markdown

2022-04-20

First, I will load in the packages (tidyverse, lemon, knitr). This code is in the script but not knit to the markdown document for readability.

I also read in the “DS-KROK4BDJ.txt” file, which is a tsv downloaded from our paper of choice and (I believe) contains all the data we will need to work with, including sample IDs, geographic information, and sequences.

## Rows: 124 Columns: 80  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: "\t"  
## chr (41): processid, sampleid, catalognum, institution\_storing, bin\_uri, phy...  
## dbl (14): recordID, phylum\_taxID, class\_taxID, order\_taxID, family\_taxID, su...  
## lgl (25): fieldnum, collection\_code, subspecies\_taxID, subspecies\_name, tax\_...  
##   
## 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.

After reading in the file, I inspect the column names. There are 80 columns in this file.

print(column\_names)

## [1] "processid" "sampleid"   
## [3] "recordID" "catalognum"   
## [5] "fieldnum" "institution\_storing"   
## [7] "collection\_code" "bin\_uri"   
## [9] "phylum\_taxID" "phylum\_name"   
## [11] "class\_taxID" "class\_name"   
## [13] "order\_taxID" "order\_name"   
## [15] "family\_taxID" "family\_name"   
## [17] "subfamily\_taxID" "subfamily\_name"   
## [19] "genus\_taxID" "genus\_name"   
## [21] "species\_taxID" "species\_name"   
## [23] "subspecies\_taxID" "subspecies\_name"   
## [25] "identification\_provided\_by" "identification\_method"   
## [27] "identification\_reference" "tax\_note"   
## [29] "voucher\_status" "tissue\_type"   
## [31] "collection\_event\_id" "collectors"   
## [33] "collectiondate\_start" "collectiondate\_end"   
## [35] "collectiontime" "collection\_note"   
## [37] "site\_code" "sampling\_protocol"   
## [39] "lifestage" "sex"   
## [41] "reproduction" "habitat"   
## [43] "associated\_specimens" "associated\_taxa"   
## [45] "extrainfo" "notes"   
## [47] "lat" "lon"   
## [49] "coord\_source" "coord\_accuracy"   
## [51] "elev" "depth"   
## [53] "elev\_accuracy" "depth\_accuracy"   
## [55] "country" "province\_state"   
## [57] "region" "sector"   
## [59] "exactsite" "image\_ids"   
## [61] "image\_urls" "media\_descriptors"   
## [63] "captions" "copyright\_holders"   
## [65] "copyright\_years" "copyright\_licenses"   
## [67] "copyright\_institutions" "photographers"   
## [69] "sequenceID" "markercode"   
## [71] "genbank\_accession" "nucleotides"   
## [73] "trace\_ids" "trace\_names"   
## [75] "trace\_links" "run\_dates"   
## [77] "sequencing\_centers" "directions"   
## [79] "seq\_primers" "marker\_codes"

Let’s look at the first 5 rows in this file. It’s a large chunk, but very interesting. Scroll side to side to see the whole chunk.

kable(bdj[1:5, ])

| processid | sampleid | recordID | catalognum | fieldnum | institution\_storing | collection\_code | bin\_uri | phylum\_taxID | phylum\_name | class\_taxID | class\_name | order\_taxID | order\_name | family\_taxID | family\_name | subfamily\_taxID | subfamily\_name | genus\_taxID | genus\_name | species\_taxID | species\_name | subspecies\_taxID | subspecies\_name | identification\_provided\_by | identification\_method | identification\_reference | tax\_note | voucher\_status | tissue\_type | collection\_event\_id | collectors | collectiondate\_start | collectiondate\_end | collectiontime | collection\_note | site\_code | sampling\_protocol | lifestage | sex | reproduction | habitat | associated\_specimens | associated\_taxa | extrainfo | notes | lat | lon | coord\_source | coord\_accuracy | elev | depth | elev\_accuracy | depth\_accuracy | country | province\_state | region | sector | exactsite | image\_ids | image\_urls | media\_descriptors | captions | copyright\_holders | copyright\_years | copyright\_licenses | copyright\_institutions | photographers | sequenceID | markercode | genbank\_accession | nucleotides | trace\_ids | trace\_names | trace\_links | run\_dates | sequencing\_centers | directions | seq\_primers | marker\_codes |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| KROK004-19 | CLPT-004 | 10924772 | CLPT-004 | NA | University of Ljubljana, Biotechnical Faculty | NA | BOLD:ABA5411 | 20 | Arthropoda | 82 | Insecta | 413 | Coleoptera | 948 | Carabidae | 2933 | Harpalinae | 403088 | Licinus | 403089 | Licinus hoffmannseggii | NA | NA | Urska Ratajc | Stereomicroscope | NA | NA | NA | NA | NA | Zan Kuralt, Neza Pajek Arambasic, Maja Ferle, Franc Kljun | NA | NA | NA | NA | NA | pitfall traps | adult | F | NA | Dinaric beech forest | NA | NA | NA | NA | 45.539 | 14.760 | NA | NA | 1108 | NA | NA | NA | Slovenia | NA | Kocevska | NA | Borovec mountain | NA | NA | NA | NA | NA | NA | NA | NA | NA | 12447477 | COI-5P | MT994111 | AACTTTATATTTCATTTTCGGGGCCTGAGCAGGAATAGTAGGTACTTCTTTAAGTATATTAATTCGAGCTGAATTAGGAAATCCAGGATCACTTATTGGTGATGATCAGATTTATAATGTTATTGTTACAGCTCATGCATTTGTAATAATTTTTTTCATAGTAATACCTATTATAATTGGGGGATTCGGAAACTGATTAGTTCCATTAATATTAGGAGCTCCTGATATAGCTTTCCCTCGAATAAATAATATAAGTTTTTGACTTCTTCCTCCTGCTTTAAGCCTACTTTTAATGAGAAGAGTAGTTGAAAGAGGAGCTGGCACCGGATGAACGGTGTACCCCCCCCTATCATCTAATATTGCCCATAGAGGTGCTTCCGTTGACTTAGCAATTTTCAGATTACACTTAGCGGGAGTATCTTCCATTTTAGGAGCTGTAAATTTTATTACCACTATTATTAATATACGATCAATAGGAATAACATTTGACCGAATACCATTATTTGTATGATCAGTAGGAATTACTGCTTTACTATTACTTTTATCATTACCAGTATTGGCTGGAGCTATTACAATACTTTTAACAGATCGAAATTTAAATACTTCATTTTTTGATCCTGCAGGAGGGGGAGACCCAATTCTTTATCAACATTTATTT | 11609303|11609302 | CLPT-004\_LCO1490.ab1|CLPT-004\_HCO2198.ab1 | <http://trace.boldsystems.org/traceIO/bold.org/13428409&#124;http://trace.boldsystems.org/traceIO/bold.org/13428408> | 2019-10-25 10:24:30|2019-10-25 12:37:54 | Macrogen, Europe|Macrogen, Europe | F|R | LCO1490|HCO2198 | COI-5P|COI-5P |
| KROK005-19 | CLPT-005 | 10924773 | CLPT-005 | NA | University of Ljubljana, Biotechnical Faculty | NA | BOLD:AAX7561 | 20 | Arthropoda | 82 | Insecta | 413 | Coleoptera | 948 | Carabidae | 2933 | Harpalinae | 2920 | Pterostichus | 373903 | Pterostichus burmeisteri | NA | NA | Urska Ratajc | Stereomicroscope | NA | NA | NA | NA | NA | Zan Kuralt, Neza Pajek Arambasic, Maja Ferle, Franc Kljun | NA | NA | NA | NA | NA | pitfall traps | adult | F | NA | Dinaric beech forest | NA | NA | NA | NA | 45.539 | 14.760 | NA | NA | 1108 | NA | NA | NA | Slovenia | NA | Kocevska | NA | Borovec mountain | NA | NA | NA | NA | NA | NA | NA | NA | NA | 12447478 | COI-5P | MT994134 | AACTTTATATTTTATTTTTGGTATATGAGCAGGAATAGTAGGTACTTCATTAAGTATATTAATTCGAGCTGAATTAGGTAATCCTGGATCATTAATTGGTGACGATCAAATTTATAATGTTATTGTTACTGCCCATGCATTTGTTATAATTTTCTTTATAGTAATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTTTAATATTAGGAGCTCCTGATATAGCTTTTCCTCGAATAAATAATATAAGTTTTTGATTACTTCCTCCTTCATTAACACTCCTTTTAATAAGAAGTATAGTTGAAAACGGATCAGGAACAGGATGAACAGTTTATCCACCTTTATCATCAGGAATTGCACATGCAGGAGCTTCAGTTGATTTAGCTATTTTTAGTTTACATTTAGCTGGAATTTCCTCTATTTTAGGAGCTGTAAATTTTATTACTACAATTATTAATATACGATCAGTAGGAATAACATTTGATCGAATACCTTTATTTGTATGATCTGTTGGAATTACTGCTTTATTATTACTTCTTTCATTACCTGTATTAGCTGGTGCCATTACAATACTATTAACAGATCGAAATTTAAATACTTCTTTTTTTGATCCTGCAGGTGGTGGAGACCCTGTTTTATATCAACATTTATTT | 11609304|11609305 | CLPT-005\_HCO2198.ab1|CLPT-005\_LCO1490.ab1 | <http://trace.boldsystems.org/traceIO/bold.org/13428410&#124;http://trace.boldsystems.org/traceIO/bold.org/13428411> | 2019-10-25 12:37:54|2019-10-25 10:24:30 | Macrogen, Europe|Macrogen, Europe | R|F | HCO2198|LCO1490 | COI-5P|COI-5P |
| KROK008-19 | CLPT-015 | 10924776 | CLPT-015 | NA | University of Ljubljana, Biotechnical Faculty | NA | BOLD:AAN9386 | 20 | Arthropoda | 82 | Insecta | 413 | Coleoptera | 948 | Carabidae | 2933 | Harpalinae | 226200 | Abax | 226201 | Abax ovalis | NA | NA | Urska Ratajc | Stereomicroscope | NA | NA | NA | NA | NA | Zan Kuralt, Neza Pajek Arambasic, Maja Ferle, Franc Kljun | NA | NA | NA | NA | NA | pitfall traps | adult | F | NA | Dinaric beech forest | NA | NA | NA | NA | 45.539 | 14.764 | NA | NA | 1113 | NA | NA | NA | Slovenia | NA | Kocevska | NA | Krokar virgin forest | NA | NA | NA | NA | NA | NA | NA | NA | NA | 12447481 | COI-5P | MT994068 | AAGATATTGGAACACTATATTTTATTTTTGGTACATGATCAGGAATAGTAGGAACATCTTTAAGAATATTAATTCGAGCTGAATTAGGAAATCCAGGATCATTAATTGGTGACGACCAAATTTATAATGTAATTGTTACTGCTCATGCATTTGTAATAATTTTTTTTATAGTAATGCCCATTATAATTGGGGGATTTGGAAACTGATTAGTTCCTTTAATATTAGGAGCTCCTGATATAGCGTTCCCTCGAATGAATAACATGAGATTTTGACTGTTACCCCCTTCTCTGACCCTTCTCCTTATAAGCAGCATAGTTGAAAGAGGAACTGGCACAGGATGGACAGTCTACCCACCACTTTCTTCTAATATTGCCCATAGAGGAGCTTCAGTTGACTTAGCTATTTTTAGTTTACATCTAGCTGGAATTTCTTCAATTCTAGGAGCTGTAAATTTTATCACTACAATTATCAATATACGATCAATAGGAATAACTTTTGATCGAATACCTCTATTTGTCTGATCAGTAGGAATCACTGCTTTATTATTATTGTTGTCACTACCTGTATTAGCTGGAGCTATCACAATACTTTTAACTGATCGAAATTTAAATACTTCATTTTTTGATCCAGCAGGAGGTGGAGATCCTATTTTATA | 11609310|11609311 | CLPT-015\_HCO2198.ab1|CLPT-015\_LCO1490.ab1 | <http://trace.boldsystems.org/traceIO/bold.org/13428416&#124;http://trace.boldsystems.org/traceIO/bold.org/13428417> | 2019-10-25 12:37:54|2019-10-25 10:24:30 | Macrogen, Europe|Macrogen, Europe | R|F | HCO2198|LCO1490 | COI-5P|COI-5P |
| KROK010-19 | CLPT-017 | 10924778 | CLPT-017 | NA | University of Ljubljana, Biotechnical Faculty | NA | BOLD:AAO0964 | 20 | Arthropoda | 82 | Insecta | 413 | Coleoptera | 948 | Carabidae | 312806 | Nebriinae | 90118 | Notiophilus | 167827 | Notiophilus biguttatus | NA | NA | Urska Ratajc | Stereomicroscope | Fabricius | NA | NA | NA | NA | Zan Kuralt, Neza Pajek Arambasic, Maja Ferle, Franc Kljun | NA | NA | NA | NA | NA | leaf litter sifting | adult | NA | NA | Dinaric beech forest | NA | NA | NA | NA | 45.539 | 14.765 | NA | NA | 1134 | NA | NA | NA | Slovenia | NA | Kocevska | NA | Krokar virgin forest | NA | NA | NA | NA | NA | NA | NA | NA | NA | 12447483 | COI-5P | MT994131 | AACTTTATATTTTATCTTTGGGGCTTGGTCAGGGATAGTTGGAACTTCTTTAAGAATACTAATTCGAGCAGAATTAGGAAATCCTGGATCTTTAATTGGAGATGACCAAATTTATAATGTAATTGTTACAGCTCATGCTTTTGTCATAATTTTTTTTATAGTTATACCTATTATAATTGGAGGATTTGGTAATTGACTTGTACCTCTAATATTAGGAGCTCCTGATATAGCTTTTCCTCGAATAAATAATATGAGTTTCTGGTTACTTCCCCCCTCTTTAACTCTCCTCCTTACAAGAAGAATAGTGGAAAGAGGAGCAGGTACAGGTTGAACTGTGTACCCTCCCCTTTCTTCTGGAATTGCCCATAGAGGAGCTTCAGTTGATTTAGCAATTTTTAGCCTTCATTTAGCTGGAGTATCTTCTATTTTAGGAGCAGTAAATTTTATTACTACTATTATTAATATACGATCAGTTGGGATATCATTTGATCGAATACCCTTATTTGTGTGATCAGTTGGAATTACAGCTCTTTTATTACTTTTATCCTTACCTGTTTTAGCTGGAGCAATTACTATATTATTAACTGATCGAAACTTAAATACTTCTTTCTTTGACCCTGCTGGGGGAGGAGATCCTATTTTATATCAACACCTATAT | 11609314|11609315 | CLPT-017\_HCO2198.ab1|CLPT-017\_LCO1490.ab1 | <http://trace.boldsystems.org/traceIO/bold.org/13428420&#124;http://trace.boldsystems.org/traceIO/bold.org/13428421> | 2019-10-25 12:37:54|2019-10-25 10:24:30 | Macrogen, Europe|Macrogen, Europe | R|F | HCO2198|LCO1490 | COI-5P|COI-5P |
| KROK011-19 | CLPT-018 | 10924779 | CLPT-018 | NA | University of Ljubljana, Biotechnical Faculty | NA | BOLD:ACR9475 | 20 | Arthropoda | 82 | Insecta | 413 | Coleoptera | 948 | Carabidae | 89439 | Carabinae | 3019 | Carabus | 660646 | Carabus creutzeri | NA | NA | Urska Ratajc | Stereomicroscope | NA | NA | NA | NA | NA | Zan Kuralt, Neza Pajek Arambasic, Maja Ferle, Franc Kljun | NA | NA | NA | NA | NA | pitfall traps | adult | F | NA | Dinaric beech forest | NA | NA | NA | NA | 45.539 | 14.765 | NA | NA | 1134 | NA | NA | NA | Slovenia | NA | Kocevska | NA | Krokar virgin forest | NA | NA | NA | NA | NA | NA | NA | NA | NA | 12447484 | COI-5P | MT994074 | AACTTTATATTTTATTTTTGGTGCTTGATCAGGAATAGTGGGAACTTCTTTAAGAATACTAATTCGAGCTGAATTAGGTAACCCTGGATCCTTAATTGGAGATGATCAAATTTATAATGTTATTGTGACAGCTCATGCTTTTGTAATAATTTTTTTTATAGTTATACCTATTATAATTGGAGGATTTGGAAATTGATTAGTTCCTTTAATATTAGGAGCACCTGATATAGCATTTCCTCGAATAAATAATATAAGATTTTGATTATTACCCCCTTCTTTAACTCTACTCCTAATAAGTTCAATAGTAGAAAGTGGAGCAGGTACTGGGTGAACAGTATACCCCCCTTTGTCATCTGGAATTGCCCATAGAGGTGCTTCTGTAGATTTAGCTATTTTTAGATTACATCTAGCTGGGATTTCTTCAATTCTAGGGGCAGTAAATTTTATTACAACAATTATTAATATACGATCAGTAGGAATAACCTTTGATCGAATACCTTTATTTGTTTGATCTGTAGGTATTACAGCTTTATTACTTTTACTTTCTCTCCCAGTATTAGCAGGAGCTATTACTATACTTCTAACAGACCGTAATTTAAATACTTCTTTTTTCGATCCAGCTGGGGGAGGTGACCCAATTTTATACCAACATTTATTT | 11609316|11609317 | CLPT-018\_HCO2198.ab1|CLPT-018\_LCO1490.ab1 | <http://trace.boldsystems.org/traceIO/bold.org/13428422&#124;http://trace.boldsystems.org/traceIO/bold.org/13428423> | 2019-10-25 12:37:54|2019-10-25 10:24:30 | Macrogen, Europe|Macrogen, Europe | R|F | HCO2198|LCO1490 | COI-5P|COI-5P |