



Published in final edited form as:

*Mol Biosyst.* 2013 January 27; 9(1): 44–54. doi:10.1039/c2mb25287b.

## A multi-omic systems approach to elucidating *Yersinia* virulence mechanisms

Charles Ansong<sup>1</sup>, Alexandra C. Schrimpe-Rutledge<sup>1</sup>, Hugh Mitchell<sup>2</sup>, Sadhana Chauhan<sup>3</sup>, Marcus B. Jones<sup>4</sup>, Young-Mo Kim<sup>1</sup>, Kathleen McAteer<sup>5</sup>, Brooke L. Deatherage Kaiser<sup>1</sup>, Jennifer L. Dubois<sup>7</sup>, Heather M. Brewer<sup>6</sup>, Bryan C. Frank<sup>4</sup>, Jason E. McDermott<sup>2</sup>, Thomas O. Metz<sup>1</sup>, Scott N. Peterson<sup>4</sup>, Richard D. Smith<sup>1</sup>, Vladimir L. Motin<sup>3</sup>, and Joshua N. Adkins<sup>1</sup>

<sup>1</sup>Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA

<sup>2</sup>Computational Sciences & Mathematics Division, Pacific Northwest National Laboratory, Richland, WA

<sup>3</sup>Departments of Microbiology and Immunology, University of Texas Medical Branch, Galveston, TX

<sup>4</sup>Infectious Disease Group, J. Craig Venter Institute, Rockville, MD

<sup>5</sup>Biology Program, Washington State University Tri-Cities, Richland, WA

<sup>6</sup>Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

<sup>7</sup>Biosciences Division, Stanford Research Institute, International, Menlo Park, CA

### Abstract

The underlying mechanisms that lead to dramatic differences between closely related pathogens are not always readily apparent. For example, the genomes of *Yersinia pestis* (YP) the causative agent of plague with a high mortality rate and *Yersinia pseudotuberculosis* (YPT) an enteric pathogen with a modest mortality rate are highly similar with some species specific differences; however the molecular causes of their distinct clinical outcomes remain poorly understood. In this study, a temporal multi-omic analysis of YP and YPT at physiologically relevant temperatures was performed to gain insights into how an acute and highly lethal bacterial pathogen, YP, differs from its less virulent progenitor, YPT. This analysis revealed higher gene and protein expression levels of conserved major virulence factors in YP relative to YPT, including the Yop virulon and the pH6 antigen. This suggests that adaptation in the regulatory architecture, in addition to the presence of unique genetic material, may contribute to the increased pathogenicity of YP relative to YPT. Additionally, global transcriptome and proteome responses of YP and YPT revealed conserved post-transcriptional control of metabolism and the translational machinery including the modulation of glutamate levels in *Yersinia*. Finally, the omics data was coupled with a computational network analysis, allowing an efficient prediction of novel *Yersinia* virulence factors based on gene and protein expression patterns.

---

Address correspondence to: Biological Sciences Division, Pacific Northwest National Laboratory, P. O. Box 999, Richland, WA 99352. Tel: 509-371-6583; Fax: 509-371-6555; Joshua.Adkins@pnl.gov.

#### Author's contributions

SC, VLM, MJB, BCF, SNP, ASR, CA contributed to transcriptomics data generation and analysis. SC, VLM, HMB, BDK, JLD, ASR, CA, JNA contributed to proteomics data generation and analysis. SC, VLM, YMK, KM, TOM, CA contributed to metabolomics data generation and analysis. HM, JEM, CA contributed to computational network analysis. ASR, HM, YMK, KM, BD, JEM, TOM, SNP, RDS participated in manuscript preparation. CA, VLM, JNA wrote the manuscript. JNA and RDS contributed partial funding. All authors read and approved the final manuscript.

## Introduction

*Yersinia pestis* (YP) is a Gram-negative bacterium and the causative agent of plague, an acute and lethal disease responsible for at least three human pandemics that resulted in an estimated 200 million deaths<sup>1</sup>. Several lines of evidence suggest that YP has evolved from the gastrointestinal pathogen *Y. pseudotuberculosis* (YPT) within the last 20,000 years<sup>2</sup>. Genomic analyses show YP and YPT to be genetically similar (~97% identity at the nucleotide level<sup>3</sup>), yet despite their close genetic relationship, the bacteria exhibit markedly different pathogenicities and modes of transmission<sup>4</sup>. YPT causes non-fatal gastrointestinal disease and is transmitted via the fecal oral route, while YP is the causative agent of typically fatal plague and is transmitted via flea bite. Based both on the ability to ferment glycerol and to reduce nitrate, YP strains have traditionally been assigned to one of three biovars: *antiqua*, *medievalis*, and *orientalis*<sup>1</sup>. Recently, the new biovar *microtus* has been identified on the basis of unique pathogenic, biochemical, and molecular features<sup>5</sup>. In laboratory studies, *microtus* strains (also known as Pestoides) are lethal to *microtus* species (voles), mice, and some other small rodents, however they are avirulent in humans and larger mammals. Whereas *antiqua*, *medievalis*, and *orientalis* biovars cause disease in humans (i.e. epidemic strains), there is no evidence that human plague can be caused by Pestoides (i.e. non-epidemic) strains<sup>6</sup>.

The availability of genome sequences for several *Yersinia* strains, including YPT and both epidemic and non-epidemic YP variants, has provided an opportunity to explore mechanisms responsible for the differences in pathogenicity. Comparative genomic analyses revealed all human pathogenic *Yersinia* strains, including YP and YPT share almost identical ~70-kb virulence plasmids that are essential for virulence<sup>7</sup>. This plasmid (pCD1 in YP) encodes two major types of virulence factors: (i) the *Yersinia* outer proteins (Yops) and V antigen and (ii) the type three secretion system (T3SS) apparatus which is required to translocate Yop effector proteins to the host cytoplasm to modulate host cell function and promote disease progression<sup>8</sup>. Additionally, comparison of the genomes of YP and its progenitor YPT reveal a modest number of species-specific chromosomal genes as well as the presence of two plasmids (pMT1 and pCP1) specific to YP that are thought to contribute to YP pathogenesis<sup>1, 3, 9-14</sup>. The pMT1 plasmid harbors genes coding for the capsular antigen F1 and murine toxin, while the pPCP1 plasmid encodes the plasminogen activator. Importantly, these species-specific attributes cannot fully account for the marked difference in pathogenicity between YP and YPT<sup>9, 15-21</sup>. One hypothesis is that the differential expression of genes common to both organisms, in addition to overt genetic differences, is an important contributing factor to the different pathogenicities and clinical outcomes of YP and YPT.

In this study we have performed a systems level multi-omic analysis of YP CO92 (YPCO) and YPT PB1/+ (YPTS) to gain an understanding as to how an acute and highly lethal bacterial pathogen, such as YP differs phenotypically from its less virulent progenitor YPT. We also compare YP CO92 (YPCO) to the non-epidemic YP strain Pestoides F (YPPF) to provide insights to the mechanism(s) underlying the virulence-restricted phenotype of non-epidemic YP strains compared to epidemic YP strains. The parallel sample-matched transcriptomics and proteomics analysis of multiple pathogenic *Yersinia* strains in a single study allows for the prediction of genes putatively involved in core pathogenic processes important for virulence mechanisms of *Yersinia* species. In the present work, cells were grown in a chemically defined medium (pH 7.2) at physiologically relevant temperatures (representative of flea vector and mammalian host environments) and sampled through an 8 hour time course. Transcription was analyzed using a multi-genome microarray and protein and metabolite levels were analyzed by mass spectrometric methods. This experimental design offered the advantage of revealing both transcriptional and post-transcriptional

responses to a temperature shift simulating the host-*Yersinia* interaction through a time course that simulates the progression of a mammalian infection. The data suggests that adaptation in the regulatory architecture, in addition to the presence of unique genetic material, may contribute to the increased pathogenicity of YP relative to YPT; and also revealed conserved post-transcriptional control of metabolism and the translational machinery in *Yersinia*.

## Experimental

### Reagents

The following reagents were used in sample preparation: Nanopure or Milli-Q quality water (~18 megohm-cm or better); ammonium bicarbonate (NH<sub>4</sub>HCO<sub>3</sub>); bicinechonic acid (BCA) or coomassie protein assay reagents (Pierce, Rockford, IL); urea; thiourea; dithiothreitol (DTT); 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate(CHAPS); calcium chloride; sequencing-grade modified trypsin (Promega); HPLC-grade methanol (MeOH); trifluoroacetic acid (TFA); acetonitrile (ACN); ammonium formate; formic acid; and ammonium hydroxide (NH<sub>4</sub>OH). All reagents were obtained from Sigma Aldrich (St. Louis, MO) unless otherwise specified.

### Bacterial strains and culture conditions

*Y. pseudotuberculosis* PB1/+ (YPTS), *Yersinia pestis* Pestoides F (YPPF), and a wild-type *Yersinia pestis* CO92 (YPCO) cured of the pPCP1 plasmid were grown in a chemically defined BCS medium<sup>22</sup> in which neutral pH 7.2 was maintained by addition of 50 mM morpholinopropanesulfonic acid (MOPS) as described previously<sup>22, 23</sup>. Bacterial cultures were grown in Erlenmeyer flasks aerated at 200 rpm at 26°C or 28°C. Briefly, a starter culture was grown, diluted to optical density OD<sub>600</sub>= 0.1 to begin overnight culture, and grown to an OD<sub>600</sub> of ~3.0. The overnight culture was diluted to OD<sub>600</sub>= 0.1 and grown in two flasks at 26°C. When the cultures reached OD<sub>600</sub> ~0.5, one flask was moved to 37°C. Similarly cultures grown at 28°C were shifted to 37°C. Aliquots from both cultures 26°C/28°C and 37°C were taken at 0, 1, 2, 4, and 8 hours, optical densities measured, and samples prepared as described below for transcriptomic, proteomic, and metabolomic analyses. The different low temperatures were used due to biofilm formation of YPCO at 26°C and more ideal suspension culturing at 28°C, results were largely similar for the purposes of our analyses.

### Transcriptomics

RNA preparation and microarray analysis of transcripts were performed as previously described in<sup>24</sup>.

### Proteomics analysis

Approximately  $2 \times 10^{10}$  bacteria were harvested from the cultures at each time point, pelleted, and immediately frozen at -80°C. Thawed cell pellets were washed with 100 mM NH<sub>4</sub>HCO<sub>3</sub> (pH 8), lysed via bead beating, and soluble and insoluble protein digestions were performed as described previously<sup>25</sup>. Peptides were concentrated in a Speed-Vac (ThermoFisher, Savant) to ~100 µL, and a BCA protein assay was performed to quantify peptide concentration prior to analysis.

Peptide samples were analyzed using the accurate mass and elution time (AMT) tag approach<sup>26</sup>, which is enabled by a number of published and in-house tools available for download at omics.pnl.gov<sup>27-31</sup>. We note that the scale of the experiment in which 24 comparisons within a single MS experiment are being considered (i.e. 3 strains × 2 temperatures X 4 time points) guided our choice of label-free intensity based quantification.

Briefly, a reference database of AMT tags for peptides from all three *Yersinia* strains employed in this study was previously generated through exhaustive 2-dimensional LC-MS/MS analyses, as described<sup>24</sup>, and was augmented with additional peptides identified in the LC-MS/MS analyses described here. Samples were blocked and randomized to minimize the effects of systematic biases and ensure the even distribution of known and unknown confounding factors across the entire experiment. Peptides from each of the soluble and insoluble protein preparations were analyzed in triplicate using a custom built capillary LC system<sup>32</sup> coupled with an LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, San Jose, CA) via an in-house manufactured electrospray ionization interface, as previously described<sup>33</sup>. RAW files for these datasets are available at <http://omics.pnl.gov> and at [www.Sysbep.org](http://www.Sysbep.org). The LC elution time and monoisotopic mass (determined using the charge state and high accuracy mass measurement) of each peptide feature observed in the analysis were matched to entries within the AMT tag database using the in-house STAC algorithm<sup>34</sup>, which calculates a probability of match. The integrated areas under the elution profiles were used as measures of peptide abundances. Each peptide included for subsequent data analysis was observed in at least one LC-MS analysis with a probability of a correct match being 0.9 and matches for the same peptide in the remaining LC-MS analyses were required to have a minimum probability of 0.5. In addition, at least two unique peptides were required per protein identification. The software program DAnTE<sup>35</sup> was employed to perform an abundance roll-up procedure to convert peptide information to protein information, thereby inferring protein abundances.

### Metabolomics analysis

For <sup>1</sup>H nuclear magnetic resonance (NMR) analysis, 540 μL of spent media were added to 60 μL of 5 mM 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) in 99.9% deuterium oxide (D<sub>2</sub>O) in 5-mm NMR tubes. DSS is used as an internal standard and to provide a <sup>1</sup>H chemical shift reference at δ 0.00 ppm. <sup>1</sup>H NMR spectra were acquired on a Varian INOVA-600 MHz NMR spectrometer (Varian Inc., Palo Alto, CA) at 298 K, using a triple resonance 5-mm HCN salt-tolerant cold probe. A one-dimensional NOE pulse sequence adapted from the two-dimensional Varian tnoesy was used. For each sample, 96 to 512 transients were collected into 64 K data points using a spectral width of 7225.4 Hz with a relaxation delay of 1.0 s, an acquisition time of 4.00 s, and a mixing time of 100 ms. Spectra were processed using Chenomx 6.1. A 0.5-Hz line-broadening function was applied to all spectra prior to Fourier transformation (FT) and baseline correction. The profiler module of Chenomx was used to identify and quantify metabolites.

For GC-MS analysis, metabolites were extracted from the cell culture suspensions using four volumes of chilled (−20°C) chloroform/methanol (2:1, v/v). The aqueous layer obtained after centrifugation (12,000 × g, 5 min) was transferred to a new vial and dried *in vacuo*. All metabolite extracts were then subjected to chemical derivatization to enhance metabolite stability and volatility during analysis<sup>36</sup>. Briefly, 20 μL of methoxyamine in pyridine (30 mg/mL) was added to each dried sample, followed by incubation at 37°C with shaking (1000 rpm) for 90 min. Next, 80 μL of *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) containing 1% trimethylchlorosilane (TMCS) was added to each vial, followed by incubation at 37°C with shaking (1000 rpm) for 30 min. The incubated samples were allowed to cool to room temperature and were then analyzed by GC-MS. The GC-MS system consisted of a 7890A GC-coupled with a single quadrupole MSD 5975C (Agilent Technologies, Santa Clara, CA), and separations were performed on a DB-5MS column (30 m × 0.25 mm × 0.25 μm; Agilent Technologies). The injection mode was splitless, and 1 μL of each sample was injected. The injection port temperature was held at 250°C throughout the analysis. The GC oven was initially maintained at 60°C for 1 min and then ramped to 325°C at 10°C/min, followed by a 5 min hold at 325°C<sup>37</sup>. The obtained GC-MS raw data

files were processed by MetaboliteDetector<sup>38</sup>. Retention indices (RI) were calculated based on the analysis of a mixture of fatty acid methyl esters (FAMES; C8 – C30), and this information was subsequently used to align the retention times of metabolite features detected across GC-MS chromatograms. The chromatographically aligned features were identified using a database, containing mass spectra and retention indices for approximately 700 metabolites<sup>37</sup>.

## Network inference

For network inference analysis transcriptomics and proteomics data from the “ambient” temperature (26°C and 28°C) samples and mammalian host temperature (37°C) samples were used (Suppl. Table 1) to enhance robustness of the network predictions.

Transcriptomics and proteomics data were log<sub>2</sub> transformed as the ratio of 37°C to 26°C samples, or 37°C to 28°C samples. Missing values in proteomics data were dealt with as follows: Proteins for which there were observations for only one strain were removed. Of the remaining 933 proteins, 670 had a complete complement of data values for all strains and conditions. For the remaining 263, if values were missing for all samples in a single strain, it was assumed the protein was not present in that strain and zeros were filled in, representing a non-changing sample for that protein. In many cases, proteins were detected across all time points at one temperature (in a given strain) but absent at another. For these scenarios, it was assumed the differential in temperature was causing the protein level to fall below the level of detection, and the minimum value for the entire data table was filled in. For instances when only 1 time point was missing for a temperature series, the value from the next earlier time point was filled in. If the missing value was from the first time point, the second time point was filled in. All other missing values were filled in with the minimum value of the entire data table.

To determine relationships between transcripts and proteins we used an approach to infer a separate coexpression network for each dataset. To infer edges between network elements, we used context likelihood of relatedness (CLR), an inference algorithm which determines similarity between gene expression profiles based on mutual information between the profiles, and then scores the relationships using a Z-score<sup>39</sup>. For each network we used default parameters for inference using 10 bins for binning data and 3 splines for curve fitting (see<sup>39</sup> for details).

Thresholds for considering a relationship to be an edge in a network were chosen to minimize false negatives and false positives. Interestingly, Z-scores for proteomics were noticeably lower than for transcriptomics (proteomics mean positive Z-score = .67, transcriptomics mean positive Z-score = 1.04) such that a lower percentage of proteomic edges were preserved than transcriptomic edges at any given cutoff. For this reason, we elected to apply differential Z-score thresholds for transcriptomics (5.0) and proteomics (3.0), yielding 3823 transcriptomic edges and 607 proteomics edges, which is a similar ratio to that of transcriptomic and proteomic entities remaining after the initial filtering process (4020/929). Since locus ID tags were common to both datasets, edge files for each were simply merged and redundant edges removed, forming a new integrated network. The full network is provided in a single Cytoscape<sup>40</sup> session file as Supplemental Data.

To find clusters in the network, we used the Louvain community-finding algorithm<sup>41</sup>, which maximizes modularity between communities and returns the corresponding cluster membership. All clusters with a membership of 10 proteins or more were analyzed for enrichment of functional categories using gene ontology. These functional clusters were visualized using the Cytoscape graph visualization package<sup>40</sup>.

## Results and Discussion

### Experimental design and omics data generation

To elucidate mechanisms that contribute to the difference in pathogenicity between YP and YPT, high throughput “omics” technologies were used to analyze YP CO92 (YPCO), YPT PB1/+ (YPTS) and YP Pestoides F (YPPF) cells sampled periodically through a time course of 8 hours (1, 2, 4, and 8 h) post-temperature shift from flea vector (26°C/28°C) to mammalian host (37°C) temperature. YPTS typically causes a mild infection in mammals and birds. The microbe is capable of a saprophytic life style, and thus alternates between the natural environment and warm blooded hosts. The non-epidemic YPPF circulates within natural plague reservoirs and is transmitted between common voles by the fleas, similar to epidemic strains of *Y. pestis* (i.e. YPCO). The only difference between non-epidemic and epidemic strains is the reduced virulence of the former strains to large animals. Thus, a temperature shift model is relevant to all three *Yersinia* used in our studies. Transcriptomics analyses were performed utilizing a *Yersinia* multi-genome microarray which has been previously described<sup>24</sup>. Unlike a traditional microarray that targets annotated genes for a single strain, this microarray incorporates 7641 unique probes designed against seven sequenced *Yersinia* strains on a single chip, including the three strains examined in this study. Proteomics analyses were performed using capillary LC-MS combined with the accurate mass and time (AMT) tag approach<sup>26</sup>. Overall, microarray-based transcriptomics analyses quantified 5005 genes, and mass spectrometry-based proteomics analyses quantified 1365 proteins across all conditions examined. The data are summarized in Suppl. Table 1 and provided as a resource for the community ([www.sysbep.org](http://www.sysbep.org)). This study focused on the response to temperature shift from flea vector temperature 28°C to mammalian host temperature 37°C, represented as log<sub>2</sub> transformed ratios of expression at 37°C relative to 28°C, unless indicated in the text.

### Comparative omics analysis of *Y. pestis* and *Y. pseudotuberculosis*

While YP and YPT have been shown to be very similar genetically (~97% identity at nucleotide level)<sup>3</sup>, the bacteria exhibit markedly different pathogenicities<sup>4</sup>. The molecular mechanism(s) underlying the acute character of the plague infection caused by YP in the mammalian host is poorly understood. We hypothesized that differential expression of genes common to both organisms, in addition to overt genetic differences contributes to their significantly different pathogenicities in the mammalian host. To test this hypothesis, thermal-shift transcript and protein profiles of YP CO92 (YPCO) and YPT PB1/+ (YPTS) were analyzed.

From our analysis, we identified 70 genes preferentially up-regulated in YPCO relative to YPTS following a shift from flea vector to mammalian host temperature, suggesting a potential functional role in the mammalian host context. Examination of these 70 genes revealed that 66 genes were located on the pCD1 plasmid, which is essential for virulence in all human pathogenic *Yersinia* strains, with 44 genes annotated as encoding Yops or components of the T3SS (Suppl. Figure 1 and Suppl. Table 2). These included secretion apparatus members, chaperones, effectors and low calcium response genes. In agreement with the transcriptional data, all pCD1-encoded Yops and T3SS proteins detected by proteomics were preferentially up-regulated in YPCO relative to YPTS in response to temperature elevation (Figure 1). In fact, protein expression was not detected for any of these proteins in YPTS across all time points. While it is possible that these proteins were simply not detected due to the stochastic nature of MS-based proteomics analysis methods, a more likely explanation is low protein expression levels in YPTS that fall beneath the detection level of the instruments employed for analysis. YadA is a factor important for YPT pathogenesis but thought to be either absent or nonfunctional in YP<sup>42</sup>. Here our microarray

data shows detection of YadA transcripts in both YPCO and YPTS suggesting that expression of this pseudogene does occur in YP (Figure 1). Examination of the corresponding proteomics data however shows presence of YadA protein in YPTS alone but not YPCO consistent with it being non-functional in YP (Figure 1). Taken together these observations suggest a role for post-transcriptional regulatory mechanisms in modulating the genotype to phenotype expression of YadA in *Yersinia*.

Three of the four chromosomally-encoded genes preferentially up-regulated in YPCO relative to YPTS following temperature increase are required for the assembly of a functional pH6 antigen (psaABC; Suppl. Figure 1). The fourth chromosomally-encoded gene is an acid resistance membrane protein (YPO0590). The pH6 antigen (Psa) is conserved across pathogenic *Yersinia* species and is thought to contribute to both YP and YPT virulence in the mammalian host. In YPT, Psa was reported to be a thermoinducible adhesin that allows binding of the organism to cultured mammalian epithelial cells<sup>43</sup>. In YP a *psa* deletion mutant strain was shown to be attenuated by the intravenous route of infection<sup>44</sup> and more recently Psa was shown to promote resistance to phagocytosis further clarifying its role in virulence<sup>45</sup>. Taken together, these observations suggest that the increased pathogenicity of YP relative to YPT is not only due to genomic differences, but stems from differences in transcriptional regulatory networks resulting in higher transcript and protein expression of common essential virulence mechanisms.

These observations encouraged examination of the extent that differential expression of other major virulence mechanisms conserved across pathogenic *Yersinia* species contribute to observed differences in pathogenicity. The high pathogenicity island (HPI)-encoded *ybt* locus is conserved across pathogenic *Yersinia* spp with 97–100% identity and is comprised of genes required for the biosynthesis and secretion of yersiniabactin, a siderophore that is essential for *Yersinia* virulence<sup>46</sup>. Transcripts of the genes encoding yersiniabactin biosynthetic proteins (YPO1907–1911), as well as the transcript for the yersiniabactin receptor (YPO1906), were down-regulated following temperature shift in YPCO compared to YPTS (Suppl. Figure 2). The ferric uptake regulator (Fur) negatively regulates transcription of *ybt* genes while transcriptional regulator YbtA positively regulates transcription of *ybt* genes<sup>47</sup>. No differences between YPCO and YPTS transcript and protein levels were observed for Fur. Similarly no differences between YPCO and YPTS transcript levels were observed for YbtA. YbtA protein expression was not detected in either organism. Taken together these results suggest that regulation of yersiniabactin biosynthesis and secretion is different in YPCO and YPTS and likely comprises additional yet unrecognized levels of transcriptional or post-transcriptional control that contribute to the differences in pathogenesis between these strains. This is supported by previous work showing that an inactivating mutation in the yersiniabactin receptor YPO1906 causes loss of siderophore production in YPT, but not in YP<sup>48</sup>.

### Comparative omics of *Y. pestis* CO92 and *Y. pestis* Pestoides F

We compared YP CO92 (YPCO) to the non-epidemic YP strain Pestoides F (YPPF) to gain insights to the mechanism underlying the virulence-restricted phenotype of non-epidemic YP strains compared to epidemic YP strains. From our analysis, we identified 66 genes preferentially expressed in one organism or the other in response to temperature shift (Suppl. Table 3) as judged by clustering analysis. A majority of these genes, 51, exhibited a coordinated down-regulation of transcript and protein levels following a shift from flea vector temperature to mammalian host temperature suggesting a potential functional role in the flea vector context. The remaining 15 genes were up-regulated following temperature shift in this case, suggesting a potential functional role in the mammalian host context. All 15 genes that were up-regulated following temperature shift were preferentially expressed in YPCO relative to YPPF, with 11 of the 15 genes encoded on the pMT1 virulence plasmid.

The pMT1 plasmid is a 100 kb YP-specific plasmid that encodes approximately 100 genes, including the two well characterized virulence factors F1 capsule protein (Caf1) and Yersinia murine toxin (Ymt). Ymt is a phospholipase D (PLD) shown to be required for survival of YP in the midgut of the flea but not required for virulence in mice<sup>12, 49</sup>. Caf1 was previously demonstrated to be an anti-phagocytic factor that inhibits the internalization of YP by macrophages<sup>9</sup>, and additionally promotes transmission by flea bite, increasing the potential for epidemic spread<sup>50</sup>. More recent results suggest that Caf1 is required for pathogenesis during bubonic and pneumonic infections<sup>51</sup>. It is clear that acquisition of the pMT1 plasmid, and thus the *ymt* and *caf* loci, enhanced YP potential for epidemic spread. It is currently unknown but likely that additional virulence determinants are encoded on the pMT1 virulence plasmid. The 11 pMT1 plasmid encoded genes preferentially expressed in YPCO relative YPPF warrant further investigation as potential molecular switches for the epidemic versus non-epidemic phenotypes of YPCO and YPPF.

### Multi-omics integration for network prediction of putative virulence factors

While the identification of factors that contribute to species-specific pathogenic attributes of YP and YPT is important, equally important is the elucidation of the full repertoire of determinants that play a role in core pathogenic processes required for the virulence mechanism of pathogenic *Yersinia* spp. Here the context likelihood of relatedness (CLR) algorithm<sup>39</sup> was employed to integrate the proteomics and transcriptomics datasets and predict proteins important for Yersinia pathogenesis in an efficient and systematic manner. Since proteomic expression profiles are often poorly correlated with transcriptomic profiles, there is likely additional regulatory information in the combined results. The value added of an integrated approach, is therefore the incorporation of potential impacts of post-transcriptional regulation on the network analysis. The robustness of the predictions was enhanced by the incorporation of omics data from bacteria grown at both flea vector temperatures, 26°C and 28°C, and at the mammalian host temperature (37°C) (Suppl. Table 1). The proteomics and transcriptomics datasets were integrated at the network level to generate a multi-omics network, and subsequent functional enrichment analyses were used to predict proteins important for pathogenesis as described below. A similar approach was previously employed to predict novel virulence factors in the related pathogen *Salmonella* Typhimurium which were subsequently verified experimentally<sup>52</sup>.

These proteomic and transcriptomic networks were then combined under specific thresholds to create an integrated multi-omics network including only high confidence protein to protein (Z score > 3.0) and transcript to transcript (Z score > 5.0) relationships (see the Materials and Methods section for the fractional contribution of transcriptomics and proteomics). The network was partitioned using a community-finding algorithm<sup>41</sup>, and each of the resulting clusters was analyzed for enrichment of functional categories using gene ontology (GO) analysis. Among the most significantly enriched functional clusters was the cluster labeled 'Type III secretion' (T3SS) as shown in Figure 2. To predict proteins potentially important for *Yersinia* pathogenesis, further analysis was focused on this cluster because of the established characterization of the T3SS as a major *Yersinia* virulence factor, and proteins co-regulated with it across multiple organisms, conditions and time points are likely to play a role in *Yersinia* pathogenesis.

Examination of the T3SS-associated functional cluster revealed 151 cluster members (Suppl. Table 4), with 34 members encoded on the pCD1 virulence plasmid, including 28 annotated as members of the Yop virulon. Fifty three cluster members are encoded on the pMT1 plasmid including the major virulence determinants Yersinia murine toxin (Ymt) and the F1 capsule protein (Caf1). Sixty four cluster members are chromosomally-encoded, the majority of which are annotated with unknown function. Among the subset of chromosome-encoded cluster members were several proteins that have been suggested to play a role in



pathogenesis in *Yersinia* including PgaA, HmsR, Ail, Asr, RpoS, and NlpD. For example, the lipoprotein NlpD was recently shown to be a novel *Yersinia pestis* virulence factor<sup>53</sup>. Chromosomal deletion of the nlpD gene sequence resulted in a drastic reduction in virulence to an LD<sub>50</sub> of at least 10<sup>7</sup> cfu for subcutaneous and airway routes of infection, and the mutant was unable to colonize mouse organs following infection<sup>53</sup>. Given the very high enrichment of virulence-related proteins and proteins important for adaptation to the host environment in this cluster, the remaining uncharacterized members of this cluster are predicted to be enriched for proteins important for *Yersinia* pathogenesis.

### Comparison of omics datasets across multiple *Yersinia* strains reveals conserved biological processes under post-transcriptional control

In most instances, gene expression in YPCO relative to YPTS or YPPF was confirmed by the proteomics data if the protein of interest was confidently detected. However, this analysis also revealed a number of cases where changes in protein expression were in contrast with the observations from the transcriptional analysis (i.e. transcript appeared unaffected by the temperature shift, while protein displayed substantial change), highlighting the importance of proteomics in addition to transcriptional analyses.

Typically, disparities between transcript and protein measurements are often attributed to post-transcriptional regulation<sup>54, 55</sup>, thus these observations prompted examination of instances of apparent post-transcriptional regulation in response to temperature switch on a global scale. First, transcript and protein responses to temperature shifts were considered for each gene at the same time-point within each of the three organisms by calculating Pearson correlations. A general trend of improved correlation was observed between transcript and protein over time across all three organisms. For example, in the case of YPCO transcript-protein correlations of 0.28, 0.55, 0.65, and 0.53 were observed at 1h, 2h, 4h, and 8h respectively (Suppl. Figure 3). It is possible that this general trend of improved correlation may be explained by adaptation to new environmental conditions as the later time points reach a steady-state with only subtle changes in macromolecules needed. This is in contrast to the 1hr time point during which there appears to be a vigorous dynamic adaptation in response to temperature shift occurring, necessitating substantial transcription and translation of newly required biomolecules.

While low transcript-protein correlation is typically interpreted as evidence of post-transcriptional regulation, it is very likely that temporal lags between dynamic changes in transcription and translation at the level of individual genes also represent an important contribution to the observed low correlation; although this has rarely been demonstrated as most studies reporting low correlation have been single time-point studies. The temporal sample-matched global transcript and protein datasets presented here allow us to begin to evaluate this in part with regards to temporal lag on a genome-scale. If the assertion that temporal lags between dynamic changes in transcription and translation represent an important contribution to the low transcript-protein correlations observed is correct, then an improvement in transcript-protein correlation would be expected as a time lag is introduced in analysis of the correlation of the temporal transcript to the protein data. Indeed this was observed across all three strains when transcript and protein response to temperature shift were compared by calculating Pearson correlations with and without a 1h time lag (Suppl. Figure 4). For YPCO a transcript-protein correlation improvement of 0.28 to 0.57 was observed. Similarly for YPTS and YPPF a transcript-protein correlation improvements of 0.49 and 0.33 to 0.56 and 0.62, respectively, are observed.

Considering the improvement in global transcript-protein correlations associated with time, genome-scale instances of apparent post-transcriptional regulation in response to temperature shift were examined by performing an ANOVA analysis. The goal of this

analysis was to identify genes with transcript response significantly different from protein response to temperature shift across all time points. The ANOVA analysis identified 173 YPCO genes with a significant difference ( $p < 0.05$ ) between mRNA and proteins responses to temperature shift across all time points, with similar numbers of differences for YPPF (194) and YPTS (128) (Figure 3 and Suppl. Table 5). The Database for Annotation, Visualization and Integrated Discovery (DAVID)<sup>56</sup> was used to identify enriched functionally-related genes representing particular biological processes from each of these 3 different gene lists with the following criteria: Benjamini-Hochberg corrected  $p$ -value  $< 0.05$  and gene count  $\geq 5$ . Biological processes functionally enriched in the list of genes are shown in Table 1. Observing apparent post-transcriptional control of biological processes across multiple organisms increases the confidence in their assignment. Thus the data were examined for those inferred biological processes under potential post-transcriptional control that were conserved across at least two of the three *Yersinia* under investigation (see Table 1; bolded text). Among these post-transcriptional controlled proteins were purine metabolism, pyrimidine metabolism, and amino-acyl tRNA biosynthesis which showed conservation across all three organisms. Pyruvate metabolism and glycolysis/gluconeogenesis were conserved across YPCO and YPPF while ribosomes were conserved across YPPF and YPTS. Interestingly, previous work in the bacterial pathogen *Salmonella* Typhimurium has also suggested that general metabolism (including purine and pyrimidine metabolism, glycolysis/gluconeogenesis, the TCA pathway, and pyruvate metabolism) and the translational machinery (including aminoacyl-tRNA synthetases) are under post-transcriptional control, mediated to a large extent by the global post-transcriptional regulator Hfq which is required for *Salmonella* virulence<sup>57-59</sup>. Post-transcriptional regulation of general metabolism and the translational machinery, among other processes, is speculated to similarly play an important role in *Yersinia* adaptation to the mammalian host intracellular environment and in pathogenesis. Indeed the global post-transcriptional regulator Hfq has been shown to be required for virulence in *Yersinia*<sup>60, 61</sup>.

### Metabolomics analyses suggest post-translational control of glutamate levels

Advances in mass spectrometry and nuclear magnetic resonance (NMR) techniques in the past decade now allow for global, high throughput profiling of metabolites produced by bacteria. Here, NMR was used to analyze spent media samples derived from two YP strains (YPCO and YPPF) and the YPT strain PB1/+ (YPTS) grown at the flea vector and mammalian host temperatures and sampled at 4hr post-culturing. These samples were prepared from the cultures used for transcriptomics and proteomics analyses. A striking observation from this analysis was the difference in glutamate concentrations across the three strains. Glutamate concentration in the YPCO sample was significantly higher than in either YPTS or YPPF samples (data not shown). As an orthogonal measure of the observed differences in this metabolite across the three strains, whole cell cultures of YPCO, YPPF and YPTS (grown at both flea vector and mammalian host temperatures and sampled at 1hr, 2hr, 4hr and 8hr post-culturing) were analyzed by gas chromatography-mass spectrometry (GC-MS). In support of the NMR results, the GC-MS data showed that glutamic acid, which was not an ingredient of the culture media (i.e. an endogenous metabolite), accumulated to high levels in the YPCO cell culture with time, while remaining at background levels in YPTS and YPPF (Suppl. Figure 5). Interestingly, the level of glutamate in YPCO sample grown at mammalian host temperature reached maximum at the 4hr time point, and dropped significantly at 8hr time point. In contrast, there was little difference in glutamate concentration in YPCO sample grown at flea vector temperature at the 4hr and 8hr time points. These data were confirmed with the measurement of glutamate by using Amplex Red Glutamic Acid/Glutamate Oxidase fluorescence assay kit from Invitrogen (data not shown). It was observed previously, that during adaptation of YP to the flea gut environment, plague bacterium may catabolize L-glutamate group of amino acids, such as glutamine, histidine,

arginine, and proline, to give rise to L-glutamate and the TCA cycle intermediates succinate, formate, and  $\alpha$ -ketoglutarate<sup>62</sup>. We have not detected differences in expression of genes involved in uptake and catabolism of these amino acids between three *Yersinia* strains, as well as in concentrations of histidine, arginine, proline, succinate, and formate. The example of this comparison for the transcript and protein levels of key enzymes in glutamate metabolism of YPCO relative YPPF and YPTS is shown on Figure 4. No significant differences were observed in the expression patterns of glutamate dehydrogenase, glutamate synthase alpha and beta subunits, or other key enzymes in glutamate metabolism across all three strains. Given that the difference in glutamate levels between the three strains were not reflected at the level of gene transcription or protein translation, these results implicate post-translational control as a previously unappreciated additional regulatory mechanism involved in modulating glutamate levels *in vivo*. Interestingly, a recent study of nitrogen metabolism in *Mycobacterium smegmatis* investigated the transcription and specific enzyme activity of glutamine synthetase and glutamate dehydrogenase. The authors found that glutamate dehydrogenase activity was not reflected at the level of gene transcription, thereby implicating post-transcriptional modification as a regulatory mechanism in response to nitrogen availability<sup>63</sup>.

## Conclusions

The molecular mechanisms underlying the enhanced pathogenesis of YP relative to YPT (~97% sequence identity) remain elusive despite several comparative genomic studies. Here using a systems biology approach using sample-matched transcriptomic and proteomic profiling appear to support the hypothesis that differential expression of genes common to both YP and YPT, as opposed to the mere presence or absence of species-specific genes, contributes to the striking difference in the diseases caused by these pathogens; and that similar mechanisms may also contribute to the virulence-restricted phenotype of the non-epidemic YP *microtus* biovar. Of particular interest, the matched omics profiling of multiple *Yersinia* strains also revealed conserved post-transcriptional control of metabolism and the translational machinery including the modulation of glutamate levels in *Yersinia*. Additionally, statistical inference modeling methods were able to predict novel *Yersinia* virulence factors. This work highlights the utility of a systems approach incorporating multiple omics measurements and computational analyses to provide novel insights into *Yersinia* biology; and provides an important resource for the *Yersinia* research community that should aid the understanding of the markedly different pathogenicities of YP and YPT.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Research described was supported by the National Institute of Allergy and Infectious Diseases NIH/DHHS through Interagency agreement Y1-AI-8401. Proteomics capabilities were developed under support from the U.S. Department of Energy (DOE) Office of Biological and Environmental Research (BER), NIH grant 5P41RR018522-10 and National Institute of General Medical Sciences grant 8 P41 GM103493-10. Metabolomics capabilities were developed under support from the U.S. DOE BER. Work was performed in the Environmental Molecular Sciences Laboratory, a DOE-BER national scientific user facility at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated by Battelle Memorial Institute for the DOE under contract DE-AC05-76RLO 1830.

## References

1. Perry RD, Fetherston JD. Clin Microbiol Rev. 1997; 10:35–66. [PubMed: 8993858]

2. Achtman M, Morelli G, Zhu P, Wirth T, Diehl I, Kusecek B, Vogler AJ, Wagner DM, Allender CJ, Easterday WR, Chenal-Francisque V, Worsham P, Thomson NR, Parkhill J, Lindler LE, Carniel E, Keim P. *Proc Natl Acad Sci U S A*. 2004; 101:17837–17842. [PubMed: 15598742]
3. Chain PS, Carniel E, Larimer FW, Lamerdin J, Stoutland PO, Regala WM, Georgescu AM, Vergez LM, Land ML, Motin VL, Brubaker RR, Fowler J, Hinnebusch J, Marceau M, Medigue C, Simonet M, Chenal-Francisque V, Souza B, Dacheux D, Elliott JM, Derbise A, Hauser LJ, Garcia E. *Proc Natl Acad Sci U S A*. 2004; 101:13826–13831. [PubMed: 15358858]
4. Brubaker, RR. *Yersinia pestis and bubonic plague*. Springer; New York: 2006.
5. Zhou D, Tong Z, Song Y, Han Y, Pei D, Pang X, Zhai J, Li M, Cui B, Qi Z, Jin L, Dai R, Du Z, Wang J, Guo Z, Huang P, Yang R. *J Bacteriol*. 2004; 186:5147–5152. [PubMed: 15262951]
6. Anisimov AP, Lindler LE, Pier GB. *Clin Microbiol Rev*. 2004; 17:434–464. [PubMed: 15084509]
7. Cornelis GR, Van Gijsegem F. *Annu Rev Microbiol*. 2000; 54:735–774. [PubMed: 11018143]
8. Brubaker RR. *Infect Immun*. 2003; 71:3673–3681. [PubMed: 12819047]
9. Du Y, Rosqvist R, Forsberg A. *Infect Immun*. 2002; 70:1453–1460. [PubMed: 11854232]
10. Ferber DM, Brubaker RR. *Infect Immun*. 1981; 31:839–841. [PubMed: 7216478]
11. Hinchliffe SJ, Isherwood KE, Stabler RA, Prentice MB, Rakin A, Nichols RA, Oyston PC, Hinds J, Titball RW, Wren BW. *Genome Res*. 2003; 13:2018–2029. [PubMed: 12952873]
12. Hinnebusch BJ, Rudolph AE, Cherepanov P, Dixon JE, Schwan TG, Forsberg A. *Science*. 2002; 296:733–735. [PubMed: 11976454]
13. Lahteenmaki K, Virkola R, Saren A, Emody L, Korhonen TK. *Infect Immun*. 1998; 66:5755–5762. [PubMed: 9826351]
14. Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebahia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeno-Tarraga AM, Chillingworth T, Cronin A, Davies RM, Davis P, Dougan G, Feltwell T, Hamlin N, Holroyd S, Jagels K, Karlyshev AV, Leather S, Moule S, Oyston PC, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. *Nature*. 2001; 413:523–527. [PubMed: 11586360]
15. Drozdov IG, Anisimov AP, Samoilova SV, Yezhov IN, Yeremin SA, Karlyshev AV, Krasilnikova VM, Kravchenko VI. *J Med Microbiol*. 1995; 42:264–268. [PubMed: 7707334]
16. Du Y, Galyov E, Forsberg A. *Contrib Microbiol Immunol*. 1995; 13:321–324. [PubMed: 8833863]
17. Filippov AA, Solodovnikov NS, Kookleva LM, Protsenko OA. *FEMS Microbiol Lett*. 1990; 55:45–48. [PubMed: 2328909]
18. Friedlander AM, Welkos SL, Worsham PL, Andrews GP, Heath DG, Anderson GW Jr, Pitt ML, Estep J, Davis K. *Clin Infect Dis*. 1995; 21(Suppl 2):S178–181. [PubMed: 8845449]
19. Kuttyrev V, Mehig RJ, Motin VL, Pokrovskaya MS, Smirnov GB, Brubaker RR. *Infect Immun*. 1999; 67:1359–1367. [PubMed: 10024583]
20. Pouillot F, Derbise A, Kukkonen M, Foulon J, Korhonen TK, Carniel E. *Microbiology*. 2005; 151:3759–3768. [PubMed: 16272397]
21. Welkos SL, Andrews GP, Lindler LE, Snellings NJ, Strachan SD. *Plasmid*. 2004; 51:1–11. [PubMed: 14711524]
22. Fowler JM, Brubaker RR. *Infect Immun*. 1994; 62:5234–5241. [PubMed: 7960099]
23. Brubaker RR. *Infect Immun*. 2005; 73:4743–4752. [PubMed: 16040987]
24. Schrimpe-Rutledge AC, Jones MB, Chauhan S, Purvine SO, Sanford JA, Monroe ME, Brewer HM, Payne SH, Ansong C, Frank BC, Smith RD, Peterson SN, Motin VL, Adkins JN. *PLoS One*. 2012; 7:e33903. [PubMed: 22479471]
25. Mottaz-Brewer HM, Norbeck AD, Adkins JN, Manes NP, Ansong C, Shi L, Rikihisa Y, Kikuchi T, Wong SW, Estep RD, Heffron F, Pasa-Tolic L, Smith RD. *J Biomol Tech*. 2008; 19:285–295. [PubMed: 19183792]
26. Zimmer JS, Monroe ME, Qian WJ, Smith RD. *Mass Spectrom Rev*. 2006; 25:450–482. [PubMed: 16429408]
27. Jaitly N, Monroe ME, Petyuk VA, Clauss TR, Adkins JN, Smith RD. *Anal Chem*. 2006; 78:7397–7409. [PubMed: 17073405]
28. Kiebel GR, Auberry KJ, Jaitly N, Clark DA, Monroe ME, Peterson ES, Tolic N, Anderson GA, Smith RD. *Proteomics*. 2006; 6:1783–1790. [PubMed: 16470653]

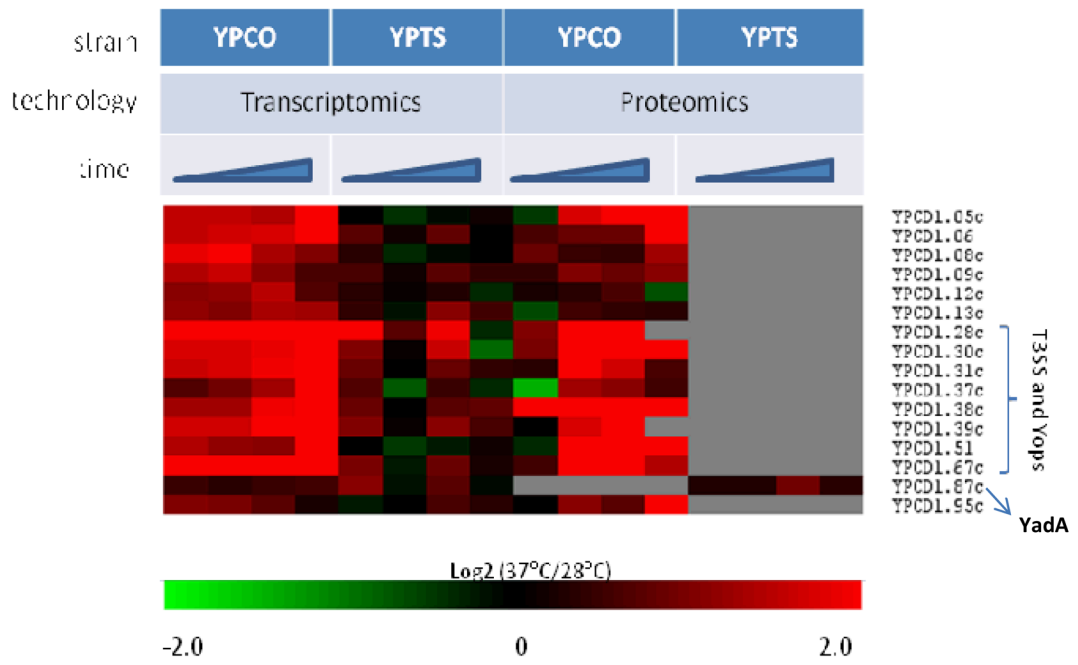
29. Monroe ME, Shaw JL, Daly DS, Adkins JN, Smith RD. *Comput Biol Chem.* 2008; 32:215–217. [PubMed: 18440872]
30. Monroe ME, Tolic N, Jaitly N, Shaw JL, Adkins JN, Smith RD. *Bioinformatics.* 2007; 23:2021–2023. [PubMed: 17545182]
31. Petritis K, Kangas LJ, Yan B, Monroe ME, Strittmatter EF, Qian WJ, Adkins JN, Moore RJ, Xu Y, Lipton MS, Camp DG 2nd, Smith RD. *Anal Chem.* 2006; 78:5026–5039. [PubMed: 16841926]
32. Livesay EA, Tang K, Taylor BK, Buschbach MA, Hopkins DF, LaMarche BL, Zhao R, Shen Y, Orton DJ, Moore RJ, Kelly RT, Udseth HR, Smith RD. *Anal Chem.* 2008; 80:294–302. [PubMed: 18044960]
33. Kelly RT, Page JS, Tang K, Smith RD. *Anal Chem.* 2007; 79:4192–4198. [PubMed: 17472340]
34. Stanley JR, Adkins JN, Slys GW, Monroe ME, Purvine SO, Karpievitch YV, Anderson GA, Smith RD, Dabney AR. *Anal Chem.* 2011; 83:6135–6140. [PubMed: 21692516]
35. Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG 2nd, Anderson GA, Smith RD. *Bioinformatics.* 2008; 24:1556–1558. [PubMed: 18453552]
36. Kim YM, Metz TO, Hu Z, Wiedner SD, Kim JS, Smith RD, Morgan WF, Zhang Q. *Rapid Commun Mass Spectrom.* 2011; 25:2561–2564. [PubMed: 21910289]
37. Kind T, Wohlgemuth G, Lee do Y, Lu Y, Palazoglu M, Shahbaz S, Fiehn O. *Anal Chem.* 2009; 81:10038–10048. [PubMed: 19928838]
38. Hiller K, Hangebrauk J, Jager C, Spura J, Schreiber K, Schomburg D. *Anal Chem.* 2009; 81:3429–3439. [PubMed: 19358599]
39. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. *PLoS Biol.* 2007; 5:e8. [PubMed: 17214507]
40. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. *Genome Res.* 2003; 13:2498–2504. [PubMed: 14597658]
41. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. *Journal of Statistical Mechanics.* 2008; 10:P10008.
42. Han YW, Miller VL. *Infect Immun.* 1997; 65:327–330. [PubMed: 8975933]
43. Yang Y, Merriam JJ, Mueller JP, Isberg RR. *Infect Immun.* 1996; 64:2483–2489. [PubMed: 8698470]
44. Lindler LE, Klempner MS, Straley SC. *Infect Immun.* 1990; 58:2569–2577. [PubMed: 2164509]
45. Huang XZ, Lindler LE. *Infect Immun.* 2004; 72:7212–7219. [PubMed: 15557646]
46. Perry RD, Fetherston JD. *Microbes Infect.* 2011; 13:808–817. [PubMed: 21609780]
47. Marceau M. *Curr Issues Mol Biol.* 2005; 7:151–177. [PubMed: 16053248]
48. Perry RD, Balbo PB, Jones HA, Fetherston JD, DeMoll E. *Microbiology.* 1999; 145(Pt 5):1181–1190. [PubMed: 10376834]
49. Hinnebusch J, Cherepanov P, Du Y, Rudolph A, Dixon JD, Schwan T, Forsberg A. *Int J Med Microbiol.* 2000; 290:483–487. [PubMed: 11111930]
50. Sebbane F, Jarrett C, Gardner D, Long D, Hinnebusch BJ. *Infect Immun.* 2009; 77:1222–1229. [PubMed: 19103769]
51. Weening EH, Cathelyn JS, Kaufman G, Lawrenz MB, Price P, Goldman WE, Miller VL. *Infect Immun.* 2011; 79:644–652. [PubMed: 21115720]
52. Yoon H, Ansong C, McDermott JE, Gritsenko M, Smith RD, Heffron F, Adkins JN. *BMC Syst Biol.* 2011; 5:100. [PubMed: 21711513]
53. Tidhar A, Flashner Y, Cohen S, Levi Y, Zauberman A, Gur D, Aftalion M, Elhanany E, Zvi A, Shafferman A, Mamroud E. *PLoS One.* 2009; 4:e7023. [PubMed: 19759820]
54. Greenbaum D, Colangelo C, Williams K, Gerstein M. *Genome Biol.* 2003; 4:117. [PubMed: 12952525]
55. Gry M, Rimini R, Stromberg S, Asplund A, Ponten F, Uhlen M, Nilsson P. *BMC Genomics.* 2009; 10:365. [PubMed: 19660143]
56. Huang da W, Sherman BT, Lempicki RA. *Nucleic Acids Res.* 2009; 37:1–13. [PubMed: 19033363]

57. Ansong C, Yoon H, Porwollik S, Mottaz-Brewer H, Petritis BO, Jaitly N, Adkins JN, McClelland M, Heffron F, Smith RD. *PLoS One*. 2009; 4:e4809. [PubMed: 19277208]
58. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, Binnewies TT, Hinton JC, Vogel J. *PLoS Genet*. 2008; 4:e1000163. [PubMed: 18725932]
59. Sittka A, Pfeiffer V, Tedin K, Vogel J. *Mol Microbiol*. 2007; 63:193–217. [PubMed: 17163975]
60. Geng J, Song Y, Yang L, Feng Y, Qiu Y, Li G, Guo J, Bi Y, Qu Y, Wang W, Wang X, Guo Z, Yang R, Han Y. *PLoS One*. 2009; 4:e6213. [PubMed: 19593436]
61. Schiano CA, Bellows LE, Lathem WW. *Infect Immun*. 2010; 78:2034–2044. [PubMed: 20231416]
62. Vadyvaloo V, Jarrett C, Sturdevant DE, Sebbane F, Hinnebusch BJ. *PLoS Pathog*. 2010; 6:e1000783. [PubMed: 20195507]
63. Harper CJ, Hayward D, Kidd M, Wiid I, van Helden P. *BMC Microbiol*. 2010; 10:138. [PubMed: 20459763]

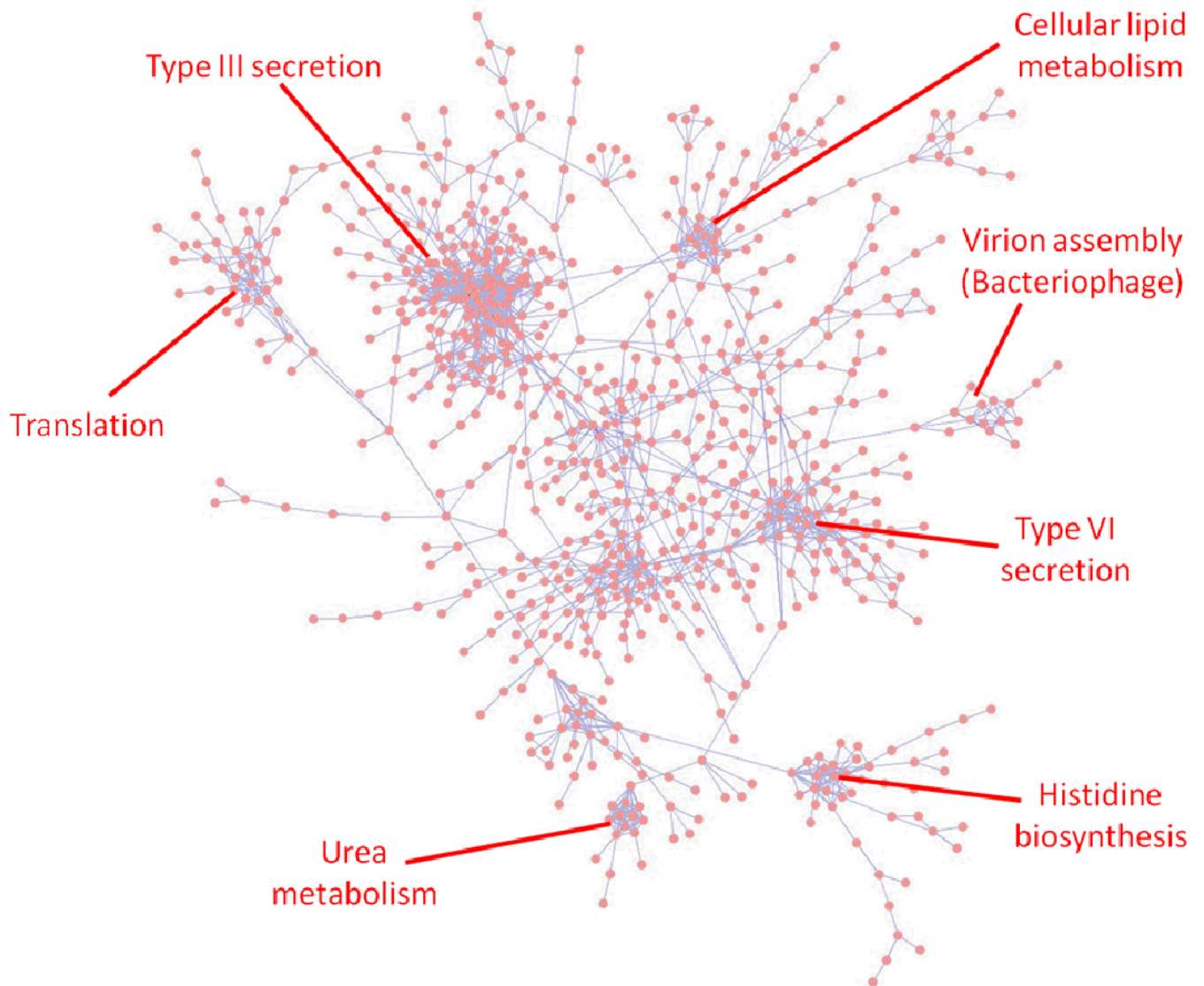
\$watermark-text

\$watermark-text

\$watermark-text

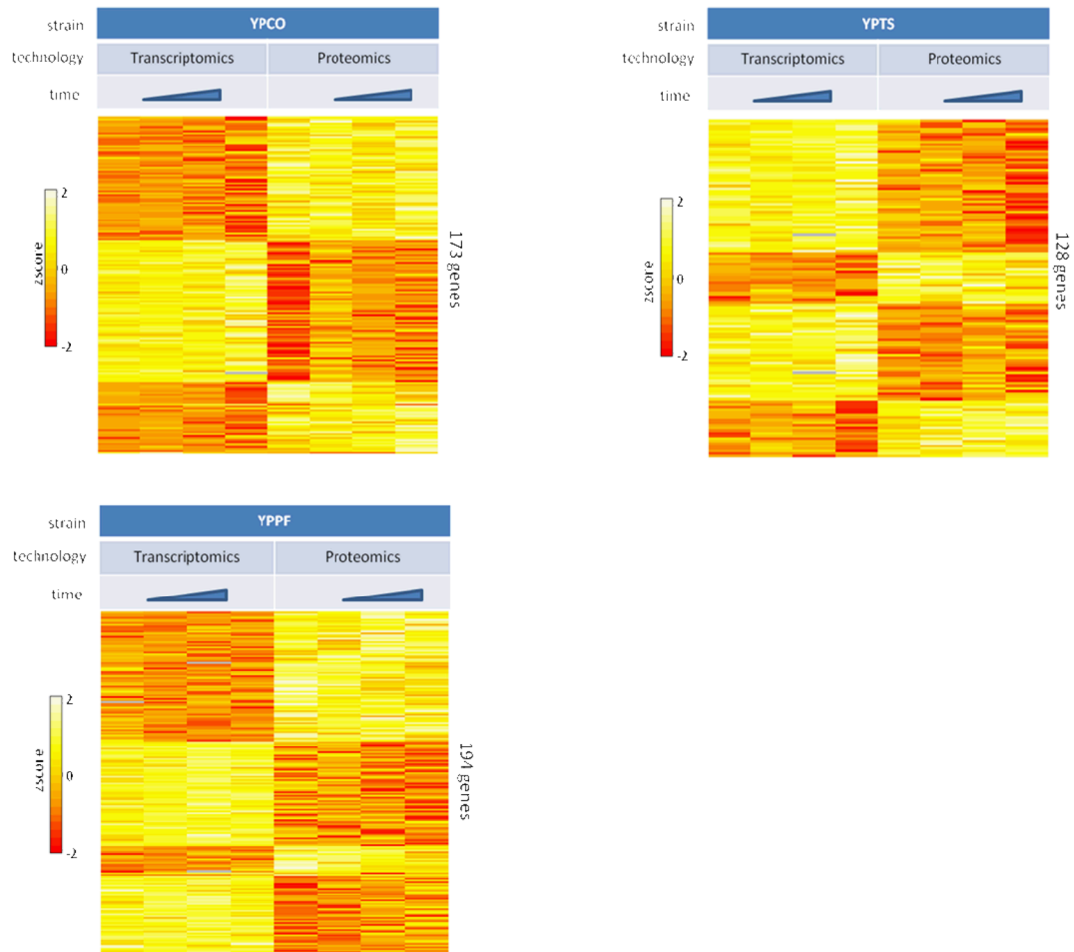


**Figure 1.** Heat map illustration of pCD1-encoded proteins, detected by both proteomics and transcriptomics, preferentially up-regulated in YPCO relative to YPTS in response to temperature elevation. Time represents sampling points at 1hr, 2hr, 4hr, 8hr. YPCO, *Yersinia pestis* CO92; YPTS, *Yersinia pseudotuberculosis* PB1/+.

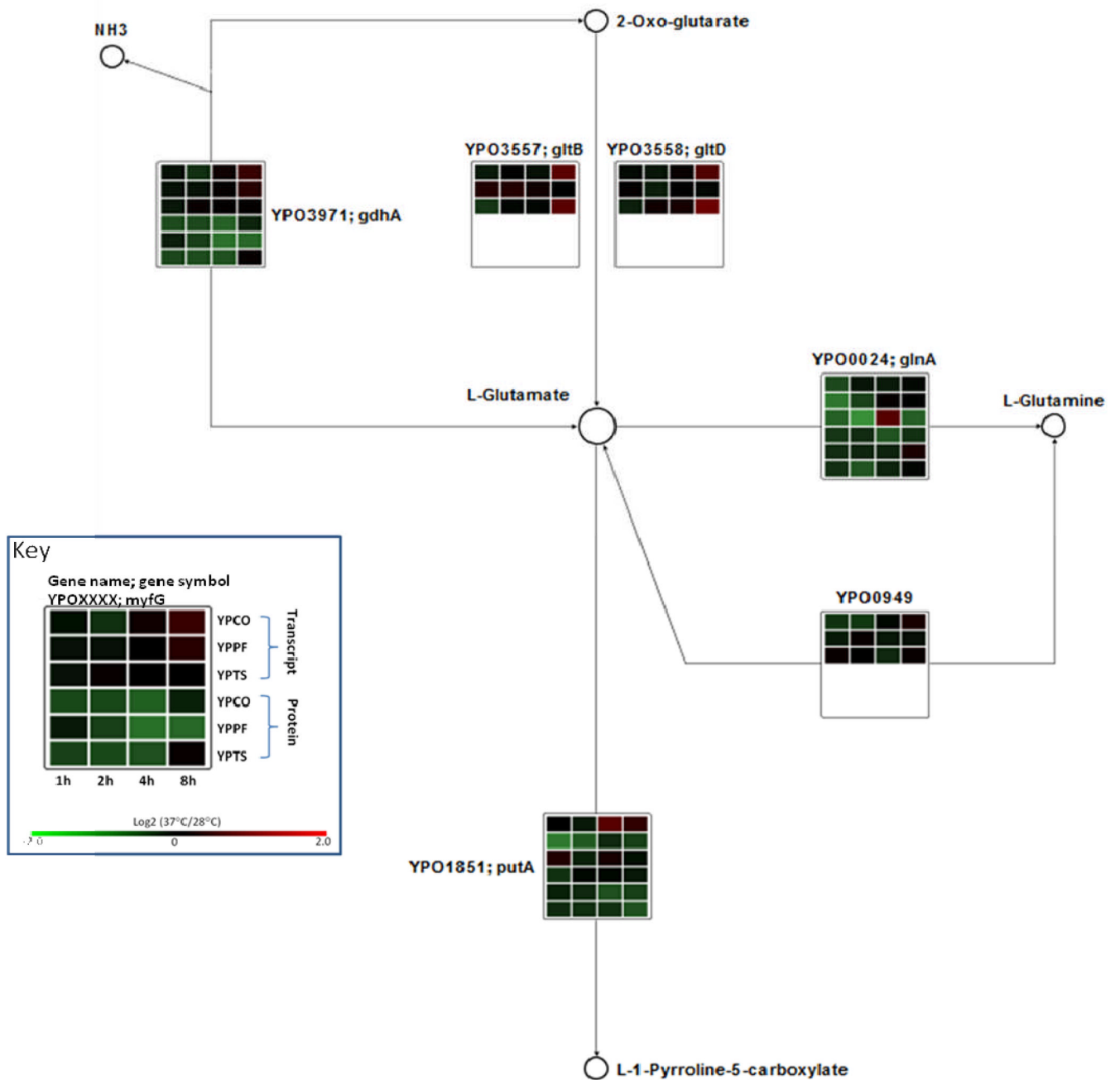


**Figure 2.** Association network inferred from integrated proteomic and transcriptomic data. The CLR method was used to infer association relationships between proteins on the basis of their abundance profiles ( $Z$  score  $> 3.0$ ). The resulting network was extended by combining with association relationships inferred from transcriptomics data ( $Z$  score  $> 5.0$ ). The network was visualized in Cytoscape. Examples of significantly enriched functional clusters are indicated in the figure.





**Figure 3.** Heat map illustration of genes with transcript response to temperature shift across all time points significantly different from protein response to temperature shift across all time points as determined by ANOVA analysis ( $p < 0.05$ ) for each organism. Time represents sampling points at 1hr, 2hr, 4hr, 8hr. YPCO, *Yersinia pestis* CO92; YPTS, *Yersinia pseudotuberculosis* PB1/+; YPPF, *Yersinia pestis* Pestoides F.



**Figure 4.** Pathway diagram illustrating comparison of transcript and protein levels of key enzymes in glutamate metabolism across all three strains. Inset panel provides a key/legend for interpreting the pathway elements. YPCO, *Yersinia pestis* CO92; YPTS, *Yersinia pseudotuberculosis* PB1/+; YPPF, *Yersinia pestis* Pestoides F.

**Table 1**  
**Post-transcriptionally regulated biological process**

Biological processes functionally enriched in the list of genes with a significant difference ( $p < 0.05$ ) between mRNA and proteins responses to temperature shift across all time points for YPCO, YPPF and YPTS.

Organism	Biological Process	Gene Count	Fold Enrichment	Benjamini PValue
YPCO	<b>Pyrimidine metabolism</b>	11	13	1.47E-06
	<b>Purine metabolism</b>	9	7	1.70E-03
	<b>Aminoacyl-tRNA biosynthesis</b>	6	13	3.12E-03
	<b>Glycolysis/Gluconeogenesis</b>	6	12	4.06E-03
	<b>Pyruvate metabolism</b>	6	10	4.72E-03
	Citrate cycle (TCA cycle)	5	11	1.46E-02
	Glutathione metabolism	4	16	2.55E-02
	Fatty acid biosynthesis	4	15	2.97E-02
	Arginine and proline metabolism	5	8	4.82E-02
YPPF	<b>Ribosome</b>	12	9	3.44E-06
	<b>Pyruvate metabolism</b>	9	15	3.04E-06
	<b>Purine metabolism</b>	9	7	9.27E-04
	<b>Aminoacyl-tRNA biosynthesis</b>	6	13	1.79E-03
	<b>Glycolysis/Gluconeogenesis</b>	6	12	2.58E-03
	<b>Pyrimidine metabolism</b>	7	8	2.91E-03
	Alanine, aspartate and glutamate metabolism	5	11	1.40E-02
	Glyoxylate and dicarboxylate metabolism	4	13	3.16E-02
YPTS	<b>Ribosome</b>	13	13	6.44E-09
	<b>Purine metabolism</b>	9	9	1.15E-04
	<b>Pyrimidine metabolism</b>	7	11	7.38E-04
	Peptidoglycan biosynthesis	5	18	1.91E-03
	<b>Aminoacyl-tRNA biosynthesis</b>	5	14	4.85E-03

Bolded text indicates biological process conserved across at least two of the three Yersinia strains. YPCO, *Yersinia pestis* CO92; YPTS, *Yersinia pseudotuberculosis* PB1/+; YPPF, *Yersinia pestis* Pestoides F.