Supplementary Information
I. NCBI SRA Accessions

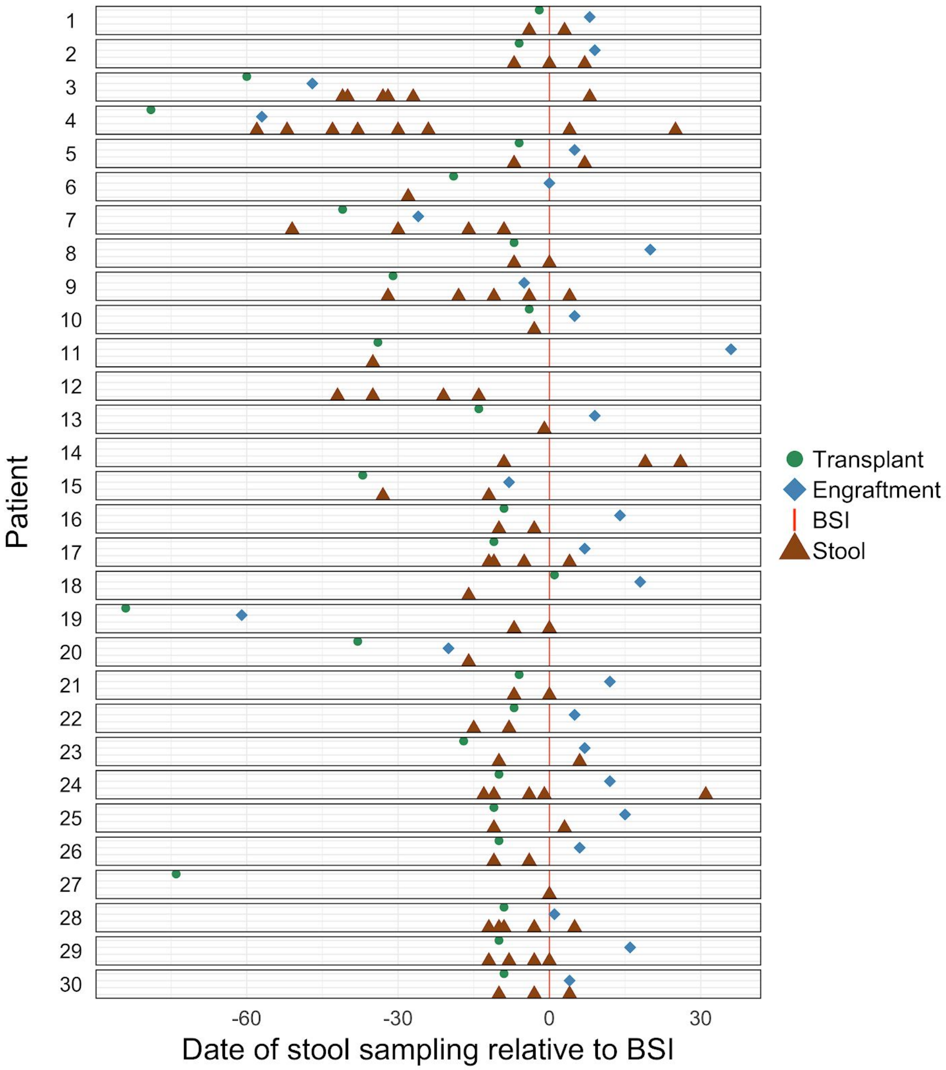
All WGS datasets are available via NCBI SRA under Bioproject PRJNA477326.

Sample	SRA BioSample Name	Category	SRA BioSample Accession
Patient 1 stool day -4	p1_meta_1	stool metagenome	SAMN09464888
Patient 1 stool day 3	p1_meta_2	stool metagenome	SAMN09464889
Patient 2 stool day -7	p2_meta_1	stool metagenome	SAMN09464890
Patient 2 stool day 0	p2_meta_2	stool metagenome	SAMN09464891
Patient 2 stool day 7	p2_meta_3	stool metagenome	SAMN09464892
Patient 3 stool day -41	p3_meta_1	stool metagenome	SAMN09464893
Patient 3 stool day -40	p3_meta_2	stool metagenome	SAMN09464894
Patient 3 stool day -33	p3_meta_3	stool metagenome	SAMN09464895
Patient 3 stool day -32	p3_meta_4	stool metagenome	SAMN09464896
Patient 3 stool day -27	p3_meta_5	stool metagenome	SAMN09464897
Patient 3 stool day 8	p3_meta_6	stool metagenome	SAMN09464898
Patient 4 stool day -58	p4_meta_1	stool metagenome	SAMN09464899
Patient 4 stool day -52	p4_meta_2	stool metagenome	SAMN09464900
Patient 4 stool day -43	p4_meta_3	stool metagenome	SAMN09464901
Patient 4 stool day -38	p4_meta_4	stool metagenome	SAMN09464902
Patient 4 stool day -30	p4_meta_5	stool metagenome	SAMN09464903
Patient 4 stool day -24	p4_meta_6	stool metagenome	SAMN09464904
Patient 4 stool day 4	p4_meta_7	stool metagenome	SAMN09464905
Patient 4 stool day 25	p4_meta_8	stool metagenome	SAMN09464906
Patient 5 stool day -7	p5_meta_1	stool metagenome	SAMN09464907
Patient 5 stool day 7	p5_meta_2	stool metagenome	SAMN09464908
Patient 6 stool day -28	p6_meta_1	stool metagenome	SAMN09464909
Patient 7 stool day -51	p7_meta_1	stool metagenome	SAMN09464910
Patient 7 stool day -30	p7_meta_2	stool metagenome	SAMN09464911
Patient 7 stool day -16	p7_meta_3	stool metagenome	SAMN09464912
Patient 7 stool day -9	p7_meta_4	stool metagenome	SAMN09464913
Patient 8 stool day -7	p8_meta_1	stool metagenome	SAMN09464914
Patient 8 stool day 0	p8_meta_2	stool metagenome	SAMN09464915
Patient 9 stool day -32	p9_meta_1	stool metagenome	SAMN09464916
Patient 9 stool day -18	p9_meta_2	stool metagenome	SAMN09464917
Patient 9 stool day -11	p9_meta_3	stool metagenome	SAMN09464918
Patient 9 stool day -4	p9_meta_4	stool metagenome	SAMN09464919
Patient 9 stool day 4	p9_meta_5	stool metagenome	SAMN09464920
Patient 10 stool day -3	p10_meta_1	stool metagenome	SAMN09464921
Patient 11 stool day -35	p11_meta_1	stool metagenome	SAMN09464922
Patient 12 stool day -42	p12_meta_1	stool metagenome	SAMN09464923
Patient 12 stool day -35	p12_meta_2	stool metagenome	SAMN09464924
Patient 12 stool day -21	p12_meta_3	stool metagenome	SAMN09464925
Patient 12 stool day -14	p12_meta_4	stool metagenome	SAMN09464926
Patient 13 stool day -1	p13_meta_1	stool metagenome	SAMN09464927
Patient 14 stool day -9	p14_meta_1	stool metagenome	SAMN09464928
Patient 14 stool day 19	p14_meta_2	stool metagenome	SAMN09464929
Patient 15 stool day -33	p15_meta_1	stool metagenome	SAMN09464930
Patient 15 stool day -12	p15_meta_2	stool metagenome	SAMN09464931
Patient 16 stool day -10	p16_meta_1	stool metagenome	SAMN09464932
Patient 16 stool day -3	p16_meta_2	stool metagenome	SAMN09464933
Patient 17 stool day -12	p17_meta_1	stool metagenome	SAMN09464934
Patient 17 stool day -11	p17_meta_2	stool metagenome	SAMN09464935
Patient 17 stool day -5	p17_meta_3	stool metagenome	SAMN09464936
Patient 17 stool day 4	p17_meta_4	stool metagenome	SAMN09464937
Patient 18 stool day -16	p18_meta_1	stool metagenome	SAMN09464938
Patient 19 stool day -7	p19_meta_1	stool metagenome	SAMN09464939
Patient 19 stool day 0	p19_meta_2	stool metagenome	SAMN09464940

Patient 20 stool day -16	p20_meta_1	stool metagenome	SAMN09464941
Patient 21 stool day -7	p21_meta_1	stool metagenome	SAMN09464942
Patient 21 stool day 0	p21_meta_2	stool metagenome	SAMN09464943
Patient 22 stool day -15	p22_meta_1	stool metagenome	SAMN09464944
Patient 22 stool day -8	p22_meta_2	stool metagenome	SAMN09464945
Patient 23 stool day -10	p23_meta_1	stool metagenome	SAMN09464946
Patient 23 stool day 6	p23_meta_2	stool metagenome	SAMN09464947
Patient 24 stool day -13	p24_meta_1	stool metagenome	SAMN09464948
Patient 24 stool day -11	p24_meta_2	stool metagenome	SAMN09464949
Patient 24 stool day -4	p24_meta_3	stool metagenome	SAMN09464950
Patient 24 stool day -1	p24_meta_4	stool metagenome	SAMN09464951
Patient 24 stool day 31	p24_meta_5	stool metagenome	SAMN09464952
Patient 25 stool day -11	p25_meta_1	stool metagenome	SAMN09464953
Patient 25 stool day 3	p25_meta_2	stool metagenome	SAMN09464954
Patient 26 stool day -11	p26_meta_1	stool metagenome	SAMN09464955
Patient 26 stool day -4	p26_meta_2	stool metagenome	SAMN09464956
Patient 27 stool day 0	p27_meta_1	stool metagenome	SAMN09464957
Patient 28 stool day -12	p28_meta_1	stool metagenome	SAMN09464958
Patient 28 stool day -10	p28_meta_2	stool metagenome	SAMN09464959
Patient 28 stool day -9	p28_meta_3	stool metagenome	SAMN09464960
Patient 28 stool day -3	p28_meta_4	stool metagenome	SAMN09464961
Patient 28 stool day 5	p28_meta_5	stool metagenome	SAMN09464962
Patient 29 stool day -12	p29_meta_1	stool metagenome	SAMN09464963
Patient 29 stool day -8	p29_meta_2	stool metagenome	SAMN09464964
Patient 29 stool day -3	p29_meta_3	stool metagenome	SAMN09464965
Patient 29 stool day 0	p29_meta_4	stool metagenome	SAMN09464966
Patient 30 stool day -10	p30_meta_1	stool metagenome	SAMN09464967
Patient 30 stool day -3	p30_meta_2	stool metagenome	SAMN09464968
Patient 30 stool day 4	p30_meta_3	stool metagenome	SAMN09464969
Patient 1 Staphylococcus aureus BSI	p1_bsi_1	bloodstream isolate	SAMN09466919
Patient 2 Enterobacter cloacae BSI	p2_bsi_1	bloodstream isolate	SAMN09466920
Patient 2 Klebsiella pneumoniae BSI	p2_bsi_2	bloodstream isolate	SAMN09466921
Patient 3 Escherichia coli BSI	p3_bsi_1	bloodstream isolate	SAMN09466922
Patient 3 Staphylococcus aureus BSI	p3_bsi_2	bloodstream isolate	SAMN09466923
Patient 4 Staphylococcus aureus BSI	p4_bsi_1	bloodstream isolate	SAMN09466924
Patient 5 Staphylococcus aureus BSI	p5_bsi_1	bloodstream isolate	SAMN09466925
Patient 6 Klebsiella oxytoca BSI	p6_bsi_1	bloodstream isolate	SAMN09466926
Patient 7 Escherichia coli BSI	p7_bsi_1	bloodstream isolate	SAMN09466927
Patient 8 Rothia mucilaginosa BSI	p8_bsi_1	bloodstream isolate	SAMN09466928
Patient 9 Escherichia coli BSI	p9_bsi_1	bloodstream isolate	SAMN09466929
Patient 10 Staphylococcus aureus BSI	p10_bsi_1	bloodstream isolate	SAMN09466930
Patient 11 Escherichia coli BSI	p11_bsi_1	bloodstream isolate	SAMN09466931
Patient 12 Staphylococcus aureus BSI	p12_bsi_1	bloodstream isolate	SAMN09466932
Patient 13 Staphylococcus epidermidis BSI	p13_bsi_1	bloodstream isolate	SAMN09466933
Patient 14 Klebsiella pneumoniae BSI	p14_bsi_1	bloodstream isolate	SAMN09466934
Patient 15 Staphylococcus epidermidis BSI	p15_bsi_1	bloodstream isolate	SAMN09466935
Patient 16 Streptococcus mitis BSI	p16_bsi_1	bloodstream isolate	SAMN09466936
Patient 17 Streptococcus mitis BSI	p17_bsi_1	bloodstream isolate	SAMN09466937
Patient 18 Staphylococcus epidermidis BSI	p18_bsi_1	bloodstream isolate	SAMN09466938
Patient 19 Pseudomonas aeruginosa BSI	p19_bsi_1	bloodstream isolate	SAMN09466939
Patient 20 Staphylococcus epidermidis BSI	p20_bsi_1	bloodstream isolate	SAMN09466940
Patient 21 Staphylococcus aureus BSI	p21_bsi_1	bloodstream isolate	SAMN09466941
Patient 22 Streptococcus mitis BSI	p22_bsi_1	bloodstream isolate	SAMN09466942
Patient 23 Enterococcus faecium BSI	p23_bsi_1	bloodstream isolate	SAMN09466943
Patient 24 Rothia mucilaginosa BSI	p24_bsi_1	bloodstream isolate	SAMN09466944
Patient 25 Enterococcus faecium BSI	p25_bsi_1	bloodstream isolate	SAMN09466945
Patient 26 Streptococcus mitis BSI	p26_bsi_1	bloodstream isolate	SAMN09466946
Patient 27 Staphylococcus aureus BSI	p27_bsi_1	bloodstream isolate	SAMN09466947
Patient 28 Streptococcus mitis BSI	p28_bsi_1	bloodstream isolate	SAMN09466948
Patient 29 Streptococcus mitis BSI	p29_bsi_1	bloodstream isolate	SAMN09466949
Patient 30 Streptococcus mitis BSI	p30_bsi_1	bloodstream isolate	SAMN09466950

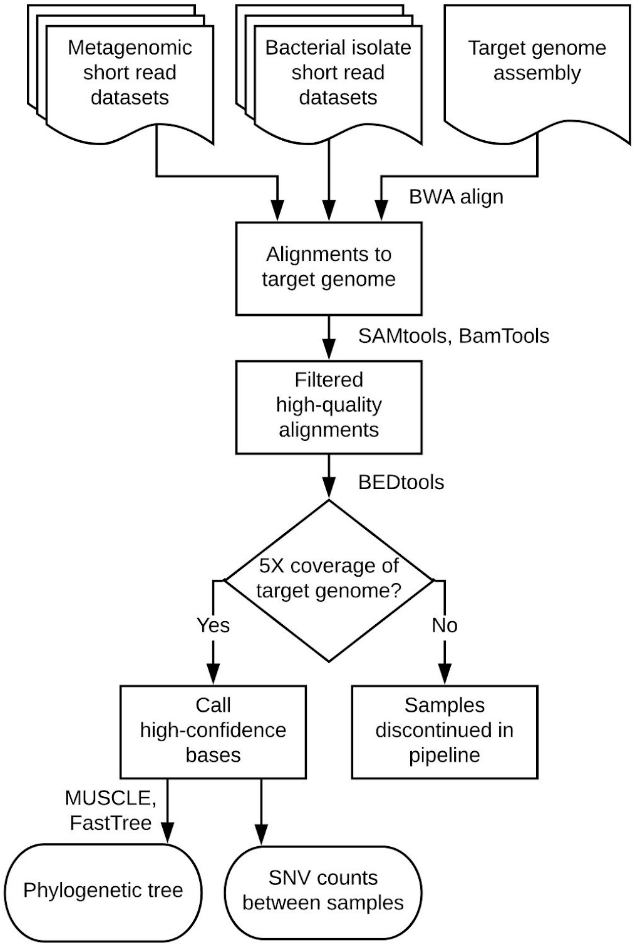
II. Supplementary Figures

Figure S1. Stool sampling relative to hematopoietic cell transplantation (HCT) and engraftment



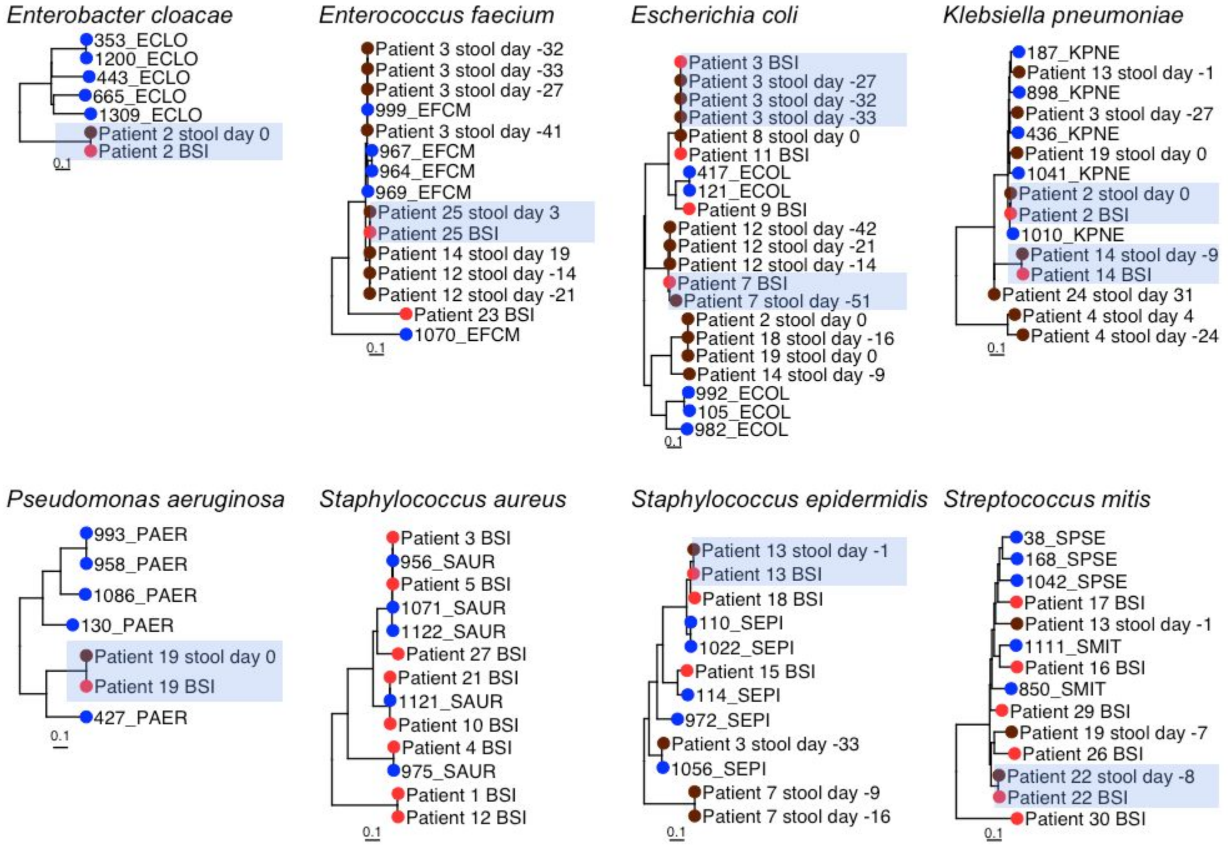
Timeline of stool sampling, transplantation, and engraftment relative to bloodstream infection (BSI) for patients in the study cohort. Engraftment and HCT occurred more than 100 days prior to BSI for patients 12 and 14 and are therefore not displayed. Engraftment could not be assessed for patient 27 as the patient did not become neutropenic.

Figure S2. StrainSifter workflow



Flowchart showing input, analysis steps, and output of the StrainSifter pipeline.

Figure S3. Phylogenetic trees of gut and bloodstream infection (BSI) strains with publicly available references



Phylogenetic relatedness between bacterial strains in our sample collection and publicly-available reference strains isolated from intensive care unit patients at the University of Washington Medical Center (NCBI BioProject PRJNA267549) as assessed by StrainSifter. Branch tip colors indicate stool metagenomes (brown), BSIs (red), and reference strains (blue). Samples from the same patient are more closely phylogenetically related to each other (blue highlight) than to samples from other patients. Days given are relative to BSI. Though patient 20's BSI is classified as *S. epidermidis*, this strain does not meet the coverage requirements for inclusion in the *S. epidermidis* phylogenetic tree.

III. Computational methods supplement

StrainSifter is implemented as a Snakemake workflow packaged with conda and is available for download at https://github.com/bhattlab/StrainSifter along with all source code. The following is an overview of the component steps in the workflow with example commands.

Computational methods

1. Align short reads to reference genome using Burrows-Wheeler aligner (BWA); filter alignments on mapping quality (mapq) >= 60; filter reads with 5 or fewer mismatches:

   ```
   bwa mem -t 8 {reference}.fna {sample}.fastq | samtools view -b
   -q 60 | bamtools filter -tag 'NM:<=5' | samtools sort --threads
   8 -o {sample}.filtered.bam
   ```

   Input:
   - Reference genome `{reference}.fna`
   - WGS fastq: `{sample}.fastq`

   Output:
   - `{sample}.filtered.bam`

2. Calculate average coverage per alignment with bedtools:

   ```
   bedtools genomecov -ibam {sample}.bam > {sample}.tsv
   ```

   Process with python to determine samples that meet the coverage requirements.

   Script:
   `getCoverage.py`

   Input: `{sample}.tsv`
   Output: `{sample}.cvg`

3. Filter samples meeting coverage requirements using bash/Python in Snakemake (default: 5X coverage over 50% of the genome)

4. Create pileup file for each sample meeting coverage requirements using samtools mpileup:

   ```
   samtools mpileup -f reference.fna -B -aa -o {sample}.pileup
   {sample}.bam
   ```

   Parameters:
   - -B -- report Phred scores instead of BAQ

-aa -- report every position in ref

Input: `{sample}.bam`
Output: `{sample}.pileup`

5. Call SNVs -- only consider positions covered by at least 5 reads; consider only bases with a minimum phred score of 20; call a base only if the frequency exceeds 0.8:

   Script:
   `callSNPs.py`

   Parameters (default):
      Min coverage (5)
      Min base frequency (0.8)
      Min phred score (20)

   Input: `{sample}.pileup`
   Output: `{sample}.tsv`

6. Tree building
   a. Obtain consensus sequence from each pileup file in bash

      Input: `{sample}.tsv`
      Output: `{sample}.txt`

   b. Find core positions among samples from each species -- positions where a non-"N" base call is made in every sample:

      Script:
      `findCoreSNPs.py`

      Input: `{sample}.txt` consensus sequence files for each sample
      Output: `{reference}.tsv`

   c. Generate fasta of concatenated SNVs:

      Script:
      `coreSNPs2fasta.py`

      Input: `{reference}.tsv`
      Output: `{reference}.fasta`

   d. Multi-alignment of SNV fasta with MUSCLE:

```
muscle -in {reference}.fasta -out {reference}.afa
```

e. Compute approximate maximum-likelihood tree with FastTree:

```
Fasttree -nt {reference}.afa > {reference}.tree
```

7. Count SNVs between samples

   Script: `pairwiseDist.py`

   Input: `{sample}.txt` consensus sequence files for each sample
   Output: `{reference}.dist.tsv`

Rationale for variant-calling frequency cutoff

We performed SNV calling and counted SNVs between a pair of gut and BSI samples, incrementing the minimum proportion to call a base from 0 to 1. By 0.7 we approach an asymptote, suggesting an appropriate minimum frequency parameter is 0.7-0.8.