



UNIVERSITÄT  
LEIPZIG

# Beerinformatics

February 12rd 2019, Bled

Jörg Fallmann, Stephan Bernhart



## ONCE UPON A TIME

### BeerDeCoded: Exploring the beer metagenome

3 November, 2017

[Jonathan Sobel](#)

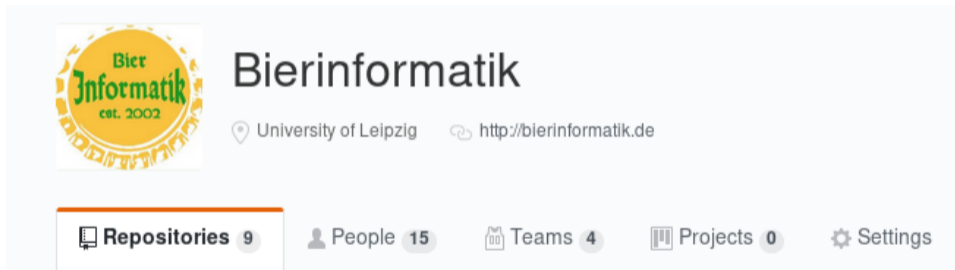


FICCO Research

*Jonathan Sobel shares his scientific quest to understand beer at a molecular level*



# BEERINFORMATICS



The screenshot shows the GitHub profile for Bierinformatik. The profile picture is a circular logo with a yellow background and green text that reads "Bier Informatik est. 2002". The name "Bierinformatik" is displayed in a large, bold font. Below the name, the location "University of Leipzig" and the website URL "http://bierinformatik.de" are shown. A navigation bar at the bottom of the profile includes the following items: "Repositories 9" (highlighted with an orange underline), "People 15", "Teams 4", "Projects 0", and "Settings".

# BRING SCIENCE BACK FROM THE STREETS

StreetScienceCommunity / StreetScienceCommunity.github.io

Watch 3 Star 2 Fork 5

Code Issues 0 Pull requests 0 Projects 0 Insights

Bring science to the "street" <https://streetscience.community>

113 commits 2 branches 0 releases 6 contributors MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

## THE PLAN



## THE SPONSOR



## THE SAVIOUR



## **INTRODUCTION**

### **SETTING**

WIMP analysis

Mapper Comparison

Replicates and yeast composition

### **Conclusion and Outlook**



## MOTIVATION

- BEER
- Use state of the art techniques in practical course to motivate and entertain our students
- ++ BEER
  - Compare nanopore results to published Illumina results
- BEER
  - Get better results than what was achieved in F.....g
- ?? did we mention BEER

# BEERINFORMATICS

Bierinformatik / **Praktikum012019** Private

Unwatch 3 Star 0 Fork 0


<> Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings




<http://www.bioinf.uni-leipzig.de/teaching/currentClasses/class254.html> Edit

Manage topics

1 commit 33 branches 0 releases 1 contributor GPL-3.0

Branch: master New pull request Create new file Upload files Find file Clone or download

 **jfallmann** Initial commit Latest commit 7c9739e on Dec 13, 2018

 <a href="#">.gitignore</a>	Initial commit	2 months ago
 <a href="#">LICENSE</a>	Initial commit	2 months ago
 <a href="#">README.md</a>	Initial commit	2 months ago

## RESEARCH TOPICS

- Why is segemehl the best mapper for nanopore reads?
- How reliable is nanopore's WIMP software?
- Why is segemehl the best mapper for nanopore reads?
- How do replicates compare, what kind of yeast do we have?
- How do results generated by 30 students in 10 days compare to results generated by experts us (on a train ride)?

## **INTRODUCTION**

### **SETTING**

WIMP analysis

Mapper Comparison

Replicates and yeast composition

### **Conclusion and Outlook**

## EXPERIMENTAL SETUP

- Cooperate with real Biochemists (The Moerl Lab)
- BEER was chosen for expected diversity and local interests
  - ▶ Gose from Bayrischer Platz (Leipzig "Beer" variant with lactic acid - explicably never really caught on outside Leipzig)
  - ▶ Weissbier (Schneider Weisse Tap 7) - to keep the bavarian experementalist in chief interested
  - ▶ Waldbier - because wild cherry was no longer available it was wild pear
- Extract DNA from BEER samples
- Use Oxford Nanopore sequencing on extracted DNA
- Use students to analyse data

## SAMPLING

- Gose and Schneider Weisse were acquired from Getränkefeinkost Leipzig
- Waldbier was ordered via the internet from the producer at <https://kiesbye.at>



## DNA EXTRACTION

- DNA extraction was performed at Mario Mörl's Lab at the Institute for Biochemistry



## SAMPLE PREPARATION

- Sample preparation was performed at Mario Mörl's Lab at the Institute for Biochemistry .....





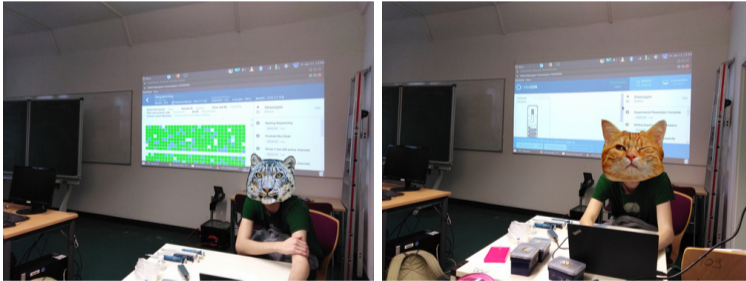
## SAMPLE PREPARATION

- ... or at the Institute for Bioinformatics

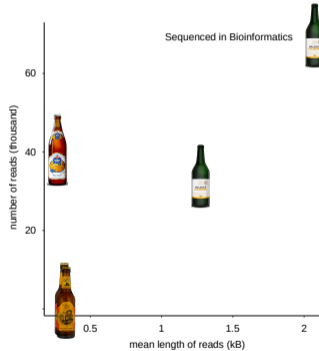


## NANOPORE SEQUENCING

- Institute for Biochemistry or Institute for Bioinformatics



# NANOPORE SEQUENCING RESULTS



## **INTRODUCTION**

### **SETTING**

WIMP analysis

Mapper Comparison

Replicates and yeast composition

### **Conclusion and Outlook**

## WIMP

- On-the-fly analysis of nanopore sequencing
- Uses centrifuge metagenomics program suite
- Result is annotated taxonomic tree of reads

Question How reliable is it?

## FIRST ANALYSIS



Figure: WIMP taxonomic tree for Gose run 1, 3%

## FIRST ANALYSIS

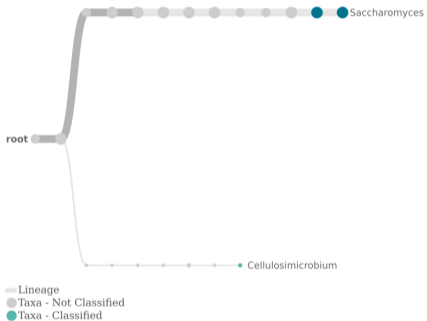


Figure: WIMP taxonomic tree for Gose run 1, 1%

## FIRST ANALYSIS

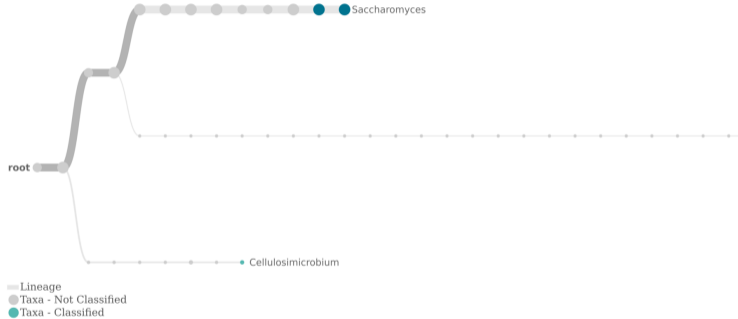


Figure: WIMP taxonomic tree for Gose run 1, 0.5%



## FIRST ANALYSIS

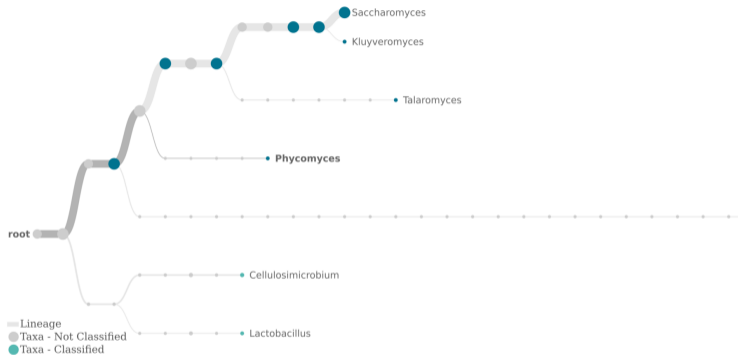


Figure: WIMP taxonomic tree for Gose run 1, 0.1%

# FIRST ANALYSIS

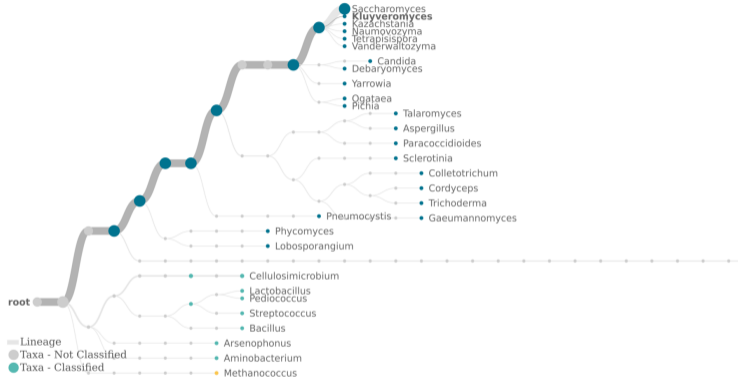


Figure: WIMP taxonomic tree for Gose run 1, all reads

## FIRST ANALYSIS

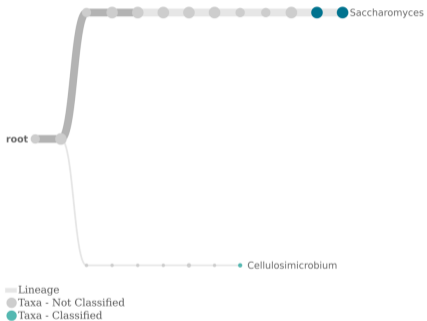


Figure: WIMP taxonomic tree for Gose run 2, 1%

## FIRST ANALYSIS

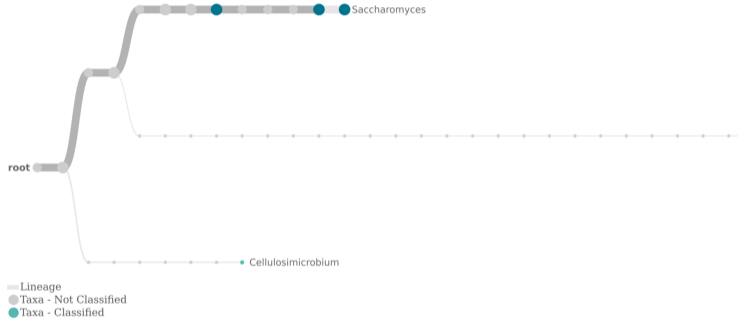


Figure: WIMP taxonomic tree for Gose run 2, 0.5%

# FIRST ANALYSIS

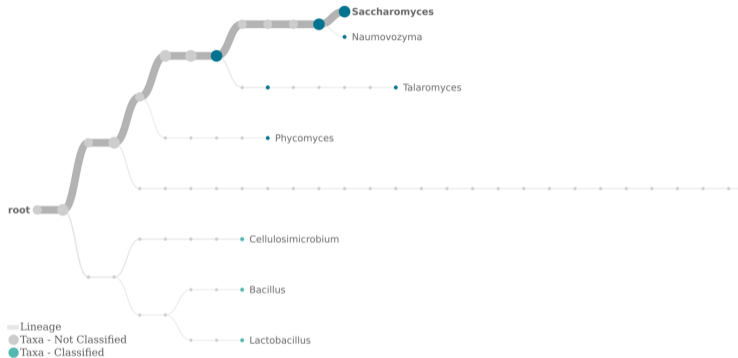


Figure: WIMP taxonomic tree for Gose run 2, 0.1%

# FIRST ANALYSIS



Figure: WIMP taxonomic tree for Gose run 2, all reads

# FIRST ANALYSIS

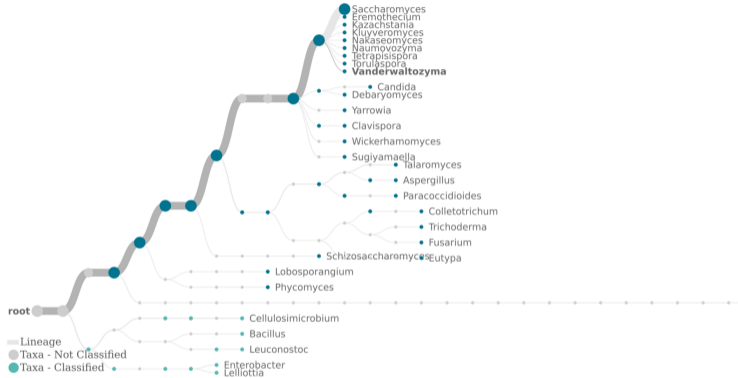


Figure: WIMP taxonomic tree for Waldbier 1, all reads

# FIRST ANALYSIS

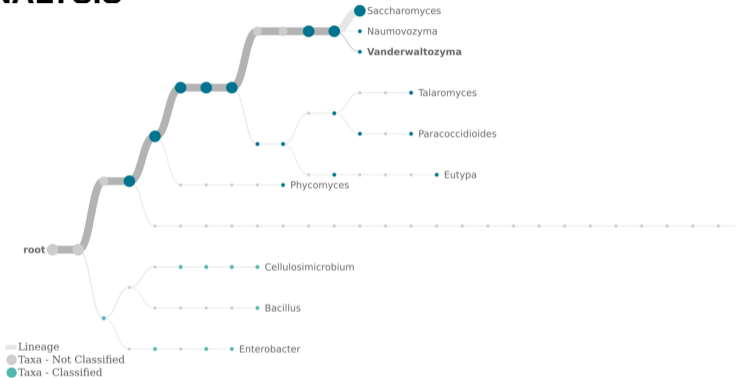


Figure: WIMP taxonomic tree for Waldbier 2, all reads



# FIRST ANALYSIS

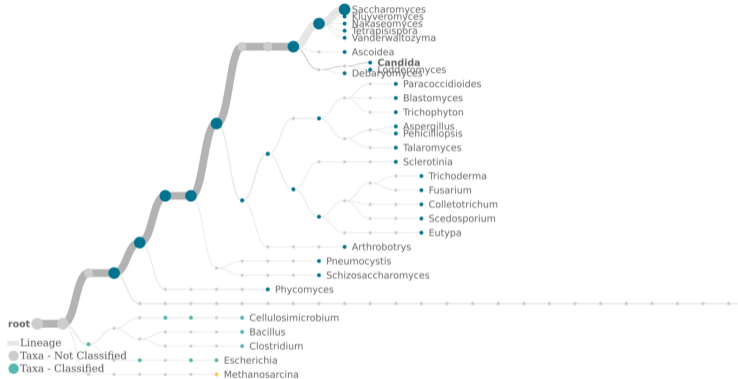


Figure: WIMP taxonomic tree for Tap 7 2, all reads

## FIRST ANALYSIS

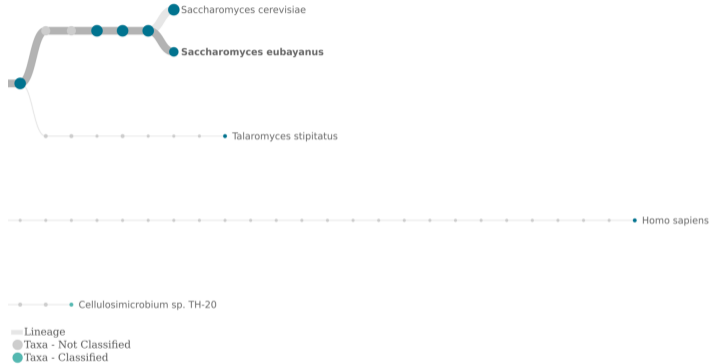


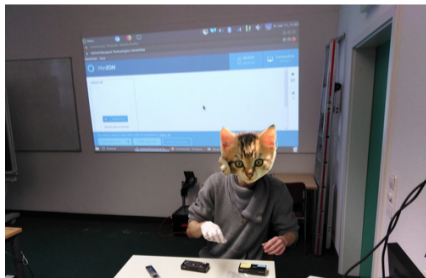
Figure: Representative example of long branch.

## FIRST ANALYSIS

- Reads from homo sapiens are suspicious
- Might be contamination
- Contamination at our side would be easily distinguishable

## FIRST ANALYSIS

- Reads from homo sapiens are suspicious
- Might be contamination
- Contamination at our side would be easily distinguishable



# STUDENTS' APPROACH

UNIVERSITÄT LEIPZIG

Introduction

Workflow

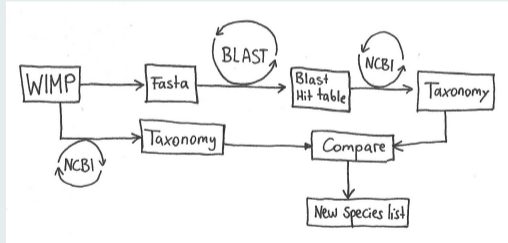
**Results**

Conclusion

Outlook

## Workflow

WIMP - BLAST comparison



25.01.2019 13

# STUDENTS' RESULTS

UNIVERSITÄT LEIPZIG

Introduction

Workflow

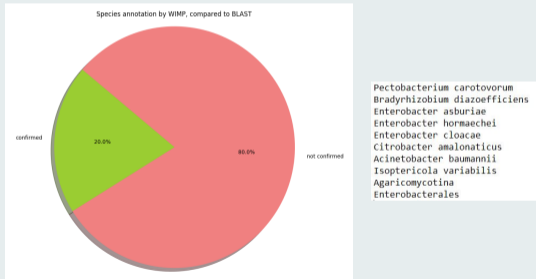
**Results**

Conclusion

Outlook

## WIMP-BLAST comparison

Species level



25.01.2019 14

# STUDENTS' RESULTS

THE „BLACK BOX“ TOOL WIMP

## MAPPED SPECIES

organism	Gose1	Gose2	Waldbier1	Waldbier2	Weißbier
<i>Talaromyces</i>	4 (3)	4 (4)	96 (704)	40 (135)	86 (19)
<i>Saccharomyces cerevisiae</i>	2382 (1812)	1492 (1200)	67117 (58778)	21089 (17798)	27376 (29224)
<i>Saccharomyces eubayanus</i>	351 (55)	244 (39)	43520 (38241)	14179 (11031)	9360 (165)
<i>Bacillus subtilis</i>	1 (-)	-	1 (-)	6 (-)	31 (10)
<i>Cellulosimicrobium</i>	52 (48)	19 (17)	1065 (921)	259 (213)	196 (165)
<i>Geobacillus</i>	-	-	1 (-)	1 (-)	2 (-)
<i>E. coli</i>	-	1 (-)	57 (3)	13 (1)	42 (20)

UNIVERSITÄT  
LEIPZIG

43

# TODO

## ES BLASTET DER FALLMANN AM RAUSCHIGEN BACH


**BLAST**® » **blastn suite** » RID-62S7XNJU016

[Edit and Resubmit](#) [Save Search Strategies](#) [▶ Formatting options](#) [▶ Download](#)

 **There was a problem with the search. Please, contact [Help Desk](#) and include RID 62S7XNJU016**

 **Informational Message: [blastsrv4.REAL]: Error: Process size limit exceeded, resulting in SIGXFSZ (25).**

**Job title: 7942 sequences (fbd4368b-aa19-41f7-b24e-c805fd09c09f)**

Results for:  

RID [62S7XNJU016](#) (Expires on 02-12 18:48 pm)

Query ID |cl|Query\_16804

Description fbd4368b-aa19-41f7-b24e-c805fd09c09f

Molecule type nucleic acid

Query Length 518



## MAPPER COMPARISON

- Compare results of state of the art mappers
- Work out why segemehl outperforms the others
- (Optimize parameters)

## STUDENTS' APPROACH

### OS and hardware

Segemehl runs nativ on linux-OS and after a long ancient yakutian shamanic dance on Windows UbuntuShell (tested on Windows 10)

On Windows 10 with 8 threads segemehl used 95% of RAM (15,1 Gb from 15,9 GB) and 58% of CPU (Intel Core i7) and consumed approx. 62 GB of the paging file



## STUDENTS' APPROACH

### Warning

Beware of cats, they can press ctrl+Z while running over the keyboard and destroy hours of work at 99% done!

24

## STUDENTS' APPROACH



## STUDENTS' RESULTS

SEARCH FOR MAPPERS | for Nanopore reads

## MAPPING STATISTICS

Elapsed time for  
indexing and mapping

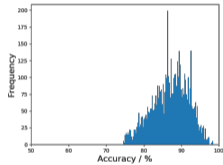
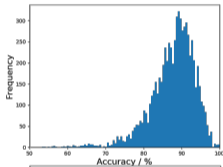
Sample	Mapping Tools	number of Reads	Time for Mapping	Time for Indexing
Gose1	segemehl	2565	60 s	6 min
	minimap2	2565	14 s	-- inclusive
	GraphMap	2565	66 s	54 s
	BLAT	2565	35 s	-- inclusive
Gose2	segemehl	1667	60 s	6 min
	minimap2	1667	15 s	-- inclusive
	GraphMap	1667	40 s	54 s
	BLAT	1667	31 s	-- inclusive
Mix	segemehl	34301	2 h 58 min	6 min
	minimap2	34301	157 s	-- inclusive
	GraphMap	34301	42 min	54 s
	BLAT	34301	8 h 54 min	-- inclusive
Waldbier	segemehl	106885	1 h 35 min	6 min
	minimap2	106885	128 s	-- inclusive
	GraphMap	106885	39 min	54 s
	BLAT	106885	13 h 20 min	-- inclusive
Weissbier	segemehl	40498	120 s	6 min
	minimap2	40498	19 s	-- inclusive
	GraphMap	40498	19 min	54 s
	BLAT	40498	8 min 48 s	-- inclusive

# STUDENTS' RESULTS

SEARCH FOR MAPPERS | for Nanopore reads  
**MAPPING ACCURACY**

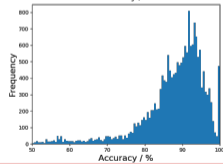
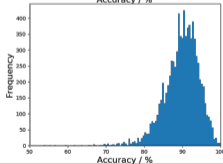
Gose1

GraphMap



Segemehl

Minimap2



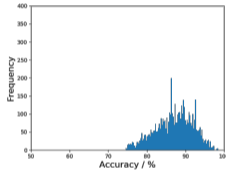
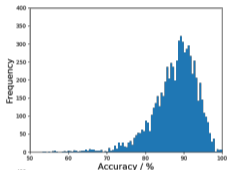
BLAT

# STUDENTS' RESULTS

SEARCH FOR MAPPERS | for Nanopore reads  
**MAPPING ACCURACY**

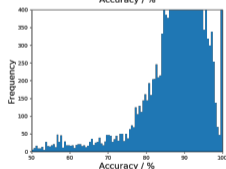
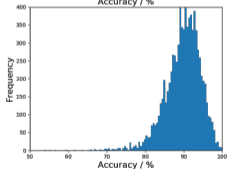
Gose1

GraphMap



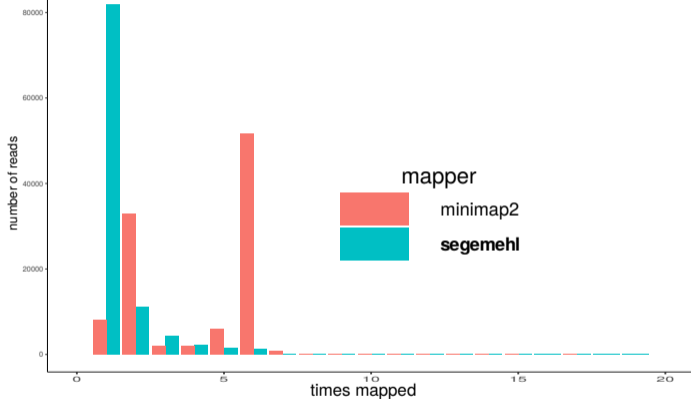
Segemehl

Minimap2



BLAT

## OUR RESULTS





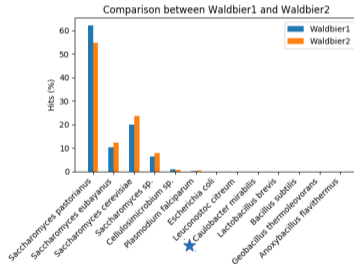
## Replicates

- We have two “technical” replicates (Waldbier, Gose)
- We treat BEER types as biological replicates
- How do they differ?
- Focus on yeast composition

# STUDENTS' RESULTS



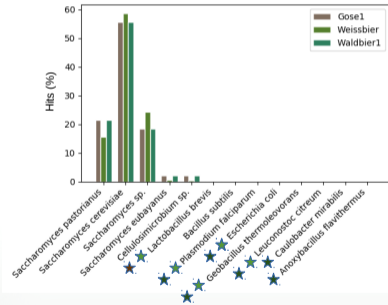
Part 1: Technical Comparison  
identified species in Waldbier1/Waldbier2 and relative amount of  
hits



## STUDENTS' RESULTS

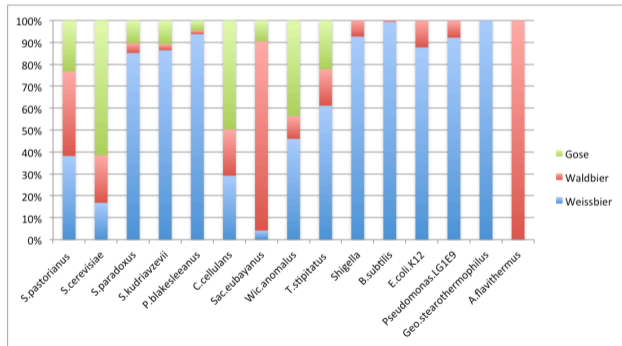


Part 2: Biological Comparison  
comparison of species/strains across the samples

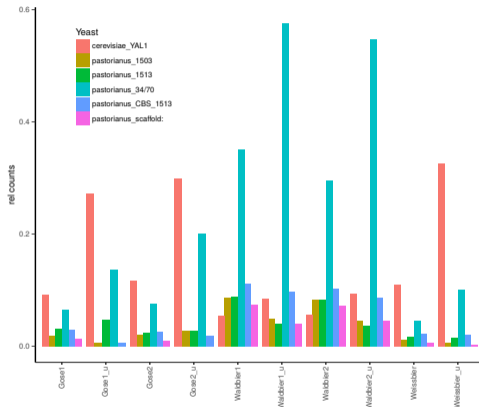


# STUDENTS' RESULTS

## Taxonomic composition



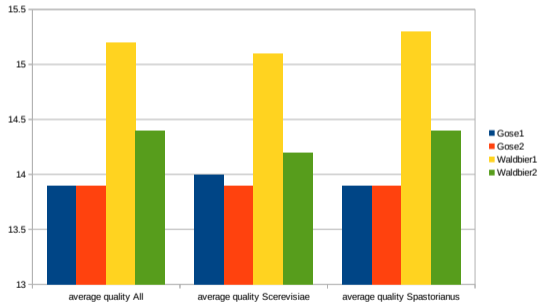
## OUR RESULTS



## WHAT DO THEY WANT TO TELL US?



### Statistics



28

## **INTRODUCTION**

### **SETTING**

WIMP analysis

Mapper Comparison

Replicates and yeast composition

### **Conclusion and Outlook**

## CONCLUSION

- Organoleptic testing of BEER can be dangerous to your health
- Students are more cost efficient than post-docs
- If you want to enjoy your beer, you probably should not analyze it

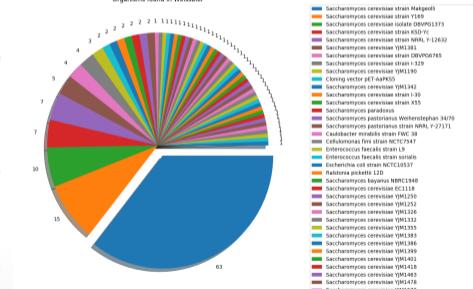


## WHY IS METAGENOMICS SO HARD



Part 3: Analysis of unmapped reads  
 metagenomic profile derived from unmapped reads found in  
 Weissbier sample (metagenome 2)

Organisms found in Weissbier



## WORSE WAYS TO WASTE BEER



Feuerwehr Aschbach Markt

## ACKNOWLEDGEMENTS

We want to thank:

- Vereinigung von Förderern und Freunden der Universität Leipzig e. V.
- The Mario Moerl Lab
- FE, BB, BG, PV, MM and **RB** in F.
- the students of the practical course
- teaching staff of the practical course (Zasha, Jan, Sandra, Alexander, Prof. Sonja Prohaska)
- His Wholiness Prof Peter F.. Stadler





UNIVERSITÄT  
LEIPZIG

**Thank You!**

**Jörg Fallmann, Stephan Bernhart**

Transcriptome Bioinformatics Group  
Institute for Bioinformatics

`berni/fall@bioinf.uni-leipzig.de`

`http://www.bioinf.uni-leipzig.de/`