Part 3: Kraken2 Statistics

Kraken2 classification results have been compiled and saved as CSV files.

The Kraken2 results data includes:

- Sample names
- Species classifications
- Unaligned clade reads
- Non−human clade reads
- Dataset information (unaligned/nonhuman)
- Merged input read counts from RRstats
- Additional optional columns (confidence, minimum hit groups, etc.)

Data source: Combined from unaligned_kreports/ and nonhuman_kreports/
Processing: output_processing.R with species filtering and merging
Merging: Combined with runtime and read statistics by sample name
Output location: Check outputs/ directory for:

  - sample_report_data_with_metadata*op.csv (combined data)
  - sample_report_data_[dataset]_*op.csv (dataset−specific files)

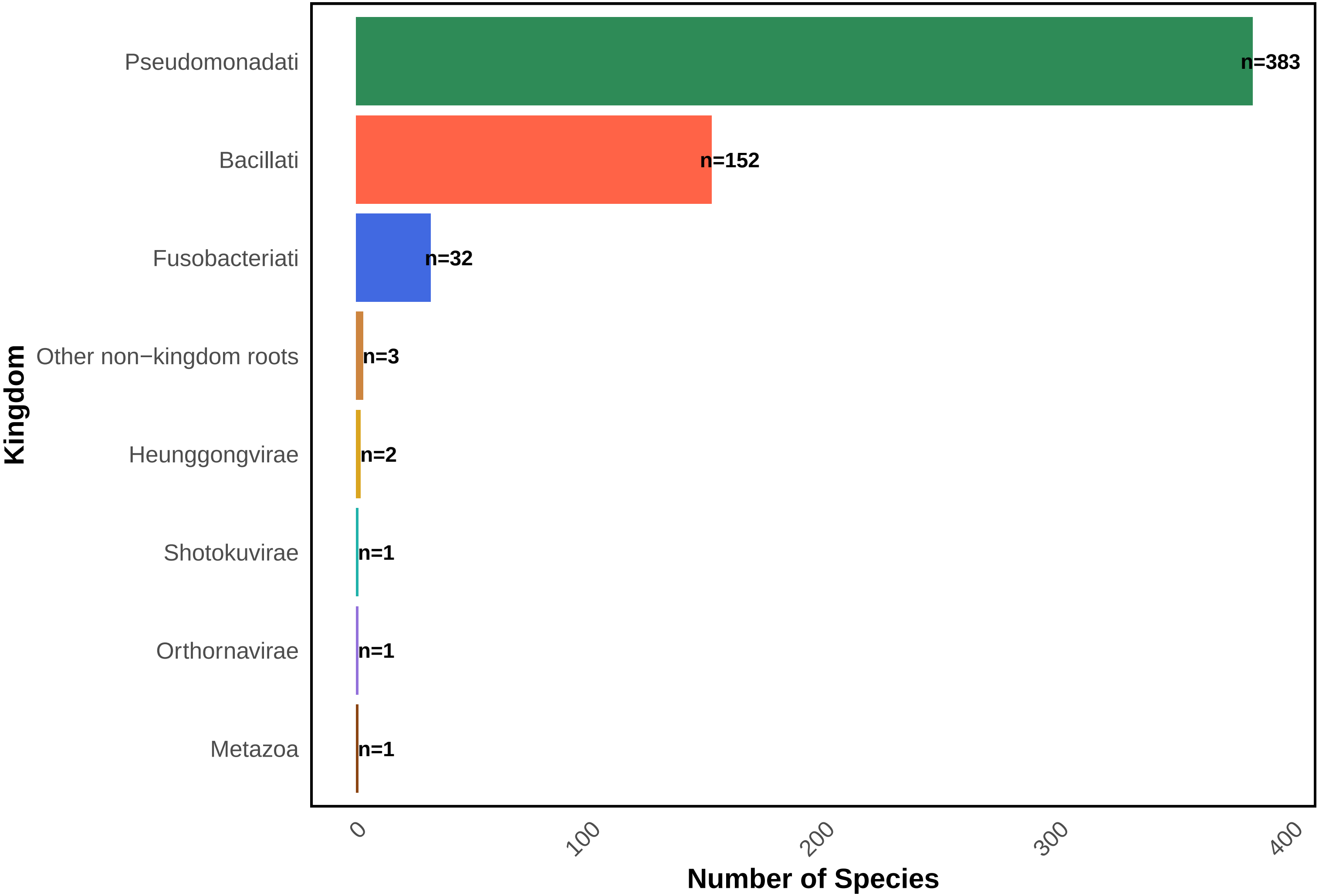
Key findings:

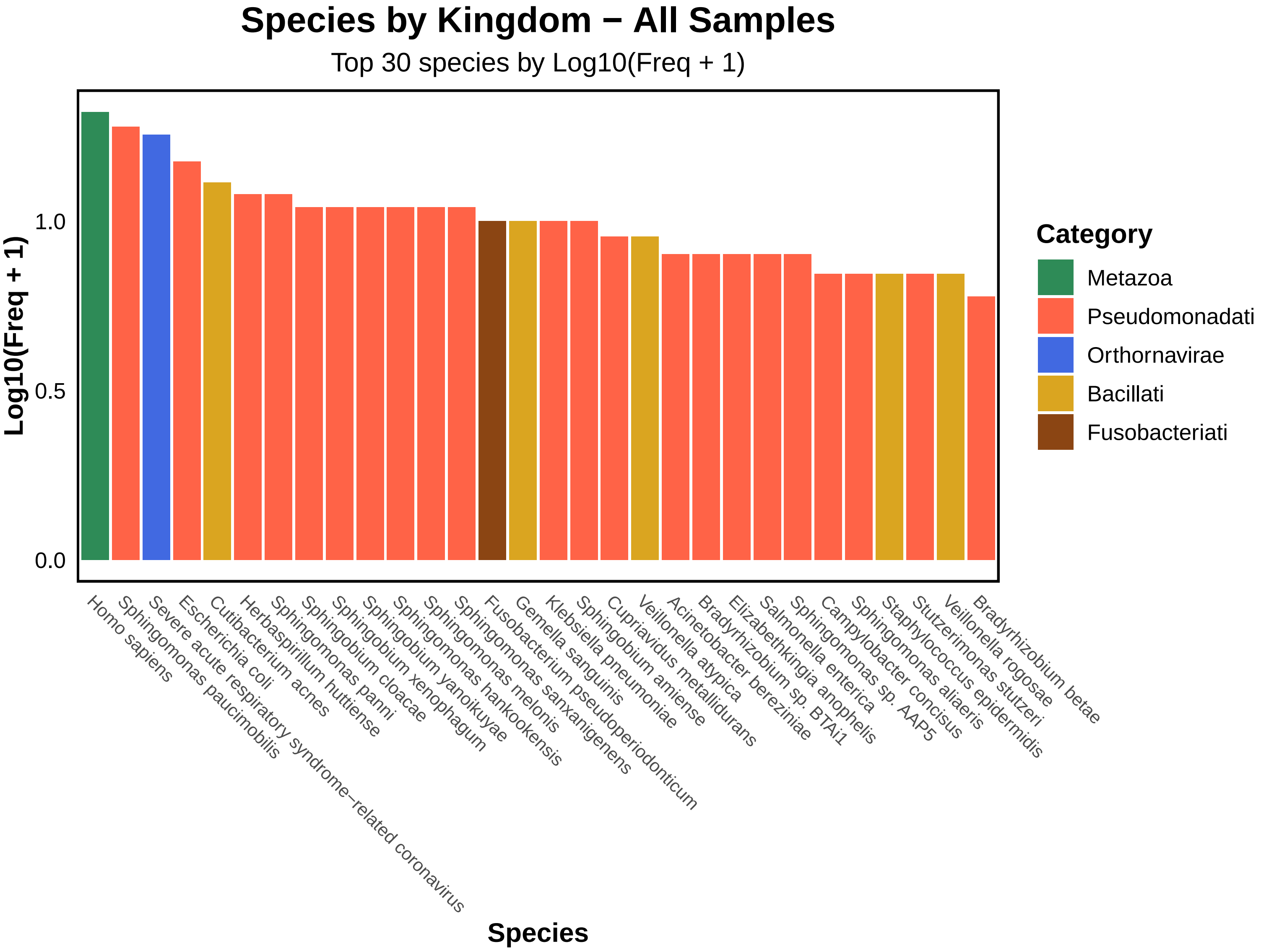
- Average number of species found per sample before secondary filtering: 611.55
- Average number of species found per sample after secondary filtering: 80.075
- Number of unique species across all unaligned samples: 579

Homo sapiens: 66784243.6
Severe acute respiratory syndrome−related coronavirus: 3677261.7
Fusobacterium pseudoperiodonticum: 541689.5
Solobacterium moorei: 460470.5
Cupriavidus pauculus: 378178.25
Veillonella atypica: 293441.4
Sphingomonas paucimobilis: 263914.894736842
Staphylococcus aureus: 223765.111111111
Selenomonas sputigena: 167271
Haemophilus parahaemolyticus: 162031

Species per Kingdom − All Samples

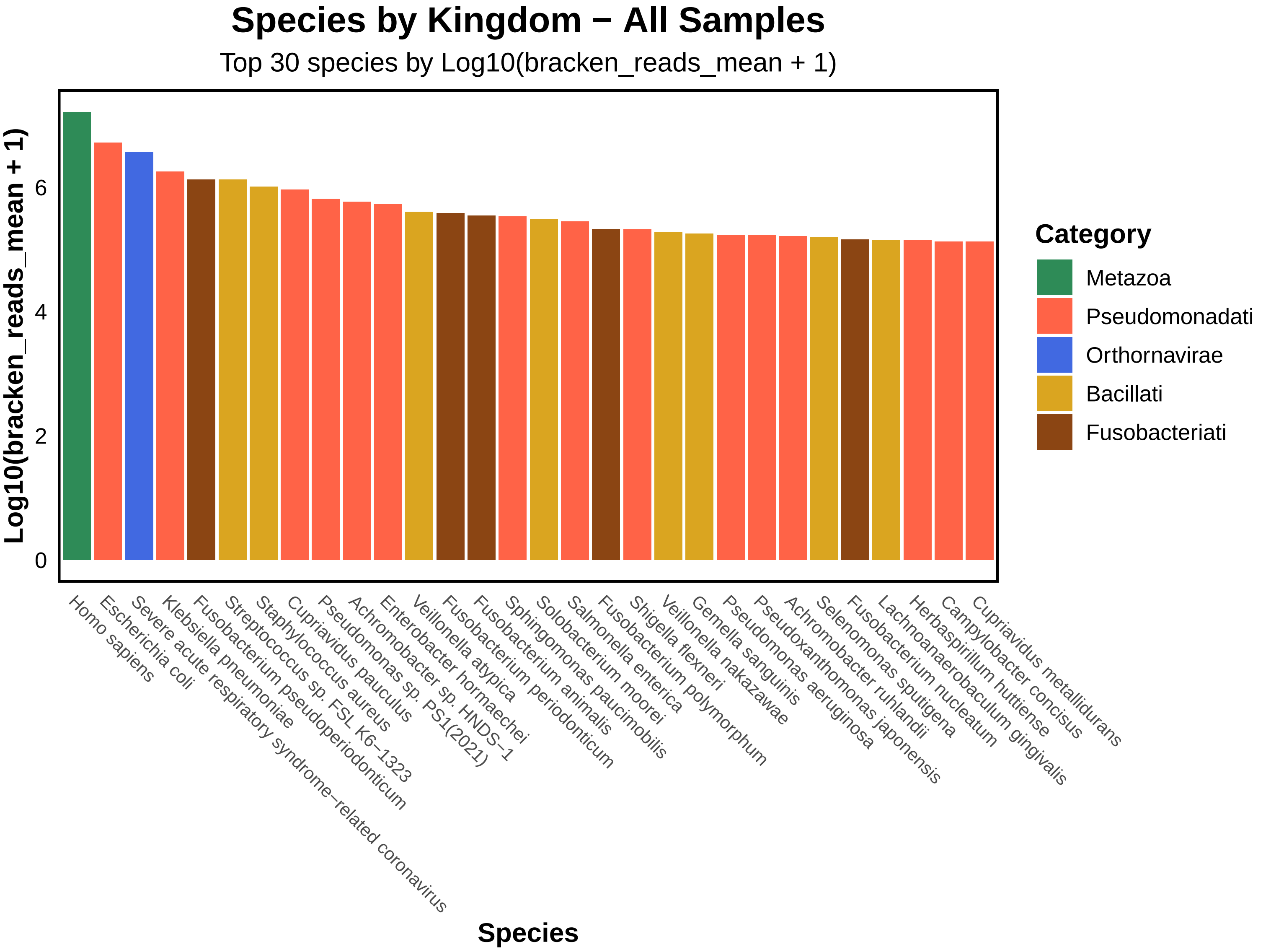
Total species: 575





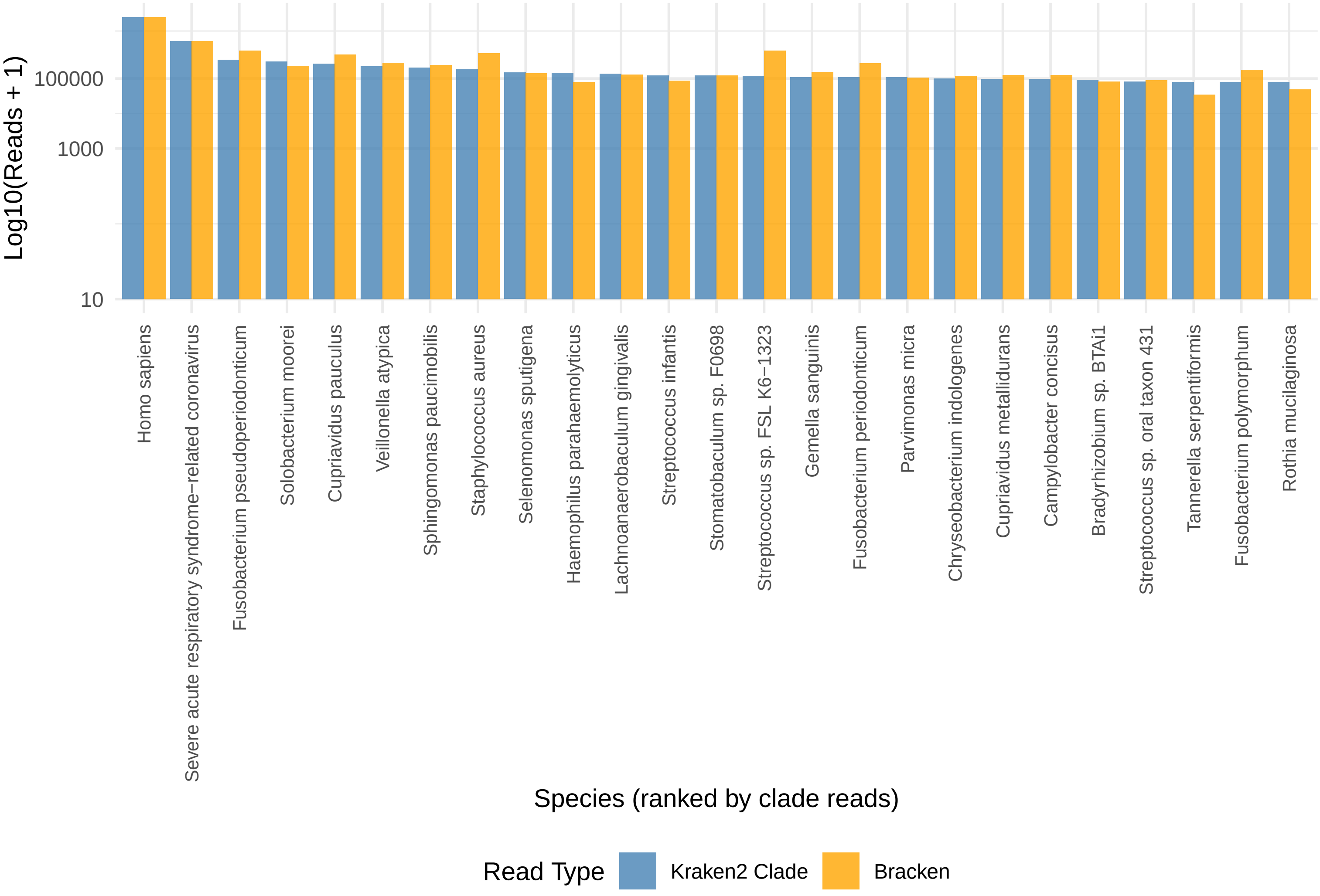
Species by Kingdom − All Samples

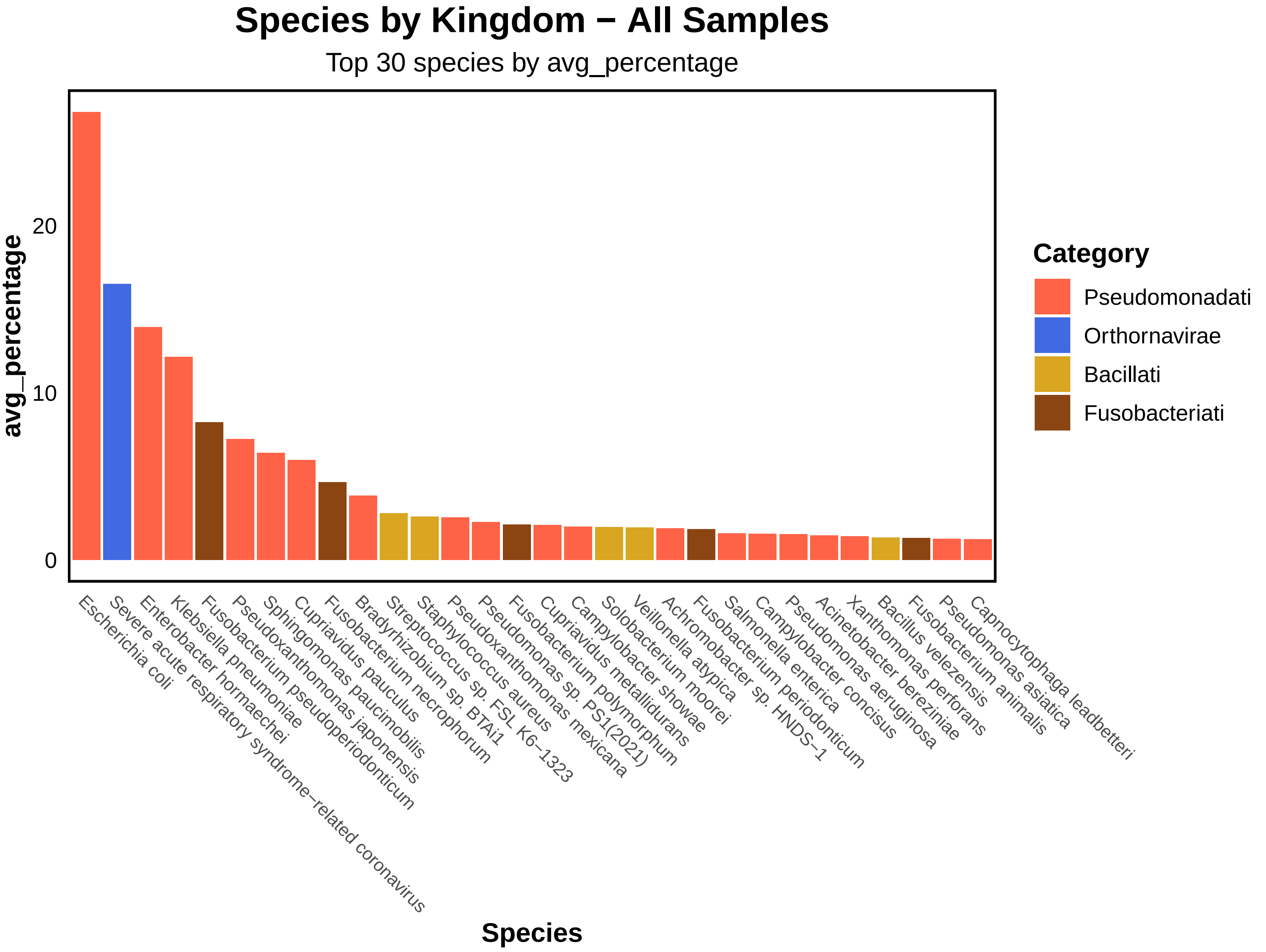
Top 30 species by Log10(Freq + 1)


Species by Kingdom − All Samples
Top 30 species by Log10(bracken_reads_mean + 1)
Log10(bracken_reads_mean + 1)
0
2
4
6
Species
Homo sapiens
Escherichia coli
Severe acute respiratory syndrome−related coronavirus
Klebsiella pneumoniae
Fusobacterium pseudoperiodonticum
Streptococcus sp. FSL K6−1323
Staphylococcus aureus
Cupriavidus pauculus
Pseudomonas sp. PS1(2021)
Achromobacter sp. HNDS−1
Enterobacter hormaechei
Veillonella atypica
Fusobacterium periodonticum
Fusobacterium animalis
Sphingomonas paucimobilis
Solobacterium moorei
Salmonella enterica
Fusobacterium polymorphum
Shigella flexneri
Veillonella nakazawae
Gemella sanguinis
Pseudomonas aeruginosa
Pseudoxanthomonas japonensis
Achromobacter ruhlandii
Selenomonas sputigena
Fusobacterium nucleatum
Lachnoanaerobaculum gingivalis
Herbaspirillum huttiense
Campylobacter concisus
Cupriavidus metallidurans
Category
Metazoa
Pseudomonadati
Orthornavirae
Bacillati
Fusobacteriati

Kraken2 Clade Reads vs Bracken Reads (Top 25 Species)

Ranked by Kraken2 clade reads (log10 scale)



Species by Kingdom − All Samples
Top 30 species by avg_percentage
avg_percentage
20
10
0
Species
Escherichia coli
Severe acute respiratory syndrome−related coronavirus
Enterobacter hormaechei
Klebsiella pneumoniae
Fusobacterium pseudoperiodonticum
Pseudoxanthomonas japonensis
Sphingomonas paucimobilis
Cupriavidus pauculus
Fusobacterium necrophorum
Bradyrhizobium sp. BTAi1
Streptococcus sp. FSL K6−1323
Staphylococcus aureus
Pseudoxanthomonas mexicana
Pseudomonas sp. PS1(2021)
Fusobacterium polymorphum
Cupriavidus metallidurans
Campylobacter showae
Solobacterium moorei
Veillonella atypica
Achromobacter sp. HNDS−1
Fusobacterium periodonticum
Salmonella enterica
Campylobacter concisus
Pseudomonas aeruginosa
Acinetobacter bereziniae
Xanthomonas perforans
Bacillus velezensis
Fusobacterium animalis
Pseudomonas asiatica
Capnocytophaga leadbetteri
Category
Pseudomonadati
Orthornavirae
Bacillati
Fusobacteriati