Serendipitous quirks of *Yersinia* biology catalyzed much of the research into the molecular basis of *Y. pestis* pathogenesis in the last three decades. For example, the finding that the addition of skim milk to the growth medium enhanced retention of virulence of *Y. pestis* cultured at 37°C (1) ultimately led to the discovery of type III secretion (2). The identification of the yersiniabactin siderophore system required for bacterial iron acquisition during infection (3, 4) was similarly prefaced by the observation that the ability of *Y. pestis* to form pigmented colonies on medium containing hemin was somehow linked to virulence (5, 6). A large fraction of current research on *Yersinia* virulence centers exclusively on the few virulence determinants discovered by these early efforts. To date, these studies have largely eclipsed more systematic attempts to define virulence determinants and other important factors contributing to severity of disease.

The limitations of previous iterations of whole-genome techniques are partly responsible for this focus. Prior genome-level studies of *Y. pestis* relied on acquisition of gene expression data under various *in vitro* and *in vivo* conditions (7–11). These results provide a deeper understanding of physiological adaptations employed by the bacterium in different environments; however, gene expression data have not proved to be a reliable predictor of virulence determinants (12–14). Attempts to circumvent this problem by directly measuring the functional contribution of *Y. pestis* genes to bacterial growth *in vivo* have been more successful but rely on measuring the fitness of a defined panel of mutants rather than implementing a truly unbiased genome-scale technique (15).

Recently, the combination of transposon mutagenesis and high-throughput sequencing has provided a more direct approach to functional genomics. This method has been assigned several...
names, including HITS (16), TraDIS (17), INSeq (18), and Tn-seq (19). Tn-seq employs deep sequencing to map and count transposon/genome junctions from a dense library of random transposon insertion mutants grown en masse under each condition of interest. The number of transposon insertions detected in a given gene reflects the prevalence of the corresponding mutants in the library following selection (20). This information can be used to estimate relative fitness for mutants in different growth environments.

Tn-seq is an increasingly popular tool in studies of bacterial pathogens (14, 16–31), although to date only five pathogens have been examined using animal infection as a selection regime (Streptococcus pneumoniae [23, 26], Haemophilus influenzae [16], Yersinia pseudotuberculosis [25], uropathogenic Escherichia coli [30], and Mycobacterium tuberculosis [31]). Here, we used Tn-seq to measure the fitness of nearly 200,000 unique Y. pestis insertion mutants during growth in vitro in rich culture medium and in mice following intravenous inoculation. We identified several genes contributing to bacterial fitness during infection that had not previously been implicated in virulence, including more than 30 genes with roles in nutrient import and metabolism. The majority of known virulence factors required for optimal bacterial growth during infection were also required in the Tn-seq data set, underscoring the suitability of this method in the context of infection studies.

RESULTS

Transposon mapping of the Yersinia pestis genome. We mutagenized the fully virulent Y. pestis strain KIM1001 (see Text S1 in the supplemental material) with a himar1-derived transposon to construct a bank of approximately 1.5 million independent insertion mutants. To select against mutants harboring insertions in genes required for growth in vitro, the insertion library was grown on TB, a complex medium adapted from tryptose blood agar base that provides a broadly permissive growth condition for insertion mutants. To select against mutants harboring insertions in genes required for growth in vivo, the insertion library was grown on TB, a complex medium adapted from tryptose blood agar base that provides a broadly permissive growth condition for insertion mutants.

To select against mutants harboring insertions in genes required for growth during mammalian infection, we performed intravenous infections of three C57BL/6 mice with 2.3 × 10^7 CFU of the insertion library, which resulted in a population of ~10^6 bacteria in each spleen 1 h postinfection. This infection route preserves library diversity by circumventing bottlenecks imposed by peripheral routes of infection. Bacteria were recovered from spleens 40 h postinfection by plating on TB agar. Bacterial populations in the spleen at the time of harvest averaged 1.63 × 10^9 CFU/spleen (ranging from 1.4 × 10^9 to 2.1 × 10^9). This population expansion is in excess of 1,000-fold, large enough to permit detection of modest reductions in fitness even in the absence of bacterial killing by host defenses.

Following selection, we mapped the transposon insertions present in the bacterial populations recovered from the in vitro and in vivo expansions via selective amplification of transposon-genome junctions and Illumina sequencing, using methods similar to those previously described (see Text S1 in the supplemental material). The himar1 transposon derivative used for mutagenesis inserts at TA dinucleotides (32, 33). Approximately 69% (187,780 of 271,354) of TAs in the Y. pestis KIM genome harbored inser-
tions, an average density of 39 insertions per kilobase. This high density of insertions allowed fine-scale mapping of functional genetic units, including intergenic regions (Fig. 1A). The density of mutations on pMT1, pCD1, and pPCP1, the three plasmids present in \textit{Yersinia pestis} KIM, was similar to that on the chromosome (78% on pMT1, 76% on pPCP1, and 63% on pPCP1, compared to 69% on the chromosome) (Fig. 1B). Although substantial large-scale heterogeneity in insertion density has been observed after similar transposon mutagenesis of other pathogens (21, 34), insertions in our library were distributed relatively evenly across the chromosome (Fig. 1A).

Identifying genes required for optimal growth \textit{in vitro}. To identify genes required for optimal growth of \textit{Y. pestis} on rich medium, we analyzed the full set of \textit{Y. pestis} KIM open reading frames (ORFs) annotated in the NCBI database by two distinct mathematical approaches for measuring the effect of insertions on fitness.

In the first approach, we estimated the relative fitness \((W)\) of mutants harboring insertions in each chromosomal ORF, using an analysis similar to that described by van Opijnen and Camilli (35). In the context of evolutionary biology, relative fitness is defined as the ratio of the rates of population expansion for the two genotypes being compared. Mutants with relative fitness of \(1.10\) are defective in growth or survival relative to the wild type, while mutants with relative fitness of \(1.0\) are more fit than the wild type. To test the statistical significance of the observed changes in fitness for each gene, we employed a nonparametric approach based on a reference distribution created by resampling insertion frequencies in TA sites in nonessential ORFs (defined as those with relative fitness between 0.95 and 1.05) and plotted relative fitness versus

![Selection of chromosomal \textit{Y. pestis} genes during growth \textit{in vitro} on rich medium. (A) The relative fitness of mutants in each chromosomal ORF plotted against significance (calculated by resampling; see Text S1 in the supplemental material). Genes harboring no sequenced insertions in the central 90% of the ORF (relative fitness = 0) are shown in the box at the left of the plot. (B) Relative fitness (histogram) correlates well to essentiality, as predicted by the hidden Markov model recently published by DeJesus and Ioerger (36) (red, essential; yellow, growth-deficient; green, nonessential). Genes harboring no sequenced insertions in the central 90% of the ORF (relative fitness = 0) are shown at the left of the plot. (C) Three examples of genes with modest reductions in relative fitness that were predicted to be essential by the hidden Markov model. Many genes in this class contain groups of consecutive TA sites that do not harbor insertions (red boxes), suggesting domain architecture. For example, the apparently essential N terminus of the \textit{rne} gene is highly homologous to the essential \textit{E. coli} gene of the same name (65, 66), while the evidently nonessential C terminus is highly divergent from the \textit{E. coli} homologue, as has been reported for several other species (67); also shown is the apparently essential C-terminal DNA-binding domain of the \textit{rcsB} gene, a transcriptional activator with multiple roles, including small RNA production and cell division (68). Inserts are shown by the Integrated Genomics Viewer software tool (63, 64) on a log scale.
significance (Fig. 2A) (for method details, see Text S1 in the supplemental material; for complete results, see Table S1).

A drawback of this first approach is its insensitivity to ORFs containing multiple functional domains with differing contributions to fitness. To address this problem, we used the hidden Markov model (HMM) recently developed by DeJesus and Loegering (36). This method classifies genes as essential, growth disadvantaged, nonessential, or growth advantageous. A complete list of classification of ORFs by the HMM algorithm is available in Table S2 in the supplemental material, and the metabolic pathways most commonly represented by this list of genes, as mapped by the Kyoto Encyclopedia of Genes and Genomes (37), are summarized in Table S3. Relative fitness values were good predictors of HMM classification of ORFs by the HMM algorithm is available in Table S2 in the supplemental material; for complete results, see Table S1).

In some cases, these results are consistent with known domain structure; this algorithm is particularly sensitive to consecutive TA sites without insertions and therefore correctly classifies ORFs containing subgenic essential domains (Fig. 2C). However, other genes in this group have insertion patterns that render their classification suspect (discussed fully in Text S2 in the supplemental material). These include genes with relatively short (<20) runs of insertion-free TAs, which are less likely to represent functional protein domains, and runs of insertion-free TAs in unusually AT-rich sequences, which may be more difficult to sequence due to base composition bias in library preparation and/or Illumina sequencing (38).

The HMM analysis also identifies some plasmid genes as essential, including genes required for plasmid replication. The apparent selection on these genes does not stem from any essential function of the plasmids but instead reflects the biological impossibility of propagating insertions that prevent the replication of their host genetic element. The plasmid pCD1, which encodes the Y. pestis type III secretion system (T3SS), harbors a second class of genes that the HMM scores as essential or growth deficient. The majority of these are involved in regulating type III secretion (T3S) activity (e.g., yscM, lcrGVH, and lcrD), and insertions in these genes are underrepresented because uncontrolled expression of the T3SS at 37°C in vitro inhibits growth of Y. pestis (e.g., see reference 39). Curiously, the pattern of T3SS genes selected in vitro deviates somewhat from current models of T3S regulation. In particular, the regulatory gene lcrE (sometimes called yopN) was not subject to significant selection under our in vitro conditions (falling into the nonessential HMM class), and we observed unexpectedly strong selection against yopB mutants in vitro. We do not conceive a simple explanation for these unexpected phenotypes and suggest that this regulatory scheme may require further investigation.

Identifying genes selected during infection. Because we were particularly interested in identifying genes required for optimal growth and survival during infection, we calculated the relative fitness of insertion mutants from the in vivo Tn-seq data set. To determine whether the period of growth in vivo significantly changed the abundance of insertion mutants for each gene, we developed a nonparametric statistic based on a reference distribution created by permuting insertion frequencies in the relevant ORF from both the in vitro and in vivo data sets (see Text S1 in the supplemental material) and plotted the relative fitness versus significance (Fig. 3A). A complete table of these values is provided in Table S1. Rare genes with substantial reductions in in vivo relative fitness of insertion mutants from the in vivo data set are represented by black circles. For each gene, relative fitness of insertion mutants in vivo is given with respect to the fitness of a wild-type strain. Significance is calculated by permuting the number of insertions sequenced at each TA in the gene between the two data sets (see Text S1 in the supplemental material). Genes selected in vivo appear in the upper-left quadrant, while any genes that harbor more insertions in vivo than in vitro should track to the upper-right quadrant. Genes harboring no sequenced insertions in vivo in the central 90% of the ORF (relative fitness = 0) are shown in the box at the left of the plot. (B) Genes experiencing strong selection (psn and brnQ) harbor many fewer insertions after growth in vivo. The same effect is present but less pronounced in genes experiencing moderate levels of selection (ptsG). Insertions are shown by the Integrated Genome Viewer tool (63, 64) on a log scale.

FIG 3 Comparison of Tn-seq data sets to find chromosomal genes required during mammalian infection. (A) Volcano plot for identification of chromosomal genes selected in vivo. Each gene was compared between two in vitro data sets (filled pale blue squares) or between an in vitro data set and the in vivo data set (gray circles). Genes experiencing significant selection in vivo are represented by black circles. For each gene, relative fitness of insertion mutants in vivo is given with respect to the fitness of a wild-type strain. Significance is calculated by permuting the number of insertions sequenced at each TA in the gene between the two data sets (see Text S1 in the supplemental material). Genes selected in vivo appear in the upper-left quadrant, while any genes that harbor more insertions in vivo than in vitro should track to the upper-right quadrant. Genes harboring no sequenced insertions in vivo in the central 90% of the ORF (relative fitness = 0) are shown in the box at the left of the plot. (B) Genes experiencing strong selection (psn and brnQ) harbor many fewer insertions after growth in vivo. The same effect is present but less pronounced in genes experiencing moderate levels of selection (ptsG). Insertions are shown by the Integrated Genome Viewer tool (63, 64) on a log scale.

In particular, the regulatory gene lcrE (sometimes called yopN) was not subject to significant selection under our in vitro conditions (falling into the nonessential HMM class), and we observed unexpectedly strong selection against yopB mutants in vitro. We do not conceive a simple explanation for these unexpected phenotypes and suggest that this regulatory scheme may require further investigation.

Identifying genes selected during infection. Because we were particularly interested in identifying genes required for optimal growth and survival during infection, we calculated the relative fitness of insertion mutants from the in vivo Tn-seq data set. To determine whether the period of growth in vivo significantly changed the abundance of insertion mutants for each gene, we developed a nonparametric statistic based on a reference distribution created by permuting insertion frequencies in the relevant ORF from both the in vitro and in vivo data sets (see Text S1 in the supplemental material) and plotted the relative fitness versus significance (Fig. 3A). A complete table of these values is provided in Table S1. Rare genes with substantial reductions in in vivo relative fitness of insertion mutants from the in vivo data set are represented by black circles. For each gene, relative fitness of insertion mutants in vivo is given with respect to the fitness of a wild-type strain. Significance is calculated by permuting the number of insertions sequenced at each TA in the gene between the two data sets (see Text S1 in the supplemental material). Genes selected in vivo appear in the upper-left quadrant, while any genes that harbor more insertions in vivo than in vitro should track to the upper-right quadrant. Genes harboring no sequenced insertions in vivo in the central 90% of the ORF (relative fitness = 0) are shown in the box at the left of the plot. (B) Genes experiencing strong selection (psn and brnQ) harbor many fewer insertions after growth in vivo. The same effect is present but less pronounced in genes experiencing moderate levels of selection (ptsG). Insertions are shown by the Integrated Genome Viewer tool (63, 64) on a log scale.
fitness that do not reach significance (Fig. 3A, lower left of plot) are associated with large growth defects both in vitro and in vivo (see Text S1).

The PCR amplification and Illumina sequencing steps required for Tn-seq introduce variability into the measured relative abundance of each mutant in the library, leading to differences in measured abundances between libraries that may reflect either biologically meaningful differences or noise introduced during sample processing. To address this problem, we independently prepared, sequenced, and analyzed a second sample of the inoculum library (see Table S1 in the supplemental material). The comparison of these two technical replicates allowed us to directly observe the extent of artifactual variation introduced by sample processing (Fig. 3A). ORFs in the in vivo dataset that met a strict cutoff of 0.05 and that fell entirely outside the range of variability observed between inoculum library replicates were considered candidates for significantly selected genes.

Fifty-three chromosomal genes met these criteria for signifi-
JG150A (w.t.), JG150A/H9004 brnQ

selected genes we constructed an in-frame deletion of (40). To validate this observed
trachomatis a role in valine uptake during intracellular growth of (Fig. 3A and B). A homologous importer has been reported to play
Y. pestis for

The putative branched-chain amino acid importer encoded by brnQ (y0981) was among the most strongly
selected genes in vivo, despite its apparent dispensability in vitro (Fig. 3A and B). A homologous importer has been reported to play a role in valine uptake during intracellular growth of Chlamydia trachomatis (40). To validate this observed in vivo fitness defect, we constructed an in-frame deletion of brnQ in the virulent strain KIM1001. Subcutaneous infection with 10^5 (Fig. 4A) or 10^5 (see Fig. S1 in the supplemental material) CFU of the KIM1001ΔbrnQ mutant failed to visibly sicken mice, while all mice infected with wild-type KIM1001 died within 8 days. Sixty percent of mice succumbed to infection with 10^5 CFU of this mutant (see Fig. S1), indicating a roughly 1-millionfold increase in mean lethal dose in the absence of brnQ (since the 50% lethal dose [LD_{50}] of wild-type KIM1001 is <50 CFU by subcutaneous infection [41]). Complementation with the brnQ gene driven by its native promoter on the low-copy-number plasmid pSP6 restored virulence (Fig. 4A). Four days after subcutaneous infection with 1,000 CFU KIM1001ΔbrnQ, no Y. pestis could be recovered from the homogenized spleens of mice (n = 5; limit of detection, <10^1 CFU/spleen).

To determine if infection with brnQ mutants could protect against subsequent infection, mice surviving infection with KIM1001ΔbrnQ (a dose of 10^5 CFU) were challenged 28 or 30 days after the initial infection with 10^3 CFU of wild-type KIM1001. Fifteen of the sixteen mice previously infected with KIM1001ΔbrnQ survived the challenge without visible symptoms, while all naive mice succumbed (Fig. 4A).

Notably, brnQ mutants were not underrepresented in the in vitro Tn-seq data set. To investigate this phenomenon, we reconstructed the in-frame deletion of brnQ in the attenuated strain JG150A, allowing us to investigate growth of the mutant under biosafety level 2 (BSL2) conditions. In accordance with the in vitro Tn-seq data, JG150AΔbrnQ grew as well as the wild-type control in the rich medium TB and had only a mild growth defect in PMH2, a defined medium commonly used for culturing Y. pestis (42) (Table 2; see also Fig. S2A and B in the supplemental material). Since brnQ is annotated as an ABC transporter that imports the branched-chain amino acids leucine, isoleucine, and valine, we formulated a new defined medium for culturing Y. pestis to probe for a deficiency in branched-chain amino acid acquisition. Serum nutritional medium (SNM) contains levels of magnesium,
TABLE 2 The brnQ deletion mutant grows poorly at low concentrations of branched-chain amino acids

<table>
<thead>
<tr>
<th>Growth medium</th>
<th>JG150A</th>
<th>JG150AΔbrnQ</th>
<th>Relative fitness (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TB</td>
<td>0.540 ± 0.023</td>
<td>0.5301 ± 0.044</td>
<td>0.973</td>
</tr>
<tr>
<td>SNM</td>
<td>0.3987 ± 0.075</td>
<td>0.0681 ± 0.023</td>
<td>0.171</td>
</tr>
<tr>
<td>SNM + 2 mM Leu, Ile, Val</td>
<td>0.4063 ± 0.002</td>
<td>0.3253 ± 0.001</td>
<td>0.801</td>
</tr>
</tbody>
</table>

* Growth rates of the wild-type JG150A strain and the deletion mutant JG150AΔbrnQ were calculated from cultures at mid-log phase. Means and standard errors from a minimum of two independent experiments are shown.

* In these experiments, relative fitness was calculated by taking the ratio of the JG150AΔbrnQ growth rate and the JG150A growth rate.

sodium, glucose, and all amino acids equivalent to those found in serum (see Text S1).

The growth rate of JG150AΔbrnQ cultured in SNM is less than one-fifth of the growth rate for wild-type JG150A in the same medium (relative fitness, 0.171). As in infection studies, complementing brnQ with a wild-type copy of the gene under its native promoter in trans restored growth of the mutant to wild-type levels (Fig. 4B). Growth could also be restored by supplementing SNM with 2 mM leucine, isoleucine, and valine—concentrations similar to those in PMH2 (Fig. 4B and Table 2). Interestingly, supplementing the medium with isoleucine and valine in the absence of leucine did not fully restore growth (see Fig. S2C in the supplemental material), despite the presence of the apparently functional alternative leucine uptake system LS in Y. pestis KIM (livKHMGF; y0422-y0426) (43) that has been reported to be expressed in vivo (7).

Glucose import by PtsG provides a competitive advantage to Y. pestis during infection. Recent work screening a defined panel of Y. pestis mutants implicated the glucose importer encoded by ptsG in bacterial fitness during infection (15). As with brnQ, the Tn-seq data set identified ptsG as significantly selected in vivo, but insertions in ptsG caused a much more modest reduction in relative fitness (0.693) (Table 1 and Fig. 3A and B). To assess the contribution of ptsG to virulence, we subcutaneously infected C57BL/6 mice with 1,000 CFU of the in-frame deletion strain KIM1001ΔptsG. The course of disease with KIM1001ΔptsG did not differ significantly from that with wild-type KIM1001 (Fig. 5). We therefore considered the hypothesis that the decrease in growth caused by inactivation of ptsG is insufficient to affect the outcome of infection but that ptsG confers a competitive fitness advantage to Y. pestis during growth in mammals.

We infected BALB/C and C57BL/6 mice intravenously with 10^4 CFU of KIM1001 and KIM1001ΔptsG in a 1:1 mixture. Spleens were homogenized 2 and 42 h postinfection, and the proportion of ptsG mutants in the spleen was determined at each time point. The relative frequency of KIM1001ΔptsG mutants declined significantly over the course of the experiment (Fig. 6A). On average, the proportion of KIM1001ΔptsG mutants in the spleen decreased from 0.434 at 2 h postinfection (i.e., 43.4% mutants) to 0.039 (3.9%) at 42 h postinfection. Assuming constant exponential growth rates during the infection, this change in frequency corresponds to a relative fitness of 0.8 for KIM1001ΔptsG relative to wild-type KIM1001, a value close to that determined from the Tn-seq experiments (Table 1).

We were able to recapitulate this phenotype in vitro in the defined serum-like medium SNM. Five replicate cultures were inoculated with equal numbers of the attenuated (T3SS-deficient) strains JG102 and JG102ΔptsG. After 42 h of continuous log-phase growth, the proportion of ptsG mutants in the culture had decreased from 0.5 to an average of 0.034 (Fig. 6B). This reduction in frequency represents a relative fitness of 0.794 for the ptsG mutant relative to the control strain, JG102, in the same medium, very similar to the value of 0.8 observed in vivo. No growth defect was detected in SNM when the JG102ΔptsG mutant was complemented with a low-copy-number plasmid carrying ptsG under control of its native promoter (see Fig. S3 and Text S1 in the supplemental material).

Because PtsG is annotated as a glucose importer, we tested whether glucose is required for the competitive advantage provided by ptsG. No fitness defect was detected for the ptsG mutant grown in the absence of glucose (Fig. 6B), consistent with the apparently neutral phenotype of ptsG insertion mutants in the in vitro Tn-seq data set (Fig. 3B). In contrast, the proportion of ptsG mutants in cultures containing glucose decreased on average from 0.592 to 0.266 (Fig. 6B). This decrease corresponds to a relative fitness of 0.92 for JG102ΔptsG relative to wild-type JG102 in this rich culture medium, supporting the hypothesis that ptsG provides a competitive advantage to Y. pestis during infection by permitting or increasing glucose import.

Intercellular complementation in vivo: genes required for siderophore-mediated iron acquisition. Growth in vivo strongly selected against insertions in the ptsG gene (y2404), which encodes the outer membrane TonB-dependent yersiniabactin importer (44), and in ybtA (y2398), the regulator required for expression of ptsG and other genes in the yersiniabactin uptake system (45) (Table 1 and Fig. 3A and B). This agrees with an extensive body of literature describing the requirement for yersiniabactin-mediated iron acquisition for optimal growth under iron-limiting conditions (44, 46), including in vivo (4, 6, 47).

However, the genes required for biosynthesis of the yersiniabactin siderophore (Ybt) were not selected in vitro or in vivo (see Table S4 in the supplemental material). This apparent contradiction almost certainly results from intercellular complementation. Extracellular yersiniabactin has been reported to complement the growth of mutants deficient in yersiniabactin biosynthesis even in low-iron environments (44). Because our experiments were performed at a relatively high bacterial density and the vast majority of the bacterial population remains functionally wild type with...
FIG 6  ptsG provides a competitive advantage in the presence of glucose. (A) A 10^4-CFU dose of wild-type KIM1001 and KIM1001ΔptsG (mixed 1:1) was injected intravenously into BALB/C or C57BL/6 mice. The proportion of KIM1001ΔptsG mutants in the population of Y. pestis in the spleen was determined 2 and 42 h postinfection by plating on TB supplemented with 2.5 mM CaCl_2, 1% glucose, and 50 µg/ml tetrazolium red. (B) The attenuated strains JG102 and JG102ΔptsG were mixed 1:1 in the rich medium TB, in TB supplemented with 5 mM glucose, and in the defined medium SNM (n = 5 for each condition). The proportion of JG102ΔptsG in each culture was measured before and after 42 h of log-phase growth.

respect to siderophore production, the concentration of extracellular yersiniabactin is likely high enough to provide iron to mutants deficient in Ybt synthesis.

Native Y. pestis plasmids. All three of the native plasmids in Y. pestis KIM1001 have been shown to have virulence functions in mice. To measure the contribution of plasmid genes to bacterial fitness during infection, we calculated the in vivo relative fitness and significance for each annotated ORF on the plasmids as described above for chromosomal genes.

pCD1 and genes of the type III secretion system. The type III secretion system encoded on the 70-kb plasmid pCD1 is essential for virulence of Y. pestis. We observed very strong reductions in relative fitness of mutants defective for many T3S genes (see Table S4 in the supplemental material). For example, yscF, which encodes the major structural subunit of the needle, was classified as strongly selected in vivo (relative fitness = 0; see Table S4). lcrV, yopB, and yopD, encoding the translocon apparatus, also had large reductions in relative fitness (0, 0.55, and 0.69). Among the secreted effectors, insertions in the genes encoding a GTPase-activating enzyme (yopE), a tyrosine phosphatase (yopH), and a regulatory effector (yopK) suffered clearly significant reductions in relative fitness (Fig. 7; see also Table S4). We did not detect a reduction in fitness for insertions in the effector-encoding genes yopM (encoding a possible inhibitor of inflammasome activation), ypkA (encoding a serine/threonine kinase), or yopT (encoding a cysteine protease). Insertions in yopJ (encoding an acetyltransferase) showed weak but not significant positive selection. These results are consistent with previous reports that some effectors are crucial while others appear to contribute very modestly or not at all to virulence (48–53).

pMT1. No genes on pMT1 experienced significant selection in vivo. pMT1 carries a gene encoding a phospholipase D, ymt (Y1069), that was initially thought to contribute to virulence in mice; however, the lack of apparent selection on ymt is in accordance with more recent reports that ymt primarily promotes bacterial survival in the flea vector (54) rather than enhancing virulence in mammals (55). pMT1 also carries a gene, caf1 (Y1100), encoding the fraction 1 capsule protein, which produces abundant pili loosely associated with the cell surface. As with ymt, we did not detect significant selection of caf1 in vivo.

pPCP1. None of the five annotated ORFs on pPCP1 experienced significant selection in vivo. pPCP1 encodes a single characterized virulence factor, the plasminogen activator encoded by pla, which plays a crucial role in progression to systemic infection following subcutaneous or intradermal infection (41, 56) but is less important in pulmonary infection (57) and has little effect following intravenous infection (58). Failure of pla mutants to show a fitness deficit in the Tn-seq data set is thus not surprising.

FIG 7  Many genes on the type III secretion plasmid pCD1 undergo strong selection during infection. Each gene on pCD1 was compared between two in vitro data sets (filled pale blue squares) or between an in vivo data set and the in vivo data set (gray circles). Genes experiencing significant selection in vivo are represented by black circles. Genes experiencing significant selection in vivo are represented by black circles. Genes experiencing no sequenced insertions in vivo in the central 90% of the ORF (relative fitness = 0) are shown in the box at the left of the plot. The effector proteins of the type III secretion system undergo variable levels of selection (red, yopH; yellow, yopE; dark blue, yopK; orange, ypkA; purple, yopM; pink, yopT; green, yopJ), while the ATPase (tan, yscN) and structural components (tan, yscF) undergo strong selection.
DISCUSSION

We detected a substantial fraction of the genes previously implicated in the virulence of *Y. pestis* (see Table S5 in the supplemental material). Some of the virulence genes not detected have been shown to be important via subcutaneous but not intravenous infection (e.g., *pla* and *yadBC*), while others may well have this property (*ail*, *caf1*, and *psa*; see Table S5). Alternatively, intercellular complementation may mask the requirement for *caf1* and/or *psa*, since both of these genes encode highly abundant secreted proteins. Our results are also generally in agreement with the recent study by Crimmans et al., which reports Tn-seq data for the closely related species *Y. pseudotuberculosis* following intravenous infection (25). As in our study, these investigators found crucial roles in vivo for purine biosynthesis (*purM, purH, purC*, and *purD*), aromatic amino acid biosynthesis (*aroA* and *aroE*), and LPS modification (*arnD*) (25). The remaining differences in our results may reflect biological differences between these two organisms, which differ in requirements for growth in vitro and cause very different diseases, but also reflect the increased sensitivity and resolution of our experiments, which employed a 100-fold-larger library.

A comparison of the Tn-seq data set to the in vitro competition data for a defined panel of *Y. pestis* mutants recently published by Pradel et al. (15) reveals several genes with reduced fitness in both experiments. Interestingly, however, there remains considerable disagreement between the candidates identified by Pradel et al. and results of the Tn-seq approach described here. In particular, Pradel et al. found that *ptsG* mutants are attenuated in a subcutaneous single-infection model, whereas we found that *ptsG* mutants remained fully virulent. It is possible that the low dose used by Pradel et al., which they report as 10 CFU, may explain this discrepancy.

Many genes experiencing significant in vivo selection during the Tn-seq experiment have roles in nutrient import or biosynthesis. The widespread use of rich media containing nutritional conditions not found in the host in both clinical and research settings may partially explain why these genes are often overlooked as therapeutic targets. As an illustration of this principle, the poor fitness of *brnQ* mutants in vivo cannot be detected on the rich medium TB but is readily apparent in media specifically designed to mimic nutritional conditions in serum. As illustrated by the strong immunity induced by infection with our *brnQ* mutant, importers required for nutrition in vivo may also be useful for the rational design of strains to be used as live vaccines.

Even in media recapitulating the in vivo environment, a growth defect in vitro does not always translate to reduced virulence. The behavior of *brnQ* mutants versus that of *ptsG* mutants makes clear the important distinction between fitness and virulence: the detection of reduced relative fitness by Tn-seq does not necessarily imply attenuation of the corresponding mutant. In *vivo* fitness and virulence can be distinct properties, especially when the fitness reduction is small.

With a few exceptions, Tn-seq studies reported to date present the observed effect of mutants in a given gene either as a ratio of the raw frequencies (e.g., see reference 22) or the log base 2 of this ratio (e.g., see reference 21). As we have done here, van Opipien and Camilli (35) use the metric of relative fitness. We suggest the adoption of this measure as a standard for the presentation of Tn-seq results. The ratio of raw frequencies yields extreme values that can be misleading. Log base 2 of these ratios is proportional to relative fitness, but the constant of proportionality will differ depending on experimental details, and hence this metric may not be directly comparable among studies. Relative fitness suffers from neither of these complications and has several additional advantages: most importantly, it is readily interpreted biological meaning and its congruence with the formal definitions for the synergy or independence of mutational effects. The latter is especially useful in genetic interaction studies, in which libraries constructed in specific mutant backgrounds are to be compared with libraries constructed in the wild-type parent.

The wealth of available data on *Yersinia pestis* pathogenesis enables the critical evaluation of the Tn-seq method in the context of infection studies. For example, the intercellular complementation of the yersiniabactin siderophore system suggests that intercellular complementation may be a common phenomenon in Tn-seq and related methods. In support of this conclusion, Price et al. observed strong intercellular complementation of the *Y. pestis* T3SS in lung tissue when they applied an alternate technique for screening transposon mutant libraries (TraSH) (59). The pattern we observe in the yersiniabactin system—a cluster of genes devoted to synthesis of a secondary metabolite not under selection, coupled to a transporter under strong selection—may be a signature for intercellular complementation by a highly diffusible small molecule. In the context of pathogenesis, a lack of an observed fitness deficit in Tn-seq studies cannot be interpreted as definitive evidence that the gene in question would not be important, or even essential, if the mutant were in pure culture.

The requirement for a large inoculum to provide good coverage of the genome is perhaps the most important limitation of the Tn-seq method. One of the most important properties of *Y. pestis* is its ability to produce systemic infection following intradermal inoculation with a few bacteria; however, this problem cannot be practically addressed by Tn-seq due to bottleneck effects that reduce diversity. As the work by Pradel et al. illustrates, a genomic-scale approach to identifying the genes required for infection via a peripheral route would be practicable only with an ordered array library orders of magnitude smaller than the random library used here and even then would require a large number of mice to limit inoculum size and overcome mouse-to-mouse variation (15). Nonetheless, as we have shown here, even very large inocula (in our case, more than 2 million LD<sub>50</sub>s) can yield a wealth of data that accurately reflect the underlying biology established over many years for *Y. pestis*. The *in vivo* application of Tn-seq is clearly an approach of extraordinary power that should greatly accelerate the understanding of many pathogens. Development of adequate libraries may be a significant obstacle in many species, but effort devoted to overcoming this obstacle is very likely to be worthwhile investment.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Strains used in this study and details of strain construction are presented in Text S1 in the supplemental material. Unless otherwise specified, *E. coli* strains were cultured in Luria broth (LB) and *Y. pestis* strains in the rich medium TB, prepared to maximize plating efficiency as previously described (60). Additional experiments were performed in the defined media PMH2 (42) or SNM (see Text S1). All *Y. pestis* cultures were supplemented with 2.5 mM CaCl<sub>2</sub> to suppress type II secretion. Where appropriate, media were supplemented with 100 µg/ml ampicillin, 25 µg/ml zeocin, or 25 µg/ml diaminopimelic acid.
ACKNOWLEDGMENTS

This work was supported by the NIH/NAID grant 5U01AI078073 to S.L. and by unrestricted research funds from the University of Massachusetts Medical School to I.D.G.

We thank Christopher Sassetti, Jeff Gawronski, and Brian Akerley for sharing their tremendously helpful Tn-seq expertise and Alice Li and Jarukit Ed Long for their technical advice. We also thank Thomas Loeger for the hidden Markov model, Ellen Kittler for Illumina sequencing support, Dieter Schifferli and Kenan Murphy for generously providing genetic tools, and Christina Baer for critical reading of the manuscript.

REFERENCES

28.7.e27.


Gallagher 21.

Khatiwara 24.

van Langridge 19.


Brodsky IE, Medzhivot R. 2006. Reduced secretion of YopJ by Yersinia


