

# Whole genome sequencing of brewer's yeast isolated in Canadian wineries

Hubert Kokoszka

# *Saccharomyces cerevisiae*

## *Saccharomyces cerevisiae* – species overview


- Used in wine and beer industry
- Genome length: 12 Mbp
- Cell size: 5-10  $\mu\text{m}$



## Bioproject PRJNA838724

- my aim: does the canadian yeasts clade exhibit genetic variants in coding sequences that affect proteins invaluable in the wine industry?
- yeast isolated in Okanagan Valley wine region in British Columbia
- Data was uploaded in 2023
- Illumina Hi-Seq: pair-ends reads (150 bp / read)
- 75 samples (I used 2 samples x 1 mln reads due to hardware limitations)

JOURNAL ARTICLE

**Whole genome sequencing of Canadian *Saccharomyces cerevisiae* strains isolated from spontaneous wine fermentations reveals a new Pacific West Coast Wine clade** 

R Alexander Marr, Jackson Moore, Sean Formby, Jonathan T Martiniuk, Jonah Hamilton, Sneha Ralli, Kishori Konwar, Nisha Rajasundaram, Aria Hahn, Vivien Measday ✉

[Author Notes](#)

# Reference genome R64

chr I-XVI  
+ mitochondrial

2014

*Saccharomyces  
cerevisiae* S288C



Data Download

Quality Control and Trimming

Alignment

Variant Calling

Biological Consequences

# Data Download

All bioproject-associated  
FASTQ files download

`-p, --project PRJNAXXXXX`

`sratoolkit:  
esearch, efetch, xtract, fastq-  
dump`

Given SRR-associated  
FASTQ files download

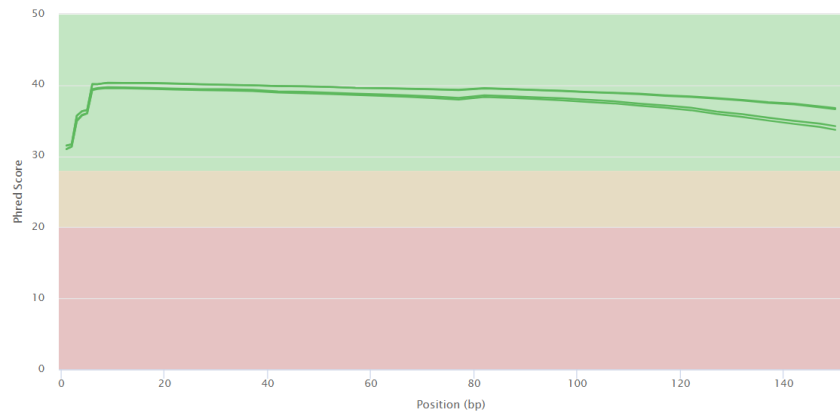
`-r, --sra SRRXXXXXX SRRXXXXXX ...`

`sratoolkit:  
fastq-dump`

# Quality Control

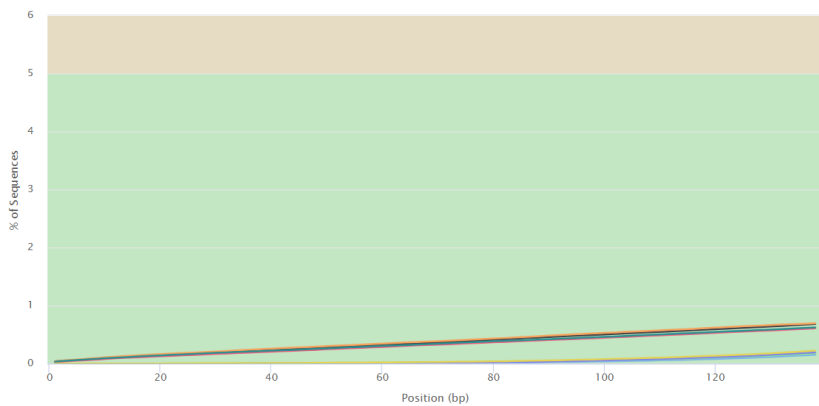
FastQC: Mean Quality Scores

[Export Plot](#)



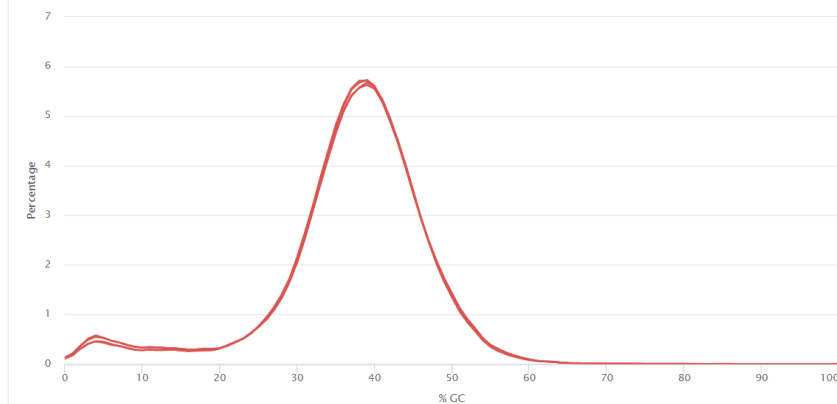
FastQC: Adapter Content

[Export Plot](#)



FastQC: Per Sequence GC Content

[Export Plot](#)

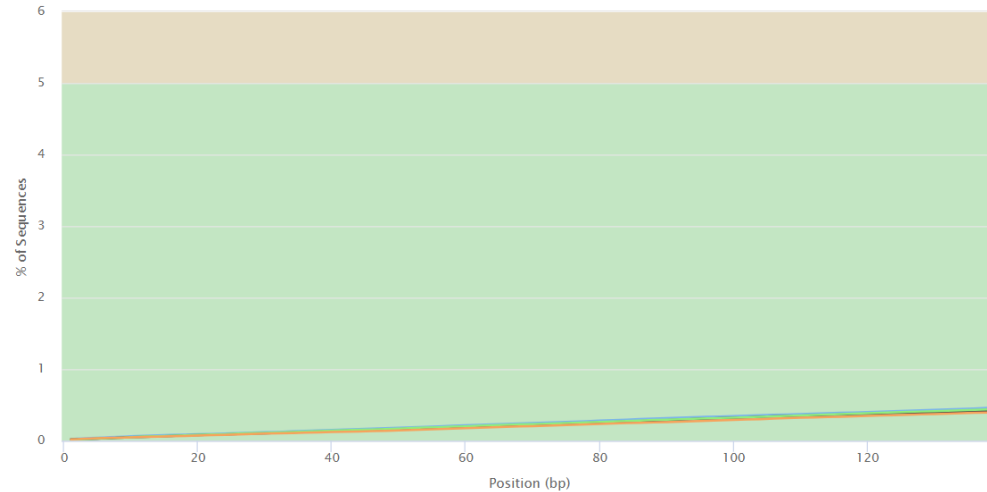


# Quality Control

polyA sequences in the genome?

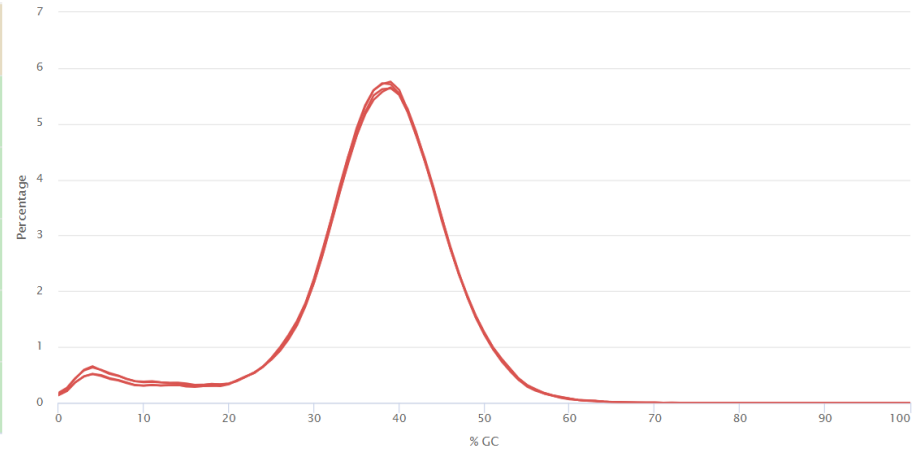
FastQC: Adapter Content

 Export Plot



FastQC: Per Sequence GC Content

 Export Plot

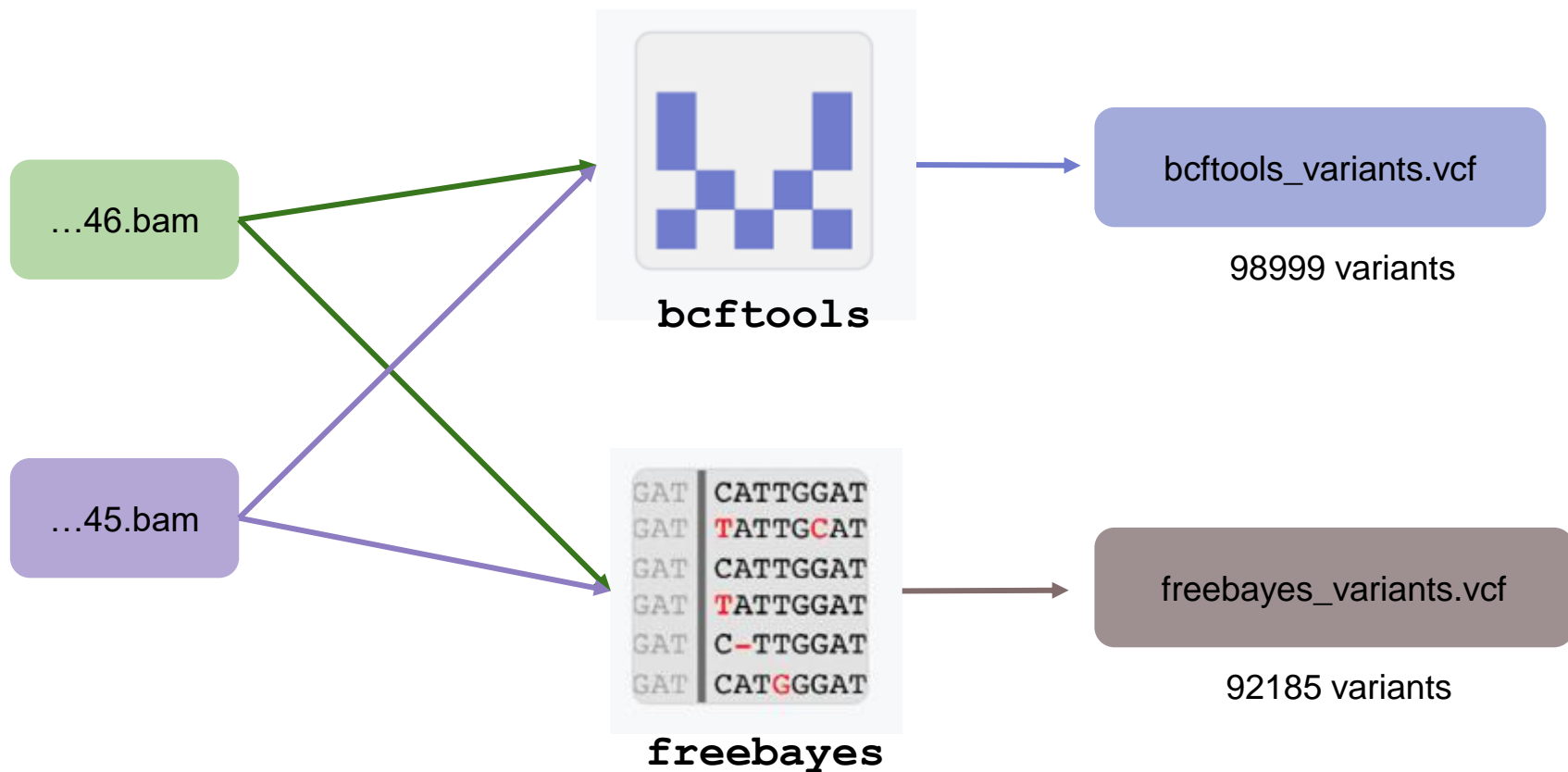




# Alignment

<b>bwa mem</b>	Sample 1	Sample 2
Mean coverage	13,88	14,47
Median coverage	11	12
Aligned reads	97,52%	98,76%
Paired	96,29%	97,57%

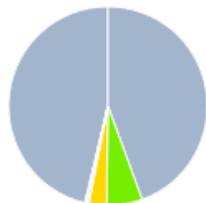
# Variant calling





bcftools

Consequences (all)



- upstream\_gene\_variant: 46%
- downstream\_gene\_variant: 44%
- synonymous\_variant: 6%
- missense\_variant: 3%
- non\_coding\_transcript\_exon\_variant: 0%
- intron\_variant: 0%
- frameshift\_variant: 0%
- inframe\_insertion: 0%
- inframe\_deletion: 0%
- Others: 0%

Coding consequences

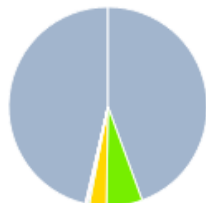


- synonymous\_variant: 65%
- missense\_variant: 32%
- frameshift\_variant: 1%
- inframe\_insertion: 0%
- inframe\_deletion: 0%
- stop\_gained: 0%
- protein\_altering\_variant: 0%
- stop\_retained\_variant: 0%
- coding\_sequence\_variant: 0%
- Others: 0%

```
GAT CATTGGAT
GAT TATTGCAT
GAT CATTGGAT
GAT TATTGGAT
GAT C-TTGGAT
GAT CATGGGAT
```

freebayes

Consequences (all)



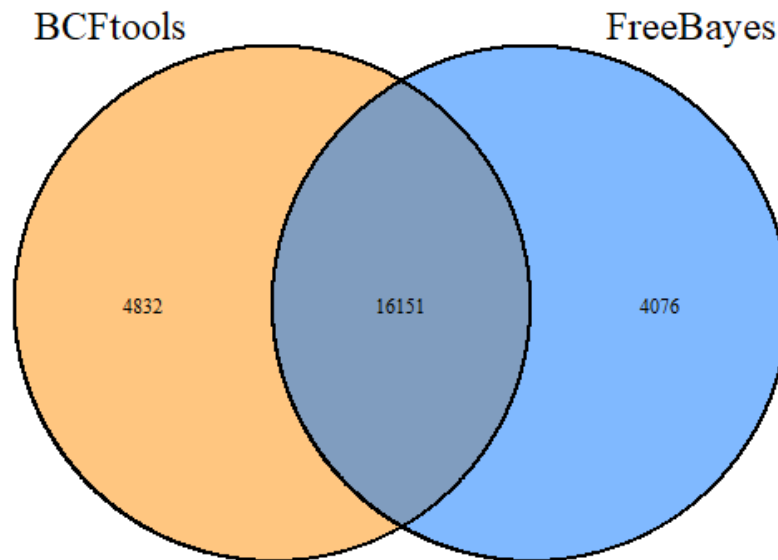
- upstream\_gene\_variant: 46%
- downstream\_gene\_variant: 44%
- synonymous\_variant: 6%
- missense\_variant: 3%
- non\_coding\_transcript\_exon\_variant: 0%
- intron\_variant: 0%
- frameshift\_variant: 0%
- inframe\_deletion: 0%
- inframe\_insertion: 0%
- Others: 0%

Coding consequences



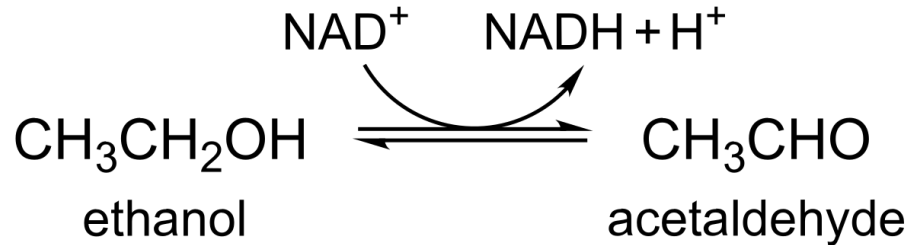
- synonymous\_variant: 66%
- missense\_variant: 31%
- frameshift\_variant: 1%
- inframe\_deletion: 0%
- inframe\_insertion: 0%
- stop\_gained: 0%
- stop\_retained\_variant: 0%
- start\_lost: 0%
- stop\_lost: 0%
- Others: 0%

Consistency of variants detected  
by the two programmes used in the project

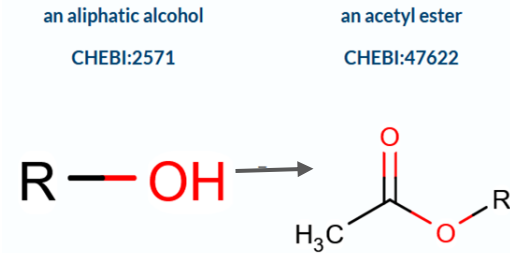


## Metabolic pathways crucial for wine industry

### Glycolysis + fermentation



### Esther metabolism



### YMR318C (ADH1) – Alcohol dehydrogenase

dehydrogenase

ADH1 gene is located on the  
chromosome XV left arm

- acetate aldehyde reduction to ethanol
- methylglyoxal reduction
- NADH oxidation
- Ethanol production out of broken amino acids

## YMR318C (ADH6)

cinnamyl dehydrogenase

- fusel alcohol synthesis
- aldehyde tolerance

## YOR377W (ATF1)

Alcohol acetyl-transferase

- acetate esters production
- terpinyl acetate synthesis

### YCR021C (HSP30) – heat shock protein 30 kDa

Cell membrane localization

H(+)-ATPase Pma1p regulator,  
which impacts pH

Stress response:

- heat shock
- high ethanol concentration,
- low organic acids concentration
- low glucose concentration



# Summary

The canadian yeast clade (oakland) apparently exhibits a number of mutations in genes crucial for the wine industry

ADH1  
(alcohol fermentation)

ATF1  
(esther metabolism)

ADH6  
(esther metabolism)

HSP30  
(stress response)