In vivo transcriptome analysis provides insights into host-dependent expression of virulence factors by *Yersinia entomophaga* MH96, during infection of *Galleria mellonella*.

Preprint Version – September 2020

Amber R. Paulson^{* †, ‡, §}, Maureen O’Callaghan[†], Xue-Xian Zhang^{**},
Paul B. Rainey^{‡, ††, ‡‡}, and Mark R.H. Hurst[†]

[†]*Forage Science, AgResearch Ltd., Lincoln, New Zealand*

[‡]*New Zealand Institute for Advanced Study, Massey University, Auckland, New Zealand*

[§]*Biology Department, Queen’s University, Kingston, Ontario, Canada*

^{**}*School of Natural and Computational Sciences, Massey University, Auckland, New Zealand*

^{††}*Laboratoire de Génétique de l’Evolution, ESPCI ParisTech, Paris France*

^{‡‡}*Department of Microbial Population Biology, Max Planck Institute for Evolutionary Biology, Plön, Germany*

Running title: *In vivo* transcriptome of *Y. entomophaga*

Keywords: Transcriptome, entomopathogenic bacteria, insecticidal toxins, *Galleria mellonella*, *Yersinia* virulence factors

Corresponding author: Amber Paulson, Department of Biology, Queen’s University,
116 Barrie Street, Kingston, Ontario Canada K7L 3N6; +1 343-364-7822;

Amber.Rose.Paulson[at]gmail[dot]com

32 Abstract

33 The function of microbes can be inferred from knowledge of genes specifically
 34 expressed in natural environments. Here we report the *in vivo* transcriptome of the
 35 entomopathogenic bacterium *Yersinia entomophaga* MH96, captured during three
 36 stages of intrahemocoelic infection in *Galleria mellonella*. A total of 1,285 genes
 37 were significantly upregulated by MH96 during infection; 829 genes responded to *in*
 38 *vivo* conditions during at least one stage of infection, 289 responded during two
 39 stages of infection and 167 transcripts responded throughout all three stages of
 40 infection. Genes upregulated during the earliest infection stage included
 41 components of the insecticidal toxin complex Yen-TC (*chi1*, *chi2* and *yenC1*), genes
 42 for rearrangement hotspot element containing protein *yenC3*, cytolethal distending
 43 toxin *cdtAB* and vegetative insecticidal toxin *vip2*. Genes more highly expressed
 44 throughout the infection cycle included the putative heat-stable enterotoxin *yenT*
 45 and three adhesins (usher-chaperone fimbria, filamentous hemagglutinin and an
 46 AidA-like secreted adhesin). Clustering and functional enrichment of gene expression
 47 data also revealed expression of genes encoding type III and VI secretion system-
 48 associated effectors. Together these data provide insight into the pathobiology of
 49 MH96 and serve as an important resource supporting efforts to identify novel
 50 insecticidal agents.

51

52 Introduction

53 For successful invasion, colonization and bioconversion of host tissues, bacterial
 54 pathogens must acclimate to the host environment via alterations in patterns of
 55 gene expression (Crawford et al., 2010; Fang et al., 2016; Heroven & Dersch, 2014;
 56 Reniere, 2018; Tran et al., 2013). Knowledge of those genes activated specifically
 57 within host tissues allows insight into the genetic determinants of *in vivo*
 58 colonization and pathogenicity. An often used strategy involves transcriptional
 59 analysis of pathogen gene expression (transcriptomics) based on patterns of mRNA
 60 transcript abundance during infection (Arenas et al., 2019; Deng et al., 2018; Fan et
 61 al., 2020; Griesenauer et al., 2019; Haueisen et al., 2019; Ibberson & Whiteley, 2019;
 62 Luo et al., 2020; Nobori et al., 2018; Sun et al., 2019; Tang et al., 2019; Valenzuela-
 63 Miranda & Gallardo-Escárate, 2018).

64 *In vivo* transcriptomic approaches have been widely applied to human pathogenic
 65 bacteria using cell lines and model hosts (Avican et al., 2015; Baddal et al., 2016;
 66 Damron et al., 2016; Nuss et al., 2017; Westermann et al., 2016; Yan et al., 2013).

67 Additionally, time-resolved *in vivo* transcriptomics have increased understanding of
 68 the function of bacterial pathogens during natural infection processes (Aprianto et
 69 al., 2016; Li et al., 2020; Luo et al., 2020; Tang et al., 2020; Westermann et al., 2019).

70 However, similar strategies are yet to be applied to pathogens infecting insects,
 71 despite the obvious potential to identify new insect-active agents.

72 *Yersinia entomophaga* MH96 is an entomopathogenic bacterium that was isolated
 73 from *Costelytra giveni* (Coleoptera: Scarabaeidae) (Hurst et al., 2011), an endemic
 74 and economically significant pasture pest in New Zealand (Ferguson et al., 2019).

75 Development of MH96 as a biopesticide has proven consistent pathogenesis by *per*
76 *os* challenge against *C. giveni*, as well as a wide range of coleopteran, lepidopteran
77 and orthopteran species (Hurst et al., 2015, 2011, 2019, 2014; Marshall et al., 2012).
78 The key virulence determinant of MH96 is an insecticidal toxin complex (TC) called
79 Yen-TC, which facilitates ingress of MH96 through the insect midgut (Hurst et al.,
80 2011, 2014; Marshall et al., 2012) and is constitutively produced and secreted by
81 MH96 when grown in broth culture at temperatures of 25 °C or lower (Hurst et al.,
82 2011). In addition to the orally active insecticidal activities of Yen-TC, structural
83 characterization and understanding the molecular mechanisms of pore-formation
84 and translocation of cytotoxin has been the focus of much research (Busby et al.,
85 2012, 2013; Landsberg et al., 2011; Piper et al., 2019).
86 Despite the importance of Yen-TC in MH96's ability to breach the insect-midgut,
87 additional investigations have suggested that as yet unknown virulence factors (VFs)
88 are deployed during intrahemocoelic infection of *G. mellonella*. As a prime example,
89 Hurst et al. 2015 showed that both wild-type MH96 and Δ TC (a Yen-TC deficient
90 strain) have the same median lethal dose (LD₅₀) of only ~ 3 cells and were equally as
91 effective at stopping the host's phenoloxidase immune response when injected
92 directly into the hemocoel of *G. mellonella* (Hurst et al., 2015). Genes for many other
93 diverse putative VFs, including known insecticidal toxins and effectors, secretion and
94 iron acquisition systems, proteolytic enzymes and adhesins have also been identified
95 in the MH96 draft genome (Hurst et al., 2016). The extensive repertoire of putative
96 VF-encoding genes represents an opportunity to utilize *in vivo* transcriptomics to
97 further characterize this entomopathogen's mode of action against insects,
98 especially during intrahemocoelic infection of *G. mellonella*.

Larvae of the greater wax moth *G. mellonella* (Lepidoptera) have emerged as a valuable model host for study of microbial infection and innate immunity (Champion et al., 2016; Desbois & McMillan, 2015; Tsai et al., 2016). Yet, to date, pathogen *in vivo* transcriptome studies from *G. mellonella* have been confined to fungal pathogens only, including *Candida albicans* (Amorim-Vaz et al., 2015) and entomopathogenic *Beauveria bassiana* (Chen et al., 2018). Since the immune system of *G. mellonella* consists of both cellular (i.e., hemocytes) and humoral components (Singkum et al., 2019; Wojda, 2017), the intrahemocoelic infection is useful to characterize host-dependent genome-wide transcriptional modulation by MH96 in response to three stages of infection using a time-resolved approach.

Here, we focus on the interaction between MH96 and *G. mellonella*, using an *in vivo* transcriptomic strategy to identify pathogen genes activated under insect immunity, septicemia and pre-cadaver stages of intrahemocoelic infection. Differential expression (DE) analysis comparing MH96 transcriptome under *in vivo* and *in vitro* conditions identified pathogen genes, including known and putative VFs, that significantly responded either at specific or multiple timepoints throughout the infection process. To further understand the pathobiology of MH96, time-resolved *in vivo* versus *in vitro* fold-changes were clustered and functionally annotated using a database of known bacterial VFs. Particularly notable was upregulation of multiple genes encoding predicted adhesins, toxins and type III secretion system (T3SS) and T6SS translocated effectors.

Materials and Methods

Intrahemocoelic infection of G. mellonella and pathogen RNA enrichment

Larval *G. mellonella* (Biosuppliers, Auckland, New Zealand) were provided with fresh diet of liquid honey, glycerol, Farex® brand baby rice cereal and granular yeast for up to one week. Healthy final instar larvae of similar lengths (> 20 mm, between 0.15 – 0.30 g) were selected and immobilized on ice. For intrahemocoelic infection, MH96 was grown overnight (18 h) in 3 ml Luria Broth Base (Invitrogen) with shaking (250 rpm) at 30 °C and then cultures were diluted in phosphate buffer solution (PBS, Sigma) and placed on ice prior to injection. Larvae were surface sterilized with 70 % ethanol and then injected below the third right leg with 10 µl of inoculum using a 30 gauge 12 mm needle on a 1 ml tuberculin syringe (Terumo) with a micro-injector. Dilutions of MH96 ranging from ~ 157 up to 2 x 10⁷ colony forming units (CFUs) were injected into *G. mellonella* depending on the stage of infection for each *in vivo* treatment (i.e., early, middle or late) (supplemental Table S1). For the early infection treatment, *G. mellonella* were injected with approximately 2 x 10⁷ CFUs and hemolymph was collected within 75 – 90 min post-infection, which was considered sufficient time to trigger a measurable host-dependent transcriptional response by MH96 under insect immunity. With respect to middle and late *in vivo* treatments, a 48 hours-post infection (HPI) growth curve was first obtained. MH96 CFU/larva were enumerated by dilution plating of *G. mellonella* homogenate on *Yersinia* semi-selective medium at specific time-points following injection of ~ 345 CFUs (Supplementary Figure S1). This enabled definition of the 18-19 and 26-30 h

incubation times for middle (i.e., septicemia) and late infection (i.e. pre-cadaver)

treatments, respectively. Also, similar to the 48 HPI growth curve, lower inocula

dosages of ~157 and 693 CFUs were injected for the middle and late infection

treatments, respectively compared to the early infection.

Following injection, the larvae were placed in petri-dishes and incubated at 25 °C

within sealed plastic bags with moistened paper towel. After incubation, hemolymph

was obtained by puncturing the dorsal cuticle near the second segment with a

needle. Hemolymph was pooled from five to six individuals, yielding approximately

150 µl total volume. Four biological replicate samples were collected from each of

the three *in vivo* infection stages (12 *in vivo* samples total). Pooled hemolymph

samples were collected directly into microcentrifuge tubes with 300 µl of

RNAprotect Bacterial Reagent (Qiagen), vortexed for five seconds, incubated at room

temperature for 5 min and then centrifuged at 300 x g for 5 min at 4 °C to pellet

hemocytes and cell debris. The supernatant was carefully separated from the host

cell pellet by pipette and bacterial cells were then pelleted by centrifugation (10 min

at 5,000 x g at ambient temperature (~ 22 °C)) and the supernatant removed and

stored at -80 °C. To estimate the CFUs per larva, serial dilutions from *G. mellonella*

homogenate were plated on a *Yersinia*-semi selective medium and all *Yersinia*-

positive colonies were counted within 24 h (5 separate larva per treatment; surface

sterilized with 70 % ethanol). The weight of each larva was measured enabling

determination of CFU per gram larval homogenate.

Control *in vitro* RNA samples were collected separately from broth cultures and two

independent biological replicate samples were collected for each of the *in vitro*

treatments – lag, exponential and stationary growth phases (6 *in vitro* samples total).

For *in vitro* growth MH96 was grown overnight in 3 ml broth at 30 °C with shaking

(250 rpm). The next day, flasks containing 50 ml fresh broth were inoculated with 1

% overnight culture and grown at 25 °C until samples were collected after 4 h 45 min

(2 ml) , 6 h (1 ml) and 9 h (1 ml) corresponding to specific cell densities of $3.2 \pm 0.1 \times$

10^7 , $2.9 \pm 0.9 \times 10^8$ and $4.4 \pm 0.1 \times 10^9$ CFU ml⁻¹ (comparable to CFU g⁻¹ assessed from

in vivo *G. mellonella* hemolymph RNA-seq samples), respectively. *In vitro* RNA

samples were stabilized in RNeasy Protect Bacterial Reagent (Qiagen) according to the

manufacturer's protocol.

RNA Extraction, library preparation and sequencing

Within one week of collection, RNA samples were re-suspended in 100 µL Tris-EDTA

buffer (10mM Tris-HCl, 1mM EDTA; pH 8.0) (Ambion) containing 15 mg ml⁻¹ lysozyme

(Roche or Sigma-Aldrich) and 10 µl proteinase K (Roche). The samples were

incubated for 10 min, interrupted by 10 s vortex every two min to lyse the cells. The

RNeasy mini kit (Qiagen) was used for RNA extraction and RNA clean-up, which

included two separate DNA digestions (on-column then off-column) with DNase

(Qiagen), all according to manufacturer's guidelines. RNA samples were then

precipitated at -20 °C overnight in 3 volumes 100 % cold isopropanol with 1/10

volumes 3M sodium acetate (pH 5.2), followed by two 70 % cold ethanol washes.

Purified RNA was re-suspended in RNase-free water and stored at -80 °C.

RNA extract from early infection hemolymph contained approximately half the

amount of total RNA required for sequencing, so early infection RNA samples were

doubled such that each sample was comprised from a pool of 10 - 12 larvae, while middle and late infection RNA samples were from five or six pooled individuals. Sufficient removal of genomic DNA from the samples was determined by confirming lack of PCR amplification from *in vivo* RNA template using primers RecA_352F 5' – TCTCAGCCAGATACCGGTGA and RecA_987R 5'- CAGCAACATTTTCACGCAGCT for the housekeeping gene *recA*. Middle and late infection *in vivo* RNA samples containing ~7.2 µg total RNA and early infection samples containing < 7.2 µg were stabilized on RNastable 1.5 micro-centrifuge tubes (Biomātrica) by SpeedVac (Savant) at ambient temperature (~ 22 °C) for 1.5 h.

Macrogen Korea carried out library preparation and sequencing. Following recovery from the RNastable tubes, RNA quantity and Integrity Number were determined using the 2200 TapeStation (Agilent Technologies). Host and bacterial ribosomal RNA (rRNA) was depleted from the *in vivo* samples using the ScriptSeq Complete Gold Kit (Epidemiology) (Illumina) then strand-specific cDNA libraries were prepared using the ScriptSeq RNA-Seq Library Preparation Kit (Illumina) according to the manufacturer's guidelines. The *in vitro* samples were prepared using the TruSeq stranded mRNA (Illumina) with the Ribo-Zero rRNA removal kit for Bacteria (Epicentre) according to standard protocol. To verify the size of the PCR enriched fragments, the template size distribution was assessed on the 2100 Bioanalyzer using a DNA 1000 chip (Agilent Technologies). The libraries were sequenced on the HiSeq2500 platform (Illumina) (Version HCS v2.2) to generate 101 bp paired-end reads.

Bioinformatic analysis

Adapter and barcode contamination were filtered using *Flexbar* (v2.4) against Illumina adapter and barcode reference sequences (supplemental Table S2 and Table S3). Reads were trimmed in ‘ANY’ mode with minimum overlap of five, maximum uncalled base of one and minimum quality threshold of 20 (Dodt et al., 2012). The *bbsplit.sh* function of the *BBMap* package was used to quantify and remove contaminating rRNA sequences with minimum ratio of 0.5, minimum hit number of one and maximum insertion/deletion length of 500 bp (Bushnell, 2015). All libraries were screened against full-length 5S, 16S and 23S rRNA sequence from the MH96 genome (see accession information below). To remove host sequences, *in vivo* libraries were also mapped against the *G. mellonella* partial 18S and 28S rRNA sequence retrieved from GenBank (Accessions: [U65198.1](#), [U65138.1](#), [AF286298.1](#), [AF423921.1](#) and [X89491.1](#)) and the *G. mellonella* immune cDNA library (JG394435 – JG406465) (Vogel et al., 2011).

Remaining paired-end reads were aligned to the MH96 genome (Accession: CP010029.1, [GI: 1034308998](#)) using *Rockhopper* (v2.03) in verbose mode. Alignment stringencies were increased from default values, such that allowable mismatch threshold was decreased from 15 to 10 % of read length and the minimum seed was increased from 33 % to 50 % of read length. Any residual count data aligning to 5S, 16S and 23S rRNA genes or any short (less than 50 bp) non-coding RNA or anti-sense RNA were not considered further in the analysis.

The early infection *in vivo* libraries were investigated whether the sequence depth was estimated to have adequately captured the majority of MH96 low-abundance transcripts using the *iNEXT* R package (nboot = 50) (v2.0.18) (Hsieh et al., 2016). The

RUUVseq R package (v1.16.0) was used to generate relative log expression (RLE) plots and conduct principle component analysis (PCA) (Risso et al., 2014). Upper-quartile normalized (Bullard et al., 2010) libraries were filtered for transcripts with greater than 5 aligned reads in at least two libraries.

Normalized expression data were then converted to log₂-counts-per-million (Log₂CPM) using the R package *limma* (v3.38.3) (Ritchie et al., 2015; Smyth, 2004) and the mean-variance relationship was modeled using the ‘voom’ method (Law et al., 2014). Genome-wide host-dependent responses by MH96 to both *in vitro* and *in vivo* conditions were visualized with *Circos* (v0.69) (Krzywinski et al., 2009) and putative genomic islands (GIs) were assigned using the webserver [IslandViewer 4](#) (Bertelli et al., 2017). To determine MH96 genes that significantly responded to *in vivo* conditions a DE analysis was undertaken by comparing average transcript abundance between *in vivo* and *in vitro* libraries sharing equivalent cell densities (i.e., early infection vs. lag growth in broth, middle infection vs. exponential growth in broth and late infection vs. stationary growth in broth). Genes with significantly DE were defined by multiple testing with false discovery rate of 0.05 using the Benjamini and Yekutieli multiple test correction method (Benjamini & Yekutieli, 2001) using *limma*.

Time-resolved average *in vivo* vs. *in vitro* log₂CPM fold-change values for all genes with significantly DE identified between the *in vivo* and *in vitro* treatments (DE analysis described above) were clustered using the fuzzy c-means (FCM) algorithm (Futschik, Matthias & Carlisle, 2005) with the *MFuzz* R package (v2.42.0) (Kumar & Futschik, 2007). A heuristic qualitative assessment was used to identify both the

optimal FCM parameter (i.e., ‘fuzzifier’) and number of clusters (Kumar & Futschik, 2007). The assessment considered both predicted operon structure (output from *Rockhopper*) and functional annotations assigned using homology based search (BLASTx, e-value cut-off = 1×10^{-8}) against the [virulence factor database](#) (VFDB) (accessed July, 2018) (Chen et al., 2016).

MH96 host-specific factors were identified among clusters containing transcripts with comparably greater median $\log_2$ CPM *in vivo* fold-change values ($\geq \sim 2$) for at least one stage of growth during intrahemocoelic infection. Next, these host-specific factors were functionally characterized as putative VFs when significant sequence homology was identified against an already characterized VF found in the VFDB. Finally, functional categories of virulence were assigned to all of these putative VFs primarily based on the VFDB-derived annotation, in accordance to our definitions found in supplemental Table S4. When VFDB-derived annotations were not informative, additional information such as, top-scoring BLASTx hits from the GenBank non-redundant protein database, were used to assign a possible function or otherwise these hypotheticals were left unclassified.

Data availability

Sequence data and both raw and normalized count tables were deposited in the GenBank Geo Submission Omnibus under accession: [GSE142509](#). All genes with significantly DE detected between *in vivo* versus *in vitro* time-resolved contrasts (early infection vs. lag growth *in vitro*; middle infection vs. exponential growth *in vitro*; late infection vs. stationary growth *in vitro*) are provided in supplementary Table S5. All functionally annotated putative VFs identified from host-specific fuzzy

280 clusters are presented in supplementary Table S6. Open-source R code for statistical
281 analysis is available at:
282 https://github.com/damselflywingz/In_vivo_infection_series_fuzzy_RNA-seq.

Results

Experimental infection and RNA sequencing

MH96 was enriched from hemolymph at cell densities of 4.4×10^7 , 4.4×10^8 and 5.0×10^9 CFU g⁻¹ (Table 1). In order to distinguish host-specific gene induction by MH96 from response to general growth conditions, *in vitro* samples were collected across three growth phases at approximately equivalent cell densities to the *in vivo* samples (Table 1).

Table 1: MH96 cell density under *in vivo* and *in vitro* conditions used for differential expression analysis.

<i>In vivo</i> infection stage/ <i>in vitro</i> growth	Mean cell density <i>in vivo</i> (i.e., <i>G. mellonella</i> homogenate) (CFU ^a g ⁻¹ ± SEM ^b)	Mean cell density <i>in vitro</i> (i.e., Luria Broth Base broth) (CFU ml ⁻¹ ± SEM)
Early infection/lag-growth	$4.4 \pm 0.5 \times 10^7$	$3.2 \pm 0.1 \times 10^7$
Middle infection/exponential growth	$4.4 \pm 1.7 \times 10^8$	$2.9 \pm 0.9 \times 10^8$
Late infection/stationary growth	$5.0 \pm 2.2 \times 10^9$	$4.4 \pm 0.1 \times 10^9$

^aCFU = colony forming units; ^bSEM = standard error of the mean.

After quality trimming, adapter removal of host and rRNA sequences, the paired-end reads were aligned to the draft MH96 genome. Only an average of 9 % of reads from early-infection libraries could be aligned to the reference, which was lower by comparison to the average reads aligned from middle (58 %) and late infection (59 %), or *in vitro* libraries (84 -93 %) (supplemental Figure S2 and Table S7). In order to determine if transcripts expressed in lower abundances were adequately captured within the early infection libraries, these libraries were analyzed by rarefaction,

which showed that only a slight improvement in total transcript detection if sequencing depth were increased (supplemental Figure S3). Based on this analysis, the *in vivo* early infection libraries were determined to have captured most MH96 transcripts produced across a dynamic range of conditions expected under these conditions and were considered saturated.

Preliminary data analysis explored the libraries for unwanted technical variation arising from batch effects and potential outliers using both RLE plots and PCA (supplementary Figure S4 and Figure S5), respectively. Variation in sequencing depth between libraries was found to have affected the distribution of RLE values (supplementary Figure S4A), which was largely controlled for by upper quartile normalization (supplementary Figure S4B). Even after normalization, the variation in RLE values from the stationary growth *in vitro* libraries was greater when compared to all other libraries indicating residual unwanted technical variation arose from this treatment. A clear separation between *in vitro* and *in vivo* libraries was determined by PCA of normalized count data (supplemental Figure S5). Libraries consistently clustered by treatment, except one *in vivo* late infection library, which clustered more closely to the middle infection *in vivo* libraries and was identified as a biological outlier and removed from subsequent analysis.

Differential expression of MH96 during early, middle and late infection stages

Comparison of average transcript abundances across three stages of growth identified 2,397 differentially expressed genes that were either up or downregulated between *in vivo* and *in vitro* libraries (Supplemental Figure S6 and Table S5). MH96 upregulated 1,285 genes during intrahemocoelic infection of *G. mellonella*; 829

323 genes responded to *in vivo* conditions during at least one stage of infection, 289
 324 responded during two stages of infection and 167 transcripts responded throughout
 325 all three stages of infection. MH96 also upregulated 1,243 genes under *in vitro*
 326 conditions, of which 131 were also found to be upregulated under *in vivo* conditions
 327 at different cell densities.

328 Regions of interest from the MH96 draft genome either known or suspected to
 329 encode VFs (supplemental Table S8) were explored with respect to time-resolved
 330 genome-wide host-dependent transcriptional responses during intrahemocoelic
 331 infection (Figure 1). Several predicted GIs partially or fully overlapped with
 332 previously identified genomic regions of interest (i.e., unique region 1 overlaps Is.2,
 333 Rhs2 overlaps Is.5, T3SSYE1 overlaps Is.6, Rhs5 overlaps Is.11, T3SSYE2 overlaps
 334 Is.12, Rhs4 overlaps Is.14, unique region 2 overlaps Is.16, T2SS overlaps Is.18 and
 335 Rhs3/T6SS overlaps Is.21).

Figure 1: Circos plot of *Y. entomophaga* MH96 genome-wide transcriptional response to *in vivo* and *in vitro* conditions during early, middle and late infection and lag, exponential and stationary growth, respectively. CPM = Counts-per-million. Mean fold-change in log counts-per-million (Log_2CPM) (*in vivo* early and middle infection: $n = 4$, *in vivo* late infection: $n = 3$ and all *in vitro*: $n = 2$). Red circles numbered 1-21 represent predicted genomic islands identified using IslandViewer 4 webserver (Refer to supplemental Table S8). Regions of interest/loci previously identified from the draft genome are numbered or labelled as required (Hurst et al., 2016).

Wide-spread host-dependent upregulation of multi-gene clusters located within several regions of interest/predicted GIs, including unique regions 1 and 2, Rhs2/Is.5, T3SSYE2/Is.12 and T2SS/Is.18, was observed in MH96. Significantly higher expression for the entire T2SS island (PL78_08930 – PL78_08970) was measured during early

and/or late infection stages compared to *in vitro* growth conditions. Significantly higher responses by genes for putative T3SSYE2 structural components (SpaS/PL78_14595, SpaP/PL78_14580, SpoA/PL78_14575, hypothetical protein/PL78_14570, SpaL/PL78_14560, InvA/PL78_14550, and PrgH/PL78_14530) to *in vivo* growth conditions compared to *in vitro* during early and/or late infection were also identified. In contrast, most of the T3SSYE1 structural component genes were significantly upregulated by MH96 either in response to *in vitro* conditions or both *in vitro* and *in vivo* conditions depending on cell density, except putative T3SSYE1-associated chaperone (PL78_18170) that was more highly expressed *in vivo* during early infection only. Genes located within a predicted phage-like island were also significantly upregulated during early stage of infection, compared to growth *in vitro* (Is.4: PL78_06170 – PL78_06360; PL78_19225 – PL78_19085, Is.8: PL78_17412 – PL78_17386 and Is. 19: PL78_06425 – PL78_06390; PL78_19127 – PL78_19160; PL78_00015 – PL78_00300).

Genes for other putative VF s from MH96 genomic regions of interest were more responsive during late stages of intrahemocoelic infection compared to other time points. Iron-acquisition gene clusters located within unique region 2 (PL78_09605, PL78_09620 – 09640, PL78_09665 – PL78_09710, PL78_09720, PL78_09730 – PL78_09745 and PL78_09780 – PL78_09785) were found to have higher expression under *in vivo* conditions, especially during later stages of infection. Genes for the flagellum (PL78_17795 – PL78_17785, PL78_17735 – PL78_17670, PL78_17655 – PL78_17610, PL78_17595 – PL78_17585) were also more highly expressed mostly during the later infection stage too. Further, phage-related and hypothetical genes encoded within a predicted GI Is.17 (PL78_09315 – PL78_0914) were also found to

have significant transcriptional response to *in vivo* conditions compared to growth *in vitro* during late infection stages as well.

An important region in the MH96 genome is Unique Region 1 (contains PAI_{YE96}) and genes for hypothetical proteins *yenU* (PL78_03775) and *yenW* (PL78_03790), putative heat-stable enterotoxin *yenT* (PL78_03785), iron-siderophore biosynthesis/transport (PL78_03905 – PL78_03955, excluding PL78_03935 and PL78_03940) and serralyisin gene cluster (PL78_03960 – PL78_03995, excluding PL78_03985), were all more highly expressed in the host at various or throughout all stages of infection compared to *in vitro* treatments. Also located in Unique region 1, Yen-TC genes *chi1* (PL78_03740), *chi2* (PL78_03755) and *yenC1* (PL78_03765), were found to have significantly higher expression during early infection only, while DE was not detected for other Yen-TC component genes.

Functional enrichment of in vivo host clusters

Qualitative assessment of fuzzy clustering results showed that 20 clusters with fuzzifier value of 1.5 was optimal for the analysis of fold-change log₂CPM values comparing MH96 transcriptome across three different stages of intrahemocoelic infection to *in vitro* growth controls. Among these 20 clusters, eight clusters (containing 798 transcripts) exhibited the most striking responses to *in vivo* conditions compared to *in vitro* growth (supplemental Figure S7). These eight '*in vivo* clusters' were further organized into three main '*in vivo* cluster types' based on shared expression profiles across multiple *in vivo* clusters (Figure 2A). *In vivo* clusters containing transcripts with elevated abundances across all three stages of infection

396 were identified as 'Host All' type clusters, while 'Host Early' and 'Host Late' type
397 clusters were characterized by containing mostly transcripts with elevated
398 abundances during early or late stages of infection only, respectively.

399
400 Figure 2: *Yersinia entomophaga* MH96 putative host-specific virulence factors (VFs)
401 sharing significant sequence similarity to known VFs from the virulence factor
402 database identified among host-specific *in vivo* clusters. A) Median log₂ counts-per-
403 million-fold-change between *in vivo* and *in vitro* libraries organized by *in vivo* cluster
404 type (each line represents a separate host-specific *in vivo* cluster, refer to
405 supplemental Figure S7). B) The proportion of genes organized by virulence function
406 classification among *in vivo* clusters types. Refer to supplementary Table S6.

407 To further classify the role of putative MH96 VFs in pathogenesis, those transcripts
408 from the eight *in vivo* clusters were functionally annotated using VFDB. Using this
409 approach, 28.7 % of the transcripts from *in vivo* clusters could be further categorized
410 by potential role in pathogenicity (i.e., adherence, defense response, horizontal gene

411 transfer, iron acquisition, metabolic adaptation, mobility, outer membrane,
 412 regulation, secretion system and toxin/effector or unclassified) (Figure 2B, Figure 3
 413 and supplemental Table S6). Transcripts related to iron acquisition, secretion
 414 systems, regulation and metabolic adaptation were commonly identified among
 415 each of the three different *in vivo* cluster types but an obvious temporal shift in the
 416 expression of iron acquisition and flagellar motility versus toxins, effectors and
 417 adhesions were identified when comparing early and late *in vivo* clusters.

Figure 3: Number of genes assigned to each virulence category for putative virulence factors identified from *in vivo* cluster types. *In vivo* clusters combined by type: Host All = constitutively higher *in vivo* expression during earl-, middle and late infection; Host Early = higher *in vivo* during early infection only; Host Late = higher *in vivo* expression during late infection only. See Figure 2A for Host All, Host Early and Host Late median expression profiles and supplemental Table S6 for more details.

Four genes encoding an usher-chaperone fimbrial cluster (PL78_12465 – PL78_12480) co-expressed with putative T3SS lipoprotein chaperone *YscW-like* (PL78_14485), were identified as some of the most host-responsive genes in the MH96 genome, especially at the early stage of infection (*in vivo* early infection vs. *in vitro* growth fold-change change values ranging from greater than 4.6 up to 7.1x Log₂CPM). Genes for filamentous hemagglutinin (PL78_11060, previously annotated as a hemolysin and its cognate two-partner secretion system (TPS) (PL78_11055) also had significantly higher expression during early and middle stages of intrahemocoelic infection. Also, a predicted fimbrial gene (PL78_08275) and chitin-

binding protein (PL78_08295) both associated with the Chi-Fim region (Hurst et al., 2016) were also captured from *in vivo* Host all type clusters but the relative expression levels of the Chi-Fim-associated fimbrial and chitin-binding protein genes were orders of magnitude lower than genes for other putative MH96 adhesions that were upregulated *in vivo*.

Transcripts for several putative toxins and effectors were identified among *in vivo* clusters (Table 2). The aforementioned heat-stable enterotoxin *yenT* was the only putative toxin gene identified to be highly expressed *in vivo* at two timepoints during infection and the transcript encoding putative AidA-like type V secretion system (T5SS) secreted adhesin (PL78_10240) clustered with other MH96 transcripts with higher *in vivo* gene expression throughout infection (Table 2).

Several other transcripts for putative toxins were identified from Host Early *in vivo* clusters, including putative cytolethal distending toxin sub-components *cdtA* and *cdtB* (PL78_18440-PL78_18445), *yenC3* (PL78_18780, a YenC-homolog also known as *rhsA* encoded within Rhs2/Is.5) and putative vegetative insecticidal protein component, *vip2* (PL78_16145) (Table 2). Transcripts for known, and putative Yen-TC components, *chi1* and *yenC3* were captured within a Host Early *in vivo* cluster, respectively, but only *yenC3*, not *chi1*, shared any significant homology against VFs known found in the VFDB using BLASTx homology search.

Table 2: Putative MH96 toxin and effector genes from host-specific gene expression clusters with significant sequence homology to other known toxins or effectors from VFDB

Toxin/Effector type	Locus tag	Designation	Host cluster type	Region of interest/Genomic island/Locus ^a
Toxin	PL78_03785	<i>yenT</i>	Host all	Unique region 1-PAI _{YE96}
Toxin	PL78_18780	<i>yenC3</i>	Host early	Rhs-associated region 2/Is. 5
Toxin	PL78_18440	<i>cdtA</i>	Host early	<i>cdtAB</i>
Toxin	PL78_18445	<i>cdtB</i>	Host early	<i>cdtAB</i>
Toxin	PL78_16145	<i>vip2</i>	Host early	<i>vip2</i>
Effector - T3SS ^b	PL78_14610	<i>sicC-like</i>	Host all	T3SSYE2/Is.12
Effector - T3SS	PL78_07345	<i>ipgD/sopB-like</i>	Host all	-
Effector - T3SS	PL78_14615	<i>sicD-like</i>	Host early	T3SSYE2/Is.12
Effector - T3SS	PL78_18085	<i>ipaB/evcA-like</i>	Host early	T3SSYE1/Is.6
Effector - T3SS	PL78_14295	<i>sopD-like</i>	Host early	-
Tube/Effector - T6SS	PL78_02751	<i>hpc</i>	Host all	-
Tube/Effector - T6SS	PL78_00900	<i>hpc</i>	Host early	Rhs-associated region 3
Effector - T6SS ^d	PL78_01040	<i>rhs-containing</i>	Host early	Rhs-associated region 3/Is.21
Tube/Effector - T6SS	PL78_03598	<i>hpc</i>	Host late	Is.1
Tube/Effector - T6SS	PL78_09211	<i>hpc</i>	Host late	Is.17
Effector-T5SS ^c	PL78_10240	<i>aidA</i>	Host all	-

^aRegions of Interest (See Figure 1): MH96 pathogenicity-associated island (PAI_{YE96}), rearrangement-hotspot (Rhs), MH96 type 3 secretion systems (T3SSs) T3SSYE1 and T3SSYE2 (Hurst et al., 2016); Predicted genomic islands see Figure 1 and supplementary Table S8; Loci: cyto-lethal distending toxin components (CdtA, CdtB), vegetative insecticidal protein 2 (Vip2) (Hurst et al., 2016); ^bT3SS = Type III secretion system; ^cT5SS = Type V secretion system and ^dT6SS = Type VI secretion system;

In addition to predicted toxins, transcripts for T2SS and T3SSYE2 structural components were also identified as putative VFs from *in vivo* clusters (supplemental Table S6). Two genes for T3SSYE2 co-located putative effectors were represented from *in vivo* clusters and found to have significantly greater expression under *in vivo*

conditions compared to *in vitro* controls, including SicC-like (PL78_14610) and SicD-like (PL78_14615), with the former being more highly expressed across all cell densities and the latter more highly expressed during early infection only. In addition to T3SSYE2-associated effector genes, two other genes for putative T3SS effectors were also found to be more highly expressed *in vivo* during early infection including T3SSYE1-associated IpaB/EvcA-like (PL78_18085) and orphan SopD-like (PL78_14295). Another gene for putative T3SS orphan effector, IpgD/SopB-like (PL78_07345) represents a fifth possible T3SS effector, with significantly increased expression throughout all three stages of infection, compared to *in vitro* controls.

With respect to T6SS, structural components for the machinery were not identified as having significantly greater expression *in vivo* compared to *in vitro* controls, but putative T6SS effector transcripts were identified from several of the *in vivo* cluster types, including Rhs-containing protein (PL78_01040) and five different hcp-type structural tube/effector proteins (PL78_02751, PL78_03598, PL78_03602, PL78_09211 and 00900). Also, the gene encoding the putative T6SS effector that shares sequence similarity to the eukaryotic transcription elongation factor domain Spt4 (PL78_00995) was found to have higher *in vivo* gene expression throughout infection but was not assigned a VFDB annotation based on amino acid sequence.

Not all MH96 predicted toxins and effectors were highly expressed under *in vivo* condition though; genes for putative PirAB binary toxin (PL78_09590 and PL78_09595), RTX toxin (PL78_16910) and adenylate cyclase (PL78_08395) were found to have higher expression during exponential and/or stationary growth *in vitro*, with no significant response to *in vivo* conditions observed. Furthermore,

489 despite having significantly greater abundance during early-infection of *G. mellonella*
 490 compared to growth *in vitro*, transcripts for some other known or suspected MH96
 491 VFs, including insecticidal TC components Chi2 and YenC1, and putative T3SS
 492 effector LopT (PL78_18760) were not found to cluster among the eight *in vivo*
 493 clusters, instead clustering with other transcripts associated with comparatively
 494 lower overall median *in vivo* vs. *in vitro* fold-change values.

495

496

Discussion

Our *in vivo* transcriptome analysis has shed light on MH96 genes activated during intrahemocoelic infection of *G. mellonella*. Major shifts in gene expression by MH96 were observed between *in vivo* and *in vitro* conditions across all three stages of growth and the mosaic nature of the MH96 genome was highlighted by wide-spread co-expression of horizontally acquired genes for Yen-TC components and putative toxins (YenT, YenC3), translocated effectors (T3SS and T6SS), secretion systems (T2SS and T3SSYE2) and iron acquisition systems in response host-dependent conditions. These findings provide evidence that complex genome-wide responses by MH96 occur throughout different stages of insect infection and reveal important insights into host-dependent transcriptional responses by entomopathogenic bacteria in nature.

Upon entry to the hemocoel, entomopathogenic bacteria must respond to the cellular and humoral components of the insect host immune system (Nielsen-LeRoux et al., 2012). Within lepidopteran larvae, the hemolymph contains different hemocytes involved in phagocytosis, nodulation and encapsulation as well as humoral factors like prophenoloxidase, antimicrobial peptides, reactive nitrogen and oxidative species or lysozyme (Casanova-torres & Goodrich-blair, 2013; Jiang et al., 2010; Wojda, 2017). In order to occupy the insect hemocoel, MH96 produces VFs that combat the host immune defenses during early stages of infection, including anti-hemocytic factors such as toxins or translocated effectors. Upon incapacitation of the host immune system, MH96 produces additional VFs that function to mobilize and acquire essential nutrients, like iron, from insect cells or tissues to fuel pathogen

metabolism and proliferation, ultimately leading to septicemia and host death (Nielsen-LeRoux et al., 2012).

During early stages of intrahemocoelic infection MH96 upregulates expression of Yen-TC components *chi1*, *chi2* and *yenC1* as well as putative Yen-TC component *yenC3* and toxins *cdtAB* and *vip2*, which are known or suspected to be anti-hemocytic factors based on functional enrichment. CdtAB is a predicted genotoxin, which are known to interfere with normal hemocyte cell cycle (Heywood et al., 2005; Lara-Tejero & Galán, 2000), while Vip2 and YenC3 may act on host cytoskeleton through (ADP)-ribosyltransferase activity targeting globular (G)-actin (Barth et al., 2004) or Rho GTPases (Lang, Alexander et al., 2010), respectively. While Yen-TC is established as the primary orally-active virulence determinant of MH96 (Hurst et al., 2011; Marshall et al., 2012), purified Yen-TC was also found to cause ruffling of *G. mellonella* hemocytes (Hurst et al., 2015) and nuclear and actin condensation and apoptosis of *Spodoptera frugiperda* insect cell line Sf9 (Marshall et al., 2012). Taken together, Yen-TC should be considered both an anti-hemocytic factor as well as an orally-active insecticidal toxin.

Unlike putative toxin genes that were upregulated only during the early stage of infection, the MH96 heat stable enterotoxin, *yenT*, was highly expressed during both the middle and late stages of infection compared *in vitro* control treatments. Similar in size YenT shares N-terminal amino acid sequence homology to the heat-stable enterotoxins YstA and YstB (Hurst et al., 2011), encoded by gastrointestinal-disease causing strains of *Y. enterocolitica* (Delor et al., 1990; Grant et al., 1998; Ramamurthy et al., 1997; Singh & Viridi, 2004), *Y. frederiksenii*, *Y. intermedia* and *Y. kristensenii*

(Imori et al., 2017). Recently the heat-stable enterotoxin, YacT, produced by *Y. frederiksenii* has been implicated in host immune modulation following intrahemocoelic injection of *G. mellonella*. The *E. coli* recombinant YacT, altered the morphology and increased the aggregation of *G. mellonella* hemocytes (Springer et al., 2018). It is plausible that, like YacT of *Y. frederiksenii*, MH96 YenT also functions as an anti-hemocytic factor given the responsiveness of *yenT* to intrahemocoelic conditions at multiple stages of infection.

Among pathogenic *Yersinia*, T3SSs are known to target host cells through tight attachment and translocate effectors into the cytoplasm to disrupt cytoskeleton, trigger apoptosis, prevent phagocytosis or interfere with immune pathways (Bliska et al., 2013; Pha & Navarro, 2016; Plano & Schesser, 2013). The *Y. enterocolitica* T3SS Ysa (found in biotype 1B strains) (Foultier et al., 2002; Thomson et al., 2006) shares the same gene order with MH96 T3SSYE2 (Hurst et al., 2016) of the predicted island Is.12, suggesting a common origin in both pathogens. In addition to interacting with mammalian cells (Bent et al., 2013; Matsumoto & Young, 2006), Ysa is essential for replication within *Drosophila melanogaster* S2 cells (Walker et al., 2013) and secretes effectors at temperatures of less than 26 °C in nutrient-rich broth with high salt concentrations (Venecia & Young, 2005). Based on the similar gene order to Ysa and the upregulation during intrahemocoelic infection, T3SSYE2 is likely another anti-hemocytic mechanism used by MH96 to overcome the insect immune system.

Also, similar to Ysa, which secretes effectors encoded by genes within distal regions of the *Y. enterocolitica* genome (Walker & Miller, 2009), the MH96 *in vivo* transcriptome revealed evidence for coordinated host-dependent regulation of

genes for T3SSYE2 structural components and several predicted effectors, some of which were not co-located within the T3SSYE2 region of the genome. For example, MH96 encodes a gene for a putative LopT-like effector, which is homologous to LopT, a T3SS effector (similar to YopT of *Y. pestis*) with known insecticidal activity in *Photobacterium luminescens* (Brugierard-Ricaud et al., 2005). Expression of the LopT-like effector gene was significantly upregulated by MH96 during early stage infection and notably, *lopT* and *yenC3* are both co-located within an effector-island, (i.e., Rhs-associated region 2/Is.5), and are possibly co-opted by respective translocation machineries T3SSYE2 and Yen-TC and reflect the multiple diverse modes of action deployed by MH96 against the host during early stages of intrahemocoelic infection.

In contrast to T3SSYE2 and its predicted effectors that were upregulated under *in vivo* conditions, genes for T3SSYE1 of the predicted island Is.6 responded mainly to *in vitro* conditions. The differing transcriptional responses by the MH96 T3SSs depending on growth under *in vivo* and *in vitro* conditions allude to a specific role of each T3SS in MH96 pathogenicity. It was also noted that MH96 genes encoding the predicted PirAB binary toxin, RTX toxin and adenylate cyclase also responded to growth in culture, with significantly higher expression occurring *in vitro* compared *in vivo* conditions. Clearly some yet unknown environmental factor appears to control expression of T3SSYE1 and other insect active genes. It is plausible that abiotic factors like pH and concentration of oxygen or nutrients occurring during *in vitro* growth could mimic signals that MH96 is likely to encounter in its natural environment, such as intracellular replication within *G. mellonella* hemocytes (Hurst et al., 2015; Paulson, A.R., personal observation), infection of insect midgut or

different insect hosts (Hurst et al., 2019, 2011) or even the occupation of the human urinary tract during asymptomatic catheter infection (Le Guern et al., 2018).

Along with T3SSs, another major class of well-studied VFs known from pathogenic *Yersinia* are adhesins, including T5SS-secreted adhesins and fimbriae (Chauhan et al., 2016). Putative adhesin genes upregulated by MH96 throughout or during specific stages of intrahemocoelic infection included T5SS-secreted AidA-like and filamentous hemagglutinin, and four-gene usher chaperone fimbrial cluster, the latter representing some of the most responsive genes to *in vivo* conditions in the entire MH96 genome. In addition to host cell attachment or biofilm formation, in other pathogenic *Yersinia* species, adhesins have also been linked to effector translocation by through enhanced attachment to host cells (Deuschle et al., 2015; Keller et al., 2015; Maldonado-arocha et al., 2013; Mühlenkamp et al., 2015). Perhaps adhesins also contribute to the effective anti-hemocytic activities of MH96 during intrahemocoelic infection of *G. mellonella* by too enhancing effector translocation efficiency like other pathogenic *Yersinia*, yet such mechanisms remain unknown among entomopathogenic bacteria.

The MH96 *in vivo* transcriptome revealed a high degree of functional redundancy, with multiple toxins, T3SS and T6SS effectors and adhesins all likely to target hemocytes during early and/or throughout multiple stages of infection. Such functional redundancy has also been observed in another highly entomopathogenic bacterium, *P. luminescens* (Wilkinson et al., 2009), that like MH96, produces many toxins, including insecticidal TC (ffrench-Constant et al., 2007), makes caterpillar floppy (Daborn et al., 2002), PirAB binary toxin (Waterfield et al., 2005; Wu & Yi,

2016) and several RTX-like toxins (Tobias et al., 2017). Such abundance of toxins carried within the genomes of obligate entomopathogenic bacteria is thought to have arisen by adaptation for over-kill of a wide variety of hosts (French-Constant et al., 2003). This is likely also true for MH96, which has broad insecticidal activities against diverse insect hosts (Hurst et al., 2015, 2019, 2011, 2014). Also, MH96 is highly virulent against *G. mellonella* during intrahemocoelic infection but the LD₅₀ of a Yen-TC deficient strain (Δ TC) was comparable to that of wild-type (~ 3 cells). It was also found that both wild-type and Δ TC had the same ability to reduce the phenoloxidase activity of *G. mellonella* hemolymph 24 HPI, findings that implicate the presence of additional hemocytic VFs during intrahemocoelic infection by MH96 (Hurst et al., 2015). Transcriptional response by MH96 to *in vivo* conditions, including wide-spread and significant host-dependent upregulation of genes for Yen-TC components and numerous other predicted toxins, secretion systems, effectors and adhesins is another indicator that the highly insecticidal nature of MH96 stems from multiple diverse modes of action.

Beyond primary VFs that can act directly to modulate or destroy host cells, functional enrichment of the MH96 *in vivo* transcriptome also revealed changes in the temporal expression of genes for other putative in-host fitness factors (i.e., motility, metabolism, iron acquisition, response to host-induced stress, etc.) versus toxins, effectors and adhesins from early to late infection. Notably during the late stage of infection, MH96 significantly increased expression of genes related to flagellum and iron acquisition, compared to early and middle infection stages, which supports a role for iron acquisition and flagellum-related motility in MH96 in-host fitness during pre-cadaveric stages of insect infection.

Specific to entomopathogenic bacteria, both the flagellum and iron acquisition systems have already been implicated in intrahemocoelic infection of *G. mellonella* by triphasic nematode symbionts *P. luminescens* and closely-related *Xenorhabdus nematophila*. Flagellum-driven motility provided a selective fitness advantage during colonization of *G. mellonella* by *P. luminescens* (Easom & Clarke, 2008). Also, iron sensing by *X. nematophila* was shown to regulate the expression of flagellin gene *fliC* and hemolysin genes *xaxAB*, which significantly increased during late stage intrahemocoelic infection of *G. mellonella* and reached maximum levels following host death (Jubelin et al., 2011). Beyond entomopathogenic bacteria, iron homeostasis and hemoglobin utilization genes of *C. albicans* were found to be upregulated during later infection of *G. mellonella* and murine models compared to an earlier infection; however, in this study the early to late increase in iron-related genes was much greater in *G. mellonella* compared to the mouse model (Amorim-Vaz et al., 2015). Taken together, pathogens infecting *G. mellonella*, whether bacterial or fungal, appear to shift into an iron scavenging mode during late stage (i.e., pre-cadaveric) infection, indicating that pathogen sensing and response to *in vivo* iron fluxes during insect infection are likely an important component of in-host fitness or represent an interspecific competitive strategy.

Acknowledgements

Research associate, Mitchell Weston, performed sample collection and RNA extraction for one of the *in vitro* samples used in this analysis. We would like to acknowledge Trystan Whang and Elena Kim for coordinating library preparation, quality control and sequencing at Macrogen Oceana. Finally, AgResearch Ltd. bioinformaticians Dr. Aurelie Laugraud, Dr. Charles Hefer and Paul Maclean helped with job execution on high-performance computing clusters.

Author Contributions

Conceptualization, A.R.P., M.O., X.-X.Z., P.B.R., M.R.H.; Methodology, A.R.P., M.O., X.-X.Z., P.B.R., M.R.H.; Formal Analysis, A.R.P.; Writing – Original Draft Preparation, A.R.P.; Writing – Review & Editing, A.R.P., M.O., X.-X.Z., P.B.R., M.R.H.; Visualization, A.R.P.; Supervision, M.O., X.-X.Z., P.B.R., M.R.H.; Funding Acquisition, M.O., X.-X.Z., P.B.R., M.R.H. All authors have read and agreed to the pre-print version of the manuscript.

Funding

This research aligned to and was supported by the Next Generation Biopesticides Programme, funded by the New Zealand Ministry for Business, Innovation and Employment (Contract C10X1310). Canadian National Science and Engineering Council for Postgraduate Scholarship-Doctoral Program and Universities New

674 Zealand – Te Pōkai Tara via Massey University provided support in the form of
675 student scholarships.

676 Conflicts of Interest: The authors declare no conflicts of interest.

677

References

- Amorim-Vaz, S., Tran, V. D. T., Pradervand, S., Pagni, M., Coste, A. T., & Sanglard, D. (2015). RNA enrichment method for quantitative transcriptional analysis of pathogens *in vivo* applied to the fungus *Candida albicans*. *MBio*, 6(5), 1–16. doi: 10.1128/mBio.00942-15
- Aprianto, R., Slager, J., Holsappel, S., & Veening, J. W. (2016). Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biology*, 17(1), 1–16. doi: 10.1186/s13059-016-1054-5
- Arenas, J., Bossers-de Vries, R., Harders-Westerveen, J., Buys, H., Ruuls-van Stalle, L. M. F., Stockhofe-Zurwieden, N., ... de Greeff, A. (2019). In vivo transcriptomes of *Streptococcus suis* reveal genes required for niche-specific adaptation and pathogenesis. *Virulence*, 10(1), 334–351. doi: 10.1080/21505594.2019.1599669
- Avican, K., Fahlgren, A., Huss, M., Heroven, A. K., Beckstette, M., Dersch, P., & Fällman, M. (2015). Reprogramming of *Yersinia* from virulent to persistent mode revealed by complex *in vivo* RNA-seq analysis. *PLoS Pathogens*, 11(1), 1–28. doi: 10.1371/journal.ppat.1004600
- Baddal, B., Muzzi, A., Censini, S., Calogero, R. A., Torricelli, G., Guidotti, S., & Taddei, A. R. (2016). Dual RNA-seq of nontypeable *Haemophilus influenzae* and host cell transcriptomes reveals novel insights into host-pathogen cross talk. *MBio*, 7(2), 1–13. doi: 10.1128/mBio.00373-16
- Barth, H., Aktories, K., Popoff, M. R., & Stiles, B. G. (2004). Binary bacterial toxins:

700 Biochemistry, biology and applications of common *Clostridium* and *Bacillus*
701 Proteins. *Microbiology and Molecular Biology Reviews*, 68(3), 373–402. doi:
702 10.1128/MMBR.68.3.373

703 Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in
704 multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165–1188.
705 doi: 10.1214/aos/1013699998

706 Bent, Z. W., Branda, S. S., & Young, G. M. (2013). The *Yersinia enterocolitica* Ysa type
707 III secretion system is expressed during infections both *in vitro* and *in vivo*.
708 *MicrobiologyOpen*, 2(6), 962–975. doi: 10.1002/mbo3.136

709 Bertelli, C., Laird, M. R., Williams, K. P., Lau, B. Y., Hoad, G., Winsor, G. L., &
710 Brinkman, F. S. L. (2017). IslandViewer 4: Expanded prediction of genomic
711 islands for larger-scale datasets. *Nucleic Acids Research*, 45(W1), W30–W35.
712 doi: 10.1093/nar/gkx343

713 Bliska, J. B., Wang, X., Viboud, G. I., & Brodsky, I. E. (2013). Modulation of innate
714 immune responses by *Yersinia* type III secretion system translocators and
715 effectors. *Cellular Microbiology*, 15(10), 1622–1631. doi:
716 10.1111/cmi.12164.Modulation

717 Brugirard-Ricaud, K., Duchaud, E., Givaudan, A., Girard, P. A., Kunst, F., Boemare, N.,
718 ... Zumbihl, R. (2005). Site-specific antiphagocytic function of the *Photorhabdus*
719 *luminescens* type III secretion system during insect colonization. *Cellular*
720 *Microbiology*, 7(3), 363–371. doi: 10.1111/j.1462-5822.2004.00466.x

721 Bullard, J. H., Purdom, E., Hansen, K. D., & Dudoit, S. (2010). Evaluation of statistical

- 722 methods for normalization and differential expression in mRNA-Seq
- 723 experiments. *BMC Bioinformatics*, 11, 94. doi: 10.1186/1471-2105-11-94
- 724 Busby, J. N., Landsberg, M. J., Simpson, R. M., Jones, S. A., Hankamer, B., Hurst, M. R.
- 725 H., & Lott, J. S. (2012). Structural analysis of Chi1 chitinase from Yen-Tc: The
- 726 multisubunit insecticidal ABC toxin complex of *Yersinia entomophaga*. *Journal*
- 727 *of Molecular Biology*, 415(2), 359–371. doi: 10.1016/j.jmb.2011.11.018
- 728 Busby, J. N., Panjikar, S., Landsberg, M. J., Hurst, M. R. H., & Lott, J. S. (2013). The BC
- 729 component of ABC toxins is an RHS-repeat-containing protein encapsulation
- 730 device. *Nature*, 501(7468), 547–550. doi: 10.1038/nature12465
- 731 Bushnell, B. (2015). *BBMap short-read aligner, and other bioinformatics tools*.
- 732 <http://sourceforge.net/projects/bbmap/>. Retrieved from
- 733 <http://sourceforge.net/projects/bbmap/>
- 734 Casanova-torres, Á. M., & Goodrich-blair, H. (2013). Immune signaling and
- 735 antimicrobial peptide expression in Lepidoptera. *Insects*, 4, 320–338. doi:
- 736 10.3390/insects4030320
- 737 Champion, O. L., Wagley, S., & Titball, R. W. (2016). *Galleria mellonella* as a model
- 738 host for microbiological and toxin research. *Virulence*, 7(7), 840–845. doi:
- 739 10.1080/21505594.2016.1203486
- 740 Chauhan, N., Wrobel, A., Skurnik, M., & Leo, J. C. (2016). *Yersinia adhesins*: An
- 741 arsenal for infection. *Proteomics - Clinical Applications*, 10(9–10), 949–963. doi:
- 742 10.1002/prca.201600012
- 743 Chen, A., Wang, Y., Shao, Y., Zhou, Q., Chen, S., Wu, Y., & Chen, H. (2018). Genes

- 744 involved in *Beauveria bassiana* infection to *Galleria mellonella*. *Archives of*
745 *Microbiology*, 200(4), 541–552. doi: 10.1007/s00203-017-1456-0
- 746 Chen, L., Zheng, D., Liu, B., Yang, J., & Jin, Q. (2016). VFDB 2016: Hierarchical and
747 refined dataset for big data analysis - 10 years on. *Nucleic Acids Research*,
748 44(D1), D694–D697. doi: 10.1093/nar/gkv1239
- 749 Crawford, J. M., Kontnik, R., & Clardy, J. (2010). Regulating alternative lifestyles in
750 entomopathogenic bacteria. *Current Biology*, 20(1), 69–74. doi:
751 10.1016/j.cub.2009.10.059
- 752 Daborn, P. J., Waterfield, N., Silva, C. P., Au, C. P. Y., Sharma, S., & Ffrench-Constant,
753 R. H. (2002). A single *Photorhabdus* gene, makes caterpillars floppy (*mcf*), allows
754 *Escherichia coli* to persist within and kill insects. *Proceedings of the National*
755 *Academy of Sciences of the United States of America*, 99(16), 10742–10747. doi:
756 10.1073/pnas.102068099
- 757 Damron, F. H., Oglesby-Sherrouse, A. G., Wilks, A., & Barbier, M. (2016). Dual-seq
758 transcriptomics reveals the battle for iron during *Pseudomonas aeruginosa*
759 acute murine pneumonia. *Scientific Reports*, 6(November), 39172. doi:
760 10.1038/srep39172
- 761 Delor, I., Kaeckenbeeck, A., Wauters, G., & Cornelis, G. U. Y. R. (1990). Nucleotide
762 sequence of *yst*, the *Yersinia enterocolitica* gene encoding the heat-stable
763 enterotoxin, and prevalence of the gene among pathogenic and nonpathogenic
764 *Yersiniae*. *Infection and Immunity*, 58(9), 2983–2988.
- 765 Deng, Z. L., Sztajer, H., Jarek, M., Bhuju, S., & Wagner-Döbler, I. (2018). Worlds apart

766 - Transcriptome profiles of key oral microbes in the periodontal pocket
767 compared to single laboratory culture reflect synergistic interactions. *Frontiers*
768 *in Microbiology*, 9(FEB), 1–15. doi: 10.3389/fmicb.2018.00124

769 Desbois, A. P., & McMillan, S. (2015). Paving the way to acceptance of *Galleria*
770 *mellonella* as a new model insect. *Virulence*, 6(5), 410–411. doi:
771 10.1080/21505594.2015.1036218

772 Deuschle, E., Keller, B., Siegfried, A., Manncke, B., Spaeth, T., Köberle, M., ... Schütz,
773 M. (2015). Role of β 1 integrins and bacterial adhesins for Yop injection into
774 leukocytes in *Yersinia enterocolitica* systemic mouse infection. *International*
775 *Journal of Medical Microbiology*, 306(2), 77–88. doi:
776 10.1016/j.ijmm.2015.12.001

777 Dodt, M., Roehr, J., Ahmed, R., & Dieterich, C. (2012). FLEXBAR—Flexible Barcode
778 and Adapter Processing for Next-Generation Sequencing Platforms. *Biology*,
779 1(3), 895–905. doi: 10.3390/biology1030895

780 Easom, C. A., & Clarke, D. J. (2008). Motility is required for the competitive fitness of
781 entomopathogenic *Photorhabdus luminescens* during insect infection. *BMC*
782 *Microbiology*, 8, 1–11. doi: 10.1186/1471-2180-8-168

783 Fan, J., Ma, L., Zhao, C., Yan, J., Che, S., Zhou, Z., ... Hu, B. (2020). Transcriptome of
784 *Pectobacterium carotovorum* subsp. *carotovorum* PccS1 infected in calla plants
785 *in vivo* highlights a spatiotemporal expression pattern of genes related to
786 virulence, adaptation, and host response. *Molecular Plant Pathology*, 21(6),
787 871–891. doi: 10.1111/mpp.12936

788 Fang, F. C., Frawley, E. R., Tapscott, T., & Vázquez-Torres, A. (2016). Bacterial stress
789 responses during host infection. *Cell Host and Microbe*, 20(2), 133–143. doi:
790 10.1016/j.chom.2016.07.009

791 Ferguson, C. M., Barratt, B. I. P., Bell, N., Goldson, S. L., Jackson, M., Jackson, T. A., ...
792 Wilson, M. (2019). Quantifying the economic cost of invertebrate pests to New
793 Zealand's pastoral industry. *New Zealand Journal of Agricultural Research*,
794 62(3), 255–315. doi: 10.1080/00288233.2018.1478860

795 French-Constant, R. H., Dowling, A., & Waterfield, N. R. (2007). Insecticidal toxins
796 from *Photorhabdus* bacteria and their potential use in agriculture. *Toxicon*,
797 49(4), 436–451. doi: 10.1016/j.toxicon.2006.11.019

798 French-Constant, R., Waterfield, N., Daborn, P., Susan, J., Bennett, H., Au, C., ...
799 Clarke, D. (2003). *Photorhabdus*: Towards a functional genomic analysis of a
800 symbiont and pathogen. *FEMS Microbiology Reviews*, 26(5), 433–456.

801 Foutier, B., Troisfontaines, P., Müller, S., Oppendoes, F. R., & Cornelis, G. R. (2002).
802 Characterization of the ysa pathogenicity locus in the chromosome of *Yersinia*
803 *enterocolitica* and phylogeny analysis of type III secretion systems. *Journal of*
804 *Molecular Evolution*, 55, 37–51. doi: 10.1007/s00239-001-0089-7

805 Futschik, Matthias, E., & Carlisle, B. (2005). Noise-robust soft clustering of gene
806 expression time-course data. *Journal of Bioinformatics and Computational*
807 *Biology*, 3(4), 965–988. doi: 10.1142/S0219720005001375

808 Grant, T., Bennett-Wood, V., & Robins-Browne, R. M. (1998). Identification of
809 virulence-associated characteristics in clinical isolates of *Yersinia enterocolitica*

810 lacking classical virulence markers. *Infection and Immunity*, 66(3), 1113–1120.

811 Griesenauer, B., Tran, T. M., Fortney, K. R., Janowicz, D. M., Johnson, P., Gao, H., ...
812 Spinola, S. M. (2019). Determination of an interaction network between an
813 extracellular bacterial pathogen and the human host. *MBio*, 10(3), 1–15. doi:
814 10.1128/mBio.01193-19

815 Haueisen, J., Möller, M., Eschenbrenner, C. J., Grandaubert, J., Seybold, H., Adamiak,
816 H., & Stukenbrock, E. H. (2019). Highly flexible infection programs in a
817 specialized wheat pathogen. *Ecology and Evolution*, 9(1), 275–294. doi:
818 10.1002/ece3.4724

819 Heroen, A. K., & Dersch, P. (2014). Coregulation of host-adapted metabolism and
820 virulence by pathogenic *yersiniae*. *Frontiers in Cellular and Infection*
821 *Microbiology*, 4(October), 1–13. doi: 10.3389/fcimb.2014.00146

822 Heywood, W., Henderson, B., & Nair, S. P. (2005). Cytolethal distending toxin:
823 Creating a gap in the cell cycle. *Journal of Medical Microbiology*, 54(3), 207–216.
824 doi: 10.1099/jmm.0.45694-0

825 Hsieh, T. C., Ma, K. H., & Chao, A. (2016). iNEXT: an R package for rarefaction and
826 extrapolation of species diversity (Hill numbers). *Methods in Ecology and*
827 *Evolution*, 7(12), 1451–1456. doi: 10.1111/2041-210X.12613

828 Hurst, M. R. H., Beattie, A., Altermann, E., Moraga, R. M., Harper, L. A., Calder, J., &
829 Laugraud, A. (2016). The draft genome sequence of the *Yersinia entomophaga*
830 entomopathogenic type strain MH96T. *Toxins*, 8(143), 1–19. doi:
831 10.3390/toxins8050143

- 832 Hurst, M. R. H., Beattie, A. K., Jones, S. A., Hsu, P.-C., Calder, J., & van Koten, C.
833 (2015). Temperature-dependent *Galleria mellonella* mortality as a result of
834 *Yersinia entomophaga* infection. *Applied and Environmental Microbiology*,
835 81(18), 6404–6414. doi: 10.1128/AEM.00790-15
- 836 Hurst, M. R. H., Becher, S. A., Young, S. D., Nelson, T. L., & Glare, T. R. (2011). *Yersinia*
837 *entomophaga* sp. nov., isolated from the New Zealand grass grub *Costelytra*
838 *zealandica*. *International Journal of Systematic and Evolutionary Microbiology*,
839 61(Pt 4), 844–849. doi: 10.1099/ijs.0.024406-0
- 840 Hurst, M. R. H., Jones, S. A., Beattie, A., van Koten, C., Shelton, A. M., Collins, H. L., &
841 Brownbridge, M. (2019). Assessment of *Yersinia entomophaga* as a control
842 agent of the diamondback moth *Plutella xylostella*. *Journal of Invertebrate*
843 *Pathology*, 162(October 2018), 19–25. doi: 10.1016/j.jip.2019.02.002
- 844 Hurst, M. R. H., Jones, S. A., Binglin, T., Harper, L. A., Jackson, T. A., & Glare, T. R.
845 (2011). The main virulence determinant of *Yersinia entomophaga* MH96 is a
846 broad-host-range toxin complex active against insects. *Journal of Bacteriology*,
847 193(8), 1966–1980. doi: 10.1128/JB.01044-10
- 848 Hurst, M. R. H., van Koten, C., & Jackson, T. A. (2014). Pathology of *Yersinia*
849 *entomophaga* MH96 towards *Costelytra zealandica* (Coleoptera; Scarabaeidae)
850 larvae. *Journal of Invertebrate Pathology*, 115(1), 102–107. doi:
851 10.1016/j.jip.2013.11.004
- 852 Ibberson, C. B., & Whiteley, M. (2019). The *Staphylococcus aureus* transcriptome
853 during cystic fibrosis lung infection. *MBio*, 10(6), e02774-19.

854 Imori, P. F. M., Passaglia, J., Souza, R. A., Rocha, L. B., & Falcão, J. P. (2017).
855 Virulence-related genes, adhesion and invasion of some *Yersinia enterocolitica*-
856 like strains suggests its pathogenic potential. *Microbial Pathogenesis*, 104, 72–
857 77. doi: 10.1016/j.micpath.2017.01.008

858 Jiang, H., Vilcinskas, A., & Kanost, M. R. (2010). Immunity in lepidopteran insects. In
859 K. Söderhäll (Ed.), *Invertebrate Immunity* (pp. 1–24). Landes Bioscience and
860 Springer Science+Business Media.

861 Jubelin, G., Pagès, S., Lanois, A., Boyer, M., Gaudriault, S., Ferdy, J., & Givaudan, A.
862 (2011). Studies of the dynamic expression of the *Xenorhabdus* FlhAZ regulon
863 reveal atypical iron-dependent regulation of the flagellin and haemolysin genes
864 during. *Environmental Microbiology*, 13(5), 1271–1284. doi: 10.1111/j.1462-
865 2920.2011.02427.x

866 Keller, B., Mühlenkamp, M., Deuschle, E., Siegfried, A., Mössner, S., Schade, J., ...
867 Bohn, E. (2015). *Yersinia enterocolitica* exploits different pathways to
868 accomplish adhesion and toxin injection into host cells. *Cellular Microbiology*,
869 17(8), 1179–1204. doi: 10.1111/cmi.12429

870 Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., ... Marra,
871 M. A. (2009). Circos: An information aesthetic for comparative genomics.
872 *Genome Research*, 19(9), 1639–1645. doi: 10.1101/gr.092759.109.19

873 Kumar, L., & Futschik, M. (2007). Mfuzz: A software package for soft clustering of
874 microarray data. *Bioinformatics*, 2(1), 5–7. doi: 10.6026/97320630002005

875 Landsberg, M. J., Jones, S. A., Rothnagel, R., Busby, J. N., Marshall, S. D. G., Simpson,

876 R. M., ... Hurst, M. R. H. (2011). 3D structure of the *Yersinia entomophaga* toxin
877 complex and implications for insecticidal activity. *Proceedings of the National*
878 *Academy of Sciences*, 108(51), 20544–20549. doi: 10.1073/pnas.1111155108

879 Lang, Alexander, E., Schmidt, G., Schlosser, A., Hey, Timothy, D., Larrinua, I. M.,
880 Sheets, J. J., ... Aktories, K. (2010). *Photobacterium luminescens* toxins ADP-
881 ribosylate actin and RhoA to force actin clustering. *Science*, 327, 1139–1143.
882 doi: 10.1126/science.1184557

883 Lara-Tejero, M., & Galán, J. E. (2000). Cell cycle progression as a deoxyribonuclease
884 I-like protein. *Science*, 290, 354–357. doi: 10.1126/science.290.5490.354

885 Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). Voom: precision weights unlock
886 linear model analysis tools for RNA-seq read counts. *Genome Biology*, 15, R29.
887 doi: 10.1186/gb-2014-15-2-r29

888 Le Guern, A. S., Savin, C., Brémont, S., Payro, G., Bon, D., Carniel, E., & Pizarro-Cerdá,
889 J. (2018). First isolation of *Yersinia entomophaga* in human urinary tract. *New*
890 *Microbes and New Infections*, 26, 3–7. doi: 10.1016/j.nmni.2018.08.002

891 Li, H., Lin, L., Chong, L., Gu, S., Wen, S., Yu, G., ... Li, C. (2020). Time-resolved mRNA
892 and miRNA expression profiling reveals crucial coregulation of molecular
893 pathways involved in epithelial – pneumococcal interactions. *Immunology &*
894 *Cell Biology*, 1–17. doi: 10.1111/imcb.12371

895 Luo, G., Sun, Y., Huang, L., Su, Y., Zhao, L., Qin, Y., ... Yan, Q. (2020). Time-resolved
896 dual RNA-seq of tissue uncovers *Pseudomonas plecoglossicida* key virulence
897 genes in host-pathogen interaction with *Epinephelus coioides*. *Environmental*

- 898 *Microbiology*, 22(2), 677–693. doi: 10.1111/1462-2920.14884
- 899 Maldonado-arocho, F. J., Green, C., Fisher, M. L., Paczosa, M. K., & Mecsas, J. (2013).
900 Adhesins and host serum factors drive yop translocation by *Yersinia* into
901 professional phagocytes during animal infection. *PLoS Pathogens*, 9(6),
902 e1003415. doi: 10.1371/journal.ppat.1003415
- 903 Marshall, S. D. G., Hares, M. C., Jones, S. A., Harper, L. A., Vernon, J. R., Harland, D.
904 P., ... Hurst, M. R. H. (2012). Histopathological effects of the Yen-Tc toxin
905 complex from *Yersinia entomophaga* MH96 (Enterobacteriaceae) on the
906 *Costelytra zealandica* (Coleoptera: Scarabaeidae) larval midgut. *Applied and*
907 *Environmental Microbiology*, 78(14), 4835–4847. doi: 10.1128/AEM.00431-12
- 908 Matsumoto, H., & Young, G. M. (2006). Proteomic and functional analysis of the
909 suite of Ysp proteins exported by the Ysa type III secretion system of *Yersinia*
910 *enterocolitica* Biovar 1B. *Molecular Microbiology*, 59(2), 689–706. doi:
911 10.1111/j.1365-2958.2005.04973.x
- 912 Mühlenkamp, M., Oberhettinger, P., Leo, J. C., Linke, D., & Schütz, M. S. (2015).
913 *Yersinia* adhesin A (YadA) – Beauty & beast. *International Journal of Medical*
914 *Microbiology*, 305(2), 252–258. doi: 10.1016/j.ijmm.2014.12.008
- 915 Nielsen-LeRoux, C., Gaudriault, S., Ramarao, N., Lereclus, D., & Givaudan, A. (2012).
916 How the insect pathogen bacteria *Bacillus thuringiensis* and
917 *Xenorhabdus/Photorhabdus* occupy their hosts. *Current Opinion in*
918 *Microbiology*, 15(3), 220–231. doi: 10.1016/j.mib.2012.04.006
- 919 Nobori, T., Velásquez, A. C., Wu, J., Kvitko, B. H., Kremer, J. M., Wang, Y., ... Tsuda, K.

920 (2018). Transcriptome landscape of a bacterial pathogen under plant immunity.
 921 *Proceedings of the National Academy of Sciences*, 115(13), E3055–E3064. doi:
 922 10.1073/pnas.1800529115

923 Nuss, A. M., Beckstette, M., Pimenova, M., Schmühl, C., Opitz, W., Pisano, F., ...
 924 Dersch, P. (2017). Tissue dual RNA-seq allows fast discovery of infection-specific
 925 functions and riboregulators shaping host–pathogen transcriptomes.
 926 *Proceedings of the National Academy of Sciences*, 114(5), E791–E800. doi:
 927 10.1073/pnas.1613405114

928 Pha, K., & Navarro, L. (2016). *Yersinia* type III effectors perturb host innate immune
 929 responses. *World Journal of Biological Chemistry*, 7(1), 1–13. doi:
 930 10.4331/wjbc.v7.i1.1

931 Piper, S. J., Brillault, L., Rothnagel, R., Croll, T. I., Box, J. K., Chassagnon, I., ...
 932 Landsberg, M. J. (2019). Cryo-EM structures of the pore-forming A subunit from
 933 the *Yersinia entomophaga* ABC toxin. *Nature Communications*, 10(1), 1–12. doi:
 934 10.1038/s41467-019-09890-8

935 Plano, G. V., & Schesser, K. (2013). The *Yersinia pestis* type III secretion system:
 936 expression , assembly and role in the evasion of host defenses. *Immunological*
 937 *Research*, 57, 237–245. doi: 10.1007/s12026-013-8454-3

938 Ramamurthy, T., Yoshino, K. I., Huang, X., Balakrish Nair, G., Carniel, E., Maruyama,
 939 T., ... Takeda, T. (1997). The novel heat-stable enterotoxin subtype gene (ystB)
 940 of *Yersinia enterocolitica*: Nucleotide sequence and distribution of the yst
 941 genes. *Microbial Pathogenesis*, 23(4), 189–200. doi: 10.1006/mpat.1997.0146

- 942 Reniere, M. L. (2018). Reduce, induce, thrive: Bacterial redox sensing during
943 pathogenesis. *Journal of Bacteriology*, 200(17), 1–12. doi: 10.1128/JB.00128-18.
- 944 Risso, D., Ngai, J., Speed, T. P., & Dudoit, S. (2014). Normalization of RNA-seq data
945 using factor analysis of control genes or samples. *Nature Biotechnology*, 32(9),
946 896–902. doi: 10.1038/nbt.2931
- 947 Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015).
948 Limma powers differential expression analyses for RNA-sequencing and
949 microarray studies. *Nucleic Acids Research*, 43(7), e47. doi: 10.1093/nar/gkv007
- 950 Singh, I., & Viridi, J. S. (2004). Production of *Yersinia* stable toxin (YST) and
951 distribution of yst genes in biotype 1A strains of *Yersinia enterocolitica*. *Journal*
952 *of Medical Microbiology*, 53(11), 1065–1068. doi: 10.1099/jmm.0.45527-0
- 953 Singkum, P., Suwanmanee, S., Pumeesat, P., & Luplertlop, N. (2019). A powerful *in*
954 *vivo* alternative model in scientific research: *Galleria mellonella*. *Acta*
955 *Microbiologica et Immunologica Hungarica*, 66(1), 31–55. doi:
956 10.1556/030.66.2019.001
- 957 Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing
958 differential expression in microarray experiments. *Statistical Applications in*
959 *Genetics and Molecular Biology*, 3(1), 1–25. doi: 10.2202/1544-6115.1027
- 960 Springer, K., Sanger, P.-A., Moritz, C., Felsl, A., Rattei, T., & Fuchs, T. M. (2018).
961 Insecticidal toxicity of *Yersinia frederiksenii* involves the novel enterotoxin YacT.
962 *Frontiers in Cellular and Infection Microbiology*, 8(November), 1–13. doi:
963 10.3389/fcimb.2018.00392

- 964 Sun, Y., Zhuang, Z., Wang, X., Huang, H., Fu, Q., & Yan, Q. (2019). Dual RNA-seq
965 reveals the effect of the flgM gene of *Pseudomonas plecoglossicida* on the
966 immune response of *Epinephelus coioides*. *Fish and Shellfish Immunology*,
967 87(October 2018), 515–523. doi: 10.1016/j.fsi.2019.01.041
- 968 Tang, R., Zhao, L., Xu, X., Huang, L., Qin, Y., Su, Y., & Yan, Q. (2019). Dual RNA-Seq
969 uncovers the function of an ABC transporter gene in the host-pathogen
970 interaction between *Epinephelus coioides* and *Pseudomonas plecoglossicida*.
971 *Fish and Shellfish Immunology*, 92(May), 45–53. doi: 10.1016/j.fsi.2019.05.046
- 972 Tang, Y., Xin, G., Zhao, L., Huang, L., Qin, Y., Su, Y., ... Lin, N. (2020). Novel insights
973 into host-pathogen interactions of large yellow croakers (*Larimichthys crocea*)
974 and pathogenic bacterium *Pseudomonas plecoglossicida* using time- resolved
975 dual RNA-seq of infected spleens. *Zoological Research*, 41(3), 314–327. doi:
976 10.24272/j.issn.2095-8137.2020.035
- 977 Thomson, N. R., Howard, S., Wren, B. W., Holden, M. T. G., Crossman, L., Challis, G.
978 L., ... Prentice, M. B. (2006). The complete genome sequence and comparative
979 genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081.
980 *PLoS Genetics*, 2(12), 2039–2051. doi: 10.1371/journal.pgen.0020206
- 981 Tobias, N. J., Heinrich, A. K., Eresmann, H., Wright, P. R., Neubacher, N., Backofen, R.,
982 & Bode, H. B. (2017). *Photorhabdus*-nematode symbiosis is dependent on hfq-
983 mediated regulation of secondary metabolites. *Environmental Microbiology*,
984 19(1), 119–129. doi: 10.1111/1462-2920.13502
- 985 Tran, S. L., Guillemet, E., Lereclus, D., & Ramarao, N. (2013). Iron regulates *Bacillus*

986 thuringiensis haemolysin hlyII gene expression during insect infection. *J*
987 *Invertebr Pathol*, 113(3), 205–208. doi: 10.1016/j.jip.2013.04.001

988 Tsai, C. J. Y., Loh, J. M. S., & Proft, T. (2016). *Galleria mellonella* infection models for
989 the study of bacterial diseases and for antimicrobial drug testing. *Virulence*,
990 7(3), 214–229. doi: 10.1080/21505594.2015.1135289

991 Valenzuela-Miranda, D., & Gallardo-Escárate, C. (2018). Dual RNA-Seq uncovers
992 metabolic amino acids dependency of the intracellular bacterium *Piscirickettsia*
993 *salmonis* infecting Atlantic salmon. *Frontiers in Microbiology*, 9, 1–11. doi:
994 10.3389/fmicb.2018.02877

995 Venecia, K., & Young, G. M. (2005). Environmental regulation and virulence
996 attributes of the Ysa type III secretion system of *Yersinia enterocolitica* biovar
997 1B. *Infection and Immunity*, 73(9), 5961–5977. doi: 10.1128/IAI.73.9.5961-
998 5977.2005

999 Vogel, H., Altincicek, B., Glöckner, G., & Vilcinskis, A. (2011). A comprehensive
1000 transcriptome and immune-gene repertoire of the lepidopteran model host
1001 *Galleria mellonella*. *BMC Genomics*, 12(1), 308. doi: 10.1186/1471-2164-12-308

1002 Walker, K. A., Maltez, V. I., Hall, J. D., Vitko, N. P., & Miller, V. L. (2013). A phenotype
1003 at last: Essential role for the *Yersinia enterocolitica* Ysa: Type III secretion
1004 system in a *Drosophila melanogaster* S2 cell model. *Infection and Immunity*,
1005 81(7), 2478–2487. doi: 10.1128/IAI.01454-12

1006 Walker, K. A., & Miller, V. L. (2009). Synchronous gene expression of the *Yersinia*
1007 *enterocolitica* Ysa type III secretion system and its effectors. *Journal of*

1008 *Bacteriology*, 191(6), 1816–1826. doi: 10.1128/JB.01402-08

1009 Waterfield, N., Kamita, S. G., Hammock, B. D., & Ffrench-Constant, R. (2005). The
 1010 *Photorhabdus* Pir toxins are similar to a developmentally regulated insect
 1011 protein but show no juvenile hormone esterase activity. *FEMS Microbiology*
 1012 *Letters*, 245(1), 47–52. doi: 10.1016/j.femsle.2005.02.018

1013 Westermann, A. J., Förstner, K. U., Amman, F., Barquist, L., Chao, Y., Schulte, L. N., ...
 1014 Vogel, J. (2016). Dual RNA-seq unveils noncoding RNA functions in host-
 1015 pathogen interactions. *Nature*, 529(7587), 496–501. doi: 10.1038/nature16547

1016 Westermann, A. J., Venturini, E., Sellin, M. E., Förstner, K. U., Hardt, W. D., & Vogel, J.
 1017 (2019). The major RNA-binding protein ProQ impacts virulence gene expression
 1018 in *Salmonella enterica* serovar Typhimurium. *MBio*, 10(1), 1–21. doi:
 1019 10.1128/mBio.02504-18

1020 Wilkinson, P., Waterfield, N. R., Crossman, L., Corton, C., Sanchez-contreras, M.,
 1021 Vlisidou, I., ... Ffrench-Constant, R. H. (2009). Comparative genomics of the
 1022 emerging human pathogen *Photorhabdus asymbiotica* with the insect pathogen
 1023 *Photorhabdus luminescens*. *BMC Genomics*, 10(302), 302. doi: 10.1186/1471-
 1024 2164-10-302

1025 Wojda, I. (2017). Immunity of the greater wax moth *Galleria mellonella*. *Insect*
 1026 *Science*, 24(3), 342–357. doi: 10.1111/1744-7917.12325

1027 Wu, G., & Yi, Y. (2016). Haemocoel injection of PirA₁B₁ to *Galleria mellonella* larvae
 1028 leads to disruption of the haemocyte immune functions. *Nature Publishing*
 1029 *Group*, (September), 1–7. doi: 10.1038/srep34996

1030 Yan, Y., Su, S., Meng, X., Ji, X., Qu, Y., Liu, Z., ... Han, Y. (2013). Determination of sRNA
 1031 expressions by RNA-seq in *Yersinia pestis* grown in vitro and during infection.
 1032 *PloS One*, 8(9), 1–13. doi: 10.1371/journal.pone.0074495
 1033

1034 Supplemental

1035

1036 Table S1: Mean number of *Yersinia entomophaga* MH96 cells from inoculum and
 1037 hours-post infection (HPI) of *Galleria mellonella* by intrahemocoelic injection
 1038 corresponding to lag, exponential and stationary in vivo growth phases at 25 °C.
 1039 (n = 3 replicate plates). CFU = colony forming units, SEM = standard error of the
 1040 mean and HPI = hours post infection.

In vivo infection stage	Mean inoculum cells (CFU ± SEM)	HPI
Early infection	$2 \pm 0.1 \times 10^7$	1 ¼ - 1 ½
Middle infection	157 ± 22	18 -19
Lag-infection	693 ± 162	26 – 30

1041

Table S2: Adapter and barcode sequences used to trim all *in vivo* samples and *in vitro* samples 'F22-1' and 'F22-2' (exponential growth phase).

```
>Illumina_D5_adapter_i5a
ATGATACGGCGACACCGAGATCTACAC
>Illumina_D5_adapter_i5b
ACACTCTTTCCCTACACGACGCTCTTCCGATCT
>Illumina_D5_adapter_i7a
GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
>Illumina_D5_adapter_i7b
ATCTCGTATGCCGTCTTCTGCTTG
>Truseq_Universal_Adapter
AATGATACGGCGACACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
CT
>Truseq_Adapter_1a
GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
>Truseq_Adapter_1b
ATCTCGTATGCCGTCTTCTGCTTG
>Illumina_D5_adapter_i5a Reversed:
GTGTAGATCTCGGTGGTCGCCGTATCAT
>Illumina_D5_adapter_i5b Reversed:
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
>Illumina_D5_adapter_i7a Reversed:
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
>Illumina_D5_adapter_i7b Reversed:
CAAGCAGAAGACGGCATACGAGAT
>Truseq_Universal_Adapter Reversed:
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCA
TT
>Truseq_Adapter_1a Reversed:
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
>Truseq_Adapter_1b Reversed:
CAAGCAGAAGACGGCATACGAGAT
>TruSeq_Universal_Adapter
AATGATACGGCGACACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
CT
>O15-1_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTT
CTGCTTG
>O15-2_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTT
CTGCTTG
>O15-3_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTT
CTGCTTG
>O19-1_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTT
CTGCTTG
>O18-3_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACACAGTGATCTCGTATGCCGTCTT
CTGCTTG
>O18-4_with_barcode
```

```

GATCGGAAGAGCACACGTCTGAACTCCAGTCACGCCAATATCTCGTATGCCGTCTT
CTGCTTG
>O18-7_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCAGATCATCTCGTATGCCGTCTT
CTGCTTG
>O18-8_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACACTTGAATCTCGTATGCCGTCTT
CTGCTTG
>O20-1_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTT
CTGCTTG
>O20-3B_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTATCTCGTATGCCGTCTT
CTGCTTG
>O20-4_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTT
CTGCTTG
>D4-2_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGGCTACATCTCGTATGCCGTCTT
CTGCTTG
>F22-1_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACGATCAGATCTCGTATGCCGTCTT
CTGCTTG
>F22-2_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTAGCTTATCTCGTATGCCGTCTT
CTGCTTG
>J24-5_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTATGCCGTCTT
CTGCTTG
>J24-6_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACCGATGTATCTCGTATGCCGTCTT
CTGCTTG
>J24-7_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTTAGGCATCTCGTATGCCGTCTT
CTGCTTG
>J24-8_with_barcode
GATCGGAAGAGCACACGTCTGAACTCCAGTCACTGACCAATCTCGTATGCCGTCTT
CTGCTTG
>TruSeq_Universal_Adapter_RC
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCA
TT
>O15-1_with_barcode_RC
CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O15-2_with_barcode_RC
CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O15-3_with_barcode_RC
CAAGCAGAAGACGGCATACGAGATGCCTAAGTGAAGTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O19-1_with_barcode_RC

```

```

CAAGCAGAAGACGGCATAACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O18-3_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O18-4_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O18-7_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O18-8_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O20-1_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O20-3B_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>O20-4_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>D4-2_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>F22-1_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>F22-2_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>J24-5_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>J24-6_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>J24-7_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC
>J24-8_with_barcode_RC
CAAGCAGAAGACGGCATAACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCT
TCCGATC

```


Table S3: Adapter and barcode sequences used to trim all *in vitro* samples, except 'F22-1' and 'F22-2' (see Table S1).

```
>E1
TCGTCAAGTTACTCGC
>E3
CGCCGGTCTGATCTCC
>H1
TTTGACACCCTAGCGT
>H3
ACACTGTGGGCTCGGA
>MF
AGCTCACGACGCTCAG
>MG
GAGAGGGACGAGCGAA
>E1_Reversed
GCGAGTAACTTGACGA
>E3_Reversed
GGAGATCAGACCGGCG
>H1_Reversed
ACGCTAGGGTGTCAA
>H3_Reversed
TCCGAGCCCACAGTGT
>MF_Reversed
CTGAGCGTCGTGAGCT
>MG_Reversed
TTCGCTCGTCCCTCTC
>Illumina_ReversedD5_Reversedadapter_Reversed15a
ATGATACGGCGACCACCGAGATCTACAC
>Illumina_ReversedD5_Reversedadapter_Reversed15b
ACACTCTTTCCCTACACGACGCTCTTCCGATCT
>Illumina_ReversedD5_Reversedadapter_Reversed17a
GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
>Illumina_ReversedD5_Reversedadapter_Reversed17b
ATCTCGTATGCCGTCTTCTGCTTG
>Truseq_ReversedUniversal_ReversedAdapter
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
CT
>Truseq_ReversedAdapter_Reversed1a
GATCGGAAGAGCACACGTCTGAACTCCAGTCAC
>Truseq_ReversedAdapter_Reversed1b
ATCTCGTATGCCGTCTTCTGCTTG
>Illumina_ReversedD5_Reversedadapter_Reversed15a_Reversed
GTGTAGATCTCGGTGGTCGCCGTATCAT
>Illumina_ReversedD5_Reversedadapter_Reversed15b_Reversed
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
>Illumina_ReversedD5_Reversedadapter_Reversed17a_Reversed
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
```

```
>Illumina_ReversedD5_Reversedadapter_Reversed17b_Reverse
d
CAAGCAGAAGACGGCATAACGAGAT
>Truseq_ReversedUniversal_ReversedAdapter_Reversed
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGATCTCGGTGGTCGCCGTATCA
TT
>Truseq_ReversedAdapter_Reversed1a_Reversed
GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC
>Truseq_ReversedAdapter_Reversed1b_Reversed
CAAGCAGAAGACGGCATAACGAGAT
>TruSeq_ReversedUniversal_ReversedAdapter
AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGAT
CT
```

Table S4: Categories of virulence factors considered in this study.

Category	Definition	Examples
<i>Primary – secreted and directly act against host cells/tissues</i>		
Toxin	Proteins that are secreted and directly target host cells/tissues, giving rise to pathogenesis.	Insecticidal toxin complex, vegetative insecticidal protein or Cry toxin.
Effector	Proteins that are translocated directly into host cells by type 3 or 6 secretion systems.	<i>Yersinia</i> spp. Yop or <i>Salmonella</i> spp. Spa type 3 secretion system and hemolysin co-regulated protein (Hcp) or polymorphic rearrangement hot-spot (Rhs) toxins.
<i>Secondary – all other non-primary factors contributing to pathogenesis (i.e., gene abrogation results in attenuated virulence).</i>		
Adherence	External structures used to attach to host cells or tissues.	Fimbriae or filamentous hemagglutinins.
Defense response	Factors used to detoxify against antimicrobial compounds or host-induced stress.	Superoxide dismutase, efflux pumps, iron-sulfur cluster.
Horizontal gene transfer	Factors that promote lateral gene acquisitions or mobilization.	Conjugative pili or integrases.
Iron acquisition	Systems used by pathogenic bacteria to acquire iron in the host environment.	Siderophores, hemolysins and iron-specific transporters.
Metabolic adaptation	Enzymes and transporters involved in bioconversion of host-derived nutrients.	Amino acid transporters or chitinases.
Mobility	Factors related to motility.	Flagella or chemotaxis.
Outer membrane	Factors related to modification of outer membrane/capsule often used to evade host-immune response.	Lipopolysaccharide or outer membrane proteins.
Regulation	Regulatory factors controlling pathogenic response to host.	Two-component sensors, quorum sensing or diguanylate cyclase.
Secretion system	Multi-component systems used for translocation and secretion of toxins and effectors by bacteria.	Type 2, 3, and 6 secretion systems.
Unclassified	Factors that could not be classified specific categories listed above.	Non-ribosomal peptide synthetase or tellurium resistance genes.

Table S5. Differentially expressed genes identified from *Yersinia entomophaga* MH96 for time-resolved *vivo* vs. *in vitro* comparison from upper quartile normalized count data using *Limma* with Benjamini and Yekutieli adjusted p-value cutoff of 0.05.

[See “Supplemental Table S5 - differentially expressed genes identified from *Yersinia entomophaga* time-resolved infection-series.xls” in the supplemental zip folder].

Table S6. In-host putative virulence factors identified in *Yersinia entomophaga* MH96 using fuzzy clustering and virulence annotations from VFDB.

[See “Supplemental Table S6 - In-host putative virulence factors identified in *Yersinia entomophaga* using FCM and VFDB.xls” in the supplemental zip folder].

Figure S1: *In vivo* growth curve of *Yersinia entomophaga* MH96 in the hemolymph of *Galleria mellonella* at 25 °C. Larvae were injected with approximately 345 cells in 10 μ l PBS (time-point 0) and CFU/larva were enumerated from 3 larvae per timepoint at 1, 2, 4, 6, 8, 12, 24, 32 and 48 hours post-infection. Points represent mean density and vertical bars represent standard error.

Figure S2. Mean proportion of paired-end reads attributed to host and pathogen based on average percent alignment to *Galleria mellonella* rRNA and EST sequence or *Yersinia entomophaga* MH96 genome or rRNA sequence (*in vivo*: n = 4, *in vitro*: n = 2). The growth phase and estimated cell number (CFU/g and CFU/ml for *in vivo* and *in vitro*, respectively) are provided in Table 1). Only percentages of 4 % and greater are shown.

Table S7: Mean percentage of trimmed paired-end reads aligning to *Yersinia entomophaga* MH96 reference genome, including protein-coding, tRNA and non-coding regions (*in vivo*: n = 4, *in vitro*: n = 2).

Infection-stage / growth-phase	Total align-ments	Protein coding region	Anti-sense protein coding region	tRNA	Non-annotated regions
<i>In vivo</i>					
Early infection	9 %	77 %	2 %	1 %	22 %
Middle infection	58 %	74 %	2 %	1 %	24 %
Late infection	59 %	75 %	1 %	0 %	24 %
<i>In vitro</i>					
Lag	93 %	84 %	2 %	1 %	14 %
Exponential	84 %	83 %	1 %	1 %	15 %
Stationary	93 %	85 %	1 %	0 %	14 %

Figure S3: Predicted transcript detection based on the number of aligned paired-end reads for *in vivo* early infection libraries by rarefaction analysis. 1, 2, 3 and 4 represent individual libraries.

Figure S4. Relative log expression of *in vivo* and *in vitro* libraries represented as median boxplots. A) unnormalized and B) upper-quartile normalized. Individual collection batches are noted where applicable.

Figure S5. Principle component analysis of *in vivo* and *in vitro* library upper quartile normalized count data. Outlier late infection *in vivo* sample is indicated and was removed from subsequent analysis.

Figure S6. Differentially expressed transcripts from *in vivo* or *in vitro* libraries during early, middle and late infection and lag, exponential and stationary growth phases, respectively. Significance was multiple-test corrected (Benjamini & Yekutieli, 2001) with an adjusted p-value threshold of 0.05. See supplemental Table S5 for more details on the contrasts represented in the Venn groups.

Table S8: Location of predicted genomic islands in the genome of *Y. entomophaga* MH96 based on Island Viewer 4 analysis.

Island	Start	Stop
Is.1	4,245,874	43,205
Is.2	63,587	68,974
Is.3	190,789	198,981
Is.4	642,845	691,291
Is.5	749,459	763,698
Is.6	840,932	892,566
Is.7	911,590	931,192
Is.8	1,042,716	1,047,262
Is.9	1,161,921	1,166,559
Is.10	1,323,793	1,330,066
Is.11	1,572,653	1,602,200
Is.12	1,693,457	1,718,980
Is.13	2,171,815	2,180,723
Is.14	2,633,449	2,651,765
Is.15	3,084,579	3,102,232
Is.16	3,204,015	3,214,229
Is.17	3,263,328	3,292,504
Is.18	3,326,808	3,340,556
Is.19	3,516,027	3,562,485
Is.20	3,655,950	3,662,717
Is.21	3,729,760	3,736,838

Figure S7: Host-specific transcripts from *Yersinia entomophaga* MH96 *in vivo* RNA-seq identified using c-means fuzzy clustering algorithm ($k = 20$, $m = 1.5$). $\log_2$ counts-per-million (CPM) fold-change between *in vivo* and *in vitro* samples for each cell density. Black line = median $\log_2$ CPM fold-change. CFU = colony forming unit.