

```
In [1]: ecoli = ! all_entities_Genome -f scientific_name | grep "Escherichia coli O104:H4"
sulfo = ! all_entities_Genome -f scientific_name | grep Thermodesulfo | grep -v 'kb|g.

ecoli_set = map( lambda x: x.split('\t'), ecoli[:3] )
sulfo_set = map( lambda x: x.split('\t'), sulfo )

genome_names = slice_column(ecoli_set+sulfo_set, 1)
genome_ids    = slice_column(ecoli_set+sulfo_set, 0)

print genome_ids
print genome_names

['kb|g.79', 'kb|g.78', 'kb|g.83', 'kb|g.3153', 'kb|g.676', 'kb|g.3387']
['Escherichia coli O104:H4 str. H112180282', 'Escherichia coli O104:H4 str. GOS2',
'Escherichia coli O104:H4 str. LB226692', 'Thermodesulfobium narugense DSM 14796',
'Thermodesulfobacterium sp. OPB45']
```

```
In [2]: genome_data = analysis.Analysis(genome_ids, 'genome', level='subsystem')
print genome_data.ids()

[u'kb|g.3153', u'kb|g.3387', u'kb|g.676', u'kb|g.78', u'kb|g.79', u'kb|g.83']
```

```
In [3]: genome_data.annotations()[:10]
```

```
Out[3]: [u'16S rRNA modification within P site of ribosome',
u'2-oxoisovalerate to 2-isopropyl-3-oxosuccinate module',
u'2-phosphoglycolate salvage',
u'271-Bsub',
u'5-FCL-like Experimental',
u'5-FCL-like protein',
u'A Gammaproteobacteria Cluster Relating to Translation',
u'A Gram-positive cluster that relates ribosomal protein L28P to a set of
uncharacterized proteins',
u'A Hypothetical Protein Related to Proline Metabolism',
u'A Hypothetical that Clusters with PEP Synthase']
```

```
In [4]: print "python desnse matrix"
print genome_data.matrix[:10]
print "r dense matrix"
print genome_data.Rmatrix
```

```
python desnse matrix
[[2, 0, 7, 2, 2, 2], [0, 0, 4, 0, 0, 0], [0, 0, 2, 0, 0, 0], [12, 13, 181, 27, 28,
34], [1, 1, 6, 1, 1, 1], [1, 1, 9, 1, 1, 1], [1, 1, 6, 1, 1, 1], [0, 0, 3, 0, 0,
0], [1, 2, 2, 2, 2, 2], [0, 0, 1, 0, 0, 0]]
r dense matrix
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,]      2      0      7      2      2      2
[2,]      0      0      4      0      0      0
[3,]      0      0      2      0      0      0
[4,]     12     13    181     27     28     34
[5,]      1      1      6      1      1      1
[6,]      1      1      9      1      1      1
[7,]      1      1      6      1      1      1
[8,]      0      0      3      0      0      0
[9,]      1      2      2      2      2      2
[10,]     0      0      1      0      0      0
[11,]     0      0      1      1      1      1
[12,]     0      0      3      0      0      0
[13,]     0      0      2      0      0      0
[14,]     0      0      1      0      0      0
[15,]     0      0      1      0      0      0
[16,]     0      0      1      0      0      0
[17,]     0      0      0      1      1      1
[18,]     0      0      3      0      0      0
[19,]     0      0      1      0      0      0
[20,]     0      0      3      0      0      0
[21,]     0      0      5      0      0      0
[22,]     0      0      1      0      0      0
[23,]     0      0      5      0      0      0
[24,]     0      0      2      0      0      0
[25,]     0      0      1      0      0      0
[26,]     0      0      1      0      0      0
[27,]      2      2      3      3      3      3
[28,]      0      1      5      3      3      3
[29,]      0      0      7      0      0      0
[30,]      1      1      6      3      3      3
[31,]      0      0     19      0      0      0
[32,]      0      0      1      0      0      0
[33,]      0      0      1      0      0      0
[34,]      0      0      1      0      0      0
[35,]      0      0     10      0      0      0
[36,]      0      0      1      0      0      0
[37,]      0      0      1      0      0      0
[38,]      0      0      8      0      0      0
[39,]      1      1      2      2      2      2
[40,]      0      0      3      0      0      0
[41,]      0      0      3      1      1      1
[42,]      1      1      2      1      1      1
[43,]      0      0      5      0      0      0
[44,]      0      0      4      1      1      1
[45,]      0      0      5      1      1      1
[46,]      3      2      7      6      6      7
[47,]      0      0      1      0      0      0
[48,]      3      4      4      5      4      4
[49,]      0      0      4      0      0      0
[50,]      0      0     20      1      1      1
```

```
In [5]: apco = genome_data.get_pco(method='bray-curtis', normalize=0)
print apco
```

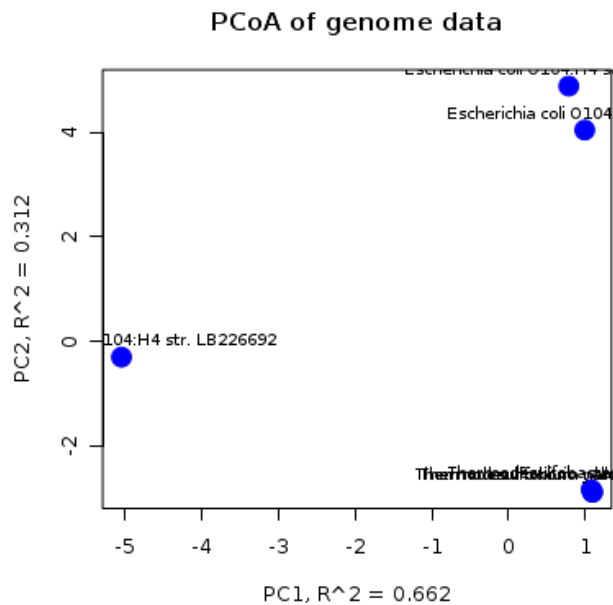
```
$values
      PC01      PC02      PC03      PC04      PC05      PC06
6.6e-01 3.1e-01 2.9e-02 7.6e-04 6.8e-05 0.0e+00

$vectors
      [,1] [,2] [,3] [,4] [,5] [,6]
[1,] -0.767 -0.96 4.84 -0.05 0.079 -0.41
[2,] -0.687 -0.58 -5.67 1.80 -2.852 -0.41
[3,] 1.160 -1.21 -0.27 -0.59 0.240 -0.41
[4,] 0.085 0.92 0.32 -24.63 -95.326 -0.41
[5,] 0.074 0.89 0.22 -13.28 118.725 -0.41
[6,] 0.135 0.94 0.57 36.75 -20.865 -0.41

$distances
      1      2      3      4      5
2 0.203
3 0.797 0.773
4 0.507 0.440 0.600
5 0.500 0.433 0.600 0.013
6 0.524 0.459 0.589 0.036 0.037
```

```
In [6]: afile = genome_data.plot_pco(labels=genome_names, title='PCoA of genome data')
Image(filename=afile)
```

Out[6]:



```
In [7]: afile = genome_data.plot_heatmap(labels=genome_names, title='Heatmap of genome data')
         Image(filename=afile)
```

```
Loading required package: gplots
Loading required package: gtools
```

```
Attaching package: 'gtools'
```

```
The following object(s) are masked from 'package:matR':
```

```
  permutations
```

```
Loading required package: gdata
gdata: read.xls support for 'XLS' (Excel 97-2004) files ENABLED.
```

```
gdata: read.xls support for 'XLSX' (Excel 2007+) files ENABLED.
```

```
Attaching package: 'gdata'
```

```
The following object(s) are masked from 'package:stats':
```

```
  nobs
```

```
The following object(s) are masked from 'package:utils':
```

```
  object.size
```

```
Loading required package: caTools
Loading required package: bitops
Loading required package: grid
Loading required package: KernSmooth
KernSmooth 2.23 loaded
Copyright M. P. Wand 1997-2009
Loading required package: MASS
```

```
Attaching package: 'gplots'
```

```
The following object(s) are masked from 'package:stats':
```

```
  lowess
```

```
Loading required package: matlab
```

```
Attaching package: 'matlab'
```

```
The following object(s) are masked from 'package:stats':
```

```
  reshape
```

```
The following object(s) are masked from 'package:utils':
```

```
  find, fix
```

```
The following object(s) are masked from 'package:base':
```

```
  sum
```

```
Out[7]:
```

Heatmapofgenomedata::ward_clustering

```
In [8]: pair_data = analysis.Analysis(['kb|g.81', 'kb|g.3153'], 'genome', level='level1')
print pair_data.ids()

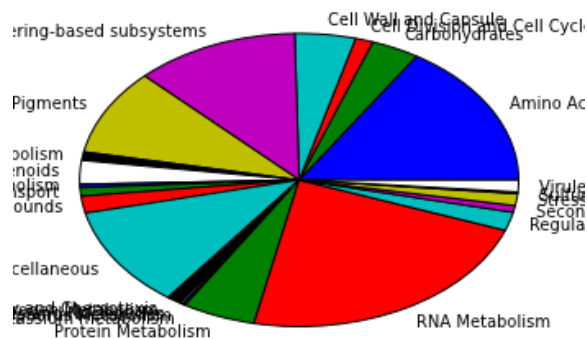
[u'kb|g.3153', u'kb|g.81']
```

```
In [9]: ecoli_slice = slice_column(pair_data.matrix, 1)
sulfo_slice = slice_column(pair_data.matrix, 0)

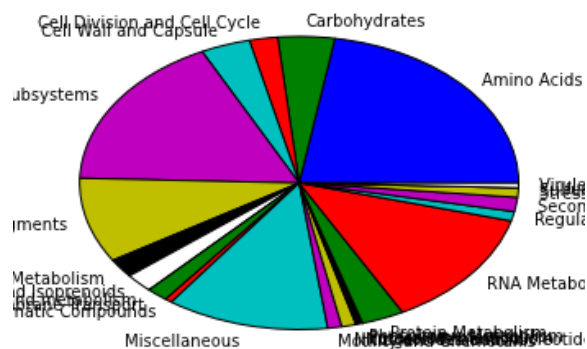
print ecoli_slice
print sulfo_slice
print pair_data.annotations()

[86, 19, 7, 24, 65, 51, 5, 14, 2, 5, 10, 60, 1, 0, 5, 1, 1, 29, 122, 11, 4, 7, 1,
7]
[43, 8, 4, 7, 33, 18, 4, 4, 0, 3, 1, 23, 2, 2, 1, 0, 0, 6, 25, 2, 3, 2, 0, 1]
[u'Amino Acids and Derivatives', u'Carbohydrates', u'Cell Division and Cell
Cycle', u'Cell Wall and Capsule', u'Clustering-based subsystems', u'Cofactors,
Vitamins, Prosthetic Groups, Pigments', u'DNA Metabolism', u'Fatty Acids, Lipids,
and Isoprenoids', u'Iron acquisition and metabolism', u'Membrane Transport',
u'Metabolism of Aromatic Compounds', u'Miscellaneous', u'Motility and Chemotaxis',
u'Nitrogen Metabolism', u'Nucleosides and Nucleotides', u'Phosphorus Metabolism',
u'Potassium metabolism', u'Protein Metabolism', u'RNA Metabolism', u'Regulation
and Cell signaling', u'Secondary Metabolism', u'Stress Response', u'Sulfur
Metabolism', u'Virulence, Disease and Defense']
```

```
In [10]: ecoli_pie = pie(ecoli_slice, labels=pair_data.annotations())
```



```
In [11]: sulfo_pie = pie(sulfo_slice, labels=pair_data.annotations())
```



In [11]: