Lab Worksheet 11 Solutions

Problem 1:

(a) Download the Medline record for the publication with pubmed id 25502413 and parse it with the `Medline.parse()` function. Then print a list of all key-value pairs returned in that record.

(b) Use an Entrez esearch query of the pubmed database to find out how many publications "Meyer AG" wrote in 2014.

(c) From the results of part (b), compile a list of all the publication titles of "Meyer AG" in 2014.

In [49]:
```
# Problem 1a

from Bio import Medline
handle = Entrez.efetch(db="pubmed", id='25502413', rettype="medline", retmode="
text")
records = Medline.parse(handle) ## Hint
record = list(records)[0] ## Hint
handle.close()

for key in record.keys():
    print(key + ":", record[key])
```

```
PMID: 25502413
OWN: NLM
STAT: MEDLINE
DA: 20141216
DCOM: 20151001
LR: 20170220
IS: 1932-6203 (Electronic) 1932-6203 (Linking)
VI: 9
IP: 12
DP: 2014
TI: Predicting growth conditions from internal metabolic fluxes in an in-silico
model of E. coli.
PG: e114608
LID: 10.1371/journal.pone.0114608 [doi]
AB: A widely studied problem in systems biology is to predict bacterial phenoty
pe from growth conditions, using mechanistic models such as flux balance analys
is (FBA). However, the inverse prediction of growth conditions from phenotype i
s rarely considered. Here we develop a computational framework to carry out thi
s inverse prediction on a computational model of bacterial metabolism. We use F
BA to calculate bacterial phenotypes from growth conditions in E. coli, and the
n we assess how accurately we can predict the original growth conditions from t
he phenotypes. Prediction is carried out via regularized multinomial regression
. Our analysis provides several important physiological and statistical insight
s. First, we show that by analyzing metabolic end products we can consistently
predict growth conditions. Second, prediction is reliable even in the presence
of small amounts of impurities. Third, flux through a relatively small number o
f reactions per growth source ( approximately 10) is sufficient for accurate pr
ediction. Fourth, combining the predictions from two separate models, one train
ed only on carbon sources and one only on nitrogen sources, performs better tha
n models trained to perform joint prediction. Finally, that separate prediction
s perform better than a more sophisticated joint prediction scheme suggests tha
t carbon and nitrogen utilization pathways, despite jointly affecting cellular
growth, may be fairly decoupled in terms of their dependence on specific assort
ments of molecular precursors.
FAU: ['Sridhara, Viswanadham', 'Meyer, Austin G', 'Rai, Piyush', 'Barrick, Jeff
rey E', 'Ravikumar, Pradeep', 'Segre, Daniel', 'Wilke, Claus O']
AU: ['Sridhara V', 'Meyer AG', 'Rai P', 'Barrick JE', 'Ravikumar P', 'Segre D',
'Wilke CO']
AD: Center for Computational Biology and Bioinformatics, The University of Texa
s at Austin, Austin, Texas, United States of America. Center for Computational
Biology and Bioinformatics, The University of Texas at Austin, Austin, Texas, U
nited States of America; Institute for Cellular and Molecular Biology, The Univ
ersity of Texas at Austin, Austin, Texas, United States of America. Department
of Computer Science, The University of Texas at Austin, Austin, Texas, United S
tates of America. Center for Computational Biology and Bioinformatics, The Univ
ersity of Texas at Austin, Austin, Texas, United States of America; Institute f
or Cellular and Molecular Biology, The University of Texas at Austin, Austin, T
exas, United States of America; Center for Systems and Synthetic Biology, The U
niversity of Texas at Austin, Austin, Texas, United States of America; Departme
nt of Molecular Biosciences, The University of Texas at Austin, Austin, Texas,
United States of America. Department of Computer Science, The University of Tex
as at Austin, Austin, Texas, United States of America. Department of Biology an
d Bioinformatics Program, Boston University, Boston, Massachusetts, United Stat
es of America. Center for Computational Biology and Bioinformatics, The Univers
ity of Texas at Austin, Austin, Texas, United States of America; Institute for
Cellular and Molecular Biology, The University of Texas at Austin, Austin, Texa
s, United States of America; Center for Systems and Synthetic Biology, The Univ
ersity of Texas at Austin, Austin, Texas, United States of America; Department
of Integrative Biology, The University of Texas at Austin, Austin, Texas, Unite
d States of America.
LA: ['eng']
PT: ['Journal Article', "Research Support, Non-U.S. Gov't"]
DEP: 20141212
```

In [50]:
```
# Problem 1b

handle = Entrez.esearch(db="pubmed",  # database to search
                        term="Meyer AG[Author] AND 2014[Date - Publication]",
# search term
                        retmax=10  # number of results that are returned
                        )
record = Entrez.read(handle)
handle.close()

# search returns PubMed IDs (pmids)
pmid_list = record["IdList"]
print("Publications found:", pmid_list)
print("Number of publications:", len(pmid_list))
```

```
Publications found: ['25502413', '25432719', '25217382', '24878198', '24847981'
, '24624315', '24239457']
Number of publications: 7
```

In [53]:
```
# Problem 1c

from Bio import Medline
handle = Entrez.efetch(db="pubmed", id=pmid_list, rettype="medline", retmode="t
ext")
records = Medline.parse(handle)

title_lst = [] # start with empty list of coauthors
for record in records:
    title = record['TI']
    title_lst.append(title)

handle.close()
print('publication titles of "Meyer AG" in 2014:')
for title in title_lst:
    print(" ", title)
```

```
publication titles of "Meyer AG" in 2014:
  Predicting growth conditions from internal metabolic fluxes in an in-silico m
odel of E. coli.
  Identifying structural variation in haploid microbial genomes from short-read
resequencing data using breseq.
  Predicting evolutionary site variability from structure in viral proteins: bu
riedness, packing, flexibility, and design.
  Naturally occurring polyphenolic inhibitors of amyloid beta aggregation.
  An iterative in silico and modular synthetic approach to aqueous soluble terc
yclic alpha-helix mimetics.
  Analyzing machupo virus-receptor binding by molecular dynamics simulations.
  Alternate splicing of dysferlin C2A confers Ca(2)(+)-dependent and Ca(2)(+)-i
ndependent binding for membrane repair.
```