# Questions for Meeting

* use of term test material

## Terminology Currently Used

**NOTE Improper use of terminology - need to discuss appropriate alternatives** \* test material - material used to evaluate a microbial detection method  
\* target genome - test material genome  
\* specificity - proportion of reads correctly classified for the single genome datasets \* sensitivity - detection of contaminants at different concentration levels \* Reproducibility - changed to bioinformatic pipeline, computational reproducibility

**Alternative Terminology** \* Target material and target genome \* Sensitivity for specificity : TP/(TP + FN) \* Detection limit for sensitivity

# Introduction

* Appropriate references for related methods
  + References for PCR based purity methods
  + IRMM genomic material quality control
  + LANL, LNNL, LGC - whole genome sequencing purity assessment?
  + CLSI - standard method?
* Bioinformatic method database limitations and assumptions
* Limitations of 16S based methods
* Add text about taxonomic and phylogenetic ambiguities and level of resolution
* Text about cells vs. DNA being detected lines 98-100

# Methods

* Each section - what was done and why
* Terminology
  + genomic purity vs. material purity
  + specificity and sensitivity \* make sure terms coincide with VIM and/or Eurochem
* Move definitions from the results to methods
* Add details to the sub-processing
* Reproducibility section
  + rename section - to what?????
  + add details
* Need better scale for specificity

# Results and Discussion

* Expand on Genus and Species level similarities
* Reiterate terminology with consistent use between methods, results, and figures

Additional Points to include \* Use of other taxonomic classification methods are likely to have different sensitivity and specificity results \* Need to evaluate the suitability of the reference database for used the genome and contaminant of interest. \* Work to further expand the taxonomic database to include genomes from uncultured organism using either metagenome datasets for single cell datasets along with efforts to address issues related to taxonomic ambiguities will help to improve the method applicability.

# Conclusions

* References/ accuracy of other classification methods
* Address suitability of reference database - make sure appropriate strains are included. A little bit of a chicken and egg problem, need pure strains to sequence for reference materials but also database to assess purity. Looking for rare variances within the population of cells used to generate the material - can cite PEPR publication
* expand on the last point
* Write conclusions based on the work that you did, not what you would like to have done or think you did.

With the drop in the cost of whole genome sequencing, viable method for evaluating the genomic purity of a microbial test material. Using similated sequening data .... that the method had appropriate level of specificity (will update with appropriate terminology) for use in detecting contaminants of different genus than the test material, and even species level detection for some genera e.g. Listeria. Using simulated whole genome sequencing data combined with a metagenomic taxonomic classifier we were able to detect the presence of contaminants at

# Terminology

Reviewed GUM and VIM as well as references from David Duewer's presentation at SPIN (book Chemical Identification and its quality assurance) p. 70 Definition of sensitivity and specificity VIM - definitions for nominal property, measurement process covers methods used for examination of a nominal property Specificity - not defined by VIM and not recommended by IUPAC David Deuwer's presentation Validation - includes everything wanted, excluded everything unwanted, measurement consistent with ID criteria Terms Detection limit Missclassification rate

Recommendations \* Figures 1 and 2, y axis - match proportion, missclassification rate??? (using inverse) \* For specificity section - what about suitability test - Ref Standards PPT

# Use Cases

1. Reference material and control material purity
2. Evaluation of inclusivity and exclusivity panels

# Conclusion

## Inclusivity and Exclusivity Panel Use Case

Use of method for general reference material, reference strain purity Pathogen detection methods used in high stakes setting including clinical, biothreat, and environmental monitoring require approriate validation to ensure confidence in the results. Inclusivity and exclusivity panels are used to validate pathogen detection methods ability to correctly detect the pathogen (inclusivity strains) of interest without triggering a false positive for closely related non-pathogenic strains (exclusivity strains). The genomic purity of the strains, presences of strains other then the target strain, included in the inclusivity and exclusivity panels must be evaluated. We have show that whole genome sequencing along with a taxonomic read classification algorithm can be used to assess target materials for the presence of genomic material from strains other then the target strain. The taxonomic classification algorithm used in the study had limited specificity and may not be appropriate for strain level differentiation. To apply this method to evaluate inclusivity and exclusivity panels, the method should first be evaluated using similated data in a similar method to that shown here. Additionally, the reference database used should include the appropriate genome sequences for all inclusivity and exclusivity strains. Finally, mixtures of DNA and cells from strains in the inclusity and exclusivity panels can be used to validate the method for use in evaluating the purity of inclusivity and exclusivity strains. Users must also consider the method detection limit, specifically in relation to the limit of detection of the pathogen detection method.