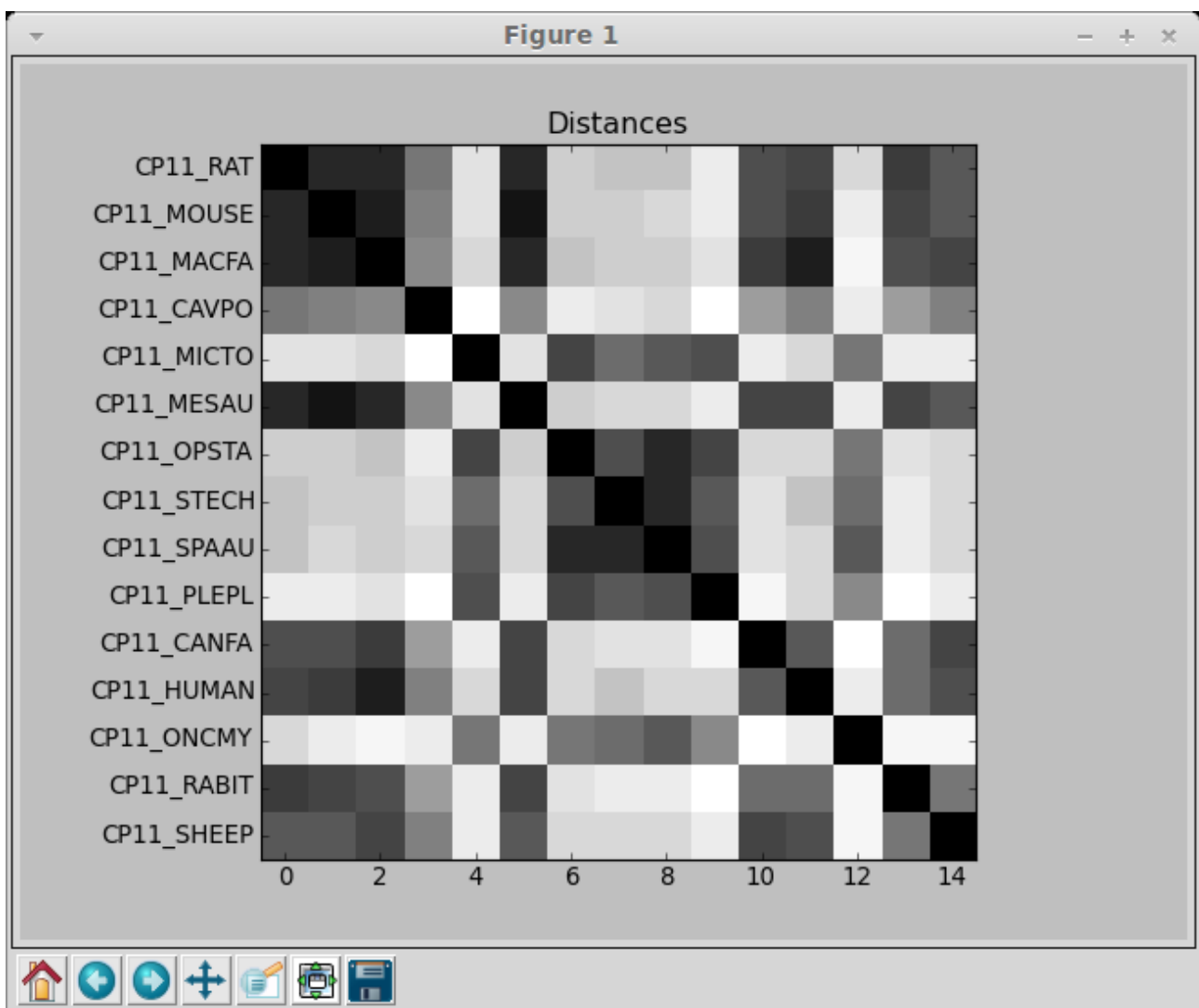


Jacek Radajewski
Student number: 43612772
SCIE2100 Prac 5

Question 1

From the heat map shown below it can be seen that various species fall into groups:

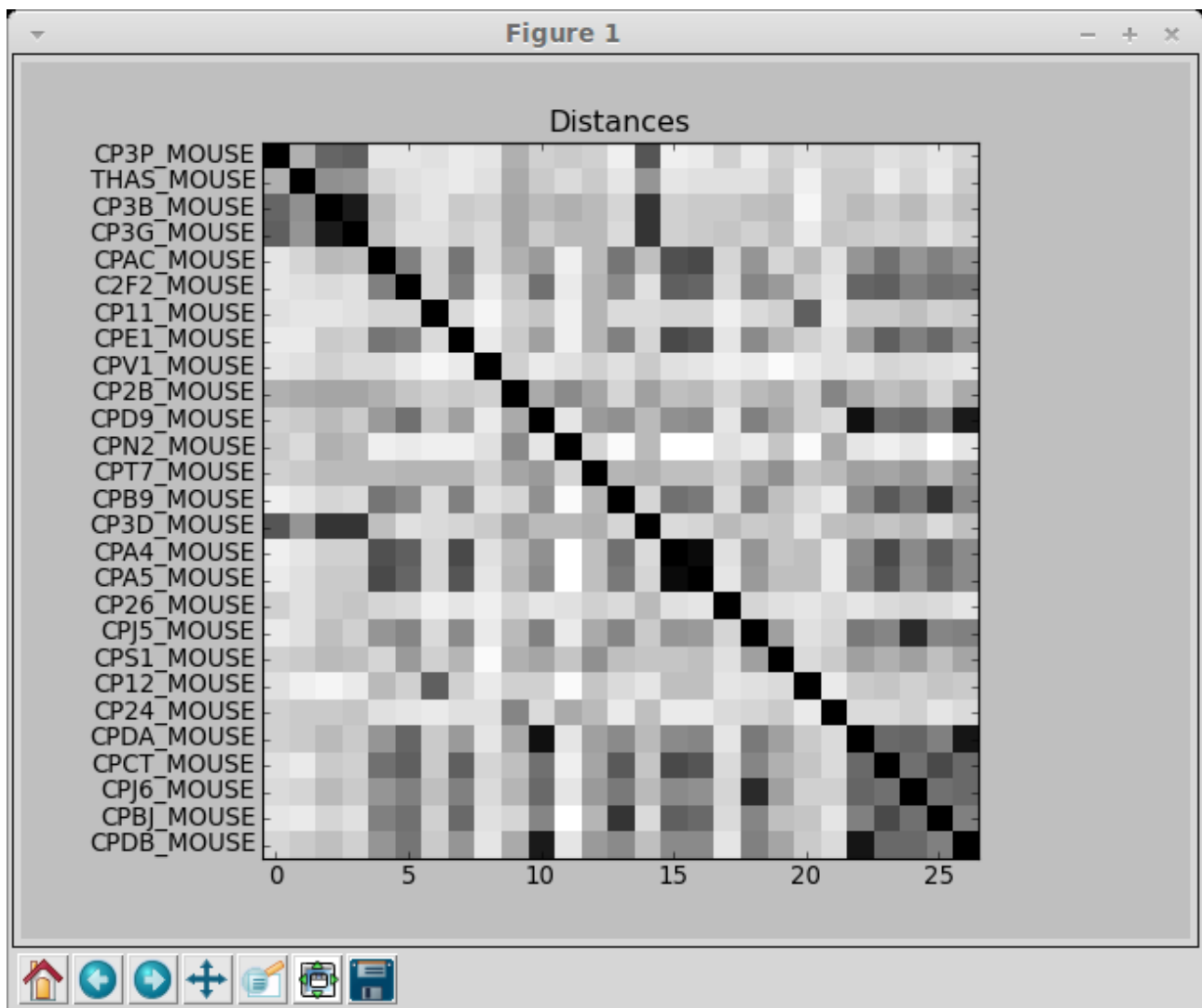
1. rat, mouse, crab-eating macaque, guinea pig (perhaps little bit further away than the rest as the shades are lighter), golden hamster, dog, human, rabbit, sheep. This is expected as all of these species are mammals.
2. Rainbow trout, European plaice, glithead sea bream, scup and oyster toad fish. These are all fish so the grouping was expected



Question 2

Short evolutionary distances can be seen for the heat map below. Highlighted prefix and uniprot functional description are show below. Not sure if CP3B & CP3G have similar function, but CPDA, CPD9 and CPDB have identical function listed in uniprot.

Prefix	Uniprot function
CP3B	Catalyzes erythromycin N-demethylation, nifedipine oxidation and testosterone 6 beta-hydroxylation.
CP3G	Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics.
CPDA	Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics.
CPD9	Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics.
CPDB	Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics.
CPBJ	Cytochromes P450 are a group of heme-thiolate monooxygenases. In liver microsomes, this enzyme is involved in an NADPH-dependent electron transport pathway. It oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids, and xenobiotics.
CP26	Plays a key role in retinoic acid metabolism. Acts on retinoids, including all-trans-retinoic acid (RA) and its stereoisomer 9-cis-RA. Capable of both 4-hydroxylation and 18-hydroxylation. Responsible for generation of several hydroxylated forms of RA, including 4-OH-RA, 4-oxo-RA and 18-OH-RA.
CPj5	Monooxygenase
CPJ6	NADPH-hemoprotein



Code

```
'''
Created on 30/03/2014

@author: mikael
'''
import numpy as np
import matplotlib.pyplot as plt

from sequence import *

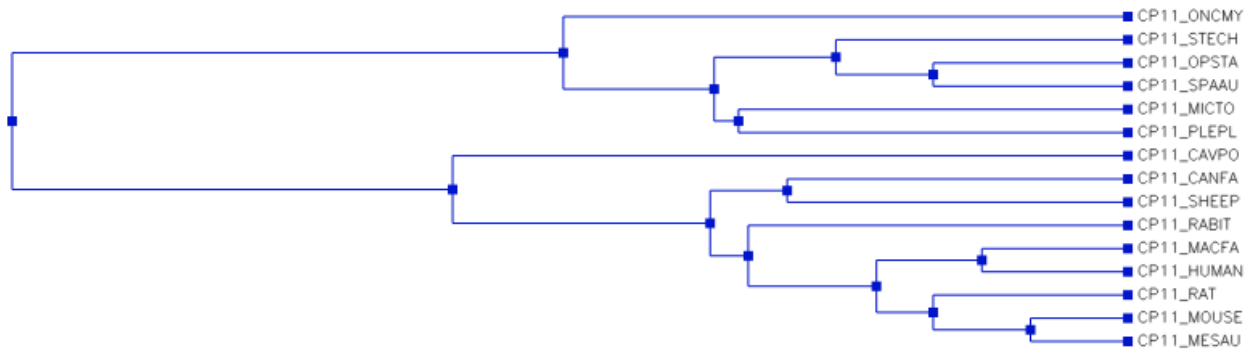
#aln = readClustalFile('cyp1a1.aln', Protein_Alphabet)
aln = readClustalFile('cyp_mouse.aln', Protein_Alphabet)
d = aln.calcDistances('fractional')

fig, ax = plt.subplots()
ax.imshow(d, plt.cm.gray, interpolation='nearest')
plt.yticks(np.arange(len(aln)), [s.name for s in aln])
plt.title('Distances')
plt.show()
```

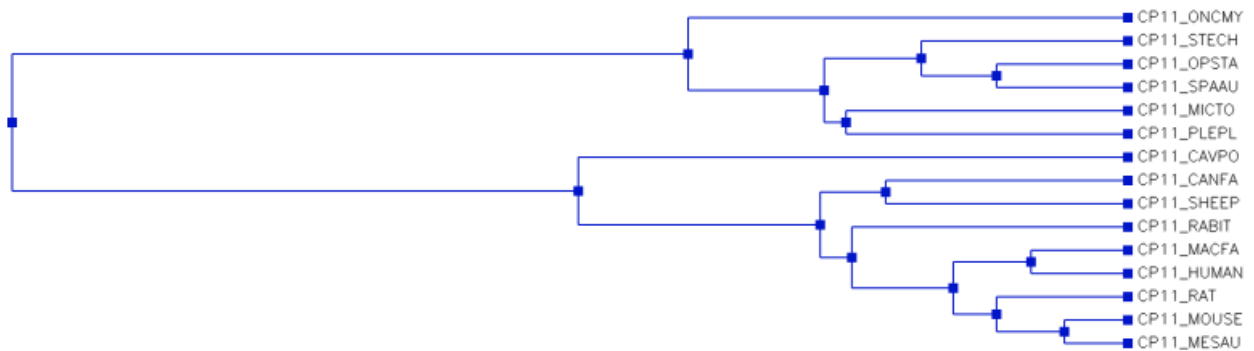
Question 3

Below are two trees fractional and k2p distance metrics

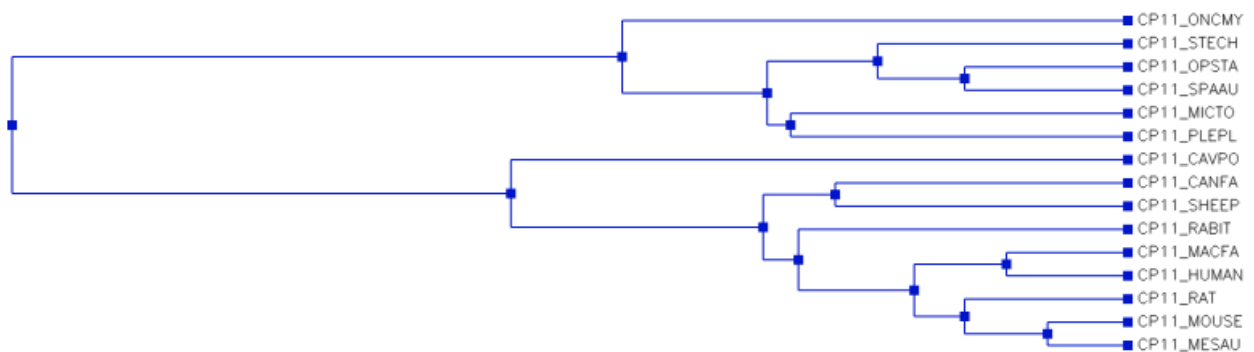
Fractional



K2P



Poisson



Appendix – Misc code

```
'''
Created on 30/04/2014

@author: s4361277
'''

from sequence import *
from phylo import *

tree = readNewick('cyp1a1.tree')
aln = readClustalFile('cyp1a1.aln', Protein_Alphabet)
for seq in aln:
    node = tree.findLabel(seq.name)
    a = tree.getAncestorsOf(node)
    print seq.name, 'has-distance', node.dist, 'from ancestor', a.label

# q 3
tree = runUPGMA(aln, "fractional")
tree2 = runUPGMA(aln, "k2p")
tree3 = runUPGMA(aln, "poisson")

print tree
print tree2
print tree3

#q4

aln2 = readClustalFile("cyp_mouse.aln", Protein_Alphabet)
q4tree = aln.calcDistances("fractional")

print q4tree

#q5

# tree = PhyloTree(PhyloNode("root"))
#
# aln = readClustalFile('cyp_mouse.aln', Protein_Alphabet)
# tree.putAlignment(aln)
# newickBeforeParsimony = tree.strSequences(10, 15)
# tree.parsimony()
```