

ORIGINAL ARTICLE OPEN ACCESS

Diagnostic Accuracy of Shotgun Metagenomics for Bloodstream Infections Is Influenced by Bioinformatics Workflow Selection

Yajing Song¹ | Christian Kjellander^{2,3} | William Robinson¹ | Lars Öhrmalm⁴ | Christian Giske² | Peter Gyarmati¹

¹University of South Carolina School of Medicine Greenville, Greenville, South Carolina, USA | ²Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden | ³Department of Internal Medicine, Södersjukhuset, Stockholm, Sweden | ⁴PRIMA Barn- och Vuxenpsykiatri, Stockholm, Sweden

Correspondence: Peter Gyarmati (gyarmati@sc.edu)

Received: 16 June 2025 | **Revised:** 27 October 2025 | **Accepted:** 30 October 2025

Funding: The authors received no specific funding for this work.

ABSTRACT

Bloodstream infection (BSI) is a severe and often fatal condition, and a major cause of mortality in patients with hematological malignancies due to underlying conditions and anticancer therapy-induced immunodeficiency. Rapid identification of the causative pathogens is essential as BSI results in worsened prognosis, extended hospitalization, delays or dose reductions in therapy, and may progress to sepsis and septic shock if untreated. Shotgun metagenomics is a culture-independent technique capable of detecting a wide range of fungal, viral, and bacterial organisms along with their antimicrobial resistance genes. Several studies showed that shotgun metagenomics enables the diagnosis of BSI, specifically in cases where conventional methods/culture-dependent techniques fail to identify the causative pathogens. However, evaluation of the accuracy of the applied bioinformatics pipelines remains incomplete. This study aimed to compare and optimize four commonly used bioinformatics pipelines (BLAST, Kraken, Metaphlan, RTG Core) for shotgun metagenomics by assessing their accuracy in identifying pathogens in blood samples from patients with hematological malignancies and suspected BSI, with blood culture serving as the reference standard. Our work shows that the selection of bioinformatics pipelines for diagnosing BSI strongly affects the precision of the findings, and an optimized BLAST pipeline was superior to the alternatives, as it was the only method that accurately identified the causative pathogens.

1 | Introduction

Bloodstream infection (BSI) is a major cause of mortality in patients with hematological malignancies due to the underlying immunological defects and therapy-induced immunosuppression (de Naurois et al. 2010). BSI results in worsened prognosis, extended hospital stay and increased healthcare costs (Garcia-Vidal et al. 2018; Dinan et al. 2015). Rapid diagnosis of BSI is essential as patient prognosis improves with early detection (Kalich et al. 2015), while organ dysfunction and septic shock can develop if the infection is not treated in time (Laupland 2013). Blood culture is the reference method to diagnose BSI. However,

this assay has a turnaround time of 4–72 h and therefore it may not be suitable in assisting primary diagnosis and selection of antimicrobials in case of fastidious organisms. In addition, blood culture cannot detect viral, and most fungal pathogens, and microbial interference may hinder polymicrobial detections. These are likely some of the reasons why 70%–90% of blood cultures from febrile neutropenia and 50% of blood cultures from septic shock are negative and the causative agents(s) remain unknown (Hämäläinen et al. 2008; Liesenfeld et al. 2014).

Molecular methods such as PlexID, MagicPlex, SeptiFast, VYOO have been available to detect BSI, but they only detect a

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *MicrobiologyOpen* published by John Wiley & Sons Ltd.

subset of pathogens using pre-designed primers and probes. These amplification-based molecular techniques are prone to false positivity in BSI diagnosis due to a large amount of human DNA background in blood samples, and false negativity due to mutations or variants not covered by primers and probes (Sinha et al. 2018; Opota et al. 2015; Ecker et al. 2010).

Standard treatment of BSI involves treatment with empirical broad-spectrum antibiotics, which are inefficient against fungal or viral pathogens, or against antimicrobial-resistant (AMR) bacteria. As a result, inappropriate antimicrobial use can contribute to an increase of AMR strains. It has been shown that 46% of the empirical antibiotic treatments were inappropriate and may have contributed to a 35% increase in overall mortality when treating BSI (Ammerlaan et al. 2013; Paul et al. 2010). Therefore, accurate characterization of the causative agents could improve the specificity and efficacy of antimicrobial treatments. Using culture-independent, untargeted techniques such as shotgun metagenomics, it has been shown that BSI in hematological malignancies is rarely monomicrobial, and caused not only by bacteria, but also by fungal and viral pathogens (Gyarmati et al. 2015, 2016; Fida et al. 2021; Ohrmalm et al. 2012). A step towards clinical validation of the method includes an optimization of the bioinformatics pipeline. Metagenomics analysis of blood samples in BSI differs from other specimen types, as blood contains a small amount of microbial DNA, but a large amount of human DNA, which can lead to false-positive detections (Gyarmati et al. 2015, 2016; Fida et al. 2021). Therefore, in this study we compared the accuracy of 4 shotgun metagenomics software to the gold standard in their accuracy of identifying the causative pathogens in blood samples from patients with hematological malignancies and suspected BSI.

2 | Methods

2.1 | Sample Collection and Processing

Blood samples were taken from patients with hematological malignancies and suspected BSI and processed for blood culture as described before (Gyarmati et al. 2016). BSI was defined as an infection manifested by the presence of bacteria in at least one blood culture bottle, or at least two blood culture bottles with the same microorganism growing in the case of common skin contaminants. Commercial BacT-Alert 3D system with 2–2 aerobic and anaerobic bottles was used (bioMérieux, Marcy l'Etoile, France) for culture. Two-hundred microliters of the blood samples were processed for DNA extraction and shotgun metagenomics sequencing using the MolYsis Complete5 kit (Molzym Life Science, Bremen, Germany). Sequencing reads were processed for quality control as described before (Gyarmati et al. 2016; Song et al. 2022). Briefly, 1 µg of the eluted DNA was processed to NebNext microbiome enrichment (New England Biolabs, Ipswich, MA, USA) following the manufacturer's instructions. Ten nanograms of DNA was subjected to multiplex displacement amplification, using the GenomiPhi V2 DNA amplification kit (GE Healthcare, Little Chalfont, United Kingdom), with 90 min amplification resulting in 2–4 µg DNA. Two micrograms DNA was used for library preparation with the Nextera XT kit and libraries were

processed to a 2*100 base pair PE sequencing on a HiSeq. 2500 instrument. Reads shorter than 30 bp and with Phred quality scores below 30 were discarded using the Fastx toolkit. Paired end reads were merged using the Flash software, while unpaired reads were discarded, with an average read count of 33.5 M/sample (Gyarmati et al. 2016).

2.2 | Data Analysis

Only sequencing data from blood samples with positive blood cultures were used for the current analysis. Fastq files for BLAST (BLASTn, 18), Kraken (Wood and Salzberg 2014) and Metaphlan (Beghini et al. 2021) were used via the Galaxy platform (Abueg et al. 2024), and all software, including RTG Core v3.12 (Cleary et al. 2014) used default settings, with the exception that the required minimum sequence similarity was set to 90% for all software where it was not the default value. All software were used with their respective native database. BLAST was used with the nt nucleotide database. Microfiltered sterile human blood ("negative blood", from Sigma-Aldrich, St. Louis, MO, USA) and no template control (NTC, sterile water) were used as negative controls, and microbial detection was considered valid only if it was not detected in either negative blood or in NTC. The ZymoBIOMICS Microbial Community Standard II (Zymo Research, Irvine, CA, USA) was used as a standardized positive control to evaluate assay performance.

Eta-squared (η^2) was used as a statistical measure to quantify the effect size in ANOVA as proportion of total variance ($\eta^2 = \text{Sum of squares between groups} / \text{Total sum of squares}$), where 0 = none of the variance is explained by the group differences, 1 = all the variance is explained by the group differences (Cohen 1988).

3 | Results

First, we evaluated the accuracies of different software using a mock microbial community standard consisting of gram-positive and gram-negative organisms and yeasts, representing clinically relevant microbes with different cell wall compositions (Zymo Microbial Community Standard (Figure S1)). Only BLAST could identify pathogens to the lowest relative abundance (8.9E-07, *Staphylococcus aureus*), while RTG Core showed a strong correlation down to 8.9E-05 of relative abundance in comparison to the theoretical distribution (Table 1).

Next, we investigated the accuracy of detection by different software in blood samples. Five blood samples were culture positive as described previously (Gyarmati et al. 2016), and all positive blood cultures showed α -hemolytic Streptococci. As expected, relative abundances for positive detections were low compared to other sample types with a native microbiome (feces, oral swab, etc.), since human genomic background greatly outweighs microbial reads in blood samples in BSI (Gyarmati et al. 2016; Song et al. 2014). Only BLAST was able to accurately detect viridans group streptococci in all blood samples from shotgun metagenomics data that were positive for these organisms on blood culture (Table 2). Even though

BLAST was the only software that correctly identified all culture-positive samples, it also detected the negative blood and NTC samples as false positives.

3.1 | Reduction of False-Positive Signals

We then investigated two different approaches to reduce or eliminate false positive signals from BLAST results. First, we tested whether the removal of human sequences improves detection due to less background noise. Sequences were mapped to the human genome (build Hg_19) before mapping reads to the nt database with e-value set to 10E-6. There were no significant differences between the relative abundances of viridans group streptococci detections ($p = 0.15$) with or without the prior removal of human sequences.

Second, we tested whether lowering the e-value improves detection specificity. E-value (Expect value) is a parameter in BLAST describing how significant the match is, and lowering the e-value leads to more stringent search criteria. Accuracy is critical for pathogen detection in blood samples, as human reads can outnumber microbial reads to 1:1 M or higher (Song et al. 2014), which can lead to false-positive detections. Lowering the e-value helped filter out false positives from no template control without affecting the accuracy of the method

(Figure 1). At e-value level 10E-12, there was a single read assigned to NTC. All culture-positive samples were positive and had the highest abundance for viridans streptococci at e-value level 10E-13, while NTC and negative blood samples were negative with 0 assigned read. Other microbial detections were in lower abundances and were not clinically relevant (e.g., *Alicyclophilus denitrificans*, *Tetragenococcus halophilus*).

4 | Discussion

Blood culture is the current reference standard for the detection of bloodstream infections, and therefore is considered as the most reliable and accurate method overall. Reference standards are not without shortcomings, as blood culture takes longer time to perform than molecular methods, and turnaround time can be up to 3 days for slow-growing/fastidious microorganisms. Using blood culture also presumes that all causative pathogens can grow in the medium, while many pathogenic microorganisms have specialized growth requirements (Stewart 2012). In addition, blood culture can result in mono-microbial detection due to microbial interference (Lagier et al. 2015), despite many studies showing that BSI is often polymicrobial (Gyarmati et al. 2015, 2016; Fida et al. 2021; Ohrmalm et al. 2012). These drawbacks may contribute to the low positivity rate of blood cultures, as only 10-30% of blood

TABLE 1 | Relative abundances of the standard microbial control as analyzed by different software, in comparison with the theoretical distribution.

	Theoretical	Blast	Metaphlan	RTG	Kraken
<i>Listeria monocytogenes</i>	8.90E-1	8.97E-1	7.23E-1	8.64E-1	3.40E-1
<i>Pseudomonas aeruginosa</i>	8.90E-2	8.03E-2	3.00E-2	7.65E-2	3.30E-3
<i>Bacillus subtilis</i>	8.90E-3	8.82E-3	2.30E-3	8.50E-3	0
<i>Salmonella enterica</i>	8.90E-4	1.84E-4	0	8.65E-4	5.17E-10
<i>Escherichia coli</i>	8.90E-4	1.07E-3	0	6.54E-4	5.17E-10
<i>Lactobacillus fermentum</i>	8.90E-5	1.53E-5	0	9.36E-5	5.17E-11
<i>Enterococcus faecalis</i>	8.90E-6	6.07E-6	0	9.68E-5	5.17E-12
<i>Staphylococcus aureus</i>	8.90E-7	1.90E-6	0	0	0
η^2	*	4.25E-9	3.00E-3	7.09E-5	3.20E-2
p-value	*	> 0.999	0.840	0.975	0.504

*not applicable.
Note: η^2 and p-value of ANOVA are shown.

TABLE 2 | Relative abundances of viridans group Streptococci as detected by different software in blood samples.

	Blast	Kraken	RTG	Metaphlan
Sample 1	1.21E-1	1.74E-5	0	0
Sample 2	2.87E-2	0	0	0
Sample 3	6.49E-1	2.03E-2	2.74E-1	2.30E-2
Sample 4	4.39E-2	1.69E-5	0	0
Sample 5	2.39E-2	0	0	0
Negative blood	1.03E-1	0	0	0
No template control	2.59E-1	0	0	0

Note: Green indicates the highest, while red indicates the lowest abundances.

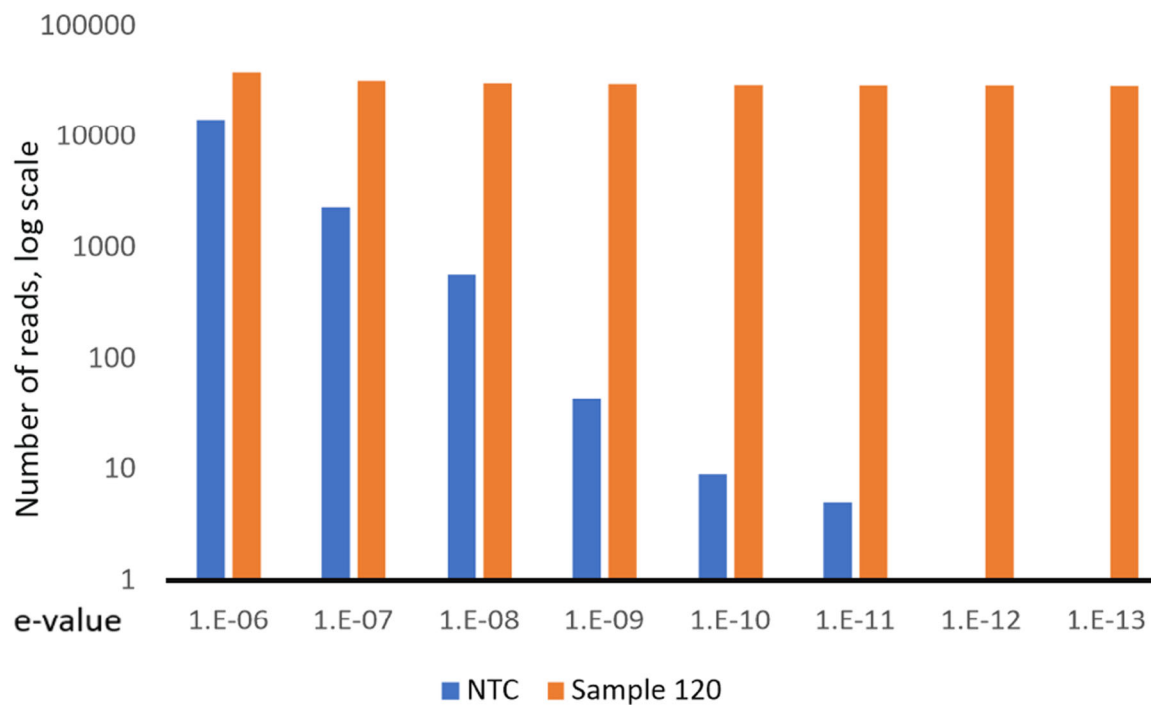


FIGURE 1 | Number of reads specific to viridans group streptococci by different e-values (from 10 to 6 to 10⁻¹³) have been tested using no template control (blue columns, NTC) and sample 3 (red columns).

cultures are positive in sepsis. Even in patients with septic shock, the causative agent remains unknown due to negative blood culture results in up to 50% of the cases (Li et al. 2021; Gupta et al. 2016).

Shotgun metagenomics emerged as a novel technology with the potential to address the challenges related to BSI detection. As a culture-independent method, it can detect a wide variety of pathogens, and their antimicrobial profiles with great precision, with a turnaround time of 4–40 h. However, blood as a sample poses special challenges to shotgun metagenomics, as it contains a large amount of host DNA and a small amount of microbial DNA, making this method less effective for BSI detection because most of the sequencing power is directed toward background signals (Gyarmati et al. 2016). Metagenomic analysis of blood samples also poses unique obstacles for bioinformatics methods, as they have been optimized for samples where microbes are present in significant proportion, such as fecal matter, bronchoalveolar fluid, and oral and skin swabs, among others. In an effort toward standardization and clinical validation of shotgun metagenomics, an optimized bioinformatics pipeline is required. We compared four bioinformatics pipelines commonly used for analyzing shotgun metagenomics data.

We only included culture-positive samples, as the aim of this study was to use culture data as benchmarks. Although all pipelines included are widely used in metagenomic analyses, they rely on distinct bioinformatic principles. Kraken classifies reads by mapping exact k-mers to a reference database and assigning taxonomy via the lowest common ancestor (LCA) algorithm; MetaPhlAn estimates community composition from clade-specific marker genes; BLAST uses a heuristic alignment approach; and RTG Core applies a dynamic-programming edit-distance model (Cock et al. 2015; Wood and Salzberg 2014; Beghini et al. 2021; Cleary et al. 2014). These

methodological differences may explain the variation in performance observed across control and test samples. BLAST requires the largest hardware and database (Salzberg and Wood 2021) of the four pipelines included in the study. It was the only one that reached 100% concordance with blood culture results, but required several optimization steps to avoid false-positive signals. These results will guide the integration of shotgun metagenomics into clinical practice, and will allow an improved detection of BSI for rapid antimicrobial treatment as part of precision medicine.

This study has several limitations. We used blood culture results as a reference since this method is the current gold standard, even though its low positivity rate and limited polymicrobial detection are well known. Twenty-two blood samples were subjected to shotgun metagenomics in the original study (Gyarmati et al. 2016), which contained 5 samples positive for blood culture. The low positivity rate of blood cultures highlights the importance of complementary diagnostic techniques in the diagnosis of BSI (Hämäläinen et al. 2008; Liesenfeld et al. 2014). The BLAST software resulted in a 95% confidence interval of 0.61–1 using the Wilson score interval (Agresti et al. 1998) by correctly identifying pathogens in blood culture positive samples, along with positive and negative controls. All blood cultures included in this study were positive for viridans group streptococci; however, streptococci are common causes of BSI. In addition, this study tested the software and its respective native database, but did not combine them due to computational or software limitations.

5 | Conclusion

The choice of metagenomics pipeline for analyzing shotgun metagenomics data for BSI samples strongly influences the result. In this study, an optimized BLAST pipeline with

stringent e-value was shown to be superior to the other tested options and provided complete concordance with the gold standard without false negative or false positive detection.

Author Contributions

Yajing Song: conceptualization (supporting), writing – review and editing (supporting), formal analysis (supporting). **Christian Kjellander:** data curation (lead), writing – review and editing (supporting), formal analysis (supporting). **William Robinson:** writing – review and editing (supporting), formal analysis (supporting). **Lars Öhrmalm:** data curation (supporting), writing – review and editing (supporting), formal analysis (supporting). **Christian Giske:** conceptualization (supporting), writing – review and editing (supporting), supervision (supporting). **Peter Gyarmati:** conceptualization (lead), writing – review and editing (lead), writing – original draft (lead), supervision (lead).

Ethics Statement

All patients provided written, informed consent. The study and all experimental protocols used in this study was approved by The Regional Ethical Review Board, Stockholm (2012/1929-31/1) and were carried out in accordance with the approved guidelines.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Sequencing data has been uploaded to NCBI Sequencing Reads Archives under accession number PRJNA1268383.

References

- Abueg, L., E. Afgan, O. Allart, et al. 2024. “The Galaxy Platform for Accessible, Reproducible, and Collaborative Data Analyses: 2024 Update.” *Nucleic Acids Research* 52, no. W1: W83–W94.
- Agresti, A., B. A. Coull, A. Agresti, and B. A. Coull. 1998. “Approximate Is Better Than ‘Exact’ for Interval Estimation of Binomial Proportions.” *American Statistician* 52: 119–126.
- Ammerlaan, H. S. M., S. Harbarth, A. G. M. Buiting, et al. 2013. “Secular Trends in Nosocomial Bloodstream Infections: Antibiotic-Resistant Bacteria Increase the Total Burden of Infection.” *Clinical Infectious Diseases* 56, no. 6: 798–805.
- Beghini, F., L. J. McIver, A. Blanco-Míguez, et al. 2021. “Integrating Taxonomic, Functional, and Strain-Level Profiling of Diverse Microbial Communities With Biobakery 3.” *eLife* 10: e65088.
- Cleary, J. G., R. Braithwaite, K. Gaastra, et al. 2014. “Joint Variant and De Novo Mutation Identification on Pedigrees From High-Throughput Sequencing Data.” *Journal of Computational Biology* 21: 405–419.
- Cock, P. J. A., J. M. Chilton, B. Grüning, J. E. Johnson, and N. Soranzo. 2015. “NCBI BLAST+ Integrated Into Galaxy.” *GigaScience* 4: 39.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Routledge. <https://doi.org/10.4324/9780203771587>.
- Dinan, M. A., B. R. Hirsch, and G. H. Lyman. 2015. “Management of Chemotherapy-Induced Neutropenia: Measuring Quality, Cost, and Value.” *Journal of the National Comprehensive Cancer Network* 13: e1–e7.
- Ecker, D. J., R. Sampath, H. Li, et al. 2010. “New Technology for Rapid Molecular Diagnosis of Bloodstream Infections.” *Expert Review of Molecular Diagnostics* 10, no. 4: 399–415.
- Fida, M., M. J. Wolf, A. Hamdi, et al. 2021. “Detection of Pathogenic Bacteria From Septic Patients Using 16S rRNA Gene Targeted Metagenomic Sequencing.” *Clinical Infectious Diseases* 73, no. 7: 1165–1172.
- Garcia-Vidal, C., C. Cardozo-Espinola, P. Puerta-Alcalde, et al. 2018. “Risk Factors for Mortality in Patients With Acute Leukemia and Bloodstream Infections in the Era of Multiresistance.” *PLoS One* 13, no. 6: e0199531.
- Gupta, S., A. Sakhuja, G. Kumar, E. McGrath, R. S. Nanchal, and K. B. Kashani. 2016. “Culture-Negative Severe Sepsis.” *Chest* 150, no. 6: 1251–1259.
- Gyarmati, P., C. Kjellander, C. Aust, M. Kalin, L. Öhrmalm, and C. G. Giske. 2015 Aug 13. “Bacterial Landscape of Bloodstream Infections in Neutropenic Patients via High Throughput Sequencing.” *PLoS One* 10, no. 8: e0135756. <https://doi.org/10.1371/journal.pone.0135756>.
- Gyarmati, P., C. Kjellander, C. Aust, Y. Song, L. Öhrmalm, and C. G. Giske. 2016. “Metagenomic Analysis of Bloodstream Infections in Patients With Acute Leukemia and Therapy-Induced Neutropenia.” *Scientific Reports* 6, no. 6: 23532.
- Hämäläinen, S., T. Kuittinen, I. Matinlahti, T. Nousiainen, I. Koivula, and E. Jantunen. 2008. “Neutropenic Fever and Severe Sepsis In Adult Acute Myeloid Leukemia (AML) Patients Receiving Intensive Chemotherapy: Causes and Consequences.” *Leukemia & Lymphoma* 49: 495–501.
- Kalich, B. A., J. M. Maguire, S. L. Campbell-Bright, et al. 2015. “Impact of an Antibiotic-Specific Sepsis Bundle on Appropriate and Timely Antibiotic Administration for Severe Sepsis in the Emergency Department.” *Journal of Emergency Medicine* S0736–4679: 937–943.
- Lagier, J. C., S. Edouard, I. Pagnier, O. Mediannikov, M. Drancourt, and D. Raoult. 2015. “Current and Past Strategies for Bacterial Culture in Clinical Microbiology.” *Clinical Microbiology Reviews* 28, no. 1: 208–236.
- Laupland, K. B. 2013. “Incidence of Bloodstream Infection: A Review of Population-Based Studies.” *Clinical Microbiology and Infection* 19: 492–500.
- Li, Y., J. Guo, H. Yang, H. Li, Y. Shen, and D. Zhang. 2021. “Comparison of Culture-Negative and Culture-Positive Sepsis or Septic Shock: A Systematic Review and Meta-Analysis.” *Critical Care* 25, no. 1: 167.
- Liesenfeld, O., L. Lehman, K. P. Hunfeld, and G. Kost. 2014. “Molecular Diagnosis of Sepsis: New Aspects and Recent Developments.” *European Journal of Microbiology and Immunology* 4: 1–25.
- de Naurois, J., I. Novitzky-Basso, M. J. Gill, et al. 2010. “Management of Febrile Neutropenia: ESMO Clinical Practice Guidelines.” supplement, *Annals of Oncology* 21, no. Suppl 5: v252–v256.
- Öhrmalm, L., M. Wong, C. Aust, et al. 2012. “Viral Findings in Adult Hematological Patients With Neutropenia.” *PLoS One* 7, no. 5: e36543. <https://doi.org/10.1371/journal.pone.0036543>.
- Opota, O., K. Jaton, and G. Greub. 2015. “Microbial Diagnosis of Bloodstream Infection: Towards Molecular Diagnosis Directly From Blood.” *Clinical Microbiology and Infection* 21, no. 4: 323–331.
- Paul, M., G. Kariv, E. Goldberg, et al. 2010. “Importance of Appropriate Empirical Antibiotic Therapy for Methicillin-Resistant *Staphylococcus aureus* Bacteraemia.” *Journal of Antimicrobial Chemotherapy* 65, no. 12: 2658–2665.
- Salzberg, S. L., and D. E. Wood. 2021. “Releasing the Kraken.” *Frontiers in Bioinformatics* 1: 808003.
- Sinha, M., J. Jupe, H. Mack, T. P. Coleman, S. M. Lawrence, and S. I. Fraley. 2018. “Emerging Technologies for Molecular Diagnosis of Sepsis.” *Clinical Microbiology Reviews* 31, no. 2: e00089-17.
- Song, Y., C. G. Giske, P. Gille-Johnson, O. Emanuelsson, J. Lundberg, and P. Gyarmati. 2014. “Nuclease-Assisted Suppression of Human DNA Background in Sepsis.” *PLoS One* 9, no. 7: e103610.

Song, Y., K. Perlman, and P. Gyarmati. 2022. "Microbial and Host Factors Contribute to Bloodstream Infection in a Pediatric Acute Lymphocytic Leukemia Mouse Model." *Heliyon* 8, no. 11: e11340.

Stewart, E. J. 2012. "Growing Unculturable Bacteria." *Journal of Bacteriology* 194, no. 16: 4151–4160. <https://doi.org/10.1128/JB.00345-12>.

Wood, D. E., and S. L. Salzberg. 2014. "Kraken: Ultrafast Metagenomic Sequence Classification Using Exact Alignments." *Genome Biology* 15, no. 3: R46.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supplementary Figure 1. Relative abundances of the microbial community standard as determined by different software, compared to the theoretical distribution (log scale).