2 Materials and Methods

2.1 General chemicals, reagents and buffers

General laboratory chemicals were purchased from Sigma unless otherwise stated. Buffers were prepared as aqueous solutions in distilled water according to standard methods [176]. Solutions were sterilised by autoclaving or by filtration (Millipore, 0.22μm).

2.2 Bacterial strains

Bacterial strains in this study were a combination of laboratory strains and genetically modified *Salmonella* and *Escherichia coli* organisms. The strains from all studies are outlined in table 2.1.

2.3 Plasmids

Plasmids were isolated from cultures using miniprep plasmid purification kit (Qiagen) according to manufacturer's instructions (section 2.5.2) and are outlined in table 2.2.

Table 2.1 Strains used in this study

Strain	Species or serovar	Parent Strain	Genotype	Selection marker	Source	Reference
BRD948	S. Typhi	Ty2	aroC, aroD, htrA	none	Gordon Dougan	[177]
TT10	S. Typhi	BRD948	ompR::kan	aph	This Study	-
SL1344	S. Typhimurium	-		L-His	Rob Kingsley	[178]
RAK83	S. Typhimurium	SL1344	ompR::kan	aph	This Study	-
RAK103	S. Typhimurium	SL1344	STM1128-29::kan	aph	This Study	-
RAK105	S. Typhimurium	SL1344	STM1130-33::kan	aph	This Study	-
RAK113	S. Typhimurium	SL1344	phoN::CmR	CmR	Rob Kingsley	[122]
TT56.1	S. Typhimurium	SL1344	STM4464-4467::kan	aph	This Study	-

T53.8	S. Typhi	BRD948	ompR3XFLAG	None	This Study	-
TOP10	E. coli	E. coli	lacZΔM15	None	Invitrogen	-
Cɪɪ8λpir	E. coli	E. coli		None	Keith Turner	[179]

Table 2.2 Plasmids used in this study

Plasmids	**Function**	**Source**	**Reference**
pCR2.1	Cloning vector	Invitrogen	-
pWT12	Suicide Vector	Keith Turner	-
pKD13	Kanamycin template	Anne Bishop	[180]
pKD46	Red recombinase	Anne Bishop	[180]
pKD3	Chloramphenicol Template	Stephen Baker	[180]
pBAD202	Protein expression and tagging	Invitrogen	-
p/c/1	ssaG::lacZ reporter	Derek Pickard	[181]

2.4 Microbiological media and techniques

2.4.1 Bacterial culture

2.4.1.1 Standard methods

Bacterial strains were grown in Luria-Bertani (LB) medium, 1% w/v tryptone (Beckton Dickinson), 0.5% w/v yeast extract (Beckton Dickinson) 0.5% NaCl in ddH_2O, with antibiotic selection when required, from a single colony. Stocks of bacteria were stored at -70° C in cryovials in a minimum of 10% glycerol (v/v). Microbiological agar plates or broth media, 10ml, were inoculated using a plastic loop (VWR - Leicestershire) from a cryovial. Plate cultures were stored at 4°C for up to a month wrapped in Parafilm. Broth cultures grown from plate cultures were inoculated from a single colony using a plastic loop.

Bacterial strains were routinely cultured in LB medium. Bacterial strains were routinely grown on LB agar plates, 1.5% Bacto-agar (Beckton Dickinson) in LB broth. Liquid cultures were grown in 50ml Falcon tubes or 20ml universal containers in a New Brunswick shaking incubator (225 rpm) at 37°C unless otherwise stated. LB-agar plate cultures were incubated statically at 37°C. When required, media were supplemented with antibiotics at the concentrations and solvents shown in table 2.3. All antibiotics were filter sterilised at the requisite stock concentration unless dissolved in 100% ethanol where required.

Table 2.3 Antibiotic concentration and solvents

Antibiotic	**Solvent**	**Concentration (μgml^{-1})**
Kanamycin	Water	40
Chloramphenicol	Ethanol	30
Ampicillin	Water	100

For the culture of *S.* Typhi BRD948 and derivatives, an auxotroph of *S.* Typhi Ty2, all media were supplemented with 1% (v/v) aro mix ($40mgL^{-1}$ phenylalanine, $40mgL^{-1}$ tryptophan, $10mgL^{-1}$ *para*-amino benzoic acid and $10mgL^{-1}$ dihydroxybenzoic acid) and 1% (v/v) tyrosine mix ($40mgL^{-1}$ tyrosine disodium salt).

2.4.1.2 Minimal media

Minimal medium, 42mM Na_2HPO_4, 24mM KH_2PO_4, 9mM NaCl, 19mM NH_4Cl, 0.2mM $MgSO_4$, 0.1mM $CaCl_2$, 1.0% glucose (w/v), was used as defined medium and supplemented as required. The carbon source is glucose and the nitrogen source is NH_4Cl. These were replaced with other sources for some minimal medium assays.

2.4.2 Harvesting of bacteria from broth cultures

Bacterial cultures smaller than 50ml and greater than 1.5ml were harvested by centrifugation at 4600 rpm for 10 mins (Sorvall legend RT). Cultures smaller than 1.5ml were harvested in a bench top microcentrifuge at 13,000 rpm for 3 mins.

2.4.3 Growth curves

Growth curves of bacterial strains in this study were performed to determine if growth of recombinant strains was comparable to the wild-type. Overnight culture was diluted 1/100 into 30ml of room temperature LB in a 50ml Falcon tube and grown at

37°C. Samples were taken hourly and the OD_{600nm} measured in 1ml cuvettes using sterile LB as the reference. If the mutation was known to affect the outer membrane or capsule serial dilutions were performed and density calculated to determine if there was any difference between the number of colony forming units (CFU) and OD_{600nm} readings.

2.4.4 Identification of bacteria

The identification of *Salmonella* serovars was performed using standard laboratory slide agglutinations according to manufacturer's instructions. O antigen and Virulence (Vi) antigen agglutination tests were used to confirm the serology for each serovar. Briefly, 10µl of antisera was pipetted into a Petri dish and mixed with approximately half a loop of fresh bacterial plate culture. A positive result was recorded if bacteria clumped together in discrete groups within 60 seconds of mixing. Antisera for alternative antigens were used as a negative control.

2.5 Molecular biology techniques

2.5.1 Bacterial genomic DNA extraction

Genomic DNA was isolated using Wizard SV Genomic DNA (Promega) kit according to the manufacturer's instructions. Briefly, cultures were harvested, the supernatant removed and the cells resuspended in 600µl of nuclei lysis solution then incubated at 80°C for 5 mins. RNA was then digested using 3µl of RNAse at 37°C for 15mins and the sample cooled to room temperature. Proteins were then precipitated with 200µl of protein precipitation solution and vortexed vigorously for 20sec before incubating on ice for 5mins. Precipitated proteins were then pelleted by 3 mins

centrifugation at 13,000xg and the DNA containing supernatant transferred to a clean tube containing 600μl of isopropanol to precipitate the DNA. DNA was then harvested by centrifugation (13,000xg for 2mins) and the supernatant discarded. The pellet was then washed with 70% ethanol (ice-cold) and harvested once more. The supernatant was discarded and the pellet allowed to dry before being resuspended in nuclease free water to the desired volume.

2.5.2 Plasmid DNA extraction

Plasmid DNA was isolated from 5ml of overnight culture using Qiagen miniprep kit (Qiagen) according to the manufacturer's instructions. Briefly, the bacterial culture is harvested and lysed by high alkaline conditions. The plasmid DNA is then adsorbed on a QIAprep membrane, washed then eluted with nuclease free water.

2.5.3 Agarose gel electrophoresis

Genomic DNA, plasmid DNA, PCR products, restriction enzyme digests and RNA were routinely separated by agarose gel electrophoresis. Agarose concentration varied from 0.8% to 2% (w/v) depending on the predicted size of DNA fragments. Agarose (Sigma ultra pure grade) was suspended in TAE at the requisite concentration and boiled. Cooled, molten agarose was then supplemented with Ethidium Bromide (Sigma) to a concentration of $1\mu g ml^{-1}$ and poured into a mould containing the requisite number of wells. DNA samples were mixed with DNA loading buffer to a ratio of 4:1 and loaded into the wells of the TAE submerged, solid agarose gel. All samples were compared to Hyperladder I™ (Bioline) or Hyperladder IV™ depending on the estimated size of the products. Application of a constant voltage across the gel

resulted in electrophoresis of the DNA and was visualised on a UV transilluminator. Photographic records were taken using UVtech hardware and software.

2.5.4 Quantification of DNA and RNA by spectroscopy

DNA and RNA preparations were quantified using the ND-1000 (NanoDrop Technologies). Briefly, the NanoDrop method determines absorbance at 260nm. A conversion factor of 5o for DNA and 40 for RNA for every unit of absorbance represents an estimated 50μgul^{-1} and 40μgul$^{-!}$, respectively.

2.5.5 DNA manipulation techniques

2.5.5.1 DNA restriction

New England Biolabs supplied all restriction endonucleases unless otherwise stated. Determination of appropriate endonuclease was done *in silico* using Macvector© (Accelerys). Restriction enzyme digests were performed according to the manufacturer's instructions. Briefly, 20 units of enzyme were used in a reaction mixture containing appropriate buffer, 10μl of DNA and incubated at 37°C for 1-2 hours in a water bath. Digests used for sub-cloning were dephosphorylated with 2μl Antarctic Shrimp Phosphatase (Invitrogen) to prevent re-ligation.

2.5.5.2 Oligonucleotides

All oligonucleotides used in this study were synthesised by Invitrogen™ and resuspended in nuclease free H_2O (Ambion) to a concentration of 100pmolμl^{-1}. Table 2.4 outlines the sequences requested for each oligonucleotide. Primers for general PCR were between 18 and 25 nucleotides. For design of general PCR primers each

cytosine or guanine was assigned four units and each adenine or thymine was assigned 2 units. Optimal primers contained a total of 64 units with a 5' and 3' terminal cytosine or guanine. Using Macvector© (Accelrys) software and the DNA sequence file for each *Salmonella* serovar, *in silico* analysis was performed to check for predicted non-specific products. Mutagenesis primers were between 48 and 55 bases long. Real Time PCR primers were between 18 and 25 bases designed by Probefinder 2.4 (Roche). Overlap extension PCR primers were between 25 and 69 bases.

Table 2.4 Primers used in this study

cDNA sequencing mechanism primers	
AACATCTGCAAG[(N)19]CAGCGACGCATC[(N)5]	DNA primer
GAUGCGUCGCUG	RNA primer
Mutagenesis	Strain
GGATCGTCTGCTGACCCGTGAATCTTTCCATCTCATGGGTGTAGGCTGGAGCTGCTTC	delta ompR
GTCTGAATATAACGCGGATGCGCCGGATCTTCTTCCACATTCCGGGGATCCGTCGACC	delta ompR
AAAACGGAGTAAAACTTCAAAATATATAAGGCGGAACTGGGTGTAGGCTGGAGCTGCTTCG	RAK103
GTCACCGTGTGCTGTGTCGGTATAGCGTGGTATCATGAAAATTCCGGGGATCCGTCGACC	RAK103
TCACTAATGATGAAGCTTTACTCCAGTTGTATTTCTTCGCGTGTAGGCTGGAGCTGCTTCG	RAK105
AGCGCCACCGGCCAATAACACCACCATCCGGCTTTAAATTATTCCGGGGATCCGTCGACC	RAK105
TCAAGCGATTAATGCATGATTTACTCATCGCAAACGGTTCATTCCGGGGATCCGTCGACC	TT56.1
ACTCCTTCTTTATTCTTGTAATTATGTAAAAGGTATAATGGTGTAGGCTGGAGCTGCTTCG	TT56.1
Mutagenesis checking primers	
TGATGACGATATGCGTCTGCG	delta ompR
GGATCTTCTTCCACCATACGGC	delta ompR
AACACTCCACAACATAATAT	RAK103
TTGTCGCCCCGCGCGTTGGC	RAK103

TGAAATTATGCACCATAAGA	RAK105	
TATCCAGTCTACATAAGCGC	RAK105	
TACTCAGCTCTGTTTTGGGA	TT56.1	
CACGCTAAATATTTGCAGCC	TT56.1	
Real Time PCR	**Gene**	**Universal Probe**
GGATCTCGGTGGTGCTCA	STY0002	82
GGCACACAGAAGCTGATGC	STY0002	82
CTCTATCCGTTCCCGCATAA	STY0779	4
ATGGACGTGAGCGTATGGTT	STY0780	4
GCGTCACCCCTGAAGAGAT	STY2281	70
CCTGTTTTAATTCGTCGCTCA	STY2281	70
AAGCCTTGCCCCTATGCT	degQ	27
CCCTCGACTTTTACGCTGAC	degQ	27
TCGTCAGTGAGCGTCTGG	STY3913	19
GGCGTTATCAGCGACCTTAAT	STY3913	19
CGTACCTGCGGATGACTTAAC	STY3941	65
TTGCATCTAAGTGCGCAAAG	STY3941	65
TGCGGCGCTTAATGATTT	waaY	62
CGTATGTTTTTACATAGCGTAATGTTT	waaY	62
CCTGAAGAATCAGTGAACAGTCA	phoN	77
AGCCGGATCATCATTACCTG	phoN	77
GCACCAGCCTCTCTTCTCC	STY0775	40
TTAACGATGAAGCCGTCCAT	STY0775	40
ACGTATGGGTTGACTACCGTTT	STY1002	62
CGTAGCTGGAGCTGTAGTCGT	STY1002	62
TTTCTGGCGCTATTCCTGAC	STY1287	23
AACGGAGATCATCTGGAAGG	STY1287	23
GTGATATGGCTGCATGGTCTT	STY1523	62
CGCAGAGCGAATGAAAGATT	STY1523	62
TTTCAACTCCGCTATCACCA	fliC	61
CTACGGGCAGAAGACAGGTT	fliC	61
AGCGTGAGCGGTGAAAAC	ompC	40
GTAACCGTCGCCGTTCTG	ompC	40
TGGTATCGACAAAACCAAAGC	STY2806	40
AGGCGCGTTTAAATTCAGTG	STY2806	40
TCAGCTACTGATTGAAAGTTATACCAA	slsA	4
GCGTTGAGAATCCAGCAGTT	slsA	4
AACTGTTCGCCATCCTTCAA	STY4402	29
GAAGCTGGCAATGGTCTTGT	STY4402	29
CGGTATAGACGAGGTAAAAATCG	tviB	23
ATTCAACTGCCAGAGGAAGC	tviB	23

Overlap extension PCR

CGTCAGGCAAACGAACTGCC	5′ to 3′ ompR bases (364:383)
CCGTCATGGTCTTTGTAGTCTGCTTTAGAACC GTCCGGTA	Full reverse primer sequence 5′ to 3′
GACTACAAAGACCATGACGGTGATTATAAAG ATCATGATATCGATTACAAGGATGACGATGA CAAGTAGGTACCGGACGGTTCTAAAGC	Concatenated primers are 5′ to 3′ Forward (1:69 FLAG + 1:20)
CGAAACGCAGGCGGCACG	Reverse for envZ is 5′ to 3′ (213:230)

2.5.5.3 Polymerase chain reaction (PCR)

PCR was used to amplify specific regions of DNA. For general PCR amplification of template DNA for colony screening or molecular cloning, Supermix (Invitrogen) was used. For high fidelity PCR, PFU polymerase (Promega) was used. For real time PCR Taqman Master Mix (Roche) was used. All PCR reactions were done in 0.2ml thin wall PCR tubes (ABgene) on a DNA engine DYAD thermal cycler (MJ Research) except for real time PCR where the Applied Biosystems ABI7900 was used. Negative controls were the same reaction without the addition of DNA template.

2.5.5.3.1 PCR supermix (Invitrogen)

DNA amplifications using PCR Supermix™ were performed as per manufacturer's instructions in a total volume of 25µl. Each reaction mix contained 22.5µl of Supermix, 0.5µl of each forward and reverse primer and 10-100ng of template DNA. Template DNA was omitted for all PCR reactions as a negative control.

DNA was amplified using the cycle parameters of one stage at 95°C for 120s then 30 cycles of 95°C for 30s, 55°C for 30s† and 72°C for 90s†† followed by an elongation step at 72°C for 300s.

† Annealing temperature is dependent on melting temperature of primers.

†† Elongation step was altered depending on the predicted size of product. 60s/kb is general rule.

2.5.5.3.2 PFU polymerase high fidelity PCR

For PCR amplicons where high fidelity of the template is required the PFU enzyme is used. This dual enzyme is a 5' to 3' polymerase and 3' to 5' proof reading exonuclease. The following components and the final concentrations were combined in a total volume of 25µl: Pfu DNA polymerase 10x buffer, 10mM dNTP mix, 200µM each, upstream primer, 1.0µM, downstream primer, 1.0µM, DNA template <0.5µg, Pfu DNA polymerase, 0.625u. Template DNA was omitted for all PCR reactions as a negative control.

DNA was amplified as in 2.5.5.3.2

2.5.5.3.3 Real time PCR

To determine DNA and cDNA template copy number fluorescent probes were used. Briefly, to make a probe the reverse complement DNA is ligated to a fluorophore and a quencher that anneals to the target DNA during PCR cycles. The Taq polymerase cleaves the quenching substrate during polymerisation of dNTPs. The fluorophore is then excitable by its specific frequency and the fluorescence is measured. When this fluorescence is plotted on a graph an exponential phase can be identified and related to the cycle number of the PCR called the cycle threshold, C_t value.

We used the Universal Probe Library (Roche) and derived amplicons and primers by submitting the target DNA sequence to ProbeFinder version 2.4 (https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp?id=UP030000).

2.5.5.3.4 Overlap extension PCR

Overlap extension PCR was used to construct large DNA fragments from two smaller DNA fragments by a ligase-independent methodology. For example, to add a peptide to the end of protein, overlap extension PCR incorporates the existing sequence at the start of a primer and appends the designed peptide sequence at the end of the primer. During PCR this single strand of DNA is extended by Taq polymerase and included in the amplicon. A further round of PCR was then used to join two such amplicons if they include a 20-25bp homologous region. This amplicon is cloned and sequenced as described in 2.5.5.5 and 2.5.7.

2.5.5.4 Purification of PCR products

PCR amplicons required for further applications were purified to remove excess primer and nucleotides using a PCR purification kit (Qiagen) according to the manufacturer's instructions. Products excised from Agarose gels were purified using a Gel Extraction kit (Qiagen) according to the manufacturer's protocol. Briefly, the agarose gel was dissolved and the DNA bound to a membrane. This was then washed and eluted in nuclease free water.

2.5.5.5 Cloning

2.5.5.5.1 Cloning PCR amplicons

PCR amplicons were cloned for further sequencing or subcloning using the TOPO® TA Cloning Kits (Invitrogen) according to the manufacturer's instructions. Briefly, 2μl of fresh PCR, 1μl of salt solution (Invitrogen) and 1μl of vector mix in a final

volume of 6µl were incubated at room temperature for five minutes. DNA was then transformed as described in 2.5.8.1.

2.5.5.5.2 Sub-cloning sequences into pWT12

Rapid DNA Ligation Kit (Roche) is used to ligate previously cloned sequences into the suicide vector pWT12. Both plasmids were digested with restriction endonucleases XbaI and SacIII, these sites flank the cloned amplicon and also flank the p15A in the suicide vector. The fragments were separated and purified from an agarose gel. Ligated DNA was then transformed into *E. coli* host strain C118λpir [179] as described in 2.5.8.2. Constructs were then isolated, screened and sequenced as described previously.

2.5.6 Precipitation of DNA

To precipitate DNA, sodium acetate is added to a final concentration of 300mM. 2.5 volumes of 96-100% ethanol is added, mixed well and centrifuged for 15 mins at a minimum of 13,000rpm at 4°C. The DNA pellet is washed with 70% ethanol at 0°C, spun at 13,000 rpm. The supernatant is discarded and the pellet air-dried.

2.5.7 DNA sequencing

2.5.7.1 Capillary-based sequencing

All capillary-based sequencing was done by Team 41, Wellcome Trust Sanger Institute (WTSI), however all experimental design and sample preparation was done by Tim Perkins.

DNA sequencing was done using big dye terminator reactions (PerkinElmer/ABI) in 0.2ml thin wall PCR tubes. 100ng of plasmid DNA was added to 4pmol of M13 forward primer, 4pmol of M13 reverse primer and 4μl of big dye reaction mixture in a total volume of 10μl. Dideoxynucleotides were incorporated by PCR during 35 cycles of 96°C for 10s, 50°C for 5s and 60°C for 240s. DNA was then precipitated using the method previously described. The air-dried pellets were loaded onto an ABI 3700 capillary sequencer. Sequence reads were checked by eye using DNA visualisation software MacVector (Accelerys).

2.5.7.2 Illumina sequencing

All experimental design and biological sample preparation was done by Tim Perkins, however, Illumina sequencing was submitted to the project co-ordinator Theresa Feltwell, Team 81 (WTSI) and libraries were prepared under the direction of Mike Quail (WTSI). Libraries were then submitted to the Illumina sequencing team (WTSI) and the quality-controlled data sent to a repository for further analysis.

Sequencing libraries for the Illumina GA platform were constructed by shearing the enriched cDNA by nebulisation (35psi, 6min) followed by end-repair with klenow polymerase, T4 DNA polymerase and T4 polynucleotide kinase (to blunt-end the DNA fragments). A single 3' adenosine moiety was added to the cDNA using klenow exo- and dATP. The Illumina adapters (containing primer sites for sequencing and flowcell surface annealing) were ligated onto the repaired ends on the cDNA. Gel-electrophoresis was used to separate ligated cDNA fragments from unligated adapters by selecting cDNA fragments between 200-250 bps in size. cDNA fragments were recovered by gel extraction at room temperature to ensure representation of AT rich sequences. Libraries were amplified by 18 cycles of PCR with Phusion polymerase.

Sequencing libraries were denatured with sodium hydroxide and diluted to 3.5 pM in hybridisation buffer for loading onto a single lane of an Illumina GA flowcell. Cluster formation, primer hybridisation and single-end, 36 cycle sequencing were performed using proprietary reagents according to manufacturer's recommended protocol (https://icom.illumina.com/).

The efficacy of each stage of library construction was ascertained in a quality control step that involved measuring the adapter-cDNA on a Agilent DNA 1000 chip. A final dilution of 2nM of the library was submitted to the sequencing machine.

2.5.8 DNA transformation

2.5.8.1 Transformation of chemically competent cells

2.5.8.1.1 Preparation of chemically competent cells

Cultures were grown to an OD_{600} of 0.4 and chilled on ice for 5min. Cells were harvested by centrifugation and resuspended in 40ml of ice cold Tbf1 (30mM of potassium acetate, 100mM potassium chloride, 10mM calcium chloride, 50mM manganese chloride, 15% (v/v) glycerol) then incubated on ice for 5min. Harvested cells were resuspend in 4ml of ice cold Tbf2 (10mM MOPS, 75mM calcium chloride, 10mM potassium chloride, 15% glycerol) and incubated on ice for 15mins. Cells were either used immediately in 100µl aliquots or snap frozen (dry ice and methanol bath).

2.5.8.1.2 Chemical transformation

Frozen cells were thawed on wet ice and 2µl of plasmid DNA was added to one aliquot of cells for 30 minutes. Cells were then placed in a water bath at 42°C for 45 s

and returned to ice for 2 minutes before the addition of 400µl of recovery media SOC (10mM $MgCl_2$, 10mM $MgSO_4$ and 20mM glucose in LB). Cells were incubated at 37°C for 1-2 hours. 100µl and 20µl aliquots were plated onto selective agar plates and incubated overnight at 37°C. An aliquot of cells without addition of DNA was used as a negative control.

2.5.8.2 Transformation by electroporation

2.5.8.2.1 Preparation of electrocompetent cells

Electroporation was used for transformation of linear and plasmid DNA into *Salmonella* or *E. coli.* Bacterial cultures were grown overnight under selection in LB, diluted 1/100 into 30ml of LB and grown to OD_{600} = 0.3 at 37°C. Strains containing heat sensitive plasmids were grown at 30°C. Cells were harvested at 4°C and washed three times in 30ml of 0°C sterile 10% glycerol. The pellet was resuspended in 0°C 10% glycerol to 1/1000 of the culture volume, 300µl, to give a cell density of circa 10^{10} CFU/ml.

2.5.8.2.2 Electroporation

Electroporation cuvettes (Equibio) were chilled on ice for 15 mins before the addition of 60µl of competent cells. 10-100ng of plasmid DNA or 100-1000ng of linear DNA was added and the total volume made up to 66µl. The cuvettes were dried and immediately electroporated using a Bio-Rad Gene Pulser set at 25µF, 600Ω and 2.4kV. Room temperature SOC, supplemented with Aro mix for BRD948, was added to the cuvettes immediately after electroporation and cells were incubated shaking at

37°C for 1-2 hours. Cells were then incubated overnight on selective LB agar plates. Cells with the addition of 6µl of nuclease-free water were used as a negative control.

2.5.8.3 Screening for transformed DNA

2.5.8.3.1 Screening for non recombinant DNA

Colonies were grown up individually in 10ml of LB and plasmid DNA was isolated using methods previously described and DNA was screened by restriction endonuclease profiles by agarose gel electrophoresis.

2.5.8.3.2 Screening for recombinant DNA

Colonies were initially screened by PCR with primers designed upstream and downstream of the predicted recombination site. Further screening was carried out on positive clones using primers internal, both forward and reverse strand, to the transformed DNA in combination with the upstream and downstream primers.

2.5.9 Bacterial conjugation

The donor and recipient strains were grown overnight in 10ml selective LB at 37°C. Cultures were then mixed and grown statically for 3 hours in a water bath at 37°C. Cells were then plated on agar plates selecting for both the donor and recipient. Colonies were screened as previously described in 2.5.8.3.

2.5.10 Transduction

2.5.10.1 Infection of donor strain with P22 phage

The P22 phage lysate 10-fold dilutions were prepared to a minimal dilution of 1/1000. 10µl of each dilution was mixed with 100µl overnight culture of the donor strain and incubated at 37°C for 20mins. Liquid top agar (0.75% agar (w/v), LB) less than 45°C was added to the broth, poured over a warm LB agar plate and incubated at 37°C until confluent plaques formed. Optimal titre was generally when confluent plaques formed by 4 hours.

2.5.10.2 Isolation of P22 phage containing genomic DNA

Top agar was scraped off the LB with a glass slide and collected in a tube with 3ml of LB. 100µl of chloroform is added to lyse the cells, vortexed and incubated at room temperature for 20mins. The supernatant was then collected after centrifugation and sterilised with 50µl of chloroform. Lysates were stored at 4°C.

2.5.10.3 Infection of recipient strain with P22 phage

The lysate from 2.5.10.2 was serially diluted as in 2.5.10.1 and mixed with 100µl of an overnight culture of the recipient strain and incubated statically at 37°C for 15mins. 1ml of LB supplemented with 10mM ethylene glycol tetraacetic acid (EGTA) was then added and incubated at 37°C for 1hour. Cells were harvested and resuspended in 100µl and plated onto selective LB agar supplemented with 10mM EGTA and incubated overnight at 37°C. P22 lysate was plated as a negative control. Colonies were screened as previously described.

2.5.11 Mutagenesis

2.5.11.1 Red recombinase mutagenesis

Allelic exchange used in this study was based on the method described by Datsenko and Wanner [180]. The basis of this methodology is as follows. The temperature sensitive plasmid, pKD46, carries an arabinose inducible lambda Red recombinase gene, and confers ampicillin resistance with β-lactamase. The recombinase mediates efficient recombination between short homologous sequences flanking an antibiotic cassette that is amplified by extension PCR. Briefly, pKD46 was transformed into the strain to be mutated and grown at 30°C. Electrocompetent cells were made as previously described with the supplementation of arabinose to a final concentration of 0.2% (w/v) one hour after dilution.

Homologous recombination was used to insert the cassette and remove the target region of DNA, to do this two primers were designed to append these homologous sequences. Firstly, the forward primer, is designed to encode 45-51bp upstream of the target DNA sequence and a region homologous to the first 18-20 bases of the resistance cassette. The reverse primer encodes homologous regions to the opposite end of the target DNA and is reverse complemented. Subsequently, PCR was used to amplify the cassette from plasmid DNA, either pKD3 (chloramphenicol) or pKD13 (kanamycin) using these primers. The resulting amplicon encoded the cassette with each end containing regions homologous to each end of the target DNA. This amplicon was then purified and electroporated as previously described. Cells were plated out and colonies were screened by colony PCR as previously described in

2.5.8.3.2. If no colonies grew the remaining broth left at room temperature overnight was plated as before and colonies screened.

The genotype was screened by PCR and desired colonies were grown in selective LB at 43°C , a non-permissive temperature for replication of pKD46. Broth was plated as previously described and clones that did not grow on ampicillin plates were maintained.

2.5.11.2 FLP recombinase mutagenesis

To remove the antibiotic resistance cassette, in order to generate an unmarked, non-polar mutation, FLP recombinase recognition sites were incorporated into the flanking regions of the resistance gene. The heat sensitive, ampicillin resistant and FLP recombinase expressing plasmid, pCP20, was transformed into the mutant and plated onto ampicillin. Colonies were selected if they were sensitive to the previously resistant antibiotic and grown at 43°C overnight to inhibit replication of pCP20. Clones sensitive to ampicillin after heat treatment were screened as previously described.

2.5.11.3 Allelic exchange mediated by suicide vector constructs

The vector unable to replicate in *Salmonella* (suicide vector) was transferred to the host strain by bacterial conjugation as described in 2.5.9. Appropriate isogenic hosts were chosen. The recombinant strain was then grown overnight and diluted 1/10 into LB supplemented with chloramphenicol and incubated at 37°C for 10 minutes. Chloramphenicol is bacteriostatic so any cells having lost the suicide vector do not grow. Ampicillin was then added to a final concentration of 500μgml^{-1}, as it only

lyses growing cells, and incubated until cells were lysed. 1ml of LB was then washed in fresh LB twice and resuspended in 200µl. This was then plated onto LB without NaCl and supplemented with sucrose and grown overnight. Colonies were screened as previously described.

2.5.12 RNA methodologies

2.5.12.1 RNA stabilisation

RNA was stabilised using RNAprotect Bacteria Reagent (Qiagen) according to the manufacturer's instructions. Briefly, 1 volume of culture was diluted in 2 volumes of RNAprotect and vortexed to precipitate RNAse enzymes. The bacteria were harvested at 4600rpm for 25 minutes in Sorvall bench top centrifuge, supernatant discarded and pellet air-dried.

2.5.12.2 RNA isolation

2.5.12.2.1 RNA isolation for microarray

RNA was isolated from bacterial cultures using RNeasy RNA purification kit (Qiagen) according to the manufacturer's instructions with the following modifications. A concentration of 1mgml-^{1} of lysozyme was used and RNA eluted using 30µl of DEPC treated water. RNA was quantified as previously described. Briefly, cells were lysed in TE containing 1mgml^{-1} lysozyme and incubated for 10mins with occasional vortexing. Cells were then lysed in buffer RLT, mixed with ethanol and bound to a membrane. The membrane was then washed and RNA eluted in DEPC water.

2.5.12.2.2 RNA isolation for Illumina sequencing

RNA was isolated from bacterial cultures using SV RNA isolation purification kit (Promega) according to the manufacturer's instructions with the following modifications. A concentration of 1mgml^{-1} of lysozyme was used and RNA eluted using 100µl of DEPC treated water. RNA was quantified as previously described.

2.5.12.2.3 rRNA removal

The 16S and 23S rRNA were removed from the total RNA population, from a final concentration of 0.83 mgml^{-1} by either oligo depletion (MicrobExpress, Ambion) according to manufacturers instructions or by incubating 10µg RNA/Unit of Terminator Exonuclease (Epibio) for 60mins at 30C. Briefly, to oligo deplete the rRNA, capture oligos were annealed to 16s and 23s sequences then annealed biotinylated oligos. Streptavidin-coated magnetic beads were then bound to the biotinylated motif and removed by magnetic capture. The beads were then washed and the RNA precipitated.

2.5.12.3 Removal of DNA from RNA samples

DNA was removed from RNA using Amplification Grade DNAse I (Invitrogen) according to the manufacturer's instructions. Briefly, 1µg of RNA was added to 1µl of DNAse Reaction Buffer and 1 unit of DNAse in a final volume of 10µl and incubated at room temperature for 15mins. DNAse was inactivated by the addition of EDTA to a final concentration of 2.5mM and incubating the sample at 65°C for 10mins. Samples were screened by PCR using to check for undigested DNA.

2.5.12.4 Reverse transcription of RNA

2.5.12.4.1 Reverse transcription of RNA for aminoallyl labelling

RNA was reverse transcribed to complementary DNA (cDNA) incorporating dUTP using Superscript III Reverse Transcriptase (Invitrogen). 16µg of RNA was incubated at 70°C for 10 mins with 1µl of $pd(N)_6$ random hexamers (GE Biotech) in a total volume of 27.7µl then cooled on ice. For the reaction, 9.0µl of First Strand buffer, 4.5µl of DTT, 1.8µl of aa-dNTP (5µl each of dATP, dCTP, dGTP (100mM; Promega), 2µl dTTP (100mM), 6µl aa-dUTP and 17µl SDW) and 2µl of Superscript III were added and incubated at 42°C for 3 hours. The RNA was then hydrolysed with 15µl of 1M NaOH, 15µl of 0.5M EDTA (pH 8.0) and incubated at 65°C for 15mins. This reaction mix was then neutralised with 15µl of 1M HCl.

2.5.12.4.2 Reverse transcription of RNA for Illumina sequencing and real time PCR

Real time PCR was used to determine copy number of mRNA to verify microarray data. RNA was reverse transcribed to cDNA as described in 2.5.12.4.1 with the following modification. A final concentration of 10mM of dTTP, dATP, dCTP and dGTP each were used instead of the aa-dNTP mix.

2.5.12.5 Labelling of RNA for microarray

2.5.12.5.1 Purification of cDNA

For purification of cDNA, MinElute Columns (Qiagen) were used with a modified protocol. 450µl of Buffer PB (Qiagen) was added to the reaction mixture and spun through a MinElute column at 13,000rpm in a microcentrifuge. The flow through was discarded and column washed with 750µl of phosphate wash buffer (4.75mM K_2HPO_4, 0.25mM KH_2PO_4 and 84.4% Ethanol) then eluted twice in 10µl each of phosphate elution buffer (0.019mM K_2HPO_4 and 0.01mM KH_2PO_4).

2.5.12.5.2 Coupling of CyDye™ ester to dUTP

CyDye™ esters were coupled to purified cDNA add 1M Na_2CO_3 (pH9.0) to a final concentration of 100mM. The reference dye was resuspended in CyDye™ Cy3 (Amersham) in 10µl of dimethyl sulfoxide (DMSO) and 1µl was added to the cDNA sample. For the target sample, CyDye™ Cy5 (Amersham) was resuspended in 10µl of DMSO, 1µl of this was added to the cDNA sample. Samples were then incubated at room temperature in the dark for 60mins.

2.5.12.5.3 Purification of labelled cDNA

Labelled cDNA was purified using MinElute Columns (Qiagen) according to the following protocol. Hydroxylamine hydrochloride was added to each reaction to a final concentration of 500mM and incubated in the dark for 30mins at room temperature. Both labelled targets were combined and 48µl of dH_2O added to 500µl of PB buffer (Qiagen) and applied to a MinElute column. Samples were spun at 13,000rpm in a microcentrifuge and washed with PE Buffer (Qiagen), the column was dried and the samples eluted twice with 10µl of phosphate elution buffer.

2.5.13 DNA microarray

2.5.13.1 Hybridisation and scanning

Microarray slides were designed and printed at the WTSI. Generation III *pan-Salmonella* array design and the protocol for hybridisation and scanning was previously published by Doyle *et al* [182]. The microarray used in this study comprises specific unique PCR products (200–500 bases) of 4097 CT18 genes, plus 345 gene segments representing the LT2-specific (relative to CT18) loci. Appropriate positive and negative controls were also printed on the arrays. The PCR were carried out using specific primers (Sigma-Genosys) in a two-step protocol. Three slides were used for each isolate, with dye reversal. The washing procedures were stringent and included 200 ml of 2x SSC at room temperature for 5 min, two washes in 200 ml of 0.1x SSC-0.1% sodium dodecyl sulfate at 65°C with gentle agitation for 30 min, and two washes in 200 ml of 0.1x SSC at 65°C with gentle agitation for 30 min (1x SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Slides were scanned using a Genepix 4000B scanner (Axon Instruments [now Molecular Devices], California), and every spot was assessed with Genepix Pro software (Axon Instruments).

2.5.13.2 Gridding and feature extraction

Spots were gridded and extracted by aligning the grid and applying the FLAG features filter in Genepix with the following parameters.

[Flags] < > [Not found] And

[F532CV] > 135 Or

[F635CV] > 135 Or

[F532 % Sat.] > 30 And

[F635 % Sat.] > 30

The information was then analysed in Genespring or using LIMMA in R(section 2.8).

2.5.14 Protein methodologies

2.5.14.1 Preparation of protein

2.5.14.1.1 Whole cell lysate

Cultures were harvested and resuspended in 200µl of PBS then mixed with 200µl of 2× SDS loading buffer (90mM Tris-HCl pH 8.45, 24% (v/v) glycerol, 4% (w/v) sodium dodecyl sulfate (SDS), 0.2% (v/v) 2-mercaptoethanol, 0.001% bromophenol blue). Samples were then boiled for 10mins and cooled on ice.

2.5.14.1.2 Cellular fractionation

Cellular fractionation was performed according to protocol previously described by Hantke *et al.* [183] and is described in detail below.

2.5.14.1.2.1 Cell lysis

Cellular fractionation is used to separate the two membrane components and the soluble cytosol for further analysis. Harvested cells were resuspended in 500µl of 100mM Tris.HCl (pH 8.0) supplemented with Complete (Roche) protease inhibitor cocktail. 1ml of 1M Sucrose dissolved in 100mM Tris.HCl (pH 8.0) was added then

100µl of 10mM EDTA and the solution mixed. 100µl of 2mgml^{-1} lysozyme dissolved in Tris.HCl (pH 8.0) was added before 3.2ml of dH_2O. Cells were incubated for 10mins at room temperature to allow formation of spheroplasts. 12.5ml of dH_2O was added to lyse cells. To the clear solution 3.5mls of 1M $MgCl_2$ was added with 100µl of 4mgml^{-1} DNAse 1 (Roche) and the solution incubated at room temperature with gentle agitation for 5mins. The solution is then centrifuged at 18,000rpm for 60mins at 4°C to separate the soluble fraction from the membrane fractions.

2.5.14.1.2.2 Precipitation of the soluble fraction

Soluble proteins were present in the supernatant and were separated from the precipitated membrane fractions. The supernatant was removed and the soluble proteins precipitated with trichloroacetic acid (TCA) to a final concentration of 10% (v/v) and centrifuged at 20,000 rpm for 30mins at 4°C. The pellet was then washed with dH_2O and resuspended in the requisite buffer.

2.5.14.1.2.3 Separation of the membrane fractions

The cytoplasmic membrane was solubilised from the remaining pellet by resuspension in 5ml of 2% Triton X-100 dissolved in 50mM Tris.HCl (pH 8.0) and centrifuged at 20,000 rpm for 30mins at 4°C and the supernatant removed. This was repeated and the supernatants were pooled leaving the precipitated outer membrane fraction in the pellet. The cytoplasmic membrane proteins, contained in the soluble fraction, were precipitated by mixing well with 10ml of chloroform, 20ml of methanol and incubated on ice for 30min. After centrifugation at 20,000 rpm for 30mins at 4°C the pelleted precipitated proteins were resuspended in the requisite buffer.

2.5.14.1.3 Heat shock

Samples were washed with PBS, resuspended in 1/1000 the volume of PBS and heat shocked for 10 mins at 60°C then diluted in an equal volume of SDS loading buffer.

2.5.14.1.4 Preparation of proteins for mass spectroscopy analysis

Samples were reduced in SDS loading buffer with 1mM dithiothreitol (DTT) at 70°C for 10mins, cooled to room temperature then alkylated for 30 minutes with 2mM iodoacetamide in the same buffer.

2.5.14.2 One-dimensional SDS-PAGE gel electrophoresis of proteins

Protein preparations were routinely separated by electrophoresis in denaturing 0.1% SDS polyacrylamide gels (PAGE). Polyacrylamide concentration affects the speed at which different sized proteins run. Suitable concentration or concentration gradient was chosen to optimise resolution of the target protein. All pre-cast gels (Invitrogen) were run at 150V for 55 mins.

2.5.14.3 Protein visualisation

2.5.14.3.1 Coomassie blue protein stain

To visualise proteins resolved by SDS-PAGE, proteins were stained with Coomassie Blue reagent. Proteins were fixed for 1 hour in 40% methanol, 2% acetic acid then stained for 1-4 hours with Brilliant Blue G (Sigma) and prepared according to the

manufacturer's instructions. The gel was rinsed for 1 min in 25% methanol, 5% acetic acid then destained in 25% methanol until the bands were visible.

2.5.14.3.2 Western blot

Resolved proteins were transferred to a nitrocellulose membrane by electrophoresis using Trans-blot semi-dry (Bio-rad). The nitrocellulose membrane was probed with a specific primary antibody and detected with Opti-4CN (Bio-rad) kit which included secondary antibody. This was done according to the manufacturer's instructions.

2.5.14.4 Immunoprecipitation for chIP-seq

Cultures were fixed with 1% formaldehyde to crosslink DNA-bound proteins and incubated at the same temperature for 20mins. The reaction was then quenched with glycine pH7 to a final concentration of 0.5M. Cells were harvested and washed twice in TBS (0.137M NaCl, 0.27mM KCl, 0.25mM Tris, pH7.5) and lysed according to cellular fractionation protocol. DNA was then sheared 3 times, 20 seconds on and 15 seconds off on ice at a 20% amplitude using a VibraCell sonicator (Sonics,Newton, USA) with a tapered probe (Model CV33). The target protein was immunoprecipitated (IP) using Protein G Immunoprecipitation Kit (Sigma) according to manufacturers instructions and uncrosslinked using pronase ($0.8 mgml^{-1}$) at 65°C overnight. DNA was then purified using Qiagen PCR purification kit. DNA was sequenced as per standard Illumina sequencing protocol.

2.5.14.5 chIP-seq data mapping and analysis

To generate a plot for use in Artemis the data were mapped using MAQ according to the same parameters as the transcriptome data mapping. The plots were then z-score

normalised and the data from the WT culture deducted from the TT53.8 (ompR:3XFLAG). This plot was then loaded into Artemis using Graph, Add User plot. Peak finder in Artemis was used to generate "misc_binding" features from all peaks that were 50bp in length with a score greater than 2. These data were manually annotated and most peaks not proximal to the start of a gene were discarded unless there were multiple sites close by. As the data do not align to a particular strand the direction of the binding site was estimated according to the closest CDS. The sequence of these sites were used to determine if there were any over-represented motifs using YMF [184] an online motif finding package. The parameters for the search were motif size = 8, Maximum of spacers in middle = 0, maximum of 2 degenerate symbols = 2. Degenerate symbols allowed in a motif are R (purine - A or G), Y (pyrimidine - C or T), W (A or T), and S (C or G).

2.6 Murine model of Salmonellosis

2.6.1 Competitive infections

Experimental design, culture, organ homogenisation and organ plating and CFU counting was done by Tim Perkins in these murine experiments. Infection by oral gavage and vivisection was performed by Robert Kingsley of the Molecular Pathogenesis group (WTSI).

Competitive infections were performed as described by previously (Kingsley *et al,* 2003). Briefly, 6 to 8 week-old female BALB/c (ByJ; Jackson Lab) mice were used. Bacteria were routinely cultured statically overnight prior to infection. In all experiments, the bacterial titre of the inoculum was determined by spreading serial 10-fold dilutions on agar plates containing the appropriate antibiotics and determining

the number of CFU. Groups of 5 mice were infected by oral gavage with an approximate 1:1 mixture of mutant and isogenic parents at a dose of approximately 10^7 CFU/strain/mouse. The caecum, 3 Peyer's patches of the terminal ileum adjacent to the caecum, the mesenteric lymph nodes, the liver and the spleen were harvested aseptically and homogenized in 5 ml of phosphate-buffered saline, pH 7.4. Dilutions of homogenized organs were plated on LB plates kanamycin and X-phos (Sigma) to distinguished between colonies expressing PhoN and colonies that were PhoN negative (RAK113). Data were normalized by dividing the output ratio (CFU of the mutant/CFU of the wild type) by the input ratio (CFU of the mutant/CFU of the wild type). In case only one bacterial strain was recovered from fecal pellets, the limit of detection was determined for the missing strain and used to calculate the minimum mutant-to-wild type ratio. All data were converted logarithmically prior to the calculation of averages and statistical analysis. A Student's *t*-test was used to determine whether the log value of the mutant-to-wild type ratio recovered from infected organs was significantly different ($p<0.05$) from the log value of the mutant-to-wild type ratio present in the inoculum.

2.7 LC-MS analysis

Experimental design, sample preparation, resolving of proteins, excision of bands and digestion of proteins was performed by Tim Perkins. Protein sequencing and database searching was done by Team 17 (WTSI).

Reduced and alkylated proteins were resolved as in 2.5.14.2. Protein bands were excised, destained completely, and digested with trypsin (sequencing grade; Roche). Peptides were extracted from gel with 5% formic acid–50% acetonitrile, dried. The

extracted peptides were redissolved in 0.5% formic acid (FA) and analyzed with on-line nano LC-MS/MS on an Ultimate 3000 Nano/Capillary LC System (Dionex) coupled to a LTQ FT Ultra mass spectrometer (ThermoElectron) equipped with a nanoelectrospray ion source (NSI). Samples were first loaded and desalted on a PepMap C18 trap (0.3 mm id x 5 mm, Dionex) at 25 $\mu lmin^{-1-}$ with 0.1% FA for 5 min then separated on a BEH C18 analytical column (75 µm id x 10 cm) (Waters) over a 30 or 45 or 60 min linear gradient of 4-32% CH_3CN/0.1% FA based on the gel band's size and intensity. The LTQ FT Ultra mass spectrometer was operated in the standard data dependent acquisition mode controlled by Xcalibur 2.0. The survey scans (m/z 400-1500) were acquired on the FT-ICR at a resolution of 100,000 at m/z 400 and one microscan per spectrum. The three most abundant doubly and triply charged ions (2+ and 3+) with a minimal intensity at 1000 counts were subject to MS/MS in the linear ion trap at an isolation width of 2 Th. Precursor activation was performed with an activation time of 30 msec and the activation Q at 0.25. The normalised collision energy was set at 35%. The dynamic exclusion width was set at ±10 ppm with 1 repeats within 45 sec and excluding for 60 sec. To achieve high mass accuracy, the automatic gain control (AGC) target value was regulated at $1x10^6$ for FT and $1x10^4$ for the ion trap, with maximum injection time at 1000 msec for FT, and 250 msec for ion trap respectively. The instrument was externally calibrated using the standard calibration mixture of caffeine, MRFA and Ultramark 1600

The Raw files were processed by BioWorks 3.3 and then submitted to a database search in Mascot server 2.2 (www.MatrixScience.com) against an in-house built Typhi Ty2 genomic 6-frame translated database using following search parameters: trypsin/P with 2 mis-cleavage, 20 ppm for MS, 0.5 Da for MS/MS, with 9 variable modification of Acetyl (N-term), Carbamidomethyl (C), Deamidated (NQ),

Dioxidation (M), Formyl (N-term), Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E), Methyl (E) and Oxidation (M).

Database search results were post-processed with an in-house tool that employs Percolator, a machine learning method for MS data analysis. This aids improved sensitivity and provides robust error measures using a random database as a null model. All peptides with a posterior error probability (probability that an individual peptide was identified by chance alone) of 1% or less were accepted for subsequent analysis, resulting in an overall false discovery rate of 0.076%.

2.8 Bioinformatics tools

2.8.1 Artemis and ACT

Artemis is a genome browsing tool developed for genome annotation and analysis at the Wellcome Trust Sanger Institute (WTSI) and the manual is available at http://www.sanger.ac.uk/Software/Artemis/. Important features of Artemis that made this study possible are outlined below. A new entry can be created and features copied, moved to or created within this feature. Features can be classified accordingly and examples include coding sequences, islands, misc_RNA, binding sites. Feature can also be selected by the details contained within the notes annotated to each feature. The published annotation for each feature usually includes these details and each feature can be easily edited. It is also possible to assign systematic gene names to each selected feature and select features on the forward or reverse strand. Keys and IDs can also be altered easily. In the CT18 annotation each feature was assigned a colour according to its functional class and these details can be easily searched for. An example of an annotated feature is below.

CDS 337..2799

/gene="thrA"

/locus_tag="t0002"

/EC_number="2.7.2.4"

/EC_number="1.1.1.13"

/note="multifunctional homotetrameric enzyme that catalyzes the phosphorylation of aspartate to form aspartyl-4-phosphate as well as conversion of aspartate semialdehyde to homoserine; functions in a number of amino acid biosynthetic pathways"

/codon_start=1

/transl_table=11

/product="bifunctional aspartokinase I/homeserine dehydrogenase I"

/protein_id="NP_803887.1"

/db_xref="GI:29140545"

/db_xref="GeneID:1066974"

/translation="MRVLKFGGTSVANAERFLRVADILESNSRQGQVATVLSAPAKIT
NHLVAMIEKTIGGQDALPNISDAERIFSDLLAGLASAQPGFPLARLKMVVEQE
FAQIKHVLHGISLLGQCPDSINAALICRGEKMSIAIMAGLLEARGHRVTVIDPV
EKLLAVGHYLESTVDIAESTRRIAASQIPADHMILMAGFTAGNEKGELVVLGR
NGSDYSAAVLAACLRADCCEIWTDVDGVYTCDPRQVPDARLLKSMSYQEA

MELSYFGAKVLHPRTITPIAQFQIPCLIKNTGNPQAPGTLIGASSDDDNLPVKGI
SNLNNMAMFSVSGPGMKGMIGMAARVFAAMSRAGISVVLITQSSSEYSISFC
VPQSDCARARRAMQDEFYLELKEGLLEPLAVTERLAIISVVGDGMRTLRGISA
KFFAALARANINIVAIAQGSSERSISVVVNNDDATTGVRVTHQMLFNTDQVIE
VFVIGVGGVGGALLEQLKRQQTWLKNKHIDLRVCGVANSKALLTNVHGLNL
DNWQAELAQANAPFNLGRLIRLVKEYHLLNPVIVDCTSSQAVADQYADFLRE
GFHVVTPNKKANTSSMDYYHQLRFAAAQSRRKFLYDTNVGAGLPVIENLQN
LLNAGDELQKFSGILSGSLSFIFGKLEEGMSLSQATALAREMGYTEPDPRDDL
SGMDVARKLLILARETGRELELSDIVIEPVLPDEFDASGDVTAFMAHLPQLDD
AFAARVAKARDEGKVLRYVGNIEEDGVCRVKIAEVDGNDPLFKVKNGENAL
AFYSHYYQPLPLVLRGYGAGNDVTAAGVFADLLRTLSWKLGV"

An important feature for this study is the graph function. It is possible to plot transcription data or chIP-seq data and it is similar to the GC plot function. Using the menu graph->Add user plot the file needs to have one value per base and every base must be represented, right clicking on the plot enables scaling of the local plot to the highest peak.

2.8.2 Mapping software

2.8.2.1 Transcriptome mapping software

The WTSI developed a data analysis pipeline for managing the enormous datasets produced by high-throughput sequencing. Each file can contain in the region of 12.5 million reads consisting of 450 million bp. For this study we map the sequence reads to the Ty2 genome using the programme MAQ [185]. The sequence data file (fastq)

and the reference genome file (fasta) were converted to binary format using the maq fastq2bfq and maq fasta2bfa command line respectively.

% maq fastq2bfq [-n nreads] <in.fastq> <out.prefix>|<out.bfq>

% maq fasta2bfa <in.fasta> <out.bfa>

Then maq map command was used to map the bfq file to the reference bfa file.

% maq map [options] <out.map> <chr.bfa> <reads_1.bfq> [reads_2.bfq]

These data were then "piled up" using maq –q 30 pileup by aligning each of the sequence reads to the genome, assigning a score to each base-pair and a description of the strand it is aligned to.

% maq pileup [options] <chr.bfa> <align.map>

The quality (-q) option was used for all mapping and was stipulated at 30. An example of the pileup output is below.

Forward strand all_bases 7887 G 45 @.....

Reverse strand all_bases 914 G 6 @,,,,,,.

Overlapping Strands all_bases 7690 G 38 @,,,,,.,.,,..,,,,,,,.....

To produce a plot readable into Artemis, we used the unix command

% awk '{print $4;}' filename.pileup > filename.plot.

2.8.2.1 Protein mapping scripts

The protein mapping script was written by Keith James, microarray facility (WTSI) for general use.

2.8.3 Perl scripts

2.8.3.1 Pileup to stranded plot - maqpileup2depth.pl

To produce a stranded plot for Artemis we derived strand information from the pileup plot using the script maqpileup2depth.pl (written by Miao He, WTSI, appendix 9.1). STOUT is a filename.plot containing two columns separated by a space representing reverse (column 1) and forward strand (column 2).

% ./maqpileup2depth STDIN.pileup > STDOUT

2.8.3.2 Local plots

Due to the shear size of each plot use the extractLines.pl (written by Kathryn Holt, WTSI, appendix 9.2) to extract specific regions of the Artemis plot.

% ./extractLines.pl –i filename.plot –start –stop > STDOUT

2.8.3.3 Extracting expression data from plot

To extract data for any feature we used a script called tram.pl (written by Sammy Assefa, WTSI, appendix 9.3) This script quantifies sequence data aligned to each feature and take into the uniqueness of the sequence in each feature. Uniqueness is

determined by mapping the entire genome back to itself, thus it filtered out any repetitive sequences.

% ./tram.pl <GFF file> < Specific strand plot file > < Uniqueness plot > 25 1 1 > STDOUT

Example output

		Start	End	Len.	Uniq	No-cov	nonU	%Ulen	GM	AM	Med	Max	Min
Misc RNA	rfam_100.1	1	189	189	148	16	24	78.31	3.51	47.99	20	195	0
Misc RNA	Thr_leader	191	311	121	83	38	0	68.6	2.48	6.89	8	21	0
Misc RNA	rfam_107c	5014	5123	110	77	33	0	70	1.03	0.76	1	2	0

Start, first base; stop, last base; Len., number of bases in feature, Uniq, number of uniquely covered bases; No-cov, number of bases with no coverage; nonU, number of bases with coverage above zero but less than the uniqueness cutoff; %Ulen, percentage of bases that are uniquely covered; GM, geometric mean; AM, arithmetic mean; Med, median; Max, maximum nucleotides one base has mapped; Min, minimum number of nucleotides one base has mapped.

2.8.4 R

R is a freeware package for statistical analysis [186]. Various analysis libraries are available freely from published packages.

2.8.4.1 Limma analysis

For differential analysis of transcriptome data we used the script outlined in Appendix 9.4 [187].

2.8.5 Genespring

For differential analysis of spotted microarray analysis we used Genespring GX (Agilent).

2.8.6 Graph pad prism 5.0

To plot and analyse data we used the programme Graph Pad Prism 5.0a for OS X. This is a self-contained package and contains standard comprehensive statistical analysis.