

Metaproteomics Analysis

Johan S. Sáenz & Bibiana Rios

2024-05-28

Table of contents

Preface	3
1 Introduction	4
2 Run quality	5
References	10

Preface

This is a Quarto book.

To learn more about Quarto books visit <https://quarto.org/docs/books>.

1 + 1

[1] 2

1 Introduction

This is a book created from markdown and executable code.

See Knuth (1984) for additional discussion of literate programming.

```
1 + 1
```

```
[1] 2
```

2 Run quality

In summary, this book has no content whatsoever.

1. Load libraries

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.1      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.1
v purrr      1.0.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
here() starts at /home/runner/work/metaproteomics-analysis/metaproteomics-analysis
```

2. Load data frame

```
quality <- read_tsv("rawdata/final_summary.tsv")
```

Rows: 21 Columns: 6

```
-- Column specification -----
Delimiter: "\t"
chr (2): Raw file, Experiment
dbl (4): MS/MS, MS/MS Identified, MS/MS Identified [%], Peptide Sequences Id...
```

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show_col_types = FALSE` to quiet this message.

3. Create a table

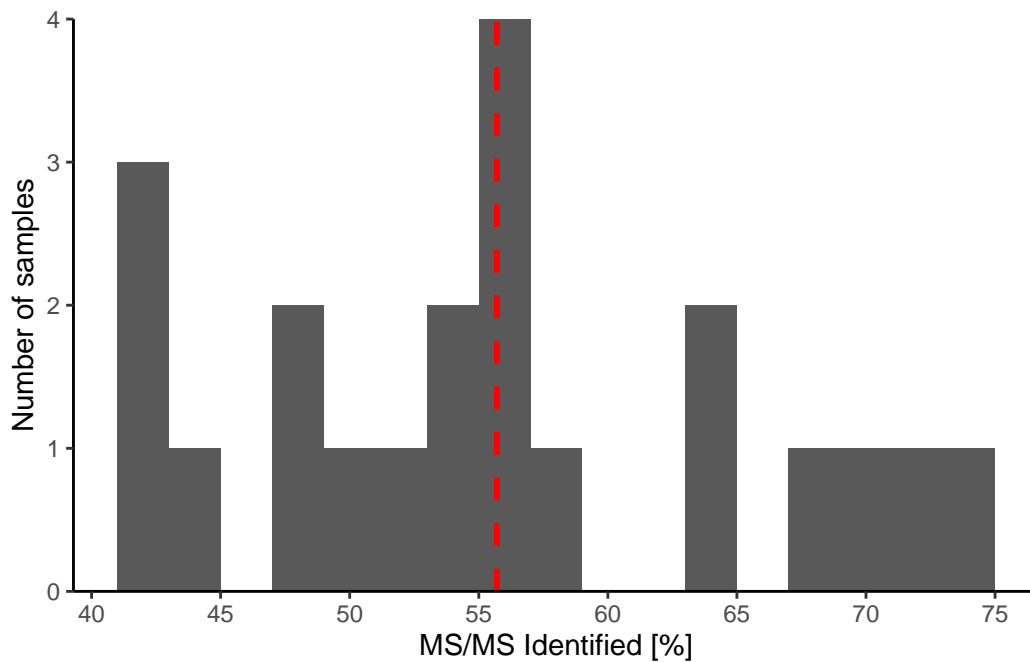
```
table %>%
  gt() %>%
  tab_stubhead(label = "landmass") %>%
  cols_label()
```

Variable	Max	Min	Mean	Sd
MS/MS	76291.0	43689.0	59796.2	10313.3
MS/MS Identified	46623.0	20545.0	32962.1	6427.9
MS/MS Identified [%]	74.1	41.4	55.8	10.4
Peptide Sequences Identified	30231.0	7042.0	16630.8	6401.0

4. Create a histogram of the identified MS/MS

```
percentage <- quality %>%
  ggplot(aes(x = `MS/MS Identified [%]`)) +
  geom_histogram(binwidth = 2) +
  geom_vline(xintercept= 55.7, linetype = "dashed",
             color = "red", linewidth = 1) +
  scale_x_continuous(breaks = seq(0, 75, by = 5)) +
  scale_y_continuous(breaks = seq(0, 5, by = 1),
                     expand = c(0, 0)) +
  labs(y = "Number of samples") +
  theme_classic()

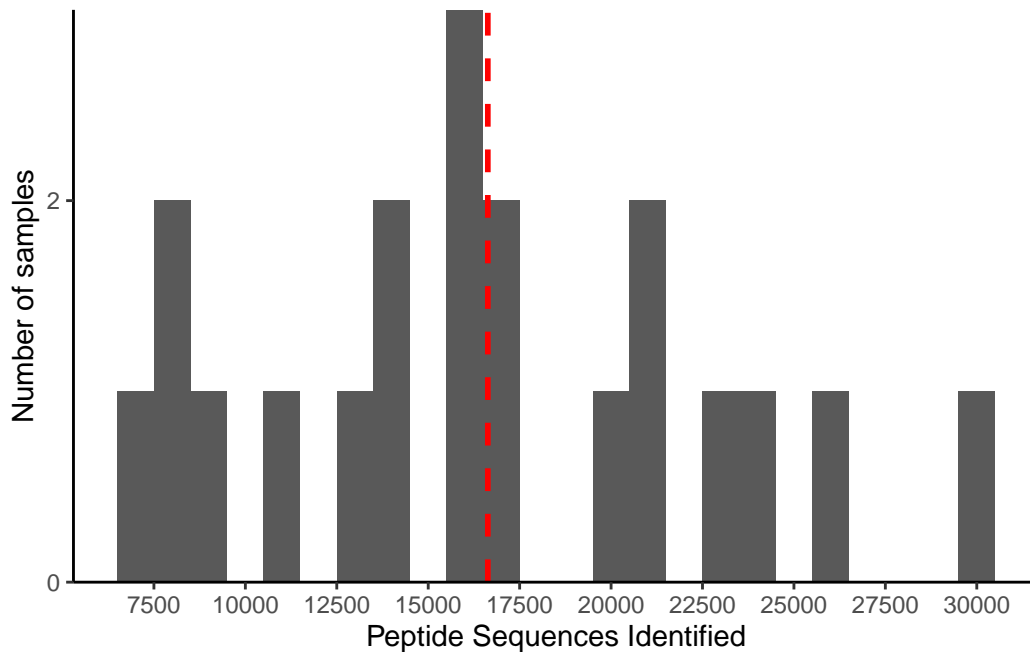
percentage
```



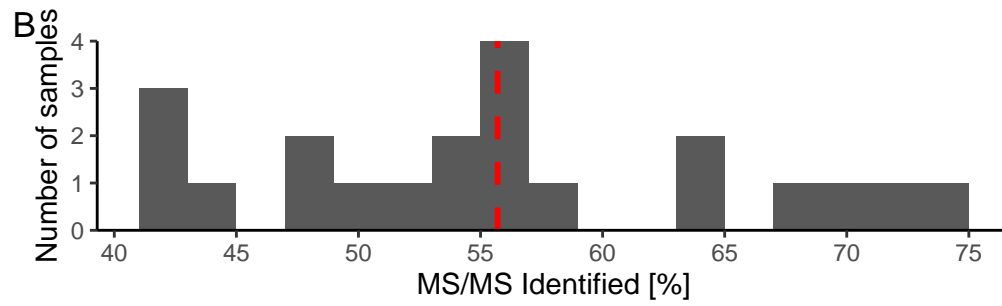
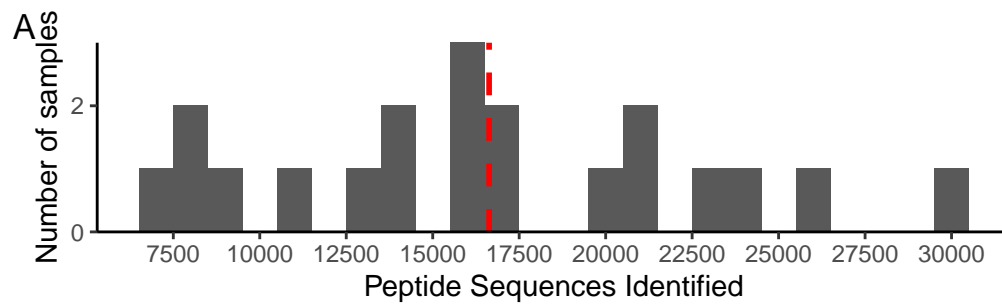
5. Create a histogram of the identified peptide sequences

```
peptides <- quality %>%
  filter(`Raw file` != "Total") %>%
  mutate(`Raw file` =str_remove(`Raw file`, "Seifert_Hammel_Nr")) %>%
  ggplot(aes(x = `Peptide Sequences Identified`)) +
  geom_histogram(binwidth = 1000) +
  geom_vline(xintercept = 16630, linetype = "dashed",
            color = "red", linewidth = 1) +
  scale_x_continuous(breaks = seq(0, 30231, by = 2500)) +
  scale_y_continuous(breaks = seq(0, 20, by = 2),
                    expand = c(0, 0)) +
  labs(y = "Number of samples") +
  theme_classic()

peptides
```



6. Create a compose plot



3

References

Knuth, Donald E. 1984. “Literate Programming.” *Comput. J.* 27 (2): 97–111. <https://doi.org/10.1093/comjnl/27.2.97>.