



Yersiniomics, a Multi-Omics Interactive Database for Yersinia Species

Pierre Lê-Bury, Karen Druart, Cyril Savin, Pierre Lechat, Guillem Mas Fiol, Mariette Matondo, Christophe Bécavin, Olivier Dussurget, Javier Pizarro-Cerdá

► To cite this version:

Pierre Lê-Bury, Karen Druart, Cyril Savin, Pierre Lechat, Guillem Mas Fiol, et al.. Yersiniomics, a Multi-Omics Interactive Database for Yersinia Species. Microbiology Spectrum, 2023, 11 (2), 10.1128/spectrum.03826-22 . pasteur-04102830

HAL Id: pasteur-04102830

<https://pasteur.hal.science/pasteur-04102830>

Submitted on 22 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Yersiniomics, a Multi-Omics Interactive Database for *Yersinia* Species

 Pierre Lê-Bury,^a Karen Druart,^b  Cyril Savin,^{a,e} Pierre Lechat,^c Guillem Mas Fiol,^a Mariette Matondo,^b Christophe Bécavin,^d  Olivier Dussurget,^a Javier Pizarro-Cerdá^{a,e}

^aInstitut Pasteur, Université Paris Cité, CNRS UMR6047, *Yersinia* Research Unit, Paris, France

^bInstitut Pasteur, Université Paris Cité, CNRS USR2000, Mass Spectrometry for Biology Unit, Proteomic Platform, Paris, France

^cInstitut Pasteur, Université Paris Cité, ALPS, Bioinformatic Hub, Paris, France

^dUniversité Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France

^eInstitut Pasteur, Université Paris Cité, *Yersinia* National Reference Laboratory, WHO Collaborating Research & Reference Centre for Plague FRA-140, Paris, France

ABSTRACT The genus *Yersinia* includes a large variety of nonpathogenic and life-threatening pathogenic bacteria, which cause a broad spectrum of diseases in humans and animals, such as plague, enteritis, Far East scarlet-like fever (FESLF), and enteric redmouth disease. Like most clinically relevant microorganisms, *Yersinia* spp. are currently subjected to intense multi-omics investigations whose numbers have increased extensively in recent years, generating massive amounts of data useful for diagnostic and therapeutic developments. The lack of a simple and centralized way to exploit these data led us to design Yersiniomics, a web-based platform allowing straightforward analysis of *Yersinia* omics data. Yersiniomics contains a curated multi-omics database at its core, gathering 200 genomic, 317 transcriptomic, and 62 proteomic data sets for *Yersinia* species. It integrates genomic, transcriptomic, and proteomic browsers, a genome viewer, and a heatmap viewer to navigate within genomes and experimental conditions. For streamlined access to structural and functional properties, it directly links each gene to GenBank, the Kyoto Encyclopedia of Genes and Genomes (KEGG), UniProt, InterPro, IntAct, and the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) and each experiment to Gene Expression Omnibus (GEO), the European Nucleotide Archive (ENA), or the Proteomics Identifications Database (PRIDE). Yersiniomics provides a powerful tool for microbiologists to assist with investigations ranging from specific gene studies to systems biology studies.

IMPORTANCE The expanding genus *Yersinia* is composed of multiple nonpathogenic species and a few pathogenic species, including the deadly etiologic agent of plague, *Yersinia pestis*. In 2 decades, the number of genomic, transcriptomic, and proteomic studies on *Yersinia* grew massively, delivering a wealth of data. We developed Yersiniomics, an interactive web-based platform, to centralize and analyze omics data sets on *Yersinia* species. The platform allows user-friendly navigation between genomic data, expression data, and experimental conditions. Yersiniomics will be a valuable tool to microbiologists.

KEYWORDS *Yersinia*, genome, transcriptome, proteome, database, RNA-Seq, mass spectrometry, microarray, multi-omics, synteny

The genus *Yersinia* comprises 26 Gram-negative bacterial species which belonged to the family *Enterobacteriaceae* until 2016 and which are now part of the new family *Yersiniaceae* (1). Although this genus includes mostly nonpathogenic environmental species, several important animal and human pathogens are also present in the group. *Yersinia enterocolitica* and *Yersinia pseudotuberculosis* are phylogenetically distant *Yersinia* species (2); however, the parallel acquisition of a diverse set of virulence factors (invasins, siderophores, and a type III secretion system) has endowed both species with the capacity

Editor Tino Polen, Forschungszentrum Jülich GmbH

Copyright © 2023 Lê-Bury et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Javier Pizarro-Cerdá, javier.pizarro-cerda@pasteur.fr, or Olivier Dussurget, olivier.dussurget@pasteur.fr.

The authors declare no conflict of interest.

Received 17 October 2022

Accepted 26 January 2023

Published 27 February 2023

to invade the gastrointestinal tract of mammals and to cause enteritis, following an oro-fecal infectious cycle (3). Enteric yersiniosis is the third most reported bacterial foodborne zoonosis in Europe (4). In the United States, *Y. enterocolitica* is classified as a priority pathogen by the National Institutes of Health (NIH). In Africa, recent evidence indicates that *Y. enterocolitica* causes human digestive disorders with a frequency similar to that reported in other continents (5). *Yersinia pestis*, on the other hand, is a clone that recently emerged from *Y. pseudotuberculosis* (6) and acquired the capacity to infect fleas and to cause plague in humans through acquisition of novel virulence factors and massive gene inactivation (7). Plague is still endemic in the Americas, Africa, and Asia (8), and the major pneumonic plague outbreak in Madagascar in 2017 is a reminder that *Y. pestis* is a severe threat to human populations (9, 10). Two other animal pathogens, *Yersinia ruckeri* and *Yersinia entomophaga*, have the capacity to cause disease in fishes and insects, respectively (11, 12). *Y. ruckeri* is responsible for enteric redmouth disease, one of the most important disease of salmonids, which leads to significant economic losses (11). *Y. entomophaga* has a commercial interest for pest management, as it can infect and kill a wide range of insects (13).

Pathogenic *Yersinia* spp. have been instrumental models to understand the evolution and mechanisms of pathogenicity in the bacterial world. The invasin of *Y. pseudotuberculosis* was the first bacterial factor reported to promote bacterial internalization within mammalian nonphagocytic epithelial cells (14, 15). The subsequent identification of $\beta 1$ integrins as receptors for invasion (16) set the general basis to understand how bacterial surface effectors subvert mammalian cellular functions (phosphoinositide metabolism, Rho GTPase signaling, and actin polymerization) to invade host cells and tissues (17–20). The *Yersinia* type III secretion system was one of the first to be thoroughly characterized, making it possible to decipher the exquisite manipulation of phagocytic and immune functions by a bacterial pathogen through injection of bacterial effectors within the cytoplasm of host neutrophils, macrophages, and dendritic cells (21, 22).

Like several other important bacterial pathogens, members of the genus *Yersinia* have been investigated using omics approaches. In the last 2 decades, the ever-growing pace of technological innovation led to generation of a massive amount of data produced by omics methods, such as whole-genome sequencing combining short (such as Illumina sequencing) and long (such as PacBio or Nanopore sequencing) reads, DNA hybridization array (macro- and microarrays) for gene expression analysis, RNA sequencing (RNA-Seq), and semiquantitative mass spectrometry (Fourier-transform ion cyclotron resonance [FTICR-MS] and liquid chromatography-tandem mass spectrometry [LC-MS/MS]). Data type-specific databases allowing the deposition of these data have been created, e.g., GenBank (23) for genomes, Gene Expression Omnibus (24) and ArrayExpress (25) for microarrays, European Nucleotide Archive (26) and Sequence Read Archive (27) for RNA-Seq, and ProteomeXchange (28) for mass spectrometry. However, integration of these data in a genus- or species-dependent manner is currently restricted to a few model organisms (29–31), and pathogenic-microorganism-specific integrated data found in databases like PATRIC (32) have not been updated with recent experiments.

Here, we present a unique curated multi-omics database gathering 200 genomic, 317 transcriptomic, and 62 proteomic data sets originating from *Yersinia* spp. since the beginning of the omics revolution. This database was constructed using the Bacnet platform (33); we contributed to improvements of this platform. Indeed, with omics technologies evolving rapidly, new data formats, such as LC-MS/MS shotgun, which was absent from the first Bacnet-based website, were implemented (31). For several reference genomes, we implemented integrated views at the gene level of the Kyoto Encyclopedia of Genes and Genomes (KEGG) for biochemical pathways (34), UniProt for protein information (35), InterPro for protein domain signature (36), IntAct for molecular interactions (37) and the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) for protein interactions (38). Raw reads from 425 RNA-Seq runs were consistently processed and analyzed with our bioinformatic pipeline, and a quality control and differential analysis report is available for each experiment which

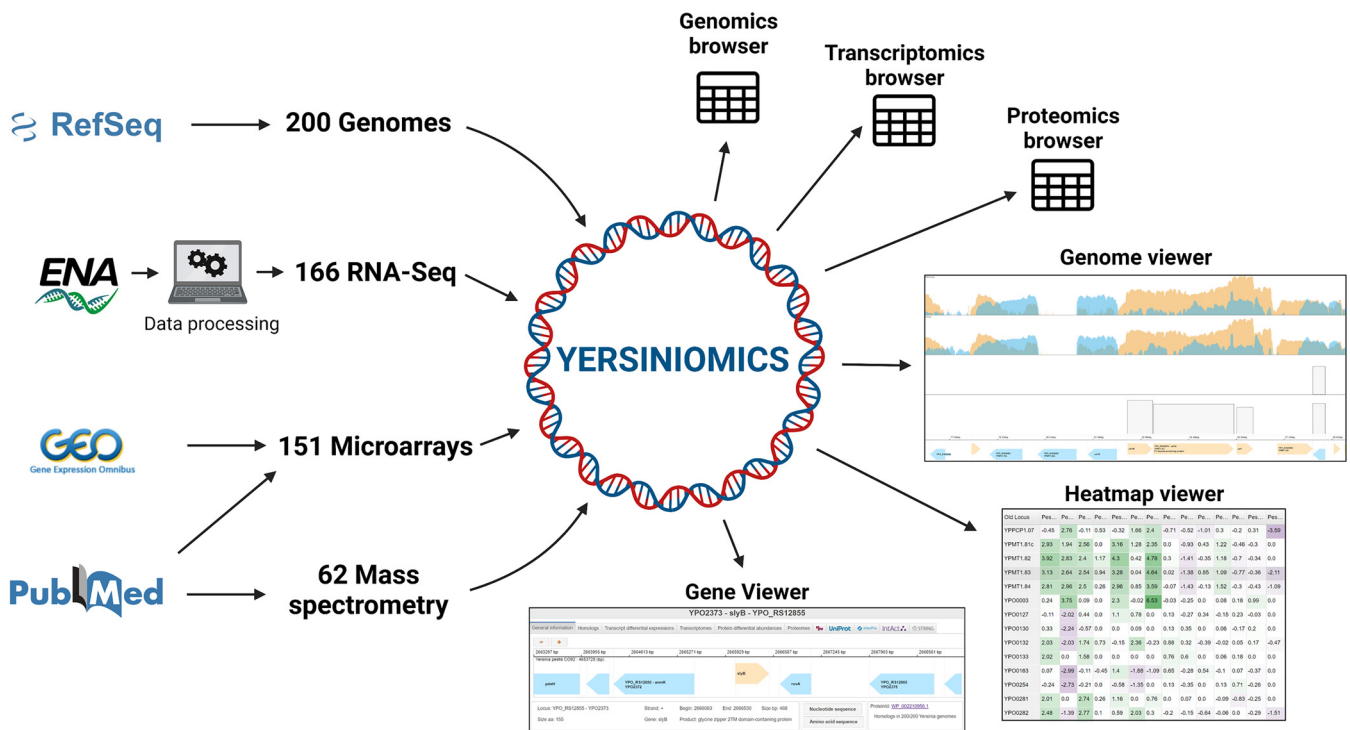


FIG 1 Yersiniomics database construction pipeline and tools. Genomes were collected on the RefSeq database (GenBank). RNA-Seq raw data (.fastq files) were collected on the ENA browser and then processed as read counts and fold change. Fold changes from microarray experiments were collected on GEO or from tables and supplemental tables in publications. Fold changes from proteomic experiments were collected from tables and supplemental tables in publications. Data sets were integrated into Yersiniomics and can be consulted via three omics browsers (genomics, transcriptomics, and proteomics) and three viewers (gene, genome, and heatmap).

encompassed replicates. Processed omics experiments are easily browsable, linked to their Gene Expression Omnibus (GEO) (24), European Nucleotide Archive (ENA) (26) or Proteomics Identifications Database (PRIDE) (39) repositories, and directly viewable at the gene level or according to experimental conditions on the dedicated Yersiniomics website that we have implemented (<https://yersiniomics.pasteur.fr/>).

RESULTS

From public databases, we collected genomic sequences, raw RNA-Seq data, microarray data, and mass spectrometry data that we processed and integrated into the Yersiniomics website (<https://yersiniomics.pasteur.fr/>). Yersiniomics relies on 6 principal tools (Fig. 1): (i) three omics browsers (“genomics,” “transcriptomics,” and “proteomics”) in a table format allowing navigation among the different genomes and biological conditions of the experiments implemented in the database; (ii) two omics data set viewers, called “genome viewer” and “heatmap viewer,” allowing navigation among transcriptomics and proteomics results; and (iii) a gene viewer allowing navigation among the genes of a specific genome, to quickly access associated entries in external databases and to browse associated omics data.

Database functionalities. (i) The omics data set browsers. From the Yersiniomics home page, three omics browsers are available: genomics, transcriptomics, and proteomics (Fig. 2, middle panel).

From the genomics browser, 200 complete *Yersinia* genomes are sorted according to their phylogenetic relatedness, based on the 500 genes of our recently proposed *Yersinia* core genome multilocus sequence typing (cgMLST) scheme (40) (Fig. 3). Displayed information related to genomes includes species name, the most recent assignment proposed by the French *Yersinia* National Reference Center and determined via our cgMLST, the number of chromosomes and plasmids, the chromosome and plasmid total size, the lineage and sublineage (when applicable) based on single nucleotide polymorphisms (SNP) analysis for

Home



Yersiniomics integrates complete **genomes**, **transcriptomes** and **proteomes** published for *Yersinia* species.

Access **enriched information** about *Yersinia* species genes in complete genomes:
Annotation, gene conservation, synteny, transcript atlas, protein atlas, integration of external databases.

Use Yersiniomics to decipher **regulatory mechanisms** of your genome element of interest,
navigating among all these datasets with **enriched metadata** in a user-friendly format.

Gene viewersⁱ

<i>Y. pestis</i>			<i>Y. pseudotuberculosis</i>		<i>Y. enterocolitica</i>		<i>Y. ruckeri</i>	<i>Y. entomophaga</i>
CO92	KIM	91001	YPIII	IP32953	8081	WA	SC09	
Pestoides F	EV76-CN		IP31758		Y1	Y11	QMA0440	MH96

Omics browsersⁱ

Genomics browser
Transcriptomics browser
Proteomics browser

Data loadingⁱ

Load genome viewer
Download processed data

Last update: January 2023

For more information on the website functionalities, please go to [Lé-Bury et al.](#)

If you use Yersiniomics, please cite our article

Contact us if you have a recently published "omics" dataset you want to be integrated to Yersiniomics:
yersiniomics@pasteur.fr

Credits

Yersinia Research Unit, Institut Pasteur, Paris






FIG 2 Yersiniomics home page. In the top panel, shortcuts to strain gene viewers of *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica*, *Y. ruckeri*, and *Y. entomophaga* are implemented. In the middle panel, three buttons lead to the genomics, transcriptomics, and proteomics browsers. The bottom panel allows the user to load previously saved genome viewers and to download omics processed data.

Y. pestis (41, 42) and on our cgMLST (40) for the other *Yersinia* species, the biotype and serotype (when relevant), the isolation source, year and country, the number of genes, rRNAs, and tRNAs, and the assembly ID, linked to GenBank, and its FTP link. A specific genome can be opened in the gene viewer (see below) by double clicking on it, and several genomes can be selected to export a summary table. Genomes are highlighted in blue in the table when searched with the search box, and genomes selected in the table are highlighted in red on the phylogenetic tree displayed on the left, which can be exported in SVG format (Fig. 3).

The transcriptomics and proteomics browsers work in a similar way and summarize the biological conditions of transcriptomic and proteomic *Yersinia* experiments (Fig. 4). Biological condition names follow a nomenclature encompassing the main important information about each experiment: species name, strain name, whether it addresses a wild-type (WT) or a mutant strain (when applicable), culture temperature, culture medium/phase/time point (when applicable), other information relevant to comparisons (when applicable), the technology used, and the year of data deposition or publication. For example, "Pestis_KIM6+_WT_37C_HIB_pH6_NextSeq500_2021" refers to an

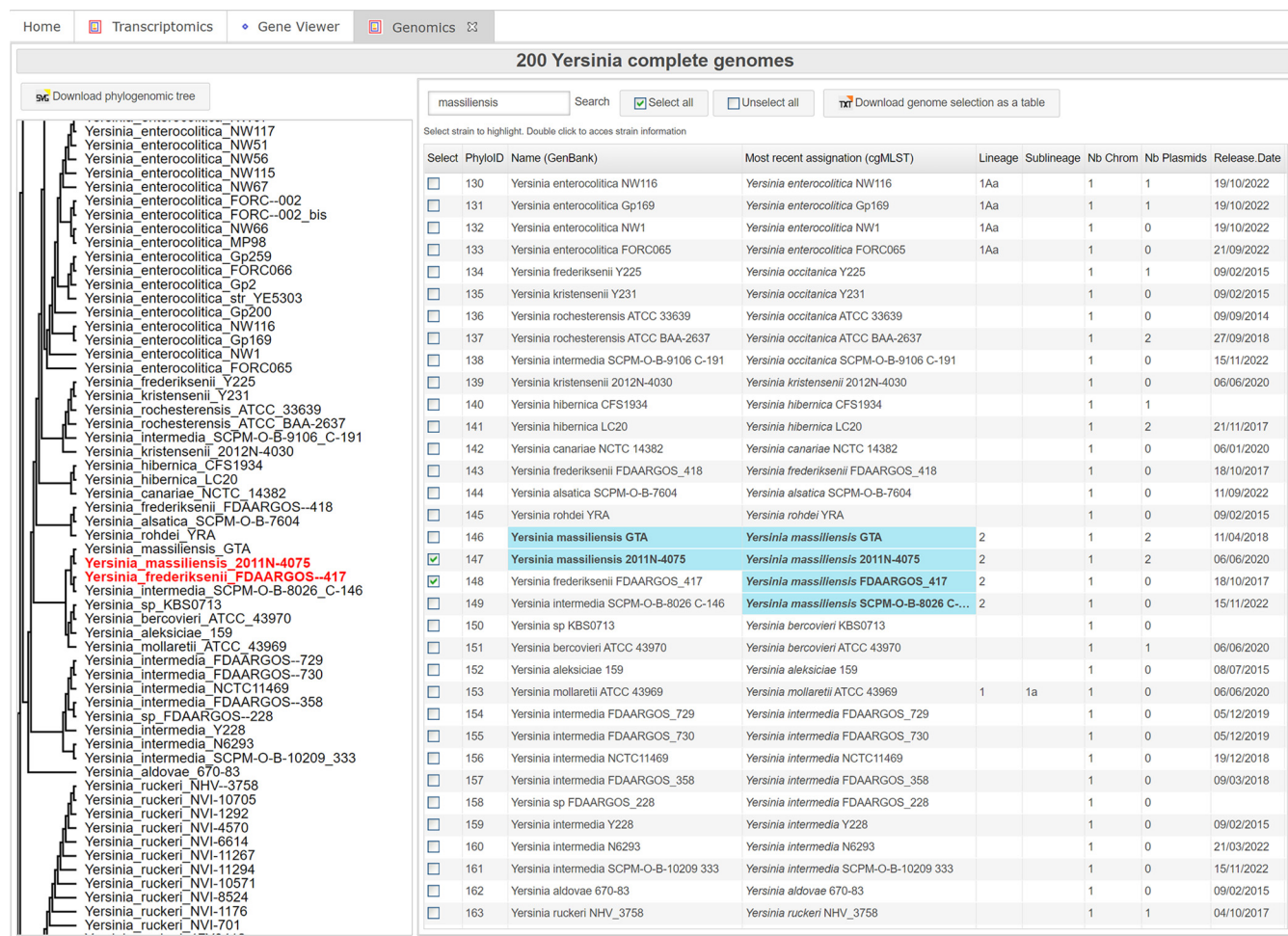


FIG 3 The genomics browser. A *Yersinia* phylogenetic tree is displayed on the left panel, and the strains information (name, lineage, isolate source, number of genes, etc.) is available on the right panel as a searchable table. Selected strains are automatically highlighted on the phylogenetic tree. An updated tree can be downloaded in SVG format.

RNA-Seq experiment using the Illumina NextSeq500 platform deposited in 2021 and measuring RNA levels of a *Y. pestis* KIM6+ strain grown at 37°C in heart infusion broth (HIB) at pH 6. “Pseudotuberculosis_YPIII_Mutant_rovA_25C_Log_Agilent_2014” refers to an experiment performed with an Agilent microarray in 2014 using a *Y. pseudotuberculosis* YPIII strain mutated at the *rovA* locus and grown at 25°C to logarithmic phase. Complementary information, such as details on the strain microarray, the genome on which the RNA-Seq experiment was mapped, the DESeq2 (43) differential analysis report for RNA-Seq data, the publication reference with a link to PubMed, and the link to the GEO (24), ENA (26), or PRIDE (39) entry, can be found in the table columns. A search bar can be used to explore the biological conditions of each experiment, and some filters are accessible, such as the reference genome used, the data type (“Gene Expression” for microarray experiments or “RNA-Seq”), as well as the selection of a specific mutant or growth phase. The proteomics browser also implements a “protein localization” column (i.e., supernatant, whole cell, or soluble or insoluble fraction), as well as the number of proteins detected in each experimental condition.

After selecting one or several biological conditions, their associated transcriptome or proteome values can be visualized in two ways: as a heatmap viewer or as a genome viewer. The heatmap viewer is a sortable table with each line corresponding to a gene in the genome and columns displaying expression comparisons of two biological conditions using a color code proportional to the fold change, facilitating tracking of highly differentially expressed genes. A $\log_2$ (fold change) cutoff, set to 1.5 by default,

Home **Transcriptomics**

Yersinia Transcriptomics Datasets

Search

Reference strain

All

Data type

☐ Array (151)

☐ RNA-Seq (166)

Mutant (Gene and sRNA)

☒ All

☐ None

☐ Choose One

DjIA+

Temperature

☐ 25°C to 28°C (147)

☐ 37°C (133)

☐ Other (44)

Growth Medium

☐ BCS (0)

☐ LB (122)

☐ Chemically Defined Media (54)

☐ DMEM/RPMI (12)

☐ *In cellulo* (10)

☐ Blood (5)

☐ *In vivo* (20)

☐ Other Media (97)

Growth Phase

☐ Logarithmic (118)

☐ Stationary (33)

Select biological conditions and:

☒ Select all ☐ Unselect all Download transcriptome selection as a table

Visualize their transcriptomics datasets with the Genome Viewer

Visualize differentially expressed genes and non-coding RNAs with the HeatMap Viewer

Select	Data Name	Type	Date	Growth	TimePoint	Temp.	Mutant	Media
☐	Enterocolitica_647176_Mutant_OAntigen_22C_Log_HiSeq2000_2015		04/03/2015	Logarithmic Phase		22	O-antigen-	LB
☐	Enterocolitica_647176_Mutant_hfq_22C_Log_HiSeq2000_2015		04/03/2015	Logarithmic Phase		22	hfq-	LB
☐	Enterocolitica_647176_Mutant_hfq_37C_Log_HiSeq2000_2015		04/03/2015	Logarithmic Phase		37	hfq-	LB
☐	Enterocolitica_647176_Mutant_rfaH_22C_Log_HiSeq2000_2015		04/03/2015	Logarithmic Phase		22	rfaH-	LB
☐	Enterocolitica_647176_Mutant_rfaH_37C_Log_HiSeq2000_2015		04/03/2015	Logarithmic Phase		37	rfaH-	LB
☐	Enterocolitica_647176_Mutant_ybeY_22C_Log_HiSeq2000_2014		23/10/2014	Logarithmic Phase		22	ybeY-	LB
☐	Enterocolitica_647176_Mutant_ybeY_37C_Log_HiSeq2000_2014		23/10/2014	Logarithmic Phase		37	ybeY-	LB
☐	Enterocolitica_647176_WT_22C_Log_HiSeq2000_2014		23/10/2014	Logarithmic Phase		22		LB
☐	Enterocolitica_647176_WT_37C_Log_HiSeq2000_2014		23/10/2014	Logarithmic Phase		37		LB
☐	Enterocolitica_647176_WT_37C_Log_HiSeq2000_2015		04/03/2015	Logarithmic Phase		37		LB
☐	Enterocolitica_8081_Mutant_yenR_26-37C_5h_SGUL_2008		17/11/2008		5h	26/37	yenR-	BHI-MOX
☐	Enterocolitica_8081_Mutant_yenR_26C_8h_SGUL_2007		06/09/2007		8h	26	yenR-	LB
☐	Enterocolitica_8081_Mutant_yenR_OverExpr_ytxR_26-37C_5h_SGUL_2008		17/11/2008		5h	26/37	yenR- ytxR+	BHI-MOX
☐	Enterocolitica_8081_Mutant_yenR_rovA_26C_8h_SGUL_2007		06/09/2007		8h	26	yenR- rovA-	LB
☐	Enterocolitica_8081_Mutant_yenR_ytxR_26C_SGUL_2008		17/11/2008			26	yenR- ytxR-	LB
☐	Enterocolitica_8081_Mutant_yenR_ytxR_OverExpr_ytxR_26C_SGUL_2008		17/11/2008			26	yenR- ytxR- ytxR+	LB
☐	Enterocolitica_8081_WT_25C_Log_LB_HiSeq2500_2019		17/03/2019	Logarithmic Phase		25		LB
☐	Enterocolitica_8081_WT_25C_Stat_LB_HiSeq2500_2019		17/03/2019	Stationary Phase		25		LB
☐	Enterocolitica_8081_WT_37C_Log_LB_HiSeq2500_2019		17/03/2019	Logarithmic Phase		37		LB
☐	Enterocolitica_8081_WT_37C_Stat_LB_HiSeq2500_2019		17/03/2019	Stationary Phase		37		LB
☐	Enterocolitica_8081_genome_Y1_WT_25C_Log_LB_HiSeq2500_2019		17/03/2019	Logarithmic Phase		25		LB
☐	Enterocolitica_8081_genome_Y1_WT_25C_Stat_LB_HiSeq2500_2019		17/03/2019	Stationary Phase		25		LB
☐	Enterocolitica_8081_genome_Y1_WT_37C_Log_LB_HiSeq2500_2019		17/03/2019	Logarithmic Phase		37		LB
☐	Enterocolitica_8081_genome_Y1_WT_37C_Stat_LB_HiSeq2500_2019		17/03/2019	Stationary Phase		37		LB
☐	Enterocolitica_ATCC23715_WT_26C_M63_Log_HiSeq2000_2020		03/06/2020	Logarithmic Phase		26	pYV-	M63
☐	Enterocolitica_ATCC23715_mutant_rcsB_26C_M63_Log_HiSeq2000_2020		03/06/2020	Logarithmic Phase		26	pYV- rcsB-	M63
☐	Enterocolitica_JB580v_WT_26C_120min_TYE_HiSeq2500_2015		20/04/2015		120min	26	pYV-	TYE
☐	Enterocolitica_JB580v_WT_26C_240min_TYE_HiSeq2500_2015		20/04/2015		240min	26	pYV-	TYE
☐	Enterocolitica_JB580v_WT_26C_30min_TYE_HiSeq2500_2015		20/04/2015		30min	26	pYV-	TYE
☐	Enterocolitica_JB580v_WT_26C_60min_TYE_HiSeq2500_2015		20/04/2015		60min	26	pYV-	TYE
☐	Enterocolitica_JB580v_WT_26C_Log_TYE_HiSeq2500_2015		20/04/2015	Log		26	pYV-	TYE

Gene expression array RNA-Seq and Ribo-Seq

FIG 4 The transcriptomics browser. Transcriptomics data sets are browsable in the main panel on the right and can be filtered through the search bar or with the preset filters on the left. Information on biological conditions, such as culture temperature, medium, growth stage, strain used, mutation, associated publication, and DESeq2 differential analysis report for RNA-Seq experiments, is accessible in the different columns of the table. After one or several lines are selected, the associated biological conditions can be viewed in the genome viewer or in the heatmap viewer by clicking on the buttons at the top.

can be set to any other value, or to 0 to display the whole genome. This view allows comparison of different results such as *Y. pestis* RNA abundance in infected C57BL/6 mouse lungs versus growth in brain heart infusion broth (BHI) (44), *Y. pestis* RNA abundance after growth in human plasma versus growth in lysogeny broth (LB) (45), and *Y. pestis* RNA abundance in infected brown Norway rat bubos versus growth in LB (46) (Fig. 5A).

The genome viewer (Fig. 5B and C) displays a value on the y axis (the number of mapped reads or the fold change in RNA or protein, for example) and corresponding genes on the x axis, allowing navigation of the genome. Depending on the data type, different information is plotted. For microarray experiments where only biological condition comparisons are available, fold change between two conditions is plotted as “relative expression data.” For RNA-Seq experiments, stranded or unstranded read coverage is plotted as “absolute expression data” (Fig. 5C) and fold changes can be plotted as “relative expression data.” For proteomics data, both relative and absolute (intensity; Fig. 5C) expression data can be plotted if available. A multi-omics view of different comparisons can be plotted this way, such as infected brown Norway rat bubos versus growth in LB (46), *Y. pestis* RNA abundance in infected C57BL/6 mouse lungs versus growth in BHI (44), and infected human plasma versus growth in LB (45) (Fig. 5B). A

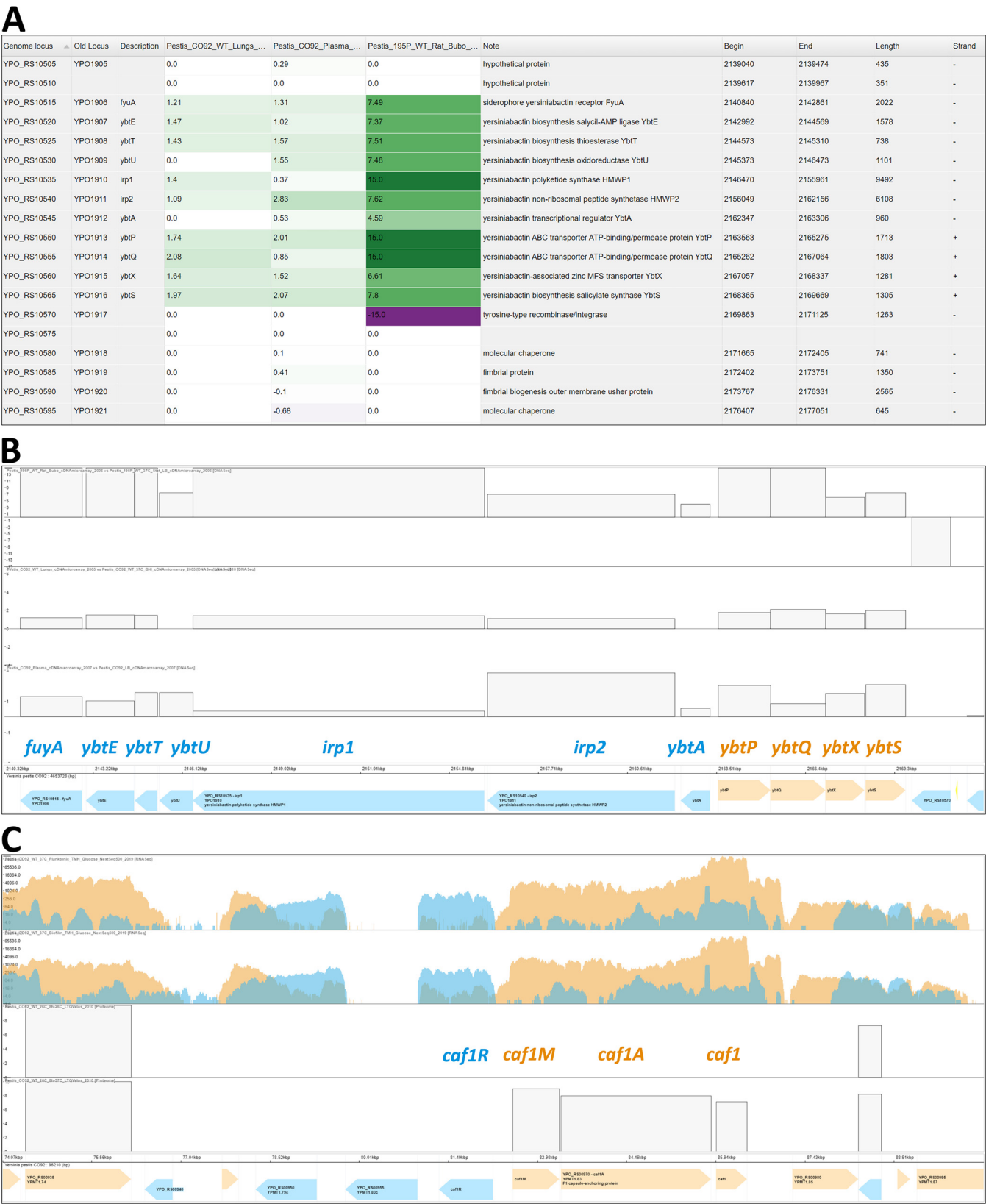


FIG 5 (A) Heatmap viewer. Heatmap view of the yersiniabactin siderophore *ybt* operon in the CO92 strain genome, displaying three *in vivo* and *ex vivo* experiments comparing (from left to right): *Y. pestis* RNA abundance in infected C57BL/6 mouse lungs versus growth in BHI (44), *Y. pestis* RNA abundance after growth in human plasma versus growth in LB (45), and *Y. pestis* RNA abundance in infected brown Norway rat bubos versus growth in LB (46). (B) Genome viewer. Multi-omics view showing the relative expression of the yersiniabactin siderophore *ybt* operon in the CO92 genome for the following (Continued on next page)

multi-omics view of absolute expression can also be plotted on this viewer, such as the two RNA-Seq stranded coverages, 37°C growth in planktonic and biofilm states in TMH medium supplemented with glucose (47), and LC-MS/MS data showing protein abundance after 8 h growth at 26°C or 37°C (48) (Fig. 5C). A specific view can be saved and downloaded in a .gview file with the “Save data selection” button. The .gview file can later be uploaded in a new session with the “Load data selection button” or from the home page.

For the 26 RNA-Seq experiments which included replicated samples, a quality control and differential analysis report is accessible from the transcriptomic browser in the “DESeq2 report” column. This column links to an interactive web page where different charts and tables make it possible to assess the overall quality of the experiment and bioinformatic pipeline. For example, the 36 *Y. pseudotuberculosis* YPIII samples generated by Avican et al. to create an RNA atlas of human pathogens exposed to different stresses can be examined (49). Mapped read counts are heterogeneous between samples, but more than 2.5 million reads could be mapped for all runs except for the run [SRR11998801](#) (Fig. 6A). Based on the feature counts, hierarchical clustering and interactive three-dimensional (3D) principal-component analysis (Fig. 6B) showed a good clustering of sample replicates. Run [SRR11998801](#) differed from the two other replicates of the same condition. Interactive volcano plots allow rapid exploration of the up- and downregulated genes in every computed comparison within the same experiment, such as the bile acid stress condition versus untreated condition YPIII (Fig. 6C). Sortable tables, which can also be filtered by locus name or gene name, allow further exploration of regulations, and the processed data can be downloaded in CSV format. We recommend using Mozilla Firefox to benefit from all the functionalities implemented in the RNA-Seq reports.

(ii) The gene viewer. From the home page, a quick access to the gene viewer for reference strains is implemented (Fig. 2, top). In the gene viewer, each gene is accessible via a gene list or a graphical view after a specific *Yersinia* genome is selected (Fig. 7). A search bar makes it possible to quickly filter the list by gene name, gene locus, or any field present in the “General information” panel as described below.

For a specific open reading frame (ORF), information such as gene locus, name, product, position, strand, and size (in base pairs and amino acids) can be found under the “General information” panel. A link to the GenBank protein and the number of genomes presenting a homolog in the Yersiniomics database is also present, in addition to the graphical view presenting the genetic environment of the gene. Other general information extracted from the “features” field of the GenBank GFF file is also displayed. A synteny based on the SynTView software (50) was also implemented for some reference genomes: *Y. pestis* CO92 and KIM, *Y. pseudotuberculosis* IP32953 and YPIII, and *Y. enterocolitica* Y11 and 8081 (Table 1). As synteny data are heavy to download and the SynTView software requires a significant amount of memory and computational power, loading is optional and conditioned to the “Show synteny” button in the “General information” panel. After the loading of SynTView, the software displays the beginning of the genome. When clicking on or searching a gene in the Yersiniomics gene viewer panel, SynTView automatically centers on the corresponding gene. The software allows the user to browse the genetic environment and conservation of each gene among other closely related genomes present in Yersiniomics. SynTView allows the user to explore syntenic organization and dynamically zoom in on a position by genomic location or go directly to specific genes by name. The view is linked with dynamic interactions with other specialized views (circular view and dot plot).

FIG 5 Legend (Continued)

conditions (from top to bottom): infected brown Norway rat bubos versus growth in LB (46), *Y. pestis* RNA abundance in infected C57BL/6 mice lungs versus growth in BHI (44) and infected human plasma versus growth in LB (45). Induction of the yersiniabactin operon is observed both *in vivo* and *ex vivo*. (C) Genome viewer. Multi-omics view showing the absolute expression of the pseudocapsule *caf* operon in the CO92 pMT1 plasmid for the following conditions (from top to bottom): two RNA-Seq stranded coverages, growth at 37°C in planktonic and biofilm states in TMH medium supplemented with glucose (47), and LC-MS/MS data showing protein abundance after 8 h growth at 26°C or 37°C (48) where the *caf* operon is induced.

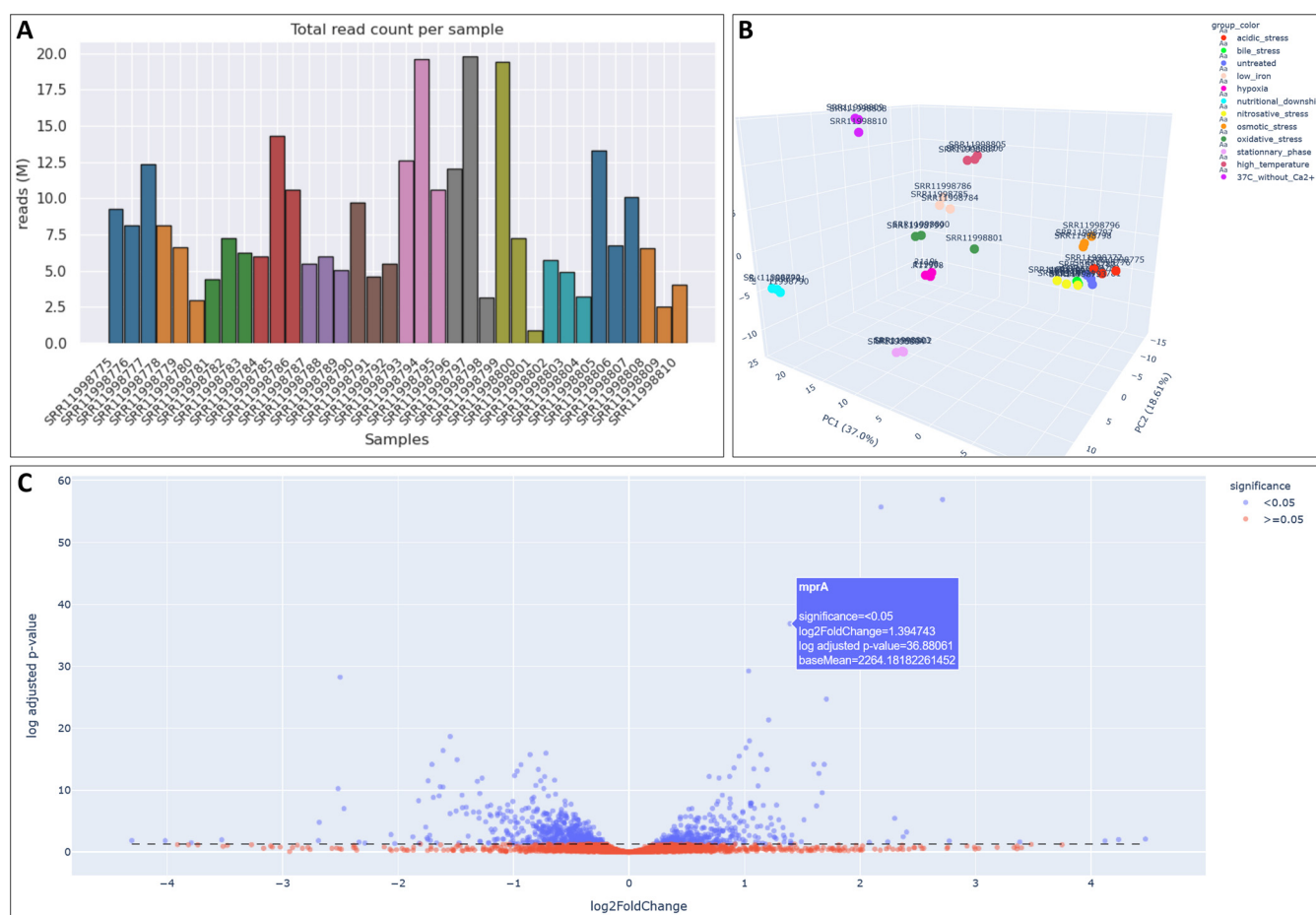


FIG 6 DESeq2 differential analysis report. (A) Mapped read counts of the 36 *Y. pseudotuberculosis* YPIII samples exposed to different stresses generated by Avican et al. (49). (B) Interactive 3D principal-component analysis based on mapped read counts showing good clustering between replicates among the 36 samples exposed to different stresses. (C) Interactive volcano plot of the bile acid stress condition versus untreated condition for the YPIII strain, allowing the user to rapidly explore the up- and downregulated genes.

The “Homologs” panel represents the 200 genomes present in the Yersiniomics database next to the phylogenetic tree (Fig. 8). For each genome, the homologous locus, old locus, and linked GenBank protein ID are displayed, next to the BLASTP results, such as the percentage of coverage and percentage of similarity on the covered region, or the BLAST E value and bit score. The “Bidirectional” column indicates if the best BLASTP hit of the homologous gene is reciprocal. Genomes are highlighted in blue in the table when searched with the search box, and genomes selected in the table are highlighted in red on the phylogenetic tree displayed on the left. The product of the percentage of coverage times the percentage of similarity is displayed alongside the gene locus in the phylogenetic tree, which can be exported in SVG format. The corresponding gene can be opened in its own gene viewer by double-clicking on its line in the homolog table, facilitating navigation between *Yersinia* homologs and access to their respective omics data.

For genomes presenting transcriptomic data sets such as microarray or RNA-Seq data, the panel “Transcript differential expressions” displays the fold change of the gene in each available transcriptome comparison (Fig. 9). These fold changes can be filtered with a cutoff to select over- or underexpressed genes over a certain threshold. A fold change of +15 or −15 corresponds to the absence of the transcript under one of the conditions. A fold change of exactly 0.0 indicates the absence of transcript under both conditions or that the information for this gene under these conditions could not be retrieved. For most of the RNA-Seq data, if replicates were measured, the *P* value

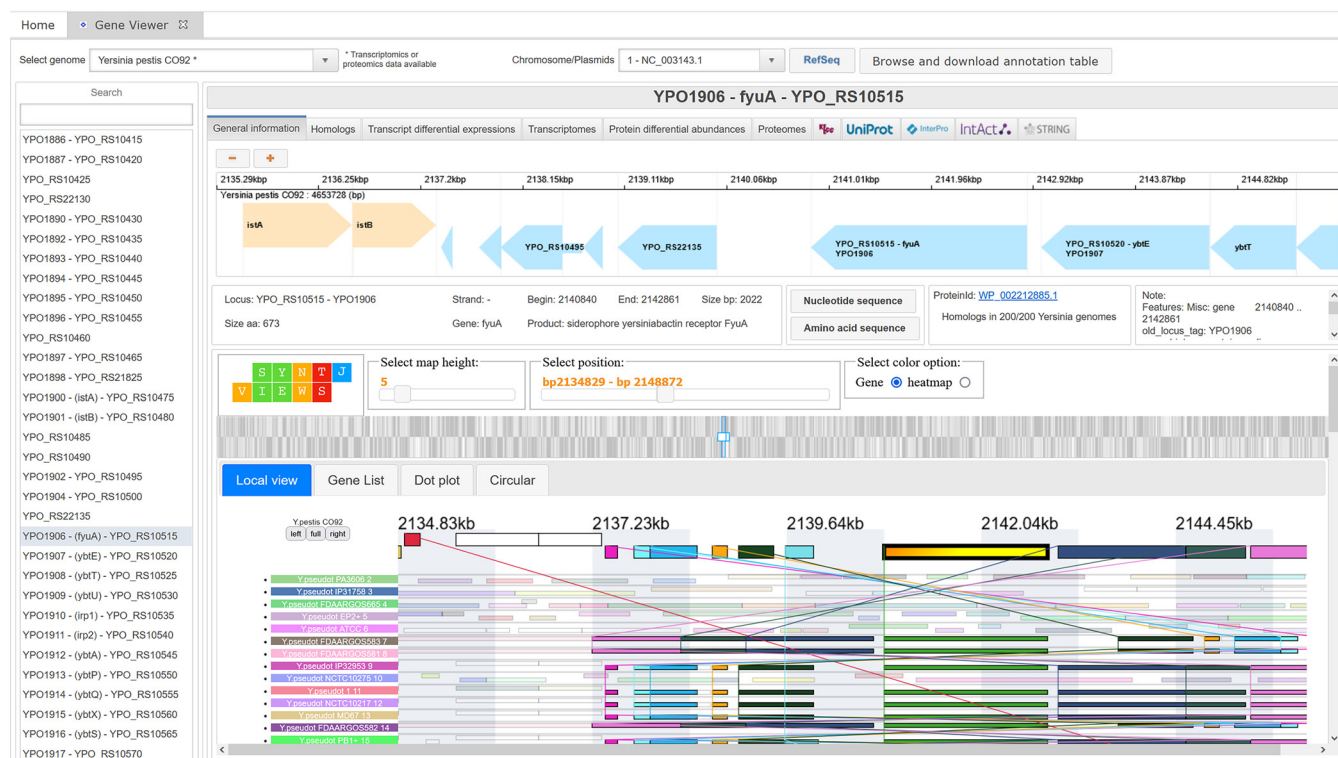


FIG 7 The gene viewer general information panel. Information about the *fyuA/psn* gene of *Yersinia pestis* CO92 (YPO1906) is displayed, such as the RefSeq new and old locus names, the length of the gene and protein, the gene position in the genome, and a graphical view of the surrounding genomic region. Nucleotide and amino acid sequences can be visualized and downloaded. A search bar to find other genes in the genome is available in the left column. A switch between the chromosome and the plasmids is available in the upper banner next to the “Chromosome/Plasmids” label. For several reference strains, synteny can be shown in the SynTV software. Selection of a gene on the Yersiniomics gene viewer automatically searches for the corresponding gene in SynTV. Genomes are aligned on the searched gene, which is shaded in yellow. Homologous genes are displayed with the same color. Genes of the reference genome can also be searched in the “Gene List” tab. Genome rearrangement between the reference and another genome can be visualized in the “Dot plot” tab. The “Circular” tab allows the user to map all the genomes on the reference genome on a circular plot.

and adjusted *P* value were computed and are displayed with the comparisons. The results can be filtered using a cutoff on the *P* value. For the microarray experiments and the RNA-Seq experiments without replicates, the *P* value and adjusted *P* value were set to 0.0 and the experiments cannot be filtered using the *P* value. Each comparison can be selected and displayed in two different ways, in the genome viewer or in the heatmap viewer, as detailed above. For RNA-Seq data, an additional panel called “Transcriptomes” displays the transcripts per million (TPM) normalized value (51) of the raw read counts mapped to the gene, allowing the user to assess whether the gene is highly transcribed.

For genomes presenting semiquantitative proteomic data, a similar panel called “Protein differential expressions” displays the fold change for the abundance of the gene-encoded protein in the available comparisons, very similarly to the “Transcript differential expressions” panel. The same customizable fold change cutoff is available, as well as the genome viewer and the heatmap viewer. A protein which is detected under one condition and not the other will also have a $\log_2(\text{fold change})$ set to +15 or −15. Experiments can also be filtered using the *P* value, and adjusted *P* values are shown next to the *P* value when present or set to 0.0 when they were not computed or not retrieved from publications. To assess for the presence of a protein in a semi-quantitative or nonquantitative proteomic experiment, the “Proteomes” panel displays the biological conditions in which the protein was detected, with its associated raw value (label-free quantitation [LFQ] or FTICR intensity) when available.

For several reference strains, dynamic access to KEGG, UniProt, InterPro, IntAct, and STRING was implemented at the gene level (Table 1). When a specific gene is selected, these tabs are automatically linked to the corresponding entries in each website. This

TABLE 1 Genomes with computed synteny and access to external databases^a

Strain	Synten	KEGG, UniProt, InterPro	STRING
<i>Y. aldovae</i> 670-83		Yes	Yes
<i>Y. aleksiciae</i> 159		Yes	
<i>Y. alsatica</i> SCPM-O-B-7604		Yes	
<i>Y. canariae</i> NCTC 14382		Yes	
<i>Y. enterocolitica</i> 1055Rr		Yes	
<i>Y. enterocolitica</i> 2516-87		Yes	
<i>Y. enterocolitica</i> 8081	Yes	Yes	Yes
<i>Y. enterocolitica</i> FORC_002		Yes	
<i>Y. enterocolitica</i> WA		Yes	
<i>Y. enterocolitica</i> Y11	Yes	Yes	
<i>Y. enterocolitica</i> YE53/03		Yes	
<i>Y. entomophaga</i> MH96			Yes
<i>Y. hibernica</i> CFS1934		Yes	
<i>Y. hibernica</i> LC20		Yes	Yes
<i>Y. intermedia</i> Y228		Yes	Yes
<i>Y. massiliensis</i> GTA		Yes	Yes
<i>Y. mollaretii</i> ATCC 43969		Yes	Yes
<i>Y. occitanica</i> ATCC 33639			Yes
<i>Y. occitanica</i> Y225		Yes	
<i>Y. occitanica</i> Y231		Yes	
<i>Y. pestis</i> 91001		Yes	
<i>Y. pestis</i> A1122		Yes	
<i>Y. pestis</i> Angola		Yes	
<i>Y. pestis</i> Antiqua		Yes	
<i>Y. pestis</i> CO92	Yes	Yes	Yes
<i>Y. pestis</i> D106004		Yes	
<i>Y. pestis</i> D182038		Yes	
<i>Y. pestis</i> El Dorado		Yes	
<i>Y. pestis</i> Harbin 35		Yes	
<i>Y. pestis</i> Harbin 35bis		Yes	
<i>Y. pestis</i> KIM10+		Yes	Yes
<i>Y. pestis</i> KIM5	Yes	Yes	Yes
<i>Y. pestis</i> Nepal516		Yes	
<i>Y. pestis</i> PBM19		Yes	
<i>Y. pestis</i> Pestoides F		Yes	
<i>Y. pestis</i> Shasta		Yes	
<i>Y. pestis</i> Z176003		Yes	
<i>Y. pseudotuberculosis</i> 1		Yes	
<i>Y. pseudotuberculosis</i> ATCC 6904		Yes	Yes
<i>Y. pseudotuberculosis</i> EP2+		Yes	
<i>Y. pseudotuberculosis</i> IP31758		Yes	
<i>Y. pseudotuberculosis</i> IP32953	Yes	Yes	
<i>Y. pseudotuberculosis</i> IP32953bis		Yes	
<i>Y. pseudotuberculosis</i> MD67		Yes	
<i>Y. pseudotuberculosis</i> PA3606		Yes	
<i>Y. pseudotuberculosis</i> PB1+		Yes	
<i>Y. pseudotuberculosis</i> YPIII	Yes	Yes	
<i>Y. rohdei</i> YRA		Yes	Yes
<i>Y. ruckeri</i> Big Creek 74		Yes	
<i>Y. ruckeri</i> SC09			Yes
<i>Y. ruckeri</i> YRB		Yes	
<i>Y. similis</i> 228		Yes	

^aSpecies nomenclature proposed by the French *Yersinia* National Reference Center.

allows very quick access to diverse information, such as the implication in biological pathways in KEGG (Fig. 10A), the AlphaFold prediction structure in UniProt (Fig. 10C), the protein domains and Pfam families in InterPro, protein-interacting partners identified with methods such as a yeast two-hybrid screen against human proteins (52) accessible in IntAct (Fig. 10D), or other metalinks such as co-occurrence in abstract computed in the STRING database (Fig. 10B). A button allows the user to directly open the corresponding website in a new tab of the web browser. Due to cookie policies, we recommend using Mozilla Firefox to access the STRING tab embedded in Yersiniomics.

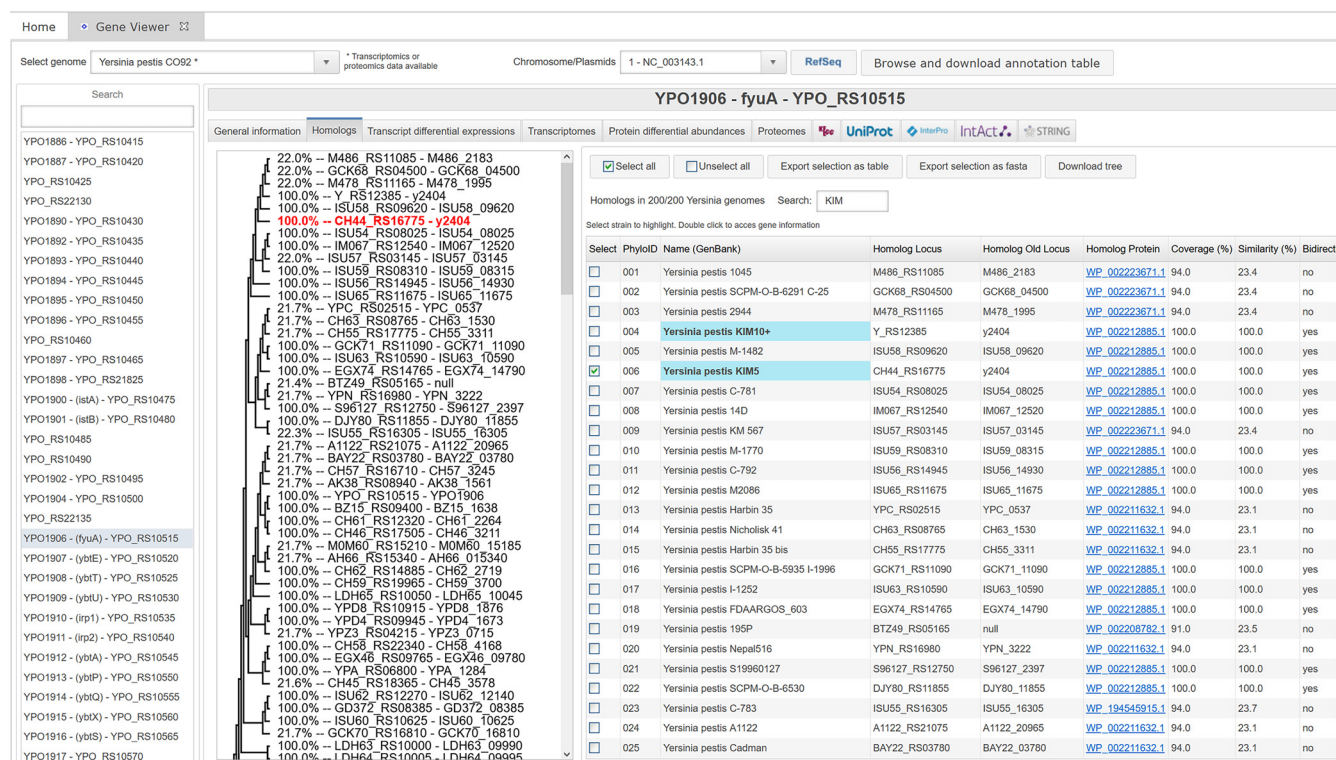


FIG 8 The gene viewer homolog panel. The table of the *fyuA* homologs in the Yersiniomics database is displayed on the right, with the gene locus, the BLASTP results (such as percent coverage and percent similarity on this coverage), the E value and bit score, and the name of the corresponding strain. The product of the percentage of similarity times the percent coverage, as well as the name of the homologous gene, are also displayed on the phylogenetic tree on the left and highlighted when selected. A multi-fasta protein file of the selected homologs can be downloaded.

Data loading. From the home page, two buttons allow the user to load a .gview file previously saved from the genome viewer and to download transcriptomics and proteomics processed data (Fig. 2, bottom panel). These data are sorted in the “Transcriptomes” and “Proteomes” directories, each one subdivided into strain-specific directories. Excel files showing genes in rows and biological conditions or comparisons of the selected strain in columns are available to download in each directory, i.e., $\log_2$ (fold change) table (Table_LOGFC_strain_name.xlsx), associated *P* value (Table_PVALUE_strain_name.xlsx), and associated adjusted *P* value (Table_PADJ_strain_name.xlsx) for the comparisons, TPM-normalized counts for RNA-Seq experiments (AllRNA-SeqTPM_strain_name.xlsx), and intensity values for proteomes (Table_Expr_strain_name.xlsx). The TPM-normalized count can then be used by each user to compare gene expression within the same sample or compute coexpression networks.

Genomic data set description. We collected 200 assemblies of *Yersinia* genomes, among which 10 were sequenced twice (annotated “bis”) by different laboratories with different technologies (such as Sanger or combination of Illumina and Nanopore or PacBio sequencing). Of the 190 unique genomes, we gathered 61 *Y. pestis*, 24 *Y. pseudotuberculosis*, and 37 *Y. enterocolitica* genomes. In addition, 37 strains of the fish pathogen *Y. ruckeri* were also collected, as well as one complete genome of the insect pathogen *Y. entomophaga* and 30 genomes of nonpathogenic *Yersinia* (Table 2).

In the genomics browser, the “Name (GenBank)” column refers to the GenBank taxonomic assignment, but particular attention should be drawn to the “Most recent assignment (cgMLST)” column, where 9 strains were reassigned by the cgMLST scheme developed in our laboratory (40).

Of note, the RefSeq-reannotated locus name, recognizable by its “...RS...” pattern, was used to access the genes differing from the initial widely used locus name, such as “YPO...” for *Y. pestis* CO92 or “YPTB...” for *Y. pseudotuberculosis* IP32953. We added this previous name in a column named “Old Locus” in the heatmaps and this old

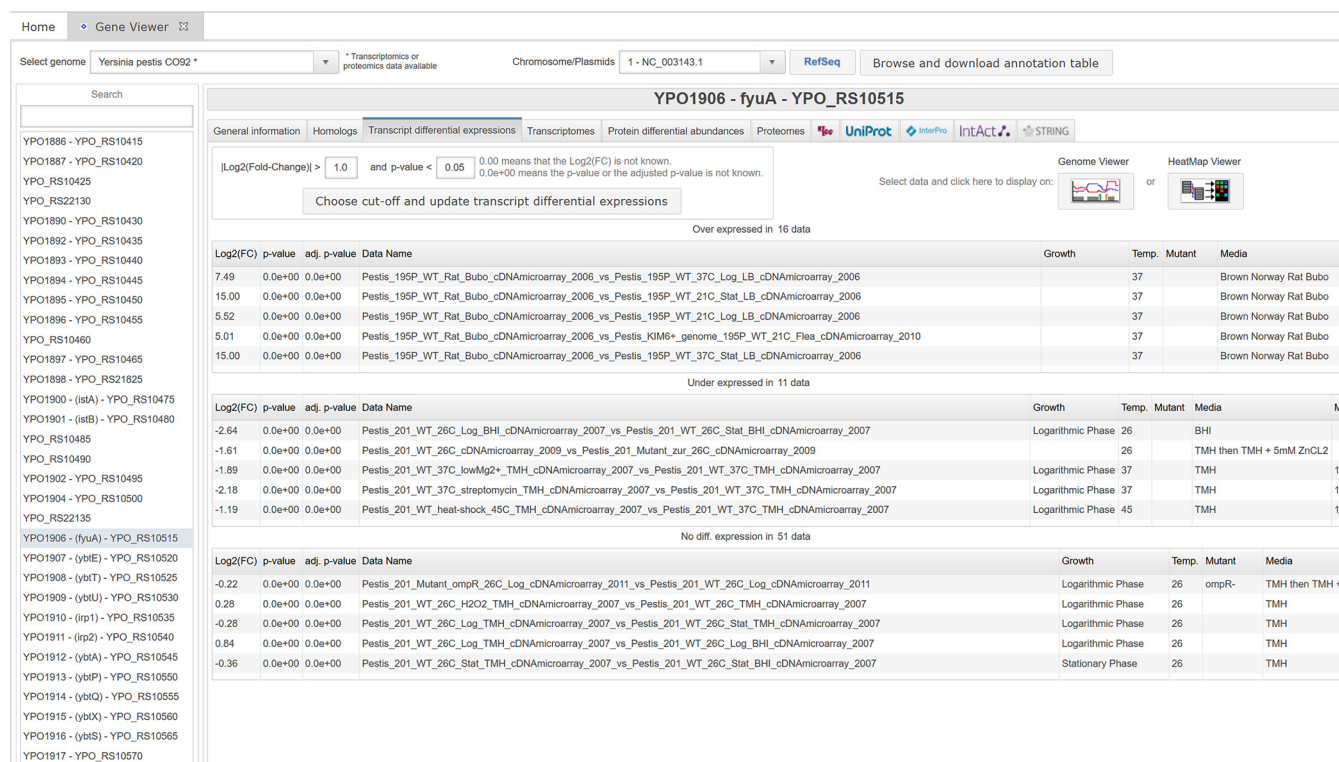


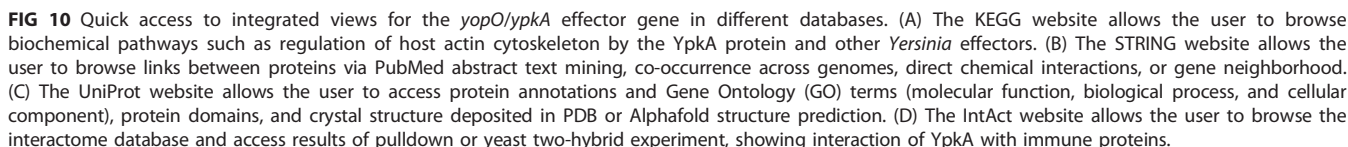
FIG 9 The gene viewer transcript differential expression panel. After selection of a log₂(fold change) cutoff, comparison of biological conditions is dispatched in three panels. For RNA-Seq experiments, results are also filtered by selecting a *P* value cutoff. The top panel shows upregulated conditions and corresponding fold change, *P* value, and adjusted *P* value. For example, *fyuA* expression is upregulated 180-fold in the rat compared to growth in LB at 37°C and upregulated 46-fold compared to growth in LB at 21°C. The middle panel shows downregulated conditions and the corresponding fold change. Expression of *fyuA* is decreased 6.2-fold in exponential phase compared to stationary phase in BHI at 26°C and decreased 4.5-fold in the presence of streptomycin at 37°C compared to the untreated condition. The lower panel displays condition comparisons in which fold change is below the selected cutoff and *P* value, when available. The conditions can be displayed with the genome viewer or the heatmap viewer via the two buttons on the top right.

locus is present after the RefSeq locus name in the gene list of a genome, separated by a dash, and followed by the gene name in parentheses. For example, the *mioC* gene in the CO92 genome can be found as “YPO_RS01005 - YPO0001 (*mioC*).”

The *Y. pestis* community widely uses the CO92 and KIM locus names to refer to genes, using the formats “YPO. . .” and “y. . .” respectively. However, the KIM “y. . .” locus can be found only in the KIM10+ genome annotation in GenBank, a strain lacking several plasmids (pCD1/pYV and pPCP1/pPla) compared to its parental strain, KIM5. To map the RNA-Seq data to the most complete genome but conserve the known locus numbers widely used by the *Y. pestis* community, we reannotated the old locus of the KIM5 genome with the KIM10+ old locus using a systematic BLAST search between both genomes’ open reading frames.

Of note, the YPIII genome in GenBank lacks the pIB1 plasmid (the YPIII pYV virulence plasmid coding for a type III secretion system). Thus, this plasmid was not included in the database.

Transcriptomics data set description. Transcriptomic experiments were collected for *Yersinia* reference strains such as *Y. pestis* CO92, KIM, 91001, and Pestoides F, *Y. pseudotuberculosis* YPIII and IP32953, *Y. enterocolitica* 8081 and Y11, *Y. ruckeri* CSF007-82, and *Y. entomophaga* MH96 (Table 3). Of note, some strains’ data are mapped on genomes of other strains, such as *Y. pestis* 201 Microtus, whose microarray data were generated using CO92 microarray, or RNA-Seq data that were mapped to strain 91001 Microtus strain, as the strain 201 sequence is assembled only at the scaffold level. This is also the case for *Y. ruckeri* CSF007-82, whose complete assembled genomic sequence is not available, and RNA-Seq data were mapped to its genetically closest sequenced strain, QMA0440. We gathered 151 biological



Of the 166 RNA-Seq biological conditions, 44 were mapped on the *Y. pseudotuberculosis* YPIII genome. A total of 30 come from the *Y. pestis* 201 Microtus strain and were mapped to the close relative strain 91001. Additionally, 24 biological conditions were mapped to the *Y. enterocolitica* 8081 genome, 17 were mapped to the *Y. pseudotuberculosis* IP32953 genome, 11 to the *Y. enterocolitica* Y11 genome, 9 to the *Y. pestis* CO92 genome, 8 to the *Y. enterocolitica* Y1 genome (of which 4 come from strain 8081), 8 to the *Y. entomophaga* MH96 genome, 7 to the *Y. pestis* KIM5 genome, 4 to the *Y. ruckeri* SC09 genome, and 2 to the vaccine strain *Y. pestis* EV76 genome, and 2 biological conditions with the *Y. ruckeri* strain CSF007-82 were mapped to its close relative QMA0440.

TABLE 2 Numbers of complete *Yersinia* genomes^a

<i>Yersinia</i> species	No. of genomes	
	Unique complete	Sequenced twice
<i>Y. pestis</i>	61	6
<i>Y. enterocolitica</i>	37	1
<i>Y. ruckeri</i>	37	
<i>Y. pseudotuberculosis</i>	24	3
<i>Y. intermedia</i>	8	
<i>Y. occitanica</i>	5	
<i>Y. massiliensis</i>	4	
<i>Y. bercovieri</i>	2	
<i>Y. hibernica</i>	2	
<i>Y. aldovae</i>	1	
<i>Y. aleksiciae</i>	1	
<i>Y. alsatica</i>	1	
<i>Y. canariae</i>	1	
<i>Y. entomophaga</i>	1	
<i>Y. frederiksenii</i>	1	
<i>Y. kristensenii</i>	1	
<i>Y. mollaretii</i>	1	
<i>Y. rohdei</i>	1	
<i>Y. similis</i>	1	
Total	190	10

^aSpecies nomenclature proposed by the French *Yersinia* National Reference Center.

A summary of the 154 *Y. pestis* experiments can be found aggregated by culture temperature and strain (Table 4), culture medium and strain (Table 5), and genetic status and strain (Table 6).

Proteomic data set description. Many *Y. pestis* proteomes have been aggregated and processed by the Pacific Northwest National Laboratory (PNNL) to create a tool that differentiates naturally occurring and laboratory strains of *Y. pestis* (53), encompassing data from different studies (48, 54, 55) as well as PNNL archives (56). These processed intensity data sets were kindly shared by Eric D. Merkley. These data represent most of the 32 biological conditions which used whole-cell lysates of *Y. pestis* reference strain CO92 and were analyzed via shotgun LC-MS/MS using linear ion trap technologies and/or Orbitrap mass spectrometry for identification (Thermo Fisher LTQ XL/LTQ Velos). Proteomes also include 5 biological conditions from one study investigating *Y. pestis* Microtus strain 201 using Orbitrap LC-MS/MS technology (Thermo

TABLE 3 Number of omics biological conditions mapped to reference strains

Reference strain	No. of conditions		
	Microarray	RNA-Seq	Mass spectrometry
<i>Y. pestis</i> CO92	100	9	32
<i>Y. pestis</i> KIM	9	7	24
<i>Y. pestis</i> Microtus 91001		30	6
<i>Y. pestis</i> Pestoides F	8		
<i>Y. pestis</i> EV76		2	
<i>Y. pseudotuberculosis</i> YPIII	20	44	
<i>Y. pseudotuberculosis</i> IP32953		17	
<i>Y. pseudotuberculosis</i> PB1 +	8		
<i>Y. enterocolitica</i> 8081	6	24	
<i>Y. enterocolitica</i> Y11		11	
<i>Y. enterocolitica</i> Y1		8	
<i>Y. ruckeri</i> SC09		4	
<i>Y. ruckeri</i> QMA0440		2	
<i>Y. entomophaga</i> MH96		8	
Total	151	166	62

TABLE 4 Number of *Y. pestis* transcriptomic experiments according to culture temperature and strain

Culture temp (°C)	<i>Y. pestis</i> strain used	No. of expts	
		Array	RNA-Seq
10	201	1	
21	195P	2	
	EV76		1
	KIM6+	5	
26	201	18	21
	CO92	4	
	Pestoides F	4	
28/37	CO92	4	
28	CO92	1	
	KIM5	4	
	KIM53	5	1
30	CO92	14	
37	195P	3	
	201	16	8
	CO92	18	9
	EV76		1
	GB	2	
	KIM53		4
	KIM6+		2
	Pestoides F	4	
45	201	1	
Multiple	201		1

Fisher QExactive), the first and only study defining the *Y. pestis* secretome (57). In addition, 24 biological conditions from 2 studies focus on the *Y. pestis* strain KIM (12 in FTICR, 12 in 2D gel electrophoresis plus LC-MS/MS) and used cell fractionation to determine protein localization (58, 59) (Table 3).

Data set exploration and validation. One dual RNA-Seq experiment providing gene expression profiles in a murine pneumonic plague model (60) did generate RNA-Seq raw data; however, gene content and differential expression were not analyzed. Our systematic processing pipeline allowed us to process and validate these data.

The *in vivo* experiment comparing transcripts of the *Y. pestis* KIM strain in OF1-infected mouse lungs versus HIB cultures confirmed most previous results described in the seminal microarray experiment in a pneumonic plague model that used the CO92 strain infecting C57BL/6 mice or grown in BHI (44): the methionine biosynthesis pathway genes *metA*, *metE*, *metF*, *metK*, and *metR* were among the most upregulated genes in the lungs in both studies. Similarly, the yersiniabactin operon (y2404 to y2394 in the KIM strain) was the most upregulated operon at 1 h, 24 h, and 48 h postinfection, highlighting the importance of metal acquisition by bacteria in their mammalian host. The pH 6 antigen *psaA* gene, encoding fimbriae required for virulence, was downregulated 48 h postinfection in both studies but was upregulated 1 h postinfection in the latter experiment. The cold shock-responsive gene *cspD* and genes involved in the detoxification of reactive oxygen species, such as *kata* and *katG* (*katY*), encoding catalases, and *sodB*, encoding superoxide dismutase, were downregulated in both studies, and nitric oxide-induced *hmpA* was even more upregulated in the experiment using the KIM strain than in the experiment based on the CO92 strain. However, the plasminogen activator *pla* was downregulated 24 h postinfection and not at 48 h postinfection in the study using the KIM strain, in contrast to the study using the CO92 strain. Similarly, *cspA1* and *cspA2* were upregulated in the lungs infected with the CO92 strain but downregulated in the study by Israeli et al. (60). Discrepancies between these results could be explained by differences in *Y. pestis* strains and mouse lines as well as culture conditions of bacteria *in vitro*.

TABLE 5 Number of *Y. pestis* transcriptomic experiments according to culture medium and strain

Culture medium ^a	<i>Y. pestis</i> strain used	No. of expts	
		Array	RNA-Seq
BAB broth	GB	2	
BHI	201	2	9
	CO92	7	1
Custom	CO92	8	
	Pestoides F	8	
DMEM	KIM5	1	
HIB	CO92	19	
	KIM53		2
	KIM6+		2
Human plasma	CO92	5	
LB	195P	4	
	201		2
	CO92	1	
	KIM6+	3	
Macrophages	CO92		2
	KIM5	3	
MHB	KIM53	5	
TMH	201	34	11
	CO92		6
<i>In vivo</i>	195P	1	
	201		2
	CO92	1	
	KIM53		3
	KIM6+	2	
Unknown	201		6
	EV76		2

^aBAB, blood agar base; DMEM, Dulbecco's modified Eagle medium; MHB, Mueller-Hinton broth.

DISCUSSION

Over the last 2 decades, the omics revolution has generated an impressive number of data sets which are often difficult to analyze in depth without any prior bioinformatic knowledge. In addition, only a fraction of generated and processed data is generally accessible and/or used in research articles. Here, we aimed at exploiting *Yersinia* omics data published over the last 20 years to make them accessible in a user-friendly way to biologists without familiarity with deep bioinformatics. We thus constructed and processed a database gathering 200 genomic, 317 transcriptomic, and 62 proteomic data sets of *Yersinia* species, which are browsable at the gene level and according to experimental conditions on the custom-made Yersiniomics website (<https://yersiniomics.pasteur.fr/>). Notably, three previous studies (60–62) did not measure differential gene expression under their biological conditions. These data were thus processed and included in Yersiniomics. One of these studies confirmed for the most part previous results obtained using a similar microarray approach. Other recent RNA-Seq studies did not make their raw sequencing data publicly available and could not be included in our database for now (63–66). To expand information on genes of interest, dynamic links to external databases were implemented for several reference strains, allowing the user to access KEGG pathways, UniProt annotation, InterPro domains, IntAct interactors and STRING networks at the gene level. These links also facilitate the access to other data, such as GO terms and AlphaFold structure predictions, implemented in UniProt.

As new data and data types continue to be published, we intend to update and upgrade the Yersiniomics database with newly published RNA-Seq and LC-MS/MS

TABLE 6 Number of *Y. pestis* transcriptomic experiment according to strain and genetic status

Strain used	Genetic status ^a	No. of expts	
		Array	RNA-Seq
195P	WT	5	
201	<i>cobB</i>		1
	<i>fur</i>	2	
	<i>fyuA</i>		2
	<i>fyuA</i> deletion GCA		2
	<i>hfq</i> pHfq		1
	<i>hfq</i> pHfq-FLAG		2
	<i>lcrG</i>	1	
	<i>ompR</i>	3	
	<i>oxyR</i>	1	
	pHfq-FLAG		1
	<i>phoP</i>	2	
	<i>rscB</i>		2
	<i>rscD</i> _{pestis} :: <i>rscD</i> _{pseudotb}		2
	WT pFLAG		3
	<i>yfiQ</i>		1
	<i>zur</i>	1	
	WT	26	13
CO92	<i>crp</i> pCD1 [−]		2
	pCD1 [−]		4
	<i>pgm</i>	15	
	<i>pgm luxS</i>	2	
	<i>pgm ypeIR</i>	2	
	<i>pgm ypeIR yspIR</i>	1	
	<i>pgm ypeIR yspIR luxS</i>	2	
	<i>pgm yspI</i>	2	
	<i>pgm yspI ypeIR</i>	1	
	pPCP1 [−]	8	
	WT	8	3
EV76-CN	WT		2
GB	<i>dam</i>	1	
	WT	1	
KIM5	WT	4	
KIM53	WT	5	5
KIM6+	pCD1 [−]	5	
	<i>psaE</i>		1
	WT		1
Pestoides F	WT	8	

^aWT, wild-type strain; mutant strain (*mutated locus*).

data, with omics data generated in our laboratory, and with already available data types such as small RNAs (61–68), transcriptional start sites (TSS) (67–69), and riboswitches (67, 69). New analyses, such as gene ranking using TPM counts for each RNA-Seq experiments, will also be implemented. Ultimately, mutant phenotypes associated with specific gene locus could also be added based on signature-tagged mutagenesis screening (70–75), high-throughput transposon site hybridization procedure (76), or more recent next-generation sequencing using techniques such as transposon-insertion sequencing or transposon-directed insertion sequencing (77–81).

MATERIALS AND METHODS

Genomic data collection. Genomic data were browsed on the NCBI database with the keyword “*Yersinia*” (<https://www.ncbi.nlm.nih.gov/data-hub/genome>). Assemblies were filtered at the “chromosome” or “complete” level, and “chromosome” assemblies were manually curated to verify data completeness. The most recent RefSeq annotated assemblies were downloaded in December 2022.

Transcriptomic data collection. Microarray data were collected on Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>), and the differential expressions calculated by data depositors were directly used when available. Alternatively, tables and supplemental tables were downloaded from articles and formatted. RNA-Seq data sets were browsed on the European Nucleotide Archive (<https://www.ebi.ac.uk/ena/>) using the keywords “Yersinia” and “transcriptome,” “RNA-Seq,” or “RNASeq” and on the Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>) using the keyword “Yersinia” and selecting the “RNA” source. Raw data, consisting of sequencing reads in fastq format, were downloaded from the SRA FTP server <http://ftp.sra.ebi.ac.uk/vol1/> via the ENA website.

Proteomic data collection. As the treatment of LC-MS/MS raw data from the PRIDE repository is more complex and highly dependent on the mass spectrometry technology used, we directly retrieved the calculated differential expression from the tables and supplemental tables published in the literature when available. Alternatively, processed intensity data were kindly provided by their authors when data were not available online.

Genomic data processing. Genomes were directly processed through the Bacnet platform (33) and implemented in Yersiniomics. From the collected genomes, a phylogenetic tree was reconstructed using the 500 genes of our *Yersinia* cgMLST scheme recently developed in our laboratory (40). The 500 genes were concatenated and aligned using MAFFT v7.453 (82). Phylogenetic reconstruction was performed using IQTREE v2.0.6 (83), and the tree was rooted on *Yersinia entomophaga* MH96 (84), as it is the most ancestral branch of the genus *Yersinia* (40). The tree was drawn using Iroki (85). The taxonomic classification of the 66 *Y. pestis* genomes was based on the presence of SNPs defining the lineages, following the nomenclature developed in previous works (41, 42). *Y. pseudotuberculosis* IP32953 genome was also included in the analysis, to recover the ancestral genotypes. Briefly, either short-read sequencing data or contigs were mapped onto the *Y. pestis* CO92 reference genome (NC_003143) using the Snippy pipeline with default parameters (<https://github.com/tseemann/snippy>) to identify variants in the chromosome. Identified variants were inspected, and those falling within repetitive sequences (i.e., IS) were excluded, while variants associated with putative recombination events were identified using Gubbins v3.2.0 (86) and filtered out. The final set of variants ($n = 14,273$) was then used to reconstruct a maximum-likelihood phylogeny using IQ-TREE 2 (83) to classify the different *Y. pestis* genomes present in Yersiniomics into known evolutionary branches (0.PE2, 0.PE4, 0.PE3, 0.ANT, 1.ORI, 1.IN, 1.ANT, 2.ANT, 2.MED, and 4.ANT). For the other species, lineages were assessed from cgMLST.

Homolog database and synteny. A protein BLAST search was automatically performed for each gene of each genome against the other genomes to construct the homolog database, using the BLAST+ command-line tool v2.13.0 (87). Synteny was constructed using a best-hit bidirectional BLAST search and implemented in the SynTVView framework (50). The tool was reprogrammed from Flash to JavaScript with the HTML5 2d Canvas JavaScript library Konva (<https://konvajs.org/>), which enables high-performance animations. Drawing strategies were developed to ensure that graphical outputs remain below the technological limits of drawing more than 10,000 dynamic graphical objects, using a mix of static and dynamic objects. SynTVViewJS can be used on a public web application at <https://plechat.pages.pasteur.fr/syntviewjs/>. The code repository is at <https://gitlab.pasteur.fr/plechat/syntviewjs>.

Transcriptomic data processing. Downloaded microarray fold change tables were processed through the Bacnet platform. RNA-Seq raw data were processed using Sequana v0.14.6, Sequana-rnaseq v0.17.0, and Sequana-pipetools v0.10.0s (88), implementing the following software with default parameters: reads were trimmed using fastp v0.20.1 (89), processed reads were then aligned on the selected reference genome with Bowtie2 v2.4.4 (90), and strandedness was automatically identified by the Sequana pipeline. Mapped reads were then quantified using featureCounts (package subread v2.0.1) (91), and differential analysis was performed on raw read counts using DESeq2 v1.38.2 with R v4.2.1 with default parameters and Cook's cutoff enabled (43). *P* values and adjusted *P* values were calculated using default DESeq2 parameters consisting, respectively, of the Wald test and the Benjamini-Hochberg correction. Fold change and adjusted *P* value tables were extracted from DESeq2 results and processed using the Bacnet platform. BAM files generated by the Sequana pipeline were converted to strand-specific .wig files using the strand_cov function from the stranded-coverage package (<https://github.com/pmenzel/stranded-coverage.git>) and then processed using the Bacnet platform. Raw read counts from featureCounts were also normalized using the TPM method (51) to be displayed in the transcriptome panel of the gene viewer.

Proteomic data processing. For the differential analyses of one condition versus another, proteins identified in the reverse and contaminant databases and proteins “only identified by site” were first discarded from the list of identified proteins. Then, only proteins with at least three quantified intensities in a condition were kept. Differential expression was performed using LFQ when at least 1 unique peptide was detected and at least 2 peptides were quantified in a condition. A normalization was applied within the same condition centered on means of the medians (92). Remaining proteins without any intensity value in one of two conditions were considered quantitatively present in a condition and absent in another. They were therefore set aside and considered differentially abundant proteins. Missing values were imputed using the SLSA method thanks to the R package imp4p (93). Differential analysis was performed with a limma *t* test and an adaptive Benjamini-Hochberg correction to adjust the *P* values with the cp4P R package (94). Proteins with a fold change less than 2.0 were considered not significantly differentially abundant. Statistical testing of the remaining proteins (having a fold change greater than 2.0) was conducted using the limma *t* test (95).

These calculated fold changes were formatted and processed using the Bacnet platform, in parallel to fold change tables downloaded from published articles.

Yersiniomics website design. Yersiniomics was constructed with the Bacnet platform (33), based on Java and Eclipse e4 RCP/RAP API, and we contributed to its most recent update.

Links to external databases. The link to the KEGG database website uses the KEGG species identifier ("ype" for *Y. pestis* CO92) and the old locus name (for instance, "YPO0001" for the *mioC* gene of *Y. pestis* CO92), dynamically queried by URL in the form "<https://www.genome.jp/entry/ype:YPO0001>." UniProt accessions are dynamically retrieved and parsed with the KEGG species identifiers and the old locus name using KEGG REST Application Programming Interface (API), queried by URL in the form "<https://rest.kegg.jp/conv/uniprot/ype:YPO0001>." When a UniProt accession is retrieved ("A0A0H2W280" for the "YPO0001" old locus name of *Y. pestis* CO92), UniProt and InterPro are dynamically queried by URL in the forms "<https://www.uniprot.org/uniprotkb/A0A0H2W280>" and "<https://www.ebi.ac.uk/interpro/protein/UniProt/A0A0H2W280>," respectively. IntAct is dynamically accessed with the old locus name by URL in the form "<https://www.ebi.ac.uk/intact/search?query=YPO0001>" or "<https://www.ebi.ac.uk/intact/search?query=YPO0001%20A0A0H2W280>" if a UniProt accession was retrieved. A STRING URL is dynamically retrieved with the STRING species identifier ("214092" for *Y. pestis* CO92) and the old locus name, querying the STRING API in the form "<https://version-11-5.string-db.org/cgi/network?taskId=bl0qgHbfGxqi&sessionId=b41GkPqd7zt>." The retrieved URL is then accessed.

Data availability. The source code of the Yersiniomics website, based on the Bacnet platform, is available on the GitHub repository <https://github.com/becavin-lab/bacnet/>. All processed data can be directly downloaded from the Yersiniomics website.

ACKNOWLEDGMENTS

The project received funding from Institut Pasteur, Agence de l'Innovation de Défense (AID-DGA), Université Paris Cité, CNRS, LabEX Integrative Biology of Emerging Infectious Diseases (ANR-10-LBX-62-IBEID), Fondation pour la Recherche Médicale (FDT202204015222), and the Inception program (Investissement d'Avenir grant ANR-16-CONV-0005). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

We thank Eric D. Merkley for sharing Pacific Northwest National Laboratory data sets. We are grateful to all members of the *Yersinia* research unit and the French national reference center for plague and other yersiniosis for insightful discussions.

We declare no conflict of interest.

REFERENCES

- Adeolu M, Alnajjar S, Naushad S, Gupta RS. 2016. Genome-based phylogeny and taxonomy of the 'Enterobacterales': proposal for *Enterobacterales* ord. nov. divided into the families *Enterobacteriaceae*, *Erwinaceae* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int J Syst Evol Microbiol* 66:5575–5599. <https://doi.org/10.1099/ijsem.0.001485>.
- Reuter S, Connor TR, Barquist L, Walker D, Feltwell T, Harris SR, Fookes M, Hall ME, Petty NK, Fuchs TM, Corander J, Dufour M, Ringwood T, Savin C, Bouchier C, Martin L, Miettinen M, Shubin M, Riehm JM, Laukkanen-Ninios R, Sihvonen LM, Siitonen A, Skurnik M, Falcão JP, Fukushima H, Scholz HC, Prentice MB, Wren BW, Parkhill J, Carniel E, Achtman M, McNally A, Thomson NR. 2014. Parallel independent evolution of pathogenicity within the genus *Yersinia*. *Proc Natl Acad Sci U S A* 111:6768–6773. <https://doi.org/10.1073/pnas.1317161111>.
- McNally A, Thomson NR, Reuter S, Wren BW. 2016. "Add, stir and reduce": *Yersinia* spp. as model bacteria for pathogen evolution. *Nat Rev Microbiol* 14:177–190. <https://doi.org/10.1038/nrmicro.2015.29>.
- European Centre for Disease Prevention and Control. 2022. Yersiniosis - annual epidemiological report for 2020.
- Saraka D, Savin C, Kouassi S, Cissé B, Koffi E, Cabanel N, Brémont S, Faye-Kette H, Dosso M, Carniel E. 2017. *Yersinia enterocolitica*, a neglected cause of human enteric infections in Côte d'Ivoire. *PLoS Negl Trop Dis* 11: e0005216. <https://doi.org/10.1371/journal.pntd.0005216>.
- 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, Linder LE, Carniel E, Keim P. 2004. Microevolution and history of the plague bacillus, *Yersinia pestis*. *Proc Natl Acad Sci U S A* 101:17837–17842. <https://doi.org/10.1073/pnas.0408026101>.
- Demeure CE, Dussurget O, Mas Fiol G, le Guern AS, Savin C, Pizarro-Cerdá J. 2019. *Yersinia pestis* and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. *Genes Immun* 20:357–370. <https://doi.org/10.1038/s41435-019-0065-0>.
- Bertherat E. 2019. Plague around the world in 2019/La peste dans le monde en 2019. *Wkly Epidemiol Rec* 94:289–293.
- Baril L, Vallès X, Stenseth NC, Rajerison M, Ratsitorahina M, Pizarro-Cerdá J, Demeure C, Belmain S, Scholz H, Girod R, Hinnebusch J, Vigan-Womas I, Bertherat E, Fontanet A, Yazadanpanah Y, Carrara G, Deuve J, D'Ortenzio E, Angulo JOC, Mead P, Horby PW. 2019. Can we make human plague history? A call to action. *BMJ Glob Health* 4:e001984. <https://doi.org/10.1136/bmjgh-2019-001984>.
- Vallès X, Stenseth NC, Demeure C, Horby P, Mead PS, Cabanillas O, Ratsitorahina M, Rajerison M, Andrianivoarimanana V, Ramasindrazana B, Pizarro-Cerdá J, Scholz HC, Girod R, Hinnebusch BJ, Vigan-Womas I, Fontanet A, Wagner DM, Telfer S, Yazadanpanah Y, Tortosa P, Carrara G, Deuve J, Belmain SR, D'Ortenzio E, Baril L. 2020. Human plague: an old scourge that needs new answers. *PLoS Negl Trop Dis* 14:e0008251. <https://doi.org/10.1371/journal.pntd.0008251>.
- Kumar G, Menanteau-Ledouble S, Saleh M, El-Matbouli M. 2015. *Yersinia ruckeri*, the causative agent of enteric redmouth disease in fish. *Vet Res* 46:e103. <https://doi.org/10.1186/s13567-015-0238-4>.
- Hurst MRH, Becher SA, Young SD, Nelson TL, Glare TR. 2011. *Yersinia entomophaga* sp. nov., isolated from the New Zealand grass grub *Costelytra zealandica*. *Int J Syst Evol Microbiol* 61:844–849. <https://doi.org/10.1099/ijso.0.024406-0>.
- Glare TR, O'Callaghan M. 2019. Microbial biopesticides for control of invertebrates: progress from New Zealand. *J Invertebr Pathol* 165:82–88. <https://doi.org/10.1016/j.jip.2017.11.014>.
- Isberg RR, Falkow S. 1985. A single genetic locus encoded by *Yersinia pseudotuberculosis* permits invasion of cultured animal cells by *Escherichia coli* K-12. *Nature* 317:262–264. <https://doi.org/10.1038/317262a0>.
- Isberg RR, Voorhis DL, Falkow S. 1987. Identification of invasins: a protein that allows enteric bacteria to penetrate cultured mammalian cells. *Cell* 50:769–778. [https://doi.org/10.1016/0092-8674\(87\)90335-7](https://doi.org/10.1016/0092-8674(87)90335-7).
- Isberg RR, Leong JM. 1990. Multiple β 1 chain integrins are receptors for invasins, a protein that promotes bacterial penetration into mammalian cells. *Cell* 60:861–871. [https://doi.org/10.1016/0092-8674\(90\)90099-z](https://doi.org/10.1016/0092-8674(90)90099-z).
- Alrutz MA, Srivastava A, Wong KW, D'Souza-Schorey C, Tang M, Ch'Ng LE, Snapper SB, Isberg RR. 2001. Efficient uptake of *Yersinia pseudotuberculosis* via integrin receptors involves a Rac1-Arp 2/3 pathway that bypasses N-WASP function. *Mol Microbiol* 42:689–703.
- Wong KW, Isberg RR. 2003. Arp6 and phosphoinositol-4-phosphate-5-kinase activities permit bypass of the Rac1 requirement for β 1 integrin-

- mediated bacterial uptake. *J Exp Med* 198:603–614. <https://doi.org/10.1084/jem.20021363>.
19. Pizarro-Cerdá J, Cossart P. 2004. Subversion of phosphoinositide metabolism by intracellular bacterial pathogens. *Nat Cell Biol* 6:1026–1033. <https://doi.org/10.1038/ncb1104-1026>.
 20. Machner MP, Isberg RR. 2007. A bifunctional bacterial protein links GDI displacement to Rab1 activation. *Science* 318:974–977. <https://doi.org/10.1126/science.1149121>.
 21. Cornelis GR, Wolf-Watz H. 1997. The *Yersinia* Yop virulon: a bacterial system for subverting eukaryotic cells. *Mol Microbiol* 23:861–867. <https://doi.org/10.1046/j.1365-2958.1997.2731623.x>.
 22. Cornelis GR. 2002. The *Yersinia* Ysc–Yop “Type III” weaponry. *Nat Rev Mol Cell Biol* 3:742–752. <https://doi.org/10.1038/nrm932>.
 23. Sayers EW, Cavanaugh M, Clark K, Pruitt KD, Schoch CL, Sherry ST, Karsch-Mizrachi I. 2022. GenBank. *Nucleic Acids Res* 50:D161–D164. <https://doi.org/10.1093/nar/gkab1135>.
 24. Clough E, Barrett T. 2016. The Gene Expression Omnibus database. *Methods Mol Biol* 1418:93–110. https://doi.org/10.1007/978-1-4939-3578-9_5.
 25. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, Snow C, Fonseca NA, Petryszak R, Papatheodorou I, Sarkans U, Brazma A. 2019. ArrayExpress update – from bulk to single-cell expression data. *Nucleic Acids Res* 47:D711–D715. <https://doi.org/10.1093/nar/gky964>.
 26. Cummins C, Ahamed A, Aslam R, Burgin J, Devraj R, Edbali O, Gupta D, Harrison PW, Haseeb M, Holt S, Ibrahim T, Ivanov E, Jayathilaka S, Kadhivelu V, Kay S, Kumar M, Lathi A, Leinonen R, Madeira F, Madhusoodanan N, Mansurova M, O’Cathail C, Pearce M, Pesant S, Rahman N, Rajan J, Rinck G, Selvakumar S, Sokolov A, Suman S, Thorne R, Totoo P, Vijayaraja S, Waheed Z, Zyouad A, Lopez R, Burdett T, Cochrane G. 2022. The European Nucleotide Archive in 2021. *Nucleic Acids Res* 50:D106–D110. <https://doi.org/10.1093/nar/gkab1051>.
 27. Katz K, Shutov O, Lapoint R, Kimelman M, Brister JR, O’Sullivan C. 2022. The Sequence Read Archive: a decade more of explosive growth. *Nucleic Acids Res* 50:D387–D390. <https://doi.org/10.1093/nar/gkab1053>.
 28. Deutsch EW, Bandeira N, Sharma V, Perez-Riverol Y, Carver JJ, Kundu DJ, García-Seisdedos D, Jarnuczak AF, Hewapathirana S, Pullman BS, Wertz J, Sun Z, Kawano S, Okuda S, Watanabe Y, Hermjakob H, Maclean B, Maccoss MJ, Zhu Y, Ishihama Y, Vizcaino JA. 2020. The ProteomeXchange consortium in 2020: enabling ‘big data’ approaches in proteomics. *Nucleic Acids Res* 48:D1145–D1152. <https://doi.org/10.1093/nar/gkz984>.
 29. Keseler IM, Gama-Castro S, Mackie A, Billington R, Bonavides-Martínez C, Caspi R, Kothari A, Krummenacker M, Midford PE, Muñoz-Rascado L, Ong WK, Paley S, Santos-Zavaleta A, Subhraveti P, Tierrafria VH, Wolfe AJ, Collado-Vides J, Paulsen IT, Karp PD. 2021. The EcoCyc Database in 2021. *Front Microbiol* 12:711077. <https://doi.org/10.3389/fmicb.2021.711077>.
 30. Cherry JM, Adler C, Ball C, Chervitz SA, Dwight SS, Hester ET, Jia Y, Juvik G, Roe T, Schroeder M, Weng S, Botstein D. 1998. SGD: *Saccharomyces* Genome Database. *Nucleic Acids Res* 26:73–79. <https://doi.org/10.1093/nar/26.1.73>.
 31. Bécavin C, Koutero M, Tchitchek N, Cerutti F, Lechat P, Maillat N, Hoede C, Chiappello H, Gaspin C, Cossart P. 2017. Listeriomics: an interactive web platform for systems biology of *Listeria*. *mSystems* 2:e00186-16. <https://doi.org/10.1128/mSystems.00186-16>.
 32. Davis JJ, Wattam AR, Aziz RK, Brettin T, Butler R, Butler RM, Chlenski P, Conrad N, Dickerman A, Dietrich EM, Gabbard JL, Gerdes S, Guard A, Kenyon RW, MacHi D, Mao C, Murphy-Olson D, Nguyen M, Nordberg EK, Olsen GJ, Olson RD, Overbeek JC, Overbeek R, Parrello B, Pusch GD, Shukla M, Thomas C, Vanoeffelen M, Vonstein V, Warren AS, Xia F, Xie D, Yoo H, Stevens R. 2020. The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. *Nucleic Acids Res* 48:D606–D612. <https://doi.org/10.1093/nar/gkz943>.
 33. Danès L, Tchitchek N, Bécavin C. 2021. Bacnet: a user-friendly platform for building multi-omics websites. *Bioinformatics* 37:1335–1336. <https://doi.org/10.1093/bioinformatics/btaa828>.
 34. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. 2017. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45:D353–D361. <https://doi.org/10.1093/nar/gkw1092>.
 35. UniProt Consortium, Bateman A, Martin M-J, Orchard S, Magrane M, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bye-A-Jee H, Cukura A, Denny P, Dogan T, Ebenezer T, Fan J, Garmiri P, da Costa Gonzales LJ, Hatton-Ellis E, Hussein A, Ignatchenko A, Insana G, Ishitani R, Joshi V, Jyothi D, Kandasamy S, Lock A, Luciani A, Lugalic M, Luo J, Lussi Y, MacDougall A, Madeira F, Mahmoudy M, Mishra A, Moulang K, Nightingale A, Pundir S, Qi G, Raj S, Raposo P, Rice DL, Saidi R, Santos R, Speretta E, Stephenson J, Totoo P, Turner E, Tyagi N, Vasudev P, Warner K, Watkins X, Zaru R, Zellner H, Bridge AJ, Aimo L, Argoud-Puy G, Auchincloss AH, Axelsen KB, Bansal P, Baratin D, Batista Neto TM, Blatter M-C, Bolleman JT, Boutet E, Breuza L, Gil BC, Casals-Casas C, Echioukh KC, Coudert E, Cuche B, de Castro E, Estreicher A, Famiglietti ML, Feuermann M, Gasteiger E, Gaudet P, Gehant S, Gerritsen V, Gos A, Gruaz N, Hulo C, Hyka-Nouspikel N, Jungo F, Kerhornou A, le Mercier P, Lieberherr D, Masson P, Morgat A, Muthukrishnan V, Paesano S, Pedruzzi I, Pilboud S, Pourcel L, Poux S, Pozzato M, Pruess M, Redaschi N, Rivoire C, Sigrist CJA, Sonesson K, Sundaram S, Wu CH, Arighi CN, Arminski L, Chen C, Chen Y, Huang H, Laiho K, McGarvey P, Natale DA, Ross K, Vinayaka CR, Wang Q, Wang Y, Zhang J. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 51:D523–D531. <https://doi.org/10.1093/nar/gkac1052>.
 36. Blum M, Chang HY, Chuguransky S, Grego T, Kandasamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD. 2021. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res* 49:D344–D354. <https://doi.org/10.1093/nar/gkaa977>.
 37. del Toro N, Shrivastava A, Ragueneau E, Meldal B, Combe C, Barrera E, Perfetto L, How K, Ratan P, Shirodkar G, Lu O, Mészáros B, Watkins X, Pundir S, Licata L, Iannuccelli M, Pellegrini M, Martin MJ, Panni S, Duesbury M, Vallet SD, Rappsilber J, Ricard-Blum S, Cesareni G, Salwinski L, Orchard S, Porras P, Panneerselvam K, Hermjakob H. 2022. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res* 50:D648–D653. <https://doi.org/10.1093/nar/gkab1006>.
 38. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryar F, Hachilif R, Gable AL, Fang T, Doncheva NT, Pyysalo S, Bork P, Jensen LJ, von Mering C. 2022. The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* 51:D638–D646. <https://doi.org/10.1093/nar/gkac1000>.
 39. Perez-Riverol Y, Bai J, Bandla C, García-Seisdedos D, Hewapathirana S, Kamatchinathan S, Kundu DJ, Prakash A, Ferriks-Zipper A, Eisenacher M, Walzer M, Wang S, Brazma A, Vizcaino JA. 2022. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res* 50:D543–D552. <https://doi.org/10.1093/nar/gkab1038>.
 40. Savin C, Criscuolo A, Guglielmini J, le Guern AS, Carniel E, Pizarro-Cerdá J, Brisse S. 2019. Genus-wide *Yersinia* core-genome multilocus sequence typing for species identification and strain characterization. *Microb Genom* 5:e000301. <https://doi.org/10.1099/mgen.0.000301>.
 41. Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, Wagner DM, Feldkamp M, Kusecek B, Vogler AJ, Li Y, Cui Y, Thomson NR, Jombart T, Leblois R, Lichtner P, Rahalison L, Petersen JM, Balloux F, Keim P, Wirth T, Ravel J, Yang R, Carniel E, Achtman M. 2010. *Yersinia pestis* genome sequencing identifies patterns of global phylogenetic diversity. *Nat Genet* 42:1140–1143. <https://doi.org/10.1038/ng.705>.
 42. Cui Y, Yu C, Yan Y, Li D, Li Y, Jombart T, Weinert LA, Wang Z, Guo Z, Xu L, Zhang Y, Zheng H, Qin N, Xiao X, Wu M, Wang X, Zhou D, Qi Z, Du Z, Wu H, Yang X, Cao H, Wang H, Wang J, Yao S, Rakin A, Li Y, Falush D, Balloux F, Achtman M, Song Y, Wang J, Yang R. 2013. Historical variations in mutation rate in an epidemic pathogen, *Yersinia pestis*. *Proc Natl Acad Sci U S A* 110:577–582. <https://doi.org/10.1073/pnas.1205750110>.
 43. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:e550. <https://doi.org/10.1186/s13059-014-0550-8>.
 44. Lathem WW, Crosby SD, Miller VL, Goldman WE. 2005. Progression of primary pneumonic plague: a mouse model of infection, pathology, and bacterial transcriptional activity. *Proc Natl Acad Sci U S A* 102:17786–17791. <https://doi.org/10.1073/pnas.0506840102>.
 45. Chauvaux S, Rosso M-L, Frangeul L, Lacroix C, Labarre L, Schiavo A, Marceau M, Dillies M-A, Foulon J, Coppee J-Y, Medigue C, Simonet M, Carniel E. 2007. Transcriptome analysis of *Yersinia pestis* in human plasma: an approach for discovering bacterial genes involved in septicemic plague. *Microbiology (Reading)* 153:3112–3124. <https://doi.org/10.1099/mic.0.2007/006213-0>.
 46. Sebbane F, Gardner D, Long D, Gowen BB, Hinnebusch BJ. 2005. Kinetics of disease progression and host response in a rat model of bubonic plague. *Am J Pathol* 166:1427–1439. [https://doi.org/10.1016/S0002-9440\(10\)62360-7](https://doi.org/10.1016/S0002-9440(10)62360-7).
 47. Ritzert JT, Minasov G, Embury R, Schipma MJ, Satchell KJF. 2019. The cyclic AMP receptor protein regulates quorum sensing and global gene expression in *Yersinia pestis* during planktonic growth and growth in biofilms. *mBio* 10:e02613-19. <https://doi.org/10.1128/mBio.02613-19>.

48. 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. 2012. Comparative omics-driven genome annotation refinement: application across *Yersiniae*. *PLoS One* 7: e33903. <https://doi.org/10.1371/journal.pone.0033903>.
49. Avican K, Aldahdooh J, Togninalli M, Mahmud AKMF, Tang J, Borgwardt KM, Rhen M, Fällman M. 2021. RNA atlas of human bacterial pathogens uncovers stress dynamics linked to infection. *Nat Commun* 12:e3282. <https://doi.org/10.1038/s41467-021-23588-w>.
50. Lechat P, Souche E, Moszer I. 2013. SynTVIEW - an interactive multi-view genome browser for next-generation comparative microorganism genomics. *BMC Bioinformatics* 14:e277. <https://doi.org/10.1186/1471-2105-14-277>.
51. Abrams ZB, Johnson TS, Huang K, Payne PRO, Coombes K. 2019. A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinformatics* 20:679. <https://doi.org/10.1186/s12859-019-3247-x>.
52. Dyer MD, Nef C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, Murali TM, Sobral BW. 2010. The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. *PLoS One* 5:e0012089. <https://doi.org/10.1371/journal.pone.0012089>.
53. Merkley ED, Sego LH, Lin A, Leiser OP, Kaiser BL, Adkins JN, Keim PS, Wagner DM, Kreuzer HW. 2017. Protein abundances can distinguish between naturally-occurring and laboratory strains of *Yersinia pestis*, the causative agent of plague. *PLoS One* 12:e0183478. <https://doi.org/10.1371/journal.pone.0183478>.
54. Ansong C, Schrimpe-Rutledge AC, Mitchell HD, Chauhan S, Jones MB, Kim Y-M, McAteer K, Deatherage Kaiser BL, Dubois JL, Brewer HM, Frank BC, McDermott JE, Metz TO, Peterson SN, Smith RD, Motin VL, Adkins JN. 2013. A multi-omic systems approach to elucidating *Yersinia* virulence mechanisms. *Mol Biosyst* 9:44–54. <https://doi.org/10.1039/c2mb25287b>.
55. Lin A, Merkley ED, Clowers BH, Hutchison JR, Kreuzer HW. 2015. Effects of bacterial inactivation methods on downstream proteomic analysis. *J Microbiol Methods* 112:3–10. <https://doi.org/10.1016/j.mimet.2015.01.015>.
56. Payne SH, Monroe ME, Overall CC, Kiebel GR, Degan M, Gibbons BC, Fujimoto GM, Purvine SO, Adkins JN, Lipton MS, Smith RD. 2015. The Pacific Northwest National Laboratory library of bacterial and archaeal proteomic biodiversity. *Sci Data* 2:e150041. <https://doi.org/10.1038/sdata.2015.41>.
57. Cao S, Chen Y, Yan Y, Zhu S, Tan Y, Wang T, Song Y, Deng H, Yang R, Du Z. 2021. Secretome and comparative proteomics of *Yersinia pestis* identify two novel E3 ubiquitin ligases that contribute to plague virulence. *Mol Cell Proteomics* 20:e100066. <https://doi.org/10.1016/j.mcpro.2021.100066>.
58. Hixson KK, Adkins JN, Baker SE, Moore RJ, Chromy BA, Smith RD, McCutchen-Maloney SL, Lipton MS. 2006. Biomarker candidate identification in *Yersinia pestis* using organism-wide semiquantitative proteomics. *J Proteome Res* 5: 3008–3017. <https://doi.org/10.1021/pr060179y>.
59. Pieper R, Huang ST, Parmar PP, Clark DJ, Alami H, Fleischmann RD, Perry RD, Peterson SN. 2010. Proteomic analysis of iron acquisition, metabolic and regulatory responses of *Yersinia pestis* to iron starvation. *BMC Microbiol* 10:e30. <https://doi.org/10.1186/1471-2180-10-30>.
60. Israeli O, Cohen-Gihon I, Aftalion M, Gur D, Vagima Y, Zauberman A, Levy Y, Zvi A, Chitlaru T, Mamroud E, Tidhar A. 2021. Novel RNA extraction method for dual RNA-seq analysis of pathogen and host in the early stages of *Yersinia pestis* pulmonary infection. *Microorganisms* 9:e2166. <https://doi.org/10.3390/microorganisms9102166>.
61. Li N, Hennelly SP, Stubben CJ, Micheva-Viteva S, Hu B, Shou Y, Vuyisich M, Tung C-S, Chain PS, Sanbonmatsu KY, Hong-Geller E. 2016. Functional and structural analysis of a highly-expressed *Yersinia pestis* small RNA following infection of cultured macrophages. *PLoS One* 11:e0168915. <https://doi.org/10.1371/journal.pone.0168915>.
62. Yan Y, Su S, Meng X, Ji X, Qu Y, Liu Z, Wang X, Cui Y, Deng Z, Zhou D, Jiang W, Yang R, Han Y. 2013. Determination of sRNA expressions by RNA-seq in *Yersinia pestis* grown in vitro and during infection. *PLoS One* 8:e74495. <https://doi.org/10.1371/journal.pone.0074495>.
63. Koo JT, Alleyne TM, Schiano CA, Jafari N, Latham WW. 2011. Global discovery of small RNAs in *Yersinia pseudotuberculosis* identifies *Yersinia*-specific small, noncoding RNAs required for virulence. *Proc Natl Acad Sci U S A* 108:E709–E717. <https://doi.org/10.1073/pnas.1101655108>.
64. Beauregard A, Smith EA, Petrone BL, Singh N, Karch C, McDonough KA, Wade JT. 2013. Identification and characterization of small RNAs in *Yersinia pestis*. *RNA Biol* 10:397–405. <https://doi.org/10.4161/rna.23590>.
65. Schiano CA, Koo JT, Schipma MJ, Caulfield AJ, Jafari N, Latham WW. 2014. Genome-wide analysis of small RNAs expressed by *Yersinia pestis* identifies a regulator of the Yop-Ysc type III secretion system. *J Bacteriol* 196: 1659–1670. <https://doi.org/10.1128/JB.01456-13>.
66. Qu Y, Bi L, Ji X, Deng Z, Zhang H, Yan Y, Wang M, Li A, Huang X, Yang R, Han Y. 2012. Identification by cDNA cloning of abundant sRNAs in a human-avirulent *Yersinia pestis* strain grown under five different growth conditions. *Future Microbiol* 7:535–547. <https://doi.org/10.2217/fmb.12.13>.
67. Schmühl C, Beckstette M, Heroven AK, Bunk B, Spröer C, McNally A, Overmann J, Dersch P. 2019. Comparative transcriptomic profiling of *Yersinia enterocolitica* O:3 and O:8 reveals major expression differences of fitness- and virulence-relevant genes indicating ecological separation. *mSystems* 4:e00239-18. <https://doi.org/10.1128/mSystems.00239-18>.
68. Nuss AM, Heroven AK, Waldmann B, Reinkensmeier J, Jarek M, Beckstette M, Dersch P. 2015. Transcriptomic profiling of *Yersinia pseudotuberculosis* reveals reprogramming of the Crp regulon by temperature and uncovers Crp as a master regulator of small RNAs. *PLoS Genet* 11:e1005087. <https://doi.org/10.1371/journal.pgen.1005087>.
69. Nuss AM, Beckstette M, Pimenova M, Schmühl C, Opitz W, Pisano F, Heroven AK, Dersch P. 2017. Tissue dual RNA-seq allows fast discovery of infection-specific functions and riboregulators shaping host-pathogen transcriptomes. *Proc Natl Acad Sci U S A* 114:E791–E800. <https://doi.org/10.1073/pnas.1613405114>.
70. Karlyshev AV, Oyston PCF, Williams K, Clark GC, Titball RW, Winzeler EA, Wren BW. 2001. Application of high-density array-based signature-tagged mutagenesis to discover novel *Yersinia* virulence-associated genes. *Infect Immun* 69:7810–7819. <https://doi.org/10.1128/IAI.69.12.7810-7819.2001>.
71. Mecas J, Bilis I, Falkow S. 2001. Identification of attenuated *Yersinia pseudotuberculosis* strains and characterization of an orogastric infection in BALB/c mice on day 5 postinfection by signature-tagged mutagenesis. *Infect Immun* 69:2779–2787. <https://doi.org/10.1128/IAI.69.5.2779-2787.2001>.
72. Darwin AJ, Miller VL. 1999. Identification of *Yersinia enterocolitica* genes affecting survival in an animal host using signature-tagged transposon mutagenesis. *Mol Microbiol* 32:51–62. <https://doi.org/10.1046/j.1365-2958.1999.01324.x>.
73. Flashner Y, Mamroud E, Tidhar A, Ber R, Aftalion M, Gur D, Lazar S, Zvi A, Bino T, Ariel N, Velan B, Shafferman A, Cohen S. 2004. Generation of *Yersinia pestis* attenuated strains by signature-tagged mutagenesis in search of novel vaccine candidates. *Infect Immun* 72:908–915. <https://doi.org/10.1128/IAI.72.2.908-915.2004>.
74. Ponnusamy D, Fitts EC, Sha J, Erova TE, Kozlova EV, Kirtley ML, Tiner BL, Andersson JA, Chopra AK. 2015. High-throughput, signature-tagged mutagenic approach to identify novel virulence factors of *Yersinia pestis* CO92 in a mouse model of infection. *Infect Immun* 83:2065–2081. <https://doi.org/10.1128/IAI.02913-14>.
75. Leigh SA, Forman S, Perry RD, Straley SC. 2005. Unexpected results from the application of signature-tagged mutagenesis to identify *Yersinia pestis* genes required for adherence and invasion. *Microb Pathog* 38: 259–266. <https://doi.org/10.1016/j.micpath.2005.02.004>.
76. Klein KA, Fukuto HS, Pelletier M, Romanov G, Grabenstein JP, Palmer LE, Ernst R, Bliska JB. 2012. A transposon site hybridization screen identifies galU and wecBC as important for survival of *Yersinia pestis* in murine macrophages. *J Bacteriol* 194:653–662. <https://doi.org/10.1128/JB.06237-11>.
77. Willcocks S, Huse KK, Stabler R, Oyston PCF, Scott A, Atkins HS, Wren BW. 2019. Genome-wide assessment of antimicrobial tolerance in *Yersinia pseudotuberculosis* under ciprofloxacin stress. *Microb Genom* 5:e000304. <https://doi.org/10.1099/mgen.0.000304>.
78. Senior NJ, Sasidharan K, Saint RJ, Scott AE, Sarkar-Tyson M, Ireland PM, Bullifant HL, Rong Yang Z, Moore K, Oyston PCF, Atkins TP, Atkins HS, Soyer OS, Titball RW. 2017. An integrated computational-experimental approach reveals *Yersinia pestis* genes essential across a narrow or a broad range of environmental conditions. *BMC Microbiol* 17:e163. <https://doi.org/10.1186/s12866-017-1073-8>.
79. Willcocks SJ, Stabler RA, Atkins HS, Oyston PF, Wren BW. 2018. High-throughput analysis of *Yersinia pseudotuberculosis* gene essentiality in optimised in vitro conditions, and implications for the speciation of *Yersinia pestis*. *BMC Microbiol* 18:e46. <https://doi.org/10.1186/s12866-018-1189-5>.
80. Yang ZR, Bullifant HL, Moore K, Paszkiewicz K, Saint RJ, Southern SJ, Champion OL, Senior NJ, Sarkar-Tyson M, Oyston PCF, Atkins TP, Titball RW. 2017. A noise trimming and positional significance of transposon insertion system to identify essential genes in *Yersinia pestis*. *Sci Rep* 7: e41923. <https://doi.org/10.1038/srep41923>.
81. Eichelberger KR, Sepúlveda VE, Ford J, Selitsky SR, Mieczkowski PA, Parker JS, Goldman WE. 2020. Tn-Seq analysis identifies genes important for *Yersinia pestis* adherence during primary pneumonic plague. *mSphere* 5:e00715-20. <https://doi.org/10.1128/mSphere.00715-20>.

82. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
83. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R, Teeling E. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
84. Hurst MRH, Beattie A, Altermann E, Moraga RM, Harper LA, Calder J, Laugraud A. 2016. The draft genome sequence of the *Yersinia entomophaga* entomopathogenic type strain MH96T. *Toxins* 8:e143. <https://doi.org/10.3390/toxins8050143>.
85. Moore RM, Harrison AO, McAllister SM, Polson SW, Eric Wommack K. 2020. Iroki: automatic customization and visualization of phylogenetic trees. *PeerJ* 8:e8584. <https://doi.org/10.7717/peerj.8584>.
86. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 43:e15. <https://doi.org/10.1093/nar/gku1196>.
87. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:e421. <https://doi.org/10.1186/1471-2105-10-421>.
88. Cokelaer T, Desvillechabrol D, Legendre R, Cardon M. 2017. “Sequana”: a set of snakemake NGS pipelines. *J Open Source Softw* 2:e352. <https://doi.org/10.21105/joss.00352>.
89. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>.
90. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
91. Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30:923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
92. Wiecek S, Combes F, Lazar C, Gai Gianetto Q, Gatto L, Dorffer A, Hesse AM, Couté Y, Ferro M, Bruley C, Burger T. 2017. DAPAR & ProStaR: software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* 33:135–136. <https://doi.org/10.1093/bioinformatics/btw580>.
93. Gai Gianetto Q, Wiecek S, Couté Y, Burger T. 2020. A peptide-level multiple imputation strategy accounting for the different natures of missing values in proteomics data. *bioRxiv*. <https://doi.org/10.1101/2020.05.29.122770>.
94. Gai Gianetto Q, Combes F, Ramus C, Bruley C, Couté Y, Burger T. 2016. Calibration plot for proteomics: a graphical tool to visually check the assumptions underlying FDR control in quantitative experiments. *Proteomics* 16:29–32. <https://doi.org/10.1002/pmic.201500189>.
95. Smyth GK. 2005. limma: linear models for microarray data, p 397–420. *In* Gentleman R, Huber W, Carey VJ, Irizarry RA, Dudoit S (ed), *Bioinformatics and computational biology Solutions using R and Bioconductor*. Springer, Heidelberg, Germany.