

# DETECTING PHENOTYPES WITH RAMAN SPECTROSCOPY



Cristina García-Timmermans<sup>1</sup>, Benjamin Buyschaert<sup>1</sup>, Peter Rubbens<sup>3</sup>, Frederiek-Maarten Kerckhof<sup>1</sup>, Andre Skirtach<sup>2</sup>, Nico Boon<sup>1</sup>

<sup>1</sup>Center for Microbial Technology and Ecology (CMET), Ghent University, Belgium

<sup>2</sup>Department of Nano & Biophotonics, Ghent University, Belgium

<sup>3</sup>KERMIT, Department for Mathematical Modelling, Statistics and Bioinformatics, Ghent University, Belgium

## PHENOTYPE DETECTION WITH RAMAN SPECTROSCOPY

Raman spectroscopy is a multiparametric tool that makes a unique fingerprint for each bacteria based on their molecular expression profile. This allows for phenotypic classification. The sensitivity of Raman calls for a standardized sample preparation. Otherwise, results could lead to a wrongful subpopulation classification.

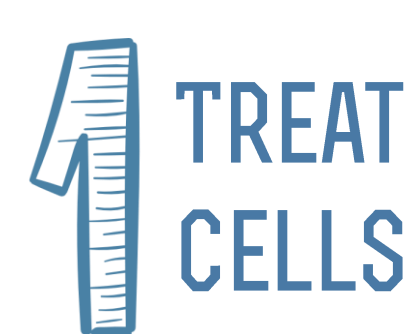
In this work, we propose a 'Raman checklist' to raise awareness of possible sources of variability, **reduce wrongful classification and improve reproducibility.**

## CHECKLIST FOR STANDARDIZATION

We have developed a checklist for identifying phenotypes in label-free bacteria using Raman spectroscopy.

Our findings demonstrate how the bacteria fingerprint can be altered by small manipulations such as the storage of a sample, the drying time on the slide or extra steps of resuspension and centrifugation. Spectral changes can be identified using supervised methods (i.e. Random Forests) and unsupervised methods (i.e. hierarchical clustering). These changes create artificial subpopulations that researchers could mistake as biological phenotypes.

## METHODS:



We tested possible sources of variability ...



...computed how this variability affects hierarchical clustering and Random Forest classification...

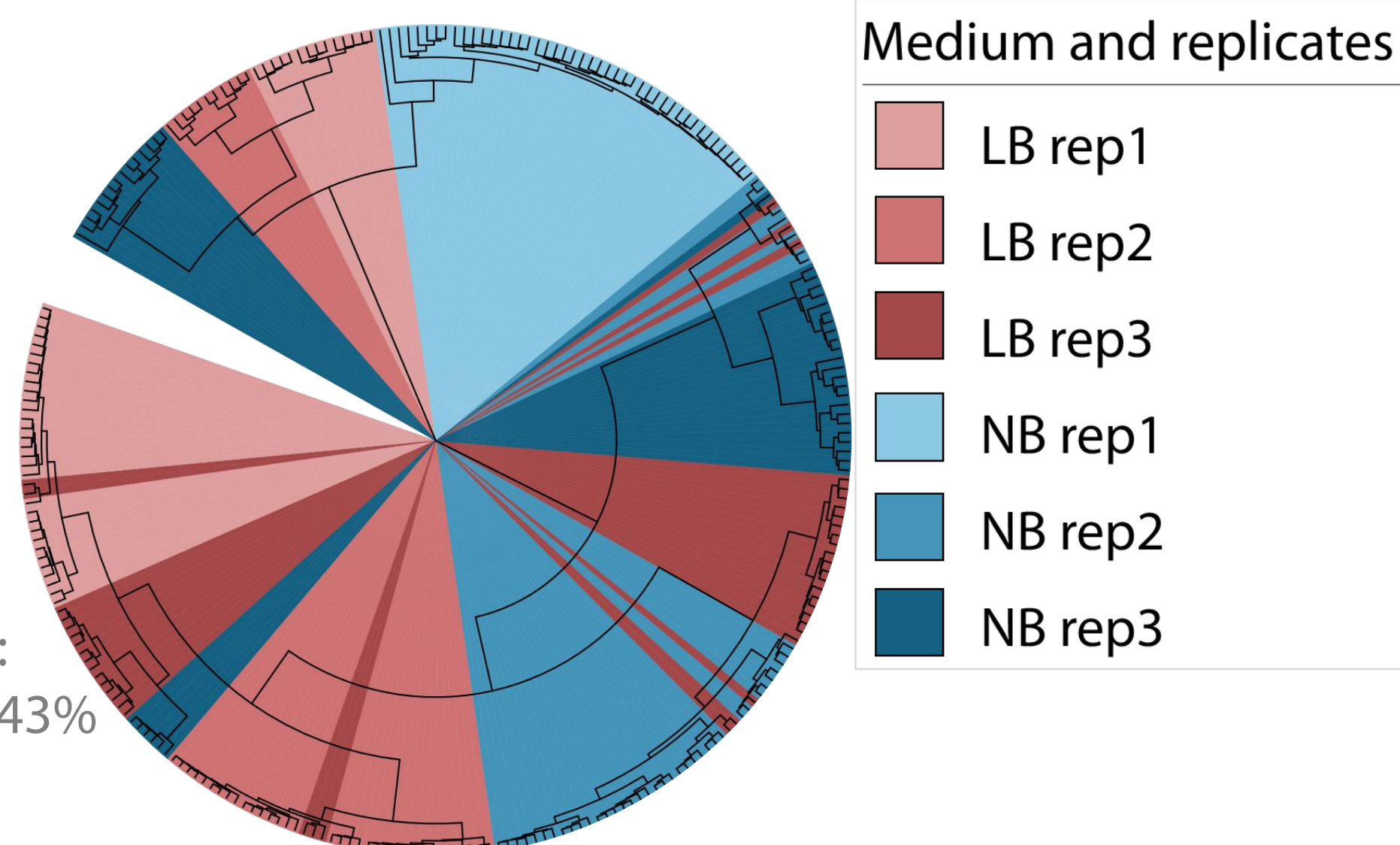


... and produced a checklist to reduce technical noise and improve reproducibility

## SOURCES OF VARIABILITY AND THEIR IMPACT ON THE CLASSIFICATION:



MEDIUM

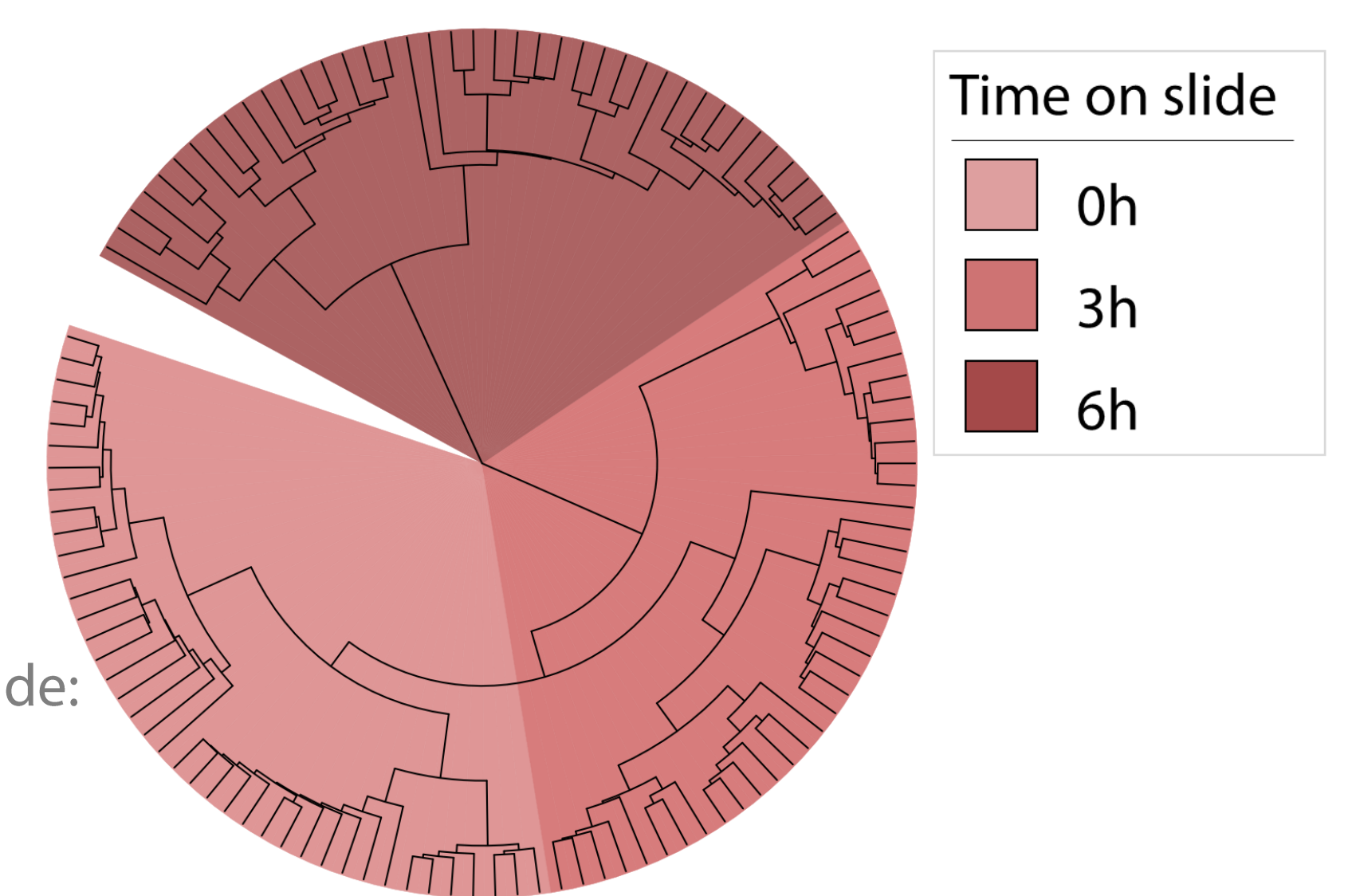


Accuracy prediction LB and NB:  
- Hierarchical clustering: 88.43%  
- Random Forest: 98.5%

We treated *E. coli* with LB and NB to induce two phenotypes. Random Forest and hierarchical clustering were able to detect these phenotypes even across replicates.



DRYING ON SLIDE

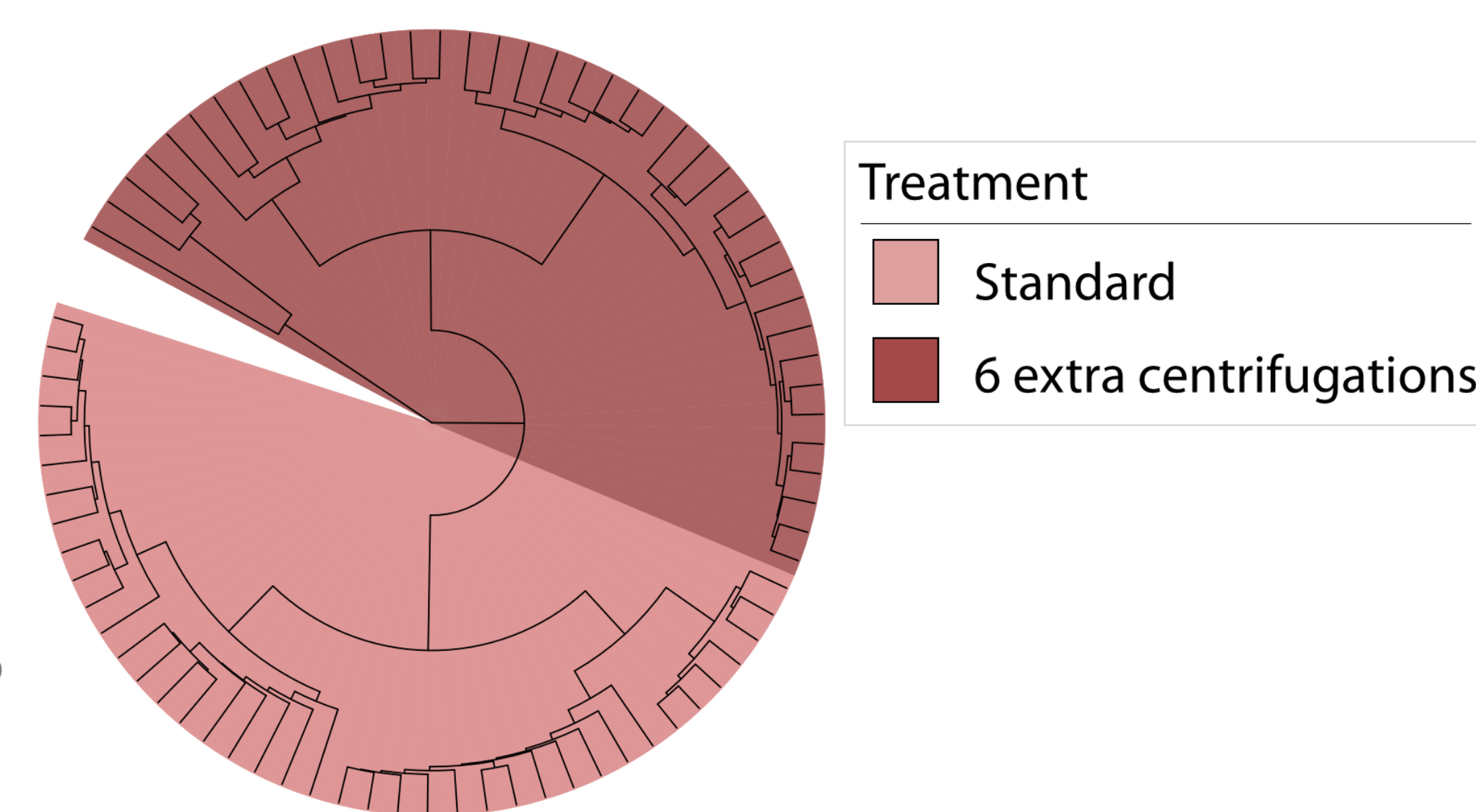


Accuracy prediction of time on slide:  
- Hierarchical clustering: 100%  
- Random Forest: 98.5%

We would recommend to analyze freshly prepared slides as the drying on the slide impacts the spectra.

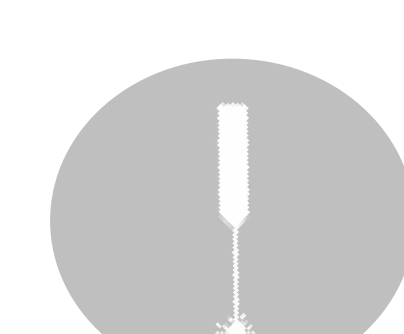


EXTRA CENTRIFUGATION & RESUSPENSION



Accuracy prediction standard vs extra manipulation:  
- Hierarchical clustering: 100%  
- Random Forest: 98.5%

These manipulations –centrifugation and resuspension- impact the spectra, and could lead to wrongful phenotypic classification.



INSTRUMENT

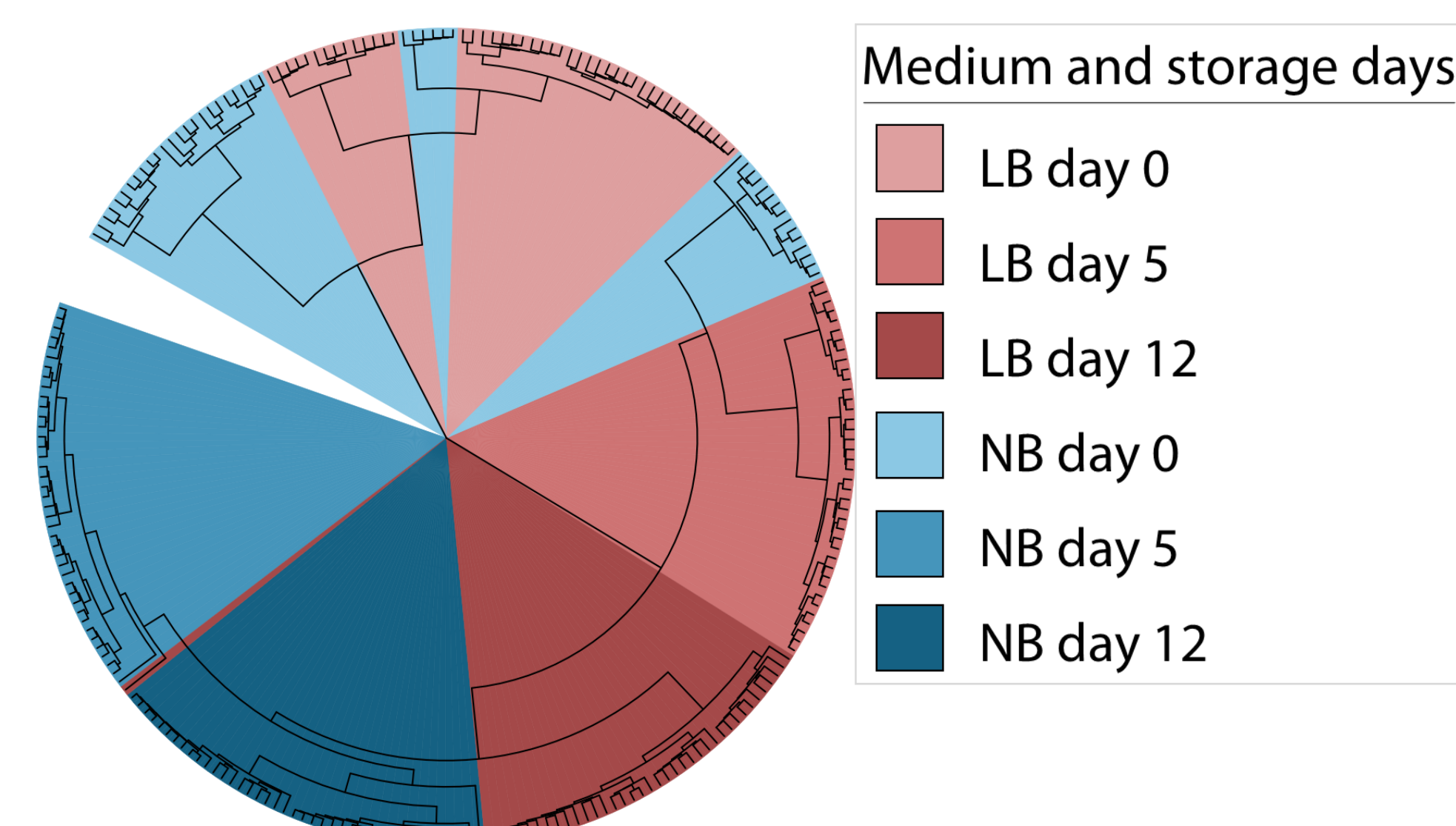


FIXATION

It has been reported how instrument variations [2] and fixation methods [3] can also have an influence on the spectra



STORAGE TIME

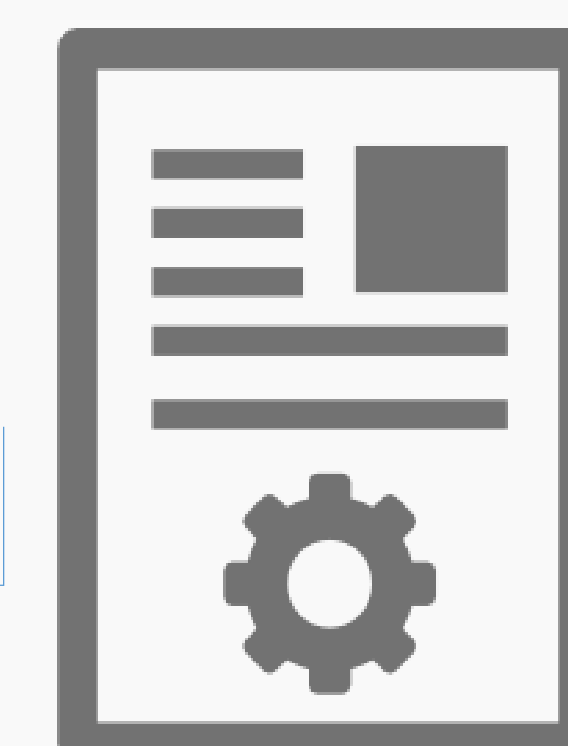


Accuracy prediction of days stored:  
- Hierarchical clustering: 100%  
- Random Forest: 98.4%

Different storage times can also influence the clustering, leading to wrongful phenotypic classification.

## RAMAN CHECKLIST

Samples



Experiment overview

Instrument specifications

Data analysis

Findings have been summarized in a checklist that researchers can use to improve data reproducibility

## REFERENCES

- [1] Garcia-Timmermans *et al.*, (manuscript in preparation)
- [2] Hutsebaut *et al.*, 2005
- [3] Read *et al.*, 2015

