Reviewer #1 (Comments for the Author):  
  
The manuscript "Multi-omic 1 Signatures of Host Response Associated with Presence, Type, and Outcome of Enterococcal Bacteremia" addresses an important scientific question. It presents several interesting results. However, there are some flaws in analytical methodology which have to be addressed to ensure which results and more importantly which conclusions are well supported.

We are deeply grateful to the reviewer for their invaluable feedback, particularly the insightful comments on the analytical methodology. These comments have been incorporated into a significantly revised version of our manuscript. Below, we respond to each comment in detail.

**Major comments**

1. Presently, individual features are identified as predictors of health vs. infection, E. Faecalis vs. E. Faecium, and survivorship vs. mortality. However, greater power could be achieved through a supervised machine learning approach, i.e., by constructing predictors that incorporate multiple factors, and the importance of individual factors could be determined from predictor composition. Moreover, training and evaluating a predictor on the same set of data can result in overfitting and an inaccurate estimate of predictor performance on new data. We therefore suggest the following:

1. For each of these comparisons, create one predictor or an ensemble of multiple predictors that utilize all markers passing some feature selection threshold.
2. Perform cross-validation on the training set after training to evaluate predictor performance within the training data.
3. Divide the samples into a training set and a test set. A commonly accepted ratio is 70% of samples in the training set and 30% of samples in the test set. Evaluate the predictor(s) on the test set to estimate predictor performance on new data. (Note: The test set would ideally be an entirely separate dataset. However, in the absence of such a dataset, this method allows predictor performance to still be estimated.)

We thank the reviewer for suggesting an alternative analysis strategy to analyze our proteomics and metabolomics data.

We agree that machine learning is an interesting alternative approach for analyzing this dataset. However, it poses challenges in datasets like ours due to the "low n, large p problem" given the relatively high number of predictors compared to the limited number of samples. While incorporating more samples would ideally help mitigate this issue, the resources required to obtain state-of-the-art proteomics and metabolomics data from over 100 subjects are not feasible for our study. Therefore, we have implemented the reviewer’s recommended approach using Lasso regression. This method allows us to build models with a reduced risk of overfitting, making it a suitable choice for our analysis.

**Incorporating a Machine Learning Approach:**

We constructed models for each of the three comparisons, incorporating the metabolite features that could be annotated by GNPS and the proteomics data. We used only features containing no missing values across all samples to build the models. The models were trained on a training set comprising 80% of the data, followed by 5-fold cross-validation to optimize performance. The accuracy of the models was then evaluated on a test set, which included 20% of the total data and was unseen by the model during training. We plotted the model coefficients greater than zero to identify the best predictors, focusing on the features that contributed most significantly to the classification outcomes. This produced the following figure, which has been added to the manuscript as a supplementary figure 11:



We observed that the predictive accuracy of the models was very similar to that of the best-performing individual features identified using our original methodology. The results for each comparison are summarized below:

* Healthy vs Infected
  + AUC of 1 for lasso regression model (Supplementary Figure X A)
  + AUC of 1 for SERPINA3, LRG1 (Figure 2C)
  + AUC of 1 for 13-OXO-ODE, and Threonylcarbamoyladenosine (Figure 3D).
* *E. faecalis* vs *E. faecium*
  + AUC of 0.81 for lasso regression model (Supplementary Figure X B)
  + 0.79 for RBP4, 0.76 for IGKV2-30 (Figure 4E)
  + 0.77 Retinol, 0.82 for PC(16:1/0:0) (Figure 5B).
* Survival vs Mortality
  + AUC of 0.83 for lasso regression model (Supplementary Figure X C)
  + 0.84 for HRG, 0.80 for FETUB (Figure 6D)
  + 0.74 for Decarnoylcarnitine, 0.78 for PC(16:0/0:0) (Figure 6F)

Furthermore, inspection of the model coefficients showed substantial overlap in the features identified as most important between the two approaches. (Supplementary Figures 4, 5, 6, 11).

While not all features were shared between the two methods, this is expected because lasso regression penalizes the addition of extra features to the model. Several models produced the same performance on the training set, implying that different features could be used interchangeably to generate models with the same performance. Overall, we believe this complementary analysis, suggested by the reviewer, shows that our findings are robust. The manuscript text has been updated to reflect this additional analysis.

**Additional Analysis: Clinical Metadata as Predictors**

The reviewer's suggestion prompted us to evaluate models using only clinical metadata to assess the relative contributions of biological data versus clinical variables in predicting outcomes.

**Type of EcB**

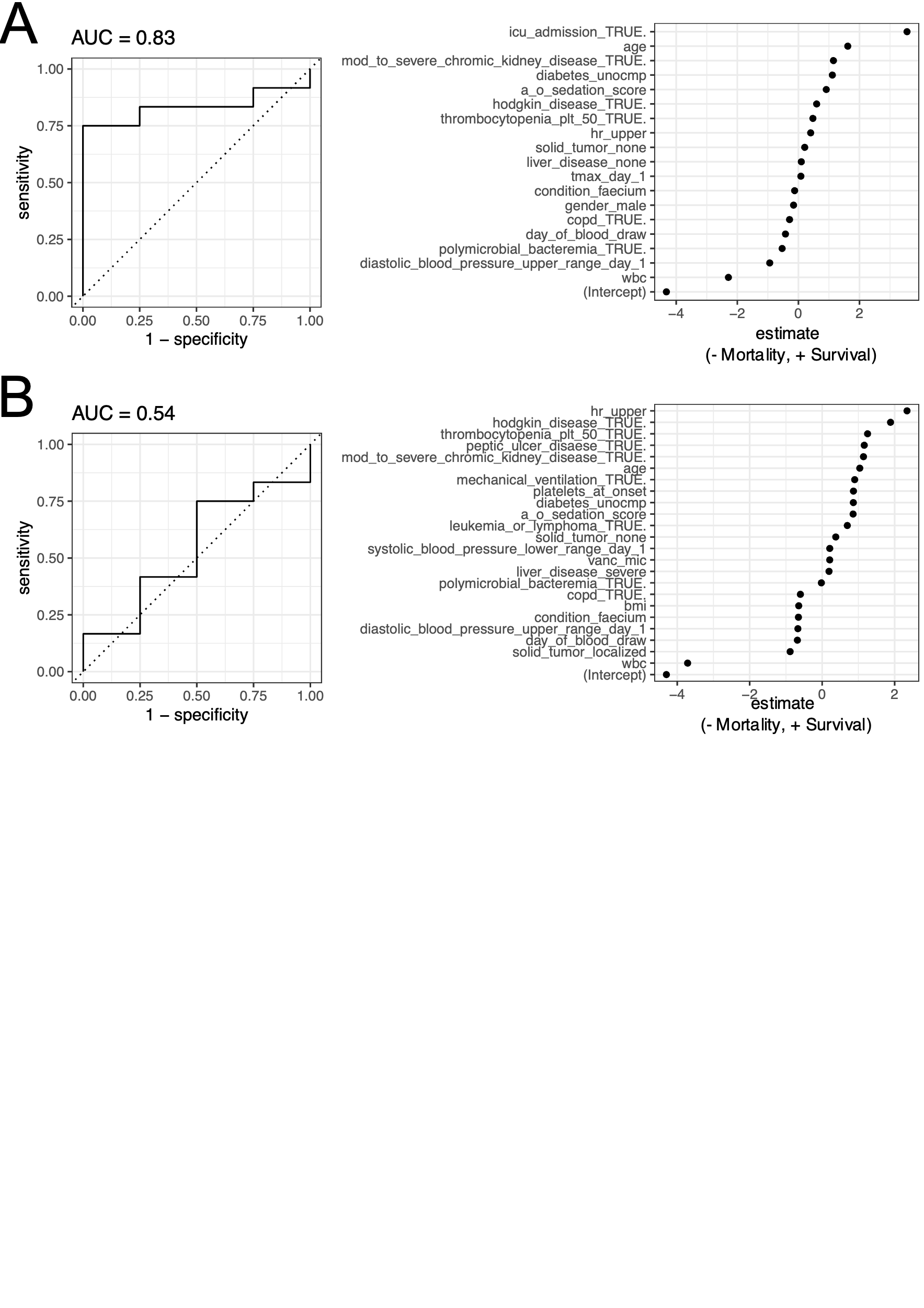
When using the full clinical metadata, we found that it predicted the type of EcB more effectively than our proteomics and metabolomics models. Upon inspecting the key features of this model, we observed that MIC values were the most essential features. This result is unsurprising, as *E. faecium* isolates are typically more resistant to ampicillin and vancomycin than *E. faecalis.* Moreover, the MIC values of *Enterococcus* isolates are not immediately available upon hospital admission and take days to detect using current techniques. At the same time, the proteomic and metabolomic profiles can be assessed sooner. When we excluded clinical metadata  available only after admission, transplant status emerged as a strong predictor of *E. faecium* or *E. faecalis*, producing a model with accuracy identical to our proteomics and metabolomics data. This finding complements our previous analysis (Supplementary Figure), where we observed that the reduction of our top individual biomarkers was significantly confounded by transplant status.



Overall, this complementary analysis strengthens our conclusion that, while the proteome and metabolome offer reasonable predictive power for distinguishing *E. faecalis* from *E. faecium* infections, they are unlikely to be clinically useful given that much greater accuracy can be achieved through traditional clinical microbiology methods. Additionally, we note that a far simpler model based on transplant status performs just as well.

**Mortality predictions**

When using the full clinical metadata to predict mortality, we found that it had similar performance to our proteomic and metabolomic models, as shown in panel A below. Unsurprisingly, upon examining the model, ICU admission emerged as a strong predictor of mortality.



When we removed ICU admission—since it is not available at the time of admission—we were left with a model that lacked predictive accuracy, as shown in panel B above. This highlights the potential clinical benefit of our proteomic and metabolomic predictors of mortality, which can be assessed upon admission. The above figure has been added to the manuscript as a supplementary figure. Overall, this complementary analysis strengthens our conclusion that biomarkers acquired on admission may be clinically useful in predicting enterococcal bacteremia patient mortality.

2. It seems there is a design problem with comparison between infection and healthy individual. Control samples consist of plasma collected from blood bank volunteers, while EcB samples were collected from patients at UW Health. This discrepancy in patient and sample collection circumstances may be a significant source of confounding. Although steps are taken to identify confounding variables from patient metadata, it is nonetheless probable that further unidentified confounding variables remain.

There is potential solution at least partially overcome this problem. If we assume that the transition between health and infection shares a molecular basis with that between survivorship and mortality, then we would expect features truly caused by EcB to be shared between the two comparisons and to have the same direction (up or down) in both. We therefore suggest the following:

1. Create a contingency table which sets the two comparisons against each other.
2. If most features share a direction of change, use these features for the construction of predictors.
3. Otherwise, it is more likely that confounding variables will play a significant role in the infected/healthy analysis, and so it may be wiser to place the greater focus on the E. Faecalis vs. E. Faecium and survivorship vs. mortality analyses performed within the hospital dataset.

We thank the reviewer for bringing up this point. We agree that the logistical constraints of sample collection when comparing healthy samples obtained via blood bank volunteers and EcB samples are not ideal. However, on a practical level, these are the only samples from healthy individuals we received for this study. Even though there were slight differences in sample collection methodology between the two sites, we believe it is reasonable to expect the differences in the underlying biological material to be the significant source of variation between these groups. We believe that the benefits of including a healthy cohort for comparison far outweigh the downsides inflicted by having slight differences in sample collection. The major confounding difference between the samples collected from the healthy volunteers and the EcB patients was the tube type. We expect this to impact the proteomics and metabolomics data differently.

**Proteomics**

When examining the proteomics data, we found little indication that the observed differences between healthy and infected individuals were due to variations in sample collection. A high-level revealed that the elevated proteins in infected patients are primarily associated with acute-phase and inflammatory responses, which aligns with expected biological reactions during infection. For example, clinically established markers of general infection, such as C-reactive protein and serum amyloid A, were among the most differentially abundant proteins (Supplementary Figure 3).

We appreciate the reviewer’s thoughtful suggestion to assess potential confounding factors. However, we respectfully suggest that the assumption that the transition from health to infection shares a molecular basis with survivorship and mortality may not be entirely accurate. While it is intent that some molecular features are could overlap between these biological processes, the extent of such overlap is unclear.

Nevertheless, we reasoned that while a weak correlation would not definitively indicate confounding, a strong correlation would provide compelling evidence supporting the assumption that the data is not substantially confounded. Therefore, we conducted an analysis similar to the reviewer's suggestion by correlating the log2 (Infected/Healthy) ratios with the log2(Mortality/Survival) ratios. This analysis was performed across all proteins, as well as limited to thoe that were significantly different (adjusted p-value ≤ 0.05) in both comparisons.

A collage of graphs and diagrams

Description automatically generatedA collage of graphs and diagrams

Description automatically generated

We observed a strong correlation, suggesting that the proteomic profiles share a molecular basis between infection and mortality. This correlation was even more pronounced when the analysis was restricted to proteins that were significantly different (adjusted p-value ≤ 0.05) in both comparisons. In conclusion, this evidence indicates that any differences in sample collection had minimal impact on the proteomics data.

**Metabolomics**

When considering the metabolomics data, we had greater concerns about confounding variables, as plastic molecules from the containers could leach into samples and be detected via our untargeted approach. Indeed, our analysis showed that several of the most differentially aboundant metabolites between healthy and EcB samples were plasticizers, as explicitly mentioned in the discussion. These molecules were almost certainly artifacts from differences in the plastics used for sample collection, rather than biologically interesting variations we set out to capture.

We conducted the same correlation analysis with the metabolomics data and observed only a slight correlation between infection and mortality, suggesting either a weak molecular connection or substantially confounded from sample collection. While certain metabolites clearly showed evidence of confounding (e.g. phthalic anhydride, dibutyl phthalate), other metabolites appeared to accurately reflect biologically relevant differences between healthy and infected patients. For example, we detected significant differences in several bile acids between healthy those with EcB (Figure 3C).

A collage of graphs and diagrams

Description automatically generated

A close inspection of the unstructured nodes in the clinical metadata shows that 17 patients (S5, S8, S11, S15, S17, S19, S20, S27, S28, S29, S32, S36, S58, S68, S69, S71, S100) had some form of bile impairment that would be expected to increase bile acid levels in plasma. Conditions included cholangitis, biliary stricture, cholecystectomy, biloma, choledocholithiasis, cholangiocarcinoma, biliary infection, and biliary leak, among others.). Combined with existing literature linking increased bile acid levels to infection, this suggests that our metabolomics data can effectively detect relevant biological differences between healthy and infected patients, potential issues caused by differences in sample collection tubes.

We propose that confounders in the metabolomics data are likely specific to individual metabolites: some differences may be artifacts of sample collection, while others reflect genuine biological variation. Therefore, we believe that the best approach is to report the metabolomic results while explicitly acknowledging that certain differential features may be influenced by the different types of tubes for storing healthy and infected samples.

To address this, we have added the following text to our discussion section.

“*These findings suggest that some metabolite features distinguishing healthy from infected samples may be confounded by differences in sample collection. Therefore, caution is advised when interpreting metabolites that show differences between healthy and infected samples in this dataset. Metabolites identified as industrial chemicals, which are not typically expected to be found in blood, are likely unwanted artifacts*.”

**Minor comments**

1. The sizes of the mortality and survival groups in Figure 1 do not sum to the same value as the E. Faecalis vs. E. Faecium groups (32 E. Faecalis + 44 E. Faecium = 76, while 17 mortality + 57 survival = 74). We suggest that this disparity be corrected if it is an error, or explained if it is not.

We thank the reviewer for bringing this discrepancy to our attention. It was caused by an unintended error in our clinical metadata, which affected our analysis. Although the issue had a limited impact on the results, it was an important oversight that needed correction.

For 2 of the 74 EcB patients, designated S49 (patient id # 78) and S76 (patient id # 79), we noticed that the mortality values were missing (NA) in our metadata, leading to the discrepancy in patient numbers. Further manual inspection revealed that other clinical metadata fields were also missing for these samples. This was unexpected, as clinical data for all specimens had been collected as part of the study. We suspected that this issue resulted from manual manipulation of Excel files during the data cleaning process and therefore reverted to the original version of the clinical metadata.

Upon comparing the original and cleaned versions, we identified a mistake in the Excel file for patient IDs S49 (ID 78) and S76 (ID 79). Although the type of EcB bacteremia had been correctly assigned (78 = faecalis, 79 = faecium), the remaining metadata fields had been inadvertently deleted. We resored the missing clinical metadata, took the opportunity to verify the accuracy of the remaining metadata, and reran our analysis using the updated, complete dataset. Figures and text have been updated accordingly.

1. The subsection "Human Plasma Samples" under the section "Materials and Methods" states that data was collected from 32 patients with E. faecium bacteremia and 44 patients with E. faecalis bacteremia, totaling 76 patients (line 128). However, the following subsection "Clinical Data Collection", states that there were 83 patients with enterococcal bacteremia (line 137). We suggest that this disparity be corrected if it is an error, or explained if it is not.

We thank the reviewer for noticing this disparity. The number 83 was written in error; it was intended to be 76. The updated version of the text has been modified to reflect this.

1. In figure 1B, we suggest either placing the text in the title of each plot beneath its x-axis or removing the x-axis label to avoid duplication. Further, the title of the plot "Sensitivity to Vancomycin" is cut off.

We thank the reviewer for this suggestion to improve the figure's aesthetics. These recommendations have been added to the new version of the figure.

1. On line 46, we suggest that the URL be made a hyperlink.

This has been corrected in the manuscript.

1. On line 46, the sentence is missing a period.

This has been corrected in the manuscript.

1. On line 134, a comma is missing after the closing paren.

This has been corrected in the manuscript.

1. On line 138, we suggest removing the comma after the word "identified".

This has been corrected in the manuscript.

1. On line 191, please check whether "cystines" should instead be "cysteines". It's possible that there may have been a misunderstanding on our end.

This has been corrected in the manuscript.

1. On line 199, we believe that the word "Analysis" should not be capitalized.

This has been corrected in the manuscript.

1. On line 208, The text "20oC" contains the character "o" rather than a degree symbol".

This has been corrected in the manuscript.

1. On line 211, the text "-80{degree sign}C{degree sign}C" contains a duplicate instance of "{degree sign}C".

This has been corrected in the manuscript.

1. Lines 218-221 use "minute" where they should say "minutes". We also suggest that semicolons are used to break up the list, such as "...following gradient: 0 to 1 minutes, 5% B; 1 to 7 minutes, a linear increase from 5% to 100% B; 7 to..."

This has been corrected in the manuscript.

1. On lines 233 and 234, the spacing around the equals signs is inconsistent.

This has been corrected in the manuscript.

1. On line 244, we suggest changing "GNPS FBMN option" to "the GNPS FBMN option".

This has been corrected in the manuscript.

1. On line 274, we believe "across any" should instead be either "for any" or "in any".

This has been corrected in the manuscript.

1. On line 281, a comma is missing after the word "randomforest".

This has been corrected in the manuscript.

1. On line 282, "R-package" should instead be "R package"

This has been corrected in the manuscript.

1. On lines 314 and 316, the temperatures "37C" and "50C" lack degree symbols.

This has been corrected in the manuscript.

1. On line 348, "GEForce" should instead be "GeForce".

This has been corrected in the manuscript.

1. On line 369, we suggest using "EcB" instead of "EB" for consistency.

This has been corrected in the manuscript.

1. On line 380, "Enterococcous" should instead be "Enterococcus".

This has been corrected in the manuscript.

1. On line 401, "1x1025" should instead be "1x10-25".

This has been corrected in the manuscript.

1. On line 417, "(30" should be "(30)".

This has been corrected in the manuscript.

1. On line 420, we believe the word "one" should be removed.

This has been corrected in the manuscript.

1. On lines 430 and 431, "gene ontology" should be capitalized to "Gene Ontology".

This has been corrected in the manuscript.

1. On line 461, a period is missing after the second close paren.

This has been corrected in the manuscript.

1. On line 484, there is an extra space before the em dash.

This has been corrected in the manuscript.

1. On line 512, the text "are leveraged" should be deleted.

This has been corrected in the manuscript.

1. On line 589, the word "of" should be inserted between the words "analysis" and "unsupervised".

This has been corrected in the manuscript.

1. On line 595, a period is missing after the close paren.

This has been corrected in the manuscript.

1. On line 662, a comma is missing between the words "biospecimen" and "and".

This has been corrected in the manuscript.

1. On line 758, a space is missing between the period and the word "Interestingly".

This has been corrected in the manuscript.  
  
Reviewer #2 (Comments for the Author):  
  
The authors of this manuscript performed a comprehensive comparative analysis of plasma samples from 3 cohorts: 29 healthy volunteers, 32 patients with E. faecium bacteremia and 44 patients with E. faecalis bacteremia, with respect to the content of proteins and metabolites. Their main goals were to characterize the host response to Enterococcus bacteremia (EcB), to determine the systemic differences between E. faecalis and E. faecium bacteremia, and to identify the systemic response markers associated with increased mortality. They used an unbiased approach based on a comparison of proteoms and metaboloms of the collected plasma samples, supported by results of characterization of Enterococcus isolates from the patients and the selected case history parameters of each patient (including gender, the Charleston comorbidity index, duration of bacteremia, day of blood draw, and mortality during admission). Based on the used methodologies they could capture from among hundreds of proteins and metabolites analyzed, those that are increased or decreased in all or in particular types of bacteremias.   
They also compared their data concerning EcB with their earlier data concerning Staphylococcus aureus bacteremia. Based on the results of these comparisons they identified among detected plasma proteins and metabolites, the proteins and metabolized increased or decreased in bacteremia patients, proteins or metabolites common for all types of bacteremia, and proteins or metabolites with changed levels specific for E. faecalis, E. faecium or S. aureus bacteremia. Gene ontology (GO) enrichment analysis of proteins differentiating all or particular bacteremias allowed the authors to identify biological processes characteristically altered in all or particular kinds of studied bacteremias. Although the authors confirm the value of two previously approved markers for the detection of infection, most of their results are novel.   
The introduction section provides sufficient background and motivation to perform the studies described by the authors. The methods are described in sufficient detail. The results are mostly presented in the form of graphs that allows one to see the differences in plasma proteomes and metaboloms observed by the authors in infected versus healthy individuals. Statistical analysis is satisfactory. Additionally, the authors discuss the differences of insufficient statistical significance that can be potentially useful if further analyzed. The manuscript is a valuable contribution to the knowledge concerning common processes in the host response to all three kinds of bacteremias studied by the authors and processes in the host response characteristic for particular kinds of bacteremias. Additionally, it has a practical value as a guide to developing plasma analysis-based diagnostic methods that could be helpful in the fast detection of EcB cases, identification of the infecting Enterococcus species and the prediction of infection fatal outcome. An important observation of the authors was that most of the E. faecalis isolates were resistant to vancomycin, while vancomycin resistance among E. faecium isolates was rare, which may help in the fast decision concerning antibiotic therapy. In general, the manuscript is interesting to read**. Only the Discussion section seems to be too wordy, and could be shortened.**

We are happy that the reviewer found our study to be a valuable contribution to the field, with practical value and an interesting read. We appreciate the reviewer's suggestion to shorten the discussion section. In the revised manuscript, we aimed to be more concise while incorporating additional information recommended by the reviewers.

My additional specific comments are below.  
  
L. 41. Please explain the abbreviations at their first use in the text

Thank you for bringing this to our attention. It has been fixed in the updated version.   
  
L. 311-312. Please provide a reference.

Please see the below comment.  
  
L. 312. "10"? What do you mean by that? Please explain.

We apologize. This was intended to be a citation, but it got disconnected from our reference manager. The updated version corrects this.  
  
L. 317. Please provide the catalog number.

This has been added as requested in the updated version.  
  
L. 370. In the classification of patients with bacteremia, the authors took into account the duration of bacteremia. Could the beginning of persistent bacteremia be precisely determined by the authors? Please explain.

Bacteremia duration was defined by the first positive culture until clearance, which is consistent with the clinical definition.

L. 380. Could the authors expect the detection of enterococcal proteins in blood samples? What was the range of titers of enterococci in patients' blood samples?

The titer of bacteria in bacteremia-derived blood is expected to be very low. The typical bacterial loads in adult septic patients are ≤100 bacterial cells (colony forming units, CFU) per mL of blood.

Please see:

Yagupsky P, Nolte FS. Quantitative aspects of septicemia. Clin Microbiol Rev. 1990 Jul;3(3):269-79. doi: 10.1128/CMR.3.3.269. PMID: 2200606; PMCID: PMC358159.

In our study, the exact levels were not attempted to be quantified. Instead, we used standard clinical microbiology methods that amplify bacteria via propagation for subsequent identification of the bacterial species. Due to their expected low abundance, it was unsurprising that bacterial proteins were not detected using our DDA proteomic methods, as their stochastic nature is more likely to detect highly abundant proteins. The text “Notably, no proteins were found to map to the *Enterococcus* proteomes” has been modified to “No proteins were found to map to the *Enterococcus* proteomes” to eliminate confusion.

L. 561. With "E" ? Please provide a complete abbreviation.

We apologize for this oversight. The text has been corrected to *E. faecium*  
  
L. 589-595. What was the reason of the differences between the initial and later findings?

Thank you for bringing this to our attention. The previous version did not accurately convey our findings. We intended to clariy that, while no high-level patterns from unsupervised hierarchical clustering could differentiate mortality by overall proteomic or metabolomic profiles, individual proteins were significantly associated with mortality outcomes.

To better reflect this, the text has been changed from:

“*Initially, our unsupervised hierarchical clustering analyses of proteomics (****Figure 1C****) or metabolomics data (****Figure 1D****) revealed no mortality association with overall proteomic or metabolomic profiles. However, upon further analysis, we identified specific proteins significantly associated with mortality outcomes*.”

to:  
  
“*Our unsupervised hierarchical clustering analyses of proteomics (****Figure 1C****) or metabolomics data (****Figure 1D****) revealed no high-level associations of mortality with overall proteomic or metabolomic profiles. However, when conducting feature-level analysis, we identified specific proteins significantly associated with mortality outcomes*.”  
  
Figure 3A. What do exactly the authors mean by "spectral match" and "no spectral match" in this Figure? This can be guessed intuitively. However, clear explanations by the author would be helpful for the readers.

We have added text to the figure legend to explain better what the spectral match vs no spectral match legend means. “No spectral match (red points) indicates features that did not produce a spectral match to any of the metabolites in the GNPS database, while spectral match (blue points) indicates features that produced a spectral match to metabolites in the GNPS database.”

Figure S2. Fonts in this Figure are too small to be readable. They should be at least slightly increased.

Due to the amount of information in this figure, it will be submitted as a vector-based image that allows the reader to zoom in on the network of interest.

The authors compared their results with the previously described results obtained with similar methods for S. aureus bacteremia. In my opinion, the conclusions based on this kind of comparison should be made with care and used as suggestive rather than indicating something. The manuscript concerning staphylococcal bacteremia was published by the authors a few years ago. Thus, in the discussion section, the authors should at least briefly, compare the methodology of studies performed for that manuscript and the current manuscript.

We thank the reviewer for raising this important point. Our intention was to present the conclusions comparing the S. aureus results to the EcB results to be suggestive rather than definitive and to highlight the limitations of comparing these two studies.

“*Technical considerations as to the experimental design of our S. aureus and EcB study prevent a direct comparison of these two datasets, but the differences we observe relative to healthy patients* ***suggest*** *the existence of features of host response that could be exploited to distinguish these types of bacteremia. A larger study designed to compare these directly and other bacteremia types is warranted to uncover these differences with greater confidence through direct comparisons*.”

In an attempt to make our intentions clearer to the reader while incorporating a brief comparison of the methodologies of the two studies we have revised the text to the following:

“*Although both the S. aureus and EcB studies were conducted in our lab using similar high-resolution TMT proteomic workflows, we cannot directly compare the two datasets due to limitations inherent in TMT-based proteomics. TMT normalization requires a 'bridge' channel that consists of a pooled aliquot from all samples in the study. No shared pooled sample was available since our studies were conducted independently at different times. Nevertheless, the observed differences concerning healthy patients suggest distinct features of the host response that could help differentiate between these types of bacteremia. A more extensive study specifically designed to compare these and other types of bacteremia is warranted to confirm this through direct comparisons*.”

The list of references requires extensive editing and supplementation with missing details concerning quite many listed publications.

L. 1022: Please provide the details of the cited publication

We apologize for this oversight. It appears that our reference manager was not functioning as intended. This reference has been fixed.

L. 1056. Does the reference cited apply to host-microbe interaction as suggested in the manuscript text? Please verify.

We believe that this citation is fitting.

The text in question is:

Host factors are widely recognized as critical determinants of the outcome of host-microbe interactions15 and have been used as prognostic biomarkers to predict patient outcomes in a variety of diseases, ranging from COVID-19 to cancer 16,17

The citation on line 1056 is:

17. Dhanasekaran, S. M. et al. Delineation of prognostic biomarkers in prostate cancer. Nature 1057 412, 822–826 (2001).

We believe that this citation is appropriate as we are claiming that host factors have been used as prognostic biomarkers in a variety of diseases, from COVID-19 to cancer. The reference is from a paper that reports prognostic biomarkers in prostate cancer.

L. 1058, and elsewhere in the text: Please italicize genus and species names

This has been corrected.

L. 1090, 1092, and elsewhere in the text. Please supplement the list of references with multiple missing details.

We apologize for this oversight. Most of these malformed citations are for GitHub repositories containing software used in the manuscript. Our reference manager did not cite them as we intended. This has been corrected in the updated version.

Fonts in Figure S2 are too small and should be increased for better visibility.

This has been addressed in the modified version of the manuscript.

Figures S8 and S9 are hardly readable. They should be provided in a better resolution.

Thank you for bringing this to our attention. We have addressed it by increasing the DPI to 300 for both of these figures.