Supporting Information Appendix

For: Genome-wide screen identifies host colonization determinants in a bacterial gut symbiont

J. Elijah Powell, Sean P. Leonard, Waldan K. Kwong, Philipp Engel, Nancy A. Moran

CONTENTS: **Page**

Materials and Methods

The following Supporting Information details experiments performed to develop and optimize a transposon mutagenesis system for *Snodgrassella alvi* strain wkB2, to show that mutagenized clones are experimentally viable, and to assess the impact of mutations on biofilm formation and gut colonization. Additionally, the bioinformatics workflows used to analyze the Tn-seq and RNA-seq results are detailed here.

Bacterial strains and culture conditions

Strains and plasmids used are listed in Table S2. Unless otherwise noted, all cultures were grown at 35°C in a 5% CO_2 environment. Additionally, *Escherichia coli* β2163 (1) with plasmid pBT20 (2) was grown in LB broth or on LB agar supplemented with ampicillin (100 μg/mL), gentamicin (30 μg/mL, for plasmid maintenance) and 0.3 mM diaminopimelic acid (DAP) (Sigma-Aldrich, St. Louis, MO, USA) (or without DAP for counterselection). *S. alvi* strain wkB2 (3) was grown *in vitro* on heart infusion agar (Hardy Diagnostics, Santa Maria, CA, USA) with 5% sheep's blood (HIA + 5% SB). For liquid culture, *S. alvi* wkB2 was grown in Insectagro DS2 (Corning, Manassas, VA, USA). For wkB2 transconjugant selection, gentamicin (12.5 μg/ml) was used.

Optimizing the conjugation of transposon-bearing *E. coli* with *S. alvi* wkB2

For the early experimental validation experiments, cultures of *S. alvi* strain wkB2 were grown from glycerol stocks for 2 days with tetracycline (25 μg/ml). The donor strain (DAP auxotrophic *E. coli* β2163) with plasmid pBT20 was grown overnight in 5 ml selective media, as above. This plasmid, which has been applied with high efficiency across a variety of bacterial lineages (2), was used to deliver the Himar1 transposon encoding gentamicin resistance. This plasmid also contains the Mariner-derived mar C-9 transposase that initiates transposition. The transposable element is inserted through a cut and paste mechanism wherein a single insertion event is expected per plasmid; hence, multiple insertion events are anticipated to be rare. Additionally, the transposon contains an outward facing Ptac promoter to minimize possible polar effects on downstream genes. An overnight culture was sub-cultured 1:100 in 5 ml of the selective media with ampicillin, gentamicin, and DAP, and grown for 3 hours to an OD_{600} of 0.26. One milliliter of this culture was centrifuged at 5,000 RPM for 5 minutes and washed twice with 1 ml sterile phosphate-buffered saline (PBS), pH 7.2. Meanwhile, we collected plated *S. alvi* wkB2 by flooding the plate with Insectagro and scraping the bacteria with a cell spreader. The recovered cells were placed in a microcentrifuge tube and measured at an OD_{600} of 6.47. The cells were then centrifuged for 5 minutes at 5,000 RPM and resuspended in 200 μl Insectagro. Twenty microliters of the donor *E. coli* strain were added to the 200 μl of *S. alvi* wkB2 cells and vortexed briefly. Fifty microliters of this ~1:10 conjugation mix was then spotted onto 4 nitrocellulose filters (Sartorious, Goettingen, Germany) placed onto TSA + 5% SB plates supplemented with 0.3 mM DAP. The conjugations were incubated for 16 hours. Then, each filter was rinsed with 1 ml PBS and the bacterial suspension was centrifuged at 5,000 RPM for 5 minutes and washed once in 1 ml PBS. The pellets were then suspended in 200 μl PBS each, combined, and 50 μl was spread on 9 × TSA + 5% SB plates containing 12.5 μg/ml gentamicin. Counterselection was performed against the donor strain by omitting DAP on these plates. Plates

were incubated for 5 days. Efficiency calculations were performed by plating 10-fold serial dilutions on non-selective TSA + 5% SB in triplicate and enumerating the average colony forming units (CFU) per ml. Conjugation/transposon integration frequency was calculated by dividing number of gentamicin resistant ($gent^R$) colonies per ml by the number of total CFU per ml. After culturing, 24 colonies were picked and passaged to fresh selective plates in ~2 cm^2 patches and cultured for ~48 hours. These patches of each passaged transconjugant clone were collected in 250 µl 25% glycerol in Insectagro, flash-frozen and placed at -80°C. Colonies were also struck to ampicillin plates to test for presence of the plasmid pBT20 backbone.

Semi-random PCR

Cells from different transposon mutants grown on selective media plates were collected with a sterile toothpick to 30 µl ddH_2O and lysed at 99°C for two minutes. One microliter of this lysate was used for semi-random nested PCR (srPCR) (2). PCR primers are listed in Table S3. PCR products were screened on a gel, and reactions from 14 clones were cleaned with Ampure beads (Agencourt, Beverly, MA, USA) and sent for unidirectional Sanger sequencing with the primer TnMseq at the Yale University DNA Analysis Facility. Returned sequences contained the junction of transposon and chromosomal sequence. The chromosomal components of the sequence were used as queries in BLASTN searches against the *S. alvi* wkB2 genome (Genbank accession CP007446). These insertion sites were examined by selecting ~500 bp of scaffold sequences from above and below the site of insertion and 1) analyzed for translation frames of putative open reading frames (ORFs) and 2) searched with BLASTX against the non-redundant database to find related protein-coding sequences. A predicted neutral transposon insertion mutant, *S. alvi* wkB2 clone c10, was selected from these analyses for inoculation experiments with microbiota-free bees; its transposon was found within a noncoding region (genome position 1,888,884).

Southern blot analysis

To examine whether the transposon was inserted into random single positions within the *S. alvi* genome, 13 clones were plated on TSA + 5% SB + 12.5 µg/ml gentamicin and grown for ~48 h at 37°C and 5% CO_2. Bacterial lawns were harvested, and high molecular weight DNA was extracted by using a bead beating, phenol-chloroform extraction method (4). From each clone, 3 µg of DNA was digested for 4 hours at 37°C with NcoI HF (New England Biolabs) under recommended reaction conditions. The reactions were inactivated with a hold at 80°C for 15 minutes, and then precipitated with ethanol and eluted in 15 µL of nuclease free water (HyClone, Logan, UT, USA). Digests were run on a 0.75% agarose gel for 19 hours at 44 V. This gel was then examined via Southern blotting with the Roche DIG kit (Roche Diagnostics, Indianapolis, IN, USA) and an insert-specific probe (see Table S3 for primers) to demonstrate single random insertion using methods and primers previously described (5). The insertion junction of these clones was also mapped to an *in silico* digest of the draft genome in order to verify band position in the resulting Southern blot (Fig. S4) using the srPCR protocol above. Two clones produced inconsistent results with the above method and were evaluated by sequencing 5 subcloned colonies of each: amplicons from clone 22 (which produced double bands in the Southern blot) and clone 39 (which did not produce resolvable sequence from srPCR) were subcloned into plasmid pGEMT and transformed into *E. coli* DH5α. Five random

clones of each were then sequenced to validate the insertion site. Of 5 colonies from a clone (clone 22) producing a double band in the Southern, all 5 matched the expected fragment inferred from mapping of sequence from srPCR, indicating that the double band observed for clone 22 resulted from an incomplete digest, as the upper band is the expected size for the anticipated lower fragment plus an adjacent one. For the second clone (clone 39), 3 of 5 sequences mapped to one insertion locus, and 2 of 5 mapped to a different but similar sized fragment, suggesting that this clone experienced a double insertion event.

The suicide plasmid pBT20 was transferred from *E. coli* into *S. alvi* wkB2 with a conjugation and/or transposon insertion efficiency of 2.9×10^{-6} (efficiency = colonies of $gent^R$ wkB2/total colonies wkB2). Transconjugants did not grow on selective media containing ampicillin (100 μg/ml) demonstrating the absence of the backbone plasmid and suggesting that the transposon had integrated into the genome of *S. alvi*. Southern blot analysis of NcoI HF (New England Biolabs) digestions from selected clones showed that most transconjugants carried a single transposon copy. These copies were located on genomic fragments of different sizes, suggesting random integration of the transposon (Fig. S4). This was confirmed by srPCR and sequencing of the transposon integration sites (5) (primers listed in Table S3). Sequences mapped to different genomic regions distributed over the genome of *S. alvi* (Fig. S5), and the fragment length expected from an *in silico* digest of the genome sequences corresponded to the actual fragment size observed on the Southern blot in 12 out 13 cases (Fig. S4). All integrations occurred at TA dinucleotides, consistent with the known mechanism of Himar1 transposon integration (6).

Bee experiments to determine mutagenized *S. alvi* wkB2 colonization efficiency

Microbiota-free newly emerged workers (NEWs) of the honey bee (*Apis mellifera*) were reared in laboratory in cup cages as previously described (7). One-day old bees (1NEWs) were fed sterile, filtered 0.5 M sucrose solution via a drip reservoir and gamma-irradiated bee bread (30 kGy) *ad libitum* in 4-well sections cut from a sterile 96-well culture plate. Control groups of non-inoculated 1NEWs were maintained, and other 1NEWs were marked immediately after emergence and placed back into their original outdoor hive (hive-reared). The experimental groups were passively inoculated at 1 day post emergence by placing a plate of a ~48-hour culture of neutral insertion clone c10 beneath their bee bread troughs for 24 hours. All bees were dissected at 8 days post emergence (7 days after exposure), and their midguts and ileums were homogenized with 50 μl of 10 mM $MgSO_4$. DNA was prepared from the homogenates with Qiagen DNeasy column prep kits (Qiagen, Valencia, CA, USA) using the Gram-negative protocol. Three microliters of the purified DNA (~50 ng) was used as template for amplification using primers for the *A. mellifera* elongation factor 1-alpha gene (EF1-a), universal bacterial 16S rRNA genes, *S. alvi*-specific 16S rRNA genes, and a clone c10 transposon-specific screen that had a forward primer situated on the transposon (TnMseq) and another on the chromosome ~1 kb from the site of insertion (BM10) (Table S3). All PCR products were screened on a 1% agarose gel run in 0.5% TBE at 95 V for 30 min, and visualized with UV after staining with ethidium bromide.

Multiple rounds of the inoculation experiment were conducted. The rate of colonization was assessed by preparing DNA from additional samples of the inoculated groups and screening them with only the clone c10-specific primers and the *A. mellifera* EF1-a primers as control

(Table S3). These screens were combined with previous ones for a final comparison of the colonization rates. For the analysis of the colonization efficiency via culturing, guts from bees of different treatment groups were dissected and homogenized at 8 days post emergence (7 days after exposure). Five microliters of each ileum was spread on TSA + 5% SB with and without 12.5 μ g/ml gentamicin. The resulting colonies from each ileum were scraped and resuspended in separate 100 µl aliquots of molecular grade H_2O, lysed at 99°C for 2 minutes, and tested with the transposon-specific screen as outlined above.

Clone c10 revealed a colonization rate of 75% of screened individuals at 7 days post-exposure (Table S4). Neither non-inoculated individuals nor hive-reared bees (which otherwise tested positive for *S. alvi)* tested positive for the mutant-specific screen. From 70% of the bees colonized with the mutant, we could isolate viable bacteria of clone c10 by plating homogenized ileums on media plates containing gentamicin 12.5 µg /ml (Table S5). Gentamicin-resistant bacteria were cultured from bees that were not colonized with clone c10. However, these bacteria formed morphologically distinct colonies and tested negative for the transposon-specific PCR, indicating that they were environmental bacteria with intrinsic gentamicin resistance.

Bee experiments to quantify bacterial colonization

Strains wkB2, c10 and c38 (an additional mutant clone with a transposon insertion that mapped to noncoding position number 1,902,630) were inoculated from frozen glycerol stocks on TSA + 5% SB and maintained at 37°C and 5% CO_2 for 48 hours. Plates were scraped with 1 ml of sterile PBS pH 7.2 and suspensions were diluted to OD_{600} ~2.0 for each strain. Standardized suspensions were diluted 1:100 in sterile-filtered 20% sucrose in PBS (w/v). Five microliters of this solution was fed to groups of 1NEWs that had been starved for 3 hours. Inoculum levels were ascertained by enumerating colonies from plated serial dilutions of the inoculum under the previously mentioned growth conditions. Each bee was fed approximately 8–10×10^3 CFU of a specific strain of *S. alvi*. Twelve bees were inoculated per condition and then split to 6 bees in 2 cup cages. A microbiota-free control group was also included in this experiment. Cup cages were maintained in previously described conditions for 7 days. At the end of this colonization period, bees in cup cages were pooled by condition, ileums of 5 bees per condition were harvested and the DNA was isolated as described in (8). DNA quality was checked by running diagnostic PCR for EF1-α, universal 16S rRNA, *S. alvi* 16S rRNA, and mutant-specific assays (Table S3) All samples tested as predicted; therefore, 1 μ l of a 1:10 dilution of DNA from each sample was used to assess *S. alvi* 16S rRNA gene copies by quantitative PCR (qPCR). Quantitative primers (Beta-1009-qtF, Beta-1115-qtR) specific to *S. alvi* 16S rRNA gene sequence were used (Table S3), and copy number was evaluated by running a 10-fold standard dilution set of the target sequence in a pGEM-T plasmid vector which was constructed using previously described methods (7). Reactions were run in triplicate, using qPCR reaction conditions as described in (7). Statistical differences between treatment groups were examined by one-way ANOVA followed by Tukey's honest significant difference (HSD) in JMP10 (SAS, Cary, NC, USA).

Bees inoculated with a defined dose of each of the 3 clones were dissected 7 days later, and their total *S. alvi* 16S rRNA gene copies were quantified. Numbers of *S. alvi* did not differ significantly between wild type wkB2 and the two tested transconjugant mutants (Tukey's HSD, $p > 0.1$) (Fig. S6). No detectable *S. alvi* were found in negative control samples (one-way

ANOVA, $p < 0.0001$). Reactions that either did not amplify or amplified later than 37 cycles (within the spread of the negative controls) were considered to have zero copies.

Construction of a saturated *S. alvi* wkB2 transconjugant library

Strain wkB2 was grown for 3 days on HIA + 5% SB with 25 μg/ml tetracycline. The culture was then restruck to 5 replicate plates of HIA + 5% SB with 25 μg/ml tetracycline for 2 days. *E. coli* β2163 + pBT20 was grown on 3 HIA plates with gentamicin, ampicillin, and DAP overnight. Plated cells were resuspended in 10 ml 1× sterile PBS, spun, and rinsed as detailed above. Final resuspensions were in 1 ml at OD_{600} ~15. Donor and recipient strains were combined in equal parts and 100 μl was spotted and briefly dried on ten conjugation filters placed on HIA + 5% SB + DAP plates. Conjugations were allowed to proceed for 19.5 hours. Bacteria from filters were resuspended in 30 ml 1× PBS. Cells were pelleted and rinsed twice with PBS to remove residual DAP for counterselection. The final pellet was resuspended in 3 ml PBS, from which 100 μl was spread to each of 29 HIA + 5% SB plates with 7.5 μg/ml tetracycline and 12.5 μg/ml gentamicin. The remaining 100 μl was used to make serial dilutions that were spread to HIA blood plates with tetracycline alone for efficiency calculations. Plates were incubated at for 5 days and resultant colonies were scraped to 50 ml Insectagro with 20% glycerol, vortexed to homogenize, distributed to 1.5 ml aliquots, flash-frozen in liquid nitrogen and stored at -80°C.

The library resulted in ~74,000 individual mutants of *S. alvi* wkB2 (Fig. 2), achieving an insertion efficiency of 4.3×10^{-5} CFU $gent^R$ wkB2/total CFU wkB2.

Tn-seq using mutant library from plates and in *A. mellifera* workers

To investigate essential genes in *S. alvi* wkB2 grown on plates, cells from 3 × 250 μl aliquots from the frozen mutant library were used. The cells were pelleted at 5,000 RCF for 10 minutes, rinsed with 1× PBS, and subjected to DNA extraction using the bead beating CTAB method as outlined in (8).

To investigate important genes in the bee ileum, pupae were removed from brood frames and allowed to develop as microbiota-free workers according to the procedure outlined in (8). Three replicate cohorts of 70–100 microbiota-free pupae were removed from the same hive on different days and inoculated on their first day post emergence. Each cohort was starved en masse for 3 hours, and then immobilized by placing at 4°C for 20–30 minutes. Individual bees were placed in feeding tubes (1.5 ml microcentrifuge tubes with the tip cut off). Each bee was fed 5 μl of the mutant library (OD_{600} = 24, suspended in Insectagro with 20% glycerol) and the cohort was placed back in the incubator (at 35°C) with sterile sucrose solution and gamma irradiated (15 kGy) bee bread *ad libitum* for 5 days. After this period, the bees were immobilized by CO_2 exposure. The guts of 50 workers per cohort were then removed aseptically, their ileums were dissected and retained as a pool in 95% ethanol at -80°C. DNA was extracted from each pool using methods detailed in (8).

Tn-seq: Illumina library preparation for transposon sequencing

The extracted DNA from transconjuant triplicate pools isolated from either HIA + 5% SB or dissected ileums of 50 bees, five days post-inoculation was used to construct Tn-seq sequencing libraries with the TdT tailing method as in (9), using modified primers for pBT20 (Table S3). The six libraries were sent for sequencing on a full 2 × 150 Illumina MiSeq lane at the Genomic Sequencing and Analysis Facility at the University of Texas at Austin. Reads are deposited in the NCBI Sequence Read Archive (SRA) under Bioproject PRJNA319191. Sequencing results are summarized in Table S1.

Tn-seq: Efficiency and reproducibility of transposon library

To evaluate the efficiency of our library, we examined total number of reads and total number of reads containing the transposon sequence (Table S1). In all cases, reads containing the transposon were > 91% of the total reads. The technical reproducibility was greater than 90% for all experiments (Fig. S7).

Tn-seq: Bioinformatics workflow summary

Reads were used to perform essential gene analysis based largely on the methods from Turner *et al*. (9). Briefly, sequenced reads containing the expected transposon sequences were trimmed of adapter sequences and mapped to the *S. alvi* wkB2 genome. In order to minimize the effect of non-disrupting insertions at the extreme 5' or 3' end of genes, we trimmed 5% of each end of the annotated gene (i.e., for a gene 100 nucleotides long, we only consider insertions in the interior 90 nucleotides). For the essential genome analysis, reads from the three replicate plate growth libraries were normalized for total number of reads. These normalized sites were then binned by gene and used to build a model of the expected number of insertions across the genome in the absence of fitness effects. Using DESeq2 (10), we compared the observed number of insertions with the predicted number.

All computational scripts and bioinformatics analyses used in this research are available at https://github.com/spleonard1/Tn-seq (commit # 1a531637dde967eb692ad2ecec2d3356f3cf0732 on May 2, 2016) and closely mirror scripts initially produced by Turner *et al*. (9). Individual steps in the bioinformatics workflow are detailed below.

Tn-seq bioinformatics workflow step 1: Identifying and mapping transposon reads

Reads were processed through Tn-seq2.sh script with the following command line:

```
>sh Tn-seq2.sh -p CACCCAGCTTTCTTG -i TACACTAGAGACCGGGGACTTATCAGCCAACCTGTTA -a
WKB2 -m 3 CONT1
```

This script uses fqgrep (http://dx.doi.org/10.5281/zenodo.45105) to search for the expected transposon sequences in raw reads. Reads containing these transposon sequences in the correct orientation were then trimmed of the transposon sequence and mapped with Bowtie 2 (11) to the genome of *S. alvi* wkB2. The location of this mapping was used to determine the insertion location, and the total sum of reads that map to each location was tallied. These files

containing transposon location and read abundance per location per sample were then further processed to identify either (A) the essential genome of *S. alvi* or (B) fitness determinants for colonization in the bee gut.

Tn-seq bioinformatics workflow step 2: Genome data preparation for sequencing analyses

All high throughput analyses were conducted with the *S. alvi* wkB2 genome (accession number NZ_CP007446.1) obtained from the NCBI Genome portal (http://www.ncbi.nlm.nih.gov/nuccore/NZ_CP007446.1) and used the GFF file NCBI_Assembly:GCF_000600005.1.

To prepare for Tn-seq analysis, only gene records were extracted from the GFF file, and 10% off the length of each gene was trimmed using GFFtrim.pl. To assign Kyoto Encyclopedia of Genes and Genomes (KEGG) numbers and KEGG pathways to each gene, we used the PullKegg.pl perl script:

```
cat wkB2.genes.gff | perl PullKegg.pl > wkB2.genes.kegg.gff
```

The "trimmed" .gff file was used for all Tn-seq analysis, and the untrimmed .gff file used for RNA-seq analysis.

Tn-seq bioinformatics workflow step 3: Determining the essential genome of *S. alvi* wkB2

Example command line:

```
>Tn-seqEssential.sh -i 50 -e 2000 -a wkB2 -o essential -c control -x 3 CONT1
CONT2 CONT3
```

An explanation of this script is detailed in Turner *et al*. (9). Mapped transposon reads from samples from plate growth (three replicates) were tallied by gene, and the 50 most abundant sites were removed to correct for any potential PCR amplification bias. These insertion locations across the entire genome were locally smoothed using weighted LOESS smoothing, and this model was then used to generate many random ("expected") data sets under a null hypothesis of no gene-specific fitness effects. We compare the observed data to these randomized data sets, searching for genes that show many fewer or no insertions compared to the generated data. Each gene was then assigned to either a "Reduced" (essential) or "Unchanged" (not-essential) category using mclust (12). In our final estimation of essential genes, we counted only those that were assigned to the "Reduced" category by mclust with an adjusted p-value < 0.05 and uncertainty < 0.1. During our analysis, we updated the Tn-seq scripts to use DESeq2 (10), but the rest of the analysis is identical to that previously described (9).

To confirm that insertions did not have polar effects on downstream genes, we verified the presence of multiple viable insertions immediately upstream of genes identified as essential.

We performed a separate analysis of essentiality in which we modeled expected insertion numbers per gene based on number of TA dinucleotides rather than on gene length (since insertion occurs at TA sites and genes can differ in their content of permissive insertion sites independently of length). This analysis resulted in reclassification of only 13 genes, 11 from "reduced" to "unchanged" and 2 from "unchanged" to "reduced". Most of these were

uncharacterized proteins. Because the length-based and TA-based analyses gave the same overall results, we present only the length-based analysis.

Tn-seq bioinformatics workflow step 4: Differential fitness analysis

Example command line:

```
>Tn-seqAnalysis.sh -i 50 -a WKB2 -o output -c CONT -x3 -t EXP -y3
CONT1 CONT2 CONT3 EXP1 EXP2 EXP3
```

Mapped transposon read locations for six samples (three replicates of the initial library and three replicates of the library prepared after bee inoculation) were tallied by gene, and the 50 most abundant sites removed as in the essential gene analysis. Raw transposon reads per gene were then treated like an RNA-seq experiment to search for differential abundance. We modified scripts to use the DESeq2 package for statistical analysis of these genes counts, based on a negative-binomial model and adjusted for multiple testing using the Benjamini-Hochberg correction. We considered genes that showed a $\log_2$-fold change ≤ -1 and an adjusted p-value ≤ 0.05 to have a biologically relevant change in fitness in the bee gut.

Tn-seq bioinformatics workflow step 5: Enriched KEGG Pathway and KEGG Module analysis

Enrichment tests for KEGG Pathways were conducted using the clusterProfiler package (13) from Bioconductor (14). This package implements a hypergeometric model to perform an over-representation test and identify KEGG pathways enriched in the set of essential, differentially fit, or differentially expressed genes. All enrichment tests are corrected for multiple testing via the Benjamini-Hochberg method and only those pathways with an adjusted p-value < 0.05 were considered enriched. These results are presented in Fig. S1 and were used as a first step in identifying biological activities important in gut colonization. We note that gene sets encompassed by groups of KEGG Pathways and Modules overlap extensively; for example, many of the same genes contribute to biosynthesis of amino acids, biosynthesis of antibiotics, and metabolic pathways. Thus, KEGG categorization was used mainly as a cursory analysis prior to more specific pathway-based analysis.

Gene categorization and placement in pathways

S. alvi genes identified as having significant fitness benefits in the gut were individually inspected and classified by BLASTP against the non-redundant protein Genbank database and the annotated *E. coli* K-12 MG1655 genome on EcoCyc (15). Pathways were constructed based on manual gene assignment to known genetic networks (16) and the *S. alvi* genome (17). Results are presented in SI Dataset 1, SI Dataset 2, and used to generate Fig. 3, Fig. S2, and Fig. S3.

RNA-seq of *S. alvi* from ileums of *A. mellifera* workers

To investigate bacterial transcripts involved in *S. alvi* persistence in the gut, microbiota-free 1-day old workers were inoculated with strain wkB2 by mixing scraped bacteria from a

culture plate with the pollen food source in the bee cup cages. Bees were kept at 34°C in the dark and given pollen and 1:1 w/v sucrose water to feed on *ad libitum*. At 5 days post emergence, bees were dissected and RNA from the ileums of the specimens was extracted with TRIzol (Ambion, Austin, TX, USA) using manufacturer recommendations. RNA was screened for quality via Bioanalyzer (Agilent, Santa Clara, CA, USA). cDNA was generated with iScript (Bio-Rad, Hercules, CA, USA) and screened with primers for *A. mellifera* Elongation Factor 1-alpha, universal 16S rRNA, and a Betaproteobacteria-specific gene to detect *S. alvi* (Table S3). High quality samples with positive screens for the three PCR assays were then used for pooling. RNA from three biological replicates of pooled ileums from 3 workers each were prepared, as well as triplicate samples from *S. alvi* wkB2 grown on a HIA + 5% SB plate for 24 hours. RNA pools from bee ileums were sent to the Yale Center for Genomic Analysis for mRNA enrichment, cDNA synthesis, and RNA-seq. Enrichment of microbial mRNA transcripts was performed with poly(A) depletion of eukaryotic mRNAs, followed by removal of eukaryotic rRNA and bacteria rRNA using Ribo-Zero rRNA Removal Kits for (Human/Mouse/Rat) and Bacteria, respectively (Epicentre, Madison, WI, USA). After cDNA synthesis, samples were run on an Illumina HiSeq for single-end, 76 bp read length sequencing. mRNA from the culture plate samples were enriched using the Ribo-Zero rRNA Removal Kit for Bacteria (Epicentre, Madison, WI, USA) and sequenced on an Illumina HiSeq for single-end 100 bp reads at the Genomic Sequencing and Analysis Facility at the University of Texas at Austin.

RNA-seq: Differential expression analysis

We used flexbar (18) to trim raw reads of adapter sequences and low quality reads. These trimmed reads were mapped to the wkB2 reference genome using Bowtie 2.1.0 (with default mapping parameters) and then tallied by gene with HTSeq (19). Differential expression was then analyzed with DESeq2 using the standard workflow from the package vignette: DESeq() function with default parameters. We considered genes with a $\log_2$-fold change ≥ 1 or ≤ -1 and an adjusted p-value ≤ 0.05 to be biologically relevant. As for the Tn-seq results, the RNA-seq data were analyzed to determine enriched KEGG pathways and modules (Fig. S1).

***In vitro* and *in vivo* experiments with insertion mutants**

Mutants with transposon insertions in two of the genes underlying the type IV pilus (T4P) apparatus as well as in the xanthine permease gene, SALWKB2_RS04310 (*pbuX*), were isolated from the insertion library. Individual clones were arrayed by transferring single colonies of the mutant library to 200 µl Insectagro + 12.5 µg/ml gentamicin in each well of a 96-well culture plate (Corning, Manassas, Virginia, USA) and growing for 3 days. Glycerol was then added to the cultures to a final concentration of 15% and cultures were frozen at -80°C. These frozen arrays were used to seed fresh 96-well plates with 200 µl Insectagro + 12.5 µg/ml gentamicin with a plate replicator (Bel-Art, Wayne, NJ, USA) and these plates were allowed to grow for 3 days. Visual inspection of growth characteristics led to the observation that some cultures did not form a complete surface biofilm but rather formed aggregates in the bottom of the well. Semi-random PCR of cultures was used to characterize the site of insertion, and those with insertions in T4P-related genes were noted for future experiments.

Pilus mutants *pilF*$^-$ and *pilG*$^-$, as well as *pbuX*,$^-$ were indicated as having a fitness cost associated with them during colonization in the Tn-seq experiment and so were used, along with mutant c10, for comparison to the wild type strain wkB2.

For *in vitro* assays, surface biofilm formation between the wild type wkB2, c10 and T4P mutants *pilF*$^-$ and *pilG*$^-$ was evaluated by starting 8 × 200 µl Insectagro cultures in a 96-well culture plate (Corning, Manassas, VA, USA). Cultures were started from 1 µl of OD_{600} ~1 glycerol stocks and grown for 3 days. Biofilms were stained, qualitatively observed and then solubilized and recorded via previously described methods (20). Mean absorbance at OD_{600} was calculated and visualized with boxplots using R (21). The statistical differences were assessed with Kruskal-Wallis in R.

Surface biofilms were observed via scanning electron microscopy by culturing 5 ml of either the wild type wkB2, c10, *pilF*$^-$ or *pilG*$^-$ strains in 50 ml Falcon conical flasks (Corning, Manassas, VA, USA) for 5 days at 35°C and 5% CO_2. Sections of the tube (~2 cm^2) were removed at the air/media interface from the flasks and observed at between 1.2–1.8K× with a Hitachi TM3030 Desktop scanning electron microscope (Hitachi High Technologies, USA).

For *in vivo* competition assays, frozen cultures were struck to HIA + 5% SB and grown for 3 days. Bacteria were transferred to 1ml 1× PBS to a final OD_{600} of 2 and then diluted 100× in 1× PBS (approximately 3.2×10^4 CFU, based on previously established OD_{600} enumeration curves). The dilutions of each mutant, as well as c10, were mixed in a 1:1 ratio with wild type wkB2. For each dual inoculum cohort (wt/c10, wt/*pilF*$^-$, wt/*pilG*$^-$ in T4P tests, and wt/c10, wt/*pbuX*$^-$ in an additional screen), 5 µl was fed to each individual in a group of 10 microbiota-free 1NEWs. These cohorts were maintained in cup cages in a growth chamber at 32°C and 80% relative humidity for 4 days. After this period, ileums were dissected out and homogenized with a sterile pestle in 100 µl Dulbecco's PBS. Ten microliters of a series of 10-fold dilutions was struck via the drop method (22) in triplicate on HIA blood plates with tetracycline (7.5 µg/ml) as well as on HIA blood plates with tetracycline (7.5 µg/ml) and gentamicin (12.5 µg/ml). Plates were grown for 3 days at 35°C and 5% CO_2, at which point colonies were counted and the proportion of the population containing the transposon insertion was calculated by subtracting the $\log_{10}$ of enumerated colonies growing on plates supplemented with tetracycline only from the $\log_{10}$ of total bacteria growing on plates with tetracycline and gentamicin. This experiment was repeated two times with 3–4 individuals assayed in each round of T4P tests, and performed once with 4 individuals in the *pbuX*$^-$ experiment. Data was visualized in a boxplot using R and statistical significance was assessed using the Kruskal-Wallis test in R. Dilutions of resuspended colonies of strains c10 and wkB2 were also plated to both tetracycline (tet) and tetracycline + gentamicin (tet + gent) plates in order to determine if there was a reduction in viability from the dual antibiotics.

Overall recovery of the control mutant c10 was lower than expected (mean = 0.004% CFU tet + gent /CFU tet). This strain was shown to colonize in equal proportions to the wild type based on the earlier single isolate colonization qPCR data. To test for effects of selective antibiotics, equivalent suspensions of both c10 and wkB2 were made and 10-fold dilutions were plated in triplicate to HIA + 5% SB with no antibiotic, with tet 7.5 µg/ml or tet 7.5 µg/ml + gent 12.5 µg/ml. There was no statistical difference between plate counts from the plates with no antibiotic to those with tet. Wild-type wkB2 does not contain a gentamicin resistance cassette, so it did not grow on the dual antibiotic media whereas c10 grew but had a ~10-fold reduction in enumerated colonies on tet + gent plates. This explains some of the observed discrepancy in

recovery versus the wild type. Since all of the Tn mutants would have the same disadvantage under gent selection, the assay can be used for direct comparison of colonization ability of the neutral c10 mutant to that of the *pil* and *pbuX* mutants, but the recovered mutants from the competition assay signify a relative scale of viability and not an absolute one.

Comparison of *S. alvi* wkB2 gene sets to those of related bacteria with different hosts

We compared gene sets of *S. alvi* wkB2 to gene sets in *S. alvi* wkB12, and also to the total gene set and essential gene set in *Neisseria gonorrhoeae* MS11, as reported by Remmele *et al*. (23). For both comparisons, genes were considered orthologs if they showed a reciprocal best BLASTP hit with E-value < 10^{-6}.

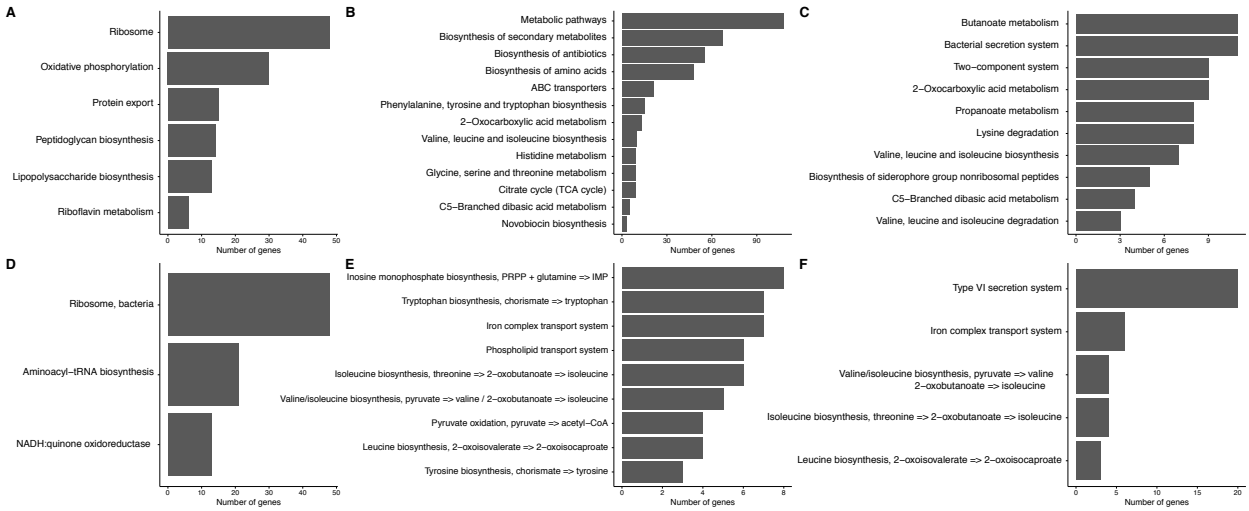

Fig. S1. Enriched KEGG pathway and KEGG module analysis of gene sets from Tn-seq and RNA-seq data. (A) Pathways significantly enriched ($p_{adj} < 0.05$) from essential gene set. (B) Pathways significantly enriched ($p_{adj} < 0.05$) from set of genes that reduce fitness when disrupted in the bee gut. (C) Enriched pathways of genes significantly upregulated (log_2-fold change > 1, $p_{adj} < 0.05$) during colonization compared to culture growth. (D) Modules significantly enriched ($p_{adj} < 0.05$) from essential gene set. (E) Modules significantly enriched ($p_{adj} < 0.05$) from set of genes that reduce fitness when disrupted in the bee gut. (F) Modules enriched in genes significantly upregulated (log_2-fold change > 1, $p_{adj} < 0.05$) during colonization compared to culture growth.

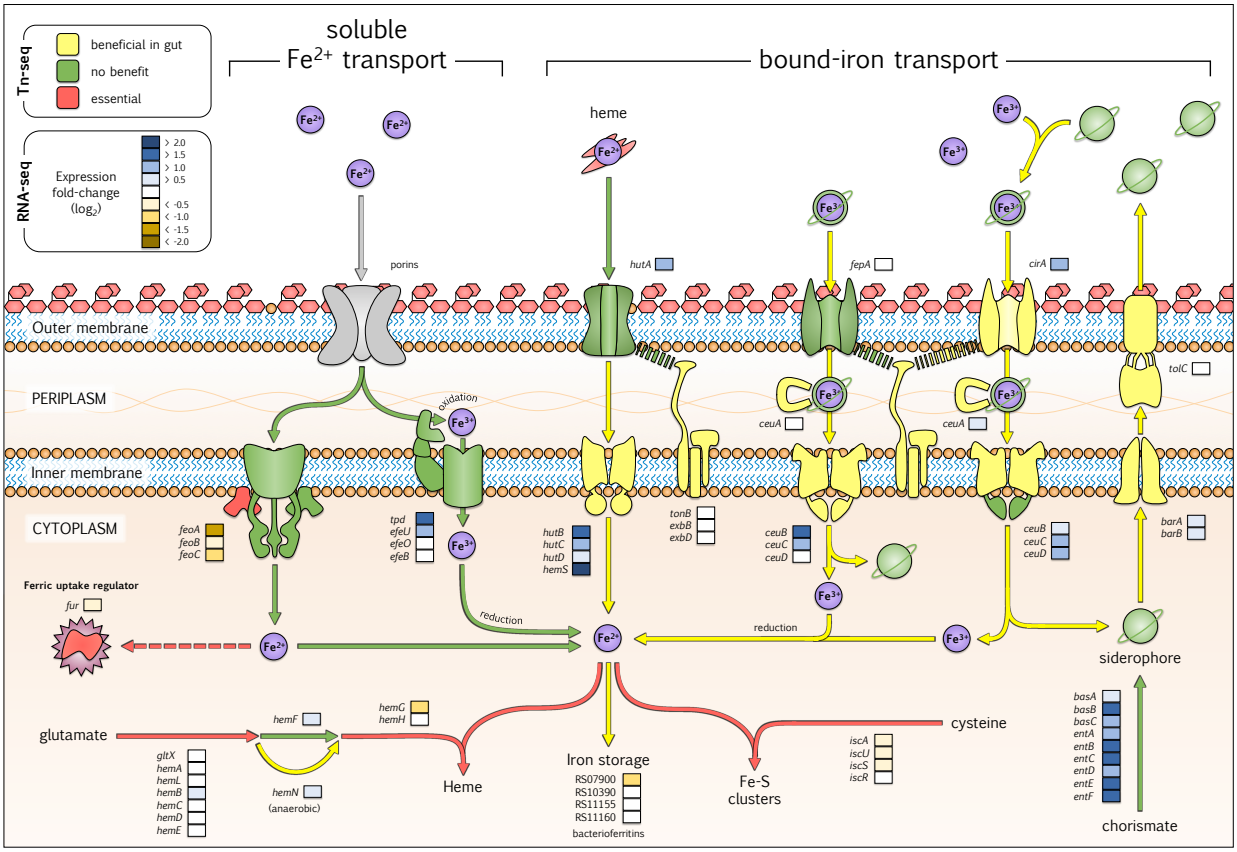

Fig. S2. Iron metabolism in *S. alvi*. Genes involved in many pathways, particularly for bound-iron acquisition, are required for *in vivo* gut colonization and are upregulated compared to *in vitro* laboratory culture.

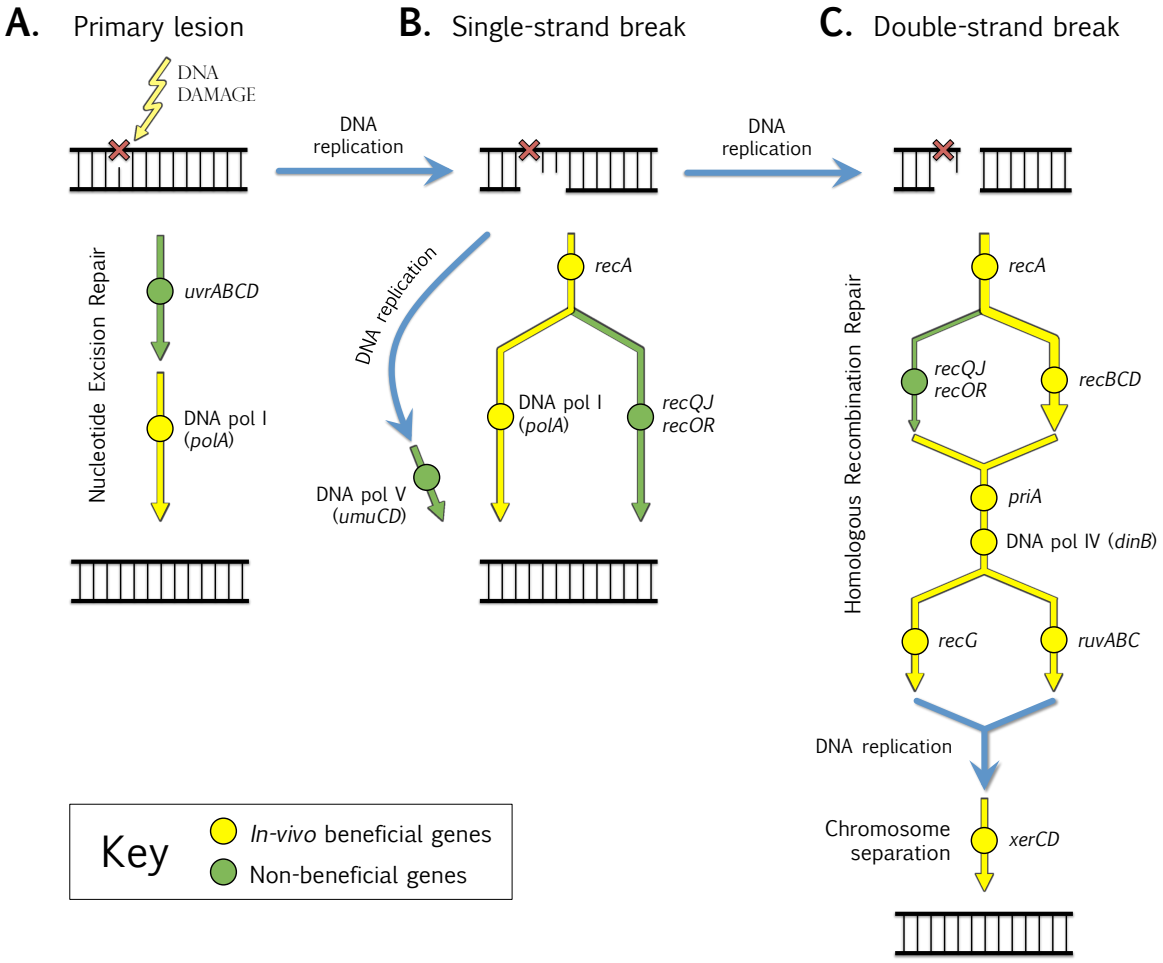

Fig. S3. *In vivo* DNA repair in *S. alvi* is geared towards double-strand break (DSB) prevention. None of the depicted repair genes are required in *in vitro* laboratory culture; however, loss of DNA pol I, which leads to an accumulation of DSBs, or loss of any genes in the major DSB repair pathway are strongly detrimental in the bee host. Failure to repair simpler DNA damage (A and B), can lead to DSBs (C) after rounds of replication.

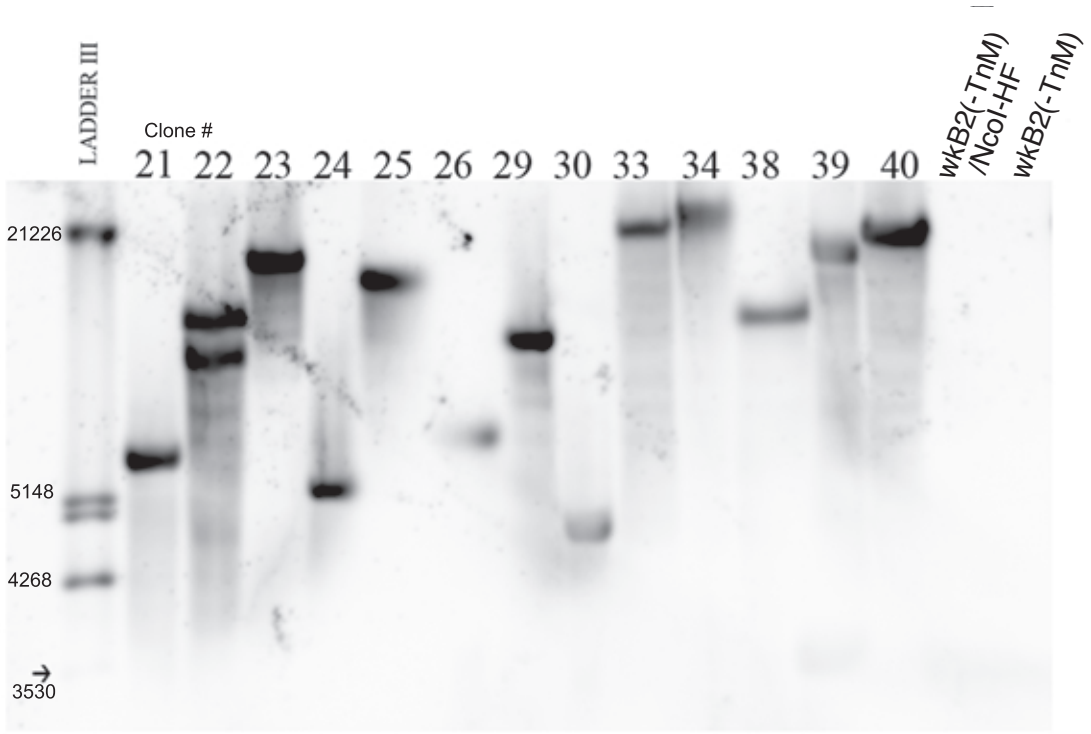

Fig. S4. Southern blot analysis of randomly selected clones bearing the gentamicin-encoding transposon from pBT20. The double band in clone 22 is believed to be an incomplete digest as the upper band is the expected size for the anticipated lower fragment plus an adjacent one. Amplicon from the srPCR reaction of clone 22 was subcloned to pGEMT and random clones were sequenced. Five out of five subclones matched the expected fragment. Clone 39 did not provide useable sequence data via the mapping srPCR procedure and was subcloned via the method above. Sequences from two fragments were detected, demonstrating that this clone may have experienced a double insertion event. Single random mutations were present in 92% of the clones queried.

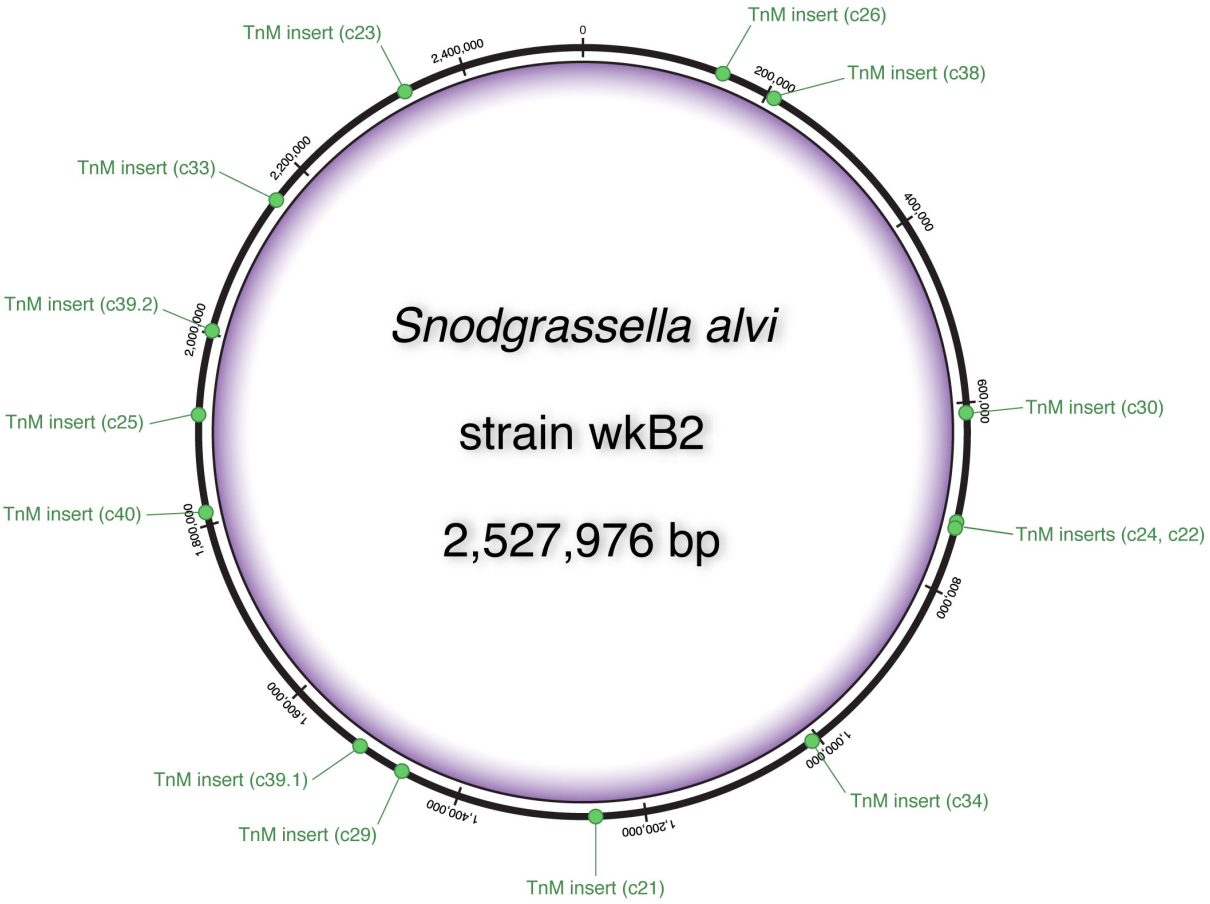

Fig. S5. Insertions of Himar1-based transposon into transconjugant clone chromosomes. Sites of insertions were mapped by sequencing amplicon from semirandom PCR. Insertion sites for clones c22 and c24 are 7,131 bp apart.

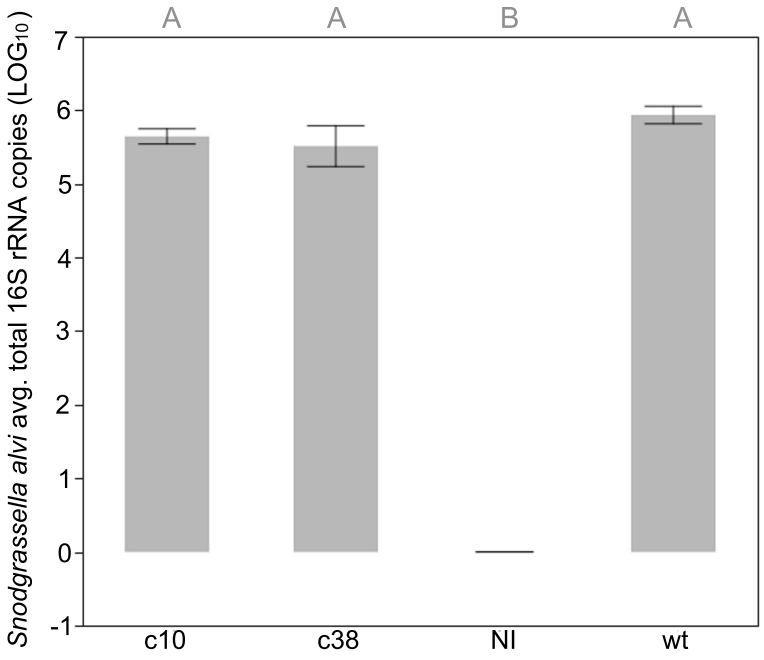

Fig. S6. Total average *S. alvi* 16S rRNA gene copies (log_{10}) of ileum DNA extracts per inoculation condition after 7 days. Tukey's HSD significance groups are noted above the chart and error bars demonstrate one standard error of the mean. Significant differences exist between the treatment groups (one-way ANOVA, $p < 0.0001$); non-inoculated bees (NI) had significantly lower bacterial load than all other groups (Tukey's HSD, $p < 0.0001$), while wild type *S. alvi* wkB2 (wt) and transconjugant clones c10 and c38 were all statistically similar (Tukey's HSD, $p > 0.1$).

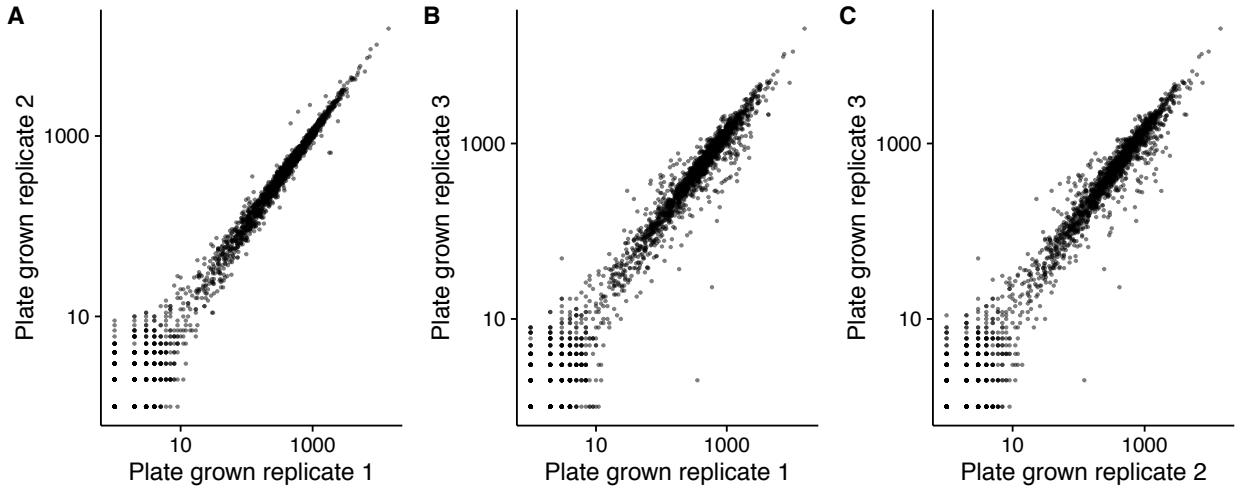

Fig. S7. Reproducibility of library preparation and sequencing, showing technical replicates of the library used for the determination of essential genes. Each point represents the number of reads mapped to a single gene. R-squared coefficients are 0.983 for (A), 0.915 for (B), and 0.915 for (C).

Table S1. Summary of read data from sequencing runs.

Library	MiSeq raw reads	Reads containing transposon sequence (%)	Alignment rate to wkB2 genome	High MAPQ mapped reads	Identified insertion sites
Plated control 1	2425350	2202448 (90%)	78.19%	1786283	74502
Plated control 2	2537725	2326996 (92%)	77.77%	1878810	75095
Plated control 3	3231396	2941256 (91%)	73.16%	2237015	78163
Bee gut experimental 1	2856080	2680334 (94%)	82.44%	2298408	29736
Bee gut experimental 2	2453927	2295149 (94%)	83.57%	1917948	23070
Bee gut experimental 3	3461575	3236851 (94%)	80.42%	2702636	32809

Table S2. Bacterial strains, clones and plasmids used in this study.

Strains, plasmids	Description	Reference
Strains		
β2163	*Escherichia coli*, DAP auxotroph, plasmid host	(1)
wkB2	*Snodgrassella alvi*, type strain	(3)
c10	wkB2 neutral transconjugant clone	This study
c38	wkB2 neutral transconjugant clone	This study
pilF-	wkB2 transconjugant clone containing attenuated *pilF* gene	This study
pilG-	wkB2 transconjugant clone containing attenuated *pilG* gene	This study
Plasmid		
pBT20	Plasmid containing Mariner C9 transposase and gentamicin resistant transposon	(2)

Table S3. Primers used in this study.

Assay name	Primer name	Sequence (5'-3')	Description	Reference
Elongation Factor 1-alpha	efa599	atctccggatggcacggYgacaa	DNA quality control assay	(24)
Elongation Factor 1-alpha	efa923	acgttcttcacgttgaaRccaa	DNA quality control assay	(24)
16S	27F-short	gagtttgatcctggctca	Universal bacterial primers	(24)
16S	1507R	taccttgttacgacttcaccccag	Universal bacterial primers	(24)
Beta	Beta-F (run w/1507R)	cttagagataggagagtg	*Snodgrassella*-specific assay	(24)
Beta-qPCR	Beta-1009-qtF	cttagagataggagagtg	*Snodgrassella*-specific quantitative assay	(7)
Beta-qPCR	Beta-1115-qtR	taatgatggcaactaatgacaa	*Snodgrassella*-specific quantitative assay	(7)
TnM round 1	ARB-1B	ggccagcgagctaacgagacnnnngat at	First round of semirandom nested PCR	(5)
TnM round 1	rnd1TnM20	tataatgtgtggaattgtgagcgg	First round of semirandom nested PCR	(5)
TnM round 2	inrARB	ggccagcgagctaacgagac	Second round of semirandom nested PCR	(5)
TnM round 2	rndRnd2TnM20	acaggaaacaggactctagagg	Second round of semirandom nested PCR	(5)
TnMseq	TnMseq	cacccagctttcttgtacac	Sequencing primer for transposon/chromosome junction	(2)
c10 screen	BM10 (run w/ TnMseq)	tgcagcgttgtaaacggata	Used to screen *Apis* workers for c10	This study
c38 screen	c38Fcand1(run w/ TnMseq)	tttgccaatattgctcatgg	Used to screen *Apis* workers for c38	This study
PCR1- Tn-seq lib	olj376	gtgactggagttcagacgtgtgctcttccg atctgggggggggggggggg	Used to amplify C tailed genomic DNA containing transposon insertion	(9)
	5prmBIO_rnd1_3	*tcgtataatgtgtggaattgtgagcgg	*=5' biotin modification	This study
PCR2- Tn-seq lib	Ill5nS_TnMseq	aatgatacggcgaccaccgagatctaca ctctttccctacacgacgctcttccgatctn nnnncacccagctttcttgtacac	Used to attach Illumina adapters and sample specific barcode	(9)
	BC-nn	caagcagaagacggcatacgagat*xxx xxx*gtgactggagttcagacgtgtg	*xxxxxx* = 6bp TruSeq barcode sequence	This study
TnM probe	TnM1-PRB	tagacgcatcccgttccatacaga	Primer for generating probe for Southern blot	This study
TnM probe	TnM2-PRB	gacctgcacagccataccacagct	Primer for generating probe for Southern blot	This study

Table S4. Summary of pooled PCR results from c10 inoculation experiments. NT, not tested.

Treatment	*n*	PCR based screens			
		EF1-α	16S	Beta (*Snodgrassella* specific assay)	BM10 (transposon specific assay)
Non-inoculated microbiota-free workers	17	17	7	4	0
Hive-reared workers	7	7	7	6	0
Snodgrassella (c10) only inoculated workers	15	15	13	12	13
Additional *Snodgrassella* (c10) only inoculated workers	36	36	NT	NT	25
Total *n* inoculated =	51			Total *n* colonized (c10) =	38
				Infection rate =	0.75

Table S5. Summary of selective plating and colony screens from *A. mellifera* ileums.

Treatment	Samples plated (*n*)	Number of samples with colonies observed on media		PCR based screen
		Gent -	Gent +	BM10 (transposon specific assay)
Non-inoculated microbiota free workers	10	9	6	0
Hive-reared workers	4	4	4	0
Snodgrassella (c10) only inoculated workers	10	10	10	7

SI Appendix References

1. Demarre G, et al. (2005) A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPα) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol* 156(2):245–255.
2. Kulasekara HD, et al. (2005) A novel two-component system controls the expression of *Pseudomonas aeruginosa* fimbrial cup genes. *Mol Microbiol* 55(2):368–380.
3. Kwong WK, Moran NA (2013) Cultivation and characterization of the gut symbionts of honey bees and bumble bees: description of *Snodgrassella alvi* gen. nov., sp. nov., a member of the family *Neisseriaceae* of the *Betaproteobacteria*, and *Gilliamella apicola* gen. nov., sp. nov., a member of *Orbaceae* fam. nov., *Orbales* ord. nov., a sister taxon to the order *'Enterobacteriales'* of the *Gammaproteobacteria*. *Int J Syst Evol Microbiol* 63(6):2008–2018.
4. Moran NA, Hansen AK, Powell JE, Sabree ZL (2012) Distinctive gut microbiota of honey bees assessed using deep sampling from individual worker bees. *PLoS One* 7(4):e36393.
5. Roden JA, Wells DH, Chomel BB, Kasten RW, Koehler JE (2012) Hemin binding protein C is found in outer membrane vesicles and protects *Bartonella henselae* against toxic concentrations of hemin. *Infect Immun* 80(3):929–942.
6. Lampe DJ, Akerley BJ, Rubin EJ, Mekalanos JJ, Robertson HM (1999) Hyperactive transposase mutants of the Himar1 mariner transposon. *Proc Natl Acad Sci U S A* 96(20):11428–33.
7. Martinson VG, Moy J, Moran NA (2012) Establishment of characteristic gut bacteria during development of the honeybee worker. *Appl Environ Microbiol* 78(8):2830–2840.
8. Powell JE, Martinson VG, Urban-Mead K, Moran NA (2014) Routes of acquisition of the gut microbiota of *Apis mellifera*. *Appl Environ Microbiol* 80(23):7378–7387.
9. Turner KH, Wessel AK, Palmer GC, Murray JL, Whiteley M (2015) Essential genome of *Pseudomonas aeruginosa* in cystic fibrosis sputum. *Proc Natl Acad Sci U S A* 112(13):4110–4115.
10. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15(12):1–34.
11. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.
12. Fraley C, Raftery AE, Murphy TB, Scrucca L (2012) mclust Version 4 for R: Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. *Tech Rep 597, Univ Washingt*:1–50.
13. Yu G, Wang L-G, Han Y, He Q-Y (2012) clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16(5):284–287.
14. Gentleman RC, et al. (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5(10):1–16.
15. Keseler IM, et al. (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res* 41:D605–D612.
16. Caspi R, et al. (2014) The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases *Nucleic Acids Res* 42:D459–D471.
17. Kwong WK, Engel P, Koch H, Moran NA (2014) Genomics and host specialization of honey bee and bumble bee gut symbionts. *Proc Natl Acad Sci U S A* 111:11509–11514.
18. Dodt M, Roehr JT, Ahmed R, Dieterich C (2012) FLEXBAR-Flexible Barcode and

Adapter Processing for Next-Generation Sequencing Platforms. *Biology (Basel)* 1(3):895–905.
19. Anders S, Pyl PT, Huber W (2015) HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31(2):166–169.
20. O'Toole GA (2011) Microtiter dish biofilm formation assay. *J Vis Exp* (47):10–11.
21. R Development Core Team (2008) *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, Vienna).
22. Herigstad B, Hamilton M, Heersink J (2001) How to optimize the drop plate method for enumerating bacteria. *J Microbiol Methods* 44:121–129.
23. Remmele CW, et al. (2014) Transcriptional landscape and essential genes of *Neisseria gonorrhoeae*. *Nucleic Acids Res* 42(16):10579–10595.
24. Martinson VG, et al. (2011) A simple and distinctive microbiota associated with honey bees and bumble bees. *Mol Ecol* 20(3):619–628.