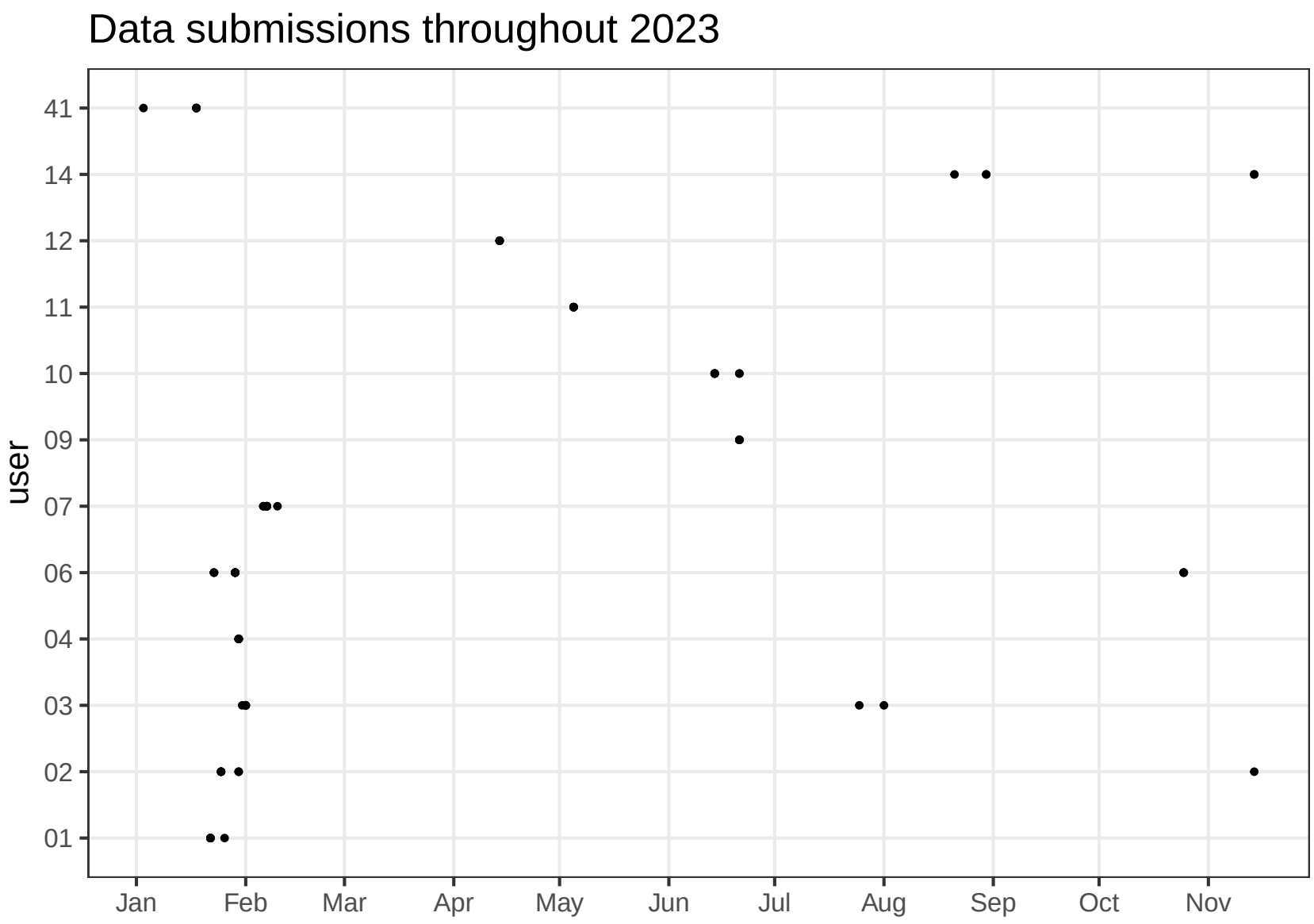
sPRG peptide quant data analysis

Submissions



Files imported

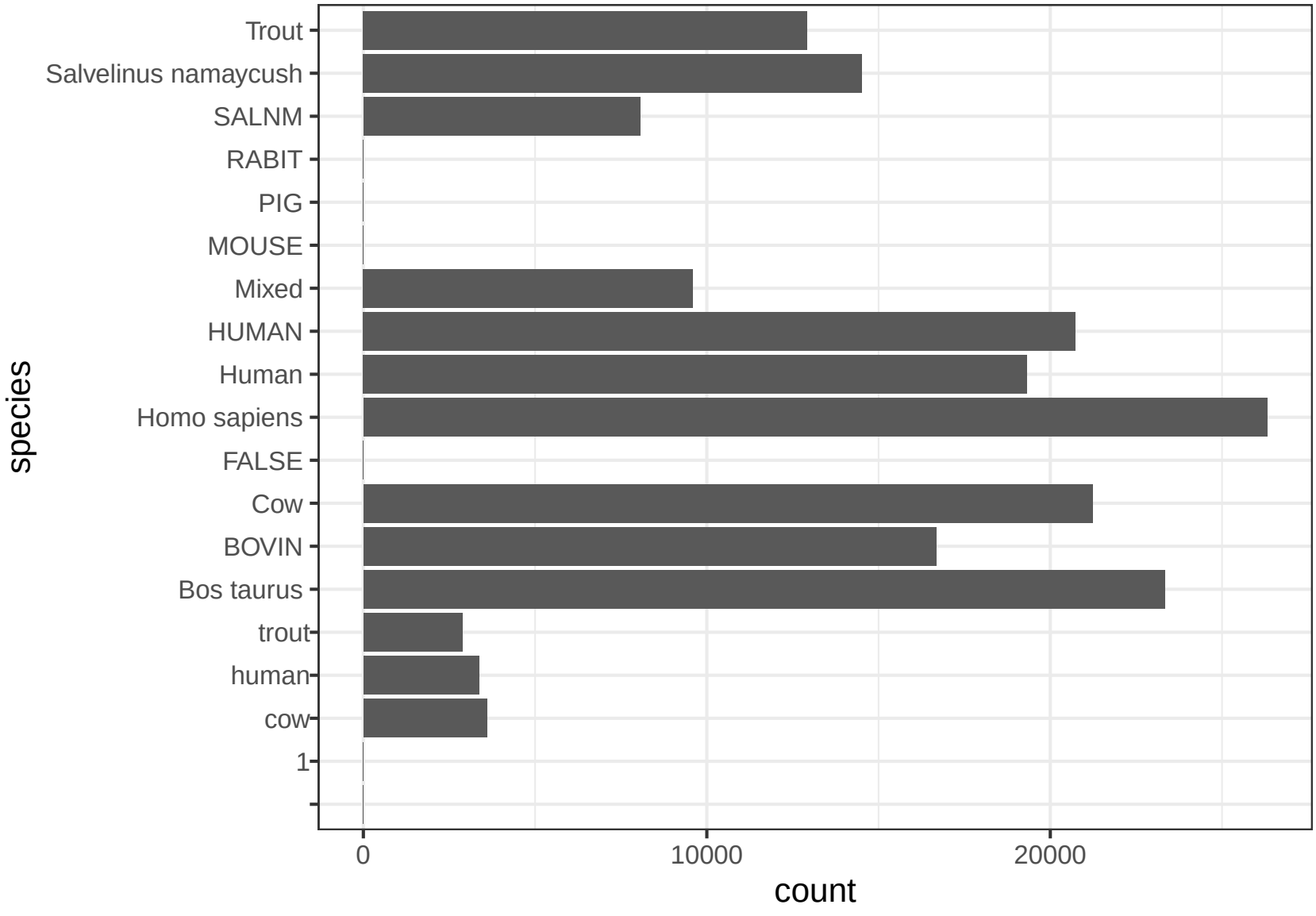
File
01-pepQuant
02-pepQuant
03-pepQuant
04-pepQuant
06-pepQuant
07-pepQuant
07-pepQuant_ShortGrad
09-pepQuant
10-Fusion_pepQuant
10-QEx_pepQuant
11-pepQuant
12-pepQuant
14_pepQuant_1x8mzStag_3x4mzGPFlibrary
14_pepQuant_2x4mzStag_Prosit_library
41-pepQuant

Data analysis strategy

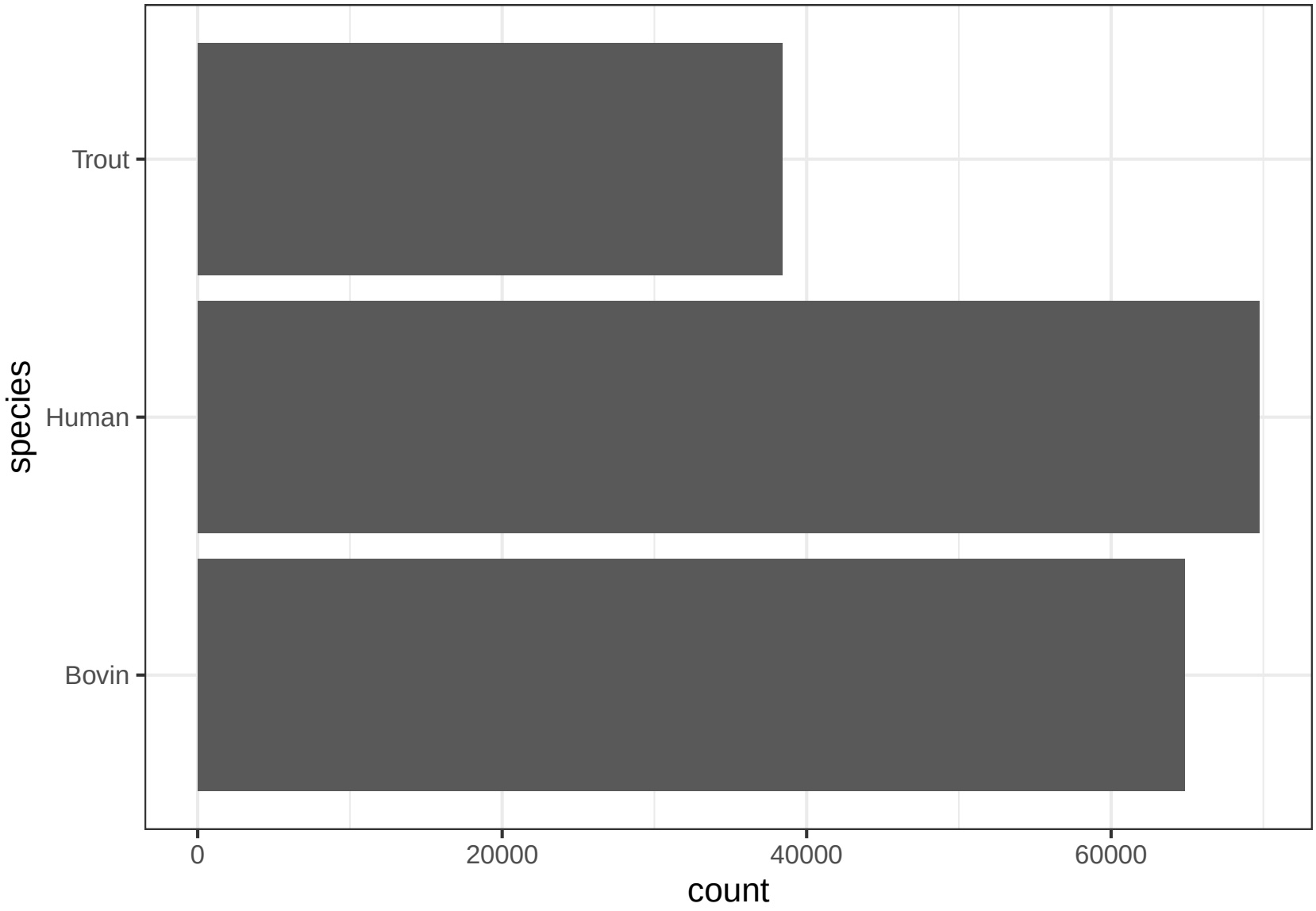
The strategy to analyze this data is to first harmonize the peptide, species, and peptide abundances. The following slides shows the various formats for peptide entries and species.

Next, I will incorporate the in-silico peptide digest used in the stability/homogeneity analysis. This in-silico list is derived by digesting the 3 fasta proteomes and selecting only the peptides unique to a single organism. The sample ratios will be calculated with and without unique peptides.

A mix of species names



After reformatting and filtering species



Peptide formats: before and after cleaning

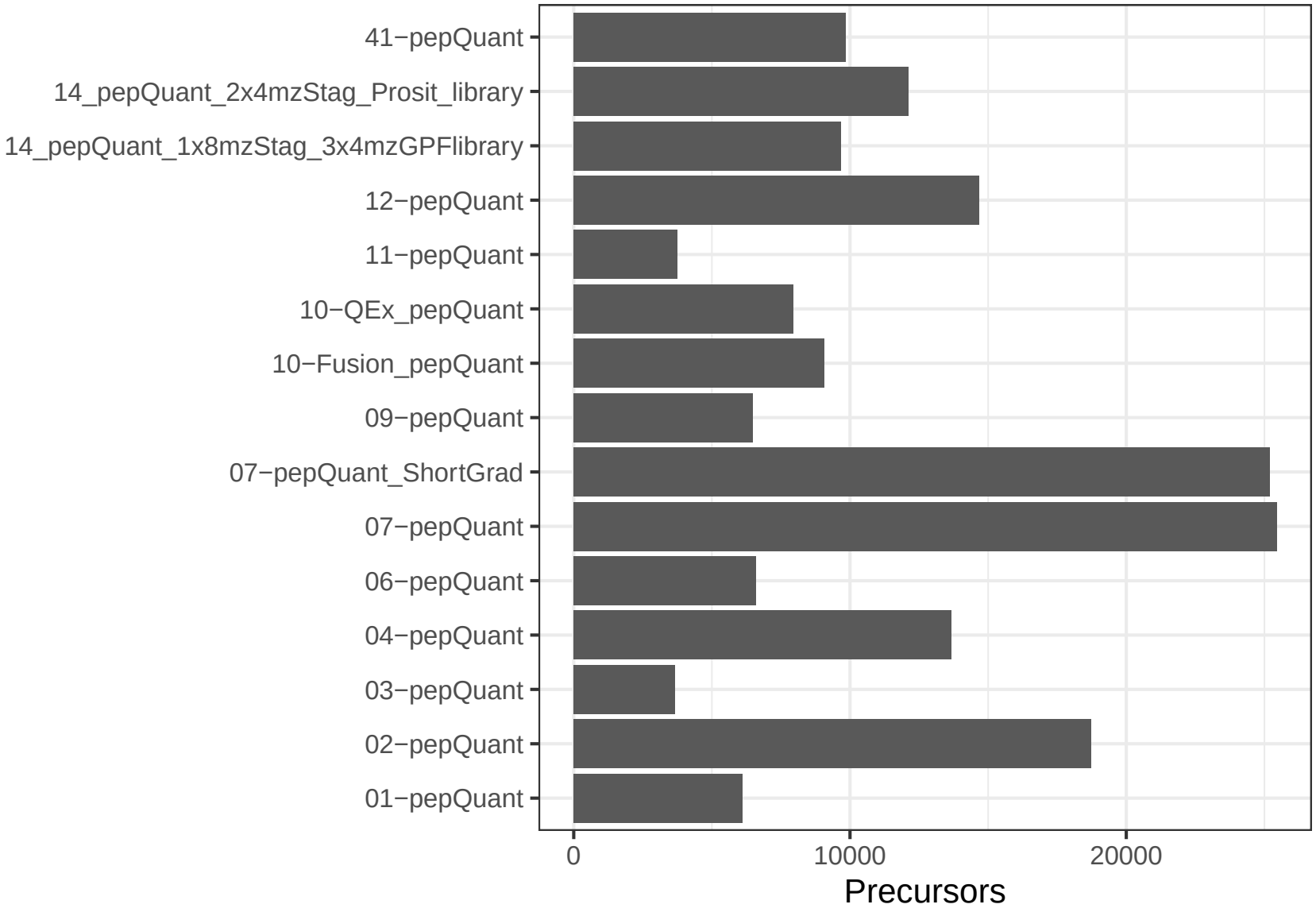
User	Peptides Before Formatting	Peptides after Formatting
01	_QAETHLM[Oxidation (M)]SDGTSR_.2	QAETHLMSDGTSR
02	DIDDLELTLAK	DIDDLELTLAK
03	HSDENLDLAR	HSDENLDLAR
04	VKLNDGHFIPVLGFGTYAPPEVAK	VKLNDGHFIPVLGFGTYAPPEVAK
06	AAQVDDVASASNFR	AAQVDDVASASNFR
07	_LC[Carbamidomethyl (C)]QPEGIHIC[Carbamidomethyl (C)]DGTEAENTATLALLEQQGLIR_.4	LCQPEGIHICDGTEAENTATLALLEQQGLIR
09	VSVWELLFYR	VSVWELLFYR
10	[R].GAEVHVVPWNHDFTK.[M]	GAEVHVVPWNHDFTK
11	ISPDLC[Common Fixed:Carbamidomethyl on C]GR	ISPDLCGR
12	EMTDLTCR	EMTDLTCR
14	AASGFNAAEDAQTLRK	AASGFNAAEDAQTLRK
41	EGFGHLSPTGNTEFWLGNEK	EGFGHLSPTGNTEFWLGNEK

Summarize precursor to peptide

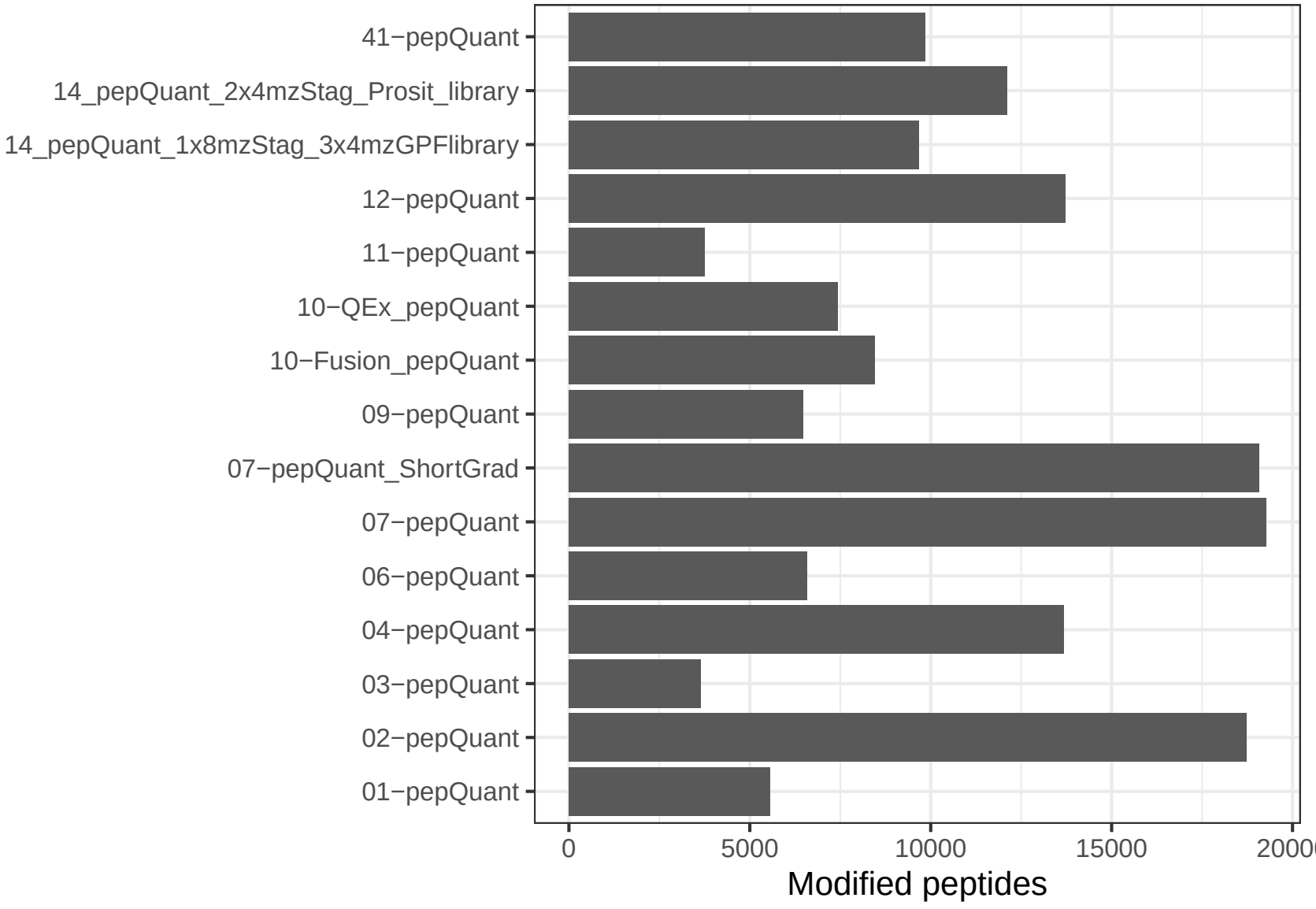
Some of the reported peptides were actually precursors. Should we aggregate precursors to peptides?

For this analysis, I did not perform and aggregation.

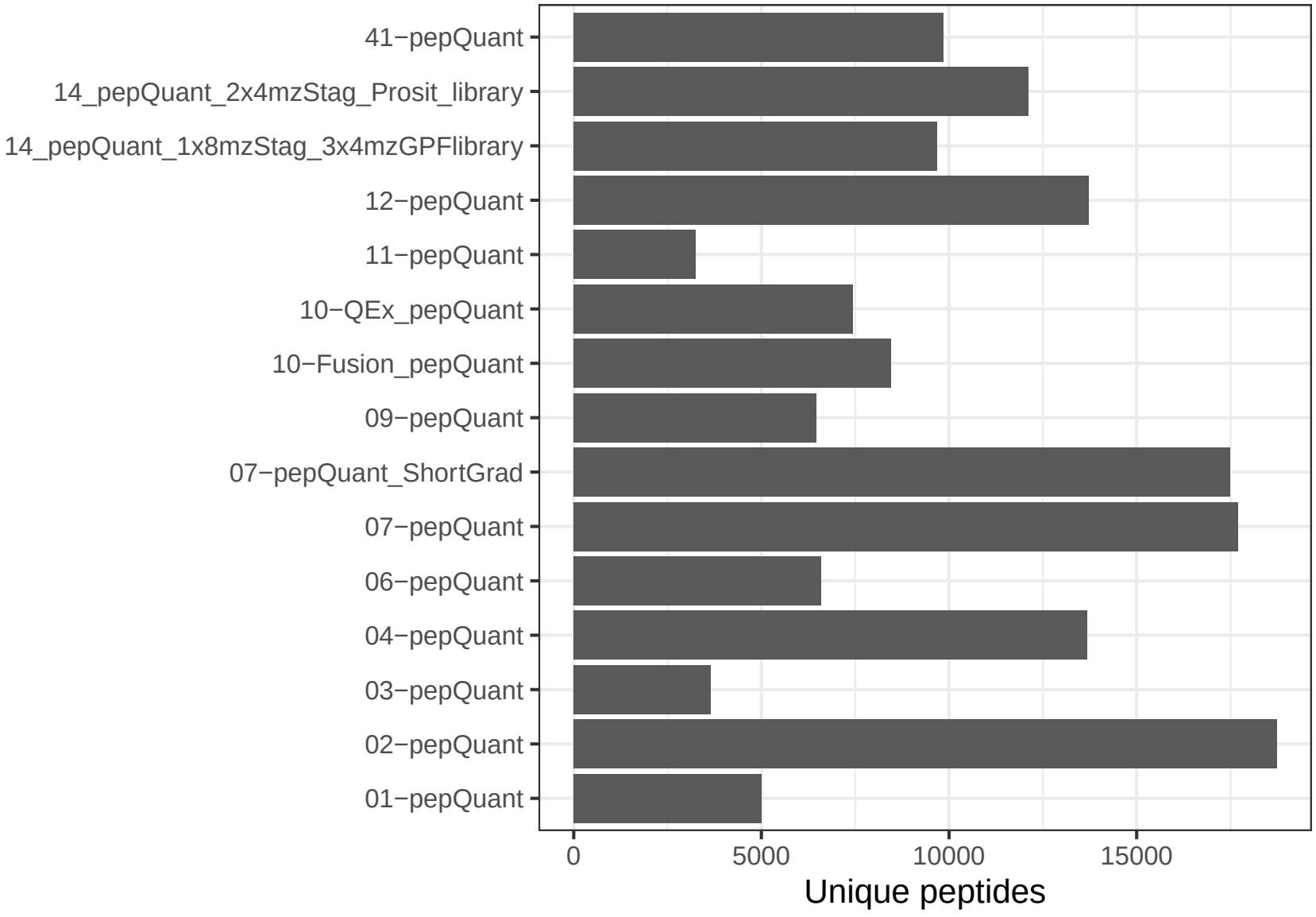
Precursors


41−pepQuant
14_pepQuant_2x4mzStag_Prosit_library
14_pepQuant_1x8mzStag_3x4mzGPFlibrary
12−pepQuant
11−pepQuant
10−QEx_pepQuant
10−Fusion_pepQuant
09−pepQuant
07−pepQuant_ShortGrad
07−pepQuant
06−pepQuant
04−pepQuant
03−pepQuant
02−pepQuant
01−pepQuant
0
10000
20000
Precursors

Modified Peptides



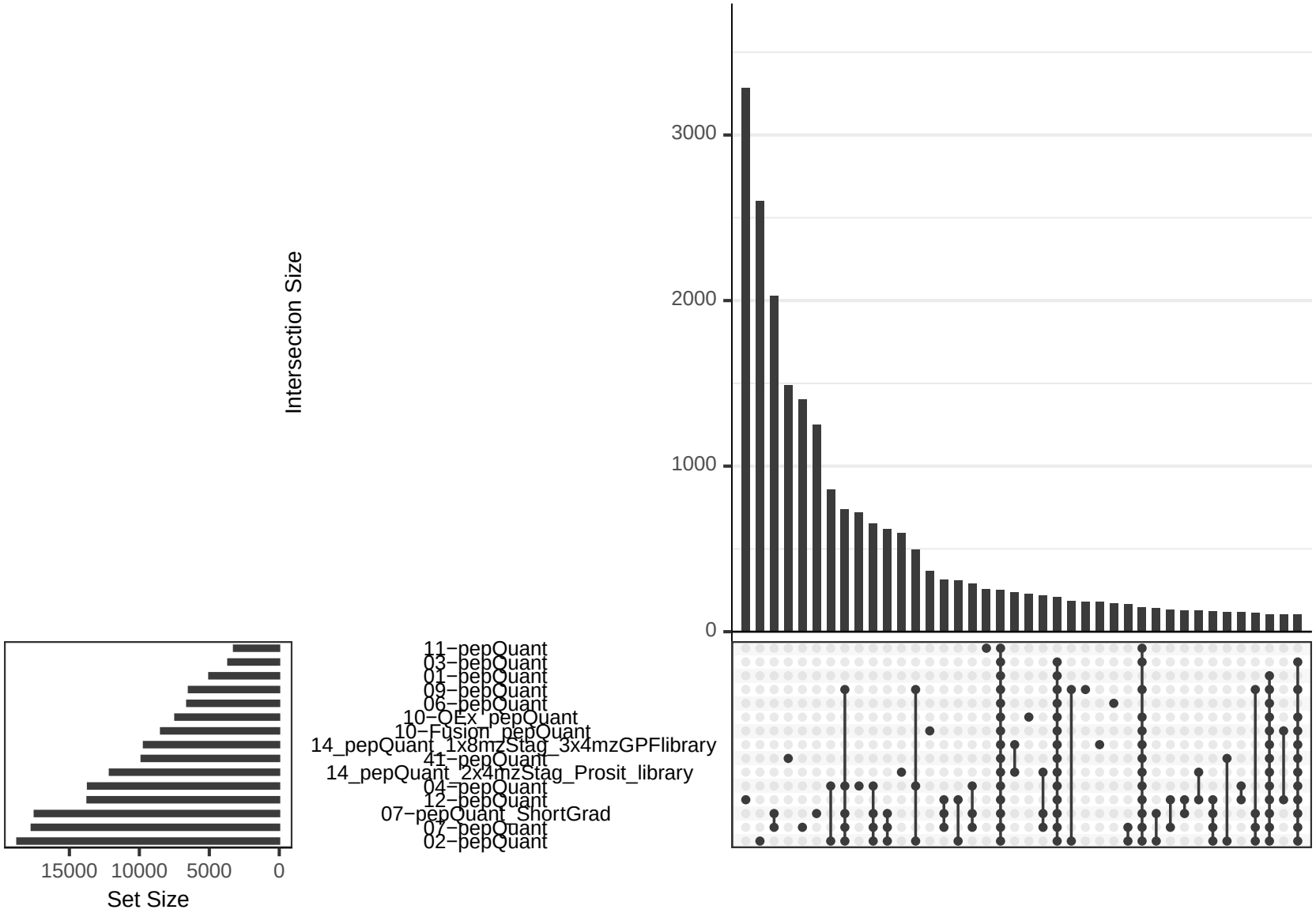
Unique Peptides



Peptide overlap across samples

Next, I used unique peptides to determine the overlap between user data.

Peptide overlap across samples

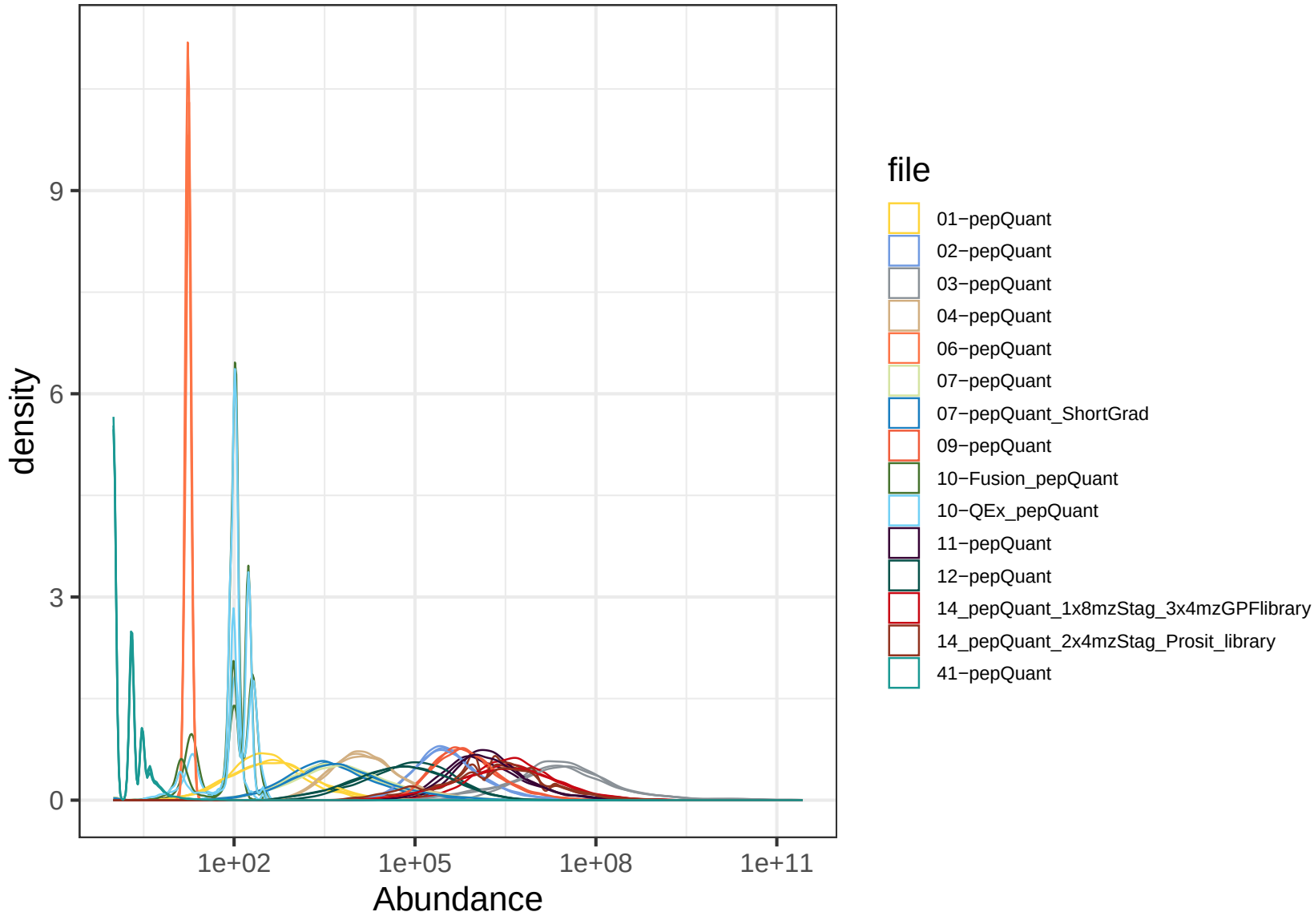


Peptide abundance distributions

To make an LFQBench-style figure, the peptide abundance values need to be non log-transformed. This is because when we generate the figure, we will do our own log transformations

- y-axis will be log(A / C)
- x-axis will be log(B)

Original abundance distributions



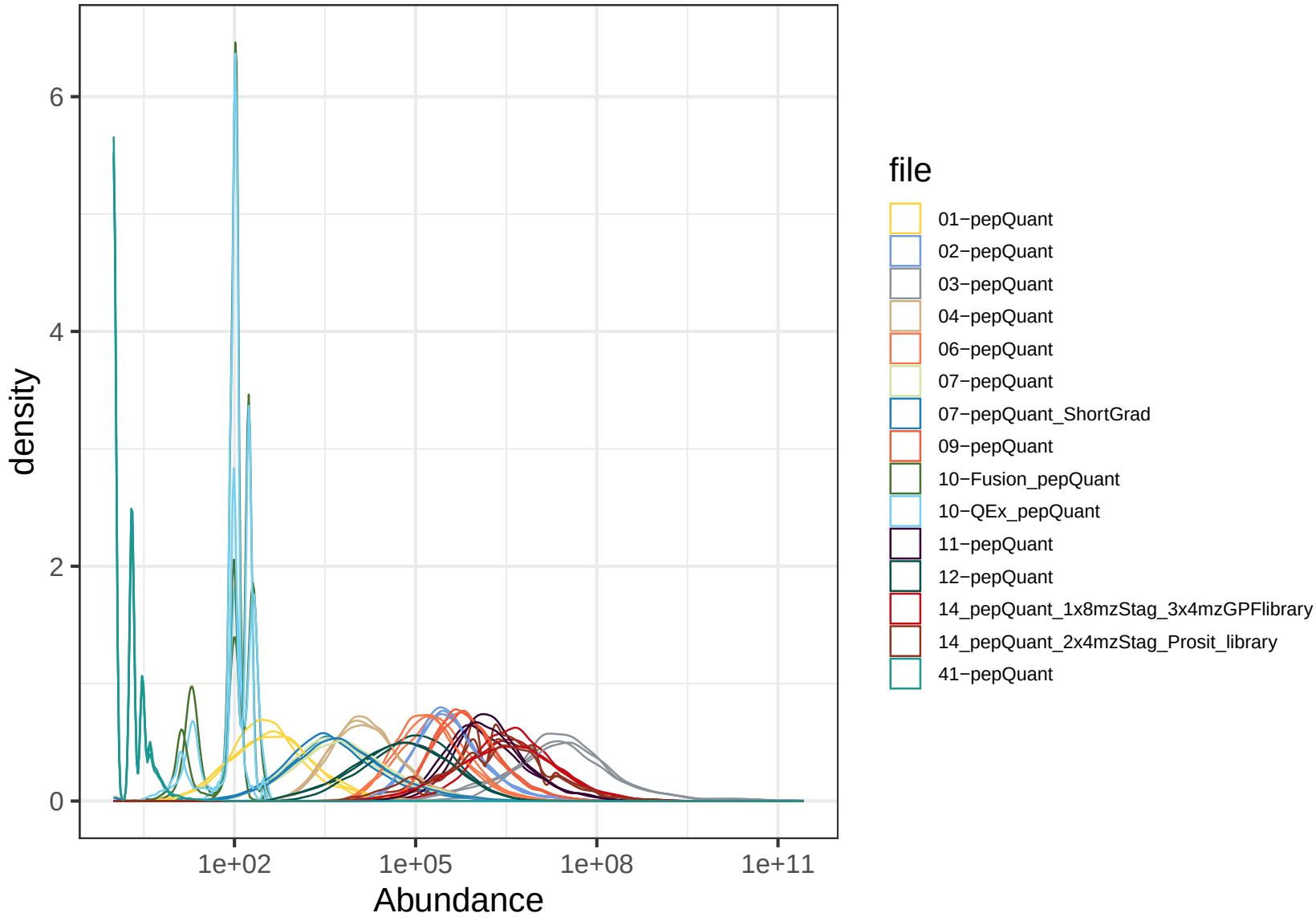
Transforming values

Most of the data entries show a distribution of non-transformed values. You can see this from the yellow curve and towards the right. However, there are three groups of distributions with some transformation.

- 06 - Log2-transformed
- 10 - Scaled quant
- 41 - Count data

Data from User 06 will be transformed back to a linear scale. 10 and 41 will be left as is.

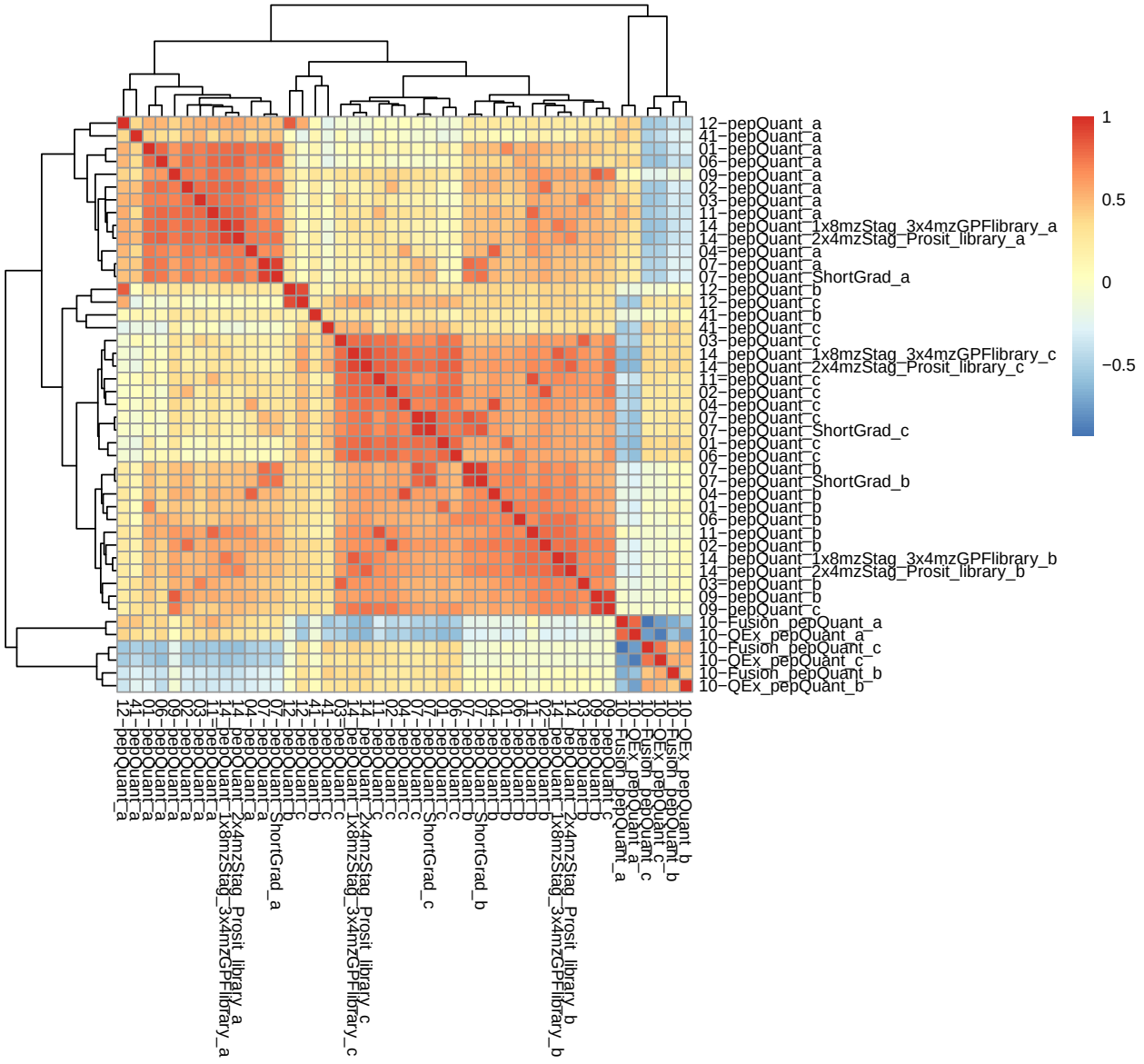
New abundance distributions



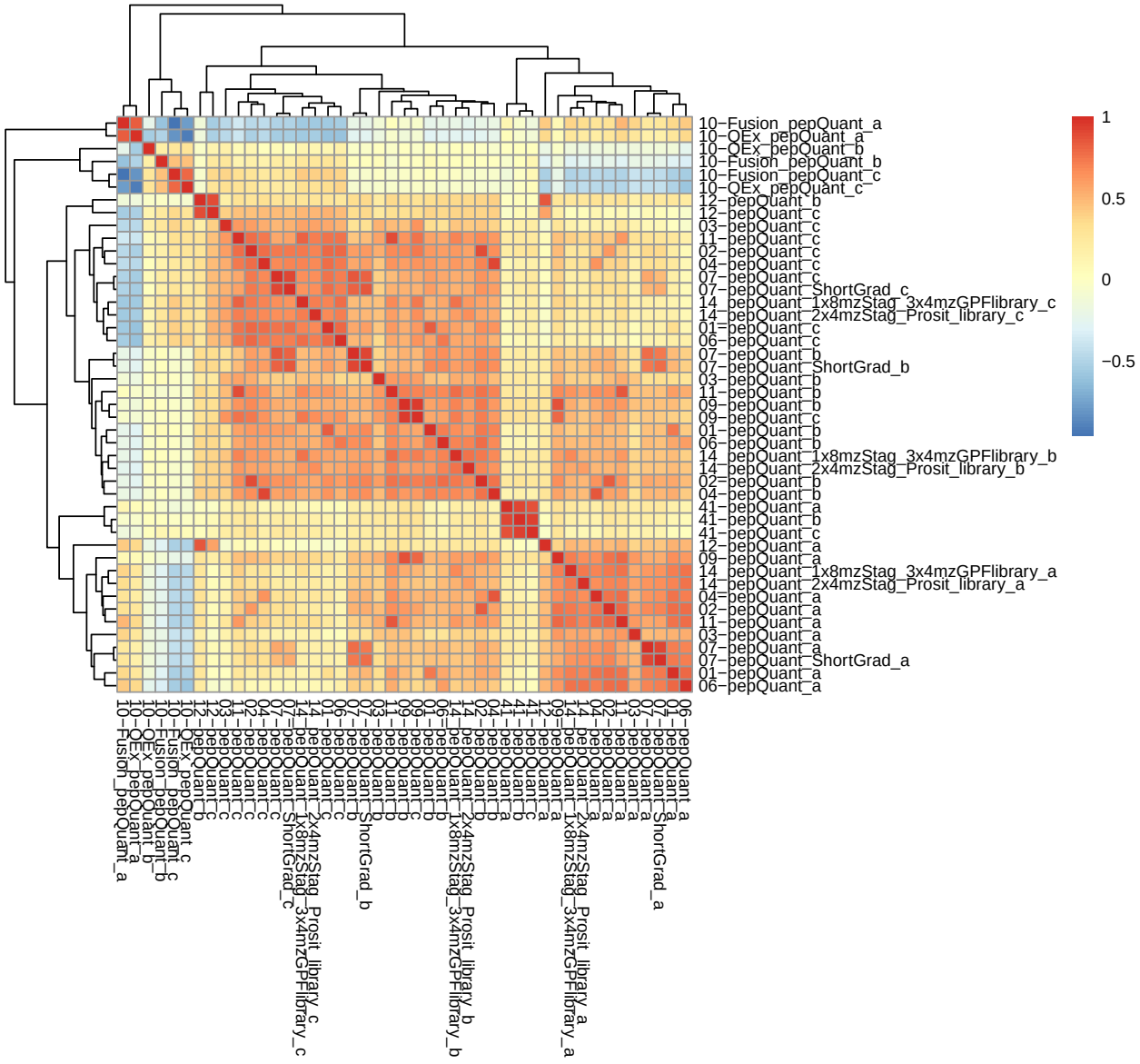
Correlation

Next, I performed a correlation analysis across samples. For this, the values were log2-transformed for all samples except 10 (scaled data) and 41 (count data).

Spearman Correlation



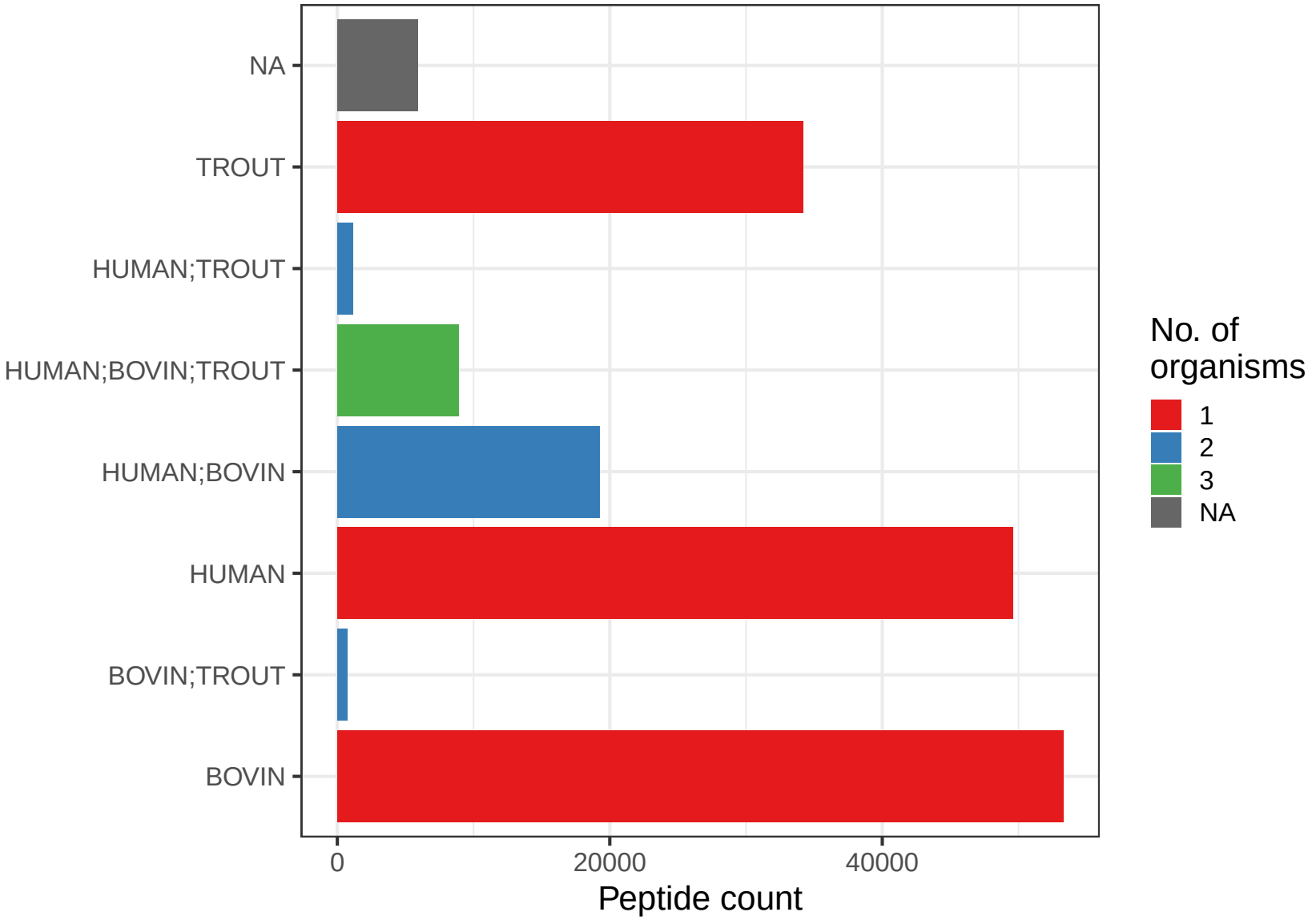
Pearson Correlation



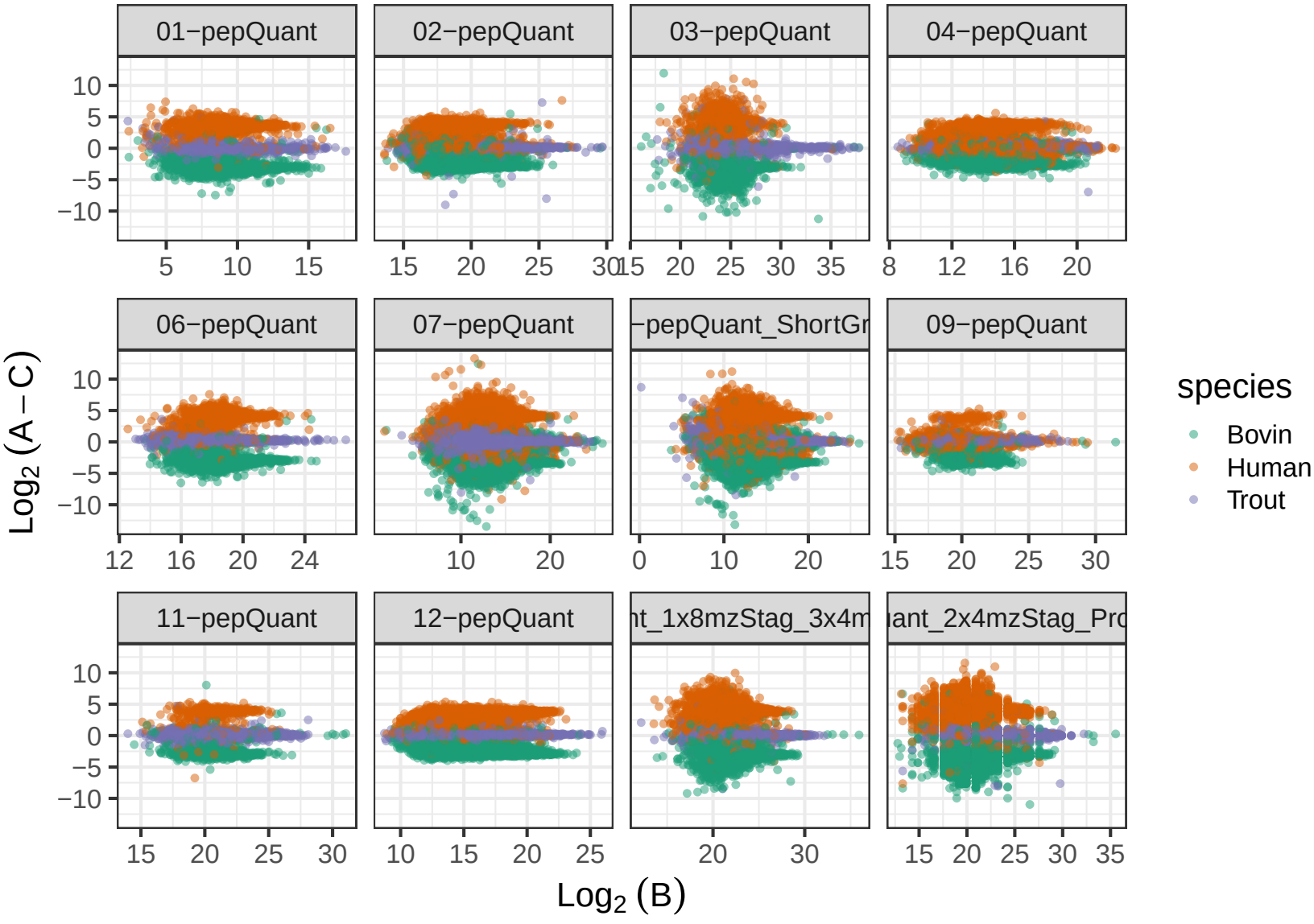
Adding the in-silico peptides

The 3 fasta proteomes were in-silico digested to peptides. Peptides were labeled according to the originating organism. The in-silico peptides are used to further filter the data

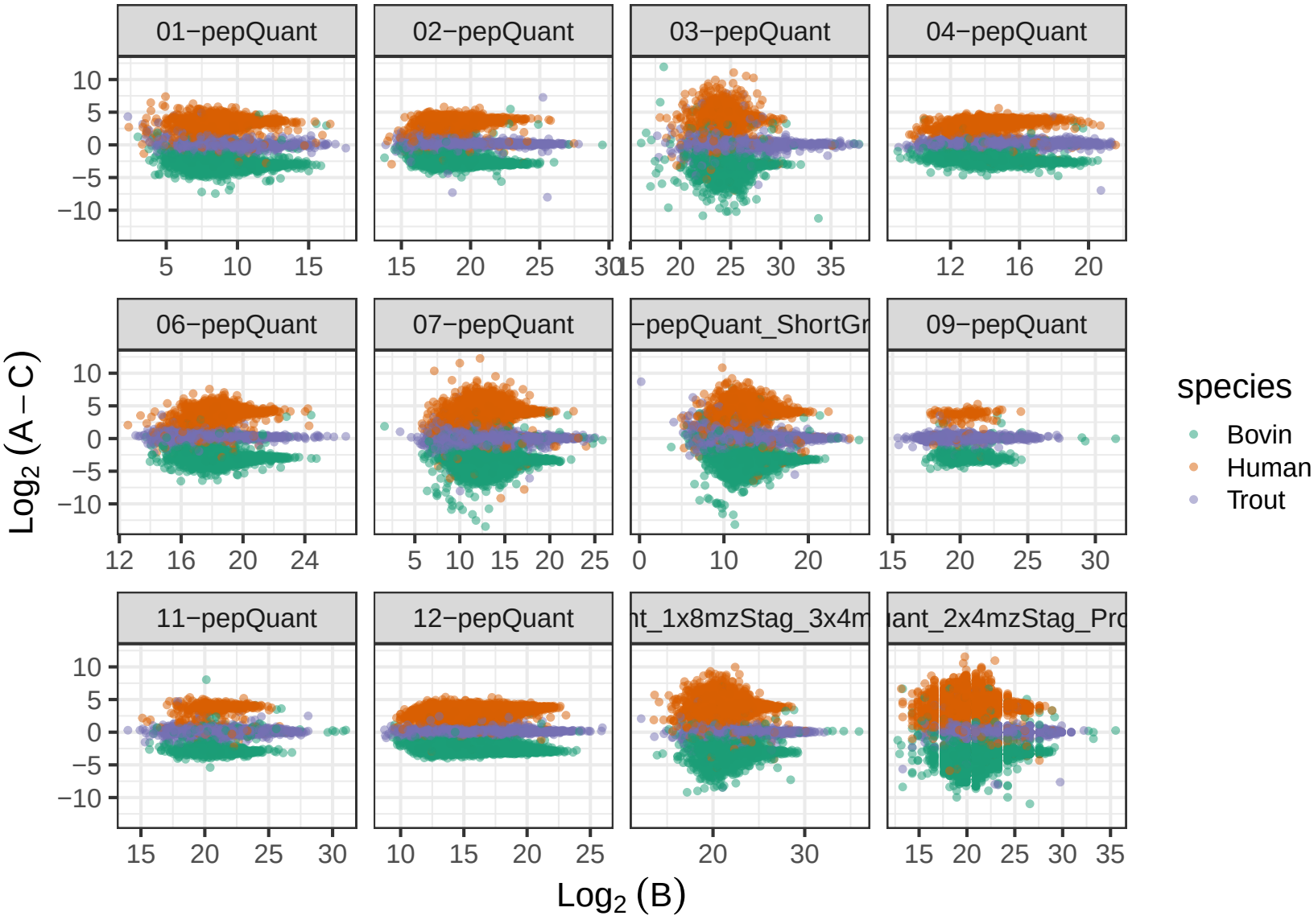
How peptides map to organisms using user data



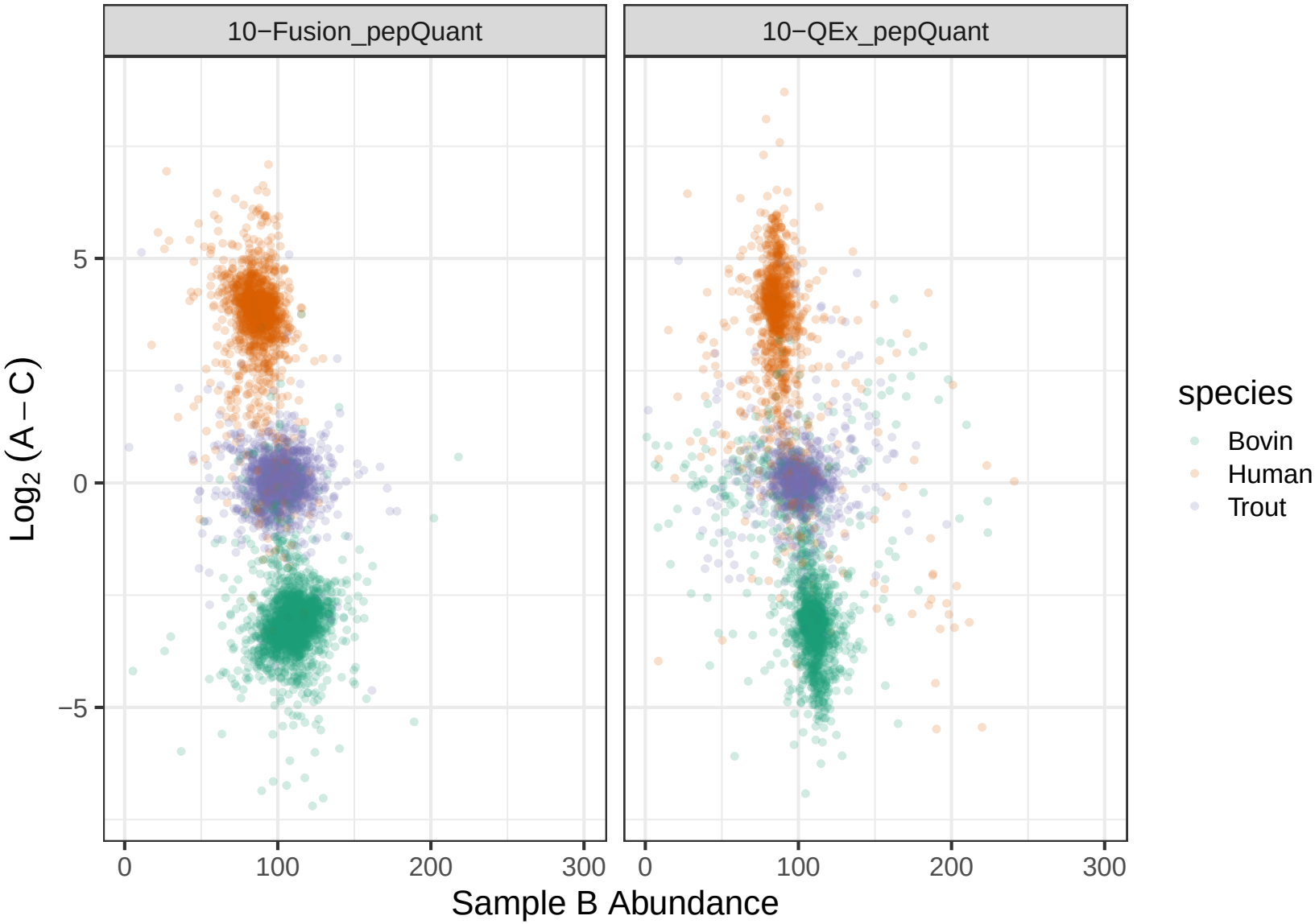
Samples - all data



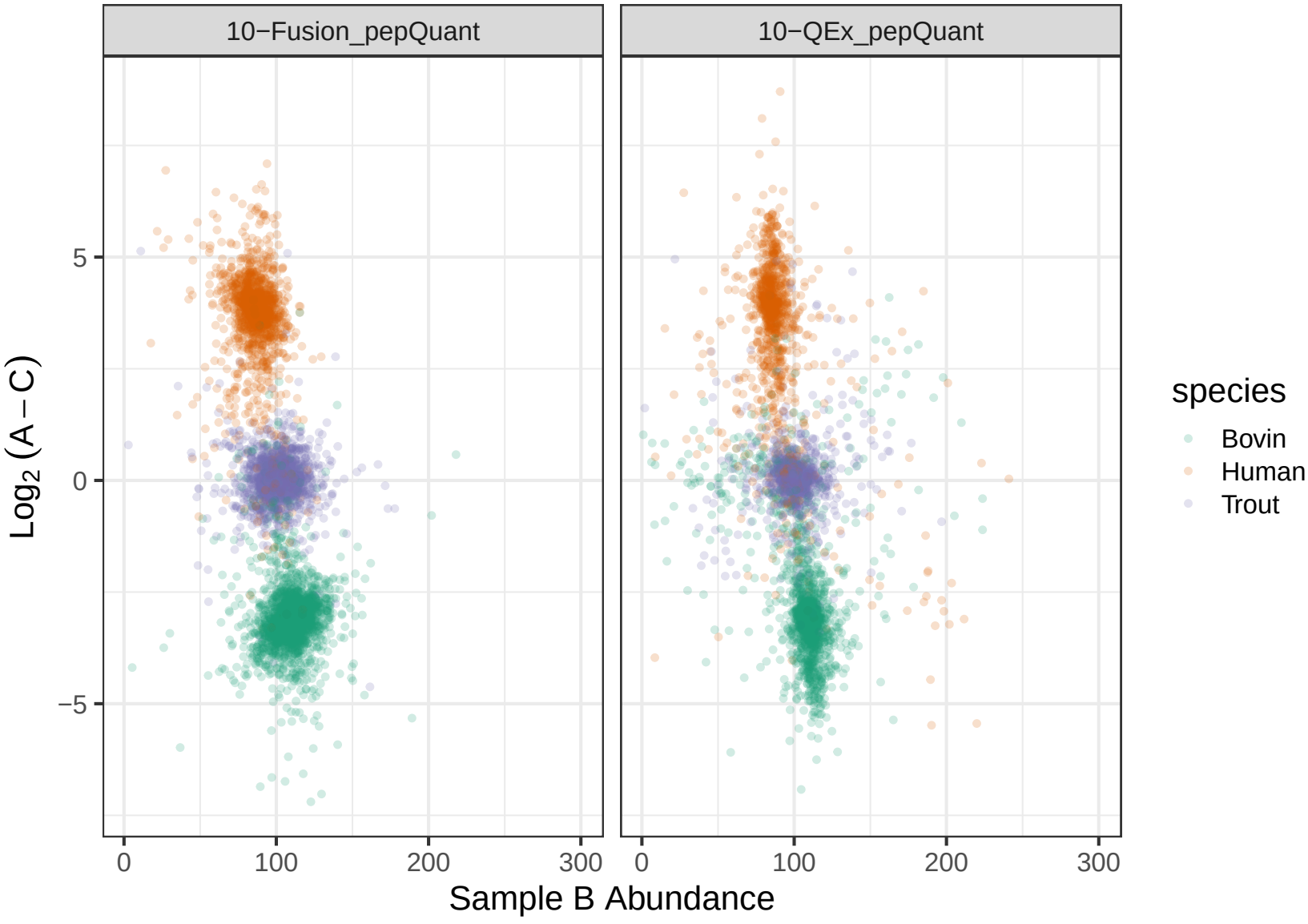
Samples - after filtering



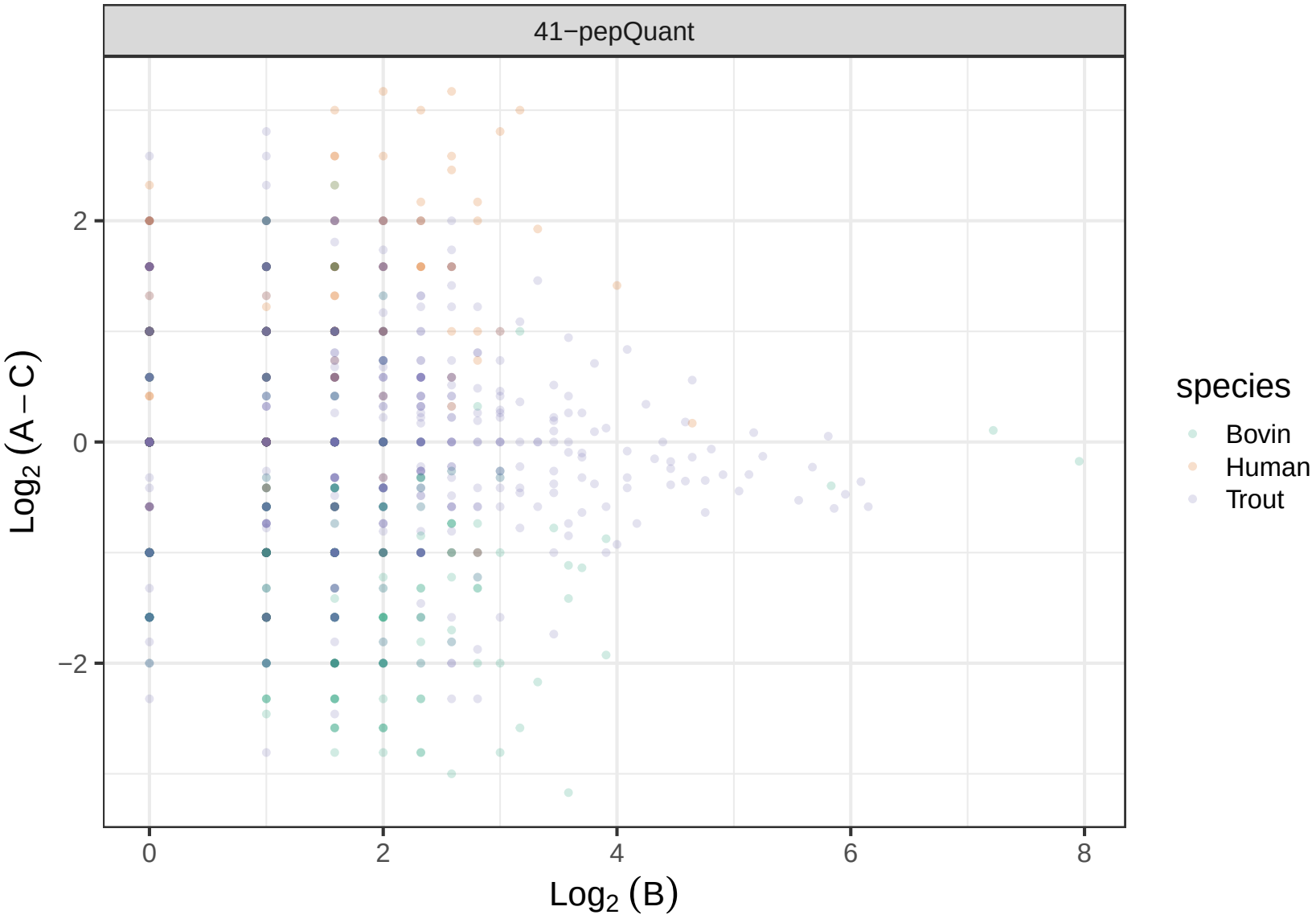
Sample 10 - all data



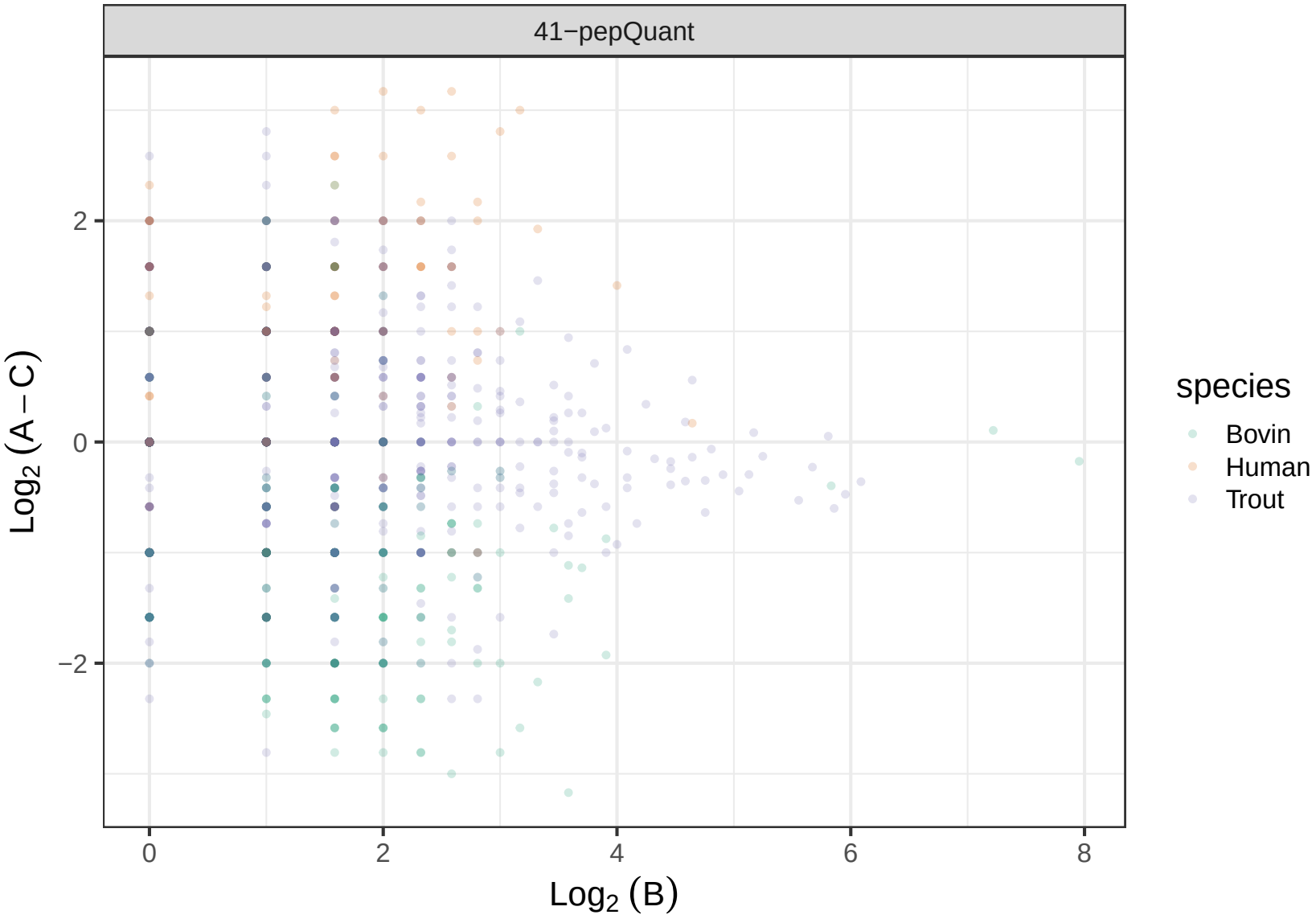
Sample 10 - after filtering



Sample 41 - all data



Sample 41 - after filtering



Calculating species ratios