison et al., 2013). In addition, as with hydrogen peroxide, it is likely there is a large inter-strain variation in sensitivity to iron chelation in *H. influenzae*. This was indeed observed in this study, where R2866 was more sensitive to the treatment with 2,2-bipyridine than Rd. While depletion of iron was not measured quantitatively in this study, a reliable indication of iron chelation was the colour change to dark red, observed in broth cultures within 1 hour after treatment. This was due to 2,2-bipyridine interacting with ferrous iron to form an iron(II) bipyridyl complex responsible for an intense dark red colour (O'Sullivan et al., 1990).

Another difference between the two strains was that R2866 had a different distribution of rRNA bands compared to Rd, most likely resulting from the fragmentation of 23S rRNA. This phenomenon has indeed been observed in *H. influenzae* and has been linked to the presence of intervening sequences (IVS) in the 23S rRNA gene (Song et al., 1999). IVSs have been previously associated with pathogenic strains in other bacteria, though their exact role remains to be characterised (Skurnik and Toivanen, 1991).

4.3.2.2 Oxidative stress: differences in strain sensitivity to hydrogen peroxide only partially translates to transcriptional response

Induction of oxidative stress resulted in the lowest number of differentially expressed genes compared to other RNA-Seq experiments. This is likely the result of a targeted bacterial response to this environmental stress. Catalase, which is a member of the OxyR regulon in the 86-028NP strain, was previously

shown to be responsible for scavenging high concentrations of hydrogen peroxide in *H. influenzae* (Pauwels et al., 2004, Harrison et al., 2007). A very high up-regulation of the catalase gene, *hktE*, in this study implies that the majority of hydrogen peroxide was scavenged by catalase in both Rd and R2866. This supports the idea that oxidative stress led to a more targeted and straightforward response than other tested infection-relevant conditions in this study. As the *oxyR* gene is an important mediator of oxidative stress defence in *H. influenzae*, it was surprising that there was no change in *oxyR* expression in this study. It is possible that the changes of the *oxyR* mRNA transcript levels were rapid and transient. As observed in a study by Whitby et al., *oxyR* transcript levels returned to normal levels within 10 minutes after treatment with hydrogen peroxide (Whitby et al., 2012).

A total of 11 members of the *oxyR* regulon were previously identified in the 86-028NP strain of *H. influenzae*, all of which were up-regulated during oxidative stress in a microarray study (Harrison et al., 2007). Of these, *hktE*, *gnd*, *pgdX* and *dps* were up-regulated in both Rd and R2866 during oxidative stress in this study. An OxyR-regulated gene, NTHI0684, in the 86-028NP strain, encoding a hypothetical protein, was also up-regulated in both strains in this study. Interestingly, it coded for a putative membrane protein in Rd and a putative CRISPR-associated protein, Cas2, in R2866. Sequence search on the InterPro database revealed no homology to known protein families or domains. Remaining members of the OxyR regulon were genes *pntA* and *pntB,* up-regulated in Rd only, as well as the *yfeABCD* locus, the expression of which was not induced in either of the strains in this study. This is different to a microarray study by Whitby et al., where the *yfeB* gene was up-regulated in Rd in response to hydrogen peroxide, while *gnd* was not (Whitby et al., 2012). The *yfeABCD* locus was also highly up-regulated in response to oxidative stress in another microarray study by Harrison et al. (Harrison et al., 2007). While there are clear inter-strain variations as well as differences between study designs, the overall OxyR-dependent response to oxidative stress seems to be similar among *H. influenzae* strains.

Several iron acquisition and ferritin-like protein genes were up-regulated during oxidative stress as well. Ferritin-like proteins sequester ferrous iron, thus preventing further oxidative damage through Fenton reaction. The up-regulation of iron uptake could be utilised for the repair of damaged iron-sulphur cluster proteins. Iron-sulphur clusters are present as cofactors in a large number of enzymes, mediating a variety of different roles in bacteria (Yoch and Carithers, 1979). There was up-regulation of genes associated with iron-sulphur cluster formation in both strains as well. The iron acquisition gene locus, *hxuABC*, as well as genes *tbp1*, *tbp2*, *hitA* and *hemR*, up-regulated in R2866 only, were shown to be part of the Fur regulon in the 86-028NP strain, suggesting that this transcriptional regulator plays a role in oxidative stress defence in R2866 as well. Iron homeostasis, including transcriptional regulation by Fur, plays an important role during oxidative stress in *H. influenzae* as shown in this and previous studies (Harrison et al., 2015).

Genes *recA*, *lexA*, *recX* and *ruvA*, up-regulated in both strains during oxidative stress, were previously shown to be involved in the SOS response in *H. influenzae* as part of the LexA regulon (Sweetman et al., 2005). The expression of genes *recA* and *recN* was also shown to be induced in the 86-028NP strain in response to hydrogen peroxide (Harrison et al., 2007). Up-regulation of these genes, along with multiple other genes with predicted roles in SOS response and protection against DNA damage, highlights the drastic damaging effect that hydrogen peroxide has on bacterial DNA (Rohwer and Azam, 2000).

Up-regulation of the acetolactate synthase gene locus, *ilvHI,* in Rd during oxidative stress was curious. This enzyme, along with an up-regulated ketol-acid reductoisomerase gene, *ilvC*, is involved in the biosynthesis of branched-chain amino acids leucine, isoleucine and valine (Ricca et al., 1988). Branched-chain amino acid supplementation has been shown to decrease oxidative stress levels in eukaryotic cells, though it is not clear whether this translates to prokaryotes as well (Iwasa et al., 2013). In addition, the lack of branched-chain amino acids in the host cells induced expression of virulence genes in *Listeria monocytogenes* (Lobel et al., 2012). Therefore, the concurrent up-regulation of

branched-chain amino acid biosynthesis during oxidative stress could represent *H. influenzae* responding to host-like conditions.

Although the induced expression of genes during oxidative stress in Rd and R2866 was largely similar, there were a lot more differences in down-regulated genes. For instance, R2866 only contained 27 down-regulated genes, whereas Rd had 97. In addition, there was no overlap in the ten most highly down-regulated genes between the two strains. It is not clear why the response was this different, but a partial explanation could be that the gene expression of one of the R2866 replicates from the mid-exponential group noticeably differed from the other two, as inferred with hierarchical clustering. Therefore, it could result in the underestimation of some differentially expressed genes. However, since the up-regulated genes in R2866 had a much greater similarity to Rd, this cannot plausibly be the whole reason. The difference in down-regulation could be simply explained by different strategies that R2866 employs during oxidative stress.

As there were a large number of up-regulated iron-related genes during oxidative stress, down-regulation of several other iron-associated genes during the same condition highlights the complex dynamic of iron homeostasis and oxidative stress in *H. influenzae*. Down-regulation of several ribosomal protein genes was most likely related to a reduction in protein synthesis resulting from a general response to a stress condition.

The arginine uptake locus, which is part of the Fur regulon in the 86-028NP strain, was down-regulated during oxidative stress in the Rd strain. In agreement with that, there was also down-regulation of other gene clusters that were Fur-regulated in the 86-028NP strain. This included DMSO and nitrite reductase loci as well as the *hbpA* gene (Harrison et al., 2013). Genes *ftnA1* and *ftnA2* were the only down-regulated genes in R2866 that are regulated by Fur. The disparate down-regulation of Fur-associated genes in Rd and R2866 possibly represents their varied response to hydrogen peroxide. This is

supported by the fact that most of up-regulated Fur-regulated genes were only present in the R2866 strain, as described earlier.

The only gene locus that was down-regulated over 3-fold during oxidative stress in R2866 contained genes encoding proteins with putative roles in carbohydrate processing. The most highly down-regulated gene, *Hgd,* encoded a putative 2-(hydroxymethyl)glutarate dehydrogenase, which is involved in nicotinate fermentation in *Eubacterium barkeri* (Reitz et al., 2008). This gene is part of the same family of β-hydroxyacid dehydrogenases as *gnd*, which was up-regulated in both strains. It is probable that the predicted annotation of this whole locus is wrong. Further investigation is required to determine the real function of this locus and its role in oxidative stress.

4.3.2.3 Different transcriptional behaviour of Rd and R2866 in response to iron-starvation

As expected, many characterised and putative iron- and metal-associated genes, including those related to iron uptake, were up-regulated during iron-starvation stress in both strains. However, a lot of these genes were induced in the R2866 strain only. In addition, those iron-related genes that were up-regulated in both strains were largely induced to a higher fold change in R2866. As demonstrated by growth kinetics, R2866 was more sensitive to iron-starvation stress than Rd in this study. Hence this can explain the increased requirement for iron acquisition and metabolism in R2866. As discussed earlier, the higher susceptibility of the R2866 stress to environmental stresses is likely surpassed by its resistance to human serum during natural infection.

It was curious that Rd had several LOS-related genes that were highly up-regulated during iron-starvation. One possibility is that Rd sensing an iron-limiting environment, which *H. influenzae* is likely to encounter in its natural niche, induced the expression of genes involved in its pathogenesis. However, apart from the *lptC* gene, the expression of LOS-related genes was not induced in R2866. This is likely due to this strain experiencing iron-starvation stress to a

higher degree than Rd and therefore employing different survival strategies by primarily inducing iron acquisition pathways. It would be interesting to examine the transcriptional response of R2866 to a lower concentration of the iron chelator in future work.

There was up-regulation of Fur-regulated genes, *hitABC*, *yfeABCD* and *hemR,* in both strains, while *hxuABC*, *tonB*, *exbBD*, *tbp1*, *tbp2* and *tnaA* were only up-regulated in R2866 (Harrison et al., 2013). In addition, the expression of these genes was induced in both strains during iron-starvation in previous microarray studies (Whitby et al., 2013). The *hitABC* locus was shown to be highly up-regulated in *H. influenzae* infecting human cells, demonstrating the requirement for iron acquisition in a limited environment (Baddal et al., 2015). LOS-related genes were not up-regulated in a microarray study where iron-starvation was induced in the Rd strain (Whitby et al., 2006). The differences between these studies are most likely due to different methodologies of inducing iron-starvation. In addition to other iron uptake genes, there was also up-regulation of multiple putative metal transporters, which are likely to have a role in iron transport. There was also up-regulation of several genes involved in heat shock response in both strains. This likely represents a dynamic and overlapping response to iron-starvation and general bacterial stress.

Consistent with up-regulated genes during iron-starvation, several of the genes down-regulated in this study were part of the Fur regulon in the 86-028NP strain (Harrison et al., 2013). These included a haem-binding lipoprotein gene, *hbpA,* as well as formate dehydrogenase and nitrite reductase operons. Moreover, the Fur regulation was again more evident in the R2866 strain, as arginine uptake and fumarate reductase operons were down-regulated in that strain only. The higher sensitivity of R2866 to iron-starvation was also highlighted by down-regulation of several translation-associated genes.

L-cystine and DL-methionine uptake genes were highly down-regulated during iron-starvation in both strains. Cysteine and methionine are both sulphur-containing amino acids, therefore the decrease in their uptake could be the

means of maintaining cellular sulphur metabolism. L-cysteine indeed acts a substrate for a sulphur transport system in *E. coli*, with implicated roles in processes like oxidative stress (Dai and Outten, 2012). This could possibly have implications for the maintenance of iron-sulphur clusters during growth at iron-limiting conditions in both Rd and R2866.

Down-regulation of *ilvC*, *ilvH* and *ilvI* genes, encoding a ketol-acid reductoisomerase and acetolactate synthase subunits, during iron-starvation in both strains was consistent with the reduced expression of these genes in a microarray study where iron-starvation was induced in the Rd strain (Whitby et al., 2006). However, their exact role in the iron-starvation response is not clear. Other notable down-regulated genes with unidentified roles in iron-starvation were histidine and tryptophan biosynthesis loci as well as genes from the ATP synthase locus, *atpABCDEFGHI*, with a role in bacterial respiration. One possible reason is simply the need to reduce respiratory and biosynthetic processes in a slow-growing bacterial community in response to stress.

4.3.2.4 Nutrient limitation in the Rd strain leads to induction of competence and iron-starvation response

The MIV medium is chemically-defined and is nutrient-limiting compared to the rich sBHI medium (Herriott et al., 1970). Therefore, it was not surprising that several nutrient metabolism- and transport-associated loci were up-regulated in Rd during nutritional stress in the MIV medium. This included glycogen biosynthesis and metabolism genes, glycerol utilisation locus as well as several other gene clusters involved in carbohydrate and amino acid metabolism and uptake. The most highly up-regulated gene during nutritional stress was *dps*, which was consistent with this gene being present intracellularly in large quantities during starvation conditions in *E. coli* (Almiron et al., 1992). The inability of the R2866 strain to grow in the MIV medium again highlighted the increased sensitivity of this strain to environmental stresses.

The up-regulation of several iron-associated genes during nutritional stress was most likely due to the reduced availability or iron in the MIV medium. Indeed, a lower amount of haemin was present in the MIV medium (10 μg/ml) compared to the sBHI medium (15 μg/ml). This is in addition to any other iron-containing compounds, which are present in the rich medium, being absent from the chemically-defined MIV medium. There was also up-regulation of multiple genes with a role in energy generation from metabolites, including a putative electron transport locus and nitrate reductase genes. This possibly represents bacteria scavenging available nutrients from the medium and inducing the expression of energy-producing genes required for the exponential growth.

As expected, a competence-associated operon, *comABCDEF*, was up-regulated during nutritional stress, as the MIV medium was originally developed to facilitate the competence process in *H. influenzae* under starvation conditions (Herriott et al., 1970). The expression of this operon was also previously shown to be induced in the Rd strain of *H. influenzae* after transfer from the sBHI medium to MIV (Redfield et al., 2005). Whilst the *comABCDEF* operon is involved in the DNA uptake, genes responsible for DNA integration were up-regulated as well, including *comM* and *dprA* (Karudapuram et al., 1995, Gwinn et al., 1998). The expression of these two genes was previously highly induced in Rd when grown in the MIV medium (Redfield et al., 2005). This highlights that both stages of the competence process, DNA uptake and integration, were initiated during nutritional stress in this study.

Depletion of purine pools in the growth medium has been previously shown to be a necessary signal for competence induction in *H. influenzae* (MacFadyen et al., 2001). This is directly related to down-regulation of multiple purine biosynthesis-associated genes during nutritional stress in Rd. Moreover, the most highly down-regulated gene in Rd encoded a permease protein with a putative role in purine import. Purine uptake and biosynthetic genes are repressed by PurR in response to available purine pools in *E. coli* (Cho et al., 2011). It is curious as to why the expression of purine-associated genes was reduced, as the MIV medium is nutrient-limiting and it is thus expected to be

starved of purines as compared to the sBHI medium. Down-regulation of purine uptake and biosynthetic pathways in this study could therefore be possibly linked to other regulatory networks responsible for competence initiation. There was also down-regulation of a DNA transformation-associated gene, *rec-2*, which was puzzling (Barouki and Smith, 1985). It was previously discovered that the *rec-2* gene contains a putative PurR binding site, though it was shown experimentally in the same study that PurR does not repress *rec-2* (Sinha et al., 2013).

Consistent with the up-regulation of Fur-regulated iron acquisition locus, *hitABC,* during nutritional stress, there were several down-regulated genes that also belonged to the Fur regulon, including arginine uptake and formate dehydrogenase gene loci. This possible regulation by Fur during nutritional stress is likely a result of reduced iron in the medium, thus mimicking the iron-starvation stress to some extent.

4.3.2.5 The metabolic response of *H. influenzae* during stationary phase

This was the first study to investigate how *H. influenzae* behaves during stationary phase in comparison to mid-exponential phase on a whole transcriptome level. Due to the limiting nature of the environmental niche that *H. influenzae* inhabits, it is likely to persist at stationary phase during most of human colonisation and infection. Of all four studied infection-relevant conditions, the highest number of differentially expressed genes was at stationary phase. This highlights the complex and multifactorial bacterial reprogramming required during transition from exponential to stationary phase growth.

As expected, a large number of gene loci, up-regulated during stationary phase, were involved in metabolic pathways. This was apparent from the up-regulation of multiple genes involved in pentose and hexose monosaccharide metabolic processes, including gene clusters responsible for L-fucose, D-ribose and

galactose metabolism. The ability to utilise alternative sugars can confer advantage during bacterial infection by improving survival rates, modifying surface structures as well as sugars acting as ligands necessary for successful colonisation. Indeed, the L-fucose metabolic pathway was previously associated with increased virulence in *Campylobacter jejuni* (Stahl et al., 2011). The galactose locus was shown to play a role in pathogenesis of *H. influenzae*, by affecting the LOS composition (Maskell et al., 1992).

Up-regulation of the whole histidine biosynthesis operon, glycerol utilisation locus as well as glycogen biosynthesis and metabolism genes during stationary phase signifies the ability to produce essential nutrients in a limiting environment, which could potentially confer important advantage to bacterial colonisation and persistence. This is exemplified by the histidine operon being previously shown to be more prevalent in *H. influenzae* strains isolated from the middle ear (Juliao et al., 2007). Glycogen acts as an important carbon source reserve and was previously shown to have a role in survival and persistence of *S. enterica* (McMeechan et al., 2005). Similarly, the up-regulation of a sugar-phosphate transport locus, *afuABC*, oligopeptide genes, *oppA* and *oppB,* as well as short-chain fatty acid utilisation locus, *atoABCD,* was likely induced by the bacterial need to scavenge any available nutrients in the exhausted medium. The presence of the *afuABC* locus was also shown to be associated with virulence in the enteric pathogen *Citrobacter rodentium* (Sit et al., 2015).

The ornithine decarboxylase gene, *speF,* was highly up-regulated during stationary phase in both strains. It is responsible for the production of putrescine, which is a polyamine compound with a wide range of functions in bacteria. Putrescine can bind directly to mRNA and its supplementation could partially restore virulence in *Shigella flexneri* mutants that were not able to produce modified nucleosides for the tRNA synthesis (Durand and Bjork, 2003). Putrescine is also responsible for enhancement of protein synthesis of a sigma factor RpoS, which is an important regulator of genes expressed during stationary phase (Igarashi and Kashiwagi, 2006). Therefore, the increased levels of putrescine in the bacterial cell could be involved in the maintenance of levels

of RpoS, thus promoting survival at stationary phase. Up-regulation of a putrescine export gene, *potE,* likely played a role in balancing the intracellular levels of putrescine. Down-regulation of putrescine and spermidine uptake genes in both strains was in agreement with up-regulation of putrescine export in this study and likely plays a role in intracellular putrescine homeostasis.

The tryptophanase locus is absent in the Rd strain, but it was highly up-regulated in the R2866 strain during growth at stationary phase. As mentioned earlier, it is responsible for indole production and was previously used as a marker for virulence in *H. influenzae* (Kilian, 1976). High levels of intracellular indole have also been implicated in the induction of stationary phase in *E. coli*, with a proposed role in promoting the long-term survival of bacterial cells (Gaimster et al., 2014).

4.3.2.6 *H. influenzae* experiences oxidative stress and iron-starvation during stationary phase

Investigation of the transcriptional behaviour of *H. influenzae* during four different infection-relevant conditions in this study revealed that it possesses overlapping responses to different stresses. In particular, Rd and R2866 strains seemed to experience oxidative stress and iron-starvation during stationary phase, as described below. Therefore, this study highlights the advantage of utilising RNA-Seq in order to analyse different bacterial responses, which can then be used to identify common trends in bacterial adaptation and survival during host colonisation and disease progression.

Multiple genes with characterised and predicted roles in oxidative stress defence, including multiple members of the OxyR regulon, were up-regulated at stationary phase in both strains. The most highly up-regulated gene in Rd was a gene encoding a putative alkylhydroperoxidase AhpD-like protein. The possession of the AhpD-like domain suggests that this protein may have alkylhydroperoxidase or other antioxidant activity in *H. influenzae*. The AhpD protein has been shown to have antioxidative activity in *Mycobacterium*

tuberculosis, with a role in protection against oxidative stress (Hillas et al., 2000). In addition, the up-regulation of several genes encoding DNA repair proteins suggests of the DNA damage, which was possibly occurring due to oxidative stress, as observed previously (Rohwer and Azam, 2000).

The inability of *E. coli* strains, lacking oxidative stress defence mechanisms, to survive at stationary phase suggests that oxidative stress indeed plays a large part during this growth phase (Dukan and Nystrom, 1999). This phenomenon is further exaggerated by the fact that in both Rd and R2866 strains there was a large number of overlapping genes that were up-regulated during both oxidative stress and stationary phase. Cell starvation in itself seems to induce protection against environmental stresses as was previously demonstrated in *E. coli*, which, when starved, was more resistant to heat and hydrogen peroxide treatment (Jenkins et al., 1988). In addition, one of the consequences of oxidative stress is protein carbonylation, which is also used as a biomarker for cell senescence (Dukan and Nystrom, 1999). Chaperone genes, *groEL* and *groES,* were up-regulated in both strains, while the expression of a heat-shock RNA polymerase sigma factor-32 gene was induced in Rd only. These genes have been previously associated with a protective role against protein carbonylation during stationary phase in *E. coli* (Fredriksson et al., 2005).

The catalase activity in the Rd strain of *H. influenzae* was previously shown to be higher at mid-exponential phase rather than stationary phase (Bishai et al., 1994). Interestingly, the opposite was true for the expression of the catalase gene in both strains in this study. As stationary phase is a complex and dynamic growth phase, this variability could have arisen due to a difference in sampling methods as well as differing laboratory conditions. The *dps* gene was highly up-regulated at stationary phase in Rd and R2866. It was indeed shown to be the most abundant protein during stationary phase in *E. coli* (Almiron et al., 1992). In addition to its ferritin-like properties, Dps is able to bind the DNA and protect it from environmental assaults, particularly in starved cells (Almiron et al., 1992). Dps also plays a role in virulence as it was shown to promote survival of *H. influenzae* in the experimental chinchilla model (Pang et al., 2012).

Several genes, related to the competence process (both DNA uptake and integration), were up-regulated in both strains at stationary phase. Competence is indeed induced by nutrient limitation and high cell density, which are both features of bacterial stationary phase (Solomon and Grossman, 1996). The availability of purine nucleotides has also been shown to repress the expression of competence genes in *H. influenzae* (MacFadyen et al., 2001). Therefore, starvation conditions during stationary phase would likely result in nutrient and purine limitation, and induction of the competence phenotype.

There were several up-regulated genes associated with iron and other metals at stationary phase, which was likely due to exhaustion of available iron in the medium. The up-regulation of iron uptake genes was greater in the R2866 strain than in Rd, again highlighting that the former strain is more sensitive to iron-limitation than the latter. Several up-regulated genes were part of the Fur regulon in the 86-028NP strain. These included a haem utilisation locus, *hxuABC*, metal transport genes, *yfeA* and *yfeB*, the tryptophanase gene, *tnaA,* as well as genes *tbp1* and *tbp2*, coding for transferrin-binding proteins (Harrison et al., 2013).

Up-regulation of genes involved in iron-sulphur cluster formation is possibly linked to the restoration of clusters damaged during oxidative stress, which is likely to occur during stationary phase as discussed above. Oxidative stress can result in the release of ferrous iron leading to increased stress levels via the Fenton reaction (Keyer and Imlay, 1996). The *iscR* gene, which encodes an iron-sulphur cluster transcriptional regulator and was up-regulated in this study, was previously shown to be involved in the oxidative stress defence in *H. influenzae* (Wong et al., 2013).

There was up-regulation of bacteriophage and ICE genes in the R2866 strain at stationary phase. In contrast, several genes from a prophage were down-regulated in the Rd strain. This highlights the diversity of these mobile genetic elements and their difference in the induction of gene expression. There was also up-regulation of several pilus and LOS modification genes in R2866, which

might contribute to its invasive phenotype by promoting survival in the limiting environment.

Multiple ribosomal protein and tRNA genes were highly down-regulated in both Rd and R2866 during stationary phase. This highlights the need for bacteria to conserve the energy during this growth phase due to reduced nutrient availability. Down-regulated ribosomal protein genes encoded members of 30S and 50S ribosomal subunits. There was no down-regulation observed for 5S, 16S and 23S rRNA genes due to the efficient depletion of rRNA transcripts during cDNA library preparation.

Genes involved in the oxidative phosphorylation pathway as well as electron transport were down-regulated in Rd and R2866 at stationary phase. This signifies reduced respiration in a non-growing bacterial community as has been previously demonstrated (Riedel et al., 2013). Down-regulation of the toxin-antitoxin locus, *vapBC-1,* in R2866 during stationary phase was consistent with a previous observation of its expression levels being inverse to *H. influenzae* cell density (Daines et al., 2007). On the other hand, it was not clear why members of both Sec and Sec-independent protein export systems were down-regulated in this study. A possible explanation for it is the concurrent reduction in protein synthesis and the need to conserve energy.

The arginine uptake operon was highly down-regulated during growth at stationary phase in both Rd and R2866. It was previously shown that the arginine uptake locus was regulated by Fur in the 86-028NP strain (Harrison et al., 2013). There was up-regulation of several genes from the Fur regulon in this study; hence down-regulation of arginine uptake could be Fur-mediated at stationary phase. In addition, ferritin-like genes, *ftnA1* and *ftnA2*, also part of the Fur regulon, were down-regulated in R2866.

4.3.2.7 Presence of genes in *H. influenzae* with roles in multiple stress responses

Among noteworthy genes that were differentially expressed in multiple conditions was *arcA*, a two-component regulator gene, which was down-regulated in oxidative stress, stationary phase and nutritional stress in Rd only. The down-regulation of *arcA* was interesting, particularly during oxidative stress, as it was implicated in the positive regulation of *dps* as well as other genes with putative roles in oxidative stress defence during anaerobic growth (Wong et al., 2007). Its consistent down-regulation in this study suggests a different role for this transcriptional regulator during aerobic growth. A glutamate dehydrogenase gene, *gdhA,* was up-regulated in all four conditions in Rd. While its exact role in tested conditions is not clear, it was previously implicated in conferring advantage to virulent strains of *Clostridium botulinum* (Hammer and Johnson, 1988). The up-regulation of several iron-associated genes in all three conditions in the R2866 strain again demonstrated its susceptibility to limited iron availability.

The data presented in this chapter provide an important foundation for future work, where genes with putative roles during infection-relevant conditions could be further investigated and characterised. This could then lead to the discovery of novel targets for vaccine or antibiotic development. The optimised invasion assay also serves as a preliminary basis for future work, where whole transcriptomes of intracellular *H. influenzae* and the human host could be studied simultaneously in a dual RNA-Seq experiment. This will help to further elucidate the important processes involved in the pathogenesis of *H. influenzae* as well as host response and adaptation to the progression of bacterial infection. In addition, RNA-Seq conditions used in this study will be essential for the subsequent identification and analysis of novel RNA transcripts (see Chapter 5).

Chapter 5: Robust identification and analysis of non-coding RNAs in *H. influenzae*

5.1 Introduction

Non-coding RNAs, including sRNAs, play a variety of important roles in bacterial cells, such as stress response and virulence regulation (see section 1.2.2). While there are available tools for computational prediction of these RNA structures in bacterial genomes, very few programs exist that use RNA-Seq data to identify ncRNAs based on their expression (Sridhar et al., 2010, McClure et al., 2013). General lack of robust online tools for ncRNA identification means researchers have to rely on either a visual inspection of the data or custom programming language scripts. Therefore, more robust and reproducible methods, specifically designed to identify putative ncRNAs from the whole-transcriptome data, are required.

The first major goal of this study was to develop a new robust and systematic tool for reproducible identification of ncRNA transcripts from RNA-Seq data. For that purpose, the transcriptomic dataset, used earlier for the analysis of the transcriptional response of *H. influenzae* during infection-relevant conditions, was utilised (see Chapter 4). This was combined with the bacterial whole-genome annotation to identify ncRNAs present in intergenic regions as well as those antisense to coding sequences. Furthermore, UTRs were detected for each gene, along with intergenic regions belonging to putative operons.

The number of sRNAs in bacteria has been suggested to be inversely proportional to the size of the genome (Guell et al., 2011). Thus, it is expected that *H. influenzae* should contain a relatively large number of ncRNAs due its small genome size (~1.8 Mb). There have only been two studies to-date that have addressed the presence of ncRNAs, including sRNAs, in *H. influenzae*. The only ncRNA that has been experimentally validated in *H. influenzae* is HrrF - a Fur-regulated sRNA in the 86-028NP strain (Santana et al., 2014). In addition,

Baddal et al. discovered only 18 putative ncRNAs in an otitis media isolate of *H. influenzae,* by manually inspecting expression of intergenic regions (Baddal et al., 2015). This highlights that there needs to be a more systematic and robust identification and characterisation of these possibly abundant and potentially important RNA elements in *H. influenzae*.

The second major aim of this study was to identify a repertoire of putative ncRNAs in *H. influenzae,* by employing the tool developed in the same study. For that purpose, all the RNA-Seq data from infection-relevant conditions used for differential gene expression studies in Chapter 4 were utilised in order to enable a more robust detection of novel RNA elements. This was facilitated by the ncRNA expression being diverse and dependent on the bacterial growth phase and response to stress, as observed previously (Rau et al., 2015). Consistent with that, the differential expression of ncRNAs during infection-relevant conditions was also investigated in this study. In addition, identified ncRNAs were examined for the possibility of encoding small proteins. Finally, selected ncRNAs were validated using the northern blotting technique, along with predicting their secondary structure and mRNA targets. This work is important for future studies where ncRNAs can be further investigated for their potential roles in the pathogenesis of *H. influenzae*.

5.2 Results

5.2.1 Development of a script to identify ncRNAs

The Python programming language was used to write a script for the identification of putative ncRNAs in a bacterial genome from the RNA-Seq data. The name for the script was chosen as "toRNAdo" (see Appendix C). As described in section 2.12.11, the script requires the RNA-Seq read coverage for every nucleotide in a genome (both DNA strands and each strand separately) as input. This is in addition to another input file containing the coordinates of annotated coding sequences in a bacterial genome. The script was developed based on the RNA-Seq data acquired in Chapter 4, specifically the nutritional

stress RNA-Seq experiment on the Rd strain. A simplified workflow of the script is shown in Figure 5.1, while a more detailed description is presented in following sections.

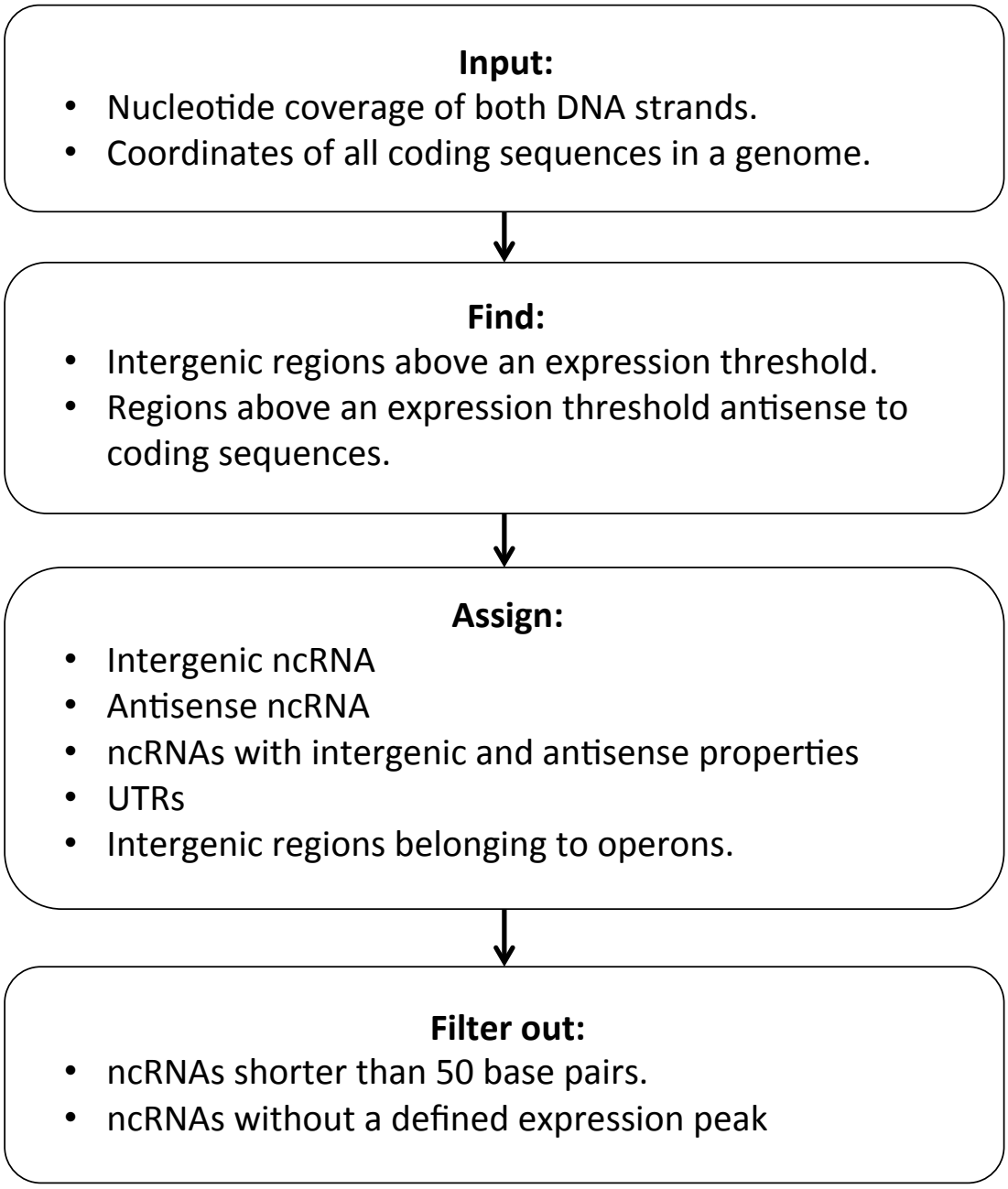

Figure 5.1: Flow chart showing a simplified workflow of the toRNAdo script used for the identification of putative ncRNAs in a bacterial genome.

5.2.1.1 Normalisation of nucleotide coverage

The nucleotide coverage data had to be normalised, so that the script could be applied to different RNA-Seq samples. First, a sum of all nucleotide coverage values in a genome was calculated. Each nucleotide coverage value was subsequently divided by that sum and multiplied by 10^{10}. The formula for the normalised nucleotide coverage (NNC) was:

$$\mathrm{NNC} = \frac{C_n}{\sum_{n \in N} C_n} \times 10^{10}$$

where N stands for a set of all nucleotides in a genome, n stands for each individual nucleotide and C_n is individual nucleotide coverage.

5.2.1.2 Initial detection of UTRs as well as intergenic and antisense regions

Next, the expression threshold needed to be set up in order to reduce background and decrease the number of false positive ncRNAs. After manual inspection of multiple NNC values, the expression threshold was chosen as 100. This value can be user-modified in order to adjust the ncRNA detection stringency.

Starting from the 5' or 3' end of every annotated coding sequence, each nucleotide was tested to see if its NNC value was above the threshold. If it was, then it was assigned to a 5' or 3' UTR and the next nucleotide was in turn tested the same way. This process continued until reaching a nucleotide below the NNC threshold, thus forming a complete UTR (see Figure 5.2A). Next, if the NNC value of a nucleotide, present in an intergenic region and not as part of an assigned UTR, was above the threshold, it was then assigned to an intergenic ncRNA (see Figure 5.2B). Similarly, if the NNC value of a nucleotide, present on the opposite strand to an annotated coding sequence, was above the threshold, it was assigned to an antisense ncRNA (see Figure 5.2C).

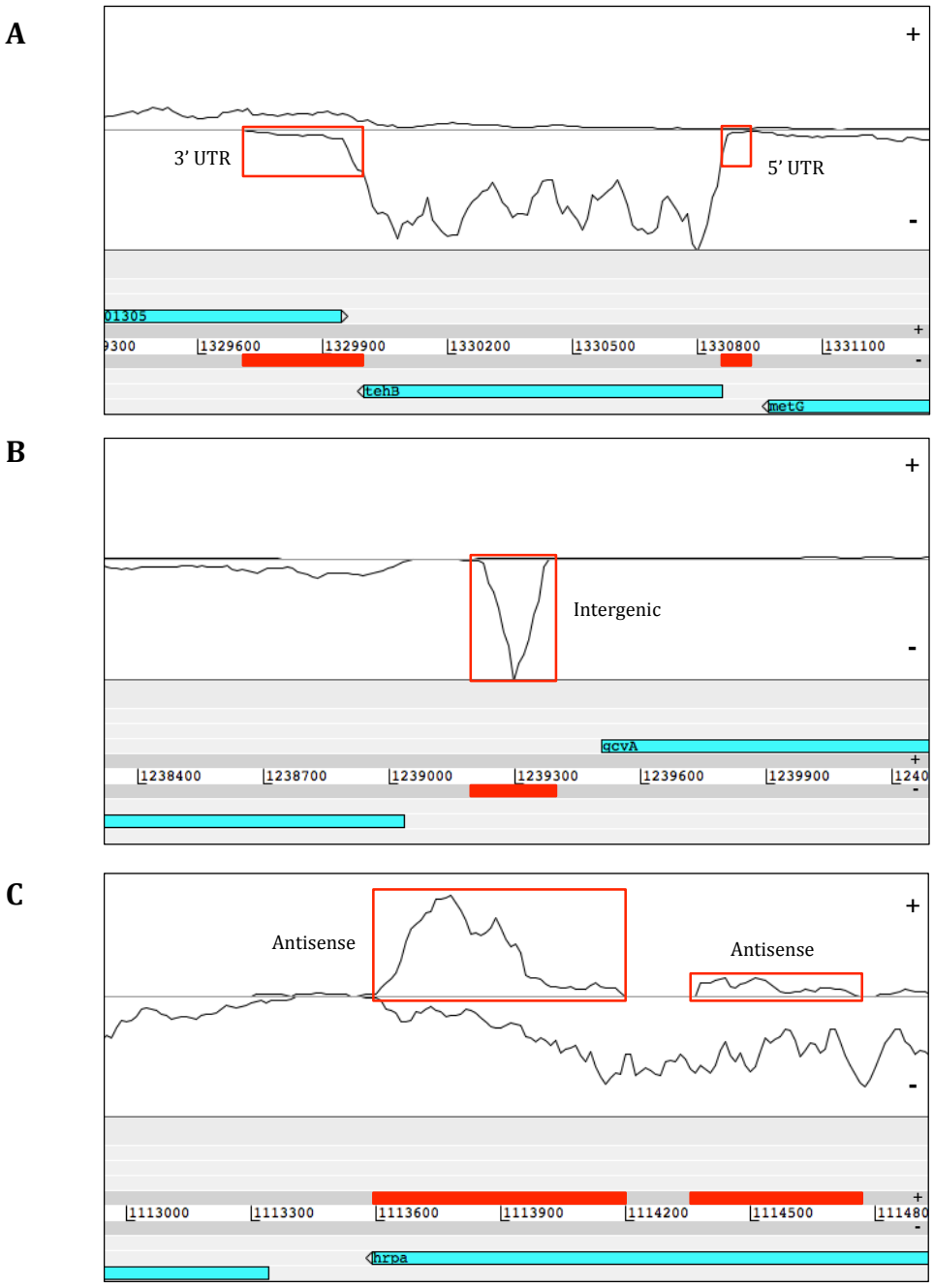

Figure 5.2: Examples of putative ncRNA elements identified by the toRNAdo script, as shown in the Artemis genome browser. Areas of mapped RNA-Seq reads, highlighted in red, represent 5' and 3' UTR (A), intergenic ncRNA (B), antisense ncRNA (C), as identified by the script. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

5.2.1.3 Identification of regions belonging to putative operons

Intergenic regions belonging to putative operons would normally have levels of expression similar to coding sequences. If the NNC value of coding sequences was above the threshold, then the intergenic operon would be assigned as part of a UTR, as described earlier. In fact, the whole intergenic region would be assigned twice: as a 3' end of one gene and as a 5' UTR of the adjacent gene. To fix that, nucleotides that were assigned to both 5' UTR and 3' UTR were thus reassigned to be part of an operon (see Figure 5.3).

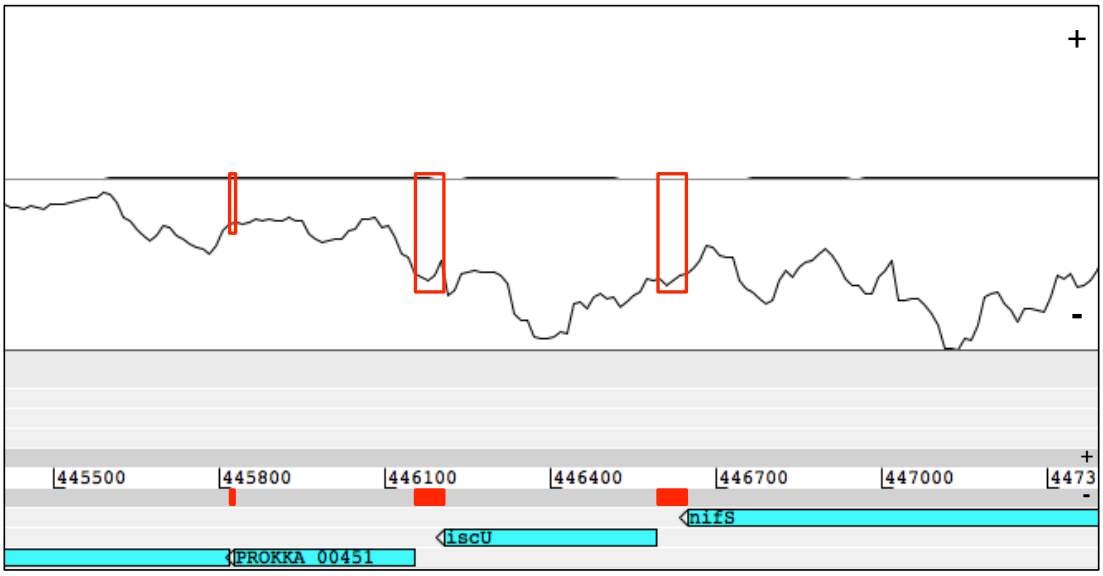

Figure 5.3: Example of an operon element identified by the toRNAdo script. Areas of mapped RNA-Seq reads (highlighted in red) represent intergenic regions that belong to putative operons, as identified by the script and shown in the Artemis genome browser. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

5.2.1.4 Further script optimisation

Quality control of assigned putative ncRNA groups revealed that there were several script errors, which resulted in multiple ncRNAs being misassigned. Therefore, the remaining steps were further improvements and adjustments to the toRNAdo script. First, there were several instances where coding sequences on opposite strands overlapped each other. In that case, the script originally assigned a UTR of one gene as antisense of another (see Figure 5.4). In addition, several UTRs were found to be very long and extended beyond an intergenic region into an antisense region of an adjacent coding sequence on the opposite strand (see Figure 5.5). Hence, the first optimisation step was to reassign misclassified antisense regions to appropriate UTR groups.

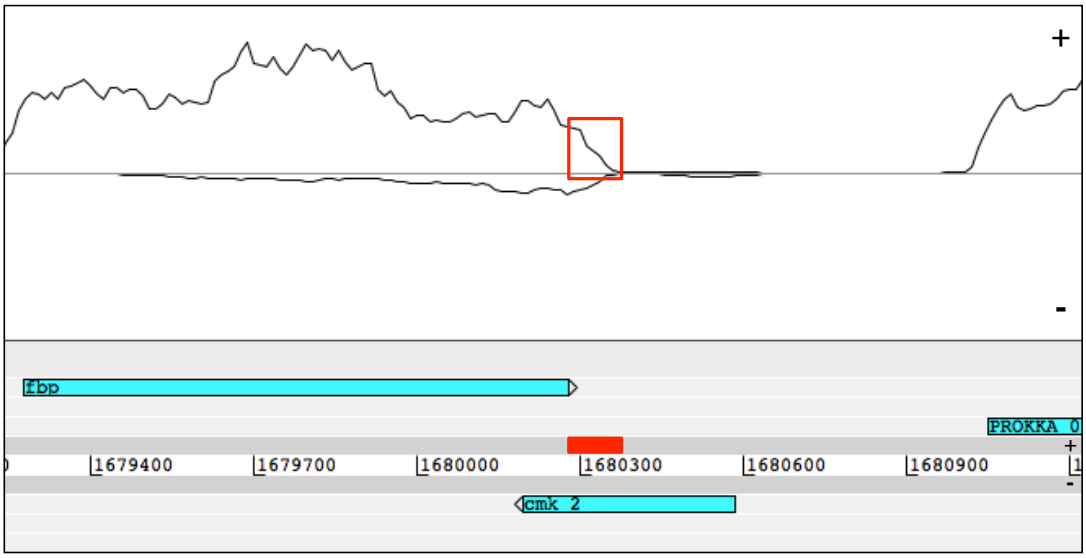

Figure 5.4: Example of a 3' UTR misassigned as an antisense ncRNA by the toRNAdo script. The misassigned 3' UTR is highlighted in red, as depicted in the Artemis genome browser. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

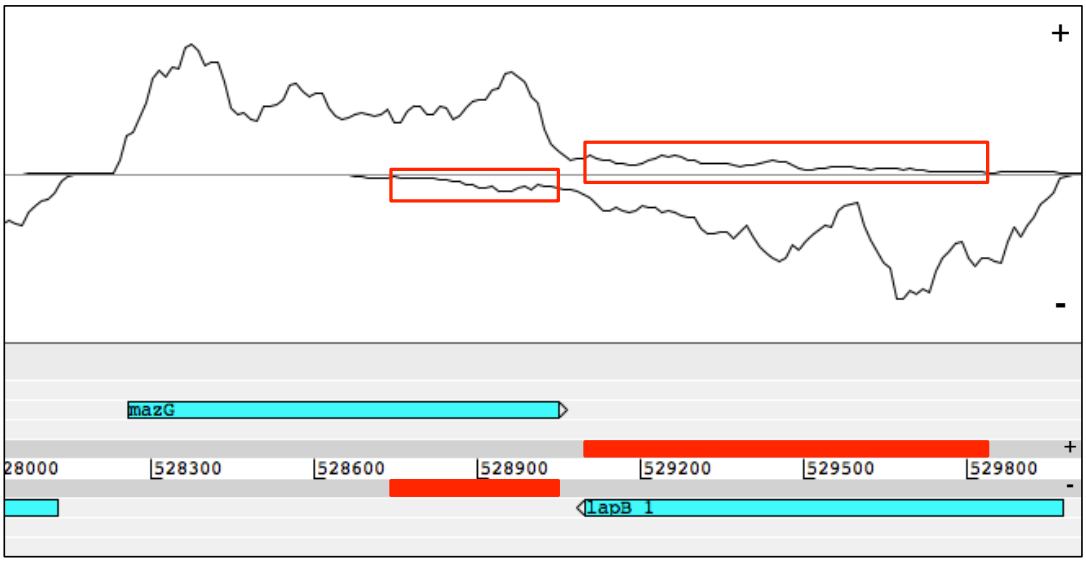

Figure 5.5: Two 3' UTRs misassigned as antisense ncRNAs by the toRNAdo script. Misassigned UTRs are highlighted in red, as depicted in the Artemis genome browser. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

There were multiple intergenic regions that bordered antisense transcripts on the same strand (see Figure 5.6). These putative "mixed" ncRNAs therefore had both intergenic and antisense properties. The toRNAdo script was further modified so that such transcripts were joined together and assigned to a "mixed" ncRNA group. Putative mixed transcripts were subsequently extended to incorporate any adjacent antisense regions that were part of the same transcript (see Figure 5.7).

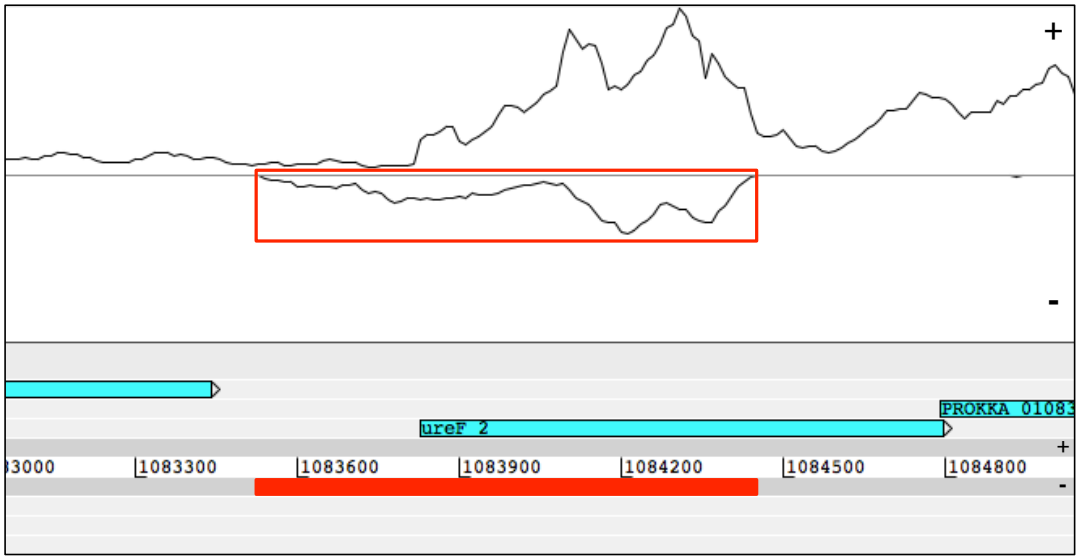

Figure 5.6: Example of a putative "mixed" ncRNA with both intergenic and antisense properties, as assigned by the toRNAdo script. The "mixed" ncRNA is highlighted in red, as shown in the Artemis genome browser. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

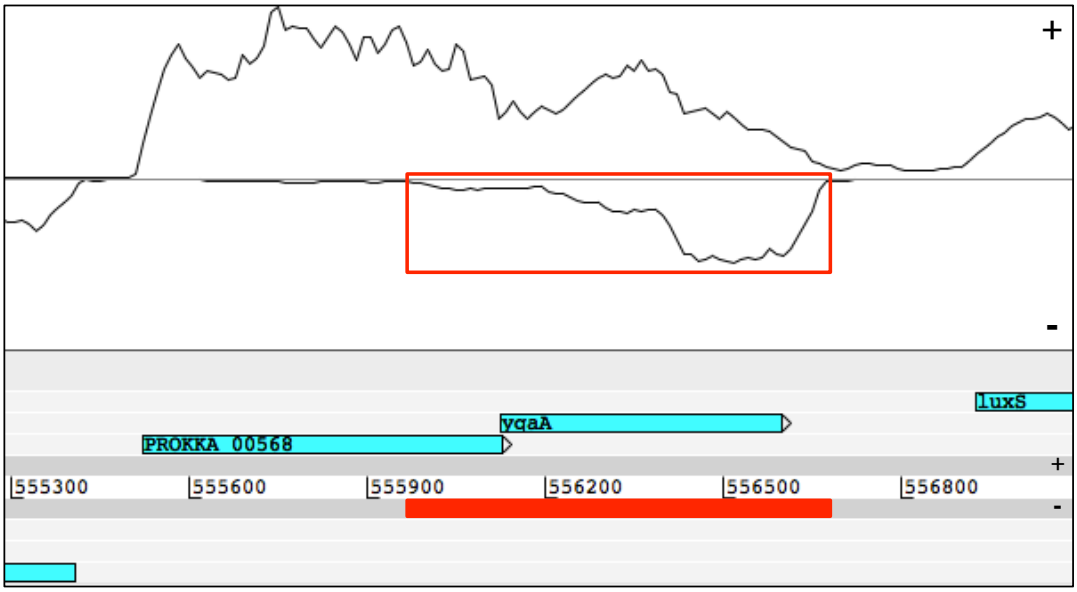

Figure 5.7: Example of a putative "mixed" ncRNA, which spans two antisense regions, as assigned by the toRNAdo script. The "mixed" ncRNA is highlighted in red, as depicted in the Artemis genome browser. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

5.2.1.5 Filtering of ncRNAs based on the presence of an expression "peak"

There were several instances where an assigned UTR contained an expression "peak", which was clearly a separate RNA element to an mRNA of the coding sequence (see Figure 5.8A). In such a case, it was necessary to reassign this peak to an appropriate ncRNA group (intergenic, antisense or mixed). However, since there was variation in the size of different peaks, a threshold needed to be set in order to decide which peaks are separate ncRNAs and which are still likely to be part of UTRs. An expression ratio between the highest and lowest points of the peak was used for choosing this threshold. Following manual inspection, it was deemed appropriate that if the NNC ratio between the highest and lowest points of the peak was at least five to one, then the peak was reassigned as intergenic, antisense or mixed ncRNA (see Figure 5.8A). A similar scenario was also occurring in intergenic regions that had been previously assigned to an operon group by the toRNAdo script. Here, the presence of the peak, expression of which was higher than that of flanking coding sequences, likely meant that it was a separate RNA element and was not part of an operon (see Figure 5.8B). Therefore, if the NNC value of the highest point of that peak was at least five times larger than its lowest points at both 3' and 5' ends, then the peak was reassigned to an appropriate ncRNA group (intergenic, antisense or mixed).

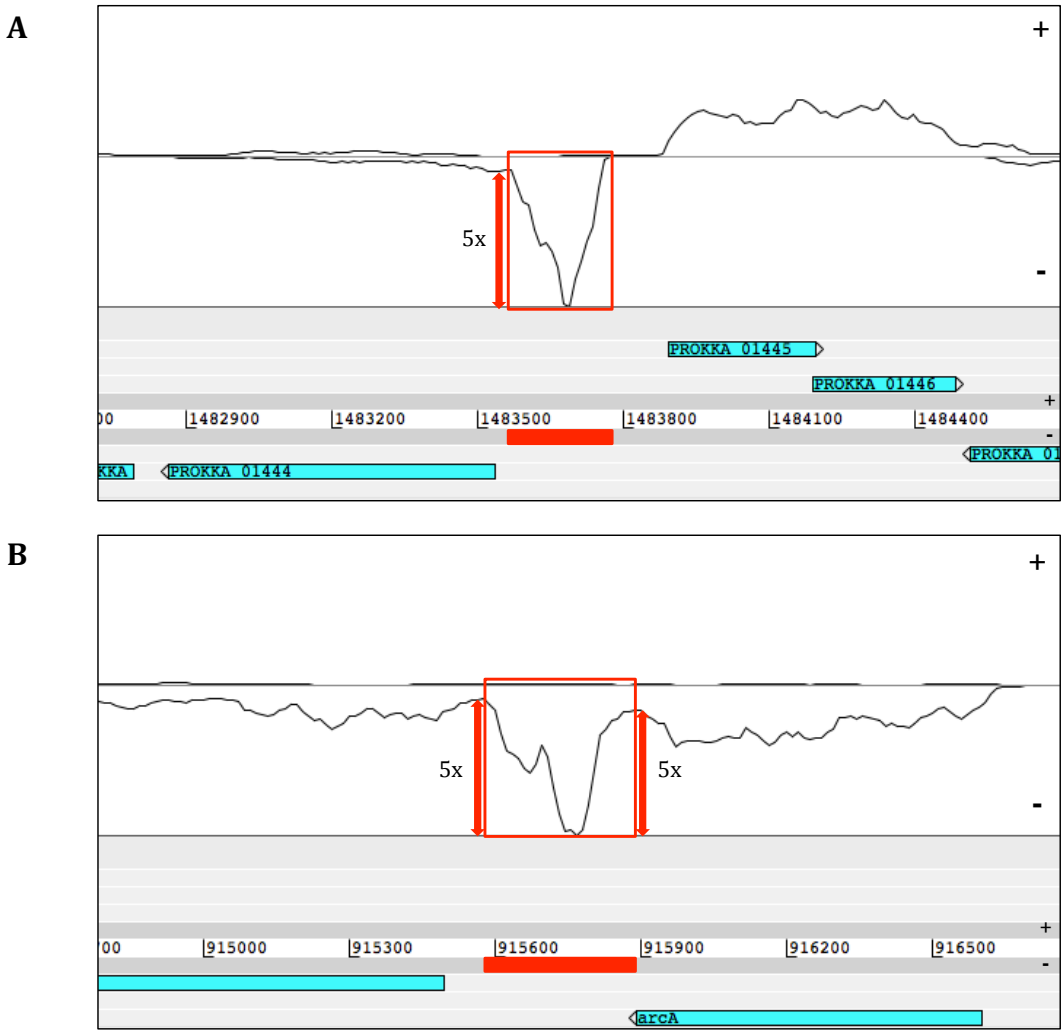

Figure 5.8: Example of expression "peaks" present in UTRs and operon regions, as identified by the toRNAdo script. The expression peaks in UTRs (A) and operon regions (B) are highlighted in red, as shown in the Artemis genome browser. The expression of the highest point of each peak is at least five times higher than the expression of the lowest point. Both peaks were reassigned to an intergenic ncRNA group. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

The presence of an expression peak meant greater transcriptional activity and thus a higher likelihood that a putative ncRNA was real and not part of background expression. With that in mind, the same peak-size threshold of a five-to-one ratio was applied to all identified antisense, intergenic and mixed putative ncRNAs. Only ncRNAs that possessed the expression peak were considered for further analyses, though putative ncRNAs below the threshold were also retained for any other user needs. The peak-size threshold can be user-modified in order to adjust the stringency of the ncRNA detection.

5.2.1.6 Further filtering by length and output file generation

Putative ncRNAs were further filtered by length to exclude anything below 50 bp, which was likely to be transcriptional noise. The minimum ncRNA length can be user-modified. The script output consisted of putative intergenic, antisense and mixed ncRNAs (below and above the peak-size threshold), predicted 5' and 3' UTRs as well as putative regions belonging to operons. All files generated by the script were in the WIG format, which can be imported in the Artemis genome browser as a "User plot" (see Figure 5.9). All expression values for each nucleotide of a putative ncRNA from the script output were NNC.

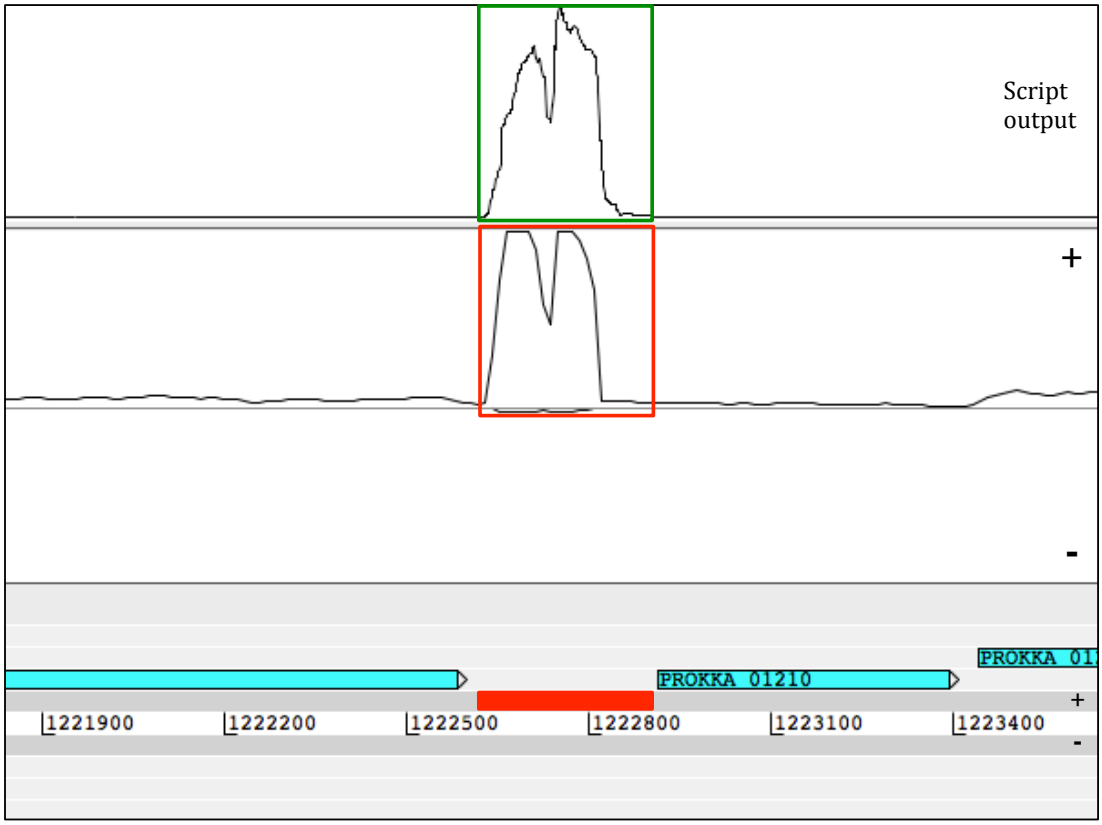

Figure 5.9: The toRNAdo script output in the WIG format, as imported into the Artemis genome browser. An area of mapped RNA-Seq reads, highlighted in red, represents a putative intergenic ncRNA, which is correctly identified by the toRNAdo script (highlighted in green). Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

5.2.2 Identification and analysis of ncRNAs in Rd and R2866 strains of *H. influenzae*

5.2.2.1 Discovery of putative ncRNAs in Rd and R2866 from the RNA-Seq data

The toRNAdo script was used to identify putative ncRNAs (intergenic, antisense and mixed) across different infection-relevant RNA-Seq conditions in both Rd and R2866 strains. These conditions were mid-exponential phase, stationary phase, oxidative stress, iron-starvation stress, and nutritional stress (Rd only) (see Chapter 4). Putative ncRNAs were only considered valid if they were present in all replicates of a particular condition. The putative ncRNA data were then merged across all tested RNA-Seq conditions for each strain. When combining ncRNAs present in multiple replicates or conditions, a minimum start position and a maximum end position were used to define the new length, in order to cover all nucleotides that were putatively part of each ncRNA.

First, 168 and 169 putative ncRNAs were identified in Rd and R2866 strains respectively across all tested RNA-Seq conditions. These were subsequently inspected manually to find any false positive sequences. Two ncRNAs in Rd and four in R2866 were filtered out, as they contained repetitive sequences, which were likely to have false expression values due to errors in RNA-Seq read alignment. This was in addition to one putative ncRNA in Rd and two in R2866, which were misclassified as ncRNAs by the toRNAdo script. Therefore, the final number of true putative ncRNAs was 165 for Rd and 163 for R2866. Of these, 36 were intergenic, 37 were antisense and 92 were mixed ncRNAs in Rd, while 39 were intergenic, 35 were antisense and 89 were mixed ncRNAs in R2866 (see Table 5.1). The majority of putative ncRNAs were identified at stationary phase. The uniform genomic distribution of all putative ncRNAs as well as in each RNA-Seq condition individually is shown in Figures 5.10 and 5.11 for Rd and R2866 respectively. Complete tables with putative ncRNA coordinates, length and type are presented in Appendix D.

Table 5.1: Number of putative ncRNAs identified in Rd and R2866 strains.

Strain	RNA-Seq condition	Total ncRNAs	Intergenic ncRNAs	Antisense ncRNAs	Mixed ncRNAs
Rd	All conditions	165	36	37	92
	Mid-exponential phase	36	14	6	16
	Stationary phase	139	28	32	79
	Iron-starvation stress	83	25	15	43
	Oxidative stress	71	21	11	39
	Nutritional stress	80	17	19	44
R2866	All conditions	163	39	35	89
	Mid-exponential phase	33	14	3	16
	Stationary phase	139	30	32	77
	Iron-starvation stress	93	31	15	47
	Oxidative stress	64	26	9	29

Figure 5.10: Positions of putative ncRNAs in the Rd genome. The locations of ncRNAs are shown in six inner concentric circles. The outer circle shows the positions of all identified ncRNAs. The remaining five circles show the positions of ncRNAs that were identified in individual conditions. The circles represent (going inward): 1) stationary phase, 2) oxidative stress, 3) iron-starvation stress, 4) nutritional stress, 5) mid-exponential phase.

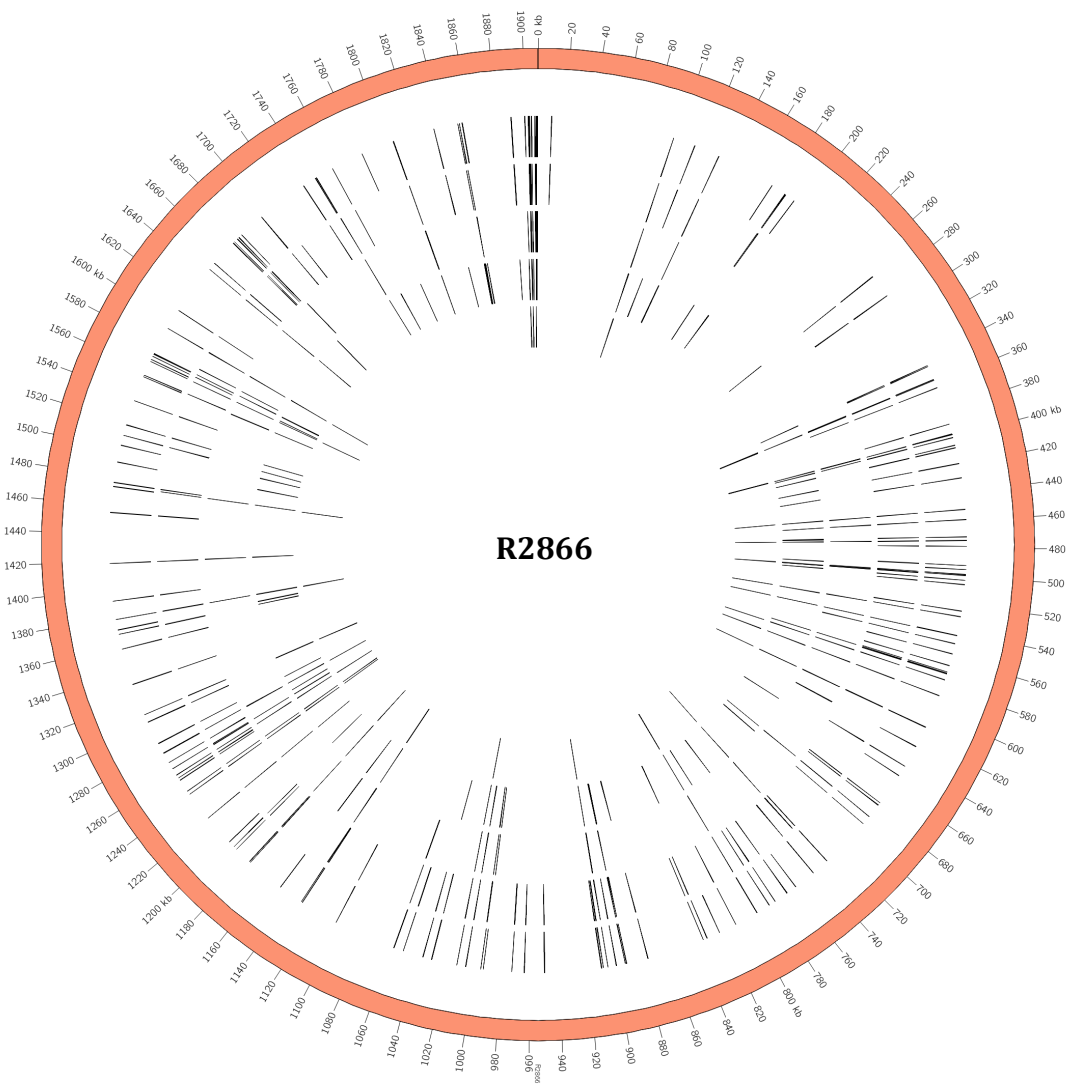

Figure 5.11: Positions of putative ncRNAs in the R2866 genome. The locations of ncRNAs are shown in six inner concentric circles. The outer circle shows the positions of all identified ncRNAs. The remaining four circles show the positions of ncRNAs that were identified in individual conditions. The circles represent (going inward): 1) stationary phase, 2) oxidative stress, 3) iron-starvation stress, 4) mid-exponential phase.

The Rfam database can be used to search for bacterial ncRNA families, including characterised sRNAs (Nawrocki et al., 2015). In this study, it was used to investigate if any of the putative ncRNAs were homologous to any known RNA elements. Three such ncRNAs were identified in both Rd and R2866: they were homologous to 6S RNA (Rd_082; R2866_132), RNase P RNA subunit (Rd_152; R2866_062) and GcvB (Rd_114; R2866_097) ncRNAs. 6S RNA was previously shown to be involved in the control of gene expression during stationary phase in *E. coli* (Wassarman and Storz, 2000). RNase P is a ribonucleic enzyme involved in the processing of tRNAs (Evans et al., 2006). Finally, GcvB is a well-characterised bacterial sRNA that has been shown to regulate amino acid uptake and metabolism (Sharma et al., 2011).

HrrF, the only discovered and validated sRNA in the 86-028NP strain of *H. influenzae*, was not identified as one of the putative ncRNAs in this study (Santana et al., 2014). Manual examination of its homologue in Rd and R2866 revealed it to have similar expression to adjacent genes in all tested RNA-Seq conditions (see Figure 5.12). This meant that the toRNAdo script classified it as part of an operon instead of a separate intergenic ncRNA.

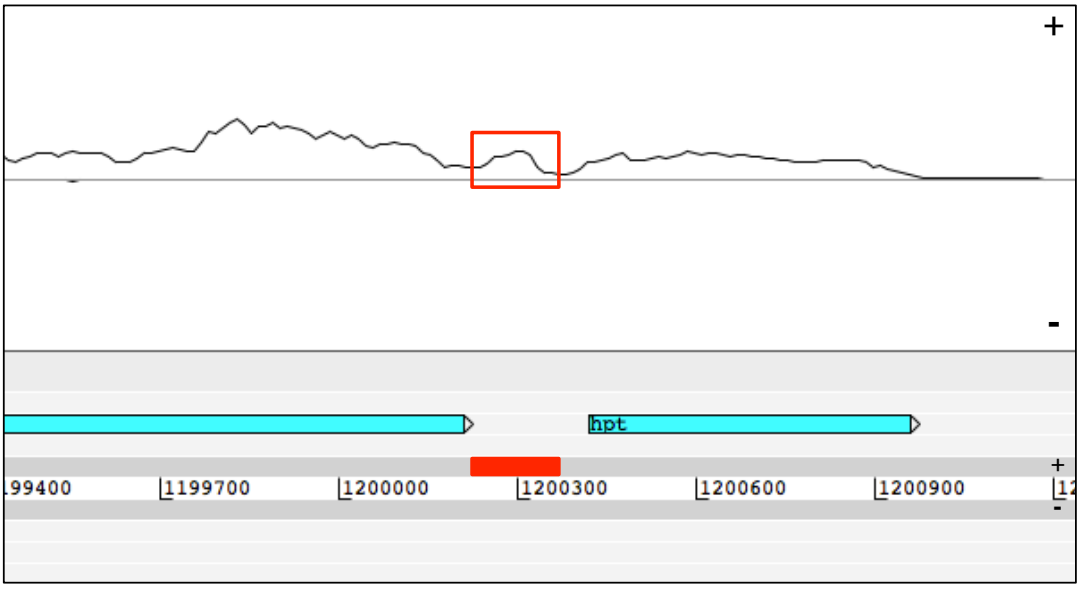

Figure 5.12: An HrrF homologue in an Rd replicate during iron-starvation stress. The homologue is highlighted in red. Light blue rectangles are coding sequences. Forward and reverse strands are represented by "+" and "-" symbols respectively.

5.2.2.2 Absolute expression analysis of putative ncRNAs

The expression of putative ncRNAs, along with all protein-coding genes, was normalised using the TPM method. The highest expression of ncRNAs in both strains was observed during stationary phase, followed by iron-starvation stress (see Figure 5.13). Analysis of separate ncRNA groups revealed that antisense ncRNAs had the lowest expression, while intergenic ncRNAs were expressed at the highest levels (see Figures 5.14; 5.15). The highest expression was again at stationary phase for each of the separate ncRNA groups. The two most highly expressed putative ncRNAs in all conditions and in both strains were 6S RNA and RNase P homologues.

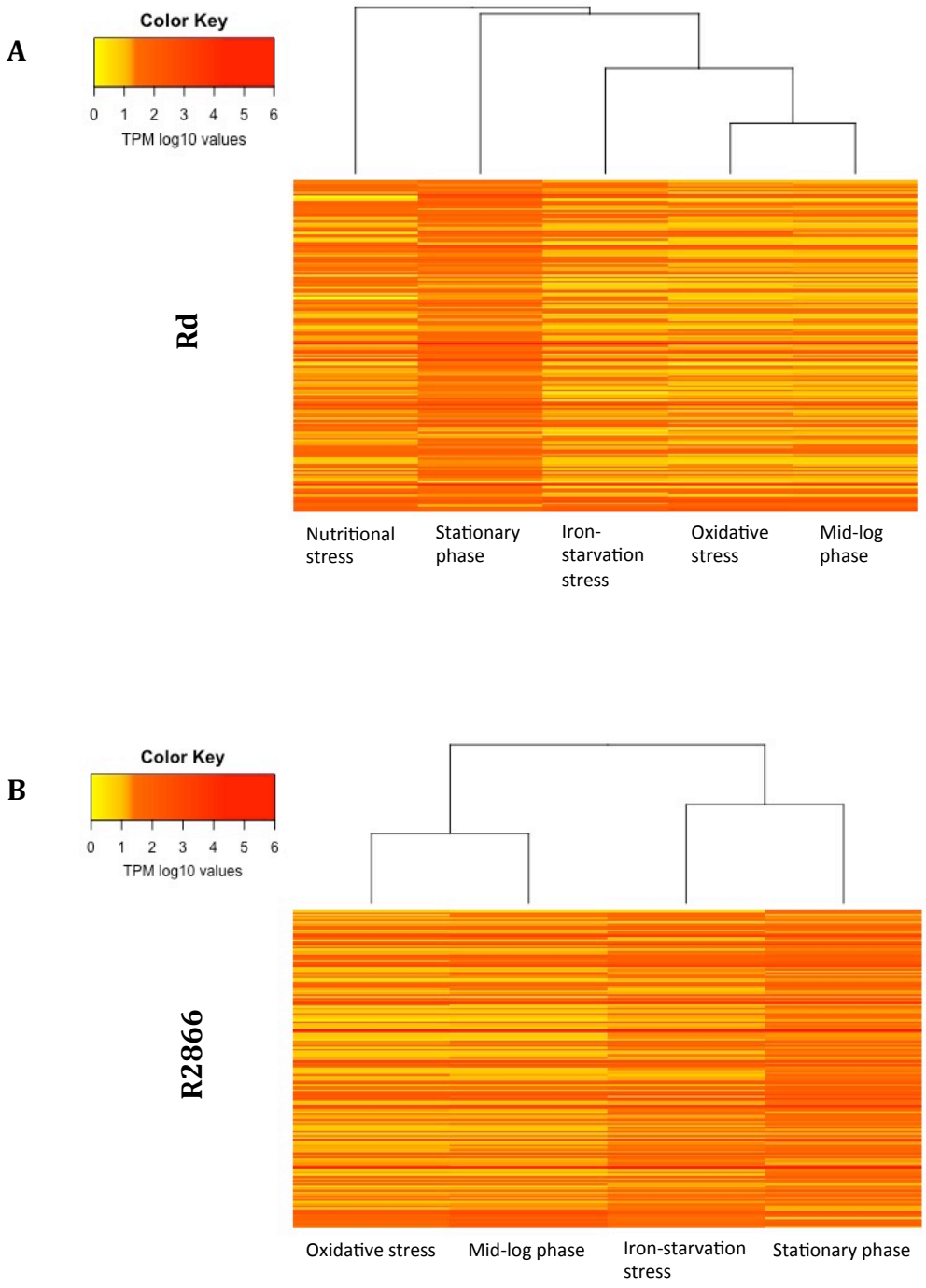

Figure 5.13: Heatmaps of log_{10}-transformed TPM expression values of all putative ncRNAs in Rd and R2866. RNA-Seq conditions were clustered based on the euclidean distance between log_{10}-transformed TPM expression of all ncRNAs in Rd (A) and R2866 (B)

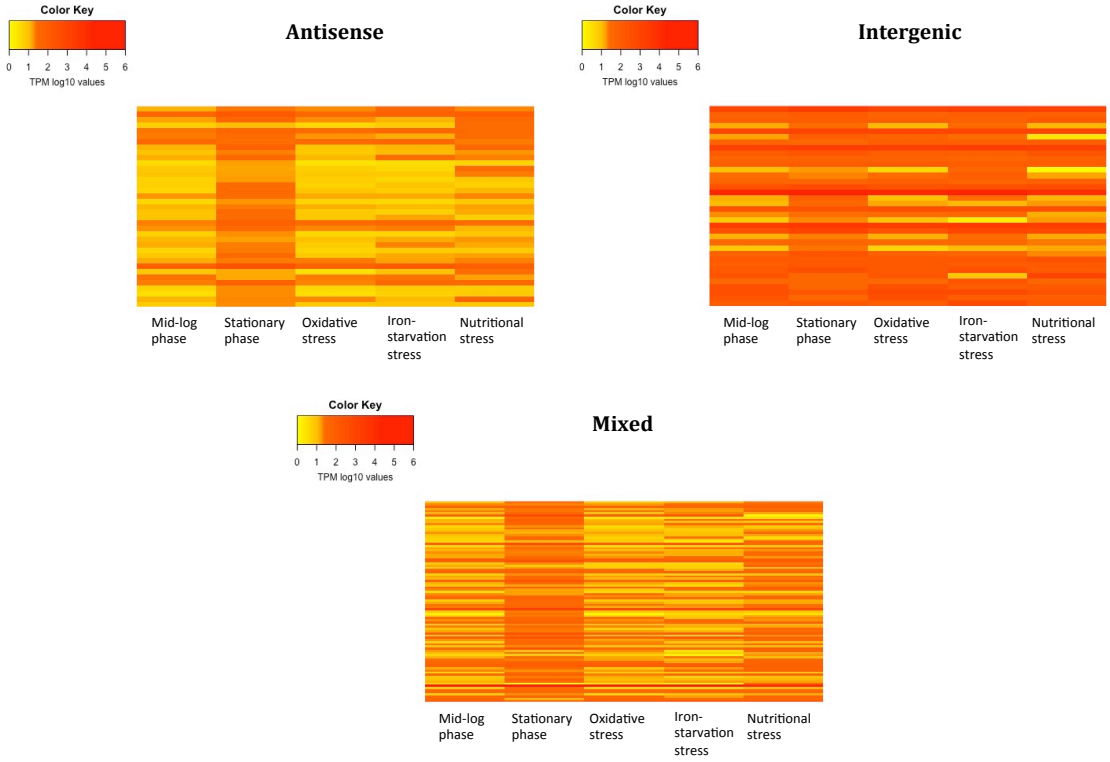

Figure 5.14: Heatmaps of log_{10}-transformed TPM expression values of antisense, intergenic and mixed putative ncRNAs in the Rd strain. RNA-Seq conditions were clustered based on the euclidean distance between log_{10}-transformed TPM expression of all ncRNAs.

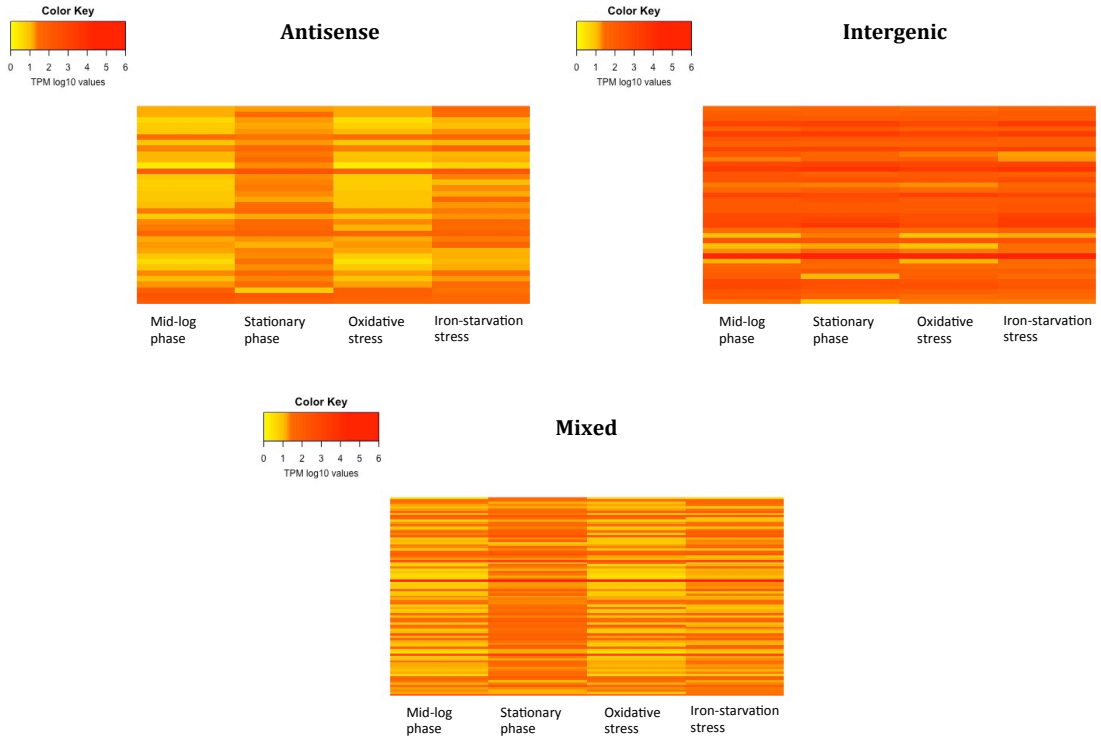

Figure 5.15: Heatmaps of log_{10}-transformed TPM expression values of antisense, intergenic and mixed putative ncRNAs in the R2866 strain. RNA-Seq conditions were clustered based on the euclidean distance between log_{10}-transformed TPM expression of all ncRNAs.

5.2.2.3 Differential expression of putative ncRNAs across infection-relevant conditions

Next, differential expression analysis of all putative ncRNAs in infection-relevant conditions was carried out. The number of differentially expressed ncRNAs, along with those common to more than one infection-relevant condition, is depicted in Figure 5.16. There was a much higher number of up-regulated ncRNAs during stationary phase, when compared to other conditions: 109 in Rd and 97 in R2866. 23 ncRNAs were down-regulated during nutritional stress in Rd, while only one was up-regulated. Figures 5.17 and 5.18 show genomic positions of all ncRNAs that were differentially expressed in each of the infection-relevant conditions. Full tables of all differentially expressed ncRNAs are presented in Appendix D.

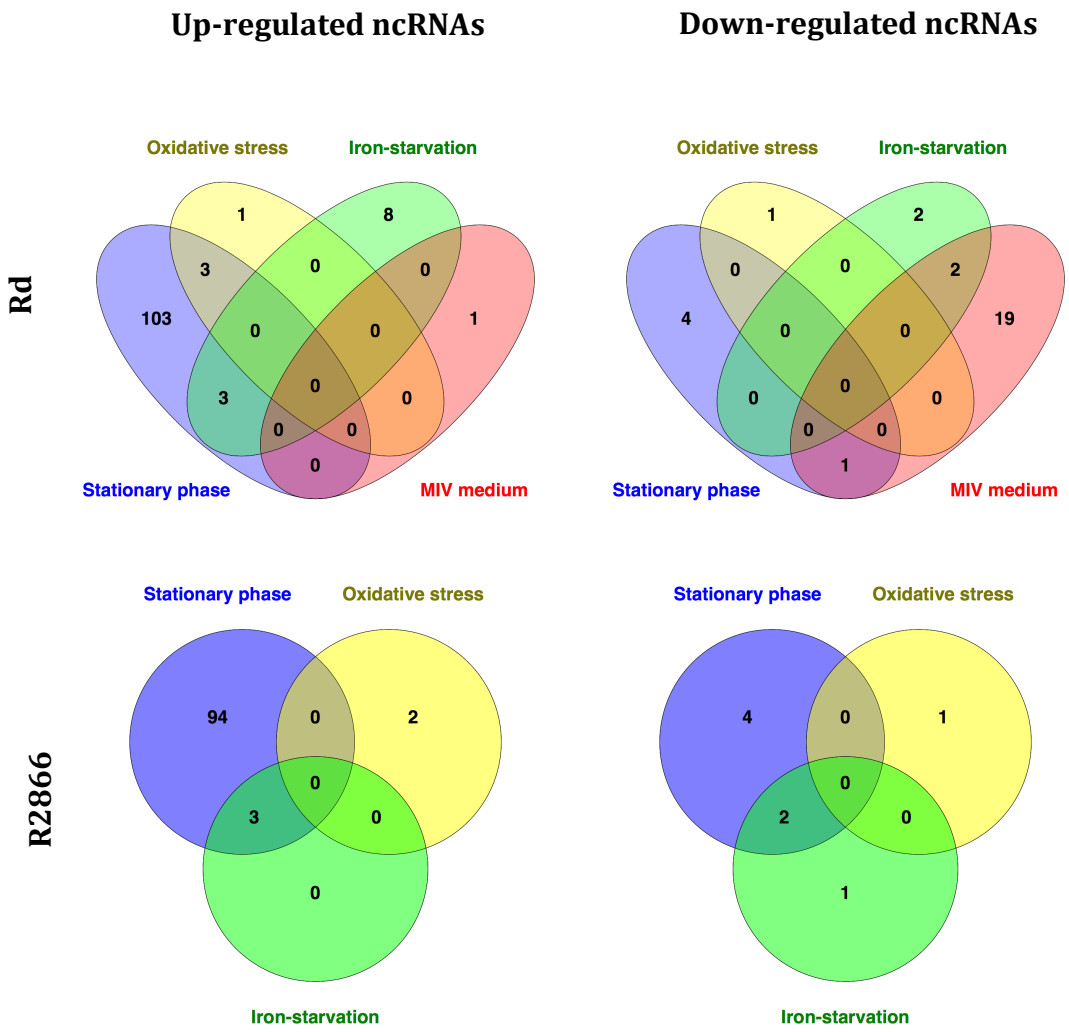

Figure 5.16: Venn diagrams of common differentially expressed ncRNAs across different RNA-Seq experiments in Rd and R2866 strains.

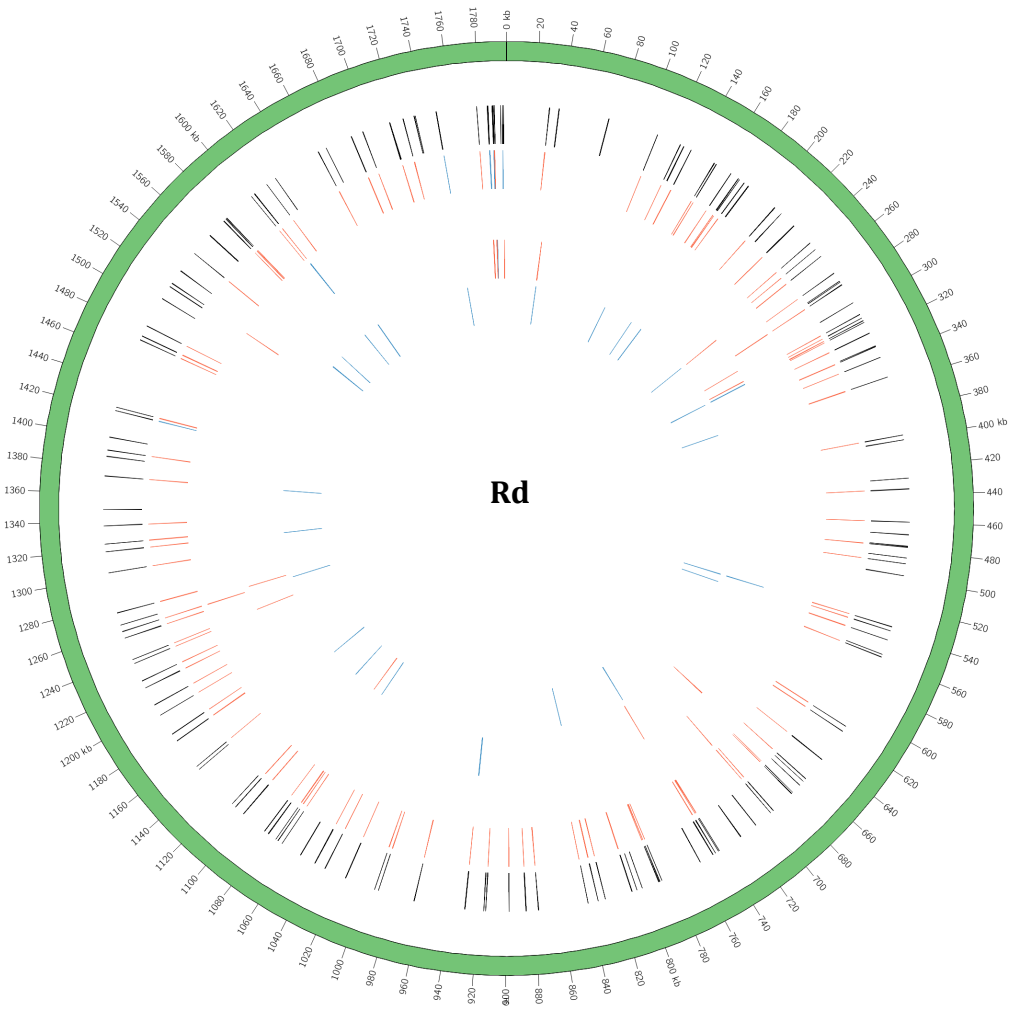

Figure 5.17: Positions of differentially expressed ncRNAs in the Rd genome. The locations of ncRNAs are shown in five inner concentric circles. The outer circle shows the positions of all identified ncRNAs. The remaining four circles show the positions of ncRNAs that were up-regulated (red strokes) and down-regulated (blue strokes) in individual conditions as compared to control. The circles represent (going inward): 1) stationary phase, 2) oxidative stress, 3) iron-starvation stress, 4) nutritional stress.

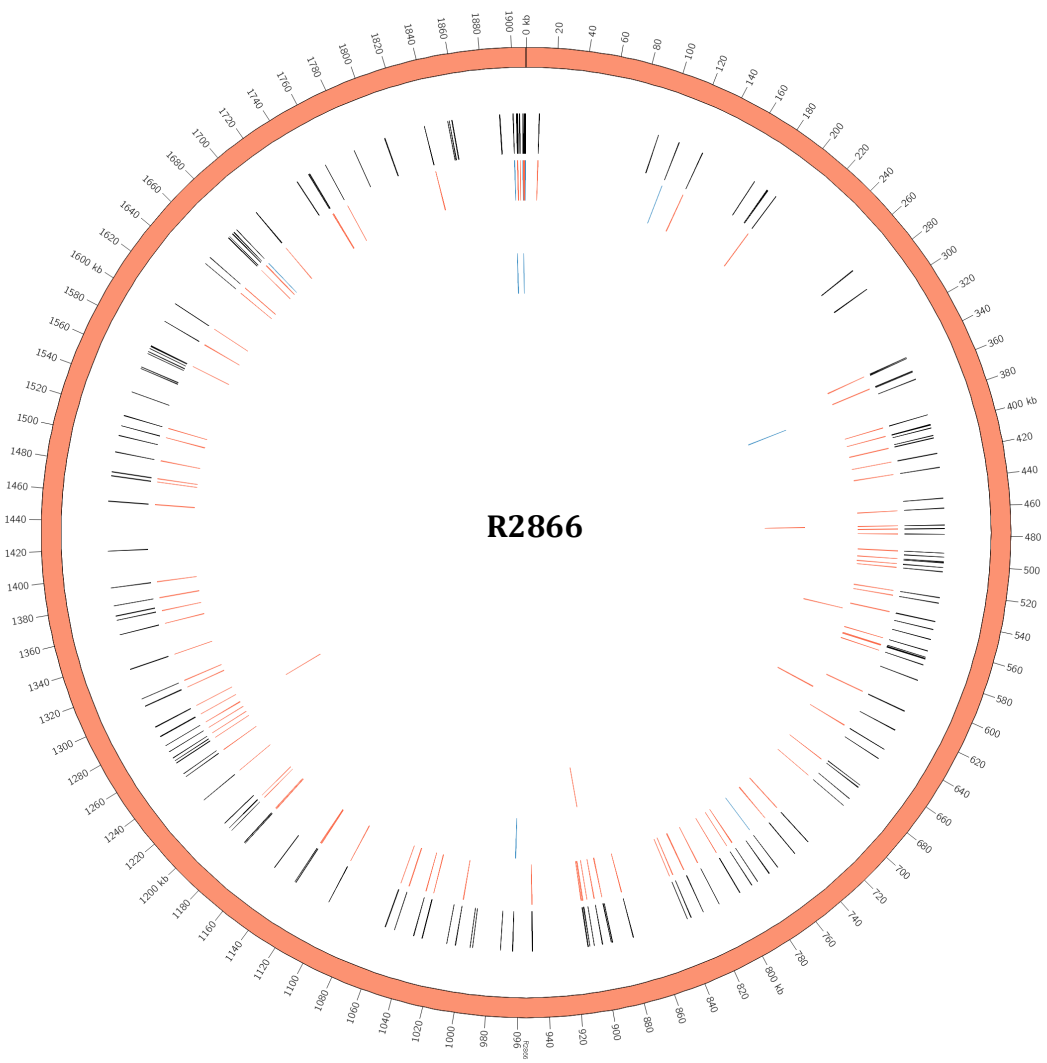

Figure 5.18: Positions of differentially expressed ncRNAs in the R2866 genome. The locations of ncRNAs are shown in four inner concentric circles. The outer circle shows the positions of all identified ncRNAs. The remaining three circles show the positions of ncRNAs that were up-regulated (red strokes) and down-regulated (blue strokes) in individual conditions as compared to control. The circles represent (going inward): 1) stationary phase, 2) oxidative stress, 3) iron-starvation stress.

Several ncRNAs were differentially expressed in more than one RNA-Seq condition (see Figure 5.16). Among these was a GcvB homologue, which was up-regulated during stationary phase and iron-starvation (nearly 5-fold) in the Rd strain (see Appendix D). In R2866, it was only up-regulated during stationary phase. None of the differentially expressed ncRNAs were common to all tested RNA-Seq conditions.

As *cis*-acting sRNAs are normally present as antisense transcripts, it was interesting to examine which putative antisense ncRNAs were differentially expressed in this study (see section 1.2.1). Several ncRNAs in the Rd strain, antisense to putative hemoglobin-binding and transferrin-binding proteins, were up-regulated during iron-starvation and stationary phase, but down-regulated during nutritional stress. Among other notable ncRNAs, which were up-regulated at stationary phase, was ncRNA antisense to the catalase gene *hktE* in Rd as well as ncRNAs antisense to the iron-acquisition gene *hitB* and several pilus genes in R2866. 6S RNA and RNase P homologues were up-regulated during stationary phase in both strains.

5.2.2.4 Homology of putative ncRNAs between Rd and R2866

BLASTn was used to identify which putative ncRNAs were conserved between Rd and R2866 strains. Just over 40% of putative ncRNAs were found in both Rd and R2866 strains by the toRNAdo script (see Table 5.2). However, over 80% of ncRNAs in Rd had homologous sequences in the R2866 genome, while over 70% of ncRNAs in R2866 had homologous sequences in Rd. Figure 5.19 shows positions of all homologous ncRNA sequences in Rd and R2866 whole genomes.

Table 5.2: Homology of putative ncRNAs between Rd and R2866 strains based on the BLASTn analysis. E-value cut-off for homology was 1e-05.

Strain	Percentage (%) of ncRNAs homologous to:	
	Putative ncRNAs in the other strain	Any sequence in the other strain
Rd	43.6	81.8
R2866	41.7	73.0

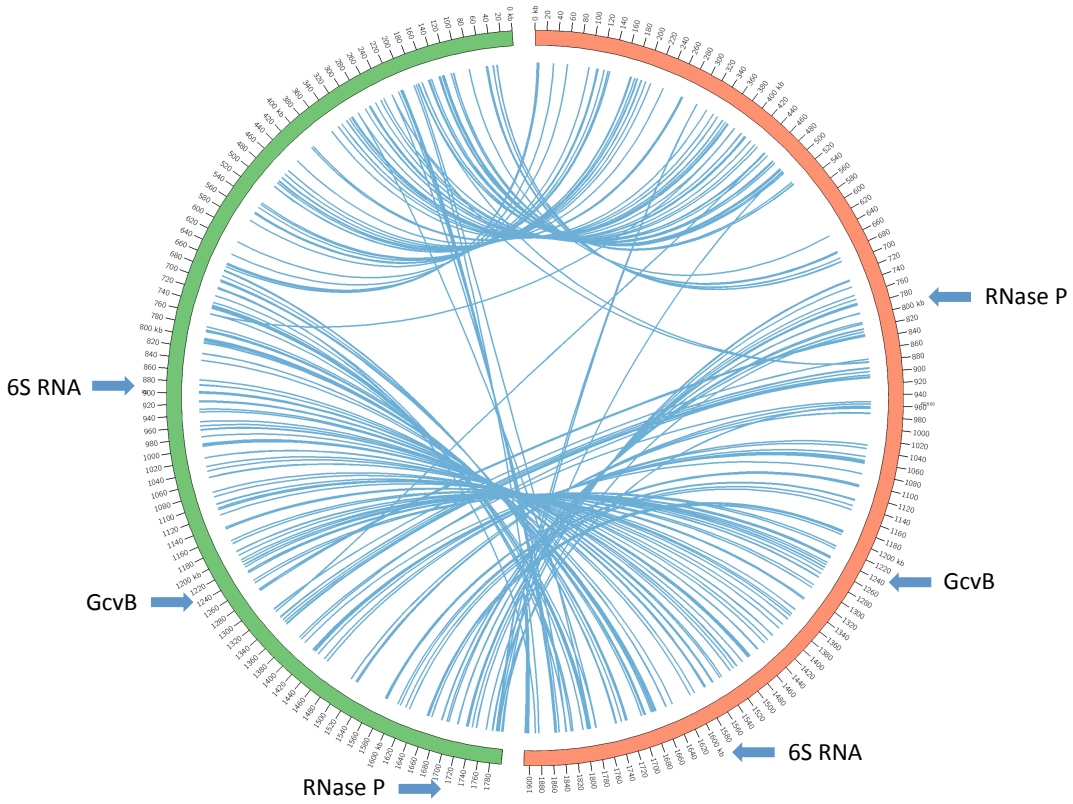

Figure 5.19: Homology of putative ncRNAs between Rd and R2866. A circular diagram shows two chromosomes of Rd (green) and R2866 (red) strains, linked by homologous ncRNAs (blue lines), based on the BLASTn analysis. Annotated ncRNA homologues were identified from the Rfam database.

5.2.2.5 Identification of potential protein-coding genes among putative ncRNAs

The Prokka software, used in this study to annotate whole re-sequenced genomes of Rd and R2866, did not predict protein-coding genes shorter than 250 bp. It is therefore possible that some true short protein-coding sequences were missed in the genome annotation. If they contained expression "peaks" in the RNA-Seq data, then they would have likely been identified as putative ncRNAs by the toRNAdo script. Thus it was important to investigate if any of the putative ncRNAs in this study were in fact potential protein-coding genes.

In order to classify a putative ncRNA as a potential protein-coding gene, an ncRNA needed to possess a start codon (AUG, GUG or UUG) within an open reading frame, covering the expression of the ncRNA as observed from the RNA-Seq data. After manually inspecting all putative ncRNAs, 25 potential protein-coding genes were identified in Rd, while 23 were identified in R2866 (see Tables 5.3; 5.4). Several sequences did not produce any hits, while the majority of top BLASTx hits for the remaining sequences were hypothetical proteins.

Table 5.3: Putative ncRNAs, present in the Rd strain, which contain a start codon and an open reading frame, suggesting the presence of a protein-coding gene. Top BLASTx are shown along with the name of the species containing a putative homologue.

ID	Start position	End position	Strand	Top BLASTx hit (species name)
Rd_009	155135	155395	+	-
Rd_018	227136	227665	-	Hypothetical protein (*H. influenzae*)
Rd_027	281280	281477	+	Hypothetical protein (*Haemophilus aegyptius*)
Rd_034	320889	321185	+	-
Rd_054	557547	558857	+	Integrase (*H. influenzae*)
Rd_059	662409	662498	-	Ornithine carbamoyltransferase (*H. influenzae*)
Rd_065	708324	708556	-	Hypothetical protein (*Haemophilus* sp. oral taxon 851)
Rd_066	721899	722015	+	Glycerophosphodiester phosphodiesterase (*H. influenzae*)
Rd_082	885347	885661	+	-
Rd_086	915521	915888	-	Hypothetical protein (*H. influenzae*)
Rd_090	994816	995745	+	-
Rd_105	1173203	1173472	-	Membrane protein (*H. influenzae*)
Rd_106	1173213	1173455	+	-
Rd_109	1203265	1203822	+	-
Rd_120	1318219	1318845	-	Hypothetical protein (*H. aegyptius*)
Rd_121	1323679	1323971	+	Hypothetical protein (*H. influenzae*)
Rd_135	1518769	1518894	-	ATP-dependent metalloprotease, partial (*H. influenzae*)
Rd_137	1545292	1545803	-	-
Rd_145	1624072	1624299	-	-
Rd_147	1666018	1666223	-	-
Rd_151	1714817	1715360	+	Hypothetical protein (*Mannheimia haemolytica*)
Rd_158	1785066	1785298	+	Hypothetical protein (*H. influenzae*)
Rd_160	1788561	1789102	-	Mercuric ion scavenger protein (*H. influenzae*)
Rd_163	1794537	1795072	+	Cytidylate kinase, partial (*H. influenzae*)
Rd_165	1796507	1796828	+	Hypothetical protein (*H. influenzae*)

Table 5.4: Putative ncRNAs, present in the R2866 strain, which contain a start codon and an open reading frame, suggesting the presence of a protein-coding gene. Top BLASTx are shown along with the name of the species containing a putative homologue.

ID	Start position	End position	Strand	Top BLASTx hit (species name)
R2866_017	397982	398500	-	Hypothetical protein (*H. aegyptius*)
R2866_018	398271	398538	+	-
R2866_024	452241	452552	+	Hypothetical protein (*H. aegyptius*)
R2866_025	452279	452786	-	Hypothetical protein, partial (*H. influenzae*)
R2866_028	474416	474668	-	-
R2866_034	499358	500645	+	Uncharacterized protein (*H. influenzae*)
R2866_040	548285	548624	-	Transcriptional regulator (*H. influenzae*)
R2866_044	570011	570339	-	-
R2866_063	791708	792149	-	hypothetical protein (*M. haemolytica*)
R2866_065	821541	822128	-	Conserved hypothetical protein (*H. influenzae*)
R2866_068	875603	876167	-	Hypothetical protein (*H. aegyptius*)
R2866_079	995263	995765	-	Hypothetical protein (*H. influenzae*)
R2866_080	1006041	1006276	+	-
R2866_105	1277352	1278405	-	-
R2866_108	1307763	1308096	+	Hypothetical protein (*H. influenzae*)
R2866_109	1307782	1308061	-	-
R2866_118	1473212	1475161	-	-
R2866_125	1554103	1554489	+	Hypothetical protein (*H. influenzae*)
R2866_128	1567612	1567841	-	Polysaccharide polymerase (*H. influenzae*)
R2866_132	1592489	1592795	-	-
R2866_133	1592505	1592681	+	-
R2866_147	1780641	1780933	-	Trk system potassium uptake protein TrkH (*H. influenzae*)
R2866_162	1908264	1908555	+	-

5.2.2.6 Validation of two ncRNAs present in R2866

To identify whether any of the putative ncRNAs were actually present in *H. influenzae* as distinct RNA transcripts, they needed to be experimentally validated. For that purpose, two putative ncRNAs from an invasive strain R2866 were chosen for validation using the northern blot technique. Selection of these ncRNAs was based on high TPM expression values, in order to maximise the probability of transcript detection. Chosen ncRNAs were R2866_101 and R2866_118, both of which were up-regulated over 4-fold during stationary phase in R2866 (see Figure 5.20) (see Appendix D). The latter ncRNA was interesting, as it formed two large adjacent peaks, one of which contained an opening reading frame with a start codon and could therefore potentially encode a small protein (see Table 5.4). Northern blotting revealed both ncRNAs to be present as RNA transcripts of 150-200 bp in size, which was consistent with the size of ncRNA sequences (see Figure 5.21). These transcripts were present in all three tested biological replicates at mid-exponential and stationary growth phases.

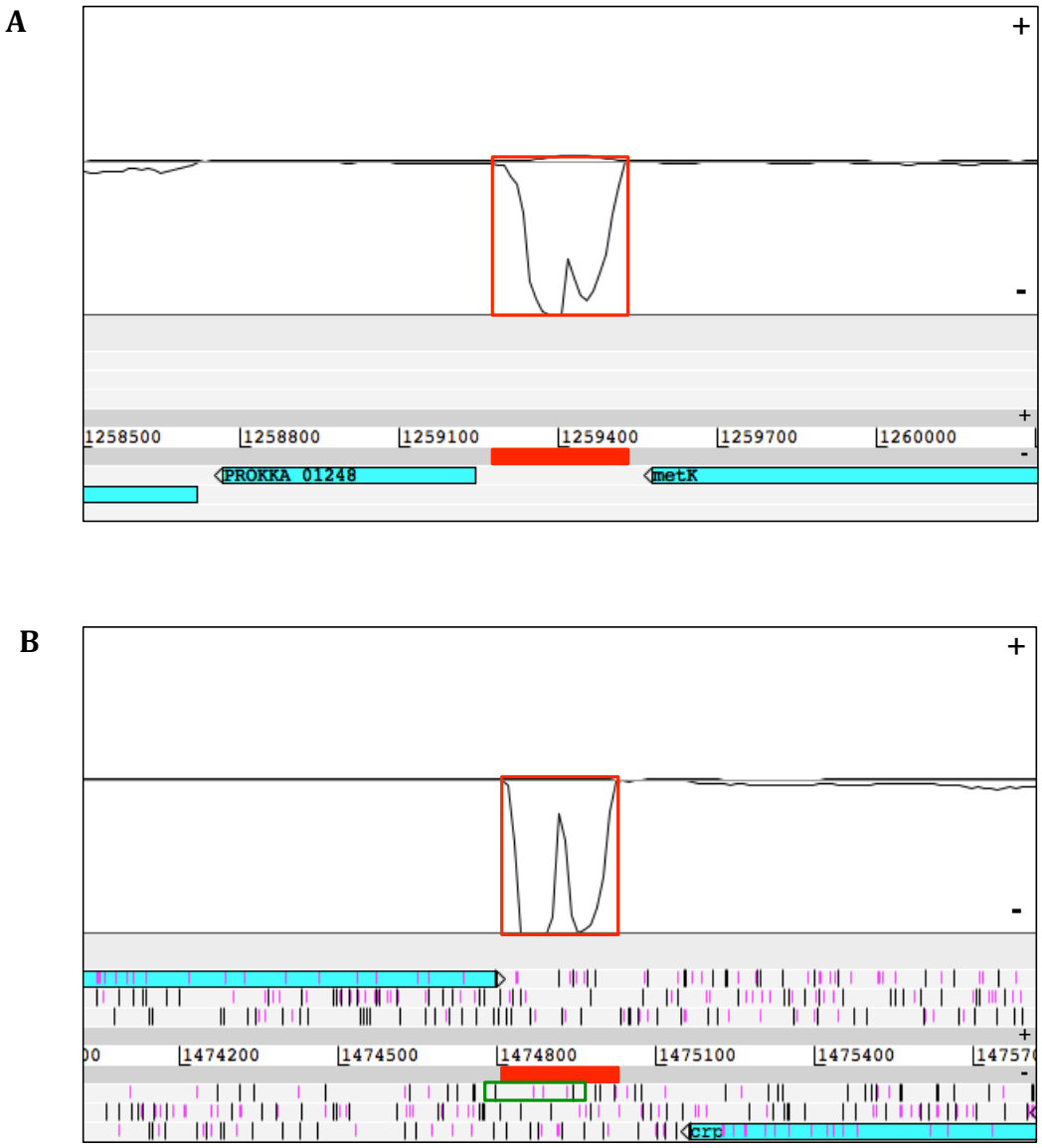

Figure 5.20: Two ncRNA candidates for validation with northern blotting, as depicted in the Artemis genome browser. Areas of mapped RNA-Seq reads, representing R2866_101 (A) and R2866_118 (B), are highlighted in red. An open reading frame, containing a start codon (blue vertical bar) and a stop codon (black vertical bar), covered one of the expression peaks of R2866_118.

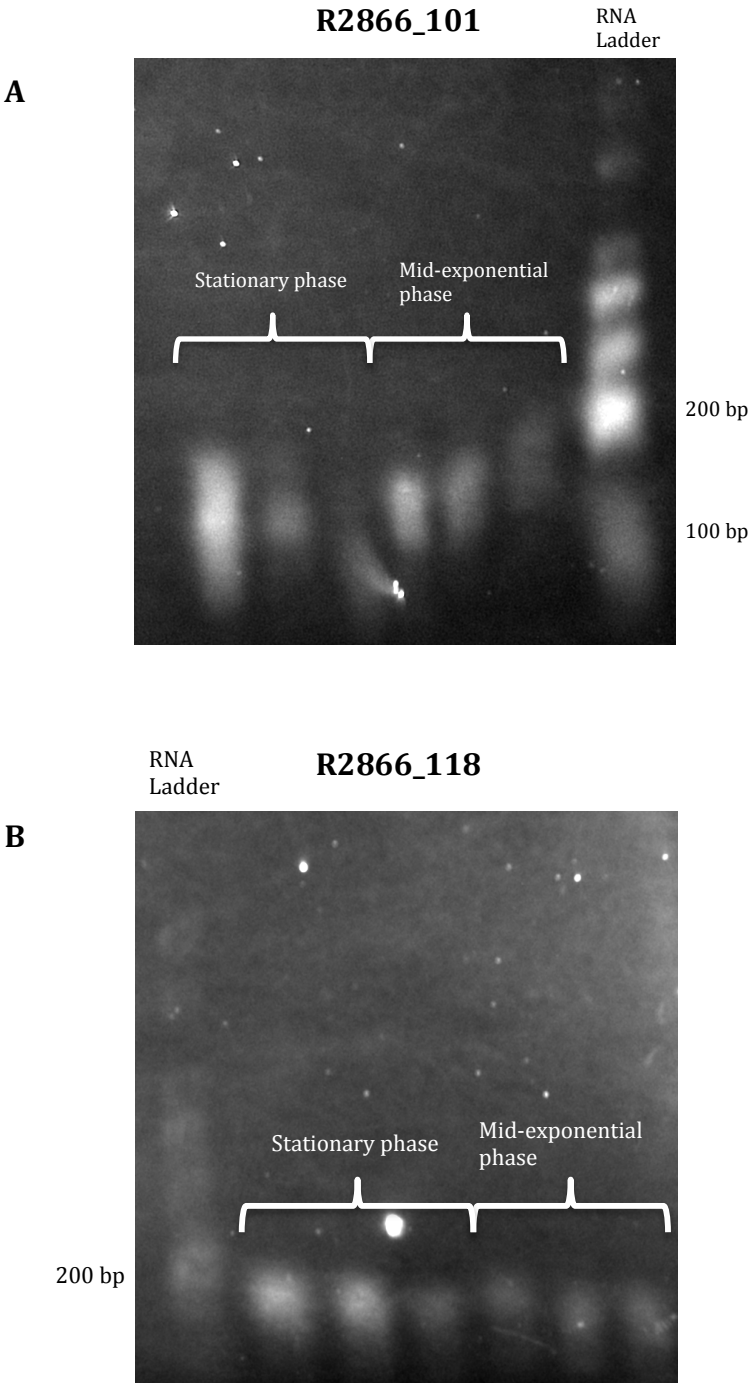

Figure 5.21: Northern blot validation of R2866_101 and R2866_118 ncRNAs. Data is shown for three biological replicates grown during mid-exponential and stationary phases for R2866_101 (A) and R2866_118 (B).

5.2.2.6.1 Secondary structure and target prediction of validated ncRNAs

Secondary RNA structures of two validated ncRNAs were predicted based on the minimum free energy for RNA folding (see Figure 5.22). Minimum free energy values were -82.20 kcal/mol for R2866_101 and -64.10 kcal/mol for R2866_118. Gene target prediction was subsequently carried out for the same two validated ncRNAs using the CopraRNA online tool (Wright et al., 2013, Wright et al., 2014). Predicted targets for R2866_101 were largely metabolic genes, with the top hit being a gene encoding a tRNA modification protein. Two tRNA-related genes were also among the top five putative targets for R2866_118. Interestingly, the top hit for R2866_118 was a bacteriophage gene, whilst another putative target gene, encoding a hypothetical protein, was part of the ICE, homologous to ICE*Hin*1056 (Juhas et al., 2007).

R2866_101 **R2866_118**

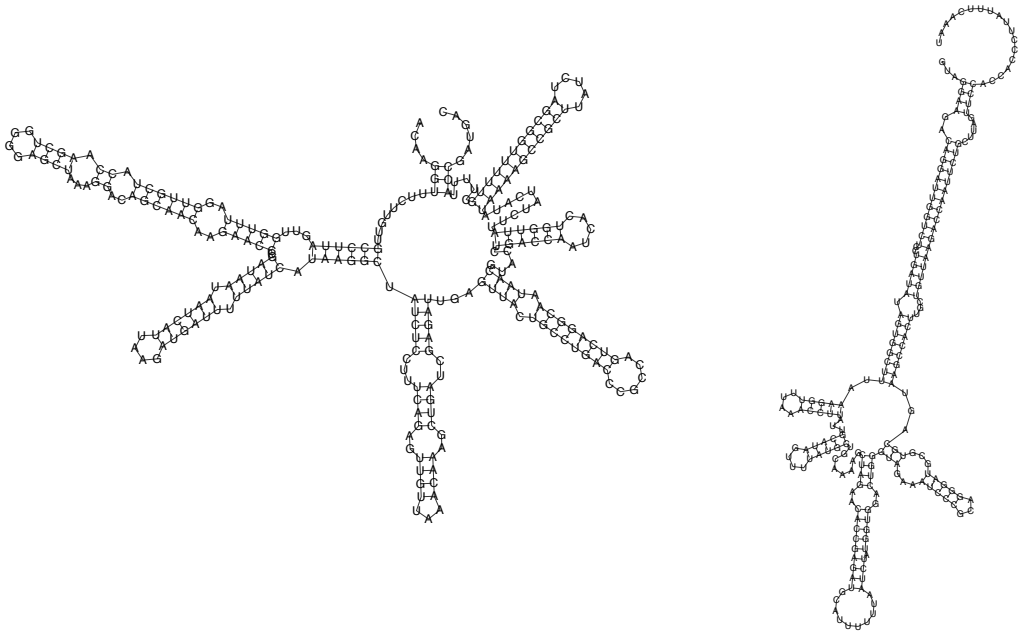

Figure 5.22: Predicted secondary RNA structures of validated ncRNAs R2866_101 and R2866_118.

Table 5.5: Top five predicted gene targets for the validated ncRNAs R2866_101 and R2866_118, based on the lowest p-values.

Target gene	Gene annotation	P-value
R2866_101		
gidA	tRNA uridine 5-carboxymethylaminomethyl modification enzyme	0.000278038
rnfD	Electron transport complex protein RnfD	0.001267742
hsdR	Type I restriction enzyme HindVIIP, R protein	0.002250713
secB	Protein export chaperone SecB	0.00311078
atpG	Membrane-bound ATP synthase, F1 sector, gamma-subunit	0.00326923
R2866_118		
-	Bacteriophage Lambda NinG protein	0.000332415
serC	Phosphoserine aminotransferase	0.000337151
ygfZ	tRNA-modifying protein	0.000757476
trmD	tRNA (guanine-N1)-methyltransferase	0.002190023
-	Hypothetical protein	0.003511625

5.3 Discussion

This study provides the most comprehensive repertoire of putative ncRNAs identified in *H. influenzae* to-date. In particular, this is the first time that antisense ncRNAs have been systematically investigated in this organism on a whole-genome scale. In addition, two novel intergenic ncRNAs, with potential roles as *trans*-acting sRNAs, were validated in an invasive strain R2866. The identification of ncRNAs from the RNA-Seq data in *H. influenzae* was made possible by the development of a robust new tool "toRNAdo". It can be used with any bacterial transcriptomic data to identify potentially important RNA elements, including UTRs, operon regions as well as intergenic and antisense ncRNAs.

The toRNAdo script, developed in this study, proved to be robust across different RNA-Seq conditions, with a very low number of false positive ncRNAs. Additional script optimisation could be carried out in the future to improve the accuracy even further. One of the biggest strengths of the toRNAdo script is that it is not limited to putative intergenic ncRNAs, but rather explores the whole complexity of a bacterial transcriptome, which also includes antisense and mixed transcripts. This is in contrast to a previous study on the putative ncRNAs in *H. influenzae*, which only focused on intergenic regions (Baddal et al., 2015). In addition, the script allows exploration of UTRs and operon regions. While identifying these RNA elements in *H. influenzae* was not the aim of this study, it will be important for future work. Finally, the script has been tested by other members of the laboratory on different bacterial species with similarly successful and reproducible outcomes. This further reaffirms the toRNAdo script as an appropriate new tool for exploring bacterial transcriptional landscape.

This study clearly demonstrated the importance of using different growth conditions when attempting to discover novel putative ncRNAs. There was a clear difference in the number of identified ncRNAs between different conditions. This supports the idea that bacterial transcriptional response is of a

very sensitive nature and is highly dependent on environmental factors. Therefore, a higher number of conditions used would likely result in a larger total number of identified ncRNAs for that organism. This approach has been previously utilised to explore the repertoire of putative sRNAs in other pathogenic bacteria like *S. enterica* (Kroger et al., 2013).

The majority of all putative ncRNAs were identified during stationary phase in both Rd and R2866 strains. Differential ncRNA expression analysis revealed most ncRNAs to be up-regulated during stationary phase as well, which was the most likely reason for their abundant discovery at that particular condition. While some sRNAs have a positive effect on gene expression, the majority of them act to inhibit translation (see section 1.2). Therefore, a subset of ncRNAs, identified in this study, could be acting in *cis* or *trans* to inhibit translation of specific mRNAs. In particular, as observed in Chapter 4, protein biosynthesis was reduced overall during stationary phase in both strains, supporting the increase in ncRNA expression and their potential role in translation inhibition. This is concurrent with a high number of intergenic and antisense ncRNA sequences being up-regulated during stationary phase in *E. coli* and *Pseudomonas aeruginosa* (Argaman et al., 2001, Gomez-Lozano et al., 2014).

A low number of ncRNAs identified during mid-exponential phase in both strains was most likely due to a high number of replicates, as this condition was used in all RNA-Seq experiments. There were a total of nine replicates for R2866 and twelve for Rd during mid-exponential phase, while all other conditions contained three replicates each. As one of the criteria for classifying ncRNAs was their appearance in all replicates, the variations in expression among a large number of replicates could therefore result in the loss of ncRNAs with low expression. However, this caveat was partially overcome using other RNA-Seq conditions in this study, enabling the detection of ncRNAs that were enriched in response to other environmental factors and growth conditions.

The only previously characterised sRNA in *H. influenzae* was the Fur-regulated HrrF (Santana et al., 2014). In this study, it did not form an expression peak in any of the RNA-Seq conditions and was therefore not identified by the toRNAdo script. In fact, it was not observed to form an expression peak during standard and iron-starvation conditions in the study by Santana et al. either. The induction of its expression was only observed in a *fur* deletion mutant. However, in the study by Baddal et al. HrrF was identified among putative intergenic ncRNAs expressed during infection of human cells (Baddal et al., 2015). This again highlights the specific nature of these RNA elements, some of which can only be identified under certain experimental conditions. Nevertheless, among 18 putative ncRNAs identified by Baddal et al. were homologues of RNase P and GcvB, both of which were present in the current dataset (Baddal et al., 2015).

The GcvB homologue was previously computationally predicted to be in the *H. influenzae* genome and this is the second time that its expression was observed in this organism (Pulvermacher et al., 2008, Baddal et al., 2015). It was up-regulated during stationary phase in both strains, which is supported by its up-regulation during this growth phase in *E. coli* (Argaman et al., 2001). However, it was particularly interesting that the GcvB homologue was up-regulated almost 5-fold during iron-starvation in the Rd strain. The lack of its up-regulation in R2866 further emphasises the difference in iron-starvation response between these strains, as explored in Chapter 4. While the role of GcvB in amino acid transport and metabolism is well documented, its significance during iron homeostasis has not been previously observed (Sharma et al., 2011). Further investigation into this ncRNA could lead to discovering potential new pathways that *H. influenzae* possesses for colonisation of iron-restricted environments.

Up-regulation of 6S RNA and RNase P homologues during stationary phase in both strains was not surprising. 6S RNA was previously shown to be abundant during stationary phase in *E. coli*, as compared to mid-exponential phase (Wassarman and Storz, 2000). It is involved in regulation of global gene

expression through specific interaction with a sigma-70 holoenzyme of RNA polymerase. RNase P plays a role in tRNA maturation, which is likely important during stationary growth phase where global changes in translation processes occur (Evans et al., 2006). It was striking that the absolute expression of these two ncRNAs was the highest among all identified ncRNAs across all RNA-Seq conditions in both strains, signifying the important functions that these RNA elements possess.

While some intergenic ncRNAs have been identified and explored in *H. influenzae* previously, there has been no systematic investigation of antisense transcripts or RNA elements that have both antisense and intergenic properties (Santana et al., 2014, Baddal et al., 2015). Therefore, this study provides the first robust identification of these RNA structures in *H. influenzae*. It was interesting that several of ncRNAs, up-regulated during stationary phase and iron-starvation, were antisense to iron-related genes as well as catalase and pilus genes. They could possibly act in *cis* to modulate the expression of these infection-relevant genes. Their further validation would provide important insight into transcriptional regulation of *H. influenzae* during infection. Additional validation and characterisation is also required for potential small proteins that were identified in the ncRNA dataset. In addition, while only intergenic ncRNAs were investigated for the presence of an open reading frame in this study, there is a possibility that some of the antisense ncRNAs encoded small proteins as well. This phenomenon has been previously observed in *Pseudomonas fluorescens*, where authors identified and characterised one such "hidden" antisense protein (Silby and Levy, 2008). Exploration of potential antisense proteins in *H. influenzae* will be part of future work.

The R2866_118 ncRNA contained two clear large expression peaks, one of which possessed an open reading frame with a start codon. However, northern blotting revealed the region covering both expression peaks to be transcribed as a single RNA molecule. It is possible that the region contains both a putative intergenic sRNA sequence and a protein-coding gene, as has been observed for

other characterised sRNAs, such as RNA III (Benito et al., 2000). Further experimental validation is required to better characterise this region in future work. It was curious that some of the top target genes for R2866_118 belonged to mobile genetic elements, such as prophage and ICE. Regulation of horizontally acquired genes was previously reported for an sRNA SgrS in *S. enterica* (Papenfort et al., 2012). Other putative target genes for both R2866_101 and R2866_118 were involved in tRNA modification, which is supported by up-regulation of these ncRNAs during stationary phase, where protein biosynthesis is reduced.

Overall, this study has significantly enlarged the known repertoire of putative ncRNAs in *H. influenzae*. Future work will involve generation of mutants of validated ncRNAs in *H. influenzae*, followed by investigation of their behaviour in different infection-relevant conditions. In addition, dRNA-Seq, a new method for identifying transcriptional start sites and primary transcripts, will be used to improve identification and annotation of novel RNA structures in *H. influenzae* (Sharma et al., 2010). Finally, the developed toRNAdo script will be used for discovery of ncRNAs in other bacteria and will help to drive important studies into these novel RNA elements.

Chapter 6: Conclusions and future work

Even after the introduction of the Hib vaccine, *H. influenzae* remains an important human pathogen, being one of the main aetiological agents of otitis media and other respiratory tract infections, such as pneumonia and bronchitis (King, 2012). Invasive disease due to NTHi, for which no effective vaccine yet exists, has also been on the rise in recent years (Langereis and de Jonge, 2015). Understanding the behaviour of *H. influenzae* during infection is an important step towards identifying novel vaccine targets, necessary for prevention of NTHi disease. Thus the primary objectives of the work presented in this thesis were to expand on the current knowledge of the response of *H. influenzae* to relevant stress conditions as well as to identify and characterise a repertoire of novel RNA elements with potentially important roles in infection.

The initial part of this study, described in Chapter 3, provided an important basis for subsequent work as well as having important implications in itself. The first sequenced whole genome of a free-living organism, the Rd strain of *H. influenzae*, was re-sequenced and re-annotated for the first time (Fleischmann et al., 1995). The discovery of multiple nucleotide-level variants between original and re-sequenced Rd genome sequences highlighted the risks in relying on old genome annotations. An important finding was that a large number of indels corrected frameshifts, which were present in pseudogenes in the original Rd genome and were likely a result of original sequencing errors. In addition, an up-to-date genome annotation of both Rd and an invasive NTHi strain, R2866, ensured that the subsequent study of differential gene expression included the latest characterised bacterial genome features.

Discovery of multiple SNPs between original and re-sequenced Rd genomes suggests sequencing errors, true genetic changes, or a combination of both. Accumulation of genotypic changes over time can lead to phenotypic differences between the same strains in different locations. Further studies should determine whether any of the SNPs in Rd actually caused changes in the phenotype, particularly the SNP present in the *mutS* gene. This work

emphasises the importance of re-sequencing reference genomes, which can be used to correct assumptions about bacterial genetic composition. Both the evolution of model bacterial strains as well as the use of old genome annotations need to be addressed in bacteriological research. This has implications for defining a model organism, as important differences may exist between the genotype of a model bacterial strain and its published genome sequence.

A detailed analysis of the accessory genome of Rd and R2866 highlighted the heterogeneity between these two strains. As expected, R2866 contained a larger number of virulence-associated genes than Rd, likely making the former a more successful pathogen. On the other hand, Rd possessed unique metabolic pathways, which may prove advantageous during growth in different environments. The genomic heterogeneity between Rd and R2866 has implications for the study of pathogenic behaviour in bacteria using model bacterial strains, which could potentially be less virulent than other strains of the same organism. In addition, this work contributed to the identification of differentially expressed genes that were present in only one of the strains, as described in Chapter 4.

One of the main objectives of this study was to explore the behaviour of *H. influenzae* during infection-relevant conditions. For that purpose, the transcriptional response of *H. influenzae* to stationary phase, nutritional stress, iron starvation and oxidative stress was characterised using the RNA-Seq technology. An intriguing finding was increased sensitivity of the invasive strain, R2866, to oxidative stress and iron starvation as compared to Rd, which is a standard laboratory strain. This was supported by differences in the transcriptional response of Rd and R2866 to these stress conditions. In particular, R2866 had a more profound Fur-mediated response to both stresses, suggesting a stronger requirement to maintain iron homeostasis in this strain.

This was the first time that the behaviour of *H. influenzae* during stationary phase and nutritional stress were examined on a whole-transcriptome level. Not

surprisingly, iron-starvation response and competence were induced in both conditions. In particular, the transcriptional profile of *H. influenzae* during stationary phase was found to be multifactorial, resembling oxidative, iron-starvation and nutritional stress responses. This highlights the advantage of utilising several infection-relevant conditions to enable identification of shared metabolic and stress response pathways, which are more likely to be exploited during natural infection.

An intriguing route for future studies would be a combination of oxidative and iron-starvation stress conditions. It was previously shown that iron chelation during oxidative stress restores the viability of *H. influenzae* (Juneau et al., 2015). This is most likely due to low intracellular iron levels leading to reduction in generated hydroxyl radicals through the Fenton reaction. Other relevant conditions, which could be explored using the RNA-Seq method, include the effect of sub-inhibitory concentrations of antibiotics as well as co-culture with other bacterial pathogens that share the same ecological niche, as discussed below.

S. pneumoniae and *Moraxella catarrhalis* have been identified alongside *H. influenzae* as the most important aetiological agents of otitis media; the significance of the multispecies biofilm, formed by these three organisms, for disease progression was previously reported (Hall-Stoodley et al., 2006). *S. pneumoniae* is a particularly important co-pathogen, as it produces hydrogen peroxide, which causes oxidative stress for *H. influenzae* and inhibits its growth (Pericone et al., 2000). While the transcriptional response during co-culture of *S. pneumoniae* and the Rd strain of *H. influenzae* has already been explored using the RNA-Seq approach, it would be interesting to study these interactions using other strains of *H. influenzae* (Tikhomirova et al., 2015b). In addition, the interactions between *H. influenzae* and another co-pathogen, *M. catarrhalis*, should also be investigated in future work.

The optimised invasion assay was used to establish that Rd and R2866 strains of *H. influenzae*, used in this laboratory, were able to infect human epithelial

cells. An attempt was also made to isolate RNA from the infected host and subsequently enrich bacterial RNA in the sample. This is important for future studies, where transcriptomic analysis of intracellular *H. influenzae* would help to elucidate potentially important pathways that occur during the invasion process. This would be a more targeted approach than the dual RNA-Seq study by Baddal et al., where adhesion and invasion processes were not separated (Baddal et al., 2015). This work could then be developed further, by using high-end sequencing instruments to characterise the transcriptional response of both the host and *H. influenzae* during invasion. Future studies would also utilise the optimised invasion model to characterise the transcriptional profile of *H. influenzae* during co-infection with a co-pathogen *S. pneumoniae*.

A detailed repertoire of putative intergenic and antisense ncRNAs was identified in *H. influenzae* in this study, as described in Chapter 5. In particular, this was the first time that antisense ncRNAs were systematically described in this organism on a whole-genome scale. This work significantly increased the current number of known ncRNAs in *H. influenzae*. Differentially expressed ncRNAs during infection-relevant conditions, with potential regulatory roles during natural infection, were also identified. In addition, two intergenic ncRNAs from the R2866 strain were validated with northern blotting, adding to a very low number of previously validated ncRNAs in *H. influenzae* (Santana et al., 2014). This work provides an important basis for future investigation and characterisation of these potentially important intergenic and antisense RNA elements. Further work is required to characterise if any of the identified ncRNAs are in fact functional sRNAs. Their potential role in the pathogenesis of *H. influenzae* should also be explored, which could be coupled with transcriptomic studies of the optimised invasion assay. The identification of virulence-associated ncRNAs could potentially facilitate the utilisation of ncRNAs for the vaccine development in the future.

The toRNAdo script, developed in this study, uses RNA-Seq data to identify multiple RNA elements in a bacterial genome, including UTRs, intergenic and antisense transcripts as well as regions belonging to operons. Investigating

UTRs in future work could potentially reveal other important regulatory RNA structures, such as riboswitches. In addition, it would be interesting to explore putative operons on a whole-genome scale and identify possible differences in operon structures across different conditions. Finally, the characterisation of transcriptional start sites on a whole-genome level using the dRNA-Seq method should also be carried out in further studies (Sharma et al., 2010). The toRNAdo script could be further modified to improve the discovery of ncRNAs from the dRNA-Seq data, based on the presence of primary transcripts containing transcriptional start sites.

Overall, this study highlighted the complex transcriptional response of *H. influenzae* to infection-relevant conditions and emphasised issues regarding old genome annotations and the evolution of bacterial strains. Importantly, a comprehensive repertoire of putative ncRNAs, with potentially important roles during natural infection, was identified and will be further characterised in the future. This study improves the current understanding of the pathogenesis of *H. influenzae* and moves the bacteriological field further towards identifying novel vaccine and antibiotic targets for this organism.

Appendix A: Recipes for media and solutions

All reagents and chemicals were purchased from Sigma, UK, unless stated otherwise.

MIV medium

MIV medium was prepared by adding 0.5 ml of solutions 22, 23, 24 and 40 each to 50 ml of solution 21.

Solution 21: 4 g of L-aspartic acid, 0.2 g of L-glutamic acid, 1 g of fumaric acid, 4.7 g of sodium chloride, 0.87 g of dipotassium phosphate, 0.67 g of potassium phosphate and 0.2 ml of Tween® 80 was added to 850 ml of distilled water.

Solution 22: 0.04 g of L-cystine and 0.1 g of L-tyrosine were dissolved in 10 ml of 1 M hydrochloric acid at 37 °C, volume adjusted to 100 ml with distilled water, followed by adding 0.06 g of L-citrulline, 0.2 g of L-phenylalanine, 0.3 g of L-serine and 0.2 g of L-alanine. The solution was then sterilised through a 0.45 µm filter.

Solution 23: 0.1 M solution of calcium chloride in distilled water.

Solution 24: 0.1 M solution of magnesium sulphate in distilled water.

Solution 40: 5% (w/v) vitamin-free casamino acids in distilled water.

Solutions 23, 24 and 40 were sterilised by autoclaving.

MIV medium was supplemented with 10 µg/ml haemin, 4 µg/ml NAD, 2 µg/ml thiamine, 2 µg/ml pantothenic acid and 20 µg/ml hypoxanthine.

TE buffer

For a 100 ml buffer, 1 ml of 1 M Tris and 0.2 ml of 0.5 M EDTA was mixed and volume adjusted to 100 ml with distilled water. The solution was sterilised by autoclaving.

MOPS buffer

10x buffer was prepared by mixing 200 mM MOPS powder, 50 mM sodium acetate, 10 mM EDTA and adjusting pH to 7.0.

RNA sample loading buffer

Buffer was prepared by mixing 6.5 ml of formamide, 1.2 ml of 34% formaldehyde, 2 ml of 10x MOPS buffer, 0.4 ml of 50% sucrose solution, 30 mg of Orange G dye. Buffer was stored at -20 °C.

Transfer buffer

Buffer was prepared by mixing 0.01 M of sodium hydroxide and 3 M sodium chloride.

Neutralisation solution

Solution was prepared by adjusting 100 mM of Tris solution to pH 7.4.

SSC solution

20x solution was prepared by adding 88.23 g of sodium acetate dihydrate and 175.32 g of sodium chloride to 850 ml of distilled water, adjusting pH to 7.0 and topping up to 1 l with distilled water. Solution was sterilised by autoclaving.

Buffer 1

10x buffer was prepared by adding 116.07 g of maleic acid, 87.66 g of sodium chloride and 72.00 g of sodium hydroxide to 850 ml of distilled water, adjusting pH to 7.5 and topping up to 1 l with distilled water.

Blocking solution

10% solution was prepared by adding blocking 20 g of Blocking reagent (Roche) to 200 ml of 1x Buffer 1.

Pre-hybridisation solution

Solution has the following composition: 0.1% sodium N-lauroylsarcosinate, 7% sodium dodecyl sulphate (SDS), 1% blocking solution, 50% formamide, topped up with 5x SSC. Solution was stored at -20 °C.

Wash solution I

Solution was prepared by mixing 100 ml of 20x SSC and 10 ml of 10% SDS, and topping up to 1 l with distilled water.

Wash solution II

Solution was prepared by mixing 10 ml of 20x SSC and 10 ml of 10% SDS, and topping up to 1 l with distilled water.

Wash buffer

Buffer was prepared by mixing 3 ml of Tween® 20 and 100 ml of 10x Buffer 1, and topping up to 1 l with distilled water.

Appendix B: Supplementary tables for Chapter 4

Table S1: Differentially expressed genes in the Rd strain during stationary phase.

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
-	Alkylhydroperoxidase AhpD family core domain protein	28.79	9.05E-201	*artP*	Arginine transporter ATP-binding protein	82.19	2.75E-24
speF	Ornithine decarboxylase	22.04	9.29E-128	*artI*	Arginine ABC transporter substrate-binding protein	40.84	1.11E-87
-	RarD protein	19.86	1.35E-114	*potD_2*	Spermidine/putrescine ABC transporter substrate-binding protein	26.39	2.70E-102
fucI	L-fucose isomerase	18.19	4.85E-126	*deaD*	ATP-dependent RNA helicase	23.54	1.38E-208
dps	DNA protection during starvation protein	13.55	2.13E-83	*artM*	Arginine transporter permease subunit ArtM	22.57	2.90E-96
potE	Putrescine transporter	11.39	2.64E-138	*cca*	Multifunctional tRNA nucleotidyl transferase/2'3'-cyclic phosphodiesterase/2'nucleotidase/phosphatase	18.02	1.00E-95
glpK	Glycerol kinase	10.59	5.22E-118	*rplL*	50S ribosomal protein L7/L12	17.69	2.59E-158
fucR	L-fucose operon activator	10.27	1.74E-98	*trpE*	Anthranilate synthase component I	16.5	7.40E-123
-	DNA polymerase V subunit UmuD	9.75	2.52E-48	*artQ*	Arginine transporter permease subunit ArtQ	16.36	8.59E-103
glpF_1	Glycerol uptake facilitator protein	9.54	2.77E-96	*nrfA*	Cytochrome c552	13.93	4.39E-12
gntP_2	Gluconate permease	9.4	2.92E-44	*rplI*	50S ribosomal protein L9	13.47	6.17E-250
rebM	Demethylrebeccamycin-D-glucose O-methyltransferase	9.01	1.18E-131	-	SH3 domain-containing protein	13.17	1.41E-92
-	Sulfatase-like protein	8.78	4.46E-43	*rpsR*	30S ribosomal protein S18	13.05	2.00E-191
fbp	Fructose-1,6-bisphosphatase	8.63	3.94E-40	*rpmC*	50S ribosomal protein L29	12.64	6.48E-141
yidK	Putative symporter YidK	8.59	6.78E-25	*rplP*	50S ribosomal protein L16	12.12	6.40E-146
msrA	Bifunctional methionine sulfoxide reductase subunits A/B	8.48	6.29E-34	-	Phosphate transport regulator	11.89	7.84E-47
afuA	Ferric ABC transporter protein	8.43	2.85E-06	-	Thiamine biosynthesis protein	11.68	5.86E-91
fucA	L-fuculose phosphate aldolase	8.01	8.28E-75	-	Primosomal replication protein N	11.48	9.09E-110
fucP	L-fucose permease	7.84	3.86E-71	*rplJ*	50S ribosomal protein L10	11.39	1.77E-95
glpC_1	sn-glycerol-3-phosphate dehydrogenase subunit C	7.79	3.77E-125	-	ABC transporter permease	11.31	1.27E-29
deoC	deoxyribose-phosphate aldolase	7.74	1.84E-156	*rplV*	50S ribosomal protein L22	11.06	1.91E-128
tehB	tellurite resistance protein TehB	7.05	2.68E-139	*rpsC*	30S ribosomal protein S3	10.91	4.57E-134
gor	glutathione reductase	6.89	9.57E-73	*rnb*	Exoribonuclease II	10.49	1.70E-43
iscR	HTH-type transcriptional regulator IscR	6.79	4.11E-94	*rpsS*	30S ribosomal protein S19	10.44	1.47E-114
-	Hypothetical protein	6.72	2.82E-94	*argG*	Argininosuccinate synthase	10.44	1.66E-85
moaD	Molybdopterin synthase small subunit	6.67	2.45E-55	*rpsQ*	30S ribosomal protein S17	10.4	1.30E-146
-	Manganese transport protein MntH	6.58	1.27E-54	*rpsI*	30S ribosomal protein S9	9.67	5.77E-190
nifS	Cysteine desulfurase	6.51	3.19E-69	*fis*	DNA-binding protein Fis	9.58	4.40E-73
fucU	Fucose operon protein	6.39	1.46E-33	-	Lipoprotein NlpI	9.26	4.72E-140
kipA	KipI antagonist	6.37	9.26E-38	*yecO*	S-adenosyl-L-methionine-dependent methyltransferase	9.22	8.70E-59

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
-	Zinc-type alcohol dehydrogenase	6.36	6.33E-05	*rpsF*	30S ribosomal protein S6	9.19	2.65E-105
fabA	3-hydroxydecanoyl-ACP dehydratase	6.28	2.90E-43	*rplB*	50S ribosomal protein L2	9.16	1.05E-109
xylG	Xylose transporter ATP-binding protein	6.26	1.72E-18	*trmD*	tRNA (guanine-N(1)-)-methyltransferase	9.15	3.28E-148
-	Peroxiredoxin hybrid Prx5	6.24	5.25E-139	*betT*	High affinity choline transport protein	8.91	1.58E-117
hisB	Imidazole glycerol-phosphate dehydratase/histidinol phosphatase	6.12	2.73E-77	*dmsA_3*	Anaerobic dimethyl sulfoxide reductase subunit A	8.8	1.33E-06
-	Pyridoxamine kinase	6.1	7.25E-75	*rplS*	50S ribosomal protein L19	8.79	1.09E-153
iscU	Scaffold protein	6.02	5.26E-121	-	Cobalt transport protein CbiM	8.65	3.35E-49
dipZ_2	Thiol:disulfide interchange protein	6.02	3.19E-15	-	Phosphate permease	8.6	3.91E-54
merT	mercuric ion transport protein	6.02	2.44E-23	*cydA*	Cytochrome D ubiquinol oxidase subunit I	8.33	1.23E-82
-	LamB/YcsF family protein	5.8	1.05E-38	*bioD_1*	Dithiobiotin synthetase	8.22	5.35E-09
moaC	Molybdenum cofactor biosynthesis protein MoaC	5.79	1.57E-65	*rplW*	50S ribosomal protein L23	8.2	2.00E-101
purL_2	Phosphoribosylformylglycinamidine synthase	5.7	8.25E-27	-	TPR repeat-containing protein precursor	8.14	1.12E-71
-	Hypothetical protein	5.61	5.03E-57	*rplD*	50S ribosomal protein L4	7.87	1.70E-106
-	Iron-sulfur cluster insertion protein ErpA	5.52	8.23E-53	*argR*	Arginine repressor	7.79	3.68E-17
fdx-1	Ferredoxin	5.47	3.92E-58	*rimM*	16S rRNA-processing protein RimM	7.77	2.29E-159
ssrA	Transfer-messenger RNA, SsrA	5.46	3.42E-48	*rplM*	50S ribosomal protein L13	7.46	2.07E-170
-	Allantoate amidohydrolase	5.44	2.99E-62	-	tRNA-Ala(tgc)	7.46	2.99E-35
xylA	Xylose isomerase	5.38	6.59E-14	-	tRNA-Ile(gat)	7.43	9.84E-39
moaE	Molybdopterin converting factor subunit 2	5.28	2.61E-40	*tbpA*	Thiamin ABC transporter substrate-binding protein	7.38	4.57E-129
yeaD	Putative glucose-6-phosphate 1-epimerase	5.25	1.42E-38	*rpsB*	30S ribosomal protein S2	7.34	3.31E-119
-	Hypothetical protein	5.22	3.97E-50	*rplC*	50S ribosomal protein L3	7.27	1.04E-115
yiaO_1	Extracytoplasmic solute receptor protein YiaO	5.19	6.87E-33	-	Hemoglobin-binding protein	7.18	3.09E-33
hisG	ATP phosphoribosyltransferase	5.18	3.38E-76	*ilvC*	Ketol-acid reductoisomerase	7.13	9.04E-54
-	Phosphate-starvation-inducible protein PsiE	5.15	5.37E-30	-	Epimerase family protein	7.02	1.19E-32
recN	DNA repair protein	5.14	6.78E-71	-	Mg2+/Co2+ transporter	6.96	1.76E-63
fucK	L-fuculokinase	5.07	8.07E-35	*tsf*	Elongation factor Ts	6.84	8.81E-114
-	Sigma factor regulatory protein	4.86	1.58E-38	*potD_1*	Spermidine/putrescine ABC transporter substrate-binding protein	6.7	1.80E-49
nhaC	Na+/H+ antiporter	4.77	2.17E-32	*rpsJ*	30S ribosomal protein S10	6.69	1.35E-115
glgC	Glucose-1-phosphate adenylyltransferase	4.74	1.12E-68	*rpsP*	30S ribosomal protein S16	6.61	6.90E-138
yadA	Adhesin YadA precursor	4.74	8.23E-10	*cydD_2*	ABC transporter ATP-binding protein	6.61	1.26E-31
-	Thioredoxin	4.72	1.37E-47	-	Hemoglobin-binding protein	6.61	2.73E-18
uspA	Universal stress protein A	4.61	0.004405658	*infB*	Translation initiation factor IF-2	6.49	6.38E-74
fnr_1	Anaerobic regulatory protein	4.6	0.000514492	*topB*	DNA topoisomerase III	6.08	5.46E-46
glgX	Glycogen operon protein	4.59	9.40E-55	-	Hypothetical protein	6	1.10E-20
gdhA	Glutamate dehydrogenase	4.57	3.54E-49	-	Transcriptional activator	5.99	1.78E-41
patB	Cystathionine beta-lyase PatB	4.56	0.00049419	*lrgA*	Antiholin-like protein LrgA	5.84	2.61E-06
-	FeS assembly protein IscX	4.54	1.42E-45	*rimP*	Ribosome maturation factor RimP	5.79	3.80E-50
glgB	Glycogen branching protein	4.52	1.44E-50	*nusA*	Transcription elongation factor NusA	5.71	1.07E-72
-	Integral membrane protein transporter	4.51	2.59E-34	-	Branched chain amino acid ABC transporter substrate-binding protein	5.62	2.82E-69
merP_2	Mercuric ion scavenger protein	4.47	1.05E-15	*argH*	Argininosuccinate lyase	5.61	7.38E-29
hisH_1	Imidazole glycerol phosphate synthase subunit HisH	4.43	5.64E-41	-	tRNA-Phe(gaa)	5.61	7.12E-18
hisC	Histidinol-phosphate	4.38	1.13E-63	-	Cobalt ABC transporter,	5.61	5.90E-23

	aminotransferase				permease protein CbiQ		
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
yiaM_2	2,3-diketo-L-gulonate TRAP transporter small permease protein YiaM	4.35	1.50E-14	*lolB*	Outer membrane lipoprotein LolB	5.57	2.15E-21
yiaO_2	Extracytoplasmic solute receptor protein YiaO	4.34	5.45E-36	-	ABC transporter ATP-binding protein	5.55	1.33E-31
kipI	Sporulation inhibitor KipI	4.31	6.13E-19	*rplO*	50S ribosomal protein L15	5.49	1.02E-62
clpB	ATP-dependent Clp protease ATPase subunit	4.31	6.62E-55	*secY*	Preprotein translocase subunit SecY	5.49	3.03E-80
mobB	Molybdopterin-guanine dinucleotide biosynthesis protein B	4.3	3.75E-36	*rplR*	50S ribosomal protein L18	5.47	1.73E-54
dapA	Dihydrodipicolinate synthase	4.26	1.93E-49	-	Lysozyme	5.44	2.74E-38
atoA	Acetate CoA-transferase subunit beta	4.2	2.01E-11	*cydB*	Cytochrome oxidase subunit II	5.42	3.64E-56
pepP	Aminopeptidase P	4.14	1.71E-40	*rpsE*	30S ribosomal protein S5	5.41	1.58E-53
rbsB	D-ribose transporter subunit RbsB	4.13	5.93E-42	-	tRNA-Asn(gtt)	5.39	4.62E-17
moaA	Molybdenum cofactor biosynthesis protein A	4.09	9.82E-34	*trpG_2*	Anthranilate synthase component II	5.37	6.22E-33
atoE	Short chain fatty acids transporter	4.05	1.58E-20	*yhhQ*	Inner membrane protein YhhQ	5.33	4.03E-14
trxM	Thioredoxin	4.03	1.20E-26	-	Hypothetical protein	5.32	4.95E-21
-	UDP-2,3-diacylglucosamine hydrolase	4.02	5.60E-45	*rplK*	50S ribosomal protein L11	5.31	5.21E-102
ribA	GTP cyclohydrolase II	3.94	3.30E-10	*recR*	Recombination protein RecR	5.15	1.31E-18
atoB	Acetyl-CoA acetyltransferase	3.94	2.39E-17	*rplA*	50S ribosomal protein L1	5.12	9.56E-95
radA	DNA repair protein RadA	3.93	1.18E-33	*rplF*	50S ribosomal protein L6	5.12	3.15E-76
hisD	Histidinol dehydrogenase	3.91	2.86E-50	-	Hypothetical protein	5.12	1.29E-48
-	Malic enzyme	3.83	0.003848486	*recG*	ATP-dependent DNA helicase RecG	4.99	5.10E-33
-	Molybdate-binding periplasmic protein	3.83	5.96E-26	-	tRNA-Glu(ttc)	4.96	1.02E-19
xylH	D-xylose ABC transporter permease	3.82	9.00E-14	*nrfB*	Cytochrome c nitrite reductase pentaheme subunit	4.95	7.89E-11
yiaM_1	2,3-diketo-L-gulonate TRAP transporter small permease protein YiaM	3.79	3.78E-13	*rpsH*	30S ribosomal protein S8	4.93	4.12E-72
pstA_2	Phosphate ABC transporter permease	3.77	2.43E-23	*rpmD*	50S ribosomal protein L30	4.91	1.01E-46
hktE	Catalase	3.75	2.69E-46	-	Short chain dehydrogenase/reductase	4.9	1.12E-40
-	Putative integral membrane protein	3.74	9.53E-38	*folK*	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	4.85	2.45E-29
-	MerR family transcriptional regulator	3.73	5.05E-18	-	Aldolase	4.74	1.32E-31
hisA	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	3.73	3.43E-34	*atpH*	F0F1 ATP synthase subunit delta	4.69	5.04E-95
pckA	Phosphoenolpyruvate carboxykinase	3.73	0.020677855	*cydD_1*	ABC transporter ATP-binding protein	4.69	1.34E-14
-	Outer membrane protein	3.71	4.95E-42	*rpmG*	50S ribosomal protein L33	4.61	3.59E-41
cysT	Sulfate transport system permease protein CysT	3.66	2.91E-10	*yidC*	Inner membrane protein translocase component YidC	4.6	5.96E-27
yqaA	Inner membrane protein YqaA	3.65	3.09E-46	*frr*	Ribosome recycling factor	4.57	3.13E-47
glpB	Anaerobic glycerol-3-phosphate dehydrogenase subunit B	3.64	1.62E-48	*mutT*	Mutator protein	4.56	1.33E-11
-	Transglutaminase-like superfamily protein	3.61	1.79E-46	*yidD*	Putative membrane protein insertion efficiency factor	4.54	1.73E-17
hscB	Co-chaperone HscB	3.59	2.99E-30	*rplQ*	50S ribosomal protein L17	4.52	1.51E-64
-	HTH-type transcriptional regulator	3.58	2.96E-19	*ccmA*	Cytochrome c biogenesis protein CcmA	4.47	8.30E-15
yhcB	Putative cytochrome d ubiquinol oxidase subunit 3	3.57	1.50E-30	-	Hypothetical protein	4.37	1.34E-51
-	Long chain fatty acid CoA ligase	3.55	3.52E-22	*cyaA*	Adenylate cyclase	4.36	3.00E-22
-	16S ribosomal RNA	3.55	1.75E-09	*rpsN*	30S ribosomal protein S14	4.36	3.27E-65
hcpC	Putative beta-lactamase HcpC precursor	3.54	3.62E-26	*psd*	Phosphatidylserine decarboxylase	4.32	1.35E-42
-	Di- and tricarboxylate transporter	3.54	0.001862815	*ttgI*	Toluene efflux pump outer membrane protein TtgI	4.3	4.24E-32

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
fhuA	Ferric hydroxamate uptake	3.53	2.24E-18	*ponB*	Penicillin-binding protein 1B	4.26	3.55E-33
-	RutC family protein	3.53	2.52E-63	-	Hypothetical protein	4.25	1.01E-17
-	2,3-diketo-L-gulonate reductase	3.51	6.55E-09	*rplX*	50S ribosomal protein L24	4.24	2.79E-76
-	Transcriptional regulator	3.45	8.56E-28	*comEA*	ComE operon protein 1	4.22	1.37E-13
lon	ATP-dependent proteinase	3.4	2.34E-46	*ygbM*	Putative hydroxypyruvate isomerase YgbM	4.22	1.92E-45
talB	Transaldolase B	3.36	5.15E-29	-	Putative protein-S-isoprenylcysteine methyltransferase	4.21	3.33E-35
psiE	Mig-7	3.35	4.27E-22	-	ABC transporter ATP-binding protein	4.13	1.16E-14
-	tRNA-seC(tca)	3.34	7.27E-09	*thiP*	Thiamine transporter membrane protein	4.1	1.38E-32
hisIE	Bifunctional phosphoribosyl-AMP cyclohydrolase/phosphoribosyl-ATP pyrophosphatase	3.33	5.86E-28	*rplE*	50S ribosomal protein L5	4.06	7.24E-70
raiA	Ribosome-associated inhibitor A	3.32	3.64E-15	-	Lipoprotein	4.02	6.19E-29
-	TonB-dependent Receptor Plug Domain protein	3.32	1.69E-16	*atpG*	F0F1 ATP synthase subunit gamma	3.99	2.76E-55
-	Integral membrane protein transporter	3.31	1.20E-34	*rpmB*	50S ribosomal protein L28	3.99	1.63E-31
-	Acetyl-CoA:acetoacetyl-CoA transferase subunit alpha	3.27	2.21E-06	*ipk*	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	3.97	5.45E-36
-	Hypothetical protein	3.27	4.31E-07	*murI*	Glutamate racemase	3.97	1.48E-24
-	Sulfite exporter TauE/SafE	3.25	4.29E-18	-	Hypothetical protein	3.91	8.28E-49
-	Glycerol-3-phosphate acyltransferase PlsY	3.24	2.84E-27	-	tRNA-Asn(gtt)	3.88	6.03E-14
hisF	Imidazole glycerol phosphate synthase subunit HisF	3.24	2.93E-25	*rpsT*	30S ribosomal protein S20	3.87	2.99E-28
rfaD	ADP-L-glycero-D-mannoheptose-6-epimerase	3.23	4.10E-18	-	ABC transporter ATP-binding protein	3.84	6.74E-20
-	N-acetylmannosamine-6-phosphate 2-epimerase	3.2	0.000825603	*atpF*	F0F1 ATP synthase subunit B	3.82	2.60E-63
rbsA	D-ribose transporter ATP binding protein	3.2	2.41E-06	*tuf_2*	Elongation factor Tu	3.82	1.35E-45
modD	Molybdenum transport protein ModD	3.2	3.81E-11	*cusC*	Cation efflux system protein CusC precursor	3.8	4.31E-22
cdsA	CDP-diglyceride synthetase	3.19	1.08E-21	*rho*	Hypothetical protein	3.79	3.39E-18
-	RNA polymerase sigma factor	3.17	1.41E-23	*pcnB*	PolyA polymerase	3.78	5.48E-20
-	TPR repeat-containing protein precursor	3.15	1.33E-07	*phnA*	Alkylphosphonate uptake protein	3.77	5.58E-22
glpA	sn-glycerol-3-phosphate dehydrogenase subunit A	3.15	0.001635678	*tuf_4*	Elongation factor Tu	3.77	5.04E-41
sucD	Succinyl-CoA synthetase subunit alpha	3.15	5.31E-21	*atpC*	F0F1 ATP synthase subunit epsilon	3.74	1.16E-38
-	Hypothetical protein	3.14	6.51E-27	-	5S ribosomal RNA	3.7	0.000799
-	Helix-turn-helix	3.13	4.95E-21	-	ABC transporter ATP-binding protein	3.66	3.04E-22
upp	Uracil phosphoribosyltransferase	3.12	1.17E-18	*tuf_3*	Elongation factor Tu	3.65	3.73E-45
queE	7-carboxy-7-deazaguanine synthase	3.1	2.85E-22	-	DksA-like zinc finger domain containing protein	3.63	2.32E-11
-	Glycine radical enzyme, YjjI family	3.1	3.35E-10	*fabH*	3-oxoacyl-ACP synthase	3.61	7.80E-15
-	Hypothetical protein	3.08	7.26E-28	*atpA*	F0F1 ATP synthase subunit alpha	3.6	1.27E-60
rbsR	RBS repressor	3.07	9.18E-23	*dcuB_2*	Anaerobic C4-dicarboxylate transporter	3.58	2.38E-15
-	Aminotransferase AlaT	3.05	2.60E-15	*secD*	Preprotein translocase subunit SecD	3.57	1.87E-51
-	23S ribosomal RNA	3.02	2.96E-18	-	tRNA-dihydrouridine synthase A	3.56	3.84E-39
-	SNARE associated Golgi protein	3.01	6.75E-35	-	Lipooligosaccharide biosynthesis protein	3.55	2.48E-10
emrB_1	Multidrug resistance protein B	3	9.37E-17	*glnE*	Bifunctional glutamine-synthetase adenylyltransferase/deadenyltransferase	3.54	3.58E-41
-	Hypothetical protein	2.98	0.00048165	*atpD*	F0F1 ATP synthase subunit beta	3.54	1.34E-48
eda	Keto-hydroxyglutarate-	2.94	5.16E-30	-	ABC transporter ATP-binding	3.54	3.75E-59

	aldolase/keto-deoxy-phosphogluconate aldolase				protein		
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
rec2	Recombination protein	2.94	8.70E-09	*tsaA*	Putative tRNA (adenine(37)-N6)-methyltransferase	3.53	1.32E-11
pstC	Phosphate ABC transporter permease	2.93	2.49E-14	-	Nucleoid-associated protein	3.51	4.79E-16
pflA_2	Pyruvate formate lyase-activating enzyme 1	2.92	7.45E-14	-	GTP-binding protein	3.51	2.14E-18
hscA	Chaperone protein HscA	2.92	7.48E-39	*miaA*	tRNA delta(2)-isopentenylpyrophosphate transferase	3.43	3.69E-26
-	DNA uptake protein	2.92	2.52E-14	*yecK*	Cytochrome C-like protein	3.39	0.007932
galK	Galactokinase	2.92	8.91E-15	*tuf_1*	Elongation factor Tu	3.36	3.09E-35
gcvA	DNA-binding transcriptional activator GcvA	2.92	1.37E-17	-	Sodium-dependent transporter	3.32	5.60E-26
-	tRNA-Lys(ctt)	2.92	5.86E-25	*infA*	Translation initiation factor IF-1	3.29	3.44E-10
-	tRNA-Arg(acg)	2.89	4.39E-22	*rplN*	50S ribosomal protein L14	3.28	1.31E-48
lyx	L-xylulose kinase	2.87	8.18E-21	*trpB*	Tryptophan synthase subunit beta	3.28	1.25E-32
-	Hypothetical protein	2.87	8.83E-23	-	Hypothetical protein	3.26	1.34E-06
-	Autonomous glycyl radical cofactor GrcA	2.85	1.28E-19	*cydC*	Cysteine/glutathione ABC transporter membrane protein/ATP-binding protein	3.22	6.53E-32
-	D-mannonate oxidoreductase	2.85	1.54E-30	*hslO*	Hsp33-like chaperonin	3.21	2.79E-30
uvrA	Excinuclease ABC subunit A	2.85	1.16E-24	*rpmF*	50S ribosomal protein L32	3.19	6.36E-34
galM	Aldose 1-epimerase	2.83	6.77E-16	*atpE*	F0F1 ATP synthase subunit C	3.19	5.75E-40
-	Hypothetical protein	2.83	4.16E-12	-	tRNA-Thr(ggt)	3.16	2.65E-37
oppA	Oligopeptide ABC transporter substrate-binding protein	2.82	6.15E-11	-	Hypothetical protein	3.12	7.25E-14
-	6-pyruvoyl tetrahydrobiopterin synthase	2.81	1.54E-11	*ycaO*	Ribosomal protein S12 methylthiotransferase accessory factor YcaO	3.11	6.55E-24
lexA	LexA repressor	2.8	1.95E-41	*fusA*	Elongation factor G	3.09	1.06E-43
sgbE	L-ribulose-5-phosphate 4-epimerase	2.8	4.52E-17	*secA*	Preprotein translocase subunit SecA	3.09	1.45E-29
murQ	N-acetylmuramic acid-6-phosphate etherase	2.79	1.57E-21	*pepT*	Peptidase T	3.09	0.013601253
atzC	N-isopropylammelide isopropyl amidohydrolase	2.79	2.08E-12	*znuA*	High-affinity zinc transporter substrate-binding protein	3.07	5.80E-19
-	Phage-associated protein, family	2.79	0.000105183	*rpoA*	DNA-directed RNA polymerase subunit alpha	3.04	4.45E-40
comD	Competence protein D	2.78	0.001498987	-	ABC transporter ATP-binding protein	3.04	1.41E-18
rbsK	Ribokinase	2.77	2.85E-30	*tig*	Trigger factor	3.03	2.18E-32
fnr_2	Fumarate/nitrate reduction transcriptional regulator	2.77	7.68E-12	*sapF*	Anti peptide resistance ABC transporter ATPase	3.03	9.52E-24
uraA	Uracil permease	2.76	4.11E-10	-	Dissimilatory sulfite reductase, desulfoviridin subunit gamma	3.01	7.37E-20
-	tRNA-Ser(gct)	2.74	1.41E-23	*coaA*	Pantothenate kinase	2.99	1.30E-20
-	Type I restriction-modification system, M subunit	2.74	1.56E-06	-	Hypothetical protein	2.99	1.95E-10
pgi	Glucose-6-phosphate isomerase	2.73	2.58E-27	*rpsD*	30S ribosomal protein S4	2.94	1.56E-43
pepA_1	Leucyl aminopeptidase	2.73	3.96E-16	*rpsA*	30S ribosomal protein S1	2.94	3.38E-45
-	tRNA-Lys(ttt)	2.73	3.84E-21	*dksA*	*dnaK* suppressor protein	2.93	3.56E-14
-	SprT-like family protein	2.72	2.54E-09	*por*	Oxidoreductase	2.92	2.19E-25
spxA	Regulatory protein spx	2.71	1.45E-30	*potC*	Spermidine/putrescine ABC transporter membrane protein	2.92	2.99E-35
-	Hypothetical protein	2.71	2.14E-17	*epmC*	Elongation factor P hydroxylase	2.91	2.14E-14
-	HTH-type transcriptional regulator	2.7	7.01E-14	-	Ribonuclease R winged-helix domain protein	2.91	5.50E-11
ndk	Nucleoside diphosphate kinase	2.7	2.71E-18	*tyrA*	Bifunctional chorismate mutase/prephenate dehydrogenase	2.89	1.83E-26
glgA	Glycogen synthase	2.7	1.47E-21	*prsA*	Ribose-phosphate pyrophosphokinase	2.89	5.16E-20
lrp	Leucine-responsive transcriptional regulator	2.7	1.05E-13	-	Branched-chain amino acid ABC transporter permease	2.85	2.89E-06
cyaY	Frataxin-like protein	2.68	2.71E-13	-	tRNA-Gly(tcc)	2.84	7.88E-32
-	Hypothetical protein	2.68	5.22E-24	-	Integrase/recombinase	2.82	3.30E-10

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
tpx	Thiol peroxidase	2.67	1.41E-16	*rpL25*	50S ribosomal protein L25	2.81	5.87E-29
sucC	Succinyl-CoA synthetase subunit beta	2.67	1.24E-15	*rnhB*	Ribonuclease HII	2.79	2.36E-14
-	Hypothetical protein	2.67	0.01119026	*dcuB_1*	Anaerobic C4-dicarboxylate transporter	2.78	8.67E-07
rbsD	D-ribose pyranase	2.66	0.044433338	-	Iron chelatin ABC transporter ATP-binding protein	2.78	9.91E-06
-	Hypothetical protein	2.66	2.20E-16	-	tRNA-Cys(gca)	2.78	1.09E-18
kdgK	2-dehydro-3-deoxygluconokinase	2.65	3.69E-19	-	Hypothetical protein	2.78	4.35E-10
frdD	Fumarate reductase subunit D	2.65	4.55E-20	*eno*	Phosphopyruvate hydratase	2.77	7.63E-09
-	Transcriptional regulator	2.65	2.01E-35	*tgt*	Queuine tRNA-ribosyltransferase	2.75	2.67E-17
-	tRNA-Arg(acg)	2.62	1.91E-22	-	Hydroxyethylthiazole kinase	2.75	2.48E-28
-	Transposase	2.62	2.45E-06	-	Nickel uptake substrate-specific transmembrane region	2.73	4.84E-15
rbsC	Ribose ABC transporter permease	2.6	1.76E-13	-	Hypothetical protein	2.72	1.71E-23
-	Plasmid RP4 TraN-like protein	2.6	5.86E-15	*pheS*	Phenylalanyl-tRNA synthetase subunit alpha	2.71	2.45E-21
-	Hypothetical protein	2.6	5.77E-16	-	Transporter	2.7	3.81E-13
menG	Ribonuclease activity regulator protein RraA	2.58	1.13E-13	*tbp2_1*	Transferrin-binding protein 2	2.7	1.30E-21
-	Xylulose kinase	2.58	1.43E-16	*pnp*	Polynucleotide phosphorylase/polyadenylase	2.69	8.76E-23
-	tRNA-Ser(gga)	2.58	9.72E-05	*rnfG*	Electron transport complex protein RnfG	2.69	1.84E-12
yabJ	Enamine/imine deaminase	2.57	2.53E-18	*mutL*	DNA mismatch repair protein	2.68	6.38E-21
serC	Phosphoserine aminotransferase	2.56	1.45E-14	*rpmE*	50S ribosomal protein L31	2.68	4.65E-21
hsdR_3	Type I restriction enzyme	2.56	2.98E-14	*arcA*	Two-component response regulator	2.68	0.006266088
-	Hypothetical protein	2.56	7.11E-05	*secF*	Preprotein translocase subunit SecF	2.66	2.64E-22
dod	Ribulose-phosphate 3-epimerase	2.54	1.41E-22	*pfs*	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	2.66	7.84E-15
yhxB_2	Tail fiber protein/phosphomannomutase	2.53	0.0182819	-	Mu-like prophage protein gp29	2.66	1.58E-13
purR	DNA-binding transcriptional repressor PurR	2.53	1.11E-14	*truB*	tRNA pseudouridine synthase B	2.65	5.08E-25
mazG	Nucleoside triphosphate pyrophosphohydrolase	2.52	4.95E-17	*thiE*	Thiamine-phosphate pyrophosphorylase	2.63	3.93E-18
-	Hypothetical protein	2.52	2.03E-11	*tyrS*	Tyrosyl-tRNA synthetase	2.63	4.25E-34
yccA	Modulator of FtsH protease YccA	2.51	1.47E-21	*brnQ*	Branched-chain amino acid ABC transporter	2.62	2.69E-08
trkH	Trk system potassium uptake protein TrkH	2.51	2.33E-16	-	Mu-like prophage FluMu G protein	2.62	2.19E-06
-	Hypothetical protein	2.51	1.07E-06	-	Hypothetical protein	2.62	0.010955
-	Sulfur transfer protein SirA	2.5	2.40E-17	*queF_2*	7-cyano-7-deazaguanine reductase	2.6	9.06E-07
-	Putative phage-encoded protein	2.5	3.95E-06	-	tRNA-Thr(tgt)	2.6	2.33E-24
oppB	Oligopeptide transporter permease	2.49	1.12E-18	-	Transglycosylase	2.59	8.41E-13
galT	Galactose-1-phosphate uridylyltransferase	2.48	2.58E-05	-	C32 tRNA thiolase	2.59	5.92E-10
emrB_2	Multidrug resistance protein B	2.48	7.70E-21	*thiI*	tRNA sulfurtransferase	2.58	5.99E-13
-	Translation initiation factor Sui1	2.47	1.64E-15	*tatB*	Sec-independent translocase	2.58	3.38E-12
anmK	Anhydro-N-acetylmuramic acid kinase	2.46	1.21E-29	-	tRNA-Leu(taa)	2.58	4.39E-12
-	Error-prone DNA polymerase	2.46	3.07E-12	-	Oligopeptide transporter, OPT family	2.57	1.08E-22
-	Hypothetical protein	2.45	0.010955231	*trpC*	Bifunctional indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase	2.55	5.75E-15
msbB	Lipid A biosynthesis (KDO)2-	2.44	1.62E-10	-	Membrane-fusion protein	2.54	1.13E-22

	(lauroyl)-lipid IVA acyltransferase						
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
nrdD	Anaerobic ribonucleoside triphosphate reductase	2.43	3.30E-16	*rpS7*	30S ribosomal protein S7	2.53	7.86E-33
nudF	ADP-ribose pyrophosphatase	2.43	0.010232	*rnpA*	Ribonuclease P	2.53	4.36E-09
comF	Competence protein F	2.43	1.74E-12	-	Hypothetical protein	2.53	6.90E-07
yigZ	IMPACT family member YigZ	2.41	4.59E-14	*hitC*	Iron(III) ABC transporter ATP-binding protein	2.52	2.76E-11
gph	Phosphoglycolate phosphatase	2.41	9.94E-20	-	Diaminobutyrate--2-oxoglutarate aminotransferase	2.52	3.30E-07
-	Holo-(acyl carrier protein) synthase 2	2.41	3.43E-16	*hrpa*	ATP-dependent RNA helicase HrpA	2.51	3.28E-25
ygiX	Transcriptional regulatory protein	2.4	2.54E-23	-	tRNA-Tyr(gta)	2.51	5.12E-22
dnaJ	Chaperone protein DnaJ	2.38	5.36E-09	*narP*	Nitrate/nitrite response regulator protein	2.49	3.75E-21
-	Hypothetical protein	2.38	0.000861431	*tatA*	Sec-independent protein secretion pathway component TatA	2.47	8.26E-09
-	SEC-C motif	2.38	1.69E-19	*kicB*	Condesin subunit F	2.47	7.90E-13
-	Hypothetical protein	2.38	0.018713902	*yohK*	Inner membrane protein YohK	2.46	0.020121282
recX	Recombination regulator RecX	2.37	7.45E-14	*thiD*	Phosphomethylpyrimidine kinase	2.46	1.68E-19
gcp_1	DNA-binding/iron metalloprotein/AP endonuclease	2.36	1.59E-20	-	Protein of unknown function, DUF	2.46	1.15E-13
groEL	Chaperonin GroEL	2.36	1.16E-11	-	Thiamin ABC transporter ATP-binding protein	2.44	3.36E-20
-	AraC family transcriptional regulator	2.36	2.14E-16	*potB*	Spermidine/putrescine ABC transporter membrane protein	2.44	4.15E-30
-	ABC transporter ATP-binding protein	2.36	0.004527839	-	EamA-like transporter family protein	2.44	8.09E-19
-	Hypothetical protein	2.36	3.11E-07	*rnt*	Ribonuclease T	2.43	5.28E-11
comE	Competence protein E	2.35	1.33E-11	-	tRNA-Val(gac)	2.43	0.000187
zwf	Glucose-6-phosphate 1-dehydrogenase	2.34	1.52E-18	*higB-1*	Toxin HigB-1	2.42	1.10E-10
hsdR_2	Type I restriction enzyme	2.34	9.52E-11	*mltA*	Murein transglycosylase A	2.42	1.17E-14
cdaR	Sugar diacid regulator	2.33	8.46E-09	*rep*	ATP-dependent DNA helicase	2.42	1.37E-16
pepD	Aminoacyl-histidine dipeptidase	2.33	7.13E-18	-	tRNA-Gly(gcc)	2.42	5.49E-12
-	Phage-related protein	2.33	8.04E-11	*rumA*	23S rRNA 5-methyluridine methyltransferase	2.41	7.94E-17
-	YecA family protein	2.33	2.81E-25	*ksgA*	Dimethyladenosine transferase	2.41	1.68E-09
-	N-acetylneuraminic acid mutarotase	2.32	1.89E-22	-	Hsf-like protein	2.41	1.28E-15
phoB	Phosphate regulon transcriptional regulatory protein PhoB	2.32	8.88E-18	*rpS11*	30S ribosomal protein S11	2.39	1.15E-26
-	Hypothetical protein	2.32	0.001297	-	Putative assembly protein	2.39	1.85E-10
rpoH	RNA polymerase factor sigma-32	2.31	1.04E-10	*fadL*	Long-chain fatty acid transport protein	2.38	3.57E-14
tbp1	Transferrin-binding protein 1	2.31	5.47E-16	*dnaX*	DNA polymerase III subunits gamma and tau	2.37	2.76E-17
glpC_2	sn-glycerol-3-phosphate dehydrogenase subunit C	2.3	4.06E-25	-	Hypothetical protein	2.37	2.20E-12
hsdM_1	Type I modification enzyme	2.3	3.28E-14	*mukB*	Cell division protein MukB	2.36	1.96E-17
malP	Maltodextrin phosphorylase	2.29	3.06E-24	*yjcD*	Putative permease YjcD	2.34	3.50E-09
vapD	Virulence-associated protein D	2.26	1.21E-12	*thrB*	Homoserine kinase	2.34	8.21E-17
-	Undecaprenyl-phosphate alpha-N-acetylglucosaminyltransferase	2.26	1.85E-15	*lpxB*	Lipid-A-disaccharide synthase	2.34	2.24E-12
groES	Co-chaperonin GroES	2.25	1.37E-14	-	Branched-chain amino acid ABC transporter permease	2.34	7.47E-09
recD	Exodeoxyribonuclease V subunit alpha	2.25	1.58E-13	-	Putative epimerase/dehydratase	2.33	1.08E-16
dld	D-lactate dehydrogenase	2.25	5.48E-14	-	Na(+)-translocating NADH-quinone reductase subunit C	2.32	1.30E-11
-	TPR repeat-containing	2.24	0.001480	*mtr*	Tryptophan-specific	2.32	8.00E-13

	protein precursor		445		transport protein		
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
afuB	Ferric transport system permease-like protein	2.24	3.39E-14	*glyS*	Glycyl-tRNA synthetase subunit beta	2.32	2.98E-22
asd	Aspartate-semialdehyde dehydrogenase	2.24	2.04E-11	*ftsL*	Cell division protein	2.32	1.24E-10
truA	tRNA pseudouridine synthase A	2.23	1.44E-13	*pykA*	Pyruvate kinase	2.31	4.84E-14
fbpC	Ferric transporter ATP-binding protein	2.22	1.64E-08	*pgpB*	Phosphatidylglycerophosphatase B	2.3	1.74E-12
-	Two component signal transduction system protein	2.22	0.016579113	*hindIIR*	Type II restriction endonuclease	2.3	2.55E-11
glpX	Fructose 1,6-bisphosphatase II	2.22	6.55E-09	*rbfA*	Ribosome-binding factor A	2.3	1.57E-12
-	RNA 2'-O-ribose methyltransferase	2.22	9.66E-16	*nhaA*	pH-dependent sodium/proton antiporter	2.29	2.73E-15
ygiY	Sensor protein QseC	2.21	1.49E-14	*polA*	DNA polymerase I	2.29	1.56E-26
-	Amino acid carrier protein	2.2	2.88E-13	-	16S ribosomal RNA methyltransferase RsmE	2.28	1.80E-13
-	Lipoprotein	2.2	1.29E-16	-	Transporter	2.27	9.92E-15
comB	Competence protein B	2.19	0.007688	*ureH*	Urease accessory protein	2.26	2.62E-15
arcC	Carbamate kinase	2.18	1.23E-19	*pta*	Phosphate acetyltransferase	2.26	1.41E-24
ybaK	Cys-tRNA(Pro)/Cys-tRNA(Cys) deacylase YbaK	2.18	1.06E-07	-	Opacity protein	2.26	1.57E-06
-	Hypothetical protein	2.18	3.45E-05	-	Uracil DNA glycosylase superfamily protein	2.26	1.64E-07
mesJ	Cell cycle protein	2.17	3.13E-15	-	Lipoprotein	2.25	1.51E-15
purN	Phosphoribosylglycinamide formyltransferase	2.17	3.38E-06	-	Protease	2.24	0.000129221
uxuA	Mannonate dehydratase	2.16	1.19E-12	*gpmA*	Phosphoglyceromutase	2.24	1.35E-06
-	Hemoglobin-binding protein	2.16	2.39E-13	-	tRNA-Met(cat)	2.24	6.98E-12
xylR	Xylose operon regulatory protein	2.15	2.48E-11	*pyrG*	CTP synthetase	2.22	2.39E-06
hsdR_1	Type I restriction enzyme EcoR124II R protein	2.15	3.29E-08	-	tRNA-Leu(tag)	2.22	6.48E-10
pflA_1	Pyruvate formate lyase-activating enzyme 1	2.14	9.82E-16	*dsbE_1*	Thiol-disulfide interchange protein	2.21	1.46E-06
hxuA	Haem-hemopexin utilization protein A	2.14	7.32E-11	*spoT*	Guanosine-3'5'-bis(diphosphate) 3'-pyrophosphohydrolase	2.21	1.68E-18
glpE	Thiosulfate sulfurtransferase GlpE	2.14	6.77E-08	-	Hypothetical protein	2.21	2.51E-13
-	RarD protein	2.13	2.71E-16	*dam*	DNA adenine methylase	2.2	4.12E-06
ilvE	Branched-chain amino acid aminotransferase	2.13	4.73E-10	*accC*	Acetyl-CoA carboxylase biotin carboxylase subunit	2.2	1.19E-11
-	Hydroxyacylglutathione hydrolase	2.13	1.03E-12	-	Electron transport complex RsxE subunit	2.2	1.49E-10
-	Glycerate dehydrogenase	2.13	5.99E-14	*fdnG*	Formate dehydrogenase, nitrate-inducible, major subunit precursor	2.19	3.05E-08
yfeB	Iron (chelated) transporter ATP-binding protein	2.12	7.17E-17	*kicA*	Condesin subunit E	2.19	1.92E-17
trxB	Thioredoxin reductase	2.12	2.30E-17	-	Mu-like prophage FluMu protein gp28	2.19	1.91E-08
-	N-acetylmannosamine kinase	2.11	4.20E-06	*fumC*	Fumarate hydratase	2.17	0.025787
fadD	Long-chain-fatty-acid--CoA ligase	2.11	3.73E-19	-	UDP-GlcNAc--lipooligosaccharide N-acetylglucosaminyl glycosyltransferase	2.17	1.14E-13
arcB	Ornithine carbamoyltransferase	2.11	7.92E-17	*nqrB*	Na(+)-translocating NADH-quinone reductase subunit B	2.16	1.89E-09
aroE_2	Shikimate 5-dehydrogenase	2.11	5.12E-14	-	Na(+)-translocating NADH-quinone reductase subunit D	2.15	6.64E-13
cysZ	Sulfate transport protein CysZ	2.11	1.54E-08	*yajC*	Preprotein translocase subunit YajC	2.15	1.47E-12
-	Putative esterase	2.11	1.87E-10	*aceE*	Pyruvate dehydrogenase subunit E1	2.15	3.07E-12
ygfZ	tRNA-modifying protein YgfZ	2.1	1.46E-12	-	tRNA-Asp(gtc)	2.15	8.96E-08
yjjV_2	Putative deoxyribonuclease YjjV	2.1	1.42E-14	*thrA*	Bifunctional aspartokinase I/homoserine dehydrogenase I	2.14	3.62E-17
aroC	Chorismate synthase	2.09	2.98E-11	-	Putative transcriptional regulatory protein	2.14	2.55E-11
purL_1	Phosphoribosylformylglycinamidine synthase	2.09	1.03E-14	-	PemK-like protein	2.14	2.35E-13
ulaD	3-keto-L-gulonate-6-phosphate decarboxylase	2.09	2.04E-11	*moeB_1*	Molybdopterin biosynthesis protein MoeB	2.13	6.64E-08

Gene	Product	Fold change	Adjusted p-value
Up-regulated genes			
-	Type II secretory pathway, component PulJ	2.09	0.00068264
rnd	Ribonuclease D	2.08	6.78E-07
birA	Biotin-protein ligase	2.07	5.13E-14
tfoX	DNA transformation protein	2.07	0.001263138
emrA	Multidrug resistance protein A	2.06	6.86E-13
dsbC	Thiol-disulfide interchange protein	2.06	9.09E-13
-	UDP-N-acetylmuramoylalanine-D-glutamate ligase-like protein	2.05	2.22E-21
icc	Cyclic 3',5'-adenosine monophosphate phosphodiesterase	2.05	7.04E-06
-	Phosphatase/phosphohexomutase	2.05	3.74E-09
iga1_1	Immunoglobin A1 protease	2.05	4.41E-10
acpD	Acyl carrier protein phosphodiesterase	2.04	2.08E-07
-	Hypothetical protein	2.04	3.26E-10
comC	Competence protein C	2.03	0.023544834
dppB	Dipeptide ABC transporter permease	2.03	1.12E-11
-	Hypothetical protein	2.03	2.18E-13
proC	Pyrroline-5-carboxylate reductase	2.02	2.91E-10
napF_2	Ferredoxin-type protein	2.02	0.000174167

Gene	Product	Fold change	Adjusted p-value
Down-regulated genes			
sapD	Peptide ABC transporter ATP-binding protein	2.12	3.44E-14
-	Isoleucyl-tRNA synthetase	2.11	7.57E-13
rnfD	Electron transport complex protein RnfD	2.1	7.31E-06
-	tRNA-Leu(taa)	2.09	5.99E-06
vacB	Virulence-associated protein	2.08	1.41E-16
folC	Folylpolyglutamate synthase/dihydrofolate synthase	2.08	2.65E-09
queF_1	7-cyano-7-deazaguanine reductase	2.08	3.36E-10
licC	Lic-1 operon protein	2.08	2.90E-12
ppc	Phosphoenolpyruvate carboxylase	2.08	0.000525693
-	Hypothetical protein	2.07	7.28E-13
era	GTP-binding protein Era	2.06	7.59E-10
acpP	Acyl carrier protein	2.06	1.01E-16
-	Sulfur transfer complex subunit TusD	2.06	1.76E-05
thrC	Threonine synthase	2.05	9.23E-14
-	Formate transporter	2.05	2.29E-05
napD	Nitrate reductase assembly protein NapD	2.05	3.58E-09
hindIIM	Modification methylase	2.05	9.80E-14
aceF	Dihydrolipoamide acetyltransferase	2.05	1.53E-10
gyrA	DNA gyrase subunit A	2.05	4.02E-18
-	Phage head morphogenesis protein, SPP1 gp7 family	2.05	8.63E-07
-	Hypothetical protein	2.05	1.82E-07
dacA	Penicillin-binding protein 5	2.04	3.14E-12
acrB	Acriflavine resistance protein	2.04	2.52E-12
nrfF	Cytochrome C-type biogenesis protein	2.04	2.92E-11
-	Sulfur transfer complex subunit TusB	2.04	0.001686842
mtgA	Monofunctional biosynthetic peptidoglycan transglycosylase	2.03	1.47E-05
modC	Molybdate transporter ATP-binding protein	2.02	2.63E-11
cstA	Carbon starvation protein A	2.01	1.57E-06
-	tRNA-Trp(cca)	2.01	0.04627
-	5S ribosomal RNA	2.01	0.00046

Table S2: Differentially expressed genes in the R2866 strain during stationary phase.

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
tnaA	Tryptophanase	123.37	0	*rpL29*	50S ribosomal protein L29	23.53	0
-	Alkylhydroperoxidase AhpD family core domain protein	56.49	0	*artP*	Arginine ABC transporter, ATP-binding protein ArtP	23.51	0
tnaB	Tryptophan permease	32.04	0	*rpL16*	50S ribosomal protein L16	22.41	0
afuA	Ferric transport system AfuABC; periplasmic-binding protein component	28.36	0	*rpS17*	30S ribosomal protein S17	20.09	0
hxuC	Haem-hemopexin utilization protein C	27.2	0	*rpS3*	30S ribosomal protein S3	18.93	0
hxuB	Haem-hemopexin utilization protein B	23.48	0	*rpL7*	50S ribosomal protein L7/L12	18.41	0
hxuA	Haem-hemopexin utilization protein A	20.61	0	*rpL22*	50S ribosomal protein L22	17.49	0
yjiG	Inner membrane protein YjiG	20.3	0	*artM*	Arginine ABC transporter, permease protein ArtM	16.75	0
-	Sporulation integral membrane protein YlbJ	17.38	0	*rpS19*	30S ribosomal protein S19	16.64	0
-	Putative peptidase	14.64	0	*potD*	Spermidine/putrescine ABC transporter, periplasmic-binding protein	16.2	0
ompU1	Putative outer membrane protein OmpU1	14.28	0	*rpL2*	50S ribosomal protein L2	15.46	0
-	Hypothetical protein	13.01	0	*rpL9*	50S ribosomal subunit protein L9	15.33	0
tbp2	Transferrin-binding protein 2	12.9	0	*cca*	tRNA nucleotidyltransferase/2'3'-cyclic phosphodiesterase/2'nucleotidase and phosphatase	14.88	0
-	Hypothetical protein	12.9	0	*artI*	Arginine ABC transporter, periplasmic-binding protein ArtI	13.75	0
dpsA	DPS ferritin-like protein	12.87	0	*phoU*	Putative phosphate regulator	13.56	0
-	Hypothetical protein	12.64	0	*rpS18*	30S ribosomal subunit protein S18	13.16	0
galT	Galactose-1-phosphate uridylyltransferase	12.63	0	*pitA*	Putative phosphate permease	12.43	0
tehB	Putative tellurite resistance protein B	12.2	0	*rpL23*	50S ribosomal protein L23	12.39	0
-	Putative zinc-type alcohol dehydrogenase	11.74	6.50E-153	*rpL10*	50S ribosomal protein L10	11.62	0
glpF	Aquaglyceroporin GlpF	11.6	0	*artQ*	Arginine ABC transporter, permease protein ArtQ	11.5	0
glpK	Glycerol kinase	11.21	0	*rpL4*	50S ribosomal protein L4	11.18	0
-	Hypothetical protein p54	11.13	2.84E-279	-	SH3 domain-containing protein	10.66	0
tbp1	Transferrin-binding protein 1	10.54	0	*priB*	Primosomal replication protein N	10.58	0
pckA	Phosphoenolpyruvate carboxykinase	9.58	6.27E-89	*rpL3*	50S ribosomal protein L3	10.22	0
-	Putative TonB-dependent transport protein	9.15	5.74E-167	*rps6*	30S ribosomal subunit protein S6	9.27	0
fucI	L-fucose isomerase	9.14	1.29E-178	*rpS10*	30S ribosomal protein S10	9.23	0
-	Hypothetical protein p55	8.92	4.89E-148	*rpL19*	50S ribosomal subunit protein L19	9.15	0
galK	Galactokinase	8.75	0	*rpL1*	50S ribosomal subunit protein L1	9.15	0
afuB	Ferric transport system AfuABC; permease component	8.41	0	*deaD*	ATP-dependent RNA helicase DeaD	8.57	0
pilA	Type II secretory pathway, major prepilin PilA	7.75	3.27E-86	*dat*	L-2,4-diaminobutyrate:2-ketoglutarate 4-aminotransferase aminotransferase	8.44	0
speF	Ornithine decarboxylase	7.69	0	-	Hypothetical protein	8.36	1.83E-21
yfiA	Ribosome binding protein Y	7.67	7.33E-87	*tsf*	Elongation factor Ts	7.97	0
pgdX	Peroxiredoxin/glutaredoxin	7.67	0	*topB_*	DNA topoisomerase III	7.95	9.67E-282

	glutathione-dependent peroxidase			*1*			
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
hitA	Iron(III) ABC transporter periplasmic-binding protein	7.42	0	*rpL11 1*	50S ribosomal subunit protein L11	7.76	0
glgB	1,4-alpha-glucan branching enzyme	6.95	0	*rpS2*	30S ribosomal protein S2	7.62	0
-	Putative permease	6.92	0	*ahpC*	Peroxiredoxin	7.57	0
gpFII	Putative bacteriophage tail tube protein	6.79	3.43E-194	*ytfL_2*	Putative hemolysin	7.57	1.49E-228
uspA	Universal stress protein A	6.68	1.38E-65	*fis*	DNA architectural protein Fis	7.29	0
galM	Galactose-1-epimerase (mutarotase)	6.62	0	*thiE*	Thiamin-phosphate pyrophosphorylase	7.27	4.13E-159
afuC	Ferric transport system AfuABC; ATP-binding component	6.55	0	*yidD*	Putative membrane protein insertion efficiency factor	6.85	1.61E-237
-	Hypothetical protein	6.55	6.64E-238	*rpS9*	30S ribosomal protein S9	6.61	0
-	Hypothetical protein	6.37	5.28E-227	*argG*	Argininosuccinate synthetase	6.46	0
-	tRNA-Leu(caa)	6.3	1.98E-33	*thiM*	Hydroxyethylthiazole kinase	6.42	2.08E-198
-	Hypothetical protein	6.18	0	*trmD*	tRNA (guanine-N1)-methyltransferase	6.35	0
-	Hypothetical protein	6.15	0	*rho*	Hypothetical protein	6.28	0
gpFI	Putative bacteriophage tail sheath protein	6.06	4.42E-238	*yhjE*	Putative permease	6.28	4.60E-242
-	Hypothetical protein	5.98	2.42E-99	*potD 2*	Spermidine/putrescine ABC transporter, periplasmic-binding protein	6.25	0
pulG	Type II secretory pathway, pseudopilin	5.93	3.97E-20	*thiD*	Phosphomethylpyrimidine kinase	6.15	4.74E-228
-	Putative antirestriction protein	5.86	7.20E-31	*tufA*	Elongation factor Tu (EF-Tu)	5.96	0
-	Putative bacteriophage lysozyme	5.8	2.87E-25	-	Putative ABC transporter, fused permease and ATP-binding components	5.86	1.04E-268
comA	Competence operon protein A	5.77	2.96E-35	-	Hypothetical protein	5.81	4.21E-168
ssrA	Transfer-messenger RNA, SsrA	5.76	0	*Hgd*	2-(hydroxymethyl)glutarate dehydrogenase	5.71	1.98E-136
ompE	Adhesin protein E (PE)	5.75	9.27E-273	*fdx-2*	Putative 4Fe-4S ferredoxin-type protein	5.69	3.11E-52
-	Hypothetical protein	5.74	3.82E-16	*rpL13*	50S ribosomal protein L13	5.67	0
glgC	Glucose-1-phosphate adenylyltransferase	5.7	5.78E-290	*rpS5*	30S ribosomal protein S5	5.62	0
-	16S ribosomal RNA	5.7	2.51E-05	*rnb*	Ribonuclease II	5.62	0
-	Hypothetical protein p56_2	5.68	8.81E-108	-	tRNA-Ile(gat)	5.6	2.36E-298
gpE	Putative bacteriophage tail protein E	5.68	1.33E-74	*rimM*	16S rRNA processing protein RimM	5.56	0
gpE+ E'	Putative bacteriophage tail protein E+E'	5.68	5.25E-42	*ygbL*	Putative sugar aldolase/epimerase	5.45	2.75E-91
gpO	Putative bacteriophage capsid scaffolding protein	5.66	2.32E-146	*argH*	Argininosuccinate lyase	5.44	2.41E-285
-	Putative TRAP-type transport system, small permease component	5.45	8.18E-57	*fabH*	Beta-ketoacyl-ACP synthase III	5.38	0
glgX	Glycogen debranching enzyme	5.41	9.41E-294	*tufB*	Elongation factor Tu (EF-Tu)	5.38	0
rbsD	D-ribose pyranase	5.35	4.15E-51	*psd*	Phosphatidylserine decarboxylase	5.37	9.63E-277
moaD	Molybdopterin synthase, small subunit	5.34	4.76E-247	*ispF*	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	5.36	1.83E-207
-	Putative TRAP-type transport system, periplasmic component	5.31	6.87E-95	*secY*	Protein translocase subunit SecY	5.35	0
comM_1	Competence protein ComM	5.28	4.80E-19	*recJ*	Single-stranded-DNA-specific exonuclease RecJ	5.33	6.98E-229
dprA	DNA processing chain A	5.26	6.26E-30	*rpL31*	50S ribosomal subunit protein L31	5.3	0
aspT	Aspartate/alanine antiporter	5.25	1.33E-38	*cmoA*	tRNA cmo(5)U34 methyltransferase, SAM-dependent	5.29	0
groES	GroES, chaperone Hsp10	5.23	3.66E-179	*bipA*	Ribosome binding GTPase BipA	5.26	0
malQ	4-alpha-glucanotransferase (amylomaltase)	5.23	4.48E-41	*rpL6*	50S ribosomal protein L6	5.23	0
fucR	L-fucose operon regulator	5.17	1.42E-194	*rpL18*	50S ribosomal protein L18	5.23	0

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
fucA	L-fuculose phosphate aldolase	5.12	3.39E-54	*mdaB*	NADPH quinone reductase	5.03	4.11E-33
fbp	Fructose-1,6-bisphosphatase	5.11	4.98E-276	*rplQ*	50S ribosomal protein L17	5.01	0
moaE	Molybdopterin synthase, large subunit	5.11	1.38E-271	-	Putative membrane protein	4.93	2.46E-82
cspD	Cold shock protein CspD	5.08	5.17E-200	*rpS8*	30S ribosomal protein S8	4.9	0
msrAB	Peptide methionine sulfoxide reductase	5.06	1.82E-59	*acrA*	Multidrug efflux system protein AcrA	4.88	0
clpB	ATP-dependent Clp protease ATPase subunit	5.02	2.12E-57	*yidC*	Inner membrane translocation protein YidC	4.87	0
trxM	Thioredoxin	5.01	0	*rpL15*	50S ribosomal protein L15	4.78	0
aspA	Aspartate ammonia-lyase (aspartase)	4.98	3.22E-37	*recR*	Recombination protein RecR	4.74	7.14E-213
gpN	Putative bacteriophage major capsid protein	4.98	2.19E-176	*infB*	Translation initiation factor 2	4.7	0
cbiK	Nickel and cobalt ABC transporter, periplasmic binding protein	4.96	2.09E-91	*rimP*	30S ribosomal maturation protein RimP	4.69	3.46E-164
-	Hypothetical protein HP2p14	4.93	1.89E-291	*ispD*	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	4.65	2.05E-199
rbsB	D-ribose ABC transporter, periplasmic-binding protein	4.9	6.93E-171	*acrB*	Multidrug efflux system protein AcrB	4.62	0
-	Putative TPR repeat protein	4.89	7.09E-34	-	tRNA-Trp(cca)	4.61	8.86E-10
-	DksA-like zinc finger domain containing protein	4.88	1.01E-12	-	tRNA-Asn(gtt)	4.57	1.27E-63
-	Hypothetical protein	4.87	1.97E-28	*folK*	7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase	4.56	9.32E-86
-	5S ribosomal RNA	4.83	0.032303	*pcnB*	Poly(A) polymerase I	4.48	4.63E-220
gpR	Putative bacteriophage tail completion protein	4.8	1.18E-12	-	tRNA-Ala(tgc)	4.47	1.16E-223
gpU	Putative bacteriophage protein U	4.78	1.78E-24	*rpS16*	30S ribosomal subunit protein S16	4.45	0
-	5S ribosomal RNA	4.69	0.035681099	*infA*	Translation initiation factor 1 (IF-1)	4.44	4.10E-187
-	Putative methyltransferase	4.67	7.75E-190	*nlpI*	Lipoprotein NlpI	4.34	8.43E-219
gpP	Putative bacteriophage terminase, ATPase subunit	4.66	5.67E-42	*dxr*	1-deoxy-D-xylulose 5-phosphate reductoisomerase	4.34	9.75E-248
yeaD	Putative glucose-6-phosphate 1-epimerase	4.62	4.83E-199	*rpL30*	50S ribosomal protein L30	4.33	0
nanM	Putative N-acetylneuraminate epimerase	4.6	1.52E-239	*rsmE*	16S rRNA methyltransferase	4.31	1.11E-134
oppA	Oligopeptide ABC transporter, periplasmic-binding protein OppA	4.6	0	-	Putative membrane transporter	4.29	3.60E-132
galR	Galactose operon regulator	4.5	1.07E-85	*nusA*	Transcription elongation factor NusA	4.28	0
pulJ	Type II secretory pathway, pseudopilin	4.49	1.41E-20	*mutT*	NTP pyrophosphohydrolase (MutT) (7,8-dihydro-8-oxoguanine-triphosphatase)	4.23	1.10E-34
gpW	Putative phage baseplate assembly protein W	4.48	1.26E-24	*rpS14*	30S ribosomal protein S14	4.2	0
-	Hypothetical protein	4.44	5.88E-06	*yegQ*	Putative protease	4.16	3.17E-256
ytfE	Iron-sulfur repair protein YtfE	4.42	1.47E-118	*hslO*	Hsp33-like chaperonin	4.13	3.63E-120
comM_2	Competence protein ComM	4.41	2.02E-54	-	Putative ABC transporter permease protein	4.1	9.09E-41
hemR	Putative TonB-dependent haem receptor	4.4	4.35E-251	*atpC*	Membrane-bound ATP synthase, F1 sector, epsilon-subunit	4.08	1.26E-211
-	LamB/YcsF family protein	4.36	2.51E-90	*betT*	Osmoprotection-related protein BetT	4.07	9.62E-160
groEL	GroEL, chaperone Hsp60	4.34	1.39E-53	*hrpA*	Putative ATP-dependent RNA helicase	4.03	0
xylB_2	Xylulose kinase	4.31	4.59E-32	*atpG*	Membrane-bound ATP synthase, F1 sector, gamma-subunit	4.02	0
-	Putative bacteriophage holin protein	4.31	5.42E-11	*ilvA*	Threonine deaminase	4.02	2.49E-180
-	Hypothetical protein	4.29	1.23E-94	*rrf*	Ribosome releasing factor	4.02	3.93E-273
-	Hypothetical protein	4.22	6.29E-282	-	tRNA-Phe(gaa)	4.01	1.39E-156
comB	Competence operon protein B	4.19	2.50E-12	*fusA*	Elongation factor G (EF-G)	3.99	0
potE	Putrescine-ornithine	4.16	0	-	Putative ABC transporter	3.97	6.77E-54

	antiporter				periplasmic binding protein		
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
gpM	Putative bacteriophage terminase subunit	4.07	1.04E-101	-	tRNA-Asn(gtt)	3.97	3.64E-144
ccmG_1	Haem lyase/disulfide oxidoreductase (DsbE)	4.04	4.10E-20	-	tRNA-Glu(ttc)	3.94	2.04E-156
gpI	Putative bacteriophage tail protein gpI	4.03	2.64E-18	*atpH*	Membrane-bound ATP synthase, F1 sector, delta-subunit	3.93	2.47E-294
-	Hypothetical protein	4.02	4.69E-07	*nrdA*	Ribonucleoside-diphosphate reductase 1, alpha subunit	3.91	2.26E-231
sdaC	Putative serine transporter	4.02	2.78E-160	*tusB*	Putative tRNA 2-thiouridine synthesizing protein B	3.9	7.30E-32
nhaC	Putative Na+/H+ antiporter	4.01	1.08E-117	*glnE*	Glutamate-ammonia-ligase adenylyltransferase (ATase)	3.89	2.93E-180
-	Hypothetical protein	4	4.24E-81	-	tRNA-Thr(ggt)	3.89	2.08E-300
dsbD_1	Thiol-disulfide interchange protein DsbD	3.98	5.61E-21	*rnt*	Ribonuclease T	3.88	2.83E-161
gpD	Putative bacteriophage protein D	3.97	1.63E-36	*fmt*	Methionyl-tRNA formyltransferase	3.86	5.56E-261
pflA_1	Pyruvate formate-lyase activating enzyme	3.94	4.28E-189	-	Putative chromosome partitioning related protein	3.85	4.75E-42
deoC	Deoxyribose-phosphate aldolase	3.94	1.73E-282	*rpS20*	30S ribosomal protein S20	3.82	0
fucP	L-fucose permease	3.88	8.35E-69	*tenA*	Thiaminase-2	3.8	6.54E-52
bphH	Putative glutathione-S-transferase	3.84	3.43E-264	*rnpA*	Ribonuclease P, protein component	3.79	1.57E-199
rbsC	D-ribose ABC transporter, permease protein	3.83	4.83E-64	*ponB*	Penicillin-binding protein 1B	3.74	2.89E-110
moaC	Molybdenum cofactor biosynthesis protein C	3.83	1.15E-213	*cya*	Adenylate cyclase	3.73	7.78E-273
mobB	Molybdopterin-guanine dinucleotide biosynthesis protein B	3.82	5.51E-287	-	Hypothetical protein	3.73	8.91E-116
-	DNA binding domain, excisionase family	3.79	3.19E-08	*dam_1*	DNA adenine methylase	3.72	3.86E-76
-	Hypothetical protein	3.77	1.17E-05	*trpX*	tRNA delta(2)-isopentenylpyrophosphate transferase (IPTase)	3.71	2.05E-138
gpS	Putative bacteriophage tail completion protein	3.76	3.29E-07	*rplE*	50S ribosomal protein L5	3.68	0
rbsA	D-ribose ABC transporter, ATP-binding protein	3.72	2.07E-59	*atpD*	Membrane-bound ATP synthase, F1 sector, beta-subunit	3.66	0
-	Hypothetical protein p13	3.69	3.60E-155	*yjjP*	Inner membrane protein YjjP	3.65	7.82E-18
gpL	Putative bacteriophage capsid completion protein	3.69	1.01E-21	*vcaM*	Putative ABC transporter, multidrug efflux pump	3.64	7.08E-234
-	Putative 5'(3')-deoxyribonucleotidase	3.67	3.23E-35	*nhaA*	Na+/H+ antiporter 1	3.63	8.19E-186
-	Putative bacteriophage tail collar protein	3.66	1.33E-13	*tgt*	tRNA-guanine transglycosylase	3.62	3.77E-166
gpJ	Phage baseplate-assembly protein J	3.65	2.59E-30	*atpF*	Membrane-bound ATP synthase, F0 sector, subunit B	3.6	3.33E-198
hktE	Catalase	3.63	7.16E-87	-	Hypothetical protein HP2p11	3.58	2.38E-13
kipA	KipI antagonist	3.63	5.14E-81	*ygbM*	Putative hydroxypyruvate isomerase YgbM	3.57	6.02E-67
-	Putative phage tail-fiber protein	3.61	2.30E-06	*lpxB*	Lipid-A-disaccharide synthetase	3.55	8.00E-211
-	23S ribosomal RNA	3.6	1.17E-06	*ftsB*	Putative cell division protein ftsB	3.54	2.14E-79
-	Hypothetical protein	3.59	1.92E-13	*mopI*	Molybdate-binding protein	3.54	6.78E-49
-	Bacteriophage Lambda NinG protein	3.58	6.59E-07	*lolB*	Outer membrane lipoprotein LolB	3.51	3.80E-149
pilB	Type II secretory pathway, ATPase component PilB	3.56	3.26E-41	-	Putative ABC transport system, ATP-binding protein	3.5	1.82E-283
comD	Competence operon protein D	3.54	5.15E-11	*rpL33*	50S ribosomal protein L33	3.48	3.14E-293
merT	Putative heavy metal transport protein	3.48	6.35E-42	*rpL24*	50S ribosomal protein L24	3.47	0
patB	putative PLP-dependent aminotransferase	3.46	1.58E-21	*rpS1*	30S ribosomal protein S1	3.46	0
-	Putative short-chain alcohol dehydrogenase	3.42	1.09E-146	*atpA*	Membrane-bound ATP synthase, F1 sector, alpha-subunit	3.45	9.39E-283
kdgA	2-keto-3-deoxygluconate 6-phosphate aldolase and 2-	3.41	2.44E-125	*pgpB*	Putative phosphatidylglycerophospha	3.44	3.47E-77

	keto-4-hydroxyglutarate aldolase				tase B		
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
comE	Outer membrane secretin ComE	3.41	1.51E-46	*accC*	Acetyl-CoA carboxylase, biotin carboxylase subunit	3.44	0
-	Putative NAD-dependent protein deacetylase	3.41	5.41E-49	-	tRNA-Gly(tcc)	3.43	0
pflB_2	Formate acetyltransferase 1	3.4	7.22E-205	*tusE*	tRNA 2-thiouridine synthesizing protein E	3.42	4.64E-64
htpG	Chaperone Hsp90	3.37	5.51E-108	-	Hypothetical protein	3.41	9.02E-13
pdxH	Putative pyridoxamine 5'-phosphate oxidase	3.37	1.38E-162	-	Nucleoid-associated protein	3.35	8.93E-141
ycsG	Putative transporter	3.37	1.11E-89	*oatA_1*	O-acetyltransferase OatA	3.32	1.09E-31
moaA	Molybdenum cofactor biosynthesis protein A	3.33	2.88E-283	*tusC*	tRNA 2-thiouridine synthesizing protein C	3.32	3.92E-47
ligA	ATP-dependent DNA ligase	3.32	2.92E-10	*rrmB*	16S rRNA m5C967 methyltransferase	3.32	9.24E-189
-	Putative NAD-dependent protein deacetylase	3.3	7.48E-27	*rimO*	Ribosomal protein S12 methylthiotransferase	3.31	1.45E-245
gpV	Phage baseplate assembly protein V	3.27	1.41E-40	*tig*	Trigger factor	3.28	0
fucU	L-fucose mutarotase	3.26	4.17E-11	*ttcA*	tRNA 2-thiocytidine biosynthesis protein TtcA	3.28	1.64E-138
-	Glycine radical enzyme, YjjI family	3.25	5.64E-139	*rnhB*	Ribonuclease HII	3.27	1.45E-94
-	hypothetical protein	3.25	1.33E-17	*rpoA*	DNA-directed RNA polymerase subunit alpha	3.26	0
-	Putative TPR repeat protein	3.23	1.46E-09	-	YheO-like PAS domain protein	3.26	3.06E-154
-	Putative type II secretory pathway, pseudopilin	3.23	1.39E-08	*ruvC*	Holliday junction resolvasome, endodeoxyribonuclease subunit	3.23	1.52E-84
gpX	Putative bacteriophage tail protein X	3.23	1.38E-10	*rpL28*	50S ribosomal protein L28	3.21	0
-	Hypothetical protein	3.23	2.10E-41	-	Hypothetical protein	3.19	2.64E-93
tnpA	Transposon Tn3 transposase	3.22	9.29E-167	*nlpD*	Putative metallopeptidase	3.18	1.94E-126
-	Hypothetical protein	3.22	9.17E-59	*sapF*	Peptide ABC transporter system, ATPase protein SapF	3.18	2.12E-87
gpQ	Putative bacteriophage capsid portal protein	3.22	4.38E-16	*tusD*	tRNA 2-thiouridine synthesizing protein D	3.17	2.85E-83
atoA	Acetyl-CoA:acetoacetyl-CoA transferase, beta subunit	3.21	1.93E-16	*rep_2*	ATP-dependent DNA helicase Rep	3.14	8.00E-167
cbiL	Nickel and cobalt ABC transporter component	3.21	7.64E-30	*pqiB*	Paraquat-inducible protein B	3.13	2.00E-150
-	Transglutaminase-like superfamily protein	3.2	1.08E-205	*rps7*	30S ribosomal subunit protein S7	3.12	1.18E-264
-	Hypothetical protein	3.2	6.93E-38	*cydD2*	Putative glutathione ABC transporter, fused ATPase and permease	3.12	1.52E-62
comC	Competence operon protein C	3.19	3.57E-12	-	Hypothetical protein	3.12	5.50E-29
merR2	Putative metal-binding transcriptional regulator	3.19	7.11E-62	*thrC*	Threonine synthase	3.1	9.01E-140
mglB	Galactoside ABC transporter, periplasmic binding protein	3.16	6.19E-241	*ispE*	4-diphosphocytidyl-2-C-methyl-D-erythritol kinase	3.1	2.05E-207
-	Hypothetical protein p47	3.16	8.30E-25	*ppiB*	Peptidyl-prolyl cis-trans isomerase B (rotamase B)	3.06	2.42E-142
talB	Transaldolase B	3.14	0	*cydA*	Cytochrome D ubiquinol oxidase subunit I	2.99	3.03E-294
iscR	Transcriptional regulator IscR	3.12	9.64E-197	-	Epimerase family protein	2.96	6.66E-68
maeB	NADP-dependent malic enzyme (NADP-ME)	3.12	1.32E-139	*brnQ*	Putative branched-chain amino acid transport system II carrier protein	2.95	4.61E-125
phoB	Phosphate regulon transcriptional regulatory protein	3.11	3.43E-199	*glmS*	Glucosamine-fructose-6-phosphate aminotransferase	2.95	2.61E-110
fdx-1	[2FE-2S] Ferredoxin, electron carrer protein	3.1	3.14E-151	*ksgA*	S-adenosylmethionine-6-N,N-adenosyl (rRNA) dimethyltransferase	2.95	1.96E-85
-	Hypothetical protein p56_1	3.1	1.66E-32	*rlmB*	23S rRNA methyltransferase	2.94	1.10E-66
-	Hypothetical protein p49	3.1	5.77E-146	-	Hypothetical outer membrane protein	2.94	2.43E-10
yciH	Translation initiation factor Sui1	3.09	1.97E-78	-	tRNA-Leu(taa)	2.94	2.07E-101

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
-	5S ribosomal RNA	3.08	0.04978	-	Hypothetical protein	2.93	2.22E-77
yccA	Modulator of FtsH protease YccA	3.06	4.25E-235	*ilvD*	Dihydroxyacid dehydratase	2.92	1.16E-157
glgA	Glycogen synthase	3.06	3.76E-162	*azlC*	Putative branched-chain amino acid permease AzlC	2.92	4.38E-37
serC	Phosphoserine aminotransferase	3.05	2.34E-144	*rumA*	23S rRNA methyltransferase	2.91	3.10E-74
pstA	Phosphate ABC transport system, permease component	3.05	8.41E-55	-	Putative permease	2.9	9.74E-97
ydeM	Hypothetical protein	3.05	7.70E-66	*pilF*	Transformation and Tfp-related protein PilF	2.9	1.17E-27
-	Hypothetical protein	3.02	5.53E-50	*fruA*	PTS system, fructose-specific IIBC component	2.89	5.03E-165
lon	ATP-dependent protease La	2.99	7.94E-197	-	Putative transport protein	2.87	2.52E-64
-	Hypothetical protein	2.99	1.82E-73	*dusB*	tRNA-dihydrouridine synthase B	2.87	6.48E-235
-	Putative bacteriophage integrase	2.98	2.38E-43	*rpS4*	30S ribosomal protein S4	2.87	0
yabJ	Enamine/imine deaminase	2.96	2.52E-64	*res*	Type III restriction-modification system restriction enzyme (HindVIP)	2.87	2.83E-202
-	Hypothetical protein	2.96	1.33E-68	-	Putative prophage antirepressor protein	2.86	5.64E-18
iscS	Cysteine desulfurase IscS	2.95	2.22E-229	*trpCF*	Indole-3-glycerol phosphate synthase/phosphoribosylanthranilate isomerase	2.85	2.03E-71
atoD	Acetyl-CoA:acetoacetyl-CoA transferase, alpha subunit	2.94	3.48E-13	*yebS*	Inner membrane protein YebS	2.84	9.60E-67
gdhA	Glutamate dehydrogenase, NADP-specific	2.92	4.55E-78	*lpxD*	UDP-3-O-(3-hydroxymyristoyl)-glucosamine N-acyltransferase	2.84	1.05E-142
-	Hypothetical protein	2.92	1.02E-37	*polA*	DNA polymerase I	2.83	8.62E-183
sixA	Phosphohistidine phosphatase SixA	2.91	1.93E-35	*prsA*	Ribose-phosphate pyrophosphokinase	2.83	1.75E-195
merP	Putative heavy metal chaperone protein	2.89	5.62E-45	*rpL14*	50S ribosomal protein L14	2.81	2.52E-283
-	Hypothetical protein	2.88	1.11E-09	*cydB*	Cytochrome d ubiquinol oxidase subunit II	2.8	4.70E-222
-	Putative type II secretory pathway, pseudopilin	2.88	2.57E-18	-	Hypothetical protein	2.8	9.09E-69
rbsK	Ribokinase	2.87	9.69E-83	*rpL25*	50S ribosomal protein L25	2.79	6.53E-217
metC	Cystathionine beta-lyase	2.86	4.88E-163	*menA*	1,4-dihydroxy-2-naphthoate octaprenyltransferase	2.78	4.98E-70
-	Hypothetical protein	2.86	0.000291	-	tRNA-Tyr(gta)	2.78	1.25E-168
glmM	Phosphoglucosamine mutase	2.85	8.99E-172	*nth*	DNA glycosylase and apyrimidinic (AP) lyase (endonuclease III)	2.77	3.92E-44
-	Hypothetical protein	2.85	1.25E-23	-	Putative transcriptional regulatory protein	2.76	5.78E-151
dapA	Dihydrodipicolinate synthetase	2.84	2.57E-243	*tdeA*	Outer membrane efflux porin TdeA	2.75	2.54E-170
fucK	L-fuculokinase	2.84	3.34E-13	*acrR*	Putative transcriptional regulator AcrR	2.73	1.46E-88
-	Hypothetical protein	2.84	4.87E-65	*pfs*	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	2.73	9.94E-54
-	Hypothetical protein	2.83	5.31E-150	-	Putative protein-S-isoprenylcysteine methyltransferase	2.73	4.77E-108
-	Hypothetical protein	2.81	6.49E-38	-	Hypothetical protein	2.72	1.58E-19
-	Hypothetical protein	2.8	5.38E-17	*rpL32*	50S ribosomal protein L32	2.71	1.19E-118
-	Putative N-carbamyl-L-amino acid amidohydrolase	2.79	1.67E-131	*azlD*	Putative branched-chain amino acid permease AzlD	2.7	1.07E-11
-	Hypothetical protein	2.79	0.002605948	*rlmN*	23S rRNA m(2) methyltransferase, SAM-dependent	2.69	6.77E-113
rec2	Recombination protein Rec2	2.75	4.10E-18	*thiI*	Thiamine biosynthesis protein ThiI	2.68	2.33E-120
-	Hypothetical protein	2.75	8.04E-157	*rpoC*	DNA-directed RNA polymerase beta' chain	2.68	2.30E-272
-	Hypothetical protein	2.75	2.10E-25	*can*	Carbonic anhydrase 2	2.68	9.46E-78
comF	Competence protein F	2.74	1.36E-19	*argR*	Arginine repressor	2.67	1.19E-35
queE	Putative 7-cyano-7-	2.73	1.78E-35	-	tRNA-Thr(tgt)	2.67	2.00E-89

	deazaguanine (preQ0) synthesis protein QueE						
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
-	Hypothetical protein	2.73	0.000320446	*greA*	Transcription elongation factor GreA	2.66	4.90E-67
glpQ	Glycerophosphoryl diester phosphodiesterase	2.72	4.93E-104	*lic3A*	Lipopolysaccharide alpha-2,3-sialyltransferase Lic3A	2.65	9.98E-24
-	Hypothetical protein p14_1	2.72	6.13E-59	-	tRNA-Val(tac)	2.65	4.98E-50
atoE	Short chain fatty acids transporter	2.71	1.60E-21	-	tRNA-Cys(gca)	2.64	8.63E-114
atoB	Acetyl-CoA acetyltransferase	2.7	1.65E-33	*cydC*	Cysteine/glutathione ABC transporter, fused ATPase and permease components	2.63	5.98E-84
folP	7,8-dihydropteroate synthase	2.7	6.33E-124	*mukE*	Condesin subunit E	2.63	6.83E-69
uxuA	Mannonate dehydratase	2.69	4.50E-131	*atpE*	Membrane-bound ATP synthase, F0 sector, subunit C	2.62	2.07E-110
purM	Phosphoribosylaminoimidazole synthetase	2.69	2.31E-33	*intA*	Putative integrase	2.61	1.23E-36
iscA	Iron-sulfur cluster assembly protein IscA	2.69	8.30E-120	-	Hypothetical protein	2.61	1.01E-61
-	Putative TRAP-type transport system, large permease component	2.67	2.69E-81	*dksA*	Regulator of rRNA transcription (DksA)	2.59	3.23E-73
-	Hypothetical protein	2.67	0.000281886	*cydC2*	Putative glutathione ABC transporter, fused ATPase and permease	2.59	9.07E-53
nanE	N-acetylmannosamine-6-phosphate 2-epimerase	2.66	1.46E-55	*rnfG*	Electron transport complex protein RnfG	2.58	9.69E-45
siaA	Lipooligosaccharide sialyltransferase SiaA	2.66	1.17E-18	*mltA*	Membrane-bound lytic murein transglycosylase A	2.57	2.33E-88
yhcB	Putative cytochrome d ubiquinol oxidase subunit 3	2.66	2.72E-67	*vapC 1*	Toxin-antitoxin locus protein VapC1	2.57	7.06E-56
-	Hypothetical protein	2.64	4.09E-64	*nrdB*	Ribonucleoside diphosphate reductase 1, beta subunit	2.56	5.48E-127
comE 1	Putative DNA uptake protein ComE1	2.63	1.98E-25	*rnfD*	Electron transport complex protein RnfD	2.56	8.51E-78
pepP	Aminopeptidase P	2.63	4.89E-179	*epmC*	Elongation factor P hydroxylase	2.56	2.25E-40
-	Putative phage tail-fiber protein	2.63	8.21E-11	*accA*	Acetyl-CoA carboxylase, subunit alpha	2.55	8.03E-150
sdaA	L-serine deaminase	2.62	2.41E-126	*deoD*	Purine-nucleoside phosphorylase (PNPase)	2.55	4.60E-183
-	Hypothetical protein	2.62	2.75E-58	*rpS21*	30S ribosomal subunit protein S21	2.55	2.40E-197
leuD	3-isopropylmalate dehydratase, small subunit	2.61	1.03E-22	-	Putative ABC transporter ATP-binding protein	2.54	3.33E-32
napF 2	Putative ferredoxin-type protein	2.61	6.42E-11	*gptA*	Xanthine-guanine phosphoribosyltransferase	2.54	2.23E-84
-	Hypothetical protein	2.6	3.07E-26	-	tRNA-Val(gac)	2.54	2.92E-12
hisG	ATP phosphoribosyltransferase	2.59	8.42E-29	-	tRNA-Val(tac)	2.53	1.74E-48
iscX	Fe-S cluster related protein IscX	2.59	3.04E-73	*secD*	General secretory pathway component SecD	2.52	3.21E-112
kdgK	2-dehydro-3-deoxygluconokinase (KDG kinase)	2.58	7.81E-68	*ftsL*	Cell division protein FtsL	2.52	9.79E-45
xylG	D-xylose ABC transporter, ATP-binding component	2.57	1.77E-16	*mukB*	Condesin subunit B	2.52	3.34E-174
-	Hypothetical protein	2.57	1.15E-10	*potC*	Spermidine/putrescine ABC transporter, permease protein	2.51	4.56E-134
aphA	Acid phosphatase/phosphotransferase	2.56	8.13E-10	-	tRNA-Leu(taa)	2.49	1.32E-59
rseC_ 1	Sigma-E factor regulatory protein RseC	2.55	6.65E-38	*rnr*	Exoribonuclease R (RNase R)	2.47	5.79E-117
hicB	Hif-contiguous protein B	2.55	1.59E-37	*apbE*	Lipoprotein ApbE	2.46	8.27E-154
-	tRNA-seC(tca)	2.55	3.37E-23	*hisS*	Histidyl-tRNA synthetase	2.46	6.03E-141
rfbP	Undecaprenyl-phosphate galactose phosphotransferase	2.53	2.86E-59	-	Hypothetical protein	2.45	9.63E-43
xylR	Xylose operon regulatory protein	2.52	3.50E-70	-	Oligopeptide transporter, OPT family	2.45	1.62E-139
lctP	L-lactate permease	2.51	4.93E-113	-	tRNA-Asp(gtc)	2.45	2.61E-30
sucA	2-oxoglutarate dehydrogenase E1 component	2.51	6.33E-64	*pyrG*	CTP synthetase	2.43	9.53E-122

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
iscU	Fe-S scaffold protein IscU	2.5	3.19E-187	*lpdA*	Dihydrolipoamide dehydrogenase	2.43	1.42E-202
kipI	Sporulation inhibitor KipI	2.49	2.15E-27	*efp*	Elongation factor P (EF-P)	2.42	1.04E-107
hisB	Imidazoleglycerolphosphate dehydratase and histidinol-phosphate phosphatase	2.49	4.48E-57	*yhhQ*	Inner membrane protein YhhQ	2.4	3.11E-52
grxA	Glutaredoxin 1	2.49	8.07E-122	-	tRNA-Gly(gcc)	2.39	1.66E-74
yfeA	Fe/Mn/Zn ABC transporter periplasmic-binding protein	2.48	6.47E-122	*thrB*	Homoserine kinase	2.38	1.91E-60
-	Putative thioredoxin	2.48	4.09E-75	*thiB*	Thiamin ABC transporter, periplasmic-binding protein	2.38	3.02E-112
hslV	Protease HslVU, peptidase subunit	2.47	9.70E-78	*rnfE*	Electron transport complex protein RnfE	2.38	5.93E-40
-	hypothetical protein	2.47	2.32E-68	*psiE*	Mig-7	2.38	5.21E-27
pilD	Type II secretory pathway, prepilin signal peptidase PilD	2.46	7.21E-09	-	tRNA-Gly(gcc)	2.38	1.31E-50
dld	D-lactate dehydrogenase	2.46	4.57E-89	-	Hypothetical protein HP2p08	2.38	1.21E-18
-	Hypothetical protein	2.45	8.85E-64	*rmuC*	DNA recombination protein rmuC-like protein	2.37	1.90E-40
-	Hypothetical protein	2.45	0.02664437	*thrA*	Aspartokinase I/homoserine dehydrogenase I	2.37	5.33E-70
rbsR	Ribose operon repressor	2.44	5.44E-38	*znuB*	Zinc transporter permease subunit ZnuB	2.37	4.86E-66
xylA	D-xylose isomerase	2.43	7.63E-22	-	EamA-like transporter family protein	2.37	4.87E-54
upp	Uracil phosphoribosyltransferase	2.43	1.34E-145	-	Hypothetical protein	2.37	4.82E-48
slp	Outer membrane protein slp precursor	2.42	2.48E-77	-	Hypothetical protein	2.36	6.33E-212
-	Putative DNA repair protein	2.41	2.59E-08	-	tRNA-Val(tac)	2.34	9.05E-88
dnaK	Molecular chaperone DnaK (Hsp70)	2.41	1.93E-15	-	tRNA-Gln(ttg)	2.34	8.88E-47
-	Putative pseudouridylate synthase	2.4	2.71E-33	*hflC*	Protease modulator complex HflKC, subunit HflC	2.33	4.20E-127
glpX	Fructose 1,6-bisphosphatase II	2.4	4.79E-44	*vapB 1*	Toxin-antitoxin locus protein VapB1	2.32	1.79E-40
pyrF	Orotidine 5'-phosphate decarboxylase	2.39	5.97E-105	*rumB*	23S rRNA m(5)U747-methyltransferase	2.32	9.86E-48
lldD	L-lactate dehydrogenase, FMN-linked	2.39	1.07E-144	*mukF*	Condesin subunit F	2.32	4.38E-69
-	Hypothetical protein	2.39	9.73E-08	*fruB*	PTS system, fructose-specific IIA/fpr component	2.31	2.98E-111
nanA	N-acetylneuraminate lyase (aldolase)	2.38	2.05E-85	*ispH*	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	2.31	3.86E-76
uup_ 4	DNA-binding ATPase Uup	2.38	7.48E-13	*mod*	Type III restriction-modification system methylase (M.HindVIP)	2.3	1.78E-124
-	Hypothetical protein	2.38	1.15E-78	*nrfA*	Nitrite reductase complex, periplasmic cytochrome C552 subunit	2.3	1.22E-32
yfbQ	Putative aminotransferase	2.35	8.13E-114	*rsmH*	16S rRNA m(4) methyltransferfase	2.3	5.67E-86
pstC	Phosphate ABC transport system, permease component	2.35	4.56E-22	*hda*	DNA replication initiation factor Had	2.29	1.26E-52
pgi	Glucose-6-phosphate isomerase	2.35	1.15E-188	*truC*	tRNA U65 pseudouridine synthase	2.28	6.83E-40
-	Hypothetical protein	2.35	4.59E-05	*fruK*	Fructose-1-phosphate kinase	2.28	1.08E-90
-	Hypothetical protein	2.35	1.32E-09	*rpoB*	DNA-directed RNA polymerase beta chain	2.28	5.25E-169
hifB	Pilin assembly chaperone HifB	2.34	2.67E-59	*rps12*	30S ribosomal subunit protein S12	2.28	4.88E-147
-	Putative alkylphosphonate utilization operon protein PhnA	2.34	1.42E-58	-	Hypothetical protein	2.28	9.21E-50
queD	Putative 7-cyano-7-deazaguanine (preQ0) synthesis protein QueD	2.34	1.32E-21	*rpS11*	30S ribosomal protein S11	2.28	1.01E-171
vapX	Toxin/antitoxin locus vapDX, antitoxin protein VapX	2.34	2.41E-62	*accB*	Acetyl-CoA carboxylase, biotin carboxyl carrier protein	2.28	2.71E-130
pdxY	Pyridoxine kinase	2.33	1.77E-104	*topA_ 1*	DNA topoisomerase I	2.27	8.50E-130
-	Hypothetical protein	2.31	0.001486	-	tRNA-Lys(ttt)	2.27	3.97E-22
-	Exopolysaccharide biosynthesis protein	2.31	5.85E-18	*acpP*	Acyl carrier protein	2.26	3.24E-158
oppB	Oligopeptide ABC	2.3	1.36E-75	-	Putative nucleoside	2.24	5.49E-167

	transporter, permease protein OppB				transporter		
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
glpC_2	Anaerobic glycerol-3-phosphate dehydrogenase subunit C	2.29	1.27E-92	-	Hypothetical protein	2.24	3.73E-78
fabA	3-hydroxydecanoyl-ACP dehydratase	2.29	6.05E-85	*yohI*	Truncated tRNA-dihydrouridine synthase YohI	2.24	5.24E-16
-	Putative integral membrane protein	2.29	2.75E-50	*srmB*	ATP-dependent RNA helicase	2.23	3.99E-82
-	Hypothetical protein	2.28	7.59E-16	*folC*	Bifunctional folylpolyglutamate synthase/dihydrofolate synthase	2.23	9.94E-84
-	Hypothetical protein	2.27	5.98E-25	*potB*	Spermidine/putrescine ABC transporter, permease protein	2.23	1.09E-114
-	Putative terminase, ATPase subunit	2.26	6.41E-57	*nqrB*	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrB	2.22	7.33E-168
-	Hypothetical protein p64	2.26	2.17E-60	*ompP2*	Outer membrane protein P2	2.21	6.77E-172
-	Hypothetical protein p59_3	2.26	4.92E-113	*ntpA*	dATP pyrophosphohydrolase	2.21	3.45E-45
-	Hypothetical protein	2.25	2.97E-17	*rseB*	Periplasmic negative regulator of sigmaE	2.21	6.68E-52
-	Hypothetical protein	2.25	2.55E-20	-	tRNA-Asp(gtc)	2.21	2.00E-24
-	Hypothetical protein	2.25	1.64E-20	*ompP1*	Outer membrane protein P1 precursor	2.2	2.05E-138
-	Hypothetical protein	2.24	3.48E-48	*thiP*	Thiamin ABC transporter, permease protein	2.2	1.70E-50
-	Hypothetical protein	2.22	3.17E-52	*rep_1*	Putative rep protein	2.2	1.14E-29
mdh	Malate dehydrogenase	2.21	7.23E-60	-	Hypothetical protein	2.2	0.000162
-	Hypothetical protein	2.21	1.49E-15	-	Hypothetical protein	2.19	4.44E-20
queC	Putative 7-cyano-7-deazaguanine (preQ0) synthesis protein QueC	2.2	2.38E-26	*ftnA2*	Ferritin protein A2	2.18	6.38E-35
-	Hypothetical protein	2.2	3.75E-06	-	tRNA-Val(tac)	2.18	4.25E-58
pepB	Aminopeptidase B	2.19	5.14E-111	*ribB*	3,4-dihydroxy-2-butanone 4-phosphate synthase	2.16	5.25E-50
nfuA	Fe/S biogenesis protein NfuA	2.18	2.53E-117	*moeB_1*	Molybdopterin biosynthesis protein MoeB	2.16	8.66E-42
xylB_1	Xylulose kinase	2.18	1.70E-36	*pnp*	Polynucleotide phosphorylase	2.15	4.60E-152
sucD	Succinyl-CoA synthetase, alpha subunit	2.18	1.36E-53	*narP*	Nitrate/nitrite response regulator protein	2.15	4.00E-62
ssb2	Single-stranded DNA-binding protein	2.17	5.30E-15	*hugZ*	Haem oxygenase	2.15	4.51E-99
hslU	Protease HslVU, ATPase subunit	2.17	2.02E-13	*trmB*	tRNA m(7)G46 methyltransferase, SAM-dependent	2.14	1.22E-58
-	Hypothetical protein p10	2.17	2.69E-17	*napD*	Periplasmic nitrate reductase assembly protein NapD	2.14	5.93E-11
yfeB	Fe/Mn/Zn ABC transporter ATP-binding protein	2.16	1.01E-111	*ftnA1*	Ferritin protein A1	2.14	9.29E-29
mgsA	Methylglyoxal synthase	2.16	4.11E-76	-	Hypothetical protein	2.14	3.05E-12
-	Hypothetical protein HP2p34	2.16	3.09E-21	*menH*	2-succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylate synthase	2.13	7.42E-24
-	Hypothetical protein	2.16	7.70E-13	*ispA2*	Intracellular septation protein A	2.13	3.19E-45
hscB	Fe-S cluster co-chaperone protein HscB	2.15	1.41E-115	*queF*	NADPH-dependent 7-cyano-7-deazaguanine reductase QueF	2.13	4.81E-46
purK	Phosphoribosylaminoimidazole carboxylase, ATPase subunit	2.15	1.47E-13	*bcr*	Putative efflux permease Bcr	2.12	9.22E-36
modE	Transcriptional regulator ModE	2.15	2.12E-35	*tyrZ*	Tyrosyl-tRNA synthetase	2.12	9.21E-98
-	Hypothetical protein	2.15	3.38E-13	*trkA*	TRK system potassium uptake protein	2.11	2.98E-76
-	Putative sulfate transport protein CysZ	2.14	1.48E-62	-	Putative ABC transport system, permease protein	2.11	7.06E-29
-	tRNA-Lys(ctt)	2.14	8.62E-24	*aceF*	Dihydrolipoamide acetyltransferase	2.1	7.49E-132
pqqL	Putative Zn-dependent	2.13	1.12E-19	*valS*	Valyl-tRNA synthetase	2.09	3.26E-134

	protease						
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
prtR_2	Pyocin repressor protein	2.13	1.85E-83	*dacA*	D-alanyl-D-alanine carboxypeptidase, penicillin-binding protein 5	2.08	1.14E-66
-	Hypothetical protein p50	2.13	7.48E-51	*mreC*	Rod shape-determining protein MreC	2.08	1.16E-42
-	Hypothetical protein	2.13	2.72E-67	*secF*	General secretory pathway component SecF	2.08	1.33E-67
-	Hypothetical protein p60	2.12	7.86E-23	*mraY*	Phospho-N-acetylmuramoyl-pentapeptide- transferase E	2.08	2.24E-58
topB_2	DNA topoisomerase III	2.1	1.20E-07	*pta*	Phosphate acetyltransferase	2.08	1.66E-81
pyrD	Dihydroorotate dehydrogenase	2.1	2.73E-59	-	tRNA-Gln(ttg)	2.08	2.46E-35
sucB	2-oxoglutarate dehydrogenase E2 component dihydrolipoamide succinyltransferase	2.1	3.02E-58	*fdnI*	Formate dehydrogenase-N, cytochrome B556(Fdn) gamma subunit, nitrate-inducible	2.07	2.21E-51
tnpR	Transposon Tn3 resolvase	2.1	6.16E-28	*aroK*	Shikimate kinase I	2.07	1.55E-73
-	Hypothetical protein	2.1	1.11E-22	-	Type I restriction enzyme EcoKI subunit R	2.07	6.44E-52
-	Putative long-chain-fatty-acid--CoA ligase (Long-chain acyl-CoA synthetase) (LACS)	2.09	5.81E-94	*hia*	Adhesin Hia	2.06	2.02E-131
uxuR	Uxu operon transcriptional regulator	2.09	1.39E-35	*rfaL*	Putative lipooligosaccharide biosynthesis protein	2.06	1.32E-29
hifD	Pilus tip protein HifD	2.09	1.28E-36	-	Hypothetical protein	2.05	3.19E-24
lpxH	UDP-2,3-diacylglucosamine hydrolase	2.08	1.30E-41	*nqrC*	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrC	2.04	6.78E-122
dmsB	Anaerobic dimethyl sulfoxide reductase, subunit B	2.08	5.04E-27	*murJ*	Peptidoglycan lipid II flippase	2.04	3.95E-41
-	Hypothetical protein	2.08	1.21E-16	*era*	GTP-binding protein	2.03	3.95E-57
-	Hypothetical protein	2.08	4.13E-06	*nqrE*	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrE	2.03	8.37E-54
nanK	N-acetylmannosamine kinase	2.07	9.13E-45	*fdnG_1*	Formate dehydrogenase-N, major subunit	2.03	1.16E-100
hscA	Fe-S cluster chaperone protein HscA	2.07	9.85E-122	*ackA*	Acetate kinase	2.03	6.60E-113
higA	Putative toxin-antitoxin locus protein (HigA-family)	2.07	9.65E-58	*nusG*	Transcription antitermination protein NusG	2.01	2.61E-88
-	Putative capsid portal protein	2.07	2.85E-18	*mtgA*	Biosynthetic peptidoglycan transglycosylase	2.01	1.57E-10
-	Hypothetical protein	2.06	4.36E-11	*pheS*	Phenylalanyl-tRNA synthetase, alpha subunit	2.01	1.27E-76
-	tRNA-Lys(ttt)	2.05	8.24E-21	*sapD*	Peptide ABC transporter system, ATPase protein SapD	2.01	3.22E-39
lsgA_3	Lipopolysaccharide biosynthesis protein LsgA	2.02	1.69E-28				
pstS	Phosphate ABC transport system, periplasmic binding protein	2.02	7.76E-20				
-	Hypothetical protein	2.02	5.73E-09				

Table S3: Differentially expressed genes in the Rd strain during oxidative stress.

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
hktE	Catalase	76.94	0	*artP*	Arginine transporter ATP-binding protein	21.55	6.15E-125
acpD	Acyl carrier protein phosphodiesterase	26.07	3.32E-153	*dmsA_3*	Anaerobic dimethyl sulfoxide reductase subunit A	11.89	2.25E-31
-	DoxX	14.58	3.42E-141	*artI*	Arginine ABC transporter substrate-binding protein	11.01	1.06E-24
ilvC	Ketol-acid reductoisomerase	11.48	9.71E-47	*artQ*	Arginine transporter permease subunit ArtQ	9.12	5.79E-11
dps	DNA protection during starvation protein	10.31	1.21E-45	*dmsB*	Anaerobic dimethyl sulfoxide reductase subunit B	7.98	9.13E-36
-	DNA polymerase V subunit UmuD	10.31	6.89E-160	*nrfC*	Nitrite reductase Fe-S protein	7.75	1.63E-27
-	Peroxiredoxin hybrid Prx5	7.56	1.06E-137	*artM*	Arginine transporter permease subunit ArtM	7.63	1.34E-08
ilvI	Acetolactate synthase 3 catalytic subunit	4.88	1.39E-33	*nrfB*	Cytochrome c nitrite reductase pentaheme subunit	7.5	2.08E-22
recN	DNA repair protein	4.68	1.28E-98	-	Twin-argninine leader-binding protein DmsD	7.24	5.31E-41
ilvH	Acetolactate synthase 3 regulatory subunit	4.52	1.21E-29	*nrfA*	Cytochrome c552	6.89	1.32E-09
-	FeS assembly protein IscX	4.51	4.91E-48	*hlpA*	D-methionine-binding lipoprotein MetQ	6.24	6.76E-46
ribA	GTP cyclohydrolase II	4.43	2.67E-45	*yecK*	Cytochrome C-like protein	6	3.63E-14
fdx-1	Ferredoxin	4.42	7.02E-52	*pstA_1*	Phosphate ABC transporter permease	5.93	1.02E-21
trxM	Thioredoxin	3.98	2.05E-51	*cydD_2*	ABC transporter ATP-binding protein	5.59	4.51E-62
gdhA	Glutamate dehydrogenase	3.98	3.07E-14	*napF_2*	Ferredoxin-type protein	5.56	7.16E-21
iscU	Scaffold protein	3.91	3.58E-58	*pepT*	Peptidase T	5.34	3.06E-13
nifS	Cysteine desulfurase	3.77	9.76E-50	*dmsC*	Anaerobic dimethyl sulfoxide reductase subunit C	5.27	9.60E-37
lexA	LexA repressor	3.69	5.85E-40	*argR*	Arginine repressor	5.26	9.67E-30
hscA	Chaperone protein HscA	3.48	3.79E-88	*cydD_1*	ABC transporter ATP-binding protein	5.2	5.94E-18
-	Transporter	3.44	3.11E-20	-	ABC transporter ATP-binding protein	4.87	1.93E-74
-	Iron-sulfur cluster insertion protein ErpA	3.39	1.07E-62	-	Hypothetical protein	4.86	3.04E-11
hscB	Co-chaperone HscB	3.3	1.72E-74	-	Hemoglobin-binding protein	4.77	1.20E-20
-	Hypothetical protein	3.19	4.41E-47	*argH*	Argininosuccinate lyase	4.72	3.88E-22
recX	Recombination regulator RecX	3.18	3.26E-46	*nrfD*	Nitrite reductase transmembrane protein	4.65	5.16E-27
-	Manganese transport protein MntH	3.08	4.51E-32	*argG*	Argininosuccinate synthase	4.56	8.00E-27
gor	Glutathione reductase	2.96	3.33E-43	*oapA_3*	Hemoglobin-binding protein	4.45	3.14E-24
uvrA	Excinuclease ABC subunit A	2.89	6.15E-45	-	Epimerase family protein	4.1	5.15E-15
iscR	HTH-type transcriptional regulator IscR	2.86	3.96E-25	*ccmA*	Cytochrome c biogenesis protein CcmA	3.81	2.27E-11
gnd	6-phosphogluconate dehydrogenase	2.83	4.05E-46	*trpE*	Anthranilate synthase component I	3.78	0.000118629
dapF	Diaminopimelate epimerase	2.78	5.45E-74	*metN*	DL-methionine transporter ATP-binding protein	3.77	0.000137417
ruvB	Holliday junction DNA helicase RuvB	2.66	1.06E-43	-	Hemoglobin-binding protein	3.68	2.91E-21
pntB	Pyridine nucleotide transhydrogenase	2.66	8.61E-106	*bisC*	Biotin sulfoxide reductase	3.6	1.20E-14
pntA	NAD(P) transhydrogenase subunit alpha	2.64	4.38E-82	*ccmB*	Haem exporter protein B	3.49	5.15E-09
kipI	Sporulation inhibitor KipI	2.62	4.37E-13	*ribB*	3,4-dihydroxy-2-butanone 4-phosphate synthase	3.47	1.48E-35
radA	DNA repair protein RadA	2.61	2.61E-41	-	Proton glutamate symport protein	3.44	0.003297323
recA	Recombinase A	2.57	4.77E-58	*bioD_1*	Dithiobiotin synthetase	3.41	6.66E-08
ftsH	Cell division protein	2.57	1.18E-32	*dppA*	Haem-binding lipoprotein	3.32	0.003442

Gene	Product	Fold change	Adjusted p-value
Up-regulated genes			
adhC	Alcohol dehydrogenase class III	2.5	1.51E-62
-	LamB/YcsF family protein	2.47	1.31E-10
ruvA	Holliday junction DNA helicase RuvA	2.45	6.73E-29
-	Cell division FtsH-like protein	2.45	1.70E-42
-	Alkylhydroperoxidase AhpD family core domain protein	2.43	4.80E-11
-	Hypothetical protein	2.42	6.04E-30
-	Putative membrane protein	2.41	5.94E-34
-	Transposase	2.38	5.95E-16
-	Esterase	2.3	6.08E-41
nagB_2	Glucosamine-6-phosphate deaminase	2.26	3.13E-28
kipA	KipI antagonist	2.25	4.66E-13
rsgA_1	Ferritin like protein 1	2.23	4.66E-13
-	Hypothetical protein	2.23	0.00227
-	Protease	2.18	7.88E-14
-	Thioredoxin	2.15	2.14E-14
uvrD	DNA-dependent helicase II	2.04	7.73E-34

Gene	Product	Fold change	Adjusted p-value
Down-regulated genes			
galR	LacI family transcriptional repressor	3.3	1.32E-05
malQ	4-alpha-glucanotransferase	3.29	1.66E-06
cdd	Cytidine deaminase	3.24	7.83E-09
aspT	Aspartate/alanine antiporter	3.14	7.18E-06
-	Formate transporter	3.12	5.08E-11
-	3-hydroxyisobutyrate dehydrogenase	3.12	2.35E-06
pckA	Phosphoenolpyruvate carboxykinase	3.07	1.23E-05
ansB	L-asparaginase II	3.04	9.02E-07
rbsD	D-ribose pyranase	2.95	1.25E-09
-	Hypothetical protein	2.93	9.26E-07
ygbM	Putative hydroxypyruvate isomerase YgbM	2.92	1.95E-09
-	Aldolase	2.91	4.22E-08
-	Hypothetical protein	2.83	0.000184
hslO	Hsp33-like chaperonin	2.73	7.14E-06
lrgA	Antiholin-like protein LrgA	2.73	1.43E-07
ccrB	Camphor resistance protein CrcB	2.72	9.16E-10
pepE	Peptidase E	2.67	5.96E-08
-	Short chain dehydrogenase/reductase	2.67	6.91E-07
tyrA	Bifunctional chorismate mutase/prephenate dehydrogenase	2.66	4.56E-13
ccmC	Haem exporter protein C	2.61	1.09E-07
mtr	Tryptophan-specific transport protein	2.55	1.52E-05
-	Putative epimerase/dehydratase	2.54	3.63E-09
-	Hemoglobin-binding protein	2.52	2.54E-11
-	Amino acid ABC transporter substrate-binding protein	2.51	0.045661216
-	N-acetylmannosamine-6-phosphate 2-epimerase	2.5	1.23E-05
-	Haloacid dehalogenase-like protein	2.47	5.52E-10
uspA	Universal stress protein A	2.45	2.97E-05
rpmC	50S ribosomal protein L29	2.43	1.30E-70
sdaC	Serine transporter	2.38	0.001485
gntP_3	Gluconate permease	2.37	1.12E-14
-	Formate acetyltransferase	2.36	4.10E-12
trpG_2	Anthranilate synthase component II	2.36	0.031678415
-	Nickel uptake substrate-specific transmembrane region	2.36	1.08E-19
dcuD	Putative cryptic C4-dicarboxylate transporter DcuD	2.35	2.48E-07
yohK	Inner membrane protein YohK	2.33	1.84E-06
nudF	ADP-ribose pyrophosphatase	2.33	0.001353
rplP	50S ribosomal protein L16	2.33	2.57E-60
galT	Galactose-1-phosphate uridylyltransferase	2.31	0.000116
rpsQ	30S ribosomal protein S17	2.27	2.41E-52
-	Autonomous glycyl radical cofactor GrcA	2.26	6.73E-09
arcA	Two-component response regulator	2.25	3.03E-07
-	Hypothetical protein	2.25	1.45E-22
rpsC	30S ribosomal protein S3	2.21	4.21E-61
-	Hypothetical protein	2.21	0.016991
vapA	Virulence-associated protein A	2.2	2.74E-30
fumC	Fumarate hydratase	2.2	1.99E-07
aspA	Aspartate ammonia-lyase	2.19	0.033786

Gene	Product	Fold change	Adjusted p-value
Down-regulated genes			
ccmD	Haem exporter protein D	2.19	6.95E-06
trpB	Tryptophan synthase subunit beta	2.19	0.000454738
yjjP	Inner membrane protein YjjP	2.13	6.70E-06
-	N-acetylmannosamine kinase	2.13	3.99E-05
rplV	50S ribosomal protein L22	2.13	5.83E-40
-	Hypothetical protein	2.11	0.000167
tyrP_1	Tyrosine-specific transport protein	2.07	1.07E-14
-	TRAP transporter, DctM subunit	2.07	2.06E-06
-	tRNA-Trp(cca)	2.07	0.004022
yhxB_1	Phosphomannomutase	2.05	0.000143546
rpsS	30S ribosomal protein S19	2.05	1.42E-30
ccmE	Cytochrome c-type biogenesis protein CcmE	2.04	3.97E-07
-	Branched chain amino acid ABC transporter substrate-binding protein	2.04	1.78E-10

Table S4: Differentially expressed genes in the R2866 strain during oxidative stress.

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
hktE	Catalase	111.22	6.53E-231	*Hgd*	2-(hydroxymethyl)glutarate dehydrogenase	7.82	8.12E-34
-	Hypothetical protein	26.61	4.74E-65	-	Hypothetical protein	6.52	1.45E-40
-	Hypothetical protein	19.57	5.72E-63	*ygbM*	Putative hydroxypyruvate isomerase YgbM	6.11	8.91E-32
-	Putative NAD-dependent protein deacetylase	17.02	1.99E-69	*ygbL*	Putative sugar aldolase/epimerase	5.98	3.38E-40
-	Hypothetical protein	15.63	1.71E-20	-	Putative sugar epimerase	4	6.23E-09
dpsA	DPS ferritin-like protein	13.68	3.16E-51	-	Putative permease	3.18	1.46E-09
-	DNA polymerase V subunit UmuD	13.26	9.21E-104	*ftnA2*	Ferritin protein A2	2.67	1.68E-07
-	Putative 5'(3')-deoxyribonucleotidase	12.87	1.22E-40	-	Hypothetical protein	2.57	0.0102554
azoR	FMN-dependent NADH-azoreductase	11.36	6.37E-29	*galR*	Galactose operon regulator	2.53	1.04E-07
-	Hypothetical protein	10.23	2.91E-59	*rpS18*	30S ribosomal subunit protein S18	2.41	2.23E-06
-	Putative NAD-dependent protein deacetylase	9.75	5.49E-18	*yjjP*	Inner membrane protein YjjP	2.35	0.038421349
recN	DNA repair protein RecN	8.78	5.27E-161	*psd*	Phosphatidylserine decarboxylase	2.35	3.54E-08
pgdX	Peroxiredoxin/glutaredoxin glutathione-dependent peroxidase	8.3	3.83E-52	-	Transglutaminase-like superfamily protein	2.35	1.91E-14
doc	Death on curing protein	5.46	1.19E-147	-	Selenium metabolism protein YedF	2.3	0.005077797
lexA	SOS-response transcriptional repressor LexA	5.3	2.78E-15	*siaT*	Sialic acid TRAP transporter, fused permease protein SiaT	2.28	0.014223162
recX	RecA regulator RecX	5.25	1.14E-102	*ftnA1*	Ferritin protein A1	2.27	2.78E-10
gnd	6-phosphogluconate dehydrogenase, decarboxylating (6PGD)	4.96	9.81E-24	-	Hypothetical protein	2.27	3.92E-08
recA	DNA recombination protein RecA	4.96	1.10E-20	*potD*	Spermidine/putrescine ABC transporter, periplasmic-binding protein	2.19	6.87E-21
msrAB	Peptide methionine sulfoxide reductase	4.64	6.08E-14	-	Hypothetical protein	2.11	0.0011555
-	Hypothetical protein	4.29	1.31E-14	*brnQ*	Putative branched-chain amino acid transport system II carrier protein	2.1	0.025729252
-	Hypothetical protein	4.27	4.40E-27	*merP*	Putative heavy metal chaperone protein	2.1	0.03213608
hxuC	Haem-hemopexin utilization protein C	3.97	7.54E-17	*ahpC*	Peroxiredoxin	2.08	7.21E-05
hemR	Putative TonB-dependent haem receptor	3.9	2.05E-31	*trpE_1*	Anthranilate synthase component I	2.07	0.025063016
-	Hypothetical protein	3.75	1.31E-32	*cstA*	Carbon starvation protein A	2.06	0.000540
ruvB	Holliday junction resolvasome, ATPase subunit	3.72	2.49E-43	*cdd*	Cytidine deaminase	2.06	0.045591017
-	DoxX	3.68	6.19E-10	*tgt*	tRNA-guanine transglycosylase	2.02	3.53E-10
hitA	Iron(III) ABC transporter periplasmic-binding protein	3.65	4.04E-12	-	Hypothetical protein	2.01	0.000521177
hxuB	Haem-hemopexin utilization protein B	3.63	1.52E-24				
ruvA	Holliday junction resolvasome, DNA-binding subunit	3.5	6.30E-49				
uvrA	Excinuclease ABC subunit A	3.47	1.49E-34				
radA	DNA repair protein RadA	3.43	4.21E-06				
-	CRISPR associated protein Cas2	3.42	1.38E-32				
fdx-1	[2FE-2S] Ferredoxin, electron carrer protein	3.37	1.38E-05				
dapF	Diaminopimelate epimerase	3.17	1.62E-42				
trxM	Thioredoxin	3.01	3.92E-08				
ribA	GTP cyclohydrolase II	2.96	1.12E-06				
uvrD	DNA helicase II	2.84	1.11E-59				

Gene	Product	Fold change	Adjusted p-value
Up-regulated genes			
mfd	Transcription-repair coupling factor	2.83	4.19E-29
hscA	Fe-S cluster chaperone protein HscA	2.81	3.50E-11
zwf	Glucose-6-phosphate dehydrogenase (G6PD)	2.79	2.30E-14
iscX	Fe-S cluster related protein IscX	2.79	2.78E-15
-	Hypothetical protein	2.78	0.000436
yeaD	Putative glucose-6-phosphate 1-epimerase	2.73	1.23E-11
iscU	Fe-S scaffold protein IscU	2.72	0.000432
bphH	Putative glutathione-S-transferase	2.68	1.27E-06
hxuA	Haem-hemopexin utilization protein A	2.61	1.64E-05
glpK	Glycerol kinase	2.57	3.92E-11
pgl	putative 6-phosphogluconolactonase (6PGL)	2.54	5.93E-05
tbp1	Transferrin-binding protein 1	2.52	1.91E-13
iscA	iron-sulfur cluster assembly protein IscA	2.48	8.14E-20
gor	Glutathione oxidoreductase	2.47	3.19E-10
iscS	Cysteine desulfurase IscS	2.43	1.28E-14
-	Putative thioredoxin	2.42	1.02E-05
fbp	Fructose-1,6-bisphosphatase	2.41	9.32E-12
artI	Arginine ABC transporter, periplasmic-binding protein ArtI	2.32	1.91E-08
ftsH	ATP-dependent protease FtsH	2.32	0.000843036
ccmG_1	Haem lyase/disulfide oxidoreductase (DsbE)	2.31	0.000713536
-	Alkylhydroperoxidase AhpD family core domain protein	2.29	5.86E-06
groES	GroES, chaperone Hsp10	2.27	0.01283014
iscR	Transcriptional regulator IscR	2.24	4.83E-09
ompU1	Putative outer membrane protein OmpU1	2.22	2.11E-05
-	Hypothetical protein	2.19	3.23E-08
talB	Transaldolase B	2.16	1.35E-08
pgi	Glucose-6-phosphate isomerase	2.13	2.13E-06
groEL	GroEL, chaperone Hsp60	2.1	0.038421349
-	Transposase	2.1	4.05E-05
dnaK	Molecular chaperone DnaK (Hsp70)	2.09	0.002122231
tbp2	Transferrin-binding protein 2	2.08	0.000230
mdaB	NADPH quinone reductase	2.03	0.008384083

Table S5: Differentially expressed genes in the Rd strain during iron-starvation stress.

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
lptF	Lipopolysaccharide export system permease protein LptF	22.71	8.87E-196	-	Proton glutamate symport protein	17.82	1.58E-59
-	Lipopolysaccharide biosynthesis protein	6.3	2.41E-20	*hlpA*	D-methionine-binding lipoprotein MetQ	9.96	2.51E-97
-	Lipopolysaccharide biosynthesis protein	5.5	1.20E-73	*pstA_1*	Phosphate ABC transporter permease	8.6	2.92E-78
-	Na(+)-translocating NADH-quinone reductase subunit E	4.95	1.55E-18	-	Amino-acid ABC transporter ATP-binding protein	8.56	3.48E-38
ppc	Phosphoenolpyruvate carboxylase	4.92	3.99E-75	-	Amino acid ABC transporter substrate-binding protein	7.8	1.63E-09
merR 2	Mercuric resistance operon regulatory protein	4.77	8.18E-26	*metN*	DL-methionine transporter ATP-binding protein	7.72	5.30E-54
-	Branched chain amino acid ABC transporter substrate-binding protein	4.59	1.43E-63	-	Amino acid ABC transporter permease	7.04	1.77E-12
-	Hypothetical protein	4.59	3.10E-60	*dppA*	Haem-binding lipoprotein	5.41	6.17E-10
-	Cobalt transport protein CbiM	4.55	1.12E-65	-	Long chain fatty acid CoA ligase	5.19	4.47E-69
-	Cobalt ABC transporter, permease protein CbiQ	4.37	7.15E-53	*merT*	Mercuric ion transport protein	4.69	2.45E-33
-	ABC transporter ATP-binding protein	4.36	9.55E-26	*merP_2*	Mercuric ion scavenger protein	4.49	7.82E-24
fiu	TonB-dependent receptor Fiu	4.25	3.26E-71	*fdhE*	Formate dehydrogenase accessory protein FdhE	4	2.00E-52
cspD	Cold shock-like protein	3.73	6.71E-13	*ilvC*	Ketol-acid reductoisomerase	3.96	4.12E-36
-	Nickel uptake substrate-specific transmembrane region	3.7	3.06E-09	-	Transglutaminase-like superfamily protein	3.72	1.53E-34
-	Hypothetical protein	3.69	2.27E-51	*ilvI*	Acetolactate synthase 3 catalytic subunit	3.58	2.17E-26
-	Hypothetical protein	3.54	2.23E-11	*ilvH*	Acetolactate synthase 3 regulatory subunit	3.33	3.83E-28
-	TonB-dependent Receptor Plug Domain protein	3.52	2.06E-11	-	Pyridoxamine kinase	3.31	1.82E-45
-	Zinc protease	3.46	4.03E-32	*cydD_2*	ABC transporter ATP-binding protein	3.29	2.57E-17
rrmJ	23S rRNA methyltransferase J	3.45	1.25E-28	*tgt*	Queuine tRNA-ribosyltransferase	3.23	1.78E-36
-	Serine protease	3.35	2.73E-33	*dmsA_3*	Anaerobic dimethyl sulfoxide reductase subunit A	3.13	0.001795 612
-	ABC-type uncharacterized transport system, periplasmic component	3.34	1.56E-14	*yecK*	Cytochrome C-like protein	3.12	0.000319 396
ltaS1	Lipoteichoic acid synthase 1	3.21	3.23E-13	*potD_2*	Spermidine/putrescine ABC transporter substrate-binding protein	3.01	1.89E-08
dnaK	Molecular chaperone DnaK	3.19	3.80E-05	*oppF*	Oligopeptide ABC transporter ATP-binding protein	3	2.15E-30
lctP	L-lactate permease	3.18	4.62E-18	*dcuD*	Putative cryptic C4-dicarboxylate transporter DcuD	2.96	1.07E-15
-	Hypothetical protein	3.12	3.35E-09	*nrfB*	Cytochrome c nitrite reductase pentaheme subunit	2.92	2.39E-10
ligA_ 2	DNA ligase	3.1	4.51E-15	*oppD*	Oligopeptide transporter ATP-binding protein	2.77	4.22E-34
artI	Arginine ABC transporter substrate-binding protein	3.08	0.000751 252	*fdxI*	Formate dehydrogenase subunit gamma	2.75	1.58E-23
yjgA	x96 protein	3.02	1.12E-09	-	Diaminobutyrate--2-oxoglutarate aminotransferase	2.75	6.00E-08
lptC	Lipopolysaccharide export system protein LptC	2.98	1.17E-31	*fdxH*	Formate dehydrogenase subunit beta	2.72	2.76E-22

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
lptA	Lipopolysaccharide export system protein LptA precursor	2.96	2.67E-15	*pepE*	Peptidase E	2.7	3.77E-08
comM	Competence protein	2.93	6.47E-29	*nrfA*	Cytochrome c552	2.64	0.004472781
xylF	D-xylose transporter subunit XylF	2.92	5.84E-08	*ilvD*	Dihydroxy-acid dehydratase	2.59	7.63E-13
-	Aminotransferase AlaT	2.89	2.30E-22	*aroA*	3-phosphoshikimate 1-carboxyvinyltransferase	2.58	8.83E-16
-	Putative epimerase/dehydratase	2.88	2.46E-12	*deaD*	ATP-dependent RNA helicase	2.55	4.57E-12
-	TPR repeat-containing protein precursor	2.87	5.78E-14	-	Hydrolase	2.53	5.07E-16
dprA	DNA processing chain A	2.85	1.64E-17	-	Aminotransferase AlaT	2.51	1.02E-15
radC	DNA repair protein RadC	2.8	2.61E-12	-	ABC transporter ATP-binding protein	2.51	1.87E-15
mutT	Mutator protein	2.75	3.93E-13	*mesJ*	Cell cycle protein	2.48	7.42E-20
-	Hypothetical protein	2.75	2.19E-15	*hisC*	Histidinol-phosphate aminotransferase	2.48	1.30E-12
acrB	Acriflavine resistance protein	2.72	1.45E-08	*hisB*	Imidazole glycerol-phosphate dehydratase/histidinol phosphatase	2.46	2.30E-17
-	Membrane-fusion protein	2.68	6.30E-30	*tsf*	Elongation factor Ts	2.43	9.81E-16
-	Transcriptional repressor	2.67	1.53E-09	*hisG*	ATP phosphoribosyltransferase	2.42	1.76E-22
fnr_1	Anaerobic regulatory protein	2.64	3.00E-14	*ilvA*	Threonine dehydratase	2.42	1.70E-13
clpB	ATP-dependent Clp protease ATPase subunit	2.64	3.53E-26	*hisD*	Histidinol dehydrogenase	2.41	4.46E-14
-	Hypothetical protein	2.63	2.26E-07	*oppC*	Oligopeptide ABC transporter permease	2.41	1.77E-28
-	Putative small periplasmic lipoprotein	2.58	0.003511826	*menD*	2-succinyl-5-enolpyruvyl-6-hydroxy-3- cyclohexene-1-carboxylate synthase	2.36	1.71E-27
-	Putative permease, DMT superfamily	2.56	9.86E-05	*asd*	Aspartate-semialdehyde dehydrogenase	2.36	4.36E-10
-	Phi X174 lysis protein	2.51	2.39E-10	*rpsB*	30S ribosomal protein S2	2.36	4.18E-15
asnA	Asparagine synthetase AsnA	2.47	1.55E-07	*nrfC*	Nitrite reductase Fe-S protein	2.36	2.41E-12
groEL	Chaperonin GroEL	2.46	2.37E-12	*trpA*	Tryptophan synthase subunit alpha	2.31	1.41E-12
groES	Co-chaperonin GroES	2.44	6.04E-12	*trpB*	Tryptophan synthase subunit beta	2.3	2.24E-16
hslU	ATP-dependent protease ATP-binding subunit HslU	2.42	2.96E-11	*greA*	Transcription elongation factor GreA	2.29	3.69E-06
hslV	ATP-dependent protease peptidase subunit	2.41	8.26E-15	*pyrG*	CTP synthetase	2.28	9.41E-09
comA	Competence protein A	2.4	0.000661	-	Hypothetical protein	2.27	5.80E-22
lon	ATP-dependent proteinase	2.4	9.82E-19	-	Lipoprotein	2.26	1.96E-19
comD	Competence protein D	2.39	2.47E-14	*speF*	Ornithine decarboxylase	2.24	7.13E-13
comB	Competence protein B	2.39	1.30E-26	*nrfD*	Nitrite reductase transmembrane protein	2.24	9.81E-09
yfeB	Iron (chelated) transporter ATP-binding protein	2.38	5.85E-07	*dmsA_1*	Anaerobic dimethyl sulfoxide reductase subunit A	2.23	3.35E-15
yfeA	Iron-chelated ABC transporter substrate-binding protein	2.38	2.58E-09	*adhC*	Alcohol dehydrogenase class III	2.21	8.43E-09
yfeC	iron (chelated) ABC transporter permease	2.37	2.65E-10	*dmsB*	Anaerobic dimethyl sulfoxide reductase subunit B	2.21	2.97E-05
yfeD_1	Iron (chelated) ABC transporter permease	2.36	5.10E-06	*purU*	Formyltetrahydrofolate deformylase	2.2	3.18E-07
-	ABC transporter permease	2.34	1.81E-06	-	Hypothetical protein	2.18	1.01E-09
-	Thiamine biosynthesis protein	2.34	2.32E-09	-	SprT-like family protein	2.17	5.31E-07
-	ABC transporter ATP-binding protein	2.33	1.28E-23	-	TPR repeat-containing protein precursor	2.16	7.62E-11
-	Lipooligosaccharide biosynthesis protein	2.32	6.63E-14	*hslO*	Hsp33-like chaperonin	2.16	4.82E-10
-	Prepilin peptidase-dependent protein D	2.31	1.73E-17	*oppB*	Oligopeptide transporter permease	2.16	3.03E-20
-	Protein transport protein	2.27	0.000189	*rplT*	50S ribosomal protein L20	2.16	7.53E-16
hopD	Type 4 prepilin-like protein specific leader peptidase	2.26	1.46E-10	*pldB*	Lysophospholipase L2	2.15	2.60E-06
-	Type IV pilin secretion	2.26	1.43E-09	*purK*	Phosphoribosylaminoimid	2.15	1.62E-06

	protein				azole carboxylase ATPase subunit		
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
merP_1	Mercuric ion scavenger protein	2.25	4.03E-21	*mtr*	Tryptophan-specific transport protein	2.13	3.77E-05
-	Mercury transport-like protein	2.23	4.72E-06	-	2,3-diketo-L-gulonate reductase	2.13	2.71E-05
-	Mercury transport-like protein	2.22	3.37E-14	*luxS*	S-ribosylhomocysteinase	2.12	2.83E-15
queA	S-adenosylmethionine:tRNA ribosyltransferase-isomerase	2.19	1.16E-05	-	N-acetyl-D-glucosamine kinase	2.1	8.26E-15
brnQ	Branched-chain amino acid ABC transporter	2.18	2.96E-14	*nusA*	Transcription elongation factor NusA	2.1	1.69E-19
-	Hypothetical protein	2.18	1.06E-05	-	Hypothetical protein	2.1	4.81E-16
-	Amino acid carrier protein	2.17	5.48E-18	*gsp*	Bifunctional glutathionylspermidine synthetase/amidase	2.09	2.21E-19
gdhA	Glutamate dehydrogenase	2.17	1.24E-25	*napC*	Cytochrome C-type protein	2.09	1.25E-14
nagB_1	Glucosamine-6-phosphate deaminase	2.16	2.99E-09	*infB*	Translation initiation factor IF-2	2.08	1.35E-14
nanA	N-acetylneuraminate lyase	2.16	2.59E-10	*betT*	High affinity choline transport protein	2.08	2.11E-07
siaP	Neu5Ac-binding protein	2.16	8.03E-08	-	Acyl-CoA thioester hydrolase YfbB	2.06	1.76E-05
hitC	Iron(III) ABC transporter ATP-binding protein	2.14	1.76E-05	*ndk*	Nucleoside diphosphate kinase	2.05	4.54E-11
htpG	Heat shock protein 90	2.14	5.04E-06	*cydD_1*	ABC transporter ATP-binding protein	2.03	0.007520324
hemR	Haemin receptor	2.14	1.24E-14	*guaA*	GMP synthase	2.02	2.96E-15
hitB	Iron(III) ABC transporter permease	2.13	0.006737257	*fabA*	3-hydroxydecanoyl-ACP dehydratase	2.01	4.74E-23
hitA	Iron-utilization periplasmic protein hFbpA	2.12	3.66E-07	*trmE*	tRNA modification GTPase TrmE	2	4.81E-16
-	Nucleotidyltransferase	2.09	0.000263				
grpE	Heat shock protein GrpE	2.08	5.28E-07				
yiaM_1	2,3-diketo-L-gulonate TRAP transporter small permease protein YiaM	2.06	0.00093868				
yiaO_1	Extracytoplasmic solute receptor protein YiaO	2.06	2.77E-09				
lldD	L-lactate dehydrogenase	2.02	1.17E-18				

Table S6: Differentially expressed genes in the R2866 strain during iron-starvation stress.

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
-	Putative TonB-dependent transport protein	35.06	3.19E-167	*ydjN*	Putative transporter	10.68	4.35E-124
hxuC	Haem-hemopexin utilization protein C	30.77	0	*metQ*	DL-methionine transporter, periplasmic binding protein MetQ	9.13	6.30E-133
hitA	Iron(III) ABC transporter periplasmic-binding protein	30.6	2.22E-216	*metI*	DL-methionine transporter, permease protein MetI	7.52	7.10E-133
ompU1	Putative outer membrane protein OmpU1	23.72	3.75E-171	*potD*	Spermidine/putrescine ABC transporter, periplasmic-binding protein	7.09	1.01E-43
hxuB	Haem-hemopexin utilization protein B	23.46	5.21E-193	*metN*	DL-methionine transporter, ATP binding protein MetN	7.01	4.59E-77
tbp2	Transferrin-binding protein 2	17.21	1.00E-106	*artP*	Arginine ABC transporter, ATP-binding protein ArtP	6.18	3.40E-07
tbp1	Transferrin-binding protein 1	15.68	3.64E-136	*tcyA*	L-cystine ABC transporter, periplasmic-binding protein TcyA	5.53	8.93E-92
hxuA	Haem-hemopexin utilization protein A	15.01	9.48E-168	*argH*	Argininosuccinate lyase	5.24	2.00E-08
copZ3_1	Copper chaperone protein	10.93	2.38E-120	*fdnG_1*	Formate dehydrogenase-N, major subunit	5.13	1.52E-46
hitB	Iron(III) ABC transporter permease protein	10.83	6.39E-99	*fdnI*	Formate dehydrogenase-N, cytochrome B556(Fdn) gamma subunit, nitrate-inducible	4.66	4.66E-34
copZ3_2	Copper chaperone protein	10.68	1.02E-113	*fdnH*	Formate dehydrogenase-N, Fe-S beta subunit, nitrate-inducible	4.58	8.05E-53
-	Putative ABC transporter, periplasmic-binding protein	7.92	2.46E-60	*nrfB*	Nitrite reductase complex, periplasmic pentaheme cytochrome subunit	4.52	3.60E-15
yfeA	Fe/Mn/Zn ABC transporter periplasmic-binding protein	6.32	4.89E-79	*ilvC*	Ketol-acid reductoisomerase	4.49	4.55E-16
yfeB	Fe/Mn/Zn ABC transporter ATP-binding protein	5.62	1.35E-92	*hbpA*	Dipeptide/Haem ABC transport system, periplasmic binding protein	4.49	1.04E-38
yfeC	Fe/Mn/Zn ABC transporter permease protein	5.47	8.53E-48	*fdnG_2*	Formate dehydrogenase-N, major subunit	4.37	1.51E-43
hemR	Putative TonB-dependent haem receptor	4.77	2.87E-46	*nrfA*	Nitrite reductase complex, periplasmic cytochrome C552 subunit	4.09	3.36E-05
-	Putative methyltransferase	4.59	1.28E-28	*fdhE*	Formate dehydrogenase formation protein FdhE	4.02	5.62E-17
tnaA	Tryptophanase	4.57	1.42E-54	*tgt*	tRNA-guanine transglycosylase	3.96	3.11E-24
hitC	Iron(III) ABC transporter ATP-binding protein	4.3	6.16E-25	*dat*	L-2,4-diaminobutyrate:2-ketoglutarate 4-aminotransferase aminotransferase	3.94	2.09E-14
yfeD	Fe/Mn/Zn ABC transporter permease protein	4.24	6.47E-40	*artI*	Arginine ABC transporter, periplasmic-binding protein ArtI	3.73	4.84E-13
copA	Copper-transporting ATPase	4.04	1.86E-29	*nrfC*	Nitrite reductase complex, Fe-S subunit NrfC	3.7	9.76E-14
tehB	Putative tellurite resistance protein B	3.95	3.05E-63	*merT*	Putative heavy metal transport protein	3.67	9.12E-18
htrA	Periplasmic serine protease HtrA	3.74	1.26E-58	*hslO*	Hsp33-like chaperonin	3.66	4.59E-16
-	Hypothetical protein	3.21	1.46E-07	*ilvD*	Dihydroxyacid dehydratase	3.64	3.62E-39
pqqL	Putative Zn-dependent protease	2.87	5.67E-14	*deaD*	ATP-dependent RNA helicase DeaD	3.49	4.45E-19
-	Hypothetical protein	2.87	1.31E-31	*ilvA*	Threonine deaminase	3.48	1.38E-27
clpB	ATP-dependent Clp protease ATPase subunit	2.77	3.48E-15	*artQ*	Arginine ABC transporter, permease protein ArtQ	3.48	5.74E-15
msrAB	Peptide methionine sulfoxide reductase	2.75	9.05E-07	*artM*	Arginine ABC transporter, permease protein ArtM	3.45	7.07E-18
-	tRNA-Leu(caa)	2.65	0.000109	*pyrG*	CTP synthetase	3.41	1.20E-10
acrR	Putative transcriptional regulator AcrR	2.59	5.39E-08	*torY*	Trimethylamine N-oxide reductase system III,	3.36	2.64E-08

Gene	Product	Fold change	Adjusted p-value
Up-regulated genes			
queA	S-adenosylmethionine:tRNA ribosyltransferase-isomerase (queuosine biosynthesis protein)	2.57	6.08E-14
mdaB	NADPH quinone reductase	2.53	8.42E-07
tnaB	Tryptophan permease	2.5	1.73E-07
dsbD_1	Thiol-disulfide interchange protein DsbD	2.48	0.000361234
dnaK	Molecular chaperone DnaK (Hsp70)	2.46	3.33E-08
exbB	Biopolymer transport protein ExbB	2.41	4.75E-21
hgpC	Hemoglobin and hemoglobin-haptoglobin binding protein C	2.41	5.17E-12
hslU	Protease HslVU, ATPase subunit	2.39	4.60E-07
-	16S ribosomal RNA	2.38	0.002201796
lon	ATP-dependent protease La	2.32	7.21E-16
lptC	Lipooligosaccharide transporter, accessory protein LptC	2.31	4.45E-17
exbD	Biopolymer transport protein ExbD	2.27	3.41E-13
hslV	Protease HslVU, peptidase subunit	2.24	2.82E-05
yfiA	Ribosome binding protein Y	2.21	5.70E-09
groEL	GroEL, chaperone Hsp60	2.17	3.09E-13
yjgA	x96 protein	2.13	1.83E-11
dam_2	Putative dam methylase	2.13	0.006054781
acrA	Multidrug efflux system protein AcrA	2.08	3.44E-08
dprA	DNA processing chain A	2.08	0.000469528
pckA	Phosphoenolpyruvate carboxykinase	2.07	6.03E-05
-	Putative TRAP-type transport system, periplasmic component	2.07	3.89E-05
afuA	Ferric transport system AfuABC; periplasmic-binding protein component	2.06	5.54E-12
groES	GroES, chaperone Hsp10	2.05	1.97E-06
tonB	TonB protein	2.01	7.60E-10
ribC	Riboflavin synthase, subunit alpha	2.01	6.73E-10

Gene	Product	Fold change	Adjusted p-value
	cytochrome c-type subunit		
Down-regulated genes			
rpL1	50S ribosomal subunit protein L1	3.28	1.71E-24
rpL11	50S ribosomal subunit protein L11	3.28	3.85E-21
nrfD	Nitrite reductase complex, transmembrane subunit NrfD	3.24	4.14E-11
glmS	Glucosamine-fructose-6-phosphate aminotransferase	3.16	7.14E-17
merP	Putative heavy metal chaperone protein	3.13	3.94E-16
-	Hypothetical protein	3.12	9.76E-19
-	Transglutaminase-like superfamily protein	3.11	1.38E-42
-	Hypothetical protein	3.06	5.00E-05
oppF	Oligopeptide ABC transporter, ATP-binding protein OppF	2.96	1.28E-15
-	tRNA-Ile(gat)	2.84	3.33E-15
ahpC	Peroxiredoxin	2.77	5.62E-14
aroA	3-phosphoshikimate-1-carboxyvinyltransferase	2.75	7.28E-11
rpL7	50S ribosomal protein L7/L12	2.71	4.24E-12
oppD	Oligopeptide ABC transporter, ATP-binding protein OppD	2.71	1.01E-13
-	Selenium metabolism protein YedF	2.69	4.30E-09
ompP2	Outer membrane protein P2	2.68	5.57E-30
rpL29	50S ribosomal protein L29	2.67	3.08E-23
-	SH3 domain-containing protein	2.67	3.50E-07
frdC	Fumarate reductase, subunit C	2.66	6.84E-26
tsf	Elongation factor Ts	2.63	6.58E-20
rimP	30S ribosomal maturation protein RimP	2.61	4.57E-06
rpL16	50S ribosomal protein L16	2.61	2.10E-23
rpS17	30S ribosomal protein S17	2.6	2.33E-22
rpL9	50S ribosomal subunit protein L9	2.59	1.77E-20
greA	Transcription elongation factor GreA	2.59	1.30E-05
tusB	Putative tRNA 2-thiouridine synthesizing protein B	2.57	0.001640568
atpC	Membrane-bound ATP synthase, F1 sector, epsilon-subunit	2.55	2.18E-25
ispF	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	2.55	1.01E-07
rpS3	30S ribosomal protein S3	2.54	9.25E-25
-	tRNA-Ala(tgc)	2.54	2.69E-09
dmsA	Anaerobic dimethyl sulfoxide reductase, subunit A	2.52	0.020883402
rpS18	30S ribosomal subunit protein S18	2.49	7.88E-16
nusA	Transcription elongation factor NusA	2.48	6.11E-17
cca	tRNA nucleotidyltransferase/2'3'-cyclic	2.48	1.04E-09

	phosphodiesterase/2'nucleotidase and phosphatase		
Gene	**Product**	**Fold change**	**Adjusted p-value**
Down-regulated genes			
rpS19	30S ribosomal protein S19	2.45	5.03E-22
frdD	Fumarate reductase, subunit D	2.45	2.51E-12
rpL19	50S ribosomal subunit protein L19	2.44	6.91E-14
rpS2	30S ribosomal protein S2	2.44	1.52E-17
deoD	Purine-nucleoside phosphorylase (PNPase)	2.43	4.97E-15
frdB	Fumarate reductase, subunit B	2.43	4.94E-19
frdA	Fumarate reductase, subunit A	2.42	4.76E-12
-	Hypothetical protein	2.42	0.000109
rpL22	50S ribosomal protein L22	2.41	3.05E-20
tcyB	L-cystine ABC transporter, permease component TcyB	2.41	6.27E-24
atpD	Membrane-bound ATP synthase, F1 sector, beta-subunit	2.39	2.22E-28
rpL2	50S ribosomal protein L2	2.39	1.51E-15
trmU	tRNA (5-methylaminomethyl-2-thiouridylate)-methyltransferase	2.38	1.13E-08
rpL10	50S ribosomal protein L10	2.35	5.14E-11
-	Hypothetical protein	2.35	1.30E-11
yjcD	Putative membrane permease	2.34	2.50E-05
nlpI	Lipoprotein NlpI	2.34	1.47E-14
-	Putative membrane protein	2.31	0.008362234
rpL23	50S ribosomal protein L23	2.3	5.94E-16
fis	DNA architectural protein Fis	2.29	0.017436
ispD	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase	2.28	1.56E-06
rpL4	50S ribosomal protein L4	2.26	1.33E-12
rplQ	50S ribosomal protein L17	2.26	2.36E-18
tilS	tRNA(Ile)-lysidine synthetase	2.25	2.57E-16
tusC	tRNA 2-thiouridine synthesizing protein C	2.25	0.000962104
cstA	Carbon starvation protein A	2.24	4.17E-09
hel	Outer membrane protein P4	2.24	5.25E-22
argR	Arginine repressor	2.24	2.43E-07
atpH	Membrane-bound ATP synthase, F1 sector, delta-subunit	2.21	3.63E-14
tcyC	L-cystine ABC transporter, ATP-binding protein TcyC	2.21	8.12E-17
rnb	Ribonuclease II	2.21	4.14E-06
menH	2-succinyl-6-hydroxy-2, 4-cyclohexadiene-1-carboxylate synthase	2.2	2.28E-05
atpF	Membrane-bound ATP synthase, F0 sector, subunit B	2.2	8.47E-13
-	Hypothetical protein	2.2	1.83E-11
rpL3	50S ribosomal protein L3	2.2	5.43E-12
prtR_2	Pyocin repressor protein	2.19	9.06E-11
menD	2-succinyl-6-hydroxy-24-cyclohexadiene-1-carboxylate synthase/2-oxoglutarate decarboxylase	2.18	9.12E-18
tusD	tRNA 2-thiouridine synthesizing protein D	2.18	1.65E-08
argG	Argininosuccinate synthetase	2.18	2.21E-13
pepE	Peptidase E	2.16	5.82E-07
atpA	Membrane-bound ATP synthase, F1 sector, alpha-	2.15	3.66E-18

	subunit		
Gene	**Product**	**Fold change**	**Adjusted p-value**
Down-regulated genes			
rpS10	30S ribosomal protein S10	2.15	6.03E-10
atpG	Membrane-bound ATP synthase, F1 sector, gamma-subunit	2.14	6.99E-16
priB	Primosomal replication protein N	2.14	9.57E-12
-	Putative peptidase/hydrolase	2.12	4.36E-12
apbE	Lipoprotein ApbE	2.11	2.77E-14
prmB	50S subunit L3 protein glutamine methyltransferase	2.11	2.45E-10
-	YheO-like PAS domain protein	2.11	4.90E-11
yccS	Inner membrane protein YccS	2.1	1.40E-08
pfs	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase	2.1	3.15E-09
oppC	Oligopeptide ABC transporter, permease protein OppC	2.09	6.04E-09
arfA	Alternative ribosome-rescue factor A	2.09	0.009565168
dcuC	Putative C4-dicarboxylate transporter	2.08	2.89E-07
rps6	30S ribosomal subunit protein S6	2.07	1.86E-09
rpL20	50S ribosomal protein L20	2.07	5.81E-12
infB	Translation initiation factor 2	2.06	4.06E-19
trmD	tRNA (guanine-N1)-methyltransferase	2.05	1.18E-12
secD	General secretory pathway component SecD	2.05	0.000117544
mnmE	tRNA modification GTPase mnmE	2.04	4.43E-08
purU	Formyltetrahydrofolate deformylase	2.03	1.08E-05
-	Epimerase family protein	2.02	1.56E-06
pdxY	Pyridoxine kinase	2.02	5.38E-09
glpC_2	Anaerobic glycerol-3-phosphate dehydrogenase subunit C	2.02	7.24E-12
rpoA	DNA-directed RNA polymerase subunit alpha	2.02	3.94E-13
lig	NAD-dependent DNA ligase	2.02	8.70E-06
pitA	Putative phosphate permease	2.01	3.83E-05
-	Hypothetical protein p38_1	2.01	0.008883
-	Putative ABC transporter, fused permease and ATP-binding components	2	0.002748801
nqrC	Na+-transporting NADH:ubiquinone oxidoreductase, subunit NqrC	2	4.84E-13

Table S7: Differentially expressed genes in the Rd strain during nutritional stress.

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
dps	DNA protection during starvation protein	11.82	8.83E-30	*yjcD*	Putative permease YjcD	66.56	2.82E-212
sdaA	L-serine deaminase	10.55	3.14E-78	*purH*	Bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	59.93	2.78E-230
sdaC	Serine transporter	10.22	5.42E-47	*purD*	Phosphoribosylamine--glycine ligase	35.82	2.86E-240
gntP_1	Gluconate permease	7.84	7.33E-39	*purM*	Phosphoribosylaminoimidazole synthetase	35.77	7.29E-161
garK	Glycerate 2-kinase	7.54	2.42E-29	*purE*	Phosphoribosylaminoimidazole carboxylase catalytic subunit	33.62	1.10E-124
-	Aldolase	7.21	4.79E-10	*purN*	Phosphoribosylglycinamide formyltransferase	22.74	1.24E-233
dprA	DNA processing chain A	7.16	3.28E-16	*purK*	Phosphoribosylaminoimidazole carboxylase ATPase subunit	19.39	2.90E-41
-	Hypothetical protein	6.85	6.57E-10	*mtr*	Tryptophan-specific transport protein	17.58	5.16E-66
-	3-hydroxyisobutyrate dehydrogenase	6.57	2.87E-09	*lctP*	L-lactate permease	11.89	3.28E-34
ygbM	Putative hydroxypyruvate isomerase YgbM	6.51	7.33E-09	-	Short chain dehydrogenase/reductase	10.51	2.21E-23
arcB	Ornithine carbamoyltransferase	6.38	3.91E-09	-	Transporter protein	10.21	2.01E-24
arcC	Carbamate kinase	6.16	4.96E-08	*hemH_2*	Phosphoribosylaminoimidazole-succinocarboxamide synthase	9.69	1.41E-57
-	Putative epimerase/dehydratase	6	1.18E-09	*cvpA*	Colicin V production protein	8.82	8.54E-87
gntP_3	Gluconate permease	5.1	2.25E-09	*trpE*	Anthranilate synthase component I	8.51	7.96E-24
hitC	Iron(III) ABC transporter ATP-binding protein	4.84	2.50E-12	*ydeM*	Hypothetical protein	8.35	2.27E-10
rnb	Exoribonuclease II	4.43	7.42E-11	*purF*	Amidophosphoribosyltransferase	7.78	1.68E-58
-	Putative cyclase	4.42	4.01E-07	*purL_1*	Phosphoribosylformylglycinamidine synthase	6.86	1.35E-67
hitA	Iron-utilization periplasmic protein hFbpA	4.19	9.22E-11	*leuB*	3-isopropylmalate dehydrogenase	6.82	7.56E-17
-	Formate acetyltransferase	4.13	9.16E-11	-	Haloacid dehalogenase-like protein	6.51	1.04E-18
-	Autonomous glycyl radical cofactor GrcA	4.05	3.09E-08	*leuC*	Isopropylmalate isomerase large subunit	6.19	1.84E-13
-	Nickel uptake substrate-specific transmembrane region	3.89	4.21E-18	*leuD*	Isopropylmalate isomerase small subunit	6.19	1.41E-10
cydD_2	ABC transporter ATP-binding protein	3.87	1.91E-06	*ilvC*	Ketol-acid reductoisomerase	5.78	0.000356275
-	ABC transporter ATP-binding protein	3.84	3.03E-06	*glpX*	Fructose 1,6-bisphosphatase II	5.24	5.43E-66
hitB	Iron(III) ABC transporter permease	3.66	2.21E-05	*ccrB*	Camphor resistance protein CrcB	5.22	2.15E-26
-	Twin-arginine leader-binding protein DmsD	3.61	0.001860645	-	Hypothetical protein	5.19	3.78E-30
-	Type IV pilin secretion protein	3.53	1.77E-07	*leuA*	2-isopropylmalate synthase	5.09	7.57E-14
comD	Competence protein D	3.39	6.30E-05	*chuR*	Anaerobic sulfatase-maturating enzyme	5.09	7.06E-06
fabA	3-hydroxydecanoyl-ACP dehydratase	3.39	1.11E-15	-	Hypothetical protein	5.05	1.61E-12
-	Lipooligosaccharide biosynthesis protein	3.34	6.62E-06	*fdnG*	Formate dehydrogenase, nitrate-inducible, major subunit precursor	4.95	1.65E-10
afuB	Ferric transport system permease-like protein	3.31	5.13E-14	*galR*	LacI family transcriptional repressor	4.95	3.70E-11
fbpC	Ferric transporter ATP-	3.29	2.22E-12	*uraA*	Uracil permease	4.65	1.65E-08

	binding protein						
Gene	**Product**	**Fold change**	**Adjusted p-value**	**Gene**	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes				**Down-regulated genes**			
-	Protein transport protein	3.29	3.29E-07	*fdhE*	Formate dehydrogenase accessory protein FdhE	4.52	6.26E-08
gdhA	Glutamate dehydrogenase	3.28	0.000226623	*dmsA_1*	Anaerobic dimethyl sulfoxide reductase subunit A	4.48	3.56E-09
glgB	Glycogen branching protein	3.12	9.55E-09	*atzC*	N-isopropylammelide isopropyl amidohydrolase	4.14	7.69E-20
comC	Competence protein C	3.11	0.000223426	*fdxH*	Formate dehydrogenase subunit beta	4.05	4.70E-11
asnA	Asparagine synthetase AsnA	3.02	0.0002759	*artQ*	Arginine transporter permease subunit ArtQ	4.03	2.73E-07
-	Hydrolase	2.99	7.58E-05	-	2,3-diketo-L-gulonate reductase	4.01	1.23E-08
hopD	Type 4 prepilin-like protein specific leader peptidase	2.95	6.26E-08	*artM*	Arginine transporter permease subunit ArtM	3.66	1.18E-07
-	Hydroxyethylthiazole kinase	2.92	5.71E-19	*ndh*	NADH dehydrogenase	3.63	2.91E-16
ccmD	Haem exporter protein D	2.89	0.000298756	*rsgA_1*	Ferritin like protein 1	3.63	0.000119608
icc	Cyclic 3',5'-adenosine monophosphate phosphodiesterase	2.86	1.52E-06	*fdxI*	Formate dehydrogenase subunit gamma	3.62	2.14E-12
cpdB	Bifunctional 2',3'-cyclic nucleotide 2'-phosphodiesterase/3'-nucleotidase	2.85	0.000173128	*rsgA_2*	Ferritin	3.59	0.000151272
-	Sulfur relay protein TusC	2.81	6.77E-06	*arcA*	Two-component response regulator	3.53	2.69E-24
-	TRAP transporter, DctM subunit	2.8	0.000424651	*arfA*	Alternative ribosome-rescue factor A	3.37	7.08E-10
-	Sulfur transfer complex subunit TusB	2.79	1.12E-05	*kipI*	Sporulation inhibitor KipI	3.35	7.67E-16
comB	Competence protein B	2.77	0.000578764	*artI*	Arginine ABC transporter substrate-binding protein	3.31	5.87E-09
comM	Competence protein	2.73	1.17E-05	-	DNA polymerase V subunit UmuD	3.15	2.89E-12
frr	Ribosome recycling factor	2.71	3.33E-07	*trpG_2*	Anthranilate synthase component II	2.95	2.43E-37
thiD	Phosphomethylpyrimidine kinase	2.7	2.05E-12	*argG*	Argininosuccinate synthase	2.91	2.76E-13
cydD_1	ABC transporter ATP-binding protein	2.69	2.20E-05	*artP*	Arginine transporter ATP-binding protein	2.85	9.12E-07
-	Electron transport complex protein RnfC	2.69	5.24E-09	*trpB*	Tryptophan synthase subunit beta	2.82	2.76E-13
ccmE	Cytochrome c-type biogenesis protein CcmE	2.66	0.00473122	*cmk_1*	Cytidylate kinase	2.78	4.51E-13
-	Prepilin peptidase-dependent protein D	2.65	0.000563633	*dnaJ*	Chaperone protein DnaJ	2.76	0.00368254
-	Glycine radical enzyme, YjjI family	2.65	5.55E-13	*glyA*	Serine hydroxymethyltransferase	2.75	2.56E-11
-	1-deoxy-D-xylulose 5-phosphate reductoisomerase	2.6	0.000108811	-	Zinc-type alcohol dehydrogenase	2.74	0.00037308
glgC	Glucose-1-phosphate adenylyltransferase	2.59	5.22E-06	-	DNA replication initiation factor	2.66	0.000489336
glpK	Flycerol kinase	2.58	2.77E-25	*cmk_2*	Cytidylate kinase	2.61	2.19E-05
znuA	High-affinity zinc transporter substrate-binding protein	2.55	2.24E-09	*ygiX*	Transcriptional regulatory protein	2.61	7.77E-11
-	Sulfur transfer complex subunit TusD	2.55	1.91E-06	-	Hypothetical protein	2.61	7.49E-07
napG	Quinol dehydrogenase periplasmic subunit	2.53	3.91E-12	-	Esterase	2.56	0.006881365
nudF	ADP-ribose pyrophosphatase	2.51	5.74E-06	-	Transcriptional regulator	2.56	4.78E-19
dmsA_2	Anaerobic dimethyl sulfoxide reductase subunit A	2.51	1.13E-10	-	Transcriptional regulator	2.56	2.89E-15
napH	Quinol dehydrogenase membrane subunit	2.5	4.19E-14	*infB*	Translation initiation factor IF-2	2.53	0.000109576
glpF_1	Glycerol uptake facilitator protein	2.5	1.13E-18	*adhC*	Alcohol dehydrogenase class III	2.5	0.00756002
hisH_2	Histidinol-phosphate aminotransferase	2.5	2.93E-05	-	YGGT family protein	2.47	1.57E-08
yhhQ	Inner membrane protein YhhQ	2.48	7.84E-24	*pepT*	Peptidase T	2.46	6.68E-05
thiE	Thiamine-phosphate pyrophosphorylase	2.47	1.51E-09	-	Putative membrane protein	2.44	1.51E-09
comE	Competence protein E	2.46	5.67E-07	*yiaO_1*	Extracytoplasmic solute receptor protein YiaO	2.43	0.003357859

Gene	Product	Fold change	Adjusted p-value	Gene	Product	Fold change	Adjusted p-value
Up-regulated genes				**Down-regulated genes**			
radC	DNA repair protein RadC	2.45	2.13E-05	*ilvH*	Acetolactate synthase 3 regulatory subunit	2.4	7.01E-05
pflA_2	Pyruvate formate lyase-activating enzyme 1	2.44	1.41E-09	*kipA*	KipI antagonist	2.39	0.006455781
glgX	Glycogen operon protein	2.44	7.41E-05	*trpR*	Trp operon repressor	2.38	2.52E-07
dmsB	Anaerobic dimethyl sulfoxide reductase subunit B	2.43	0.014864629	*dksA*	*dnaK* suppressor protein	2.36	7.19E-29
-	Electron transport complex RsxE subunit	2.42	9.79E-05	*yiaM_2*	2,3-diketo-L-gulonate TRAP transporter small permease protein YiaM	2.36	0.007970529
nrfD	Nitrite reductase transmembrane protein	2.41	0.013469307	-	Transporter	2.35	0.002653115
ccmF	Cytochrome C-type biogenesis protein	2.41	0.001555007	*yccA*	Modulator of FtsH protease YccA	2.34	2.21E-27
oppB	Oligopeptide transporter permease	2.41	1.26E-06	*potD_2*	Spermidine/putrescine ABC transporter substrate-binding protein	2.34	7.28E-08
rebM	Demethylrebeccamycin-D-glucose O-methyltransferase	2.41	0.000418649	*rpsA*	30S ribosomal protein S1	2.32	8.50E-14
napB	Nitrate reductase	2.4	2.63E-16	*cdd*	Cytidine deaminase	2.31	6.44E-11
yohK	Inner membrane protein YohK	2.39	1.27E-08	*lysC*	Lysine-sensitive aspartokinase 3	2.26	4.22E-18
-	Epimerase family protein	2.38	9.25E-19	*lldD*	L-lactate dehydrogenase	2.25	1.96E-13
-	Peroxiredoxin hybrid Prx5	2.37	0.000773221	*galT*	Galactose-1-phosphate uridylyltransferase	2.24	4.36E-05
tgt	Queuine tRNA-ribosyltransferase	2.36	1.21E-09	*uxuR*	Uxu operon regulator	2.18	3.83E-11
-	Selenocysteine lyase	2.36	5.02E-10	*ilvI*	Acetolactate synthase 3 catalytic subunit	2.18	0.000653146
yecO	S-adenosyl-L-methionine-dependent methyltransferase	2.34	1.94E-18	*rec2*	Recombination protein	2.13	0.000166948
comA	Competence protein A	2.34	0.001503	*hslO*	Hsp33-like chaperonin	2.1	0.015907
glgA	Glycogen synthase	2.32	0.000830233	*nhaA*	pH-dependent sodium/proton antiporter	2.09	7.30E-10
-	Hypothetical protein	2.32	0.000860	*yecK*	Cytochrome C-like protein	2.09	0.010197
menC	O-succinylbenzoate synthase	2.31	6.00E-09	*lexA*	LexA repressor	2.09	3.55E-06
-	YheO-like PAS domain protein	2.31	0.001038293	*ygiY*	Sensor protein QseC	2.09	0.000103423
afuA	Ferric ABC transporter protein	2.29	1.69E-17	*yiaM_1*	2,3-diketo-L-gulonate TRAP transporter small permease protein YiaM	2.08	0.048416645
-	hypothetical protein	2.29	1.17E-08	*tsf*	Elongation factor Ts	2.07	3.19E-13
glpA	sn-glycerol-3-phosphate dehydrogenase subunit A	2.27	1.72E-05	-	Di- and tricarboxylate transporter	2.04	0.000678349
dmsC	Anaerobic dimethyl sulfoxide reductase subunit C	2.27	0.039164522	-	Transposase	2.03	1.29E-07
-	ABC transporter permease	2.25	0.001678	-	RarD protein	2.02	0.00013
napC	Cytochrome C-type protein	2.22	4.79E-20	-	ABC transporter ATP-binding protein	2.02	0.000419663
ccmC	Haem exporter protein C	2.22	0.012288	-	Hypothetical protein	2.01	0.000426
-	Fimbrial biogenesis and twitching motility protein	2.21	2.69E-05				
-	Serine/threonine transporter SstT	2.21	0.003067276				
-	Electron transport complex protein RnfB	2.19	5.71E-19				
comF	Competence protein F	2.18	0.000276				
-	Hemoglobin-binding protein	2.17	0.014907				
dacB	D-alanyl-D-alanine carboxypeptidase/endopeptidase	2.17	2.00E-09				
-	Hemoglobin-binding protein	2.16	0.045916				
ygdK	Putative SufE-like protein YgdK	2.15	2.35E-05				
-	TonB	2.15	6.68E-08				
nrfC	Nitrite reductase Fe-S protein	2.15	0.01587				
-	YcgL domain protein	2.15	8.39E-06				
malQ	4-alpha-glucanotransferase	2.14	5.28E-20				
-	dsDNA-mimic protein	2.14	5.04E-14				
-	Selenium metabolism protein YedF	2.14	1.62E-05				
udp	Uridine phosphorylase	2.12	6.86E-05				
aspC	Aromatic amino acid aminotransferase	2.12	0.033669452				
glpB	Anaerobic glycerol-3-	2.11	0.009098				

	phosphate dehydrogenase subunit B		
Gene	**Product**	**Fold change**	**Adjusted p-value**
Up-regulated genes			
ccmH_1	Cytochrome C-type biogenesis protein	2.11	1.43E-08
-	Long chain fatty acid CoA ligase	2.1	0.034807476
-	Amino acid carrier protein	2.1	0.001785
deoD	Purine nucleoside phosphorylase	2.1	3.76E-12
-	Hypothetical protein	2.1	1.14E-06
glpC_1	sn-glycerol-3-phosphate dehydrogenase subunit C	2.09	0.01752299
-	Hemoglobin-binding protein	2.09	2.26E-08
kdsB	3-deoxy-manno-octulosonate cytidylyltransferase	2.08	5.72E-13
glpF_2	Glycerol uptake facilitator protein	2.07	0.013920544
oppC	Oligopeptide ABC transporter permease	2.06	1.29E-07
-	Aminodeoxychorismate lyase	2.06	1.36E-07
rnfG	Electron transport complex protein RnfG	2.04	0.000434589
mnmC	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis bifunctional protein MnmC	2.04	2.34E-06
-	Hypothetical protein	2.04	5.18E-06
-	Hypothetical protein	2.04	2.56E-06
nrdD	Anaerobic ribonucleoside triphosphate reductase	2.03	0.002097318
-	Amino-acid ABC transporter ATP-binding protein	2.03	0.007643933
-	Integrase/recombinase	2.03	5.87E-08
pflA_1	Pyruvate formate lyase-activating enzyme 1	2.02	3.01E-07
oppD	Oligopeptide transporter ATP-binding protein	2.02	6.29E-12
-	tRNA-Leu(gag)	2.02	1.95E-05
moeB_2	Molybdopterin biosynthesis protein MoeB	2.01	5.22E-06

Table S8: Differentially expressed genes common to three or more conditions in Rd.

Gene	Product
Up-regulated genes	
All four conditions	
gdhA	Glutamate dehydrogenase
Stationary phase, oxidative stress, nutritional stress	
-	Peroxiredoxin hybrid Prx5
dps	DNA protection during starvation protein
Stationary phase, iron starvation, nutritional stress	
-	Amino acid carrier protein
comD	Competence protein D
comB	Competence protein B
Down-regulated genes	
All four conditions	
yecK	Cytochrome C-like protein
mtr	Tryptophan-specific transport protein
hslO	Hsp33-like chaperonin
trpB	Tryptophan synthase subunit beta
Stationary phase, oxidative stress, iron starvation	
cydD_1	ABC transporter ATP-binding protein
cydD_2	ABC transporter ATP-binding protein
-	ABC transporter ATP-binding protein
dmsA_3	Anaerobic dimethyl sulfoxide reductase subunit A
nrfB	Cytochrome c nitrite reductase pentaheme subunit
nrfA	Cytochrome c552
Stationary phase, iron starvation, nutritional stress	
ilvC	Ketol-acid reductoisomerase
tsf	Elongation factor Ts
infB	Translation initiation factor IF-2
potD_2	Spermidine/putrescine ABC transporter substrate-binding protein
Stationary phase, oxidative stress, nutritional stress	
arcA	Two-component response regulator
artM	Arginine transporter permease subunit ArtM
artQ	Arginine transporter permease subunit ArtQ
artI	Arginine ABC transporter substrate-binding protein
artP	Arginine transporter ATP-binding protein
pepT	Peptidase T
trpE	Anthranilate synthase component I
trpG_2	Anthranilate synthase component II
-	Short chain dehydrogenase/reductase
argG	Argininosuccinate synthase

Table S9: Differentially expressed genes common to all three conditions in R2866.

Gene	Product
Up-regulated genes	
groEL	GroEL, chaperone Hsp60
groES	GroES, chaperone Hsp10
hxuA	Haem-hemopexin utilization protein A
hxuB	Haem-hemopexin utilization protein B
hxuC	Haem-hemopexin utilization protein C
hemR	Putative TonB-dependent haem receptor
hitA	Iron(III) ABC transporter periplasmic-binding protein
dnaK	Molecular chaperone DnaK (Hsp70)
msrAB	Peptide methionine sulfoxide reductase
ompU1	Putative outer membrane protein OmpU1
tbp2	Transferrin-binding protein 2
tbp1	Transferrin-binding protein 1
Down-regulated genes	
rpS18	30S ribosomal subunit protein S18
tgt	tRNA-guanine transglycosylase
ahpC	Peroxiredoxin
potD	Spermidine/putrescine ABC transporter, periplasmic-binding protein

Appendix C: The toRNAdo script

```
import numpy
from copy import deepcopy
import os
from sys import argv

# filters a dictionary by length
def filter_by_length(unfiltered_list):
    copy1 = deepcopy(unfiltered_list)
    for nuc1 in copy1:
        count1 = 1
        position1 = nuc1 + 1
        while position1 in copy1:
            count1 += 1
            position1 += 1
        else:
            if count1 < 50 and nuc1 - 1 not in unfiltered_list:
                del unfiltered_list[nuc1]

# function to create text files
def write_to_file(sequences, filename):
    out_file = open(output_folder + filename + '.wig','w')
    list2 = sorted(sequences.keys())
    lines1 = ["%s\t%f" % (g, sequences[g]) for g in list2]
    lines1.insert(0, "variableStep\tchrom=Rd")
    out_file.write('\n'.join(lines1))
    out_file.write('\n')
    out_file.close()

# function to find ncRNAs to the "right" of UTR, and reclassify it
def find_RNA_right(UTR_dict):
    temp_dict = {}
    another_temp_dict = {}
    copy_UTR = deepcopy(UTR_dict)
    for nuc in copy_UTR:
        if nuc - 1 not in copy_UTR:
            while nuc + 1 in copy_UTR and copy_UTR[nuc + 1] <= copy_UTR[nuc]:
                nuc += 1
            else:
                if nuc + 1 not in copy_UTR:
                    None
                else:
                    temp_thresh = copy_UTR[nuc]
                    temp_dict[nuc] = copy_UTR[nuc]
                    while nuc + 1 in copy_UTR:
                        temp_dict[nuc + 1] = copy_UTR[nuc + 1]
                        nuc += 1
                    else:
                        if any((x / 5) > temp_thresh for x in temp_dict.values()):
                            for nuc1 in temp_dict:
                                if UTR_dict == UTR5_minus:
                                    RNA_minus[nuc1] = temp_dict[nuc1]
                                    del UTR_dict[nuc1]
                                if UTR_dict == UTR3_plus:
                                    RNA_plus[nuc1] = temp_dict[nuc1]
                                    del UTR_dict[nuc1]
                        else:
                            min_key = min(temp_dict, key=lambda k: temp_dict[k])
                            min_thresh = temp_dict[min_key]
                            pos = min_key + 1
                            another_temp_dict[min_key] = temp_dict[min_key]
                            while pos in temp_dict:
                                another_temp_dict[pos] = temp_dict[pos]
                                pos += 1
                            else:
                                if any((x1 / 5) > min_thresh for x1 in another_temp_dict.values()):
                                    for nuc2 in another_temp_dict:
                                        if UTR_dict == UTR5_minus:
                                            RNA_minus[nuc2] = another_temp_dict[nuc2]
                                            del UTR_dict[nuc2]
                                        if UTR_dict == UTR3_plus:
                                            RNA_plus[nuc2] = another_temp_dict[nuc2]
                                            del UTR_dict[nuc2]
                                another_temp_dict.clear()
                    temp_dict.clear()

# function to find ncRNA to the "left" of UTR, and reclassify it
def find_RNA_left(UTR_dict1):
    temp_dict = {}
    another_temp_dict = {}
```

```
    copy_UTR = deepcopy(UTR_dict1)
    for nuc in copy_UTR:
        if nuc + 1 not in copy_UTR:
            while nuc - 1 in copy_UTR and copy_UTR[nuc - 1] <= copy_UTR[nuc]:
                nuc -= 1
            else:
                if nuc - 1 not in copy_UTR:
                    None
                else:
                    temp_thresh = copy_UTR[nuc]
                    temp_dict[nuc] = copy_UTR[nuc]
                    while nuc - 1 in copy_UTR:
                        temp_dict[nuc - 1] = copy_UTR[nuc - 1]
                        nuc -= 1
                    else:
                        if any((x / 5) > temp_thresh for x in temp_dict.values()):
                            for nuc1 in temp_dict:
                                if UTR_dict1 == UTR3_minus:
                                    RNA_minus[nuc1] = temp_dict[nuc1]
                                    del UTR_dict1[nuc1]
                                if UTR_dict1 == UTR5_plus:
                                    RNA_plus[nuc1] = temp_dict[nuc1]
                                    del UTR_dict1[nuc1]
                        else:
                            min_key = min(temp_dict, key=lambda k: temp_dict[k])
                            min_thresh = temp_dict[min_key]
                            pos = min_key - 1
                            another_temp_dict[min_key] = temp_dict[min_key]
                            while pos in temp_dict:
                                another_temp_dict[pos] = temp_dict[pos]
                                pos -= 1
                            else:
                                if any((x1 / 5) > min_thresh for x1 in another_temp_dict.values()):
                                    for nuc2 in another_temp_dict:
                                        if UTR_dict1 == UTR3_minus:
                                            RNA_minus[nuc2] = another_temp_dict[nuc2]
                                            del UTR_dict1[nuc2]
                                        if UTR_dict1 == UTR5_plus:
                                            RNA_plus[nuc2] = another_temp_dict[nuc2]
                                            del UTR_dict1[nuc2]
                                another_temp_dict.clear()
                    temp_dict.clear()

# function to find ncRNA in the intergenic region, which has been originally classified as part of an operon
def find_RNA_operon(operon_dict):
    temp_dict = {}
    temp_dict_right = {}
    temp_dict_left = {}
    new_left = {}
    new_right = {}
    copy_operon = deepcopy(operon_dict)
    for nuc in copy_operon:
        if nuc - 1 not in copy_operon:
            while nuc + 1 in copy_operon:
                temp_dict[nuc] = copy_operon[nuc]
                nuc += 1
            else:
                temp_dict[nuc] = copy_operon[nuc]
                min_key = min(temp_dict, key=lambda k: temp_dict[k])
                min_thresh = temp_dict[min_key]
                if any((x / 5) > min_thresh for x in temp_dict.values()):
                    while min_key + 1 in temp_dict:
                        temp_dict_right[min_key + 1] = temp_dict[min_key + 1]
                        min_key += 1
                    else:
                        min_key = min(temp_dict, key=lambda k: temp_dict[k])
                        temp_dict_left[min_key] = temp_dict[min_key]
                        while min_key - 1 in temp_dict:
                            temp_dict_left[min_key - 1] = temp_dict[min_key - 1]
                            min_key -= 1
                        else:
                            if any((x1 / 5) > min_thresh for x1 in temp_dict_left.values()):
                                max_key_left = max(temp_dict_left, key=lambda k1: temp_dict_left[k1])
                                max_key_left_thresh = temp_dict_left[max_key_left]
                                while max_key_left - 1 in temp_dict_left:
                                    new_left[max_key_left - 1] = temp_dict_left[max_key_left - 1]
                                    max_key_left -= 1
                                else:
                                    if any((max_key_left_thresh / 5) > x2 for x2 in new_left.values()):
                                        min_key_left = min(new_left, key=lambda k2: new_left[k2])
                                        while min_key_left + 1 in temp_dict_left:
                                            if operon_dict == operon_plus:
                                                RNA_plus[min_key_left + 1] = temp_dict_left[min_key_left + 1]
                                                del operon_dict[min_key_left + 1]
                                            if operon_dict == operon_minus:
```

```
                                        RNA_minus[min_key_left + 1] = temp_dict_left[min_key_left + 1]
                                        del operon_dict[min_key_left + 1]
                                    min_key_left += 1
                            if any((x3 / 5) > min_thresh for x3 in temp_dict_right.values()):
                                max_key_right = max(temp_dict_right, key=lambda k3: temp_dict_right[k3])
                                max_key_right_thresh = temp_dict_right[max_key_right]
                                while max_key_right + 1 in temp_dict_right:
                                    new_right[max_key_right + 1] = temp_dict_right[max_key_right + 1]
                                    max_key_right += 1
                                else:
                                    if any((max_key_right_thresh / 5) > x4 for x4 in new_right.values()):
                                        min_key_right = min(new_right, key=lambda k4: new_right[k4])
                                        while min_key_right - 1 in temp_dict_right:
                                            if operon_dict == operon_plus:
                                                RNA_plus[min_key_right - 1] = temp_dict_right[min_key_right - 1]
                                                del operon_dict[min_key_right - 1]
                                            if operon_dict == operon_minus:
                                                RNA_minus[min_key_right - 1] = temp_dict_right[min_key_right - 1]
                                                del operon_dict[min_key_right - 1]
                                            min_key_right -= 1
                            new_left.clear()
                            new_right.clear()
                        temp_dict_right.clear()
                        temp_dict_left.clear()
                temp_dict.clear()

# function to reclassify former operon regions into UTRs and delete those regions from the operon dictionary
def operon_to_UTR(operon_dict1):
    temp_dict = {}
    operon_copy = deepcopy(operon_dict1)
    for nuc in operon_copy:
        if nuc - 1 not in operon_copy:
            if operon_dict1 == operon_plus:
                if nuc - 1 not in dict_strand:
                    UTR5_plus[nuc] = operon_copy[nuc]
                    del operon_dict1[nuc]
                    while nuc + 1 in operon_copy:
                        UTR5_plus[nuc + 1] = operon_copy[nuc + 1]
                        del operon_dict1[nuc + 1]
                        nuc += 1
                else:
                    while nuc + 1 in operon_copy:
                        nuc += 1
                    else:
                        if nuc + 1 not in dict_strand:
                            UTR3_plus[nuc] = operon_copy[nuc]
                            del operon_dict1[nuc]
                            while nuc - 1 in operon_copy:
                                UTR3_plus[nuc - 1] = operon_copy[nuc - 1]
                                del operon_dict1[nuc - 1]
                                nuc -= 1
            if operon_dict1 == operon_minus:
                if nuc - 1 not in dict_strand:
                    UTR3_minus[nuc] = operon_copy[nuc]
                    del operon_dict1[nuc]
                    while nuc + 1 in operon_copy:
                        UTR3_minus[nuc + 1] = operon_copy[nuc + 1]
                        del operon_dict1[nuc + 1]
                        nuc += 1
                else:
                    while nuc + 1 in operon_copy:
                        nuc += 1
                    else:
                        if nuc + 1 not in dict_strand:
                            UTR5_minus[nuc] = operon_copy[nuc]
                            del operon_dict1[nuc]
                            while nuc - 1 in operon_copy:
                                UTR5_minus[nuc - 1] = operon_copy[nuc - 1]
                                del operon_dict1[nuc - 1]
                                nuc -= 1


# function to correct intergenic RNA into antisense of border RNAs
def RNA_to_antisense_border(rna_dict):
    RNA_copy = deepcopy(rna_dict)
    temp_dict = {}
    anti_dict = {}
    for rna in RNA_copy:
        if rna - 1 not in RNA_copy:
            temp_dict[rna] = RNA_copy[rna]
            while rna + 1 in RNA_copy:
                temp_dict[rna + 1] = RNA_copy[rna + 1]
                rna += 1
            else:
                if any(x in dict_strand for x in temp_dict.keys()):
```

```
                    if any(y not in dict_strand for y in temp_dict.keys()):
                        for rna1 in temp_dict:
                            if rna_dict == RNA_plus:
                                border_plus[rna1] = temp_dict[rna1]
                                del RNA_plus[rna1]
                            if rna_dict == RNA_minus:
                                border_minus[rna1] = temp_dict[rna1]
                                del RNA_minus[rna1]
                    else:
                        for rna2 in temp_dict:
                            if rna_dict == RNA_plus:
                                antisense_of_minus[rna2] = temp_dict[rna2]
                                del RNA_plus[rna2]
                            if rna_dict == RNA_minus:
                                antisense_of_plus[rna2] = temp_dict[rna2]
                                del RNA_minus[rna2]
            temp_dict.clear()

# function to look at all ncRNA dictionaries and find RNAs with a "5x" expression peak
def find_peaks(RNAdict, cooldict):
    temp_dict = {}
    copy_RNAdict = deepcopy(RNAdict)
    for abc in copy_RNAdict:
        if abc - 1 not in copy_RNAdict:
            paul = abc
            temp_dict[paul] = copy_RNAdict[paul]
            while paul + 1 in copy_RNAdict:
                temp_dict[paul + 1] = copy_RNAdict[paul]
                paul += 1
            else:
                min_key = min(temp_dict, key=lambda k: temp_dict[k])
                min_thresh = temp_dict[min_key]
                if any((x / 5) > min_thresh for x in temp_dict.values()):
                    for abcd in temp_dict:
                        cooldict[abcd] = temp_dict[abcd]
                        del RNAdict[abcd]
            temp_dict.clear()

# file1 - nucleotide coverage for both strands; file2 - nucleotide coverage for minus strand; file 3 - nucleotide coverage for plus strand; file
4 - nucleotide coverage for all annotated genome features and their coordinates.
script, file1, file2, file3, file4, output_folder = argv

# open all files needed for analysis
data_both = open(file1).readlines()
data_minus = open(file2).readlines()
data_plus = open(file3).readlines()
data_genes = open(file4).readlines()

# empty lists that will have data appended to
realdata_minus = []
realdata_plus = []
realdata_genes = []
realdata_both = []
gene_ncoverage = []
both_ncoverage = []

# loops for appending into empty lists and changing lists of strings into lists of integers
for line_both in data_both:
    realdata_both.append(line_both.strip().split())
realdata_both = [[float(ab) for ab in bb] for bb in realdata_both]  # turns everything in a list of lists into integers
cov_both_list = [a[1] for a in realdata_both] # makes new list with just the coverage
sum_both = float(sum(cov_both_list)) # produces the sum of all coverage values, used for normalization later

for line_minus in data_minus:
    realdata_minus.append(line_minus.strip().split())
realdata_minus = [[float(ab) for ab in bb] for bb in realdata_minus] # turns everything in a list of lists into integers
for i in realdata_minus:                    # a loop to normalize each data based on the sum of coverage values calculated earlier
    i[1] = i[1] / sum_both * 10000000000

for line_plus in data_plus:
    realdata_plus.append(line_plus.strip().split())
realdata_plus = [[float(ab) for ab in bb] for bb in realdata_plus]
for e in realdata_plus:
    e[1] = e[1] / sum_both * 10000000000

for line_genes in data_genes:
    realdata_genes.append(line_genes.strip().split())

for value in realdata_genes:
    gene_ncoverage.append(value[3])   # makes a new list with just the gene coverage values. To calculate mean and standard deviation
below. Not really used here yet!!
gene_ncoverage = [float(b) for b in gene_ncoverage]

threshold = 100.0 # expression threshold
```

```
# modifies a list of gene coordinates, removing the nucleotide column and
# turning the start position in the first column into nucleotide coordinate
for f in realdata_genes:
    f[0] = int(f[0])
    f[2] = int(f[2])
    f[3] = int(f[3])
    f[0] = f[0] + f[2] - 1
    f.pop(2)

# Dictionaries!! Easier to work with here than with lists...
dict_minus = {}
dict_plus = {}
dict_both = {}
dict_strand = {}

# Turning lists into dictionaries
for line1 in realdata_minus:
    dict_minus[line1[0]] = line1[1]
for line2 in realdata_plus:
    dict_plus[line2[0]] = line2[1]
for line3 in realdata_both:
    dict_both[line3[0]] = line3[1]
for line4 in realdata_genes:
    dict_strand[line4[0]] = line4[1]

# empty dictionaries
RNA_minus = {}
RNA_plus = {}
UTR3_minus = {}
UTR5_minus = {}
UTR3_plus = {}
UTR5_plus = {}
antisense_of_minus = {}
antisense_of_plus = {}
operon_minus = {}
operon_plus = {}

for key1 in dict_both:
    if key1 in dict_strand:
        if dict_strand[key1] == "-":
            #UTR3_minus - puts all 3' UTR regions above the threshold on a minus strand into a new dictionary
            if key1 - 1 not in dict_strand and key1 - 1 in dict_both:
                while dict_minus[key1 - 1] > threshold and key1 - 1 not in dict_strand:
                    UTR3_minus[key1 - 1] = dict_minus[key1 - 1]
                    key1 -= 1
        elif dict_strand[key1] == "+":
            #UTR5_plus - puts all 5' UTR regions above the threshold on a plus strand into a new dictionary
            if key1 - 1 not in dict_strand and key1 - 1 in dict_both:
                while dict_plus[key1 - 1] > threshold and key1 - 1 not in dict_strand:
                    UTR5_plus[key1 - 1] = dict_plus[key1 - 1]
                    key1 -= 1

for key11 in dict_both:
    if key11 in dict_strand:
        if dict_strand[key11] == "-":
            #UTR5_minus - puts all 5' UTR regions above the threshold on a minus strand into a new dictionary
            if key11 + 1 not in dict_strand and key11 + 1 in dict_both:
                while dict_minus[key11 + 1] > threshold and key11 + 1 not in dict_strand:
                    UTR5_minus[key11 + 1] = dict_minus[key11 + 1]
                    key11 += 1
        elif dict_strand[key11] == "+":
            #UTR3_plus - puts all 3' UTR regions above the threshold on a plus strand into a new dictionary
            if key11 + 1 not in dict_strand and key11 + 1 in dict_both:
                while dict_plus[key11 + 1] > threshold and key11 + 1 not in dict_strand:
                    UTR3_plus[key11 + 1] = dict_plus[key11 + 1]
                    key11 += 1

# finds any RNA that is antisense to a coding region and is above a threshold
for key21 in dict_both:
    if key21 in dict_strand:
        if dict_strand[key21] == "-":
            #antisense_minus
            if dict_plus[key21] > threshold:
                antisense_of_minus[key21] = dict_plus[key21]
        elif dict_strand[key21] == "+":
            #antisense_plus
            if dict_minus[key21] > threshold:
                antisense_of_plus[key21] = dict_minus[key21]

# finds any RNA that is in an intergenic region and above the threshold
for key2 in dict_both:
    if key2 not in dict_strand:
        if key2 not in UTR3_plus and key2 not in UTR5_plus:
            if dict_plus[key2] > threshold:
                RNA_plus[key2] = dict_plus[key2]
```

```
        if key2 not in UTR3_minus and key2 not in UTR5_minus :
            if dict_minus[key2] > threshold:
                RNA_minus[key2] = dict_minus[key2]


# creates a copy of a dictionary that will be used below. This is done so that I can modify the real dictionary while looping through the
copy
copy_UTR3_minus = deepcopy(UTR3_minus)
# finds regions betwen genes that are likely to belong to an operon
for key3 in copy_UTR3_minus:
    if key3 in UTR5_minus:
        operon_minus[key3] = UTR3_minus[key3]
        del UTR3_minus[key3]
        del UTR5_minus[key3]

copy_UTR3_plus = deepcopy(UTR3_plus)
for key4 in copy_UTR3_plus:
    if key4 in UTR5_plus:
        operon_plus[key4] = UTR3_plus[key4]
        del UTR3_plus[key4]
        del UTR5_plus[key4]

# the loop below correctly classifies UTRs that have been misclassified as antisense RNA due to overlapping coding regions
cop1_antisense_minus = deepcopy(antisense_of_minus)
cop1_antisense_plus = deepcopy(antisense_of_plus)
for pete in dict_strand:
    if dict_strand[pete] == "+":
        if pete + 1 in cop1_antisense_minus:
            pos9 = pete + 1
            while pos9 in cop1_antisense_minus:
                UTR3_plus[pos9] = cop1_antisense_minus[pos9]
                del antisense_of_minus[pos9]
                pos9 += 1
            cop1_antisense_minus = deepcopy(antisense_of_minus)
        elif pete - 1 in cop1_antisense_minus:
            pos10 = pete - 1
            while pos10 in cop1_antisense_minus:
                UTR5_plus[pos10] = cop1_antisense_minus[pos10]
                del antisense_of_minus[pos10]
                pos10 -= 1
            cop1_antisense_minus = deepcopy(antisense_of_minus)
    elif dict_strand[pete] == "-":
        if pete + 1 in cop1_antisense_plus:
            pos11 = pete + 1
            while pos11 in cop1_antisense_plus:
                UTR5_minus[pos11] = cop1_antisense_plus[pos11]
                del antisense_of_plus[pos11]
                pos11 += 1
            cop1_antisense_plus = deepcopy(antisense_of_plus)
        elif pete - 1 in cop1_antisense_plus:
            pos12 = pete - 1
            while pos12 in cop1_antisense_plus:
                UTR3_minus[pos12] = cop1_antisense_plus[pos12]
                del antisense_of_plus[pos12]
                pos12 -= 1
            cop1_antisense_plus = deepcopy(antisense_of_plus)

# joins the UTR region with antisense region if they are part of the same UTR
temp_UTR5_minus = {}
copy_antisense_of_plus = deepcopy(antisense_of_plus)
for key51 in UTR5_minus:
    if key51 + 1 in copy_antisense_of_plus:
        while key51 + 1 in copy_antisense_of_plus:
            temp_UTR5_minus[key51 + 1] = antisense_of_plus[key51 + 1]
            del antisense_of_plus[key51 + 1]
            key51 += 1

# merges two dictionaries together: will update the first dictionary with the second one, if any keys match
UTR5_minus.update(temp_UTR5_minus)

temp_UTR3_plus = {}
copy_antisense_of_minus = deepcopy(antisense_of_minus)
for key53 in UTR3_plus:
    if key53 + 1 in copy_antisense_of_minus:
        while key53 + 1 in copy_antisense_of_minus:
            temp_UTR3_plus[key53 + 1] = antisense_of_minus[key53 + 1]
            del antisense_of_minus[key53 + 1]
            key53 += 1

UTR3_plus.update(temp_UTR3_plus)

temp_UTR5_plus = {}
copy1_antisense_of_minus = deepcopy(antisense_of_minus)
for key54 in UTR5_plus:
    if key54 - 1 in copy1_antisense_of_minus:
```

```
            while key54 - 1 in copy1_antisense_of_minus:
                temp_UTR5_plus[key54 - 1] = antisense_of_minus[key54 - 1]
                del antisense_of_minus[key54 - 1]
                key54 -= 1

UTR5_plus.update(temp_UTR5_plus)

temp_UTR3_minus = {}
copy1_antisense_of_plus = deepcopy(antisense_of_plus)
for key52 in UTR3_minus:
    if key52 - 1 in copy1_antisense_of_plus:
        while key52 - 1 in copy1_antisense_of_plus:
            temp_UTR3_minus[key52 - 1] = antisense_of_plus[key52 - 1]
            del antisense_of_plus[key52 - 1]
            key52 -= 1

UTR3_minus.update(temp_UTR3_minus)

# loops below merge any intergenic RNA with antisense RNA if they are part of the same transcript. Calls this new joined transcript either
border_plus or border_minus ("mixed" ncRNAs)
border_minus = {}
for key5 in antisense_of_plus:
    if key5 - 1 in RNA_minus:
        border_minus[key5] = antisense_of_plus[key5]
        pos1 = key5
        while pos1 - 1 in RNA_minus:
            border_minus[pos1 - 1] = RNA_minus[pos1 - 1]
            pos1 -= 1
        pos2 = key5
        while pos2 + 1 in antisense_of_plus:
            border_minus[pos2 + 1] = antisense_of_plus[pos2 + 1]
            pos2 += 1
    elif key5 + 1 in RNA_minus:
        border_minus[key5] = antisense_of_plus[key5]
        pos3 = key5
        while pos3 + 1 in RNA_minus:
            border_minus[pos3 + 1] = RNA_minus[pos3 + 1]
            pos3 += 1
        pos4 = key5
        while pos4 - 1 in antisense_of_plus:
            border_minus[pos4 - 1] = antisense_of_plus[pos4 - 1]
            pos4 -= 1

border_plus = {}
for key7 in antisense_of_minus:
    if key7 - 1 in RNA_plus:
        border_plus[key7] = antisense_of_minus[key7]
        pos5 = key7
        while pos5 - 1 in RNA_plus:
            border_plus[pos5 - 1] = RNA_plus[pos5 - 1]
            pos5 -= 1
        pos6 = key7
        while pos6 + 1 in antisense_of_plus:
            border_plus[pos6 + 1] = antisense_of_minus[pos6 + 1]
            pos6 += 1
    elif key7 + 1 in RNA_plus:
        border_plus[key7] = antisense_of_minus[key7]
        pos7 = key7
        while pos7 + 1 in RNA_plus:
            border_plus[pos7 + 1] = RNA_plus[pos7 + 1]
            pos7 += 1
        pos8 = key7
        while pos8 - 1 in antisense_of_minus:
            border_plus[pos8 - 1] = antisense_of_minus[pos8 - 1]
            pos8 -= 1

# these loops just delete keys from antisense and RNA dictionaries that match the keys from border_plus or border_minus
for key9 in border_minus:
    if key9 in antisense_of_plus:
        del antisense_of_plus[key9]
    if key9 in RNA_minus:
        del RNA_minus[key9]

for key10 in border_plus:
    if key10 in antisense_of_minus:
        del antisense_of_minus[key10]
    if key10 in RNA_plus:
        del RNA_plus[key10]

# the following two loops joins mixed RNAs with antisense RNA if they are a part of the same transcript
border_plus_temp = {}
cop_antisense_minus = deepcopy(antisense_of_minus)
for a1 in border_plus:
    if a1 + 1 in cop_antisense_minus:
        apos1 = a1
```

```
            while apos1 + 1 in cop_antisense_minus:
                border_plus_temp[apos1 + 1] = cop_antisense_minus[apos1 + 1]
                del antisense_of_minus[apos1 + 1]
                apos1 += 1
        elif a1 - 1 in cop_antisense_minus:
            apos2 = a1
            while apos2 - 1 in cop_antisense_minus:
                border_plus_temp[apos2 - 1] = cop_antisense_minus[apos2 - 1]
                del antisense_of_minus[apos2 - 1]
                apos2 -= 1

border_plus.update(border_plus_temp)

border_minus_temp = {}
cop_antisense_plus = deepcopy(antisense_of_plus)
for a2 in border_minus:
    if a2 + 1 in cop_antisense_plus:
        apos3 = a2
        while apos3 + 1 in cop_antisense_plus:
            border_minus_temp[apos3 + 1] = cop_antisense_plus[apos3 + 1]
            del antisense_of_plus[apos3 + 1]
            apos3 += 1
    elif a2 - 1 in cop_antisense_plus:
        apos4 = a2
        while apos4 - 1 in cop_antisense_plus:
            border_minus_temp[apos4 - 1] = cop_antisense_plus[apos4 - 1]
            del antisense_of_plus[apos4 - 1]
            apos4 -= 1

border_minus.update(border_minus_temp)

# here I call the functions to find ncRNAs that have been misclassified as operon or UTR
find_RNA_right(UTR5_minus)
find_RNA_right(UTR3_plus)
find_RNA_left(UTR5_plus)
find_RNA_left(UTR3_minus)
find_RNA_operon(operon_minus)
find_RNA_operon(operon_plus)
operon_to_UTR(operon_minus)
operon_to_UTR(operon_plus)

RNA_to_antisense_border(RNA_plus)
RNA_to_antisense_border(RNA_minus)

# here I call the function which filters dictionaries by length
filter_by_length(antisense_of_minus)
filter_by_length(antisense_of_plus)
filter_by_length(RNA_minus)
filter_by_length(RNA_plus)
filter_by_length(border_minus)
filter_by_length(border_plus)

proper_antisense_of_minus = {}
proper_antisense_of_plus = {}
proper_RNA_minus = {}
proper_RNA_plus = {}
proper_border_minus = {}
proper_border_plus = {}

find_peaks(antisense_of_minus, proper_antisense_of_minus)
find_peaks(antisense_of_plus, proper_antisense_of_plus)
find_peaks(RNA_minus, proper_RNA_minus)
find_peaks(RNA_plus, proper_RNA_plus)
find_peaks(border_minus, proper_border_minus)
find_peaks(border_plus, proper_border_plus)

write_to_file(antisense_of_minus, "antisense_of_minus")
write_to_file(antisense_of_plus, "antisense_of_plus")
write_to_file(RNA_minus, "intergenic_minus")
write_to_file(RNA_plus, "intergenic_plus")
write_to_file(operon_minus, "operon_minus")
write_to_file(operon_plus, "operon_plus")
write_to_file(border_minus, "mixed_minus")
write_to_file(border_plus, "mixed_plus")
write_to_file(UTR3_plus, "UTR3_plus")
write_to_file(UTR5_plus, "UTR5_plus")
write_to_file(UTR3_minus, "UTR3_minus")
write_to_file(UTR5_minus, "UTR5_minus")
write_to_file(proper_antisense_of_minus, "filtered_antisense_of_minus")
write_to_file(proper_antisense_of_plus, "filtered_antisense_of_plus")
write_to_file(proper_border_minus, "filtered_mixed_minus")
write_to_file(proper_border_plus, "filtered_mixed_plus")
write_to_file(proper_RNA_minus, "filtered_intergenic_minus")
write_to_file(proper_RNA_plus, "filtered_intergenic_plus")
```

Appendix D: Supplementary tables for Chapter 5

Table S10: Characteristics of all identified putative ncRNAs in Rd and R2866 genomes.

Name	Start position	Stop position	Length	Type	Name	Start position	Stop position	Length	Type
Rd					R2866				
Rd_001	30539	31309	771	mixed	R2866_001	9716	10442	727	mixed
Rd_002	37519	37954	436	antisense	R2866_002	97846	98356	511	mixed
Rd_003	73608	73816	208	mixed	R2866_003	113760	114721	962	mixed
Rd_004	109964	110715	752	mixed	R2866_004	132320	132529	210	intergenic
Rd_005	127630	128163	534	mixed	R2866_005	175384	175658	274	mixed
Rd_006	130258	131358	1101	mixed	R2866_006	186358	186567	210	intergenic
Rd_007	136228	138127	1900	mixed	R2866_007	186977	187800	824	mixed
Rd_008	154965	155378	414	intergenic	R2866_008	194341	194992	651	mixed
Rd_009	155135	155395	261	mixed	R2866_009	272052	272292	241	intergenic
Rd_010	156713	156995	283	mixed	R2866_010	289003	290293	1291	mixed
Rd_011	169023	169556	534	mixed	R2866_011	346335	346699	365	mixed
Rd_012	175352	175818	467	antisense	R2866_012	347170	347384	215	antisense
Rd_013	176336	176520	185	intergenic	R2866_013	357196	358018	823	mixed
Rd_014	180355	181025	671	antisense	R2866_014	357774	358115	342	intergenic
Rd_015	184195	184851	656	antisense	R2866_015	363380	363853	474	mixed
Rd_016	208435	209046	612	antisense	R2866_016	390735	391522	788	mixed
Rd_017	214643	214968	326	antisense	R2866_017	397982	398500	519	intergenic
Rd_018	227136	227665	530	intergenic	R2866_018	398271	398538	268	intergenic
Rd_019	227394	227616	223	intergenic	R2866_019	400501	401715	1215	mixed
Rd_020	243893	244749	857	mixed	R2866_020	406115	406693	578	mixed
Rd_021	249824	250087	264	mixed	R2866_021	408041	408417	377	mixed
Rd_022	256153	256897	745	mixed	R2866_022	419392	419860	469	antisense
Rd_023	271291	271628	338	mixed	R2866_023	429795	430126	332	mixed
Rd_024	278059	279280	1222	mixed	R2866_024	452241	452552	312	intergenic
Rd_025	278913	279873	961	mixed	R2866_025	452279	452786	508	intergenic
Rd_026	281235	281514	280	intergenic	R2866_026	459566	460100	535	mixed
Rd_027	281280	281477	198	intergenic	R2866_027	471758	472743	986	mixed
Rd_028	296327	296520	194	intergenic	R2866_028	474416	474668	253	mixed
Rd_029	305850	306227	378	mixed	R2866_029	474428	474748	321	intergenic
Rd_030	308744	309586	843	mixed	R2866_030	478478	479142	665	mixed
Rd_031	310894	311696	803	mixed	R2866_031	492230	494186	1957	mixed
Rd_032	312222	313247	1025	mixed	R2866_032	495129	495444	315	mixed
Rd_033	320793	321132	340	intergenic	R2866_033	498570	498891	322	intergenic
Rd_034	320889	321185	297	mixed	R2866_034	499358	500645	1288	mixed
Rd_035	330677	330927	251	antisense	R2866_035	502749	503057	309	mixed
Rd_036	331381	331544	164	antisense	R2866_036	505734	506225	492	antisense
Rd_037	339574	341083	1510	antisense	R2866_037	524761	526365	1605	mixed
Rd_038	354436	355486	1051	mixed	R2866_038	528692	529532	841	mixed
Rd_039	397157	398551	1395	mixed	R2866_039	542421	543162	742	mixed
Rd_040	400512	400738	227	intergenic	R2866_040	548285	548624	340	intergenic
Rd_041	427943	428118	176	intergenic	R2866_041	555815	556104	290	mixed
Rd_042	435624	436053	430	mixed	R2866_042	564343	565334	992	antisense
Rd_043	459360	459961	602	antisense	R2866_043	569053	569691	639	mixed
Rd_044	468285	469071	787	mixed	R2866_044	570011	570339	329	intergenic
Rd_045	476500	477190	691	mixed	R2866_045	570544	571082	539	antisense
Rd_046	477271	477797	526	mixed	R2866_046	574965	575734	770	mixed
Rd_047	484909	485234	326	intergenic	R2866_047	586813	587344	532	antisense
Rd_048	488999	490088	1090	antisense	R2866_048	611151	611641	491	intergenic
Rd_049	497643	499189	1547	mixed	R2866_049	626576	627214	638	mixed
Rd_050	534853	535313	460	antisense	R2866_050	642578	643219	642	mixed
Rd_051	537894	538102	209	mixed	R2866_051	650312	650559	247	mixed
Rd_052	544821	545640	820	mixed	R2866_052	673871	674425	555	antisense
Rd_053	555630	556800	1170	mixed	R2866_053	675975	676626	652	mixed

Name	Start position	Stop position	Length	Type	Name	Start position	Stop position	Length	Type
Rd					R2866				
Rd_054	557547	558857	1311	mixed	R2866_054	686375	686787	413	antisense
Rd_055	613015	613361	346	mixed	R2866_055	692900	693767	868	mixed
Rd_056	617165	618118	954	mixed	R2866_056	729897	732565	2669	mixed
Rd_057	641974	643467	1494	mixed	R2866_057	742675	743213	539	mixed
Rd_058	659943	661422	1479	antisense	R2866_058	758611	758847	236	intergenic
Rd_059	662409	662498	89	intergenic	R2866_059	766185	766446	262	antisense
Rd_060	666223	666468	246	mixed	R2866_060	776385	776842	458	antisense
Rd_061	673236	674403	1168	mixed	R2866_061	781437	782175	739	mixed
Rd_062	674424	674962	539	mixed	R2866_062	791588	792501	914	mixed
Rd_063	692219	692619	401	mixed	R2866_063	791708	792149	442	intergenic
Rd_064	694639	695370	732	mixed	R2866_064	808837	809046	210	mixed
Rd_065	708324	708556	233	mixed	R2866_065	821541	822128	587	mixed
Rd_066	721899	722015	116	intergenic	R2866_066	831407	832623	1217	mixed
Rd_067	739932	740404	473	antisense	R2866_067	834258	835311	1054	antisense
Rd_068	740746	741363	618	antisense	R2866_068	875603	876167	565	mixed
Rd_069	743129	744301	1173	antisense	R2866_069	891340	892159	820	intergenic
Rd_070	745331	745785	454	mixed	R2866_070	892244	892573	330	antisense
Rd_071	755770	755977	207	intergenic	R2866_071	898563	898961	399	antisense
Rd_072	785976	786677	702	antisense	R2866_072	904640	905021	382	intergenic
Rd_073	787692	788471	780	antisense	R2866_073	908067	909070	1004	mixed
Rd_074	788487	789222	736	mixed	R2866_074	909244	910626	1383	antisense
Rd_075	801028	801557	530	mixed	R2866_075	949794	949955	162	mixed
Rd_076	804927	805709	783	mixed	R2866_076	964326	964722	397	mixed
Rd_077	808749	809804	1056	mixed	R2866_077	973191	974050	860	mixed
Rd_078	828325	829313	989	mixed	R2866_078	993545	993783	238	intergenic
Rd_079	833686	834447	762	mixed	R2866_079	995263	995765	503	mixed
Rd_080	840869	841324	456	antisense	R2866_080	1006041	1006276	236	intergenic
Rd_081	876552	876924	372	intergenic	R2866_081	1012493	1013174	682	mixed
Rd_082	885347	885661	315	intergenic	R2866_082	1030821	1031179	359	antisense
Rd_083	885469	885645	177	intergenic	R2866_083	1037230	1038444	1215	mixed
Rd_084	897457	897766	310	intergenic	R2866_084	1051711	1052380	670	antisense
Rd_085	914086	914547	462	mixed	R2866_085	1058738	1059734	997	mixed
Rd_086	915521	915888	367	intergenic	R2866_086	1103736	1104765	1030	mixed
Rd_087	929086	930970	1885	antisense	R2866_087	1131442	1131731	290	mixed
Rd_088	965560	966963	1404	mixed	R2866_088	1132270	1132860	591	mixed
Rd_089	992036	992369	334	antisense	R2866_089	1150324	1150566	242	intergenic
Rd_090	994816	995745	930	mixed	R2866_090	1177936	1178325	390	mixed
Rd_091	1016881	1017852	972	mixed	R2866_091	1178739	1181057	2319	mixed
Rd_092	1033013	1033875	863	mixed	R2866_092	1191418	1191868	451	mixed
Rd_093	1041098	1041478	381	antisense	R2866_093	1194076	1194461	386	mixed
Rd_094	1052866	1053413	548	antisense	R2866_094	1198769	1198991	222	intergenic
Rd_095	1068302	1068890	589	mixed	R2866_095	1221439	1221709	271	intergenic
Rd_096	1071796	1072503	708	antisense	R2866_096	1221502	1222055	554	mixed
Rd_097	1073701	1074177	477	antisense	R2866_097	1242786	1243183	398	intergenic
Rd_098	1080264	1080879	615	mixed	R2866_098	1245476	1246096	621	mixed
Rd_099	1083534	1084437	904	mixed	R2866_099	1254839	1255962	1124	mixed
Rd_100	1102893	1104117	1225	mixed	R2866_100	1256714	1257862	1149	mixed
Rd_101	1110417	1111758	1342	mixed	R2866_101	1259245	1259534	290	intergenic
Rd_102	1113627	1114077	451	antisense	R2866_102	1259344	1259817	474	mixed
Rd_103	1147693	1148130	437	mixed	R2866_103	1264431	1265888	1458	mixed
Rd_104	1150191	1150430	239	mixed	R2866_104	1269607	1270022	416	antisense
Rd_105	1173203	1173472	269	intergenic	R2866_105	1277352	1278405	1054	mixed
Rd_106	1173213	1173455	243	intergenic	R2866_106	1284990	1286304	1315	mixed
Rd_107	1179087	1179305	219	antisense	R2866_107	1301932	1302148	217	antisense
Rd_108	1194933	1196201	1269	mixed	R2866_108	1307763	1308096	334	intergenic
Rd_109	1203265	1203822	558	mixed	R2866_109	1307782	1308061	280	intergenic
Rd_110	1217011	1217468	458	mixed	R2866_110	1330706	1330920	215	antisense
Rd_111	1222538	1222811	274	mixed	R2866_111	1356981	1357494	514	mixed
Rd_112	1222610	1222912	303	intergenic	R2866_112	1367615	1367940	326	mixed
Rd_113	1236131	1236896	766	mixed	R2866_113	1370928	1371628	700	antisense
Rd_114	1239034	1239442	409	intergenic	R2866_114	1378432	1378907	476	mixed
Rd_115	1255321	1255696	376	antisense	R2866_115	1391570	1393605	2036	mixed
Rd_116	1260213	1260718	506	mixed	R2866_116	1418314	1418931	618	antisense
Rd_117	1265023	1265765	743	antisense	R2866_117	1454600	1455239	640	mixed
Rd_118	1274013	1276134	2122	mixed	R2866_118	1473212	1475161	1950	mixed
Rd_119	1303104	1304060	957	mixed	R2866_119	1476038	1476384	347	mixed
Rd_120	1318219	1318845	627	mixed	R2866_120	1490768	1491332	564	mixed
Rd_121	1323679	1323971	293	intergenic	R2866_121	1503072	1504125	1053	mixed
Rd_122	1336677	1337031	355	antisense	R2866_122	1510202	1511152	951	mixed
Rd_123	1348269	1348666	398	intergenic	R2866_123	1517911	1518282	372	antisense
Rd_124	1372243	1372984	742	antisense	R2866_124	1535931	1536542	612	mixed
Rd_125	1386245	1387027	782	mixed	R2866_125	1554103	1554489	386	intergenic

Name	Start position	Stop position	Length	Type	Name	Start position	Stop position	Length	Type
Rd					R2866				
Rd_126	1390722	1392443	1722	mixed	R2866_126	1555300	1555903	604	mixed
Rd_127	1399972	1400704	732	mixed	R2866_127	1566085	1566450	366	antisense
Rd_128	1419419	1419850	431	antisense	R2866_128	1567612	1567841	229	intergenic
Rd_129	1421775	1422207	433	antisense	R2866_129	1570086	1570511	426	antisense
Rd_130	1472704	1473265	562	mixed	R2866_130	1571782	1571932	150	antisense
Rd_131	1475674	1475988	315	intergenic	R2866_131	1572288	1572544	257	intergenic
Rd_132	1483534	1483759	226	intergenic	R2866_132	1592489	1592795	307	intergenic
Rd_133	1505845	1506661	817	mixed	R2866_133	1592505	1592681	177	intergenic
Rd_134	1515886	1516381	496	mixed	R2866_134	1607433	1607912	480	mixed
Rd_135	1518769	1518894	126	mixed	R2866_135	1643962	1645106	1145	mixed
Rd_136	1529438	1529910	473	mixed	R2866_136	1649473	1650470	998	mixed
Rd_137	1545292	1545803	512	mixed	R2866_137	1669077	1670181	1105	mixed
Rd_138	1562150	1562462	313	mixed	R2866_138	1670185	1670563	379	antisense
Rd_139	1576249	1577329	1081	mixed	R2866_139	1673212	1674054	843	mixed
Rd_140	1578359	1579198	840	antisense	R2866_140	1674588	1675134	547	antisense
Rd_141	1579222	1579848	627	antisense	R2866_141	1677370	1677857	487	mixed
Rd_142	1602676	1602924	248	antisense	R2866_142	1696211	1696464	254	antisense
Rd_143	1605548	1606363	816	mixed	R2866_143	1733584	1734196	613	mixed
Rd_144	1616521	1616858	338	mixed	R2866_144	1743386	1743894	508	mixed
Rd_145	1624072	1624299	228	intergenic	R2866_145	1744080	1745121	1042	antisense
Rd_146	1659756	1660684	929	mixed	R2866_146	1757230	1757710	481	antisense
Rd_147	1666018	1666223	205	intergenic	R2866_147	1780641	1780933	292	intergenic
Rd_148	1685708	1686593	886	mixed	R2866_148	1804546	1805373	828	mixed
Rd_149	1694403	1694923	521	antisense	R2866_149	1834710	1835250	541	intergenic
Rd_150	1714534	1714829	295	mixed	R2866_150	1851986	1852172	187	antisense
Rd_151	1714817	1715360	544	intergenic	R2866_151	1853122	1853918	797	antisense
Rd_152	1714856	1715470	615	mixed	R2866_152	1855188	1855517	329	mixed
Rd_153	1724508	1725675	1168	mixed	R2866_153	1890125	1891074	950	mixed
Rd_154	1732610	1733522	913	mixed	R2866_154	1899828	1900062	234	intergenic
Rd_155	1733417	1733874	458	mixed	R2866_155	1902432	1903046	615	intergenic
Rd_156	1748493	1748724	232	intergenic	R2866_156	1903042	1904111	1070	intergenic
Rd_157	1777537	1779667	2131	mixed	R2866_157	1904502	1904767	266	mixed
Rd_158	1785066	1785298	233	intergenic	R2866_158	1906897	1907094	197	antisense
Rd_159	1785416	1785659	244	intergenic	R2866_159	1907095	1907225	131	antisense
Rd_160	1788561	1789102	542	intergenic	R2866_160	1907338	1907455	118	antisense
Rd_161	1789118	1789884	767	mixed	R2866_161	1908114	1908244	131	intergenic
Rd_162	1789913	1790132	219	mixed	R2866_162	1908264	1908555	292	intergenic
Rd_163	1794537	1795072	536	mixed	R2866_163	1908916	1909084	168	intergenic
Rd_164	1795845	1796096	251	intergenic					
Rd_165	1796507	1796828	322	intergenic					

Table S11: Differentially expressed ncRNAs in Rd during stationary phase.

ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs		
Rd_011	22.31	1.79E-52
Rd_092	13.77	1.59E-32
Rd_010	13.37	5.73E-21
Rd_141	9.59	8.76E-22
Rd_146	9.55	2.83E-51
Rd_120	9.22	6.62E-116
Rd_106	9.00	6.07E-13
Rd_078	8.79	1.55E-45
Rd_082	8.48	1.35E-77
Rd_139	7.81	6.73E-27
Rd_031	7.74	7.65E-29
Rd_036	7.71	1.69E-13
Rd_062	7.33	4.66E-43
Rd_116	7.21	1.94E-32
Rd_153	6.90	4.28E-25
Rd_007	6.88	1.65E-30
Rd_034	6.80	3.01E-64
Rd_030	6.79	2.11E-34
Rd_115	6.66	3.13E-19
Rd_048	6.28	3.01E-16
Rd_107	6.14	1.40E-09
Rd_091	6.11	7.55E-33
Rd_083	6.02	8.81E-17
Rd_157	5.99	2.32E-40
Rd_022	5.92	1.68E-18
Rd_038	5.88	1.63E-32
Rd_009	5.77	9.58E-53
Rd_096	5.77	2.79E-22
Rd_140	5.73	1.51E-23
Rd_109	5.49	2.55E-17
Rd_090	5.29	8.87E-51
Rd_069	5.26	1.31E-36
Rd_068	5.25	2.79E-22
Rd_122	5.23	1.09E-20
Rd_020	5.20	9.00E-25
Rd_095	5.13	1.48E-20
Rd_143	5.10	6.00E-23
Rd_074	4.94	2.87E-15
Rd_023	4.71	1.26E-11
Rd_057	4.55	4.69E-28
Rd_137	4.47	3.71E-20
Rd_012	4.31	2.04E-28
Rd_064	4.28	1.40E-14
Rd_087	4.20	1.02E-36
Rd_056	4.12	1.13E-27
Rd_073	4.07	1.62E-13
Rd_129	4.03	7.14E-27
Rd_054	3.96	3.19E-16
Rd_161	3.91	6.61E-36
Rd_149	3.87	1.21E-11
Rd_032	3.77	1.52E-11
Rd_077	3.75	4.51E-31
Rd_001	3.62	7.15E-14
Rd_131	3.59	9.84E-08
Rd_132	3.59	6.37E-20
Rd_072	3.55	4.67E-11
Rd_052	3.51	2.33E-14
Rd_037	3.47	1.03E-15
Rd_100	3.43	6.55E-30
Rd_051	3.41	3.02E-30
Rd_124	3.39	3.56E-08
Rd_084	3.35	2.77E-10
Rd_061	3.29	2.02E-18
Rd_050	3.25	8.34E-09
Rd_063	3.21	2.52E-13
Rd_110	3.20	1.52E-10
Rd_112	3.19	9.72E-39
Rd_058	3.15	8.82E-18

ncRNA	Fold change	Adjusted p-value
Down-regulated ncRNAs		
Rd_164	13.76	5.59E-88
Rd_159	13.19	4.26E-87
Rd_156	7.66	1.74E-30
Rd_160	4.11	6.96E-32
Rd_128	2.14	5.93E-07

ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs		
Rd_104	3.11	3.49E-07
Rd_126	3.11	1.33E-22
Rd_014	3.10	2.75E-21
Rd_008	3.08	3.13E-30
Rd_080	3.05	2.67E-09
Rd_144	3.05	2.80E-07
Rd_101	3.02	3.71E-13
Rd_119	3.02	3.83E-14
Rd_099	3.02	3.17E-21
Rd_027	2.97	2.02E-12
Rd_043	2.96	9.89E-15
Rd_067	2.95	9.68E-08
Rd_093	2.92	2.45E-12
Rd_039	2.89	4.15E-22
Rd_148	2.87	1.44E-14
Rd_005	2.73	2.07E-09
Rd_089	2.69	1.04E-07
Rd_118	2.65	3.73E-18
Rd_042	2.64	4.29E-08
Rd_088	2.62	4.31E-15
Rd_081	2.57	1.11E-17
Rd_114	2.56	5.87E-15
Rd_111	2.50	5.13E-14
Rd_046	2.42	4.23E-08
Rd_130	2.41	6.75E-09
Rd_085	2.40	1.16E-07
Rd_142	2.40	0.000675838
Rd_079	2.40	2.96E-10
Rd_021	2.34	2.09E-06
Rd_016	2.33	7.78E-13
Rd_113	2.28	2.57E-07
Rd_019	2.27	0.000202062
Rd_108	2.24	3.65E-05
Rd_151	2.20	3.07E-07
Rd_029	2.18	0.000321933
Rd_013	2.15	2.46E-06
Rd_121	2.11	0.000178688
Rd_097	2.11	0.003194726
Rd_152	2.11	2.55E-08
Rd_055	2.03	1.45E-06
Rd_004	2.00	3.26E-11

Table S12: Differentially expressed ncRNAs in R2866 during stationary phase.

ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs		
R2866_112	14.28	1.21E-131
R2866_075	13.45	1.27E-36
R2866_067	10.39	1.46E-60
R2866_132	8.64	5.56E-88
R2866_001	8.12	1.32E-59
R2866_144	7.39	8.78E-143
R2866_031	7.19	3.85E-256
R2866_036	6.90	4.58E-27
R2866_057	6.57	6.83E-24
R2866_050	6.28	3.89E-51
R2866_133	6.27	2.85E-28
R2866_136	6.09	1.11E-53
R2866_101	5.35	5.86E-277
R2866_056	5.12	5.31E-124
R2866_117	4.99	1.06E-24
R2866_142	4.92	4.90E-15
R2866_109	4.83	3.62E-15
R2866_064	4.56	7.36E-10
R2866_046	4.53	3.88E-173
R2866_092	4.50	3.28E-72
R2866_118	4.43	4.80E-18
R2866_083	4.42	2.13E-55
R2866_086	4.36	1.76E-51
R2866_055	4.34	4.65E-44
R2866_028	4.32	1.23E-84
R2866_087	4.31	1.39E-20
R2866_093	4.28	9.27E-24
R2866_070	4.28	2.22E-94
R2866_082	4.08	8.45E-19
R2866_053	4.01	4.32E-21
R2866_103	3.95	1.51E-103
R2866_061	3.94	2.66E-32
R2866_017	3.73	1.22E-93
R2866_119	3.71	9.24E-16
R2866_137	3.69	7.03E-176
R2866_074	3.68	1.61E-49
R2866_068	3.64	4.04E-45
R2866_157	3.60	5.73E-23
R2866_091	3.60	3.94E-143
R2866_066	3.50	1.82E-40
R2866_011	3.47	1.45E-11
R2866_159	3.40	2.69E-27
R2866_027	3.39	2.94E-81
R2866_099	3.28	4.76E-33
R2866_073	3.22	2.16E-36
R2866_026	3.17	4.83E-60
R2866_090	3.15	1.01E-13
R2866_134	3.12	3.44E-17
R2866_088	3.10	6.45E-45
R2866_048	3.10	3.07E-141
R2866_013	3.08	8.79E-47
R2866_060	3.07	6.38E-20
R2866_021	3.04	3.62E-12
R2866_096	3.03	2.86E-20
R2866_033	3.02	5.42E-16
R2866_035	2.98	1.99E-22
R2866_107	2.90	8.02E-07
R2866_161	2.87	2.51E-18
R2866_123	2.86	3.78E-23
R2866_149	2.85	3.57E-62
R2866_140	2.84	2.16E-16
R2866_145	2.84	1.33E-24
R2866_155	2.82	2.43E-14
R2866_084	2.77	2.39E-16
R2866_022	2.76	2.49E-18

ncRNA	Fold change	Adjusted p-value
Down-regulated ncRNAs		
R2866_158	12.31	2.82E-65
R2866_154	9.82	2.79E-96
R2866_163	2.74	6.14E-06
R2866_003	2.60	2.24E-35
R2866_058	2.52	8.30E-35
R2866_141	2.51	2.79E-16

ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs		
R2866_039	2.73	6.76E-55
R2866_097	2.72	4.31E-82
R2866_023	2.70	9.64E-12
R2866_030	2.69	3.29E-18
R2866_071	2.66	2.31E-11
R2866_146	2.66	2.28E-20
R2866_085	2.64	2.45E-35
R2866_004	2.61	1.05E-11
R2866_160	2.54	2.44E-08
R2866_072	2.54	4.80E-28
R2866_105	2.51	1.23E-59
R2866_135	2.50	2.68E-36
R2866_110	2.48	2.07E-11
R2866_115	2.41	3.32E-30
R2866_122	2.39	4.75E-27
R2866_162	2.36	5.67E-11
R2866_120	2.34	2.01E-10
R2866_045	2.33	2.32E-11
R2866_016	2.33	2.03E-12
R2866_037	2.28	2.44E-31
R2866_042	2.27	2.33E-14
R2866_008	2.22	2.38E-18
R2866_038	2.22	5.89E-43
R2866_111	2.20	1.44E-08
R2866_065	2.19	2.39E-09
R2866_080	2.18	3.52E-27
R2866_131	2.14	4.73E-09
R2866_062	2.12	6.60E-37
R2866_044	2.11	1.36E-85
R2866_106	2.09	3.56E-25
R2866_104	2.02	3.16E-05
R2866_114	2.02	2.44E-20

Table S13: Differentially expressed ncRNAs in Rd during oxidative stress.

ncRNA	Fold change	Adjusted p-value	ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs			Down-regulated ncRNAs		
Rd_116	3.58	2.31E-19	Rd_143	2.17	5.20E-09
Rd_135	2.46	2.59E-10			
Rd_027	2.36	3.27E-09			
Rd_064	2.09	0.000599706			

Table S14: Differentially expressed ncRNAs in R2866 during oxidative stress.

ncRNA	Fold change	Adjusted p-value	ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs			Down-regulated ncRNAs		
R2866_049	2.90	1.42E-15	R2866_076	2.09	0.031810371
R2866_040	2.30	3.78E-05			

Table S15: Differentially expressed ncRNAs in Rd during iron-starvation stress.

ncRNA	Fold change	Adjusted p-value	ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs			**Down-regulated ncRNAs**		
Rd_114	4.94	1.19E-27	Rd_087	2.61	1.88E-11
Rd_165	4.86	1.24E-54	Rd_032	2.11	0.000658856
Rd_158	4.18	1.31E-38	Rd_050	2.06	0.000805925
Rd_070	2.61	1.35E-06	Rd_161	2.04	5.47E-10
Rd_117	2.45	5.86E-11			
Rd_060	2.30	7.88E-05			
Rd_002	2.21	7.77E-07			
Rd_160	2.20	5.31E-16			
Rd_028	2.15	9.13E-06			
Rd_030	2.13	2.86E-06			
Rd_022	2.03	5.10E-07			

Table S16: Differentially expressed ncRNAs in R2866 during iron-starvation stress.

ncRNA	Fold change	Adjusted p-value	ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs			Down-regulated ncRNAs		
R2866_104	3.19	1.66E-06	R2866_154	5.00	4.56E-21
R2866_027	2.18	1.28E-08	R2866_158	4.26	1.86E-13
R2866_071	2.12	0.003649679	R2866_015	2.85	1.22E-05

Table S17: Differentially expressed ncRNAs in Rd during nutritional stress.

ncRNA	Fold change	Adjusted p-value	ncRNA	Fold change	Adjusted p-value
Up-regulated ncRNAs			Down-regulated ncRNAs		
Rd_098	2.08	2.99E-09	Rd_032	7.76	9.76E-35
			Rd_145	4.38	3.98E-19
			Rd_015	4.21	2.54E-11
			Rd_069	3.21	1.16E-08
			Rd_116	2.95	9.68E-06
			Rd_120	2.91	1.64E-09
			Rd_050	2.83	1.66E-09
			Rd_101	2.81	1.19E-06
			Rd_142	2.69	2.21E-05
			Rd_078	2.67	1.29E-05
			Rd_104	2.61	1.26E-05
			Rd_022	2.52	0.000435058
			Rd_137	2.47	5.42E-14
			Rd_095	2.39	3.45E-06
			Rd_006	2.28	0.000124837
			Rd_052	2.26	3.63E-14
			Rd_124	2.22	4.70E-05
			Rd_156	2.19	0.00056713
			Rd_138	2.10	2.02E-08
			Rd_038	2.10	0.000203088
			Rd_002	2.09	1.45E-05
			Rd_011	2.01	0.000109819

References

ABU KWAIK, Y., MCLAUGHLIN, R. E., APICELLA, M. A. & SPINOLA, S. M. 1991. Analysis of *Haemophilus influenzae* type b lipooligosaccharide-synthesis genes that assemble or expose a 2-keto-3-deoxyoctulosonic acid epitope. *Mol Microbiol,* 5**,** 2475-80.

AGRAWAL, A. & MURPHY, T. F. 2011. *Haemophilus influenzae* infections in the *H. influenzae* type b conjugate vaccine era. *J Clin Microbiol,* 49**,** 3728-32.

AHMED, A. S., KHAN, N. Z., HUSSAIN, M., AMIN, M. R., HANIF, M., MAHBUB, M., EL-ARIFEEN, S., BAQUI, A. H., QAZI, S. A. & SAHA, S. K. 2013. Follow-up of cases of *Haemophilus influenzae* type b meningitis to determine its long-term sequelae. *J Pediatr,* 163**,** S44-9.

ALEXANDER, H. E. & LEIDY, G. 1951. Determination of inherited traits of *H. influenzae* by desoxyribonucleic acid fractions isolated from type-specific cells. *J Exp Med,* 93**,** 345-59.

ALMIRON, M., LINK, A. J., FURLONG, D. & KOLTER, R. 1992. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev,* 6**,** 2646-54.

ALTUVIA, S., WEINSTEIN-FISCHER, D., ZHANG, A., POSTOW, L. & STORZ, G. 1997. A small, stable RNA induced by oxidative stress: role as a pleiotropic regulator and antimutator. *Cell,* 90**,** 43-53.

ANDERS, S. & HUBER, W. 2010. Differential expression analysis for sequence count data. *Genome Biol,* 11**,** R106.

ANDERSON, R., WANG, X., BRIERE, E. C., KATZ, L. S., COHN, A. C., CLARK, T. A., MESSONNIER, N. E. & MAYER, L. W. 2012. *Haemophilus haemolyticus* isolates causing clinical disease. *J Clin Microbiol,* 50**,** 2462-5.

ANDERSSON, M., RESMAN, F., EITREM, R., DROBNI, P., RIESBECK, K., KAHLMETER, G. & SUNDQVIST, M. 2015. Outbreak of a beta-lactam resistant non-typeable *Haemophilus influenzae* sequence type 14 associated with severe clinical outcomes. *BMC Infect Dis,* 15**,** 581.

ARGAMAN, L., HERSHBERG, R., VOGEL, J., BEJERANO, G., WAGNER, E. G., MARGALIT, H. & ALTUVIA, S. 2001. Novel small RNA-encoding genes in the intergenic regions of *Escherichia coli*. *Curr Biol,* 11**,** 941-50.

BADDAL, B., MUZZI, A., CENSINI, S., CALOGERO, R. A., TORRICELLI, G., GUIDOTTI, S., TADDEI, A. R., COVACCI, A., PIZZA, M., RAPPUOLI, R., SORIANI, M. & PEZZICOLI, A. 2015. Dual RNA-seq of Nontypeable *Haemophilus influenzae* and Host Cell Transcriptomes Reveals Novel Insights into Host-Pathogen Cross Talk. *MBio,* 6**,** e01765-15.

BAKALETZ, L. O., BAKER, B. D., JURCISEK, J. A., HARRISON, A., NOVOTNY, L. A., BOOKWALTER, J. E., MUNGUR, R. & MUNSON, R. S., JR. 2005. Demonstration of Type IV pilus expression and a twitching phenotype by *Haemophilus influenzae*. *Infect Immun,* 73**,** 1635-43.

BALLEZA, E., LOPEZ-BOJORQUEZ, L. N., MARTINEZ-ANTONIO, A., RESENDIS-ANTONIO, O., LOZADA-CHAVEZ, I., BALDERAS-MARTINEZ, Y. I., ENCARNACION, S. & COLLADO-VIDES, J. 2009. Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol Rev,* 33**,** 133-51.

BANKEVICH, A., NURK, S., ANTIPOV, D., GUREVICH, A. A., DVORKIN, M., KULIKOV, A. S., LESIN, V. M., NIKOLENKO, S. I., PHAM, S., PRJIBELSKI, A. D., PYSHKIN, A. V., SIROTKIN, A. V., VYAHHI, N., TESLER, G., ALEKSEYEV, M. A. & PEVZNER, P. A. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol,* 19**,** 455-77.

BARBE, V., CRUVEILLER, S., KUNST, F., LENOBLE, P., MEURICE, G., SEKOWSKA, A., VALLENET, D., WANG, T., MOSZER, I., MEDIGUE, C. & DANCHIN, A. 2009. From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology,* 155**,** 1758-75.

BARENKAMP, S. J. & ST GEME, J. W., 3RD 1996. Identification of a second family of high-molecular-weight adhesion proteins expressed by non-typable *Haemophilus influenzae*. *Mol Microbiol,* 19**,** 1215-23.

BAROUKI, R. & SMITH, H. O. 1985. Reexamination of phenotypic defects in rec-1 and rec-2 mutants of *Haemophilus influenzae* Rd. *J Bacteriol,* 163**,** 629-34.

BARRANGOU, R., FREMAUX, C., DEVEAU, H., RICHARDS, M., BOYAVAL, P., MOINEAU, S., ROMERO, D. A. & HORVATH, P. 2007. CRISPR provides acquired resistance against viruses in prokaryotes. *Science,* 315**,** 1709-12.

BAYLISS, C. D., FIELD, D. & MOXON, E. R. 2001. The simple sequence contingency loci of *Haemophilus influenzae* and *Neisseria meningitidis*. *J Clin Invest,* 107**,** 657-62.

BEAM, C. E., SAVESON, C. J. & LOVETT, S. T. 2002. Role for radA/sms in recombination intermediate processing in *Escherichia coli*. *J Bacteriol,* 184**,** 6836-44.

BENITO, Y., KOLB, F. A., ROMBY, P., LINA, G., ETIENNE, J. & VANDENESCH, F. 2000. Probing the structure of RNAIII, the *Staphylococcus aureus* agr regulatory RNA, and identification of the RNA domain involved in repression of protein A expression. *RNA,* 6**,** 668-79.

BENT, Z. W., POOREY, K., BRAZEL, D. M., LABAUVE, A. E., SINHA, A., CURTIS, D. J., HOUSE, S. E., TEW, K. E., HAMBLIN, R. Y., WILLIAMS, K. P., BRANDA, S. S., YOUNG, G. M. & MEAGHER, R. J. 2015. Transcriptomic Analysis of *Yersinia enterocolitica* Biovar 1B Infecting Murine Macrophages Reveals New Mechanisms of Extracellular and Intracellular Survival. *Infect Immun,* 83**,** 2672-85.

BENTLEY, D. R., BALASUBRAMANIAN, S., SWERDLOW, H. P., SMITH, G. P., MILTON, J., BROWN, C. G., HALL, K. P., EVERS, D. J., BARNES, C. L., BIGNELL, H. R., BOUTELL, J. M., BRYANT, J., CARTER, R. J., KEIRA CHEETHAM, R., COX, A. J., ELLIS, D. J., FLATBUSH, M. R., GORMLEY, N. A., HUMPHRAY, S. J., IRVING, L. J., KARBELASHVILI, M. S., KIRK, S. M., LI, H., LIU, X., MAISINGER, K. S., MURRAY, L. J., OBRADOVIC, B., OST, T., PARKINSON, M. L., PRATT, M. R., RASOLONJATOVO, I. M., REED, M. T., RIGATTI, R., RODIGHIERO, C., ROSS, M. T., SABOT, A., SANKAR, S. V., SCALLY, A., SCHROTH, G. P., SMITH, M. E., SMITH, V. P., SPIRIDOU, A., TORRANCE, P. E., TZONEV, S. S., VERMAAS, E. H., WALTER, K., WU, X., ZHANG, L., ALAM, M. D., ANASTASI, C., ANIEBO, I. C., BAILEY, D. M., BANCARZ, I. R., BANERJEE, S., BARBOUR, S. G., BAYBAYAN, P. A., BENOIT, V. A., BENSON, K. F., BEVIS, C., BLACK, P. J., BOODHUN, A., BRENNAN, J. S., BRIDGHAM, J. A., BROWN, R. C., BROWN, A. A., BUERMANN, D. H., BUNDU, A. A., BURROWS, J. C., CARTER, N. P., CASTILLO, N., CHIARA, E. C. M., CHANG, S., NEIL COOLEY, R., CRAKE, N. R., DADA, O. O., DIAKOUMAKOS, K. D., DOMINGUEZ-FERNANDEZ, B., EARNSHAW, D. J.,

EGBUJOR, U. C., ELMORE, D. W., ETCHIN, S. S., EWAN, M. R., FEDURCO, M., FRASER, L. J., FUENTES FAJARDO, K. V., SCOTT FUREY, W., GEORGE, D., GIETZEN, K. J., GODDARD, C. P., GOLDA, G. S., GRANIERI, P. A., GREEN, D. E., GUSTAFSON, D. L., HANSEN, N. F., HARNISH, K., HAUDENSCHILD, C. D., HEYER, N. I., HIMS, M. M., HO, J. T., HORGAN, A. M., et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature,* 456**,** 53-9.

BISHAI, W. R., SMITH, H. O. & BARCAK, G. J. 1994. A peroxide/ascorbate-inducible catalase from *Haemophilus influenzae* is homologous to the *Escherichia coli* katE gene product. *J Bacteriol,* 176**,** 2914-21.

BOUCHER, M. B., BEDOTTO, M., COUDERC, C., GOMEZ, C., REYNAUD-GAUBERT, M. & DRANCOURT, M. 2012. *Haemophilus pittmaniae* respiratory infection in a patient with siderosis: a case report. *J Med Case Rep,* 6**,** 120.

BROUQUI, P. & RAOULT, D. 2001. Endocarditis due to rare and fastidious bacteria. *Clin Microbiol Rev,* 14**,** 177-207.

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K. & MADDEN, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics,* 10**,** 421.

CAMPOS, J., ROMAN, F., PEREZ-VAZQUEZ, M., ARACIL, B., OTEO, J., CERCENADO, E. & SPANISH STUDY GROUP FOR, H. I. T. F. 2003. Antibiotic resistance and clinical significance of *Haemophilus influenzae* type f. *J Antimicrob Chemother,* 52**,** 961-6.

CARDINES, R., GIUFRE, M., POMPILIO, A., FISCARELLI, E., RICCIOTTI, G., DI BONAVENTURA, G. & CERQUETTI, M. 2012. *Haemophilus influenzae* in children with cystic fibrosis: antimicrobial susceptibility, molecular epidemiology, distribution of adhesins and biofilm formation. *Int J Med Microbiol,* 302**,** 45-52.

CARVER, T. J., RUTHERFORD, K. M., BERRIMAN, M., RAJANDREAM, M. A., BARRELL, B. G. & PARKHILL, J. 2005. ACT: the Artemis Comparison Tool. *Bioinformatics,* 21**,** 3422-3.

CASEY, J. R. & PICHICHERO, M. E. 2004. Changes in frequency and pathogens causing acute otitis media in 1995-2003. *Pediatr Infect Dis J,* 23**,** 824-8.

CHEN, S., ZHANG, A., BLYN, L. B. & STORZ, G. 2004. MicC, a second small-RNA regulator of Omp protein expression in *Escherichia coli*. *J Bacteriol,* 186**,** 6689-97.

CHO, B. K., FEDEROWICZ, S. A., EMBREE, M., PARK, Y. S., KIM, D. & PALSSON, B. O. 2011. The PurR regulon in *Escherichia coli* K-12 MG1655. *Nucleic Acids Res,* 39**,** 6456-64.

CHONOLES IMLAY, K. R., KORSHUNOV, S. & IMLAY, J. A. 2015. Physiological Roles and Adverse Effects of the Two Cystine Importers of *Escherichia coli*. *J Bacteriol,* 197**,** 3629-44.

CHRISTENSEN, S. K. & GERDES, K. 2003. RelE toxins from bacteria and Archaea cleave mRNAs on translating ribosomes, which are rescued by tmRNA. *Mol Microbiol,* 48**,** 1389-400.

CLARKE, D. M. & BRAGG, P. D. 1985. Cloning and expression of the transhydrogenase gene of *Escherichia coli*. *J Bacteriol,* 162**,** 367-73.

COLLINS, S., RAMSAY, M., CAMPBELL, H., SLACK, M. P. & LADHANI, S. N. 2013. Invasive *Haemophilus influenzae* type b (Hib) disease in England and Wales: who is at risk after two decades of routine childhood vaccination? *Clin Infect Dis*.

COPE, E. K., GOLDSTEIN-DARUECH, N., KOFONOW, J. M., CHRISTENSEN, L., MCDERMOTT, B., MONROY, F., PALMER, J. N., CHIU, A. G., SHIRTLIFF, M. E., COHEN, N. A. & LEID, J. G. 2011. Regulation of virulence gene expression resulting from *Streptococcus pneumoniae* and nontypeable *Haemophilus influenzae* interactions in chronic disease. *PLoS One,* 6**,** e28523.

DAI, Y. & OUTTEN, F. W. 2012. The *E. coli* SufS-SufE sulfur transfer system is more resistant to oxidative stress than IscS-IscU. *FEBS Lett,* 586**,** 4016-22.

DAINES, D. A., COHN, L. A., COLEMAN, H. N., KIM, K. S. & SMITH, A. L. 2003. *Haemophilus influenzae* Rd KW20 has virulence properties. *J Med Microbiol,* 52**,** 277-82.

DAINES, D. A., JARISCH, J. & SMITH, A. L. 2004. Identification and characterization of a nontypeable *Haemophilus influenzae* putative toxin-antitoxin locus. *BMC Microbiol,* 4**,** 30.

DAINES, D. A., WU, M. H. & YUAN, S. Y. 2007. VapC-1 of nontypeable *Haemophilus influenzae* is a ribonuclease. *J Bacteriol,* 189**,** 5041-8.

DARLING, A. C., MAU, B., BLATTNER, F. R. & PERNA, N. T. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res,* 14**,** 1394-403.

DAVIS, J., SMITH, A. L., HUGHES, W. R. & GOLOMB, M. 2001. Evolution of an autotransporter: domain shuffling and lateral transfer from pathogenic *Haemophilus* to *Neisseria*. *J Bacteriol,* 183**,** 4626-35.

DAWID, S., BARENKAMP, S. J. & ST GEME, J. W., 3RD 1999. Variation in expression of the *Haemophilus influenzae* HMW adhesins: a prokaryotic system reminiscent of eukaryotes. *Proc Natl Acad Sci U S A,* 96**,** 1077-82.

DE BOLLE, X., BAYLISS, C. D., FIELD, D., VAN DE VEN, T., SAUNDERS, N. J., HOOD, D. W. & MOXON, E. R. 2000. The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases. *Mol Microbiol,* 35**,** 211-22.

DE SOUZA-HART, J. A., BLACKSTOCK, W., DI MODUGNO, V., HOLLAND, I. B. & KOK, M. 2003. Two-component systems in *Haemophilus influenzae*: a regulatory role for ArcA in serum resistance. *Infect Immun,* 71**,** 163-72.

DEADMAN, M. E., HERMANT, P., ENGSKOG, M., MAKEPEACE, K., MOXON, E. R., SCHWEDA, E. K. & HOOD, D. W. 2009. Lex2B, a phase-variable glycosyltransferase, adds either a glucose or a galactose to *Haemophilus influenzae* lipopolysaccharide. *Infect Immun,* 77**,** 2376-84.

DICARLO, R. P., ARMENTOR, B. S. & MARTIN, D. H. 1995. Chancroid epidemiology in New Orleans men. *J Infect Dis,* 172**,** 446-52.

DILLIES, M. A., RAU, A., AUBERT, J., HENNEQUET-ANTIER, C., JEANMOUGIN, M., SERVANT, N., KEIME, C., MAROT, G., CASTEL, D., ESTELLE, J., GUERNEC, G., JAGLA, B., JOUNEAU, L., LALOE, D., LE GALL, C., SCHAEFFER, B., LE CROM, S., GUEDJ, M., JAFFREZIC, F. & FRENCH STATOMIQUE, C. 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief Bioinform,* 14**,** 671-83.

DUANE, P. G., RUBINS, J. B., WEISEL, H. R. & JANOFF, E. N. 1993. Identification of hydrogen peroxide as a *Streptococcus pneumoniae* toxin for rat alveolar epithelial cells. *Infect Immun,* 61**,** 4392-7.

DUKAN, S. & NYSTROM, T. 1999. Oxidative stress defense and deterioration of growth-arrested *Escherichia coli* cells. *J Biol Chem,* 274**,** 26027-32.

DURAND, J. M. & BJORK, G. R. 2003. Putrescine or a combination of methionine and arginine restores virulence gene expression in a tRNA modification-deficient mutant of *Shigella flexneri*: a possible role in adaptation of virulence. *Mol Microbiol,* 47**,** 519-27.

DURFEE, T., HANSEN, A. M., ZHI, H., BLATTNER, F. R. & JIN, D. J. 2008. Transcription profiling of the stringent response in *Escherichia coli*. *J Bacteriol,* 190**,** 1084-96.

DWORKIN, M. S., PARK, L. & BORCHARDT, S. M. 2007. The changing epidemiology of invasive *Haemophilus influenzae* disease, especially in persons > or = 65 years old. *Clin Infect Dis,* 44**,** 810-6.

EARL, C. S., KEONG, T. W., AN, S. Q., MURDOCH, S., MCCARTHY, Y., GARMENDIA, J., WARD, J., DOW, J. M., YANG, L., O'TOOLE, G. A. & RYAN, R. P. 2015. *Haemophilus influenzae* responds to glucocorticoids used in asthma therapy by modulation of biofilm formation and antibiotic resistance. *EMBO Mol Med,* 7**,** 1018-33.

EDWARDS, A. M. & MASSEY, R. C. 2011. Invasion of human cells by a bacterial pathogen. *J Vis Exp*.

ELWELL, L. P., DE GRAAFF, J., SEIBERT, D. & FALKOW, S. 1975. Plasmid-linked ampicillin resistance in *haempohilus influenza* type b. *Infect Immun,* 12**,** 404-10.

ERWIN, A. L., ALLEN, S., HO, D. K., BONTHUIS, P. J., JARISCH, J., NELSON, K. L., TSAO, D. L., UNRATH, W. C., WATSON, M. E., JR., GIBSON, B. W., APICELLA, M. A. & SMITH, A. L. 2006. Role of lgtC in resistance of nontypeable *Haemophilus influenzae* strain R2866 to human serum. *Infect Immun,* 74**,** 6226-35.

ERWIN, A. L., NELSON, K. L., MHLANGA-MUTANGADURA, T., BONTHUIS, P. J., GEELHOOD, J. L., MORLIN, G., UNRATH, W. C., CAMPOS, J., CROOK, D. W., FARLEY, M. M., HENDERSON, F. W., JACOBS, R. F., MUHLEMANN, K.,

SATOLA, S. W., VAN ALPHEN, L., GOLOMB, M. & SMITH, A. L. 2005. Characterization of genetic and phenotypic diversity of invasive nontypeable *Haemophilus influenzae*. *Infect Immun,* 73**,** 5853-63.

EVANS, D., MARQUEZ, S. M. & PACE, N. R. 2006. RNase P: interface of the RNA and protein worlds. *Trends Biochem Sci,* 31**,** 333-41.

EVERS, S., DI PADOVA, K., MEYER, M., FOUNTOULAKIS, M., KECK, W. & GRAY, C. P. 1998. Strategies towards a better understanding of antibiotic action: folate pathway inhibition in *Haemophilus influenzae* as an example. *Electrophoresis,* 19**,** 1980-8.

EYRE, D. W., GOLUBCHIK, T., GORDON, N. C., BOWDEN, R., PIAZZA, P., BATTY, E. M., IP, C. L., WILSON, D. J., DIDELOT, X., O'CONNOR, L., LAY, R., BUCK, D., KEARNS, A. M., SHAW, A., PAUL, J., WILCOX, M. H., DONNELLY, P. J., PETO, T. E., WALKER, A. S. & CROOK, D. W. 2012. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open,* 2**,** e001124.

FARJO, R. S., FOXMAN, B., PATEL, M. J., ZHANG, L., PETTIGREW, M. M., MCCOY, S. I., MARRS, C. F. & GILSDORF, J. R. 2004. Diversity and sharing of *Haemophilus influenzae* strains colonizing healthy children attending day-care centers. *Pediatr Infect Dis J,* 23**,** 41-6.

FAYET, O., ZIEGELHOFFER, T. & GEORGOPOULOS, C. 1989. The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J Bacteriol,* 171**,** 1379-85.

FINCH, P. W., CHAMBERS, P. & EMMERSON, P. T. 1985. Identification of the *Escherichia coli* recN gene product as a major SOS protein. *J Bacteriol,* 164**,** 653-8.

FINK, D. L., GREEN, B. A. & ST GEME, J. W., 3RD 2002. The *Haemophilus influenzae* Hap autotransporter binds to fibronectin, laminin, and collagen IV. *Infect Immun,* 70**,** 4902-7.

FLEISCHMANN, R. D., ADAMS, M. D., WHITE, O., CLAYTON, R. A., KIRKNESS, E. F., KERLAVAGE, A. R., BULT, C. J., TOMB, J. F., DOUGHERTY, B. A., MERRICK, J. M. & ET AL. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science,* 269**,** 496-512.

FLINT, D. H., SMYK-RANDALL, E., TUMINELLO, J. F., DRACZYNSKA-LUSIAK, B. & BROWN, O. R. 1993. The inactivation of dihydroxy-acid dehydratase in *Escherichia coli* treated with hyperbaric oxygen occurs because of the destruction of its Fe-S cluster, but the enzyme remains in the cell in a form that can be reactivated. *J Biol Chem,* 268**,** 25547-52.

FREDRIKSSON, A., BALLESTEROS, M., DUKAN, S. & NYSTROM, T. 2005. Defense against protein carbonylation by DnaK/DnaJ and proteases of the heat shock regulon. *J Bacteriol,* 187**,** 4207-13.

FURUCHI, T., KASHIWAGI, K., KOBAYASHI, H. & IGARASHI, K. 1991. Characteristics of the gene for a spermidine and putrescine transport system that maps at 15 min on the *Escherichia coli* chromosome. *J Biol Chem,* 266**,** 20928-33.

GAIMSTER, H., CAMA, J., HERNANDEZ-AINSA, S., KEYSER, U. F. & SUMMERS, D. K. 2014. The indole pulse: a new perspective on indole signalling in *Escherichia coli*. *PLoS One,* 9**,** e93168.

GARCIA-ALCALDE, F., OKONECHNIKOV, K., CARBONELL, J., CRUZ, L. M., GOTZ, S., TARAZONA, S., DOPAZO, J., MEYER, T. F. & CONESA, A. 2012. Qualimap: evaluating next-generation sequencing alignment data. *Bioinformatics,* 28**,** 2678-9.

GERCEKER, A. A., ZAIDI, T., MARKS, P., GOLAN, D. E. & PIER, G. B. 2000. Impact of heterogeneity within cultured cells on bacterial invasion: analysis of *Pseudomonas aeruginosa* and *Salmonella enterica* serovar typhi entry into MDCK cells by using a green fluorescent protein-labelled cystic fibrosis transmembrane conductance regulator receptor. *Infect Immun,* 68**,** 861-70.

GIANNOUKOS, G., CIULLA, D. M., HUANG, K., HAAS, B. J., IZARD, J., LEVIN, J. Z., LIVNY, J., EARL, A. M., GEVERS, D., WARD, D. V., NUSBAUM, C., BIRREN, B. W. & GNIRKE, A. 2012. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol,* 13**,** R23.

GILSDORF, J. R., TUCCI, M. & MARRS, C. F. 1996. Role of pili in *Haemophilus influenzae* adherence to, and internalization by, respiratory cells. *Pediatr Res,* 39**,** 343-8.

GOLEBIEWSKA, A., KUCH, A., GAWRONSKA, A., ALBRECHT, P., SKOCZYNSKA, A., RADZIKOWSKI, A., KUTYLOWSKA, E. & FELESZKO, W. 2016. Invasive *Haemophilus influenzae* Serotype f Case Reports in Mazovia Province, Poland. *Medicine (Baltimore),* 95**,** e2671.

GOMEZ-LOZANO, M., MARVIG, R. L., TULSTRUP, M. V. & MOLIN, S. 2014. Expression of antisense small RNAs in response to stress in *Pseudomonas aeruginosa*. *BMC Genomics,* 15**,** 783.

GONG, H., VU, G. P., BAI, Y., CHAN, E., WU, R., YANG, E., LIU, F. & LU, S. 2011. A *Salmonella* small non-coding RNA facilitates bacterial invasion and intracellular replication by modulating the expression of virulence factors. *PLoS Pathog,* 7**,** e1002120.

GONZALEZ-FLECHA, B. & DEMPLE, B. 1995. Metabolic sources of hydrogen peroxide in aerobically growing *Escherichia coli*. *J Biol Chem,* 270**,** 13681-7.

GORIS, J., KONSTANTINIDIS, K. T., KLAPPENBACH, J. A., COENYE, T., VANDAMME, P. & TIEDJE, J. M. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol,* 57**,** 81-91.

GRANICK, S. & GILDER, H. 1946. The Porphyrin Requirements of *Haemophilus influenzae* and Some Functions of the Vinyl and Propionic Acid Side Chains of Heme. *J Gen Physiol,* 30**,** 1-13.

GRAY-OWEN, S. D. & SCHRYVERS, A. B. 1995. Characterization of transferrin binding proteins 1 and 2 in invasive type b and nontypeable strains of *Haemophilus influenzae*. *Infect Immun,* 63**,** 3809-15.

GRIFFITHS-JONES, S., BATEMAN, A., MARSHALL, M., KHANNA, A. & EDDY, S. R. 2003. Rfam: an RNA family database. *Nucleic Acids Res,* 31**,** 439-41.

GUELL, M., YUS, E., LLUCH-SENAR, M. & SERRANO, L. 2011. Bacterial transcriptomics: what is beyond the RNA horiz-ome? *Nat Rev Microbiol,* 9**,** 658-69.

GUREVICH, A., SAVELIEV, V., VYAHHI, N. & TESLER, G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics,* 29**,** 1072-5.

GWINN, M. L., RAMANATHAN, R., SMITH, H. O. & TOMB, J. F. 1998. A new transformation-deficient mutant of *Haemophilus influenzae* Rd with normal DNA uptake. *J Bacteriol,* 180**,** 746-8.

HALL-STOODLEY, L., HU, F. Z., GIESEKE, A., NISTICO, L., NGUYEN, D., HAYES, J., FORBES, M., GREENBERG, D. P., DICE, B., BURROWS, A., WACKYM, P. A., STOODLEY, P., POST, J. C., EHRLICH, G. D. & KERSCHNER, J. E. 2006. Direct detection of bacterial biofilms on the middle-ear mucosa of children with chronic otitis media. *JAMA,* 296**,** 202-11.

HAMMER, B. A. & JOHNSON, E. A. 1988. Purification, properties, and metabolic roles of NAD+-glutamate dehydrogenase in *Clostridium botulinum* 113B. *Arch Microbiol,* 150**,** 460-4.

HARGREAVES, R. M., SLACK, M. P., HOWARD, A. J., ANDERSON, E. & RAMSAY, M. E. 1996. Changing patterns of invasive *Haemophilus influenzae* disease in England and Wales after introduction of the Hib vaccination programme. *BMJ,* 312**,** 160-1.

HARRISON, A., BAKALETZ, L. O. & MUNSON, R. S., JR. 2012. *Haemophilus influenzae* and oxidative stress. *Front Cell Infect Microbiol,* 2**,** 40.

HARRISON, A., BAKER, B. D. & MUNSON, R. S., JR. 2015. Overlapping and complementary oxidative stress defense mechanisms in nontypeable *Haemophilus influenzae*. *J Bacteriol,* 197**,** 277-85.

HARRISON, A., RAY, W. C., BAKER, B. D., ARMBRUSTER, D. W., BAKALETZ, L. O. & MUNSON, R. S., JR. 2007. The OxyR regulon in nontypeable *Haemophilus influenzae*. *J Bacteriol,* 189**,** 1004-12.

HARRISON, A., SANTANA, E. A., SZELESTEY, B. R., NEWSOM, D. E., WHITE, P. & MASON, K. M. 2013. Ferric uptake regulator and its role in the pathogenesis of nontypeable *Haemophilus influenzae*. *Infect Immun,* 81**,** 1221-33.

HASEGAWA, K., CHIBA, N., KOBAYASHI, R., MURAYAMA, S. Y., IWATA, S., SUNAKAWA, K. & UBUKATA, K. 2004. Rapidly increasing prevalence of beta-lactamase-nonproducing, ampicillin-resistant *Haemophilus influenzae* type b in patients with meningitis. *Antimicrob Agents Chemother,* 48**,** 1509-14.

HAYASHI, M., NAKAYAMA, Y. & UNEMOTO, T. 1996. Existence of Na+-translocating NADH-quinone reductase in *Haemophilus influenzae*. *FEBS Lett,* 381**,** 174-6.

HEATH, P. T., BOOY, R., AZZOPARDI, H. J., SLACK, M. P., BOWEN-MORRIS, J., GRIFFITHS, H., RAMSAY, M. E., DEEKS, J. J. & MOXON, E. R. 2000. Antibody concentration and clinical protection after Hib conjugate vaccination in the United Kingdom. *JAMA,* 284**,** 2334-40.

HEMPEL, R. J., MORTON, D. J., SEALE, T. W., WHITBY, P. W. & STULL, T. L. 2013. The role of the RNA chaperone Hfq in *Haemophilus influenzae* pathogenesis. *BMC Microbiol,* 13**,** 134.

HERRIOTT, R. M., MEYER, E. M. & VOGT, M. 1970. Defined nongrowth media for stage II development of competence in *Haemophilus influenzae*. *J Bacteriol,* 101**,** 517-24.

HILLAS, P. J., DEL ALBA, F. S., OYARZABAL, J., WILKS, A. & ORTIZ DE MONTELLANO, P. R. 2000. The AhpC and AhpD antioxidant defense system of *Mycobacterium tuberculosis*. *J Biol Chem,* 275**,** 18801-9.

HITZEMANN, R., BOTTOMLY, D., DARAKJIAN, P., WALTER, N., IANCU, O., SEARLES, R., WILMOT, B. & MCWEENEY, S. 2013. Genes, behavior and next-generation RNA sequencing. *Genes Brain Behav,* 12**,** 1-12.

HOEFLING, D. C. 1991. Acute myositis associated with *Hemophilus parasuis* in primary SPF sows. *J Vet Diagn Invest,* 3**,** 354-5.

HOFACKER, I. L. & STADLER, P. F. 2006. Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics,* 22**,** 1172-6.

HOGG, J. S., HU, F. Z., JANTO, B., BOISSY, R., HAYES, J., KEEFE, R., POST, J. C. & EHRLICH, G. D. 2007. Characterization and modeling of the *Haemophilus influenzae* core and supragenomes based on the complete genomic sequences of Rd and 12 clinical nontypeable strains. *Genome Biol,* 8**,** R103.

HOLMES, K. A. & BAKALETZ, L. O. 1997. Adherence of non-typeable *Haemophilus influenzae* promotes reorganization of the actin cytoskeleton in human or chinchilla epithelial cells in vitro. *Microb Pathog,* 23**,** 157-66.

HOOD, D. W., DEADMAN, M. E., ENGSKOG, M. K., VITIAZEVA, V., MAKEPEACE, K., SCHWEDA, E. K. & MOXON, R. 2010. Genes required for the synthesis of heptose-containing oligosaccharide outer core extensions in *Haemophilus influenzae* lipopolysaccharide. *Microbiology,* 156**,** 3421-31.

HOTTES, A. K., FREDDOLINO, P. L., KHARE, A., DONNELL, Z. N., LIU, J. C. & TAVAZOIE, S. 2013. Bacterial adaptation through loss of function. *PLoS Genet,* 9**,** e1003617.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009a. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res,* 37**,** 1-13.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

HUMPHRYS, M. S., CREASY, T., SUN, Y., SHETTY, A. C., CHIBUCOS, M. C., DRABEK, E. F., FRASER, C. M., FAROOQ, U., SENGAMALAY, N., OTT, S., SHOU, H., BAVOIL, P. M., MAHURKAR, A. & MYERS, G. S. 2013. Simultaneous transcriptional profiling of bacteria and their host cells. *PLoS One,* 8**,** e80597.

IGARASHI, K. & KASHIWAGI, K. 2006. Polyamine Modulon in *Escherichia coli*: genes involved in the stimulation of cell growth by polyamines. *J Biochem,* 139**,** 11-6.

INZANA, T. J., JOHNSON, J. L., SHELL, L., MOLLER, K. & KILIAN, M. 1992. Isolation and characterization of a newly identified *Haemophilus* species from cats: "*Haemophilus felis*". *J Clin Microbiol,* 30**,** 2108-12.

ISHAK, N., TIKHOMIROVA, A., BENT, S. J., EHRLICH, G. D., HU, F. Z. & KIDD, S. P. 2014. There is a specific response to pH by isolates of *Haemophilus influenzae* and this has a direct influence on biofilm formation. *BMC Microbiol,* 14**,** 47.

IWASA, M., KOBAYASHI, Y., MIFUJI-MOROKA, R., HARA, N., MIYACHI, H., SUGIMOTO, R., TANAKA, H., FUJITA, N., GABAZZA, E. C. & TAKEI, Y. 2013. Branched-chain amino acid supplementation reduces oxidative stress and prolongs survival in rats with advanced liver cirrhosis. *PLoS One,* 8**,** e70309.

IWASAKI, H., SHIBA, T., NAKATA, A. & SHINAGAWA, H. 1989. Involvement in DNA repair of the ruvA gene of *Escherichia coli*. *Mol Gen Genet,* 219**,** 328-31.

JACKSON, R. W., VINATZER, B., ARNOLD, D. L., DORUS, S. & MURILLO, J. 2011. The influence of the accessory genome on bacterial pathogen evolution. *Mob Genet Elements,* 1**,** 55-65.

JAROSIK, G. P. & HANSEN, E. J. 1994. Identification of a new locus involved in expression of *Haemophilus influenzae* type b lipooligosaccharide. *Infect Immun,* 62**,** 4861-7.

JAROSIK, G. P. & HANSEN, E. J. 1995. Cloning and sequencing of the *Haemophilus influenzae* exbB and exbD genes. *Gene,* 152**,** 89-92.

JAROSIK, G. P., MACIVER, I. & HANSEN, E. J. 1995. Utilization of transferrin-bound iron by *Haemophilus influenzae* requires an intact tonB gene. *Infect Immun,* 63**,** 710-3.

JAROSIK, G. P., SANDERS, J. D., COPE, L. D., MULLER-EBERHARD, U. & HANSEN, E. J. 1994. A functional tonB gene is required for both utilization of heme and virulence expression by *Haemophilus influenzae* type b. *Infect Immun,* 62**,** 2470-7.

JENKINS, D. E., SCHULTZ, J. E. & MATIN, A. 1988. Starvation-induced cross protection against heat or H2O2 challenge in *Escherichia coli*. *J Bacteriol,* 170**,** 3910-4.

JIANG, D., TIKHOMIROVA, A., BENT, S. J. & KIDD, S. P. 2016a. A discrete role for FNR in the transcriptional response to moderate changes in oxygen by *Haemophilus influenzae* Rd KW20. *Res Microbiol,* 167**,** 103-13.

JIANG, D., TIKHOMIROVA, A. & KIDD, S. P. 2016b. *Haemophilus influenzae* strains possess variations in the global transcriptional profile in response to oxygen levels and this influences sensitivity to environmental stresses. *Res Microbiol,* 167**,** 13-9.

JOHNSON, A. P. & INZANA, T. J. 1986. Loss of ciliary activity in organ cultures of rat trachea treated with lipo-oligosaccharide from *Haemophilus influenzae*. *J Med Microbiol,* 22**,** 265-8.

JONES, P. A., SAMUELS, N. M., PHILLIPS, N. J., MUNSON, R. S., JR., BOZUE, J. A., ARSENEAU, J. A., NICHOLS, W. A., ZALESKI, A., GIBSON, B. W. &

APICELLA, M. A. 2002. *Haemophilus influenzae* type b strain A2 has multiple sialyltransferases involved in lipooligosaccharide sialylation. *J Biol Chem,* 277**,** 14598-611.

JUHAS, M., POWER, P. M., HARDING, R. M., FERGUSON, D. J., DIMOPOULOU, I. D., ELAMIN, A. R., MOHD-ZAIN, Z., HOOD, D. W., ADEGBOLA, R., ERWIN, A., SMITH, A., MUNSON, R. S., HARRISON, A., MANSFIELD, L., BENTLEY, S. & CROOK, D. W. 2007. Sequence and functional analyses of *Haemophilus* spp. genomic islands. *Genome Biol,* 8**,** R237.

JULIAO, P. C., MARRS, C. F., XIE, J. & GILSDORF, J. R. 2007. Histidine auxotrophy in commensal and disease-causing nontypeable *Haemophilus influenzae*. *J Bacteriol,* 189**,** 4994-5001.

JUNEAU, R. A., PANG, B., ARMBRUSTER, C. E., MURRAH, K. A., PEREZ, A. C. & SWORDS, W. E. 2015. Peroxiredoxin-glutaredoxin and catalase promote resistance of nontypeable *Haemophilus influenzae* 86-028NP to oxidants and survival within neutrophil extracellular traps. *Infect Immun,* 83**,** 239-46.

KALIES, H., SIEDLER, A., GRONDAHL, B., GROTE, V., MILDE-BUSCH, A. & VON KRIES, R. 2009. Invasive *Haemophilus influenzae* infections in Germany: impact of non-type b serotypes in the post-vaccine era. *BMC Infect Dis,* 9**,** 45.

KANEMORI, M., NISHIHARA, K., YANAGI, H. & YURA, T. 1997. Synergistic roles of HslVU and other ATP-dependent proteases in controlling in vivo turnover of sigma32 and abnormal proteins in *Escherichia coli*. *J Bacteriol,* 179**,** 7219-25.

KARUDAPURAM, S., ZHAO, X. & BARCAK, G. J. 1995. DNA sequence and characterization of *Haemophilus influenzae* dprA+, a gene required for chromosomal but not plasmid DNA transformation. *J Bacteriol,* 177**,** 3235-40.

KASHIWAGI, K., MIYAMOTO, S., SUZUKI, F., KOBAYASHI, H. & IGARASHI, K. 1992. Excretion of putrescine by the putrescine-ornithine antiporter encoded by the potE gene of *Escherichia coli*. *Proc Natl Acad Sci U S A,* 89**,** 4529-33.

KAWAMOTO, H., KOIDE, Y., MORITA, T. & AIBA, H. 2006. Base-pairing requirement for RNA silencing by a bacterial small RNA and acceleration of duplex formation by Hfq. *Mol Microbiol,* 61**,** 1013-22.

KEYER, K. & IMLAY, J. A. 1996. Superoxide accelerates DNA damage by elevating free-iron levels. *Proc Natl Acad Sci U S A,* 93**,** 13635-40.

KILIAN, M. 1976. A taxonomic study of the genus *Haemophilus*, with the proposal of a new species. *J Gen Microbiol,* 93**,** 9-62.

KING, P. 2012. *Haemophilus influenzae* and the lung (*Haemophilus* and the lung). *Clin Transl Med,* 1**,** 10.

KOBOLDT, D. C., ZHANG, Q., LARSON, D. E., SHEN, D., MCLELLAN, M. D., LIN, L., MILLER, C. A., MARDIS, E. R., DING, L. & WILSON, R. K. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res,* 22**,** 568-76.

KOLEY, D. & BARD, A. J. 2010. Triton X-100 concentration effects on membrane permeability of a single HeLa cell by scanning electrochemical microscopy (SECM). *Proc Natl Acad Sci U S A,* 107**,** 16783-7.

KOLKER, E., PURVINE, S., GALPERIN, M. Y., STOLYAR, S., GOODLETT, D. R., NESVIZHSKII, A. I., KELLER, A., XIE, T., ENG, J. K., YI, E., HOOD, L., PICONE, A. F., CHERNY, T., TJADEN, B. C., SIEGEL, A. F., REILLY, T. J., MAKAROVA, K. S., PALSSON, B. O. & SMITH, A. L. 2003. Initial proteome analysis of model microorganism *Haemophilus influenzae* strain Rd KW20. *J Bacteriol,* 185**,** 4593-602.

KOLTER, R., SIEGELE, D. A. & TORMO, A. 1993. The stationary phase of the bacterial life cycle. *Annu Rev Microbiol,* 47**,** 855-74.

KORESSAAR, T. & REMM, M. 2007. Enhancements and modifications of primer design program Primer3. *Bioinformatics,* 23**,** 1289-91.

KOZMIN, S. G., STEPCHENKOVA, E. I., CHOW, S. C. & SCHAAPER, R. M. 2013. A critical role for the putative NCS2 nucleobase permease YjcD in the sensitivity of *Escherichia coli* to cytotoxic and mutagenic purine analogs. *MBio,* 4**,** e00661-13.

KRESS-BENNETT, J. M., HILLER, N. L., EUTSEY, R. A., POWELL, E., LONGWELL, M. J., HILLMAN, T., BLACKWELL, T., BYERS, B., MELL, J. C., POST, J. C., HU, F. Z., EHRLICH, G. D. & JANTO, B. A. 2016. Identification and

Characterization of msf, a Novel Virulence Factor in *Haemophilus influenzae*. *PLoS One,* 11**,** e0149891.

KROGER, C., COLGAN, A., SRIKUMAR, S., HANDLER, K., SIVASANKARAN, S. K., HAMMARLOF, D. L., CANALS, R., GRISSOM, J. E., CONWAY, T., HOKAMP, K. & HINTON, J. C. 2013. An infection-relevant transcriptomic compendium for *Salmonella enterica* Serovar Typhimurium. *Cell Host Microbe,* 14**,** 683-95.

KROLL, J. S., LANGFORD, P. R. & LOYNDS, B. M. 1991. Copper-zinc superoxide dismutase of *Haemophilus influenzae* and H. parainfluenzae. *J Bacteriol,* 173**,** 7449-57.

KROLL, J. S., LANGFORD, P. R., SAAH, J. R. & LOYNDS, B. M. 1993. Molecular and genetic characterization of superoxide dismutase in *Haemophilus influenzae* type b. *Mol Microbiol,* 10**,** 839-48.

KRZYWINSKI, M., SCHEIN, J., BIROL, I., CONNORS, J., GASCOYNE, R., HORSMAN, D., JONES, S. J. & MARRA, M. A. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res,* 19**,** 1639-45.

KUBIET, M., RAMPHAL, R., WEBER, A. & SMITH, A. 2000. Pilus-mediated adherence of *Haemophilus influenzae* to human respiratory mucins. *Infect Immun,* 68**,** 3362-7.

KULESUS, R. R., DIAZ-PEREZ, K., SLECHTA, E. S., ETO, D. S. & MULVEY, M. A. 2008. Impact of the RNA chaperone Hfq on the fitness and virulence potential of uropathogenic *Escherichia coli*. *Infect Immun,* 76**,** 3019-26.

KULLAS, A. L., MCCLELLAND, M., YANG, H. J., TAM, J. W., TORRES, A., PORWOLLIK, S., MENA, P., MCPHEE, J. B., BOGOMOLNAYA, L., ANDREWS-POLYMENIS, H. & VAN DER VELDEN, A. W. 2012. L-asparaginase II produced by *Salmonella typhimurium* inhibits T cell responses and mediates virulence. *Cell Host Microbe,* 12**,** 791-8.

KUMAR, R., LAWRENCE, M. L., WATT, J., COOKSEY, A. M., BURGESS, S. C. & NANDURI, B. 2012. RNA-seq based transcriptional map of bovine respiratory disease pathogen "*Histophilus somni* 2336". *PLoS One,* 7**,** e29435.

KUNG, V. L., OZER, E. A. & HAUSER, A. R. 2010. The accessory genome of *Pseudomonas aeruginosa*. *Microbiol Mol Biol Rev,* 74**,** 621-41.

LADHANI, S. N. 2012. Two decades of experience with the *Haemophilus influenzae* serotype b conjugate vaccine in the United Kingdom. *Clin Ther,* 34**,** 385-99.

LANGEREIS, J. D. & DE JONGE, M. I. 2015. Invasive Disease Caused by Nontypeable *Haemophilus influenzae*. *Emerg Infect Dis,* 21**,** 1711-8.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.

LECLERC, J. E., LI, B., PAYNE, W. L. & CEBULA, T. A. 1996. High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science,* 274**,** 1208-11.

LEE, E. H., LEWIS, R. F., MAKUMBI, I., KEKITIINWA, A., EDIAMU, T. D., BAZIBU, M., BRAKA, F., FLANNERY, B., ZUBER, P. L. & FEIKIN, D. R. 2008. *Haemophilus influenzae* type b conjugate vaccine is highly effective in the Ugandan routine immunization program: a case-control study. *Trop Med Int Health,* 13**,** 495-502.

LETOURNEAU, J., LEVESQUE, C., BERTHIAUME, F., JACQUES, M. & MOUREZ, M. 2011. In vitro assay of bacterial adhesion onto mammalian epithelial cells. *J Vis Exp*.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LIBEREK, K., GALITSKI, T. P., ZYLICZ, M. & GEORGOPOULOS, C. 1992. The DnaK chaperone modulates the heat shock response of *Escherichia coli* by binding to the sigma 32 transcription factor. *Proc Natl Acad Sci U S A,* 89**,** 3516-20.

LINN, S. 1978. The 1978 Nobel prize in physiology or medicine. *Science,* 202**,** 1069-71.

LITTLE, J. W., MOUNT, D. W. & YANISCH-PERRON, C. R. 1981. Purified lexA protein is a repressor of the recA and lexA genes. *Proc Natl Acad Sci U S A,* 78**,** 4199-203.

LIU, G., ZHOU, J., FU, Q. S. & WANG, J. 2009. The *Escherichia coli* azoreductase AzoR Is involved in resistance to thiol-specific stress caused by electrophilic quinones. *J Bacteriol,* 191**,** 6394-400.

LIU, M. Y., GUI, G., WEI, B., PRESTON, J. F., 3RD, OAKFORD, L., YUKSEL, U., GIEDROC, D. P. & ROMEO, T. 1997. The RNA molecule CsrB binds to the global regulatory protein CsrA and antagonizes its activity in *Escherichia coli*. *J Biol Chem,* 272**,** 17502-10.

LOBEL, L., SIGAL, N., BOROVOK, I., RUPPIN, E. & HERSKOVITS, A. A. 2012. Integrative genomic analysis identifies isoleucine and CodY as regulators of *Listeria monocytogenes* virulence. *PLoS Genet,* 8**,** e1002887.

LOEB, M. R., CONNOR, E. & PENNEY, D. 1988. A comparison of the adherence of fimbriated and nonfimbriated *Haemophilus influenzae* type b to human adenoids in organ culture. *Infect Immun,* 56**,** 484-9.

LOEWEN, P. C. & TRIGGS, B. L. 1984. Genetic mapping of katF, a locus that with katE affects the synthesis of a second catalase species in *Escherichia coli*. *J Bacteriol,* 160**,** 668-75.

LOMAN, N. J., CONSTANTINIDOU, C., CHAN, J. Z., HALACHEV, M., SERGEANT, M., PENN, C. W., ROBINSON, E. R. & PALLEN, M. J. 2012. High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol,* 10**,** 599-606.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol,* 15**,** 550.

MACFADYEN, L. P., CHEN, D., VO, H. C., LIAO, D., SINOTTE, R. & REDFIELD, R. J. 2001. Competence development by *Haemophilus influenzae* is regulated by the availability of nucleic acid precursors. *Mol Microbiol,* 40**,** 700-7.

MAKELA, P. H., TAKALA, A. K., PELTOLA, H. & ESKOLA, J. 1992. Epidemiology of invasive *Haemophilus influenzae* type b disease. *J Infect Dis,* 165 Suppl 1**,** S2-6.

MANDAL, M., BOESE, B., BARRICK, J. E., WINKLER, W. C. & BREAKER, R. R. 2003. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell,* 113**,** 577-86.

MANN, B., VAN OPIJNEN, T., WANG, J., OBERT, C., WANG, Y. D., CARTER, R., MCGOLDRICK, D. J., RIDOUT, G., CAMILLI, A., TUOMANEN, E. I. & ROSCH,

J. W. 2012. Control of virulence by small RNAs in *Streptococcus pneumoniae*. *PLoS Pathog,* 8**,** e1002788.

MARTIN, K., MORLIN, G., SMITH, A., NORDYKE, A., EISENSTARK, A. & GOLOMB, M. 1998. The tryptophanase gene cluster of *Haemophilus influenzae* type b: evidence for horizontal gene transfer. *J Bacteriol,* 180**,** 107-18.

MARTINEZ, A. & KOLTER, R. 1997. Protection of DNA during oxidative stress by the nonspecific DNA-binding protein Dps. *J Bacteriol,* 179**,** 5188-94.

MASKELL, D. J., SZABO, M. J., DEADMAN, M. E. & MOXON, E. R. 1992. The gal locus from *Haemophilus influenzae*: cloning, sequencing and the use of gal mutants to study lipopolysaccharide. *Mol Microbiol,* 6**,** 3051-63.

MASSE, E. & GOTTESMAN, S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proc Natl Acad Sci U S A,* 99**,** 4620-5.

MATIC, V., BOZDOGAN, B., JACOBS, M. R., UBUKATA, K. & APPELBAUM, P. C. 2003. Contribution of beta-lactamase and PBP amino acid substitutions to amoxicillin/clavulanate resistance in beta-lactamase-positive, amoxicillin/clavulanate-resistant *Haemophilus influenzae*. *J Antimicrob Chemother,* 52**,** 1018-21.

MCCLURE, R., BALASUBRAMANIAN, D., SUN, Y., BOBROVSKYY, M., SUMBY, P., GENCO, C. A., VANDERPOOL, C. K. & TJADEN, B. 2013. Computational analysis of bacterial RNA-Seq data. *Nucleic Acids Res,* 41**,** e140.

MCCONNELL, A., TAN, B., SCHEIFELE, D., HALPERIN, S., VAUDRY, W., LAW, B., EMBREE, J. & OF THE CANADIAN IMMUNIZATION MONITORING PROGRAM, A. 2007. Invasive infections caused by *Haemophilus influenzae* serotypes in twelve Canadian IMPACT centers, 1996-2001. *Pediatr Infect Dis J,* 26**,** 1025-31.

MCMEECHAN, A., LOVELL, M. A., COGAN, T. A., MARSTON, K. L., HUMPHREY, T. J. & BARROW, P. A. 2005. Glycogen production by different *Salmonella enterica* serotypes: contribution of functional glgC to virulence, intestinal colonization and environmental survival. *Microbiology,* 151**,** 3969-77.

MERLIN, C., GARDINER, G., DURAND, S. & MASTERS, M. 2002. The *Escherichia coli* metD locus encodes an ABC transporter which includes Abc (MetN), YaeE (MetI), and YaeC (MetQ). *J Bacteriol,* 184**,** 5513-7.

MHLANGA-MUTANGADURA, T., MORLIN, G., SMITH, A. L., EISENSTARK, A. & GOLOMB, M. 1998. Evolution of the major pilus gene cluster of *Haemophilus influenzae*. *J Bacteriol,* 180**,** 4693-703.

MITCHELL, A., CHANG, H. Y., DAUGHERTY, L., FRASER, M., HUNTER, S., LOPEZ, R., MCANULLA, C., MCMENAMIN, C., NUKA, G., PESSEAT, S., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., RATO, C., YONG, S. Y., BATEMAN, A., PUNTA, M., ATTWOOD, T. K., SIGRIST, C. J., REDASCHI, N., RIVOIRE, C., XENARIOS, I., KAHN, D., GUYOT, D., BORK, P., LETUNIC, I., GOUGH, J., OATES, M., HAFT, D., HUANG, H., NATALE, D. A., WU, C. H., ORENGO, C., SILLITOE, I., MI, H., THOMAS, P. D. & FINN, R. D. 2015. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res,* 43**,** D213-21.

MIZUNO, T., CHOU, M. Y. & INOUYE, M. 1984. A unique mechanism regulating gene expression: translational inhibition by a complementary RNA transcript (micRNA). *Proc Natl Acad Sci U S A,* 81**,** 1966-70.

MOHD-ZAIN, Z., TURNER, S. L., CERDENO-TARRAGA, A. M., LILLEY, A. K., INZANA, T. J., DUNCAN, A. J., HARDING, R. M., HOOD, D. W., PETO, T. E. & CROOK, D. W. 2004. Transferable antibiotic resistance elements in *Haemophilus influenzae* share a common evolutionary origin with a diverse family of syntenic genomic islands. *J Bacteriol,* 186**,** 8114-22.

MONASTA, L., RONFANI, L., MARCHETTI, F., MONTICO, M., VECCHI BRUMATTI, L., BAVCAR, A., GRASSO, D., BARBIERO, C. & TAMBURLINI, G. 2012. Burden of disease caused by otitis media: systematic review and global estimates. *PLoS One,* 7**,** e36226.

MOREY, P., CANO, V., MARTI-LLITERAS, P., LOPEZ-GOMEZ, A., REGUEIRO, V., SAUS, C., BENGOECHEA, J. A. & GARMENDIA, J. 2011. Evidence for a non-replicative intracellular stage of nontypable *Haemophilus influenzae* in epithelial cells. *Microbiology,* 157**,** 234-50.

MORFELDT, E., TAYLOR, D., VON GABAIN, A. & ARVIDSON, S. 1995. Activation of alpha-toxin translation in *Staphylococcus aureus* by the trans-encoded antisense RNA, RNAIII. *EMBO J,* 14**,** 4569-77.

MORGAN, G. J., HATFULL, G. F., CASJENS, S. & HENDRIX, R. W. 2002. Bacteriophage Mu genome sequence: analysis and comparison with Mu-

like prophages in *Haemophilus*, *Neisseria* and *Deinococcus*. *J Mol Biol,* 317**,** 337-59.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods,* 5**,** 621-8.

MORTON, D. J., TURMAN, E. J., HENSLEY, P. D., VANWAGONER, T. M., SEALE, T. W., WHITBY, P. W. & STULL, T. L. 2010. Identification of a siderophore utilization locus in nontypeable *Haemophilus influenzae*. *BMC Microbiol,* 10**,** 113.

MORTON, D. J., WHITBY, P. W., JIN, H., REN, Z. & STULL, T. L. 1999. Effect of multiple mutations in the hemoglobin- and hemoglobin-haptoglobin-binding proteins, HgpA, HgpB, and HgpC, of *Haemophilus influenzae* type b. *Infect Immun,* 67**,** 2729-39.

MOXON, E. R. & VAUGHN, K. A. 1981. The type b capsular polysaccharide as a virulence determinant of *Haemophilus influenzae*: studies using clinical isolates and laboratory transformants. *J Infect Dis,* 143**,** 517-24.

MURPHY, T. F. & BRAUER, A. L. 2011. Expression of urease by *Haemophilus influenzae* during human respiratory tract infection and role in survival in an acid environment. *BMC Microbiol,* 11**,** 183.

MURPHY, T. F., BRAUER, A. L., SCHIFFMACHER, A. T. & SETHI, S. 2004. Persistent colonization by *Haemophilus influenzae* in chronic obstructive pulmonary disease. *Am J Respir Crit Care Med,* 170**,** 266-72.

NAIR, S., MILOHANIC, E. & BERCHE, P. 2000. ClpC ATPase is required for cell adhesion and invasion of *Listeria monocytogenes*. *Infect Immun,* 68**,** 7061-8.

NAKAMURA, K., OSHIMA, T., MORIMOTO, T., IKEDA, S., YOSHIKAWA, H., SHIWA, Y., ISHIKAWA, S., LINAK, M. C., HIRAI, A., TAKAHASHI, H., ALTAF-UL-AMIN, M., OGASAWARA, N. & KANAYA, S. 2011. Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res,* 39**,** e90.

NAWROCKI, E. P., BURGE, S. W., BATEMAN, A., DAUB, J., EBERHARDT, R. Y., EDDY, S. R., FLODEN, E. W., GARDNER, P. P., JONES, T. A., TATE, J. & FINN, R. D. 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res,* 43**,** D130-7.

NIZET, V., COLINA, K. F., ALMQUIST, J. R., RUBENS, C. E. & SMITH, A. L. 1996. A virulent nonencapsulated *Haemophilus influenzae*. *J Infect Dis,* 173**,** 180-6.

NOEL, G. J., HOISETH, S. K. & EDELSON, P. J. 1992. Type b capsule inhibits ingestion of *Haemophilus influenzae* by murine macrophages: studies with isogenic encapsulated and unencapsulated strains. *J Infect Dis,* 166**,** 178-82.

NOGUEIRA, T. & SPRINGER, M. 2000. Post-transcriptional control by global regulators of gene expression in bacteria. *Curr Opin Microbiol,* 3**,** 154-8.

NORSKOV-LAURITSEN, N. 2014. Classification, identification, and clinical significance of *Haemophilus* and *Aggregatibacter* species with host specificity for humans. *Clin Microbiol Rev,* 27**,** 214-40.

NORSKOV-LAURITSEN, N., BRUUN, B., ANDERSEN, C. & KILIAN, M. 2012. Identification of haemolytic *Haemophilus* species isolated from human clinical specimens and description of *Haemophilus sputorum* sp. nov. *Int J Med Microbiol,* 302**,** 78-83.

O'NEILL, J. M., ST GEME, J. W., 3RD, CUTTER, D., ADDERSON, E. E., ANYANWU, J., JACOBS, R. F. & SCHUTZE, G. E. 2003. Invasive disease due to nontypeable *Haemophilus influenzae* among children in Arkansas. *J Clin Microbiol,* 41**,** 3064-9.

O'ROURKE, E. J., CHEVALIER, C., PINTO, A. V., THIBERGE, J. M., IELPI, L., LABIGNE, A. & RADICELLA, J. P. 2003. Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. *Proc Natl Acad Sci U S A,* 100**,** 2789-94.

O'SULLIVAN, D. J., MORRIS, J. & O'GARA, F. 1990. Identification of an additional ferric-siderophore uptake gene clustered with receptor, biosynthesis, and fur-like regulatory genes in fluorescent *Pseudomonas* sp. strain M114. *Appl Environ Microbiol,* 56**,** 2056-64.

PAGES, V., KOFFEL-SCHWARTZ, N. & FUCHS, R. P. 2003. recX, a new SOS gene that is co-transcribed with the recA gene in *Escherichia coli*. *DNA Repair (Amst),* 2**,** 273-84.

PANG, B., HONG, W., KOCK, N. D. & SWORDS, W. E. 2012. Dps promotes survival of nontypeable *Haemophilus influenzae* in biofilm communities in vitro and resistance to clearance in vivo. *Front Cell Infect Microbiol,* 2**,** 58.

PAPAKOSTAS, K., BOTOU, M. & FRILLINGOS, S. 2013. Functional identification of the hypoxanthine/guanine transporters YjcD and YgfQ and the adenine transporters PurP and YicO of *Escherichia coli* K-12. *J Biol Chem,* 288**,** 36827-40.

PAPENFORT, K., PODKAMINSKI, D., HINTON, J. C. & VOGEL, J. 2012. The ancestral SgrS RNA discriminates horizontally acquired *Salmonella* mRNAs through a single G-U wobble pair. *Proc Natl Acad Sci U S A,* 109**,** E757-64.

PAULI, G. & OVERATH, P. 1972. ato Operon: a highly inducible system for acetoacetate and butyrate degradation in *Escherichia coli*. *Eur J Biochem,* 29**,** 553-62.

PAUWELS, F., VERGAUWEN, B. & VAN BEEUMEN, J. J. 2004. Physiological characterization of *Haemophilus influenzae* Rd deficient in its glutathione-dependent peroxidase PGdx. *J Biol Chem,* 279**,** 12163-70.

PAUWELS, F., VERGAUWEN, B., VANROBAEYS, F., DEVREESE, B. & VAN BEEUMEN, J. J. 2003. Purification and characterization of a chimeric enzyme from *Haemophilus influenzae* Rd that exhibits glutathione-dependent peroxidase activity. *J Biol Chem,* 278**,** 16658-66.

PEANO, C., PIETRELLI, A., CONSOLANDI, C., ROSSI, E., PETITI, L., TAGLIABUE, L., DE BELLIS, G. & LANDINI, P. 2013. An efficient rRNA removal method for RNA sequencing in GC-rich bacteria. *Microb Inform Exp,* 3**,** 1.

PERICONE, C. D., OVERWEG, K., HERMANS, P. W. & WEISER, J. N. 2000. Inhibitory and bactericidal effects of hydrogen peroxide production by *Streptococcus pneumoniae* on other inhabitants of the upper respiratory tract. *Infect Immun,* 68**,** 3990-7.

PERICONE, C. D., PARK, S., IMLAY, J. A. & WEISER, J. N. 2003. Factors contributing to hydrogen peroxide resistance in *Streptococcus pneumoniae* include pyruvate oxidase (SpxB) and avoidance of the toxic effects of the fenton reaction. *J Bacteriol,* 185**,** 6815-25.

PFEIFER, Y., MEISINGER, I., BRECHTEL, K. & GROBNER, S. 2013. Emergence of a multidrug-resistant *Haemophilus influenzae* strain causing chronic pneumonia in a patient with common variable immunodeficiency. *Microb Drug Resist,* 19**,** 1-5.

PFEIFFER, R. 1892. I.-Preliminary Communication on the Exciting causes of Influenza. *Br Med J,* 1**,** 128.

PFEIFFER, V., PAPENFORT, K., LUCCHINI, S., HINTON, J. C. & VOGEL, J. 2009. Coding sequence targeting by MicC RNA reveals bacterial mRNA silencing downstream of translational initiation. *Nat Struct Mol Biol,* 16**,** 840-6.

PITTMAN, M. 1931. Variation and Type Specificity in the Bacterial Species *Hemophilus Influenzae*. *J Exp Med,* 53**,** 471-92.

PITTMAN, M. 1953. A classification of the hemolytic bacteria of the genus *Haemophilus*: *Haemophilus haemolyticus* Bergey et al. and *Haemophilus parahaemolyticus* nov spec. *J Bacteriol,* 65**,** 750-1.

PITTMAN, M. & DAVIS, D. J. 1950. Identification of the Koch-Weeks bacillus (*Hemophilus aegyptius*). *J Bacteriol,* 59**,** 413-26.

POSTLE, K. 1990. TonB and the gram-negative dilemma. *Mol Microbiol,* 4**,** 2019-25.

PRASADARAO, N. V., LYSENKO, E., WASS, C. A., KIM, K. S. & WEISER, J. N. 1999. Opacity-associated protein A contributes to the binding of *Haemophilus influenzae* to chang epithelial cells. *Infect Immun,* 67**,** 4153-60.

PRYSAK, M. H., MOZDZIERZ, C. J., COOK, A. M., ZHU, L., ZHANG, Y., INOUYE, M. & WOYCHIK, N. A. 2009. Bacterial toxin YafQ is an endoribonuclease that associates with the ribosome and blocks translation elongation through sequence-specific and frame-dependent mRNA cleavage. *Mol Microbiol,* 71**,** 1071-87.

PULVERMACHER, S. C., STAUFFER, L. T. & STAUFFER, G. V. 2008. The role of the small regulatory RNA GcvB in GcvB/mRNA posttranscriptional regulation of oppA and dppA in *Escherichia coli*. *FEMS Microbiol Lett,* 281**,** 42-50.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

RADMAN, M. 1975. SOS repair hypothesis: phenomenology of an inducible DNA repair which is accompanied by mutagenesis. *Basic Life Sci,* 5A**,** 355-67.

RAFFEL, F. K., SZELESTEY, B. R., BEATTY, W. L. & MASON, K. M. 2013. The *Haemophilus influenzae* Sap transporter mediates bacterium-epithelial cell homeostasis. *Infect Immun,* 81**,** 43-54.

RAU, M. H., BOJANOVIC, K., NIELSEN, A. T. & LONG, K. S. 2015. Differential expression of small RNAs under chemical stress and fed-batch fermentation in *E. coli*. *BMC Genomics,* 16**,** 1051.

REDDY, M. S., BERNSTEIN, J. M., MURPHY, T. F. & FADEN, H. S. 1996. Binding between outer membrane proteins of nontypeable *Haemophilus influenzae* and human nasopharyngeal mucin. *Infect Immun,* 64**,** 1477-9.

REDFIELD, R. J., CAMERON, A. D., QIAN, Q., HINDS, J., ALI, T. R., KROLL, J. S. & LANGFORD, P. R. 2005. A novel CRP-dependent regulon controls expression of competence genes in *Haemophilus influenzae*. *J Mol Biol,* 347**,** 735-47.

REIS, J. N., LIMA, J. B., RIBEIRO, G. S., CORDERIO, S. M., SALGADO, K., REIS, M. G. & KO, A. I. 2002. Antimicrobial resistance in *Haemophilus influenzae* isolated during population-based surveillance for meningitis in Salvador, Brazil. *Antimicrob Agents Chemother,* 46**,** 3641-3.

REITZ, S., ALHAPEL, A., ESSEN, L. O. & PIERIK, A. J. 2008. Structural and kinetic properties of a beta-hydroxyacid dehydrogenase involved in nicotinate fermentation. *J Mol Biol,* 382**,** 802-11.

REUTER, J. A., SPACEK, D. V. & SNYDER, M. P. 2015. High-throughput sequencing technologies. *Mol Cell,* 58**,** 586-97.

RICCA, E., LIMAURO, D., LAGO, C. T. & DE FELICE, M. 1988. Enhanced acetohydroxy acid synthase III activity in an ilvH mutant of *Escherichia coli* K-12. *J Bacteriol,* 170**,** 5197-9.

RICHARDSON, E. J. & WATSON, M. 2013. The automatic annotation of bacterial genomes. *Brief Bioinform,* 14**,** 1-12.

RIEDEL, T. E., BERELSON, W. M., NEALSON, K. H. & FINKEL, S. E. 2013. Oxygen consumption rates of bacteria under nutrient-limited conditions. *Appl Environ Microbiol,* 79**,** 4921-31.

RISSMAN, A. I., MAU, B., BIEHL, B. S., DARLING, A. E., GLASNER, J. D. & PERNA, N. T. 2009. Reordering contigs of draft genomes using the Mauve aligner. *Bioinformatics,* 25**,** 2071-3.

RODRIGUEZ-MARTINEZ, J. M., LOPEZ, L., GARCIA, I. & PASCUAL, A. 2006. Characterization of a clinical isolate of *Haemophilus influenzae* with a high level of fluoroquinolone resistance. *J Antimicrob Chemother,* 57**,** 577-8.

ROGERS, Y. C., MUNK, A. C., MEINCKE, L. J. & HAN, C. S. 2005. Closing bacterial genomic sequence gaps with adaptor-PCR. *Biotechniques,* 39**,** 31-2, 34.

ROHWER, F. & AZAM, F. 2000. Detection of DNA damage in prokaryotes by terminal deoxyribonucleotide transferase-mediated dUTP nick end labeling. *Appl Environ Microbiol,* 66**,** 1001-6.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M. A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-5.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-7.

SANTANA, E. A., HARRISON, A., ZHANG, X., BAKER, B. D., KELLY, B. J., WHITE, P., LIU, Y. & MUNSON, R. S., JR. 2014. HrrF is the Fur-regulated small RNA in nontypeable *Haemophilus influenzae*. *PLoS One,* 9**,** e105644.

SCHEIFELE, D. W., FUSSELL, S. J. & ROBERTS, M. C. 1982. Characterization of ampicillin-resistant *Haemophilus parainfluenzae*. *Antimicrob Agents Chemother,* 21**,** 734-9.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SCRIVER, S. R., WALMSLEY, S. L., KAU, C. L., HOBAN, D. J., BRUNTON, J., MCGEER, A., MOORE, T. C. & WITWICKI, E. 1994. Determination of antimicrobial susceptibilities of Canadian isolates of *Haemophilus influenzae* and characterization of their beta-lactamases. Canadian Haemophilus Study Group. *Antimicrob Agents Chemother,* 38**,** 1678-80.

SEAVER, L. C. & IMLAY, J. A. 2001. Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in *Escherichia coli*. *J Bacteriol,* 183**,** 7173-81.

SEEMAN, P., CHENG, D. & ILES, G. H. 1973. Structure of membrane holes in osmotic and saponin hemolysis. *J Cell Biol,* 56**,** 519-27.

SEEMANN, T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics,* 30**,** 2068-9.

SETLOW, J. K., SPIKES, D. & GRIFFIN, K. 1988. Characterization of the rec-1 gene of *Haemophilus influenzae* and behavior of the gene in *Escherichia coli*. *J Bacteriol,* 170**,** 3876-81.

SHARMA, C. M., HOFFMANN, S., DARFEUILLE, F., REIGNIER, J., FINDEISS, S., SITTKA, A., CHABAS, S., REICHE, K., HACKERMULLER, J., REINHARDT, R., STADLER, P. F. & VOGEL, J. 2010. The primary transcriptome of the major human pathogen *Helicobacter pylori*. *Nature,* 464**,** 250-5.

SHARMA, C. M., PAPENFORT, K., PERNITZSCH, S. R., MOLLENKOPF, H. J., HINTON, J. C. & VOGEL, J. 2011. Pervasive post-transcriptional control of genes involved in amino acid metabolism by the Hfq-dependent GcvB small RNA. *Mol Microbiol,* 81**,** 1144-65.

SHENDURE, J. & JI, H. 2008. Next-generation DNA sequencing. *Nat Biotechnol,* 26**,** 1135-45.

SHUEL, M., HOANG, L., LAW, D. K. & TSANG, R. 2011. Invasive *Haemophilus influenzae* in British Columbia: non-Hib and non-typeable strains causing disease in children and adults. *Int J Infect Dis,* 15**,** e167-73.

SILBY, M. W. & LEVY, S. B. 2008. Overlapping protein-encoding genes in Pseudomonas fluorescens Pf0-1. *PLoS Genet,* 4**,** e1000094.

SINGH, B., BRANT, M., KILIAN, M., HALLSTROM, B. & RIESBECK, K. 2010. Protein E of *Haemophilus influenzae* is a ubiquitous highly conserved adhesin. *J Infect Dis,* 201**,** 414-9.

SINGH, N. K., KUNDE, D. A. & TRISTRAM, S. G. 2016. Effect of epithelial cell type on in vitro invasion of non-typeable *Haemophilus influenzae*. *J Microbiol Methods,* 129**,** 66-69.

SINHA, S., MELL, J. & REDFIELD, R. 2013. The availability of purine nucleotides regulates natural competence by controlling translation of the competence activator Sxy. *Mol Microbiol,* 88**,** 1106-19.

SIT, B., CROWLEY, S. M., BHULLAR, K., LAI, C. C., TANG, C., HOODA, Y., CALMETTES, C., KHAMBATI, H., MA, C., BRUMELL, J. H., SCHRYVERS, A. B., VALLANCE, B. A. & MORAES, T. F. 2015. Active Transport of Phosphorylated Carbohydrates Promotes Intestinal Colonization and Transmission of a Bacterial Pathogen. *PLoS Pathog,* 11**,** e1005107.

SITTKA, A., PFEIFFER, V., TEDIN, K. & VOGEL, J. 2007. The RNA chaperone Hfq is essential for the virulence of *Salmonella typhimurium*. *Mol Microbiol,* 63**,** 193-217.

SKAARE, D., ANTHONISEN, I. L., KAHLMETER, G., MATUSCHEK, E., NATAS, O. B., STEINBAKK, M., SUNDSFJORD, A. & KRISTIANSEN, B. E. 2014. Emergence of clonally related multidrug resistant *Haemophilus influenzae* with penicillin-binding protein 3-mediated resistance to extended-spectrum cephalosporins, Norway, 2006 to 2013. *Euro Surveill,* 19.

SKURNIK, M. & TOIVANEN, P. 1991. Intervening sequences (IVSs) in the 23S ribosomal RNA genes of pathogenic *Yersinia enterocolitica* strains. The IVSs in *Y. enterocolitica* and *Salmonella typhimurium* have a common origin. *Mol Microbiol,* 5**,** 585-93.

SMITH, C. B., GOLDEN, C., KLAUBER, M. R., KANNER, R. & RENZETTI, A. 1976. Interactions between viruses and bacteria in patients with chronic bronchitis. *J Infect Dis,* 134**,** 552-61.

SMITH, H. O. & WILCOX, K. W. 1970. A restriction enzyme from *Hemophilus influenzae*. I. Purification and general properties. *J Mol Biol,* 51**,** 379-91.

SOLOMON, J. M. & GROSSMAN, A. D. 1996. Who's competent and when: regulation of natural genetic competence in bacteria. *Trends Genet,* 12**,** 150-5.

SONDERGAARD, A., WITHERDEN, E. A., NORSKOV-LAURITSEN, N. & TRISTRAM, S. G. 2015. Interspecies transfer of the penicillin-binding protein 3-encoding gene ftsI between *Haemophilus influenzae* and *Haemophilus haemolyticus* can confer reduced susceptibility to beta-lactam antimicrobial agents. *Antimicrob Agents Chemother,* 59**,** 4339-42.

SONG, X. M., FORSGREN, A. & JANSON, H. 1999. Fragmentation heterogeneity of 23S ribosomal RNA in *Haemophilus* species. *Gene,* 230**,** 287-93.

SONG, X. M. & JANSON, H. 2003. Differences in genetic and transcriptional organization of the glpTQ operons between *Haemophilus influenzae* type b and nontypeable strains. *J Bacteriol,* 185**,** 7285-90.

SOUPENE, E., VAN HEESWIJK, W. C., PLUMBRIDGE, J., STEWART, V., BERTENTHAL, D., LEE, H., PRASAD, G., PALIY, O., CHARERNNOPPAKUL, P. & KUSTU, S. 2003. Physiological studies of *Escherichia coli* strain MG1655: growth defects and apparent cross-regulation of gene expression. *J Bacteriol,* 185**,** 5611-26.

SPANIOL, V., WYDER, S. & AEBI, C. 2013. RNA-Seq-based analysis of the physiologic cold shock-induced changes in *Moraxella catarrhalis* gene expression. *PLoS One,* 8**,** e68298.

SRIDHAR, J., SAMBATURU, N., SABARINATHAN, R., OU, H. Y., DENG, Z., SEKAR, K., RAFI, Z. A. & RAJAKUMAR, K. 2010. sRNAscanner: a computational tool for intergenic small RNA detection in bacterial genomes. *PLoS One,* 5**,** e11970.

SRIKHANTA, Y. N., MAGUIRE, T. L., STACEY, K. J., GRIMMOND, S. M. & JENNINGS, M. P. 2005. The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes. *Proc Natl Acad Sci U S A,* 102**,** 5547-51.

ST GEME, J. W., 3RD & CUTTER, D. 1996. Influence of pili, fibrils, and capsule on in vitro adherence by *Haemophilus influenzae* type b. *Mol Microbiol,* 21**,** 21-31.

ST GEME, J. W., 3RD & CUTTER, D. 2000. The *Haemophilus influenzae* Hia adhesin is an autotransporter protein that remains uncleaved at the C terminus and fully cell associated. *J Bacteriol,* 182**,** 6005-13.

ST GEME, J. W., 3RD, DE LA MORENA, M. L. & FALKOW, S. 1994. A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. *Mol Microbiol,* 14**,** 217-33.

ST GEME, J. W., 3RD & FALKOW, S. 1990. *Haemophilus influenzae* adheres to and enters cultured human epithelial cells. *Infect Immun,* 58**,** 4036-44.

ST GEME, J. W., 3RD & FALKOW, S. 1991. Loss of capsule expression by *Haemophilus influenzae* type b results in enhanced adherence to and invasion of human cells. *Infect Immun,* 59**,** 1325-33.

ST GEME, J. W., 3RD, FALKOW, S. & BARENKAMP, S. J. 1993. High-molecular-weight proteins of nontypable *Haemophilus influenzae* mediate attachment to human epithelial cells. *Proc Natl Acad Sci U S A,* 90**,** 2875-9.

STAHL, M., FRIIS, L. M., NOTHAFT, H., LIU, X., LI, J., SZYMANSKI, C. M. & STINTZI, A. 2011. L-fucose utilization provides *Campylobacter jejuni* with a competitive advantage. *Proc Natl Acad Sci U S A,* 108**,** 7194-9.

STEWART, V. & BLEDSOE, P. J. 2005. Fnr-, NarP- and NarL-dependent regulation of transcription initiation from the *Haemophilus influenzae* Rd napF (periplasmic nitrate reductase) promoter in *Escherichia coli* K-12. *J Bacteriol,* 187**,** 6928-35.

STOTHARD, P. & WISHART, D. S. 2006. Automated bacterial genome analysis and annotation. *Curr Opin Microbiol,* 9**,** 505-10.

SUTTON, A., SCHNEERSON, R., KENDALL-MORRIS, S. & ROBBINS, J. B. 1982. Differential complement resistance mediates virulence of *Haemophilus influenzae* type b. *Infect Immun,* 35**,** 95-104.

SWEETMAN, W. A., MOXON, E. R. & BAYLISS, C. D. 2005. Induction of the SOS regulon of *Haemophilus influenzae* does not affect phase variation rates at tetranucleotide or dinucleotide repeats. *Microbiology,* 151**,** 2751-63.

SWORDS, W. E., BUSCHER, B. A., VER STEEG II, K., PRESTON, A., NICHOLS, W. A., WEISER, J. N., GIBSON, B. W. & APICELLA, M. A. 2000. Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol Microbiol,* 37**,** 13-27.

TAKAHATA, S., IDA, T., SENJU, N., SANBONGI, Y., MIYATA, A., MAEBASHI, K. & HOSHIKO, S. 2007. Horizontal gene transfer of ftsI, encoding penicillin-binding protein 3, in *Haemophilus influenzae*. *Antimicrob Agents Chemother,* 51**,** 1589-95.

TAMARIT, J., CABISCOL, E. & ROS, J. 1998. Identification of the major oxidatively damaged proteins in *Escherichia coli* cells exposed to oxidative stress. *J Biol Chem,* 273**,** 3027-32.

TARGOWSKI, S. & TARGOWSKI, H. 1979. Characterization of a *Haemophilus paracuniculus* isolated from gastrointestinal tracts of rabbits with mucoid enteritis. *J Clin Microbiol,* 9**,** 33-7.

TIKHOMIROVA, A., JIANG, D. & KIDD, S. P. 2015a. A new insight into the role of intracellular nickel levels for the stress response, surface properties and twitching motility by *Haemophilus influenzae*. *Metallomics,* 7**,** 650-61.

TIKHOMIROVA, A., TRAPPETTI, C., PATON, J. C. & KIDD, S. P. 2015b. The outcome of *H. influenzae* and *S. pneumoniae* inter-species interactions depends on pH, nutrient availability and growth phase. *Int J Med Microbiol,* 305**,** 881-92.

TIRADO-LEE, L., LEE, A., REES, D. C. & PINKETT, H. W. 2011. Classification of a *Haemophilus influenzae* ABC transporter HI1470/71 through its cognate molybdate periplasmic binding protein, MolA. *Structure,* 19**,** 1701-10.

TOBE, T., SASAKAWA, C., OKADA, N., HONMA, Y. & YOSHIKAWA, M. 1992. vacB, a novel chromosomal gene required for expression of virulence genes on the large plasmid of *Shigella flexneri*. *J Bacteriol,* 174**,** 6359-67.

TOMB, J. F., EL-HAJJ, H. & SMITH, H. O. 1991. Nucleotide sequence of a cluster of genes involved in the transformation of *Haemophilus influenzae* Rd. *Gene,* 104**,** 1-10.

TOMIZAWA, J., ITOH, T., SELZER, G. & SOM, T. 1981. Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc Natl Acad Sci U S A,* 78**,** 1421-5.

TONDELLA, M. L., QUINN, F. D. & PERKINS, B. A. 1995. Brazilian purpuric fever caused by *Haemophilus influenzae* biogroup aegyptius strains lacking the 3031 plasmid. *J Infect Dis,* 171**,** 209-12.

TOUATI, D. 2000. Iron and oxidative stress in bacteria. *Arch Biochem Biophys,* 373**,** 1-6.

TREANGEN, T. J. & SALZBERG, S. L. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet,* 13**,** 36-46.

TRISTRAM, S., JACOBS, M. R. & APPELBAUM, P. C. 2007. Antimicrobial resistance in *Haemophilus influenzae*. *Clin Microbiol Rev,* 20**,** 368-89.

TSANG, R. S., LI, Y. A., MULLEN, A., BAIKIE, M., WHYTE, K., SHUEL, M., TYRRELL, G., ROTONDO, J. A., DESAI, S. & SPIKA, J. 2016. Laboratory characterization of invasive *Haemophilus influenzae* isolates from Nunavut, Canada, 2000-2012. *Int J Circumpolar Health,* 75**,** 29798.

TSAO, D., NELSON, K. L., KIM, D. & SMITH, A. L. 2012. Infant rat infection modifies phenotypic properties of an invasive nontypeable *Haemophilus influenzae*. *Microbes Infect,* 14**,** 509-16.

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40**,** e115.

VAN HAM, S. M., VAN ALPHEN, L., MOOI, F. R. & VAN PUTTEN, J. P. 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell,* 73**,** 1187-96.

VAN SCHILFGAARDE, M., VAN ALPHEN, L., EIJK, P., EVERTS, V. & DANKERT, J. 1995. Paracytosis of *Haemophilus influenzae* through cell layers of NCI-H292 lung epithelial cells. *Infect Immun,* 63**,** 4729-37.

VAN WESSEL, K., RODENBURG, G. D., VEENHOVEN, R. H., SPANJAARD, L., VAN DER ENDE, A. & SANDERS, E. A. 2011. Nontypeable *Haemophilus influenzae* invasive disease in The Netherlands: a retrospective surveillance study 2001-2008. *Clin Infect Dis,* 53**,** e1-7.

VANWAGONER, T. M., ATACK, J. M., NELSON, K. L., SMITH, H. K., FOX, K. L., JENNINGS, M. P., STULL, T. L. & SMITH, A. L. 2016. The modA10 phasevarion of nontypeable *Haemophilus influenzae* R2866 regulates multiple virulence-associated traits. *Microb Pathog,* 92**,** 60-7.

VASU, K. & NAGARAJA, V. 2013. Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol Mol Biol Rev,* 77**,** 53-72.

VEAL, W. L. & SHAFER, W. M. 2003. Identification of a cell envelope protein (MtrF) involved in hydrophobic antimicrobial resistance in *Neisseria gonorrhoeae*. *J Antimicrob Chemother,* 51**,** 27-37.

VERGAUWEN, B., HERBERT, M. & VAN BEEUMEN, J. J. 2006. Hydrogen peroxide scavenging is not a virulence determinant in the pathogenesis of *Haemophilus influenzae* type b strain Eagan. *BMC Microbiol,* 6**,** 3.

VERGAUWEN, B., PAUWELS, F., JACQUEMOTTE, F., MEYER, T. E., CUSANOVICH, M. A., BARTSCH, R. G. & VAN BEEUMEN, J. J. 2001. Characterization of clutathione amide reductase from *Chromatium gracile*. Identification of a novel thiol peroxidase (Prx/Grx) fueled by glutathione amide redox cycling. *J Biol Chem,* 276**,** 20890-7.

VERGAUWEN, B., PAUWELS, F. & VAN BEEUMEN, J. J. 2003a. Glutathione and catalase provide overlapping defenses for protection against respiration-generated hydrogen peroxide in *Haemophilus influenzae*. *J Bacteriol,* 185**,** 5555-62.

VERGAUWEN, B., PAUWELS, F., VANEECHOUTTE, M. & VAN BEEUMEN, J. J. 2003b. Exogenous glutathione completes the defense against oxidative stress in *Haemophilus influenzae*. *J Bacteriol,* 185**,** 1572-81.

VIDA, T. A. & EMR, S. D. 1995. A new vital stain for visualizing vacuolar membrane dynamics and endocytosis in yeast. *J Cell Biol,* 128**,** 779-92.

WAGNER, G. P., KIN, K. & LYNCH, V. J. 2012. Measurement of mRNA abundance using RNA-seq data: RPKM measure is inconsistent among samples. *Theory Biosci,* 131**,** 281-5.

WASSARMAN, K. M. & STORZ, G. 2000. 6S RNA regulates *E. coli* RNA polymerase activity. *Cell,* 101**,** 613-23.

WATERS, L. S. & STORZ, G. 2009. Regulatory RNAs in bacteria. *Cell,* 136**,** 615-28.

WEISER, J. N., CHONG, S. T., GREENBERG, D. & FONG, W. 1995. Identification and characterization of a cell envelope protein of *Haemophilus influenzae* contributing to phase variation in colony opacity and nasopharyngeal colonization. *Mol Microbiol,* 17**,** 555-64.

WEISER, J. N., LOVE, J. M. & MOXON, E. R. 1989. The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. *Cell,* 59**,** 657-65.

WESTERMANN, A. J., FORSTNER, K. U., AMMAN, F., BARQUIST, L., CHAO, Y., SCHULTE, L. N., MULLER, L., REINHARDT, R., STADLER, P. F. & VOGEL, J. 2016. Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions. *Nature,* 529**,** 496-501.

WHITBY, P. W., MORTON, D. J., VANWAGONER, T. M., SEALE, T. W., COLE, B. K., MUSSA, H. J., MCGHEE, P. A., BAUER, C. Y., SPRINGER, J. M. & STULL, T. L.

2012. *Haemophilus influenzae* OxyR: characterization of its regulation, regulon and role in fitness. *PLoS One,* 7**,** e50588.

WHITBY, P. W., SEALE, T. W., VANWAGONER, T. M., MORTON, D. J. & STULL, T. L. 2009. The iron/heme regulated genes of *Haemophilus influenzae*: comparative transcriptional profiling as a tool to define the species core modulon. *BMC Genomics,* 10**,** 6.

WHITBY, P. W., VANWAGONER, T. M., SEALE, T. W., MORTON, D. J. & STULL, T. L. 2006. Transcriptional profile of *Haemophilus influenzae*: effects of iron and heme. *J Bacteriol,* 188**,** 5640-5.

WHITBY, P. W., VANWAGONER, T. M., SEALE, T. W., MORTON, D. J. & STULL, T. L. 2013. Comparison of transcription of the *Haemophilus influenzae* iron/heme modulon genes in vitro and in vivo in the chinchilla middle ear. *BMC Genomics,* 14**,** 925.

WHITE, D. C. & GRANICK, S. 1963. Hemin Biosynthesis in *Haemophilus*. *J Bacteriol,* 85**,** 842-50.

WILLIAMS, B. J., GOLOMB, M., PHILLIPS, T., BROWNLEE, J., OLSON, M. V. & SMITH, A. L. 2002. Bacteriophage HP2 of *Haemophilus influenzae*. *J Bacteriol,* 184**,** 6893-905.

WILLIAMS, B. J., MORLIN, G., VALENTINE, N. & SMITH, A. L. 2001. Serum resistance in an invasive, nontypeable *Haemophilus influenzae* strain. *Infect Immun,* 69**,** 695-705.

WINSLOW, C. E., BROADHURST, J., BUCHANAN, R. E., KRUMWIEDE, C., ROGERS, L. A. & SMITH, G. H. 1920. The Families and Genera of the Bacteria: Final Report of the Committee of the Society of American Bacteriologists on Characterization and Classification of Bacterial Types. *J Bacteriol,* 5**,** 191-229.

WISSENBACH, U., SIX, S., BONGAERTS, J., TERNES, D., STEINWACHS, S. & UNDEN, G. 1995. A third periplasmic transport system for L-arginine in *Escherichia coli*: molecular characterization of the artPIQMJ genes, arginine binding and transport. *Mol Microbiol,* 17**,** 675-86.

WOLLSTEIN, M. 1911. Serum Treatment of Influenzal Meningitis. *J Exp Med,* 14**,** 73-82.

WONG, S. M., ALUGUPALLI, K. R., RAM, S. & AKERLEY, B. J. 2007. The ArcA regulon and oxidative stress resistance in *Haemophilus influenzae*. *Mol Microbiol,* 64**,** 1375-90.

WONG, S. M., BERNUI, M., SHEN, H. & AKERLEY, B. J. 2013. Genome-wide fitness profiling reveals adaptations required by *Haemophilus* in coinfection with influenza A virus in the murine lung. *Proc Natl Acad Sci U S A,* 110**,** 15413-8.

WONG, S. M., ST MICHAEL, F., COX, A., RAM, S. & AKERLEY, B. J. 2011. ArcA-regulated glycosyltransferase lic2B promotes complement evasion and pathogenesis of nontypeable *Haemophilus influenzae*. *Infect Immun,* 79**,** 1971-83.

WRIGHT, P. R., GEORG, J., MANN, M., SORESCU, D. A., RICHTER, A. S., LOTT, S., KLEINKAUF, R., HESS, W. R. & BACKOFEN, R. 2014. CopraRNA and IntaRNA: predicting small RNA targets, networks and interaction domains. *Nucleic Acids Res,* 42**,** W119-23.

WRIGHT, P. R., RICHTER, A. S., PAPENFORT, K., MANN, M., VOGEL, J., HESS, W. R., BACKOFEN, R. & GEORG, J. 2013. Comparative genomics boosts target prediction for bacterial small RNAs. *Proc Natl Acad Sci U S A*.

YANG, C. J., CHEN, T. C., WANG, C. S., WANG, C. Y., LIAO, L. F., CHEN, Y. H., HWANG, J. J., SIU, L. K., LU, P. L. & HUANG, M. S. 2010. Nosocomial outbreak of biotype I, multidrug-resistant, serologically non-typeable *Haemophilus influenzae* in a respiratory care ward in Taiwan. *J Hosp Infect,* 74**,** 406-9.

YEW, W. S. & GERLT, J. A. 2002. Utilization of L-ascorbate by *Escherichia coli* K-12: assignments of functions to products of the yjf-sga and yia-sgb operons. *J Bacteriol,* 184**,** 302-6.

YOCH, D. C. & CARITHERS, R. P. 1979. Bacterial iron-sulfur proteins. *Microbiol Rev,* 43**,** 384-421.

YOON, H., ANDERSON, C. D. & ANDERSON, B. M. 1989. Kinetic studies of *Haemophilus influenzae* 6-phosphogluconate dehydrogenase. *Biochim Biophys Acta,* 994**,** 75-80.

ZHANG, Y., MORAR, M. & EALICK, S. E. 2008. Structural biology of the purine biosynthetic pathway. *Cell Mol Life Sci,* 65**,** 3699-724.

ZHENG, M., DOAN, B., SCHNEIDER, T. D. & STORZ, G. 1999. OxyR and SoxRS regulation of fur. *J Bacteriol,* 181**,** 4639-43.

ZHENG, M., WANG, X., TEMPLETON, L. J., SMULSKI, D. R., LAROSSA, R. A. & STORZ, G. 2001. DNA microarray-mediated transcriptional profiling of the *Escherichia coli* response to hydrogen peroxide. *J Bacteriol,* 183**,** 4562-70.

ZULTY, J. J. & BARCAK, G. J. 1995. Identification of a DNA transformation gene required for com101A+ expression and supertransformer phenotype in *Haemophilus influenzae*. *Proc Natl Acad Sci U S A,* 92**,** 3616-20.