UCSC, Entrez, KEGG, and Reactome

> **UCSC** genome database is very informative and offers a lot of options that narrow down our search. The genome browser is easy to use and provides good user interface that allows users to look into details by simpily clicking a specific locus on the chromosome.
> **Entrez** allows users to search information in cross databases. It reduces works when users try to compare data in different databases. It allows users to select and sort datas in python, which simplifies data manipulation. However, it stores uncurated data, therefore it could provide undesirable information.
> **KEGG** presents the pathways graphically and also allows users to interact with the graph. It clarifies the interaction among pathways. The pathways are categorized.
> **Reactome** also presents pathways in colorful graphs that makes easier to see. It's easy to zoom in and out.
> I prefer Reactome because of its simplicity, but Entrez seems to be the most complete because it has info in all databases.

In [1]:

```python
from Bio import Entrez
from Bio import SeqIO
Entrez.email = 'jinghuawu@berkeley.edu'
```

In [2]:

```python
orgs = ['Drosophila', 'E.coli', 'human']

glycolysis = ['pyruvate kinase', 'enolase',
              'Phosphoglycerate mutase', 'phosphofructokinase']

TCA = ['malate dehydrogenase', 'citrate synthase',
       'aconitase', 'isocitrate dehydrogenase']

pentose_phosphate = ['glucose 6-phosphate dehydrogenase', 'ribose 5-phospl
                     'transketolase', 'transaldolase']

all_enzymes = glycolysis + TCA + pentose_phosphate
```

Data retrieval

Variables "organism", "ids", "enzyme", "path", "description" and "ntsql" are lists that store the categorized information, which can be converted to tuples later.

In [3]:

```
#gene data
organism = []
ids = []
enzyme = []
path = []

for o in orgs: #for each organism
    for e in all_enzymes:
        if e in glycolysis:
            path.append("glycolysis")
        elif e in TCA:
            path.append("TCA")
        elif e in pentose_phosphate:
            path.append("pentose phosphate pathway")

        organism.append(o)
        enzyme.append(e)
        handle = Entrez.esearch(db='nucleotide', #can be anything on Entre
                                term= o + '[Orgn]' + ' AND ' + e + '[Prot
                                sort='relevance',
                                idtype='acc', #types of record IDs are re
                                retmax=1)
        for i in Entrez.read(handle)['IdList']:
            ids.append(i)
```

In [4]:

```
description = []
ntseq = []
for i in ids:
    handle = Entrez.efetch(db='nucleotide',
                           id=i, #the first one is the most relevant one
                           rettype='gb', #Retrieval type.
                           retmode='text')
    record = SeqIO.read(handle, "gb")
    description.append(str(record.description))
    ntseq.append(str(record.seq[:30]))
```

Turn data into tuples for later use.

In [5]:

```
gene_tuple = []
for n in range(len(ids)):
    gene_tuple.append(tuple([ids[n], description[n], organism[n], enzyme[1
```

Import data to sqlite

In [6]:

```
import sqlite3 #provide interface
conn = sqlite3.connect('my.db') #create a Connection object that represen
c = conn.cursor() #create cursor object for method calls later.
```

Gene table (36 different genes)

In [7]:

```
#gene table
c.execute(
    """CREATE TABLE gene (id INT,
                          name TEXT,
                          description TEXT,
                          organism TEXT,
                          enzyme TEXT,
                          pathway TEXT,
                          ntseq VARCHAR(20));""")
```

Out[7]:

```
<sqlite3.Cursor at 0x104d69880>
```

In [8]:

```
for i in gene_tuple:
    temp = i
    c.execute("""INSERT INTO gene (id, description, organism, enzyme, patl
conn.commit()
```

Pathway table (3 pathways total)

In [9]:

```
c.execute("""CREATE TABLE pathway (name TEXT, description TEXT)""")
conn.commit()
```

In [10]:

```
c.execute("""INSERT INTO pathway (name, description) VALUES ('glycolysis
c.execute("""INSERT INTO pathway (name, description) VALUES ('TCA', 'a me
c.execute("""INSERT INTO pathway (name, description) VALUES ('pentose_pho
conn.commit()
```

enzyme table (12 enzymes total)

In [11]:

```
#enzyme table
c.execute("""CREATE TABLE enzyme (name TEXT, function TEXT, enzyme_commis
conn.commit()
```

In [12]:

```
c.execute("""INSERT INTO enzyme (name, function, enzyme_commission, path
             VALUES  ('pyruvate kinase', 'catalyzes the final step of gly
                     ('enolase', 'metalloenzyme responsible for the catal
                     ('Phosphoglycerate mutase', 'any enzyme that cataly:
                     ('phosphofructokinase', 'a kinase enzyme that phospl
                     ('malate dehydrogenase', 'an enzyme that reversibly
                     ('citrate synthase', 'pace-making enzyme in the firs
                     ('aconitase', 'an enzyme that catalyses the stereo-s
                     ('isocitrate dehydrogenase', 'an enzyme that cataly:
                     ('glucose 6-phosphate dehydrogenase', 'a cytosolic e
                     ('ribose 5-phosphate isomerase', 'catalyzes the conv
                     ('transketolase', 'catalyzes two important reactions
                     ('transaldolase', 'an enzyme (EC 2.2.1.2) of the nor
conn.commit()
```

In [13]:

```
c.execute("SELECT * FROM gene;")
print(c.fetchall())
```

```
[('XM_023317917.1', None, 'PREDICTED: Drosophila hydei pyruvate kinase
(LOC111601379), mRNA', 'Drosophila', 'pyruvate kinase', 'glycolysis',
'TTTCAATACTTAAAAAAAACAAAGTTAATA'), ('XM_023310563.1', None, 'PREDICTED
: Drosophila hydei enolase (LOC111596366), mRNA', 'Drosophila', 'enola
se', 'glycolysis', 'TTATTTTGATATATTCAATTCTTAGTTTA'), ('NT_033777.3',
None, 'Drosophila melanogaster chromosome 3R', 'Drosophila', 'Phosphog
lycerate mutase', 'glycolysis', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('N
T_033778.4', None, 'Drosophila melanogaster chromosome 2R', 'Drosophil
a', 'phosphofructokinase', 'glycolysis', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNN
NN'), ('NT_033779.5', None, 'Drosophila melanogaster chromosome 2L', '
Drosophila', 'malate dehydrogenase', 'TCA', 'NNNNNNNNNNNNNNNNNNNNNNNNN
NNNNN'), ('NC_004354.4', None, 'Drosophila melanogaster chromosome X',
'Drosophila', 'citrate synthase', 'TCA', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNN
NN'), ('NT_033779.5', None, 'Drosophila melanogaster chromosome 2L', '
Drosophila', 'aconitase', 'TCA', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('
LC058675.1', None, 'Drosophila nasuta Idh gene for isocitrate dehydrog
enase, partial cds, strain: G7', 'Drosophila', 'isocitrate dehydrogena
se', 'TCA', 'CTCATTCTGCCATTCCTCGACATTGAGTTG'), ('AY364534.1', None, 'D
rosophila mojavensis strain MJS 71 glucose-6-phosphate dehydrogenase (
G6pd) gene, exon 4 and partial cds', 'Drosophila', 'glucose 6-phosphat
e dehydrogenase', 'pentose phosphate pathway', 'GGCTGGAATCGCGTGATCGTCG
AGAAGCCC'), ('XM_023305862.1', None, 'PREDICTED: Drosophila hydei ribo
se-5-phosphate isomerase (LOC111593212), mRNA', 'Drosophila', 'ribose
5-phosphate isomerase', 'pentose phosphate pathway', 'TTTCAAATAGATGTCA
AATTCACTGTGAAA'), ('NT_033777.3', None, 'Drosophila melanogaster chrom
osome 3R', 'Drosophila', 'transketolase', 'pentose phosphate pathway',
'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('XM_002005838.2', None, 'Drosophil
a mojavensis uncharacterized protein (Dmoj\\GI18849), mRNA', 'Drosophi
la', 'transaldolase', 'pentose phosphate pathway', 'GTCCAGACTAAAGATTTA
GTTGCACCGGGT'), ('NZ_CP009050.1', None, 'Escherichia coli NCCP15648, c
omplete genome', 'E.coli', 'pyruvate kinase', 'glycolysis', 'NNNNNNNNN
NNNNNNNNNNNNNNNNNNNNN'), ('NZ_CP009050.1', None, 'Escherichia coli NCC
P15648, complete genome', 'E.coli', 'enolase', 'glycolysis', 'NNNNNNNN
NNNNNNNNNNNNNNNNNNNNNN'), ('PDAC01000032.1', None, 'Escherichia coli s
```

```
train TVS 353 NODE_32_length_28275_cov_33.7839, whole genome shotgun s
equence', 'E.coli', 'Phosphoglycerate mutase', 'glycolysis', 'AGAAAGGC
AGTCCGCTGCATAAATCTACGC'), ('CP026473.1', None, 'Escherichia coli strai
n KBN10P04869 chromosome, complete genome', 'E.coli', 'phosphofructoki
nase', 'glycolysis', 'CCAGCATGGCGCGCCGGGTGGAGGATTATA'), ('NZ_CP009050.
1', None, 'Escherichia coli NCCP15648, complete genome', 'E.coli', 'ma
late dehydrogenase', 'TCA', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('NZ_PD
AP01000234.1', None, 'Escherichia coli strain 2016C-3325 NODE_234_leng
th_464_cov_0.697329_ID_21090, whole genome shotgun sequence', 'E.coli'
, 'citrate synthase', 'TCA', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('QREF
01000006.1', None, 'Escherichia coli strain 333 Ga0213564_106, whole g
enome shotgun sequence', 'E.coli', 'aconitase', 'TCA', 'CTTGAGACTTGGTA
TTCATTTTTCGTCTTG'), ('NZ_NDCE01000057.1', None, 'Escherichia coli stra
in 39913 39913_NODE_70.ctg_1, whole genome shotgun sequence', 'E.coli'
, 'isocitrate dehydrogenase', 'TCA', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN')
, ('NZ_CP009050.1', None, 'Escherichia coli NCCP15648, complete genome
', 'E.coli', 'glucose 6-phosphate dehydrogenase', 'pentose phosphate p
athway', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('NZ_CP009050.1', None, 'E
scherichia coli NCCP15648, complete genome', 'E.coli', 'ribose 5-phosp
hate isomerase', 'pentose phosphate pathway', 'NNNNNNNNNNNNNNNNNNNNNNN
NNNNNNN'), ('NZ_CP009050.1', None, 'Escherichia coli NCCP15648, comple
te genome', 'E.coli', 'transketolase', 'pentose phosphate pathway', 'N
NNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('NZ_CP009050.1', None, 'Escherichia 
coli NCCP15648, complete genome', 'E.coli', 'transaldolase', 'pentose 
phosphate pathway', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN'), ('FUIG01000043.
1', None, 'Homo sapiens genome assembly, contig: BQ8482_Contig_35, who
le genome shotgun sequence', 'human', 'pyruvate kinase', 'glycolysis',
 'AATTACGCGATCATGACACTAGCACGATGC'), ('FUIG01000070.1', None, 'Homo sapi
ens genome assembly, contig: BQ8482_Contig_6, whole genome shotgun seq
uence', 'human', 'enolase', 'glycolysis', 'GAACTTGACGCACACAACTACAATCAG
TCG'), ('FUIG01000013.1', None, 'Homo sapiens genome assembly, contig:
 BQ8482_Contig_11, whole genome shotgun sequence', 'human', 'Phosphogly
cerate mutase', 'glycolysis', 'GGCCGAGAAGGCGCTGACCGCCGTCATCCA'), ('AH0
02936.2', None, 'Homo sapiens phosphofructokinase (PFKM) gene, partial
 cds', 'human', 'phosphofructokinase', 'glycolysis', 'AGTGGTTCGCACACAGT
GGCTGTGATGAAC'), ('NM_006623.3', None, 'Homo sapiens phosphoglycerate 
dehydrogenase (PHGDH), mRNA', 'human', 'malate dehydrogenase', 'TCA', 
'GCAGGGATTTGGCAACCTCAGAGCCGCGAG'), ('FUIG01000002.1', None, 'Homo sapi
ens genome assembly, contig: BQ8482_Contig_10, whole genome shotgun se
quence', 'human', 'citrate synthase', 'TCA', 'CAGAACTTGACGCACACAACTCGA
GACTGG'), ('AH007467.3', None, 'Homo sapiens chromosome 22 aconitase (
ACO2) gene, complete cds', 'human', 'aconitase', 'TCA', 'GATGGCGGAGATA
ACTAAAATTTGTTCTTG'), ('KU639670.1', None, 'Homo sapiens voucher NGX277 
isocitrate dehydrogenase (IDH2) gene, partial cds', 'human', 'isocitra
te dehydrogenase', 'TCA', 'TCCCAATGGAACTATCCGGAACATCCTGGG'), ('FUIG010
00013.1', None, 'Homo sapiens genome assembly, contig: BQ8482_Contig_1
1, whole genome shotgun sequence', 'human', 'glucose 6-phosphate dehyd
rogenase', 'pentose phosphate pathway', 'GGCCGAGAAGGCGCTGACCGCCGTCATCC
A'), ('NM_144563.2', None, 'Homo sapiens ribose 5-phosphate isomerase 
A (RPIA), mRNA', 'human', 'ribose 5-phosphate isomerase', 'pentose pho
sphate pathway', 'CGGGGGCGGGACTTCAGCGGAGGCCGGAGC'), ('BC009970.2', Non
e, 'Homo sapiens transketolase, mRNA (cDNA clone MGC:15349 IMAGE:43103
96), complete cds', 'human', 'transketolase', 'pentose phosphate pathw
```

```
ay', 'GCCTGTCGCCGCGGGAGCAGCCGCTATCTC'), ('NM_006755.1', None, 'Homo sa
piens transaldolase 1 (TALDO1), mRNA', 'human', 'transaldolase', 'pent
ose phosphate pathway', 'CGCGCCCGTCCCGTCGCCGCCGCCGCCGCC')]
```

```
In [14]: c.execute("SELECT * FROM pathway;")
         print(c.fetchall())
```

```
[('glycolysis', 'a metabolic process that occurs during aerobic and an
aerobic respiration of living organisms within the cytoplasm.'), ('TCA
', 'a metabolic process that occurs during aerobic and anaerobic respi
ration of living organisms within the cytoplasm.'), ('pentose_phosphat
e_pathway', 'hexose monophosphate shunt) is a metabolic pathway parall
el to glycolysis.')]
```

```
In [15]: c.execute("SELECT * FROM enzyme;")
         print(c.fetchall())
```

```
[('pyruvate kinase', 'catalyzes the final step of glycolysis', '2.7.1.
40', 'glycolysis'), ('enolase', 'metalloenzyme responsible for the cat
alysis of the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpy
ruvate (PEP)', '4.2.1.11', 'glycolysis'), ('Phosphoglycerate mutase',
'any enzyme that catalyzes step 8 of glycolysis', '5.4.2.11', 'glycoly
sis'), ('phosphofructokinase', 'a kinase enzyme that phosphorylates fr
uctose 6-phosphate in glycolysis', '2.7.1.11', 'glycolysis'), ('malate
dehydrogenase', 'an enzyme that reversibly catalyzes the oxidation of
malate to oxaloacetate', '1.1.1.37', 'TCA'), ('citrate synthase', 'pac
e-making enzyme in the first step of the citric acid cycle', '2.3.3.1'
, 'TCA'), ('aconitase', 'an enzyme that catalyses the stereo-specific
isomerization of citrate to isocitrate via cis-aconitate in the tricar
boxylic acid cycle', '4.2.1.3', 'TCA'), ('isocitrate dehydrogenase', '
an enzyme that catalyzes the oxidative decarboxylation of isocitrate',
'1.1.1.42', 'TCA'), ('glucose 6-phosphate dehydrogenase', 'a cytosolic
enzyme that catalyzes D-glucose 6-phosphate', '1.1.1.49', 'pentose_pho
sphate_pathway'), ('ribose 5-phosphate isomerase', 'catalyzes the conv
ersion between ribose-5-phosphate (R5P) and ribulose-5-phosphate (Ru5P
)', '5.3.1.6', 'pentose_phosphate_pathway'), ('transketolase', 'cataly
zes two important reactions, which operate in opposite directions in t
hese two pathways', '2.2.1.1', 'pentose_phosphate_pathway'), ('transal
dolase', 'an enzyme (EC 2.2.1.2) of the non-oxidative phase of the pen
tose phosphate pathway', '2.2.1.2', 'pentose_phosphate_pathway')]
```

Associative Table

- gene table x pathway table x enzyme table *

In [16]:
```
c.execute("""
CREATE TABLE assoc AS
SELECT gene.id, gene.name AS gene_name, gene.description AS gene_descript
       pathway.name AS pathway, pathway.description AS pathway_descriptio
       enzyme.name AS enzyme, enzyme.function AS function, enzyme.enzyme_
FROM gene, pathway, enzyme
WHERE gene.pathway == pathway.name AND gene.enzyme == enzyme.name AND pat
conn.commit()
```

In [ ]:

Comments:

1. To avoid crashing the website, I only limited the sequence to 30 nucleotides here.
2. As one of the disadvantages of Entrez, the search results from Entrez are inconsistent, therefore some sequences (especially those in e.coli) do not contain useful information.

1. There are relationships between enzymes and pathways—some enzymes belong to certain pathways. Do any belong to multiple pathways? Is this a one-to-many or many-to-many relationship?
To my knowledge, the enzymes I found here belong to only one pathway, therefore it's a one-to-one relationship. But if there are, it will be a many-to-many relationship because some enzymes might belong to many pathways and a pathway might contain enzymes that are also in its parental table.

2. There is an order to enzymes within pathways. How can the order be represented in a table?
Assigning enzymes a name or index, then sort the table by names or index.

3. Genes in the gene table encode enzymes in the enzyme table. How can this be represented? Is this a one-to-one, one-to-many, or many-to-many relationship, and in which direction?
This can be represented by adding a columns that indicates the pathway(s) the enzyme involved to the end of the gene table, and use boolean statement to filter the rows. Gene to enzyme is a one-to-one relationship because one gene corresponds to one enzyme.

Some manual modifications on the table

In [17]:

```
c.execute("""UPDATE assoc SET gene_name = "AC02" WHERE id == 'AH007467.3
c.execute("""UPDATE assoc SET gene_name = "G6pd" WHERE id == 'AY364534.1
c.execute("""UPDATE assoc SET gene_name = "PFKM" WHERE id == 'AH002936.2
c.execute("""UPDATE assoc SET gene_name = "PHGDH" WHERE id == 'NM_006623
c.execute("""UPDATE assoc SET gene_name = "IDH2" WHERE id == 'KU639670.1
c.execute("""UPDATE assoc SET gene_name = "TALDO1" WHERE id == 'NM_006755
conn.commit()
```

> **Comment**: Due to the inconsistentcy of Entrez search results, not all given sequences are mRNA sequence, thus not all have gene_name.

In [18]:

```
c.execute("SELECT * FROM assoc;")
print(c.fetchall())
```

```
[('XM_023317917.1', None, 'PREDICTED: Drosophila hydei pyruvate kinase
(LOC111601379), mRNA', 'Drosophila', 'TTTCAATACTTAAAAAAAACAAAGTTAATA',
'glycolysis', 'a metabolic process that occurs during aerobic and anae
robic respiration of living organisms within the cytoplasm.', 'pyruvat
e kinase', 'catalyzes the final step of glycolysis', '2.7.1.40'), ('XM
_023310563.1', None, 'PREDICTED: Drosophila hydei enolase (LOC11159636
6), mRNA', 'Drosophila', 'TTATTTTGATATATTCAATTCTTAGTTTA', 'glycolysis
', 'a metabolic process that occurs during aerobic and anaerobic respi
ration of living organisms within the cytoplasm.', 'enolase', 'metallo
enzyme responsible for the catalysis of the conversion of 2-phosphogly
cerate (2-PG) to phosphoenolpyruvate (PEP)', '4.2.1.11'), ('NT_033777.
3', None, 'Drosophila melanogaster chromosome 3R', 'Drosophila', 'NNNN
NNNNNNNNNNNNNNNNNNNNNNNNNN', 'glycolysis', 'a metabolic process that o
ccurs during aerobic and anaerobic respiration of living organisms wit
hin the cytoplasm.', 'Phosphoglycerate mutase', 'any enzyme that catal
yzes step 8 of glycolysis', '5.4.2.11'), ('NT_033778.4', None, 'Drosop
hila melanogaster chromosome 2R', 'Drosophila', 'NNNNNNNNNNNNNNNNNNNNN
NNNNNNNNN', 'glycolysis', 'a metabolic process that occurs during aero
bic and anaerobic respiration of living organisms within the cytoplasm
.', 'phosphofructokinase', 'a kinase enzyme that phosphorylates fructo
se 6-phosphate in glycolysis', '2.7.1.11'), ('NT_033779.5', None, 'Dro
sophila melanogaster chromosome 2L', 'Drosophila', 'NNNNNNNNNNNNNNNNNN
NNNNNNNNNNNN', 'TCA', 'a metabolic process that occurs during aerobic
and anaerobic respiration of living organisms within the cytoplasm.',
'malate dehydrogenase', 'an enzyme that reversibly catalyzes the oxida
tion of malate to oxaloacetate', '1.1.1.37'), ('NC_004354.4', None, 'D
rosophila melanogaster chromosome X', 'Drosophila', 'NNNNNNNNNNNNNNNNN
NNNNNNNNNNNNN', 'TCA', 'a metabolic process that occurs during aerobic
and anaerobic respiration of living organisms within the cytoplasm.',
'citrate synthase', 'pace-making enzyme in the first step of the citri
c acid cycle', '2.3.3.1'), ('NT_033779.5', None, 'Drosophila melanogas
ter chromosome 2L', 'Drosophila', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN', 'T
```

```
CA', 'a metabolic process that occurs during aerobic and anaerobic res
piration of living organisms within the cytoplasm.', 'aconitase', 'an
enzyme that catalyses the stereo-specific isomerization of citrate to
isocitrate via cis-aconitate in the tricarboxylic acid cycle', '4.2.1.
3'), ('LC058675.1', None, 'Drosophila nasuta Idh gene for isocitrate d
ehydrogenase, partial cds, strain: G7', 'Drosophila', 'CTCATTCTGCCATTC
CTCGACATTGAGTTG', 'TCA', 'a metabolic process that occurs during aerob
ic and anaerobic respiration of living organisms within the cytoplasm.
', 'isocitrate dehydrogenase', 'an enzyme that catalyzes the oxidative
decarboxylation of isocitrate', '1.1.1.42'), ('NZ_CP009050.1', None, '
Escherichia coli NCCP15648, complete genome', 'E.coli', 'NNNNNNNNNNNNN
NNNNNNNNNNNNNNNNN', 'glycolysis', 'a metabolic process that occurs dur
ing aerobic and anaerobic respiration of living organisms within the c
ytoplasm.', 'pyruvate kinase', 'catalyzes the final step of glycolysis
', '2.7.1.40'), ('NZ_CP009050.1', None, 'Escherichia coli NCCP15648, c
omplete genome', 'E.coli', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN', 'glycolys
is', 'a metabolic process that occurs during aerobic and anaerobic res
piration of living organisms within the cytoplasm.', 'enolase', 'metal
loenzyme responsible for the catalysis of the conversion of 2-phosphog
lycerate (2-PG) to phosphoenolpyruvate (PEP)', '4.2.1.11'), ('PDAC0100
0032.1', None, 'Escherichia coli strain TVS 353 NODE_32_length_28275_c
ov_33.7839, whole genome shotgun sequence', 'E.coli', 'AGAAAGGCAGTCCGC
TGCATAAATCTACGC', 'glycolysis', 'a metabolic process that occurs durin
g aerobic and anaerobic respiration of living organisms within the cyt
oplasm.', 'Phosphoglycerate mutase', 'any enzyme that catalyzes step 8
of glycolysis', '5.4.2.11'), ('CP026473.1', None, 'Escherichia coli st
rain KBN10P04869 chromosome, complete genome', 'E.coli', 'CCAGCATGGCGC
GCCGGGTGGAGGATTATA', 'glycolysis', 'a metabolic process that occurs du
ring aerobic and anaerobic respiration of living organisms within the
cytoplasm.', 'phosphofructokinase', 'a kinase enzyme that phosphorylat
es fructose 6-phosphate in glycolysis', '2.7.1.11'), ('NZ_CP009050.1',
None, 'Escherichia coli NCCP15648, complete genome', 'E.coli', 'NNNNNN
NNNNNNNNNNNNNNNNNNNNNNNN', 'TCA', 'a metabolic process that occurs dur
ing aerobic and anaerobic respiration of living organisms within the c
ytoplasm.', 'malate dehydrogenase', 'an enzyme that reversibly catalyz
es the oxidation of malate to oxaloacetate', '1.1.1.37'), ('NZ_PDAP010
00234.1', None, 'Escherichia coli strain 2016C-3325 NODE_234_length_46
4_cov_0.697329_ID_21090, whole genome shotgun sequence', 'E.coli', 'NN
NNNNNNNNNNNNNNNNNNNNNNNNNNNN', 'TCA', 'a metabolic process that occurs
during aerobic and anaerobic respiration of living organisms within th
e cytoplasm.', 'citrate synthase', 'pace-making enzyme in the first st
ep of the citric acid cycle', '2.3.3.1'), ('QREF01000006.1', None, 'Es
cherichia coli strain 333 Ga0213564_106, whole genome shotgun sequence
', 'E.coli', 'CTTGAGACTTGGTATTCATTTTTCGTCTTG', 'TCA', 'a metabolic pro
cess that occurs during aerobic and anaerobic respiration of living or
ganisms within the cytoplasm.', 'aconitase', 'an enzyme that catalyses
the stereo-specific isomerization of citrate to isocitrate via cis-aco
nitate in the tricarboxylic acid cycle', '4.2.1.3'), ('NZ_NDCE01000057
.1', None, 'Escherichia coli strain 39913 39913_NODE_70.ctg_1, whole g
enome shotgun sequence', 'E.coli', 'NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN', '
TCA', 'a metabolic process that occurs during aerobic and anaerobic re
spiration of living organisms within the cytoplasm.', 'isocitrate dehy
drogenase', 'an enzyme that catalyzes the oxidative decarboxylation of
```

```
isocitrate', '1.1.1.42'), ('FUIG01000043.1', None, 'Homo sapiens genom
e assembly, contig: BQ8482_Contig_35, whole genome shotgun sequence',
'human', 'AATTACGCGATCATGACACTAGCACGATGC', 'glycolysis', 'a metabolic
process that occurs during aerobic and anaerobic respiration of living
organisms within the cytoplasm.', 'pyruvate kinase', 'catalyzes the fi
nal step of glycolysis', '2.7.1.40'), ('FUIG01000070.1', None, 'Homo s
apiens genome assembly, contig: BQ8482_Contig_6, whole genome shotgun
sequence', 'human', 'GAACTTGACGCACACAACTACAATCAGTCG', 'glycolysis', 'a
metabolic process that occurs during aerobic and anaerobic respiration
of living organisms within the cytoplasm.', 'enolase', 'metalloenzyme
responsible for the catalysis of the conversion of 2-phosphoglycerate
(2-PG) to phosphoenolpyruvate (PEP)', '4.2.1.11'), ('FUIG01000013.1',
None, 'Homo sapiens genome assembly, contig: BQ8482_Contig_11, whole g
enome shotgun sequence', 'human', 'GGCCGAGAAGGCGCTGACCGCCGTCATCCA', 'g
lycolysis', 'a metabolic process that occurs during aerobic and anaero
bic respiration of living organisms within the cytoplasm.', 'Phosphogl
ycerate mutase', 'any enzyme that catalyzes step 8 of glycolysis', '5.
4.2.11'), ('AH002936.2', 'PFKM', 'Homo sapiens phosphofructokinase (PF
KM) gene, partial cds', 'human', 'AGTGGTTCGCACACAGTGGCTGTGATGAAC', 'gl
ycolysis', 'a metabolic process that occurs during aerobic and anaerob
ic respiration of living organisms within the cytoplasm.', 'phosphofru
ctokinase', 'a kinase enzyme that phosphorylates fructose 6-phosphate
in glycolysis', '2.7.1.11'), ('NM_006623.3', 'PHGDH', 'Homo sapiens ph
osphoglycerate dehydrogenase (PHGDH), mRNA', 'human', 'GCAGGGATTTGGCAA
CCTCAGAGCCGCGAG', 'TCA', 'a metabolic process that occurs during aerob
ic and anaerobic respiration of living organisms within the cytoplasm.
', 'malate dehydrogenase', 'an enzyme that reversibly catalyzes the ox
idation of malate to oxaloacetate', '1.1.1.37'), ('FUIG01000002.1', No
ne, 'Homo sapiens genome assembly, contig: BQ8482_Contig_10, whole gen
ome shotgun sequence', 'human', 'CAGAACTTGACGCACACAACTCGAGACTGG', 'TCA
', 'a metabolic process that occurs during aerobic and anaerobic respi
ration of living organisms within the cytoplasm.', 'citrate synthase',
'pace-making enzyme in the first step of the citric acid cycle', '2.3.
3.1'), ('AH007467.3', 'AC02', 'Homo sapiens chromosome 22 aconitase (A
CO2) gene, complete cds', 'human', 'GATGGCGGAGATAACTAAAATTTGTTCTTG', '
TCA', 'a metabolic process that occurs during aerobic and anaerobic re
spiration of living organisms within the cytoplasm.', 'aconitase', 'an
enzyme that catalyses the stereo-specific isomerization of citrate to
isocitrate via cis-aconitate in the tricarboxylic acid cycle', '4.2.1.
3'), ('KU639670.1', 'IDH2', 'Homo sapiens voucher NGX277 isocitrate de
hydrogenase (IDH2) gene, partial cds', 'human', 'TCCCAATGGAACTATCCGGAA
CATCCTGGG', 'TCA', 'a metabolic process that occurs during aerobic and
anaerobic respiration of living organisms within the cytoplasm.', 'iso
citrate dehydrogenase', 'an enzyme that catalyzes the oxidative decarb
oxylation of isocitrate', '1.1.1.42')]
```