

Running wsl commands using system2() function in R

Dan Li

2020-08-12

Using “wsl” command in system2() to run commands in wsl

```
system2("wsl", "cd ~/bioinfor/; ls", stdout = TRUE)
```

```
## [1] "AF086833.gb"          "NC_045512-version1.fa" "RNASeqByExample"
## [4] "chr22.fa"             "runinfo.csv"
```

We can retrieve the SARS-coronavirus 2 gene sequences using efetch

```
system2("wsl", "efetch -db=nuccore -format=gb -id=NC_045512", stdout = "../../NC_045512.gb")
```

Accession number NC_045512 in Fasta format.

```
system2("wsl", "efetch -db=nuccore -format=fasta -id=NC_045512 > NC_045512.fasta", stdout = TRUE)
```

```
## character(0)
```

```
system2("wsl", "cat ../NC_045512.gb | head", stdout = TRUE)
```

```
## [1] "LOCUS      NC_045512      29903 bp ss-RNA      linear      VRL 18-JUL-2020"
## [2] "DEFINITION Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,"
## [3] "          complete genome."
## [4] "ACCESSION  NC_045512"
## [5] "VERSION   NC_045512.2"
## [6] "DBLINK     BioProject: PRJNA485481"
## [7] "KEYWORDS   RefSeq."
## [8] "SOURCE     Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)"
## [9] "  ORGANISM Severe acute respiratory syndrome coronavirus 2"
## [10] "          Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;"
```

- GO annotations for the human genes that may be relevant for SARS-CoV-2 infection

```
system2("wsl", "wget http://geneontology.org/data/sars-cov-2_targets.gaf")
```

- GO annotations for the SARS-CoV-2 virus genes

```
system2("wsl", "wget http://geneontology.org/data/uniprot_sars-cov-2.gpa")
```

```
system2("wsl", "cat sars-cov-2_targets.gaf | head -20", stdout = TRUE)
```

```
## [1] "!gaf-version: 2.1"
## [2] "!"
## [3] "!Generated by GoExporter package"
## [4] "!"
## [5] "!Created on: Tue Apr 07 2020 09:52:20 GMT+0000 (Coordinated Universal Time)"
```

```
## [6] "!"
## [7] "!Contact Email: laurent.albou@lbl.gov"
## [8] "!"
## [9] "!Documentation about this header can be found here: http://geneontology.org/docs/go-annotation"
## [10] "!"
## [11] "UniProtKB\tP50897\tPPT1\t\t\tGO:0002084\tGO_REF:0000024\tISS\tUniProtKB:P45478\tP\tPalmitoyl-pro
## [12] "UniProtKB\tP50897\tPPT1\t\t\tGO:0002084\tPMID:10658183\tIDA\t\tP\tPalmitoyl-protein thioesterase
## [13] "UniProtKB\tP50897\tPPT1\t\t\tGO:0002084\tPMID:10737604\tIDA\t\tP\tPalmitoyl-protein thioesterase
## [14] "UniProtKB\tP50897\tPPT1\t\t\tGO:0005515\tPMID:17237713\tIPI\tUniProtKB:O14773\tF\tPalmitoyl-prot
## [15] "UniProtKB\tP50897\tPPT1\t\t\tGO:0005515\tPMID:19941651\tIPI\tUniProtKB:Q3UMW8\tF\tPalmitoyl-prot
## [16] "UniProtKB\tP50897\tPPT1\t\t\tGO:0005515\tPMID:25544563\tIPI\tUniProtKB:Q76RH3\tF\tPalmitoyl-prot
## [17] "UniProtKB\tP50897\tPPT1\t\t\tGO:0005576\tGO_REF:0000024\tISS\tUniProtKB:P45478\tC\tPalmitoyl-pro
## [18] "UniProtKB\tP50897\tPPT1\t\t\tGO:0005576\tPMID:8895569\tIDA\t\tC\tPalmitoyl-protein thioesterase
## [19] "UniProtKB\tP50897\tPPT1\t\t\tGO:0005623\tGO_REF:0000108\tIEA\tGO:0007042\tC\tPalmitoyl-protein th
## [20] "UniProtKB\tP50897\tPPT1\t\t\tGO:0005634\tPMID:10992246\tIDA\t\tC\tPalmitoyl-protein thioesterase

system2("wsl", "cat uniprot_sars-cov-2.gpa | head -30", stdout = TRUE)
```

```
## [1] "!gpa-version: 1.1"
## [2] "!Columns:"
## [3] "!"
## [4] "!" name required? cardinality GAF column #"
## [5] "!" DB required 1 1"
## [6] "!" DB_Object_ID required 1 2 / 17"
## [7] "!" Qualifier required 1 or greater 4"
## [8] "!" GO ID required 1 5"
## [9] "!" DB:Reference(s) required 1 or greater 6"
## [10] "!" ECO evidence code required 1 7 + 6 (GO evidence code + reference)"
## [11] "!" With optional 0 or greater 8"
## [12] "!" Interacting taxon ID optional 0 or 1 13"
## [13] "!" Date required 1 14"
## [14] "!" Assigned_by required 1 15"
## [15] "!" Annotation Extension optional 0 or greater 16"
## [16] "!" Annotation Properties optional 0 or 1 n/a"
## [17] "!"
## [18] "!Generated: 2020-03-24 16:33"
## [19] "!GO-version: http://purl.obolibrary.org/obo/go/releases/2020-03-22/extensions/go-plus.owl"
## [20] "!"
## [21] "UniProtKB\tP0DTC1\ttenables\tGO:0003723\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0694\t\t2
## [22] "UniProtKB\tP0DTC1\ttenables\tGO:0004518\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0540\t\t2
## [23] "UniProtKB\tP0DTC1\ttenables\tGO:0004519\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0255\t\t2
## [24] "UniProtKB\tP0DTC1\ttenables\tGO:0008233\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0645\t\t2
## [25] "UniProtKB\tP0DTC1\ttenables\tGO:0008234\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0788\t\t2
## [26] "UniProtKB\tP0DTC1\ttenables\tGO:0016787\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0378\t\t2
## [27] "UniProtKB\tP0DTC1\ttenables\tGO:0036459\tGO_REF:0000003\tECO:0000501\tEC:3.4.19.12\t\t20200321\t\t2
## [28] "UniProtKB\tP0DTC1\ttenables\tGO:0046872\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0479\t\t2
## [29] "UniProtKB\tP0DTC1\tinvolved_in\tGO:0006508\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0645\t\t2
## [30] "UniProtKB\tP0DTC1\tinvolved_in\tGO:0016032\tGO_REF:0000043\tECO:0000322\tUniProtKB-KW:KW-0945\t\t2
```

GOLGA2 is the most highly annotated proteins in the GO dataset.

```
system2("wsl", "cat sars-cov-2_targets.gaf | cut -f 3 | sort | uniq -c | sort -r | head", stdout = TRUE)

## [1] " 437 GOLGA2" " 270 RHOA" " 260 ITGB1" " 217 RIPK1"
## [5] " 211 RBX1" " 187 NUP62" " 171 TBK1" " 156 PRKACA"
## [9] " 155 HDAC2" " 153 TLE5"
```