trnL reference

Contents

Stepping through this notebook will build a *trnL* reference database, trimmed to sequences amplified by the desired primer set (here, *trnL*g and *trnL*h).

This notebook use

Read in data

```
# Primer sequences
trnLG <- DNAString('GGGCAATCCTGAGCCAA')
trnLH <- DNAString('CCATTGAGTCTCTGCACCTATC')

primers <- list(trnLG, trnLH)
```

```
# Manually curated list of dietary and medicinal plants
plants <-
     here('data', 'inputs', 'human-foods.csv') |>
     read.csv(stringsAsFactors = FALSE) |>
     filter(category == 'plant') |>
     pull(scientific_name)

length(plants)
```

```
## [1] 1497
```

```
head(plants)
```

```
## [1] "Asphodelus tenuifolius" "Lagerstroemia speciosa" "Prosopis cineraria"
## [4] "Abelmoschus esculentus" "Abelmoschus manihot"    "Abutilon indicum"
```

```
tail(plants)
```

```
## [1] "Zizania latifolia"   "Zizania palustris"   "Ziziphus jujuba"
## [4] "Ziziphus mauritiana" "Medicago sativa"     "Cinchona pubescens"
```

```
# Manual edits
edits <-
     here('data', 'inputs', 'Manual renaming.csv') |>
     read_csv()
```

```
## Rows: 30 Columns: 8
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (8): type, accession, name_initial, name_update, sequence, date, by, notes
##
## 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.
```

```
edits
```

```
## # A tibble: 30 x 8
##    type   accession  name_initial         name_update sequence date  by    notes
##    <chr>  <chr>      <chr>                <chr>       <chr>    <chr> <chr> <chr>
##  1 rename LR031876.1 Brassica oleracea    Brassica o~ <NA>     1/4/~ BP    Iden~
##  2 rename AB213010.1 Brassica oleracea    Brassica o~ <NA>     1/7/~ BP    Iden~
##  3 rename AC183493.1 Brassica oleracea    Brassica o~ <NA>     1/7/~ BP    Iden~
##  4 rename LR031874.1 Brassica oleracea    Brassica o~ <NA>     1/7/~ BP    Iden~
##  5 rename LR031875.1 Brassica oleracea    Brassica o~ <NA>     <NA>  <NA>  <NA>
##  6 omit   JX390727.1 Murraya koenigii     <NA>        GGGTAAT~ 2/11~ BP    Susp~
##  7 omit   KF550171.1 Allium schoenoprasum <NA>        GGGCAAT~ 2/17~ BP    Susp~
##  8 omit   GU595140.1 Secale cereale       <NA>        GGGCAAT~ 5/20~ BP    Susp~
##  9 add    LR031876.1 Brassica oleracea    <NA>        <NA>     <NA>  <NA>  <NA>
## 10 add    LR031875.1 Brassica oleracea    <NA>        <NA>     <NA>  <NA>  <NA>
## # i 20 more rows
```

```
# SQL reference
sql <- here('accessionTaxa.sql')
```

```
# Parsed RefSeq data (last organized Jan 2023)
plastid <-
     readDNAStringSet(
          here('data',
               'outputs',
               'parsed-refs',
               'RefSeq',
               'refseq_plastid_species.fasta'))

plastid
```

```
## DNAStringSet object of length 11122:
##           width seq                                         names
##     [1]  130584 GGCATAAGCTATCTTCCCAA...GATTCAAACATAAAAGTCCT NC_018523.1 Sacch...
##     [2]  161592 ATGGGCGAACGACGGGAATT...AGAAAAAAAAAATAGGAGTAA NC_022431.1 Ascle...
##     [3]  117672 ATGAGTACAACTCGAAAGTC...GATTTCATCCACAAACGAAC NC_022259.1 Nanno...
##     [4]  154731 TTATCCATTTGTAGATGGAA...TATACACTAAGACAAAAGTC NC_022417.1 Cocos...
##     [5]  156618 GGGCGAACGACGGGAATTGA...TTTTGTAGCGAATCCGTTAT NC_022459.1 Camel...
##     ...     ... ...
## [11118]  114882 ATGAGTACAACTCGAAAGTC...GATTTTATTCACAAACGAAC NC_022261.1 Nanno...
## [11119]  117557 ATGAGTACAACTCGAAAGTC...GATTTCATCTACAAACGAAC NC_022263.1 Nanno...
## [11120]  156974 GGGCGAACGACGGGAATTGA...TTTTGTAGCGAATCCGTTAT NC_022264.1 Camel...
## [11121]  117806 ATGAGTACAACTCGAAAGTC...GATTTCATCCACAAACGAAC NC_022262.1 Nanno...
## [11122]  155461 TAATGGGCGAACGACGGGAA...CACAAAAGCAGAAAAAGAAA AC_000188.1 Solan...
```

Submit query

RefSeq (local)

Find primers

```r
# Note that there are lots of sequences that include Ns
length(plastid)
```

```
## [1] 11122
```

```r
length(clean(plastid))
```

```
## [1] 9777
```

```r
refseq.trnL <- find_primer_pair(plastid,
                                fwd = primers[[1]],
                                rev = primers[[2]])
```

```
## Warning in inner_join(fwd_matches, rev_rc_matches, by = "group", multiple = "all"): Detected an unex
## i Row 460 of `x` matches multiple rows in `y`.
## i Row 13 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
## Warning in inner_join(rev_matches, fwd_rc_matches, by = "group", multiple = "all"): Detected an unex
## i Row 4 of `x` matches multiple rows in `y`.
## i Row 2760 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```r
cat(length(refseq.trnL), 'sequences have the primer set')
```

```
## 10468 sequences have the primer set
```

Subset to foods

Just doing this by a simple grep for now Could imagine it would be cleaner to process RefSeq to a comparable name as in our query Keep accessions from raw files, look up taxonomy, then assign a “lowest level” name that would correspond to entries in human-foods.txt.

```r
# Find indices of entries matching
plants.i <-
     lapply(plants, grep, x = names(refseq.trnL)) %>%
     unlist()

cat('There are', length(plants), 'food plants in our query\n')
```

```
## There are 1497 food plants in our query
```

```r
# Subset
refseq.trnL <- refseq.trnL[plants.i]
cat(length(refseq.trnL), 'have a trnL sequence in the RefSeq plastid database')
```

```
## 691 have a trnL sequence in the RefSeq plastid database
```

```r
# Strip name to only NCBI accession
names(refseq.trnL) %>% head()
```

```
## [1] "NC_031414.1 Lagerstroemia speciosa" "NC_049133.1 Prosopis cineraria"
## [3] "NC_035234.1 Abelmoschus esculentus" "NC_053353.1 Abelmoschus manihot"
## [5] "NC_053702.1 Abutilon theophrasti"   "NC_051960.1 Acer saccharum"
```

```
names(refseq.trnL) <-
     gsub(names(refseq.trnL),
          pattern = ' .*$',
          replacement = '')

head(names(refseq.trnL))
```

```
## [1] "NC_031414.1" "NC_049133.1" "NC_035234.1" "NC_053353.1" "NC_053702.1"
## [6] "NC_051960.1"
```

NCBI (remote)

```
# Pull sequences from NCBI
ncbi.trnL <- query_ncbi(marker = 'trnL',
                        organisms = plants)
```

```
## Warning: package 'rentrez' was built under R version 4.3.2
```

```
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 1 species processed...
## Equal lengths: TRUE
## 15977 sequences processed for trnL
```

This is from the total number of available sequences

```r
length(ncbi.trnL)
```

```
## [1] 15977
```

```r
length(clean(ncbi.trnL))
```

```
## [1] 14889
```

Find primers

Now look for primer binding sites within retrieved sequences.

```r
ncbi.trnL <- find_primer_pair(ncbi.trnL,
                              fwd = primers[[1]],
                              rev = primers[[2]])

cat(length(ncbi.trnL), 'sequences have the primer set')
```

```
## 7220 sequences have the primer set
```

```r
# Note some entries are marked as unverified
names(ncbi.trnL)[grepl('UNVERIFIED', names(ncbi.trnL))] |>
     head(5)
```

```
## [1] ">KM385468.1 UNVERIFIED: Achillea millefolium isolate Ach_mil tRNA-Leu (trnL) gene, partial seque
## [2] ">KM113357.1 UNVERIFIED: Amorphophallus paeoniifolius var. paeoniifolius voucher ARG-37 tRNA-Leu
## [3] ">KM113356.1 UNVERIFIED: Amorphophallus paeoniifolius var. campanulatus voucher ARG-34 tRNA-Leu
## [4] ">JQ352618.1 UNVERIFIED: Antigonon leptopus isolate 440 tRNA-Leu (trnL) gene and trnL-trnF interg
## [5] ">KM385472.1 UNVERIFIED: Avena sativa isolate Ave_sat tRNA-Leu (trnL) gene, partial sequence; tr
```

```r
# Remove
length(ncbi.trnL)
```

```
## [1] 7220
```

```r
ncbi.trnL <- ncbi.trnL[!(grepl('UNVERIFIED', names(ncbi.trnL)))]
length(ncbi.trnL)
```

```
## [1] 7027
```

Want to convert these long, descriptive names to just an accession and taxon. Strip to accession number, and then use this to do a taxonomic lookup.

```r
# Strip name to only NCBI accession
names(ncbi.trnL) |> head()
```

```
## [1] ">AB933512.1 Asphodelus tenuifolius chloroplast DNA, tRNA-Leu (trnL) and trnL-trnF intergenic spa
## [2] ">NC_031414.1 Lagerstroemia speciosa chloroplast, complete genome"
## [3] ">KU821692.1 Lagerstroemia speciosa chloroplast, complete genome"
## [4] ">AY905468.1 Lagerstroemia speciosa voucher Shi 99060103 (SYS) tRNA-Leu (trnL) gene and trnL-trnF
## [5] ">AF354180.1 Lagerstroemia speciosa tRNA-Leu gene and trnL-trnF spacer, partial sequence; chlorop
## [6] ">EF165292.1 Prosopis cineraria tRNA-Leu (trnL) gene, partial sequence; trnL-trnF intergenic spa
```

```r
names(ncbi.trnL) <-
     names(ncbi.trnL) |>
     gsub(pattern = ' .+$', replacement = '') |>
     gsub(pattern = '^>', replacement = '')

head(names(ncbi.trnL))
```

```
## [1] "AB933512.1"  "NC_031414.1" "KU821692.1"  "AY905468.1"  "AF354180.1"
## [6] "EF165292.1"
```

Combine

Check overlap

```r
length(refseq.trnL)
```

```
## [1] 691
```

```r
length(ncbi.trnL)
```

```
## [1] 7027
```

```r
# Named as accession numbers:
intersect(names(ncbi.trnL), names(refseq.trnL)) |> length()
```

```
## [1] 59
```

```r
setdiff(names(refseq.trnL), names(ncbi.trnL)) |> length()
```

```
## [1] 614
```

```r
setdiff(names(ncbi.trnL), names(refseq.trnL)) |> length()
```

```
## [1] 6741
```

Theoretically, RefSeq is entirely contained within NCBI's nucleotide record, but there are entries that are unique to RefSeq here. Think this is because I restrict the query to "big" NCBI to have the term "trnL" in the record name– not impossible to overcome, but currently don't have a strategy for handling the # and length of records that get pulled down without that filter in our query term. This can be an area for future updates.

Merge

```r
# Data frame of results
seqs.df <-
     data.frame(source = 'RefSeq',
                accession = names(refseq.trnL),
                seq = as.character(refseq.trnL))

seqs.df <-
     data.frame(source = 'GenBank',
                accession = names(ncbi.trnL),
                seq = as.character(ncbi.trnL)) |>
     bind_rows(seqs.df)

head(seqs.df)
```

```
##    source   accession
## 1 GenBank  AB933512.1
## 2 GenBank NC_031414.1
## 3 GenBank  KU821692.1
## 4 GenBank  AY905468.1
## 5 GenBank  AF354180.1
## 6 GenBank  EF165292.1
##                                                                                                  seq
## 1           GGGCAATCCTGAGCCAAATCTTTTTTTTTTTGAAAAACTGATTAATCGGACAAGAATAAAAAAGGATAGGTGCAGAGACTCAATGG
## 2  GGGCAATCCTGAGCCAAATCCTATTTTGTACGAAAACCAACAACAAGGGTTTAGAAAGCGAGAATAAAAAAAGGATAGGTGCAGAGACTCAACGG
## 3  GGGCAATCCTGAGCCAAATCCTATTTTGTACGAAAACCAACAACAAGGGTTTAGAAAGCGAGAATAAAAAAAGGATAGGTGCAGAGACTCAACGG
## 4 GGGCAATCCTGAGCCAAATCCTATTTTTGTACGAAAACCAACAACAAGGGTTTAGAAAGCGAGAATAAAAAAAGGATAGGTGCAGAGACTCAACGG
## 5 GGGCAATCCTGAGCCAAATCCTATTTTTTTACGAAAACCAACAACAAGGGTTTAGAAAGCGAGAATAAAAAAAGGATAGGTGCAGAGACTCAACGG
## 6       GGGCAATCCTGAGCCAAATCCTGTTTTCCGAAAACCAAGAAGAGTTCAGAAAGGGAGAATAAAAAAAGGATAGGTGCANANACNCAACGG
```

```r
# Also add manual additions here
additions <- filter(edits, type == 'add')
additions
```

```
## # A tibble: 22 x 8
##    type  accession   name_initial         name_update sequence date  by    notes
##    <chr> <chr>       <chr>                <chr>       <chr>    <chr> <chr> <chr>
##  1 add   LR031876.1  Brassica oleracea    <NA>        <NA>     <NA>  <NA>  <NA>
##  2 add   LR031875.1  Brassica oleracea    <NA>        <NA>     <NA>  <NA>  <NA>
##  3 add   MN082371.1  Corylus avellana     <NA>        <NA>     <NA>  <NA>  <NA>
##  4 add   LR031874.1  Brassica oleracea    <NA>        <NA>     <NA>  <NA>  <NA>
##  5 add   HG994366.1  Brassica napus       <NA>        <NA>     <NA>  <NA>  <NA>
##  6 add   MK187055.1  Eclipta prostrata    <NA>        <NA>     <NA>  <NA>  Not ~
##  7 add   NC_039346.1 Sphagneticola calen~ <NA>        <NA>     <NA>  <NA>  Not ~
##  8 add   MZ997428.1  Rheum rhabarbarum    <NA>        <NA>     <NA>  <NA>  <NA>
##  9 add   LT606769.1  Vaccinium uliginosum <NA>        <NA>     <NA>  <NA>  <NA>
## 10 add   LT606432.1  Vaccinium vitis-ida~ <NA>        <NA>     <NA>  <NA>  <NA>
## # i 12 more rows
```

```
# Note that these don't have primers currently
# To get the most accurate sequence, let's just pull these records from NCBI by their accession number
# Note some of these returned sequences are whole genomes-- takes a few mins.
seqs <-
     entrez_fetch(db='nucleotide',
                  id = additions$accession,
                  rettype='fasta') %>%
     # This returns concatenated sequence strings; split apart
     # so we can re-name inline
     strsplit('\n{2,}') %>% # Usually two newline chars, but sometimes more
     unlist()

# Save this to ultimately combine with taxonomy data, as want to
# be able to identify these sequences after the fact
ex <- '[^>]\\S*'
accs <- str_extract(seqs, ex)

# Keep full header for descriptive name
headers <- str_extract(seqs, '^[^\n]*')

seqs <-
     seqs %>%
     # Now update seqs to sequence only, stripping header
     sub('^[^\n]*\n', '', .) %>%
     # And removing separating \n characters
     gsub('\n', '', .)

# Now add to DNAStringSet
seqs <- DNAStringSet(seqs)
names(seqs) <- accs

seqs <- find_primer_pair(seqs,
                         fwd = primers[[1]],
                         rev = primers[[2]])
```

```
## Warning in inner_join(fwd_matches, rev_rc_matches, by = "group", multiple = "all"): Detected an unex
## i Row 1 of `x` matches multiple rows in `y`.
## i Row 1 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.

## Warning in inner_join(rev_matches, fwd_rc_matches, by = "group", multiple = "all"): Detected an unex
## i Row 1 of `x` matches multiple rows in `y`.
## i Row 1 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

```
# This leaves 'source' labeled as NA for these entries
seqs.df <-
     data.frame(seq = as.character(seqs),
                accession = names(seqs)) |>
     bind_rows(seqs.df)
```

Taxonomy

```
# Look up accession taxonomy using taxonomizr-formatted SQL database
ids <- taxonomizr::accessionToTaxa(seqs.df$accession, sql)
```

```
## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bca767031', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bca767031', reason
## 'Permission denied'
```

```
# Any names missing?
any(is.na(ids))
```

```
## [1] TRUE
```

```
sum(is.na(ids))
```

```
## [1] 3
```

```
# Which ones?
missing.df <- seqs.df[is.na(ids), c('source', 'accession')]
missing.df
```

```
##       source   accession
## 1849 GenBank  OQ942624.1
## 1850 GenBank  OQ942623.1
## 7172  RefSeq NC_027223.1
```

Missing entries are sequence records that have been added to NCBI in the time between making the taxonomic SQL file and now. Note the output of the previous chunk. The first column describes the index of ids, the third column is the accession ID, which you will manually query in the NCBI website. From the result, find the taxon ID. This will be the value at the right-hand side below.

```
source(here('code', 'functions', 'query_ncbi_accession.R'))

for(i in seq_len(nrow(missing.df))) {
     idx <- rownames(missing.df)[i]
     idx <- as.integer(idx)
     acc <- missing.df[i,2]
     ids[idx] <- query_ncbi_accession(acc)
}

# Manually add these by querying in browser with accession
# IMPORTANT: If repeating, check these to make sure they still line up
# If any additions/changes to input files, likely they will not
#ids[5742] <- 3483 # NC_027223.1 Cannabis sativa
```

```
# Confirm got them all
any(is.na(ids))
```

```
## [1] FALSE
```

```
taxonomy.raw <- taxonomizr::getRawTaxonomy(ids, sql)
```

```
## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc3a5910e6', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc718d38a4', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc26715941', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc3bd913ea', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc795a788f', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc430f4723', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc1c5813db', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc35a12e4b', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc5876a9a', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc2fbf2522', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc50b65c31', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc3aca58d1', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc5a7866b8', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc610a5cd0', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc11792fe', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc29a3707', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc720f1cd', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc10c55734', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc60114e6f', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc50246d8c', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc52ec3a70', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bcba854a5', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc7f336a6', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc36652293', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc2ef71d6f', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc43cd3552', reason
## 'Permission denied'

## Warning in file.remove(tmp): cannot remove file
## 'C:\Users\PAUL\AppData\Local\Temp\RtmpC8qvwx\file13bc4518723a', reason
## 'Permission denied'
```

```r
# Pull desired levels from this structure
# Not working within getTaxonomy function
vars <- c("superkingdom",
          "phylum",
          "class",
          "order",
          "family",
          "genus",
          "species",
          "subspecies",
          "varietas",
          "forma")

taxonomy <- data.frame(superkingdom = NULL,
                       phylum = NULL,
                       class = NULL,
                       order = NULL,
                       family = NULL,
                       genus = NULL,
                       species = NULL,
                       subspecies = NULL,
                       varietas = NULL,
                       forma = NULL)

for (i in seq_along(taxonomy.raw)){
     row.i <-
          taxonomy.raw[[i]] |>
          t() |>
          data.frame()

     # Pick columns we're interested in
     shared <- intersect(vars, names(row.i))
     row.i <- select(row.i, one_of(shared))

     taxonomy <- bind_rows(taxonomy, row.i)
}

# Add taxon ID
taxonomy$taxid <-
     names(taxonomy.raw) |>
     trimws() |>
     as.integer()

taxonomy <- select(taxonomy, taxid, everything())
```

```r
head(taxonomy)
```

```
##   taxid superkingdom       phylum         class       order       family
## 1  3712    Eukaryota Streptophyta Magnoliopsida Brassicales Brassicaceae
## 2  3712    Eukaryota Streptophyta Magnoliopsida Brassicales Brassicaceae
## 3  3712    Eukaryota Streptophyta Magnoliopsida Brassicales Brassicaceae
## 4 13451    Eukaryota Streptophyta Magnoliopsida     Fagales   Betulaceae
## 5  3712    Eukaryota Streptophyta Magnoliopsida Brassicales Brassicaceae
## 6  3708    Eukaryota Streptophyta Magnoliopsida Brassicales Brassicaceae
##      genus           species varietas subspecies forma
## 1 Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 2 Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 3 Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 4  Corylus  Corylus avellana     <NA>       <NA>  <NA>
## 5 Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 6 Brassica    Brassica napus     <NA>       <NA>  <NA>
```

```r
# Join back to accession
nrow(taxonomy) == nrow(seqs.df)
```

```
## [1] TRUE
```

```r
seqs.df <-
     bind_cols(seqs.df,
               taxonomy)
```

Manual edits

There are two types of edits we're making: - **omissions**: High likelihood an included sequence has an incorrect taxon label, so we exclude it, and - **renaming**: The labeled taxon name can be specified more precisely (currently, happens only for *B. oleracea* spp.)

```r
edits
```

```
## # A tibble: 30 x 8
##    type   accession  name_initial         name_update sequence date  by    notes
##    <chr>  <chr>      <chr>                <chr>       <chr>    <chr> <chr> <chr>
##  1 rename LR031876.1 Brassica oleracea    Brassica o~ <NA>     1/4/~ BP    Iden~
##  2 rename AB213010.1 Brassica oleracea    Brassica o~ <NA>     1/7/~ BP    Iden~
##  3 rename AC183493.1 Brassica oleracea    Brassica o~ <NA>     1/7/~ BP    Iden~
##  4 rename LR031874.1 Brassica oleracea    Brassica o~ <NA>     1/7/~ BP    Iden~
##  5 rename LR031875.1 Brassica oleracea    Brassica o~ <NA>     <NA>  <NA>  <NA>
##  6 omit   JX390727.1 Murraya koenigii     <NA>        GGGTAAT~ 2/11~ BP    Susp~
##  7 omit   KF550171.1 Allium schoenoprasum <NA>        GGGCAAT~ 2/17~ BP    Susp~
##  8 omit   GU595140.1 Secale cereale       <NA>        GGGCAAT~ 5/20~ BP    Susp~
##  9 add    LR031876.1 Brassica oleracea    <NA>        <NA>     <NA>  <NA>  <NA>
## 10 add    LR031875.1 Brassica oleracea    <NA>        <NA>     <NA>  <NA>  <NA>
## # i 20 more rows
```

```r
# Handle omissions
omit <- filter(edits, type=='omit')

seqs.df <-
     filter(seqs.df,
            !(accession %in% omit$accession & seq %in% omit$sequence))
```

```r
# Handle renaming
name.update <- filter(edits, type=='rename')

filter(seqs.df,
       accession %in% name.update$accession)
```

```
##                                                                              seq
## 1 GGGCAATCCTGAGCCAAATCCTGGGTTACGCGAACAAACCAGAGTTTAGAAAGCGGGATAGGTGCAGAGACTCAATGG
## 2 GGGCAATCCTGAGCCAAATCCTGGGTTACGCGAACAAACCAAAGTTTAGAAAGCGGGATAGGTGCAGAGACTCAATGG
## 3 GGGCAATCCTGAGCCAAATCATGGGTTACGCGAACAAACCAAAGTTTAGAAAGCGGGATAGGTGCAGAGACTCAATGG
## 4 GGGCAATCCTGAGCCAAATCTTGGGTTACGCGAACAAACCAGAGTTTAGAAAGCGGGATAGGTGCAGAGACTCAATGG
## 5 CCATTGAGTCTCTGCACCTATCCCGCTTTCTAAACTTTGGTTTGTTCGCGTAACCTAGGATTTGGCTCAGGATTGCCC
## 6 CCATTGAGTCTCTGCACCTATCCCGCTTTCTAAACTTTGGTTTGTTCGCGTAACCCAGGATTTGGCTCAGGATTGCCC
## 7 GGGCAATCCTGAGCCAAATCCTGGGTTACGCGAACAAAACAGAGTTTAGAAAGCGGGATAGGTGCAGAGACTCAATGG
##    accession  source taxid superkingdom       phylum         class       order
## 1 LR031876.1    <NA>  3712    Eukaryota Streptophyta Magnoliopsida Brassicales
## 2 LR031875.1    <NA>  3712    Eukaryota Streptophyta Magnoliopsida Brassicales
## 3 LR031875.1    <NA>  3712    Eukaryota Streptophyta Magnoliopsida Brassicales
## 4 LR031874.1    <NA>  3712    Eukaryota Streptophyta Magnoliopsida Brassicales
## 5 LR031876.1    <NA>  3712    Eukaryota Streptophyta Magnoliopsida Brassicales
## 6 AC183493.1    <NA>  3712    Eukaryota Streptophyta Magnoliopsida Brassicales
## 7 AB213010.1 GenBank  3712    Eukaryota Streptophyta Magnoliopsida Brassicales
##         family    genus           species varietas subspecies forma
## 1 Brassicaceae Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 2 Brassicaceae Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 3 Brassicaceae Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 4 Brassicaceae Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 5 Brassicaceae Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 6 Brassicaceae Brassica Brassica oleracea     <NA>       <NA>  <NA>
## 7 Brassicaceae Brassica Brassica oleracea     <NA>       <NA>  <NA>
```

```
# Note that original sequence 'AC183493.1' not found, leaving off for now
# Will need to generalize this later, but now can just update specifically
seqs.df$varietas[seqs.df$accession == 'AB213010.1'] <- 'Brassica oleracea var. capitata'
seqs.df$varietas[seqs.df$accession == 'AC183493.1'] <- 'Brassica oleracea var. alboglabra'
seqs.df$varietas[seqs.df$accession == 'LR031874.1'] <- 'Brassica oleracea var. italica'
seqs.df$varietas[seqs.df$accession == 'LR031875.1'] <- 'Brassica oleracea var. italica'
seqs.df$varietas[seqs.df$accession == 'LR031876.1'] <- 'Brassica oleracea var. italica'
```

```
# Get lowest-level taxon name
seqs.df <-
     seqs.df |>
     MButils::lowest_level() |>
     rename(taxon = 'name') |>
     select(source, accession, taxon, taxid, superkingdom:forma, seq)
```

QC Check for common errors

```
# Check for degenerate nucleotide characters
grep('[AGCT]*[^AGCT]+', seqs.df$seq)
```

```
##  [1]   31   84  410  411  412  413  414  415  418  419  420  421  426  427  428
## [16]  431  432  433  434  435  436  535  711  822  824  825  826  827  828  829
## [31]  830  831  832  833  834  835  841  842  843 1353 1726 1737 1856 1859 1860
## [46] 1971 2011 2039 2125 2190 2293 2296 2301 2366 3204 3315 3419 3433 4265 4514
## [61] 4515 4558 4747 4793 4911 5883 6103 6202 6861
```

```
# Add a flag to these taxa, to see if there's a back-up sequence
seqs.df$N <- grepl('[AGCT]*[^AGCT]+', seqs.df$seq)

with_n <-
     seqs.df$taxon[grepl('[AGCT]*[^AGCT]+', seqs.df$seq)] |>
     unique()
```

```
seqs.df |>
     filter(taxon %in% with_n) |>
     group_by(taxon, N) |>
     count() |>
     ungroup() |>
     group_by(taxon) |>
     summarize(any(!N)) |>
     arrange(`any(!N)`)
```

```
## # A tibble: 39 x 2
##    taxon                  `any(!N)`
##    <chr>                  <lgl>
##  1 Amaranthus cruentus    FALSE
##  2 Amaranthus dubius      FALSE
##  3 Amaranthus thunbergii FALSE
##  4 Cordia africana        FALSE
##  5 Costus erythrophyllus FALSE
##  6 Vaccinium tenellum     FALSE
##  7 Victoria amazonica     FALSE
##  8 Amaranthus caudatus    TRUE
##  9 Amaranthus graecizans TRUE
## 10 Amaranthus hybridus    TRUE
## # i 29 more rows
```

Okay, so for all taxa but - Amaranthus cruentus - Amaranthus dubius - Costus erythrophyllus - Vaccinium tenellum - Victoria amazonica

we are covered by a second sequence.

```
# Remove sequences containing Ns
seqs.df <-
     filter(seqs.df,
            !grepl(pattern = '[AGCT]*[^AGCT]+', seq))
```

```

# Get orientation of sequence by finding primers
# How many mismatches are allowed?
fwd_err <- floor(0.2*length(trnLG))
rev_err <- floor(0.2*length(trnLH))

fwd_err
```

Get reads in same orientation

```
## [1] 3
```

```
rev_err
```

```
## [1] 4
```

```
ref <- DNAStringSet(seqs.df$seq)
names(ref) <- paste(seqs.df$accession, seqs.df$taxon)
```

```
# Forward primer at start of read
fwd_matches <-
     vmatchPattern(trnLG,
                   ref,
                   max.mismatch = fwd_err,
                   fixed = TRUE) |>
     as.data.frame() |>
     filter(start <= 1) |>
     mutate(type = 'forward') |>
     select(group, type)

# Reverse primer at start of read
rev_matches <-
     vmatchPattern(trnLH,
                   ref,
                   max.mismatch = rev_err,
                   fixed = TRUE) |>
     as.data.frame() |>
     filter(start <= 1) |>
     mutate(type = 'reverse') |>
     select(group, type)
```

```
seqs.df <-
     bind_rows(fwd_matches,
          rev_matches) |>
     arrange(group) |>
     bind_cols(seqs.df)

seqs.df |>
     group_by(type) |>
     count()
```

```
## # A tibble: 2 x 2
## # Groups:   type [2]
##   type        n
##   <chr>   <int>
## 1 forward  6983
## 2 reverse   688
```

```
nrow(rev_matches) + nrow(fwd_matches) == nrow(seqs.df)
```

```
## [1] TRUE
```

Now, if a read is reversed, we want to replace it with its reverse complement

```
seqs.df <-
     mutate(seqs.df,
            seq = ifelse(type == 'reverse',
                         yes = seq |>
                              DNAStringSet() |>
                              reverseComplement() |>
                              as.character(),
                         no = seq)) |>
     select(-c(type, group))
```

Clean results

- Sequences that are the same and that come from the same species can be de-duplicated
- Sequences that are different and come from the same species must be preserved

```
seqs.df |>
     group_by(taxon, seq)
```

```
## # A tibble: 7,671 x 16
## # Groups:   taxon, seq [1,844]
##    source accession   taxon   taxid superkingdom phylum class order family genus
##    <chr>  <chr>       <chr>   <int> <chr>        <chr>  <chr> <chr> <chr>  <chr>
##  1 <NA>   LR031876.1  Brass~ 3.71e3 Eukaryota    Strep~ Magn~ Bras~ Brass~ Bras~
##  2 <NA>   LR031875.1  Brass~ 3.71e3 Eukaryota    Strep~ Magn~ Bras~ Brass~ Bras~
##  3 <NA>   LR031875.1  Brass~ 3.71e3 Eukaryota    Strep~ Magn~ Bras~ Brass~ Bras~
##  4 <NA>   MN082371.1  Coryl~ 1.35e4 Eukaryota    Strep~ Magn~ Faga~ Betul~ Cory~
##  5 <NA>   LR031874.1  Brass~ 3.71e3 Eukaryota    Strep~ Magn~ Bras~ Brass~ Bras~
##  6 <NA>   HG994366.1  Brass~ 3.71e3 Eukaryota    Strep~ Magn~ Bras~ Brass~ Bras~
##  7 <NA>   HG994366.1  Brass~ 3.71e3 Eukaryota    Strep~ Magn~ Bras~ Brass~ Bras~
##  8 <NA>   NC_039346.1 Sphag~ 1.12e6 Eukaryota    Strep~ Magn~ Aste~ Aster~ Spha~
##  9 <NA>   MZ997428.1  Rheum~ 3.62e3 Eukaryota    Strep~ Magn~ Cary~ Polyg~ Rheum
## 10 <NA>   LT606769.1  Vacci~ 1.91e5 Eukaryota    Strep~ Magn~ Eric~ Erica~ Vacc~
## # i 7,661 more rows
## # i 6 more variables: species <chr>, varietas <chr>, subspecies <chr>,
## #   forma <chr>, seq <chr>, N <lgl>
```

```
dups <-
     seqs.df |>
     group_by(taxon) |>
     summarize(n = sum(duplicated(seq)))

arrange(dups, desc(n))
```

```
## # A tibble: 1,331 x 2
##    taxon                                      n
##    <chr>                                  <int>
##  1 Brassica nigra                           287
##  2 Arbutus unedo                            246
##  3 Morus rubra                               95
##  4 Coffea arabica                            91
##  5 Eleusine indica                           91
##  6 Achillea millefolium subsp. millefolium   86
##  7 Chrysanthemum indicum                     75
##  8 Melilotus albus                           67
##  9 Vitis vinifera                            67
## 10 Theobroma cacao                           65
## # i 1,321 more rows
```

```
sum(dups$n)
```

```
## [1] 5827
```

So the number of sequences we expect after filtering is

```
dim(seqs.df)[1] - sum(dups$n)
```

```
## [1] 1844
```

```
seqs.df <-
     seqs.df |>
     group_by(superkingdom,
              phylum,
              class,
              order,
              family,
              genus,
              species,
              subspecies,
              varietas,
              forma,
              taxon,
              seq) |>
     arrange(desc(source), accession) |> # Puts RefSeq accessions first
```

```r
    summarize(accession = first(accession)) # Choose the first accession number
```

```
## `summarise()` has grouped output by 'superkingdom', 'phylum', 'class', 'order',
## 'family', 'genus', 'species', 'subspecies', 'varietas', 'forma', 'taxon'. You
## can override using the `.groups` argument.
```

```r
dim(seqs.df)
```

```
## [1] 1870   13
```

Save

DADA2

```r
# Sort alphabetically (first by species name, and then accession number)
seqs.df <- arrange(seqs.df,
                   taxon,
                   accession)

# Convert back to DNAStringSet object
trnL <- seqs.df$seq
names(trnL) <- paste(seqs.df$accession, seqs.df$taxon)

trnL <- DNAStringSet(trnL)
trnL
```

```
## DNAStringSet object of length 1870:
##        width seq                                              names
##    [1]   113 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAATGG NC_035234.1 Abelm...
##    [2]   113 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAATGG NC_053353.1 Abelm...
##    [3]   106 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAATGG HQ696727.1 Abutil...
##    [4]   113 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAATGG NC_053702.1 Abuti...
##    [5]    96 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAACGG EU440012.1 Acacia...
##    ...   ... ...
## [1866]    91 GGGCAATCCTGAGCCAAATCCG...ATAGGTGCAGAGACTCAATGG NC_066144.1 Zizan...
## [1867]    80 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAATGG KR083150.1 Ziziph...
## [1868]    89 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAATGG NC_030299.1 Zizip...
## [1869]    89 GGGCAATCCTGAGCCAAATCCT...ATAGGTGCAGAGACTCAATGG NC_037151.1 Zizip...
## [1870]    91 GGGCAATCCTGAGCCAAATCCC...ATAGGTGCAGAGACTCAATGG MN871703.1 [Penni...
```

```r
# Save to file
writeXStringSet(trnL,
            here('data',
                 'outputs',
                 'dada2-compatible',
                 'trnL',
                 'trnLGH.fasta'))
```

```r
seqs.df <-
     seqs.df |>
     unite(col = 'name',
           superkingdom:forma,
           sep = ';')

names(trnL) <- seqs.df$name
```

```
writeXStringSet(trnL,
            here('data',
                 'outputs',
                 'dada2-compatible',
                 'trnL',
                 'trnLGH_taxonomy.fasta'))
```

QIIME2

Taxonomic assignment in QIIME2 with `classify-consensus-vsearch` requires two files: `- i-reference-reads`, a QIIME artifact containing a FASTA file of sequences identified by their accession `- i-reference-taxonomy`, a QIIME artifact containing a TSV (tab-separated value) file with two columns: "Feature ID", the same accessions, and "Taxon", the taxonomic lineage with ranks separated by semicolons

```
outdir <- here('data',
               'outputs',
               'qiime2-compatible')

Sys.setenv(QIIME2 = outdir)
```

Note that QIIME2 requires that the FASTA file have unique accessions, so make these unique before proceeding

```
seqs.df <- mutate(seqs.df,
                  accession = make.unique(accession,
                                          sep = '_'))

names(trnL) <- seqs.df$acc
```

```
## Warning: Unknown or uninitialised column: `acc`.
```

```
writeXStringSet(trnL,
            file.path(outdir,
            'trnL-sequences.fasta'))
```

```
seqs.df |>
 select(`Feature ID` = accession,
        Taxon = name) |>
 write_delim(delim = '\t',
             file.path(outdir,
                       'trnL-taxonomy.tsv'))
```

```
## Adding missing grouping variables: `taxon`
```

```
# Running this block requires a QIIME2 conda environment on local machine
# Note that yours may differ from the "qiime2-2022.8"; you can find correct name by running conda env l
# Convert both of the above objects to QIIME2 artifacts (QZA)
cd "$QIIME2"
conda activate qiime2-2022.8

qiime tools import \
     --input-path trnL-sequences.fasta \
     --output-path trnL-sequences.qza \
     --type 'FeatureData[Sequence]'

qiime tools import \
     --input-path trnL-taxonomy.tsv \
     --output-path trnL-taxonomy.qza \
     --type 'FeatureData[Taxonomy]'
```