Different routes you could take to analyze “18S” datasets

Questions are:
What do you lose in the analysis? What form of payment? How much flexibility / conformity?

Engage with a consultant:
James White Resphera biosciences: paid consultant. Does not require publication/ requires $.

Engage with a publicly financed project -- like the one we are working on.

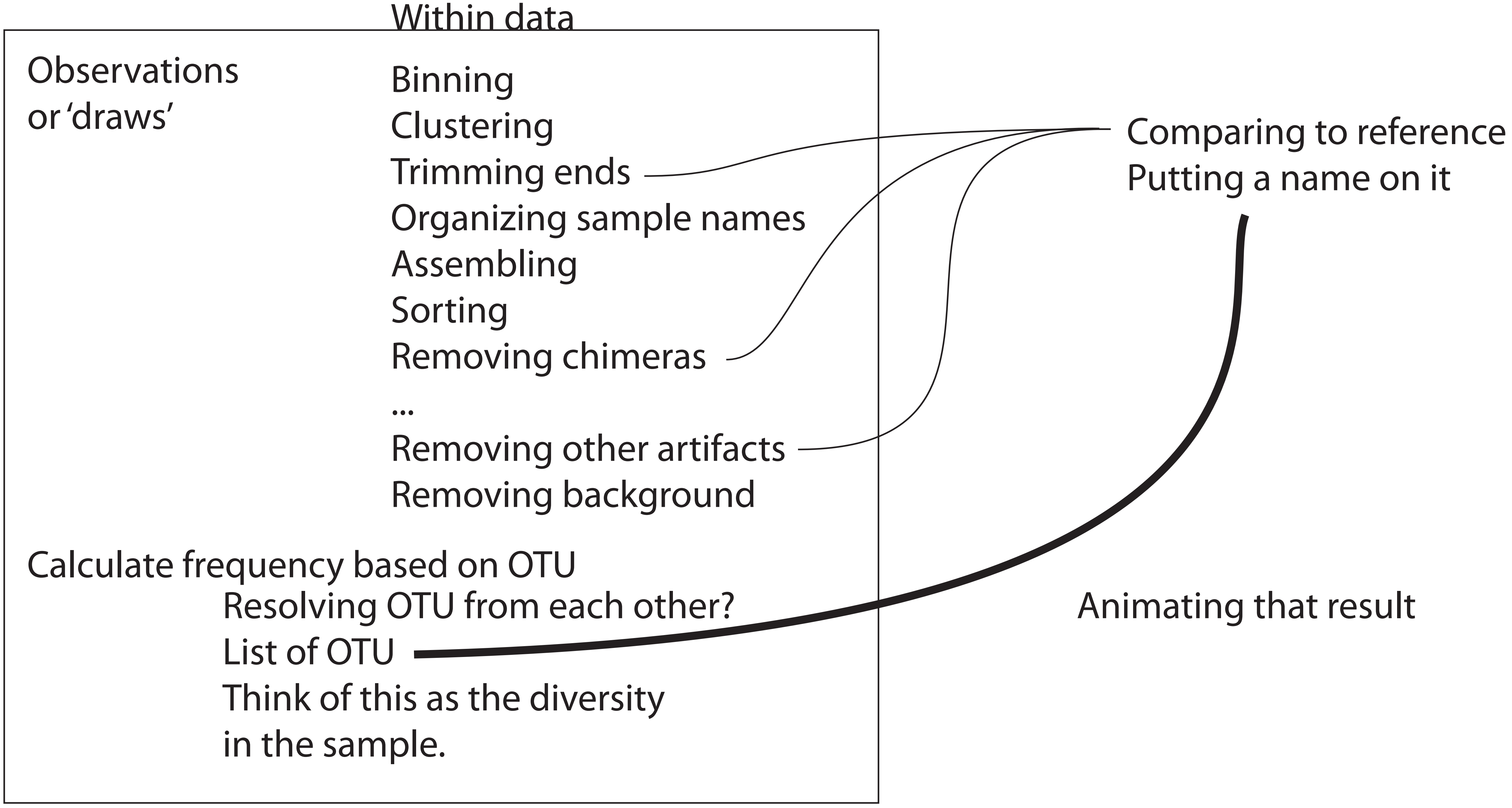
Upload your data to a website:
https://www.arb-silva.de/ngs/

Use a AWS instance of qiime or mothur (or other software).
https://forum.qiime2.org/t/basic-ec2-aws-qiime2-install-help/5655
OR
Use a ‘server’ in some other context.
What the above approaches involve is not working with the data directly on your own computer.

However we can use qiime / mothur / usearch if you are able to install them
(and can deal with command line).

Other options
Commercial software
CLC genomics has a java implimentation of _____ and if you have $ you can use it. Sometimes you c

MacVector: Concatenate reference database and map reads to it.



Within data

Observations
or 'draws'

Binning
Clustering
Trimming ends
Organizing sample names
Assembling
Sorting
Removing chimeras
...
Removing other artifacts
Removing background

Comparing to reference
Putting a name on it

Calculate frequency based on OTU
Resolving OTU from each other?
List of OTU
Think of this as the diversity
in the sample.

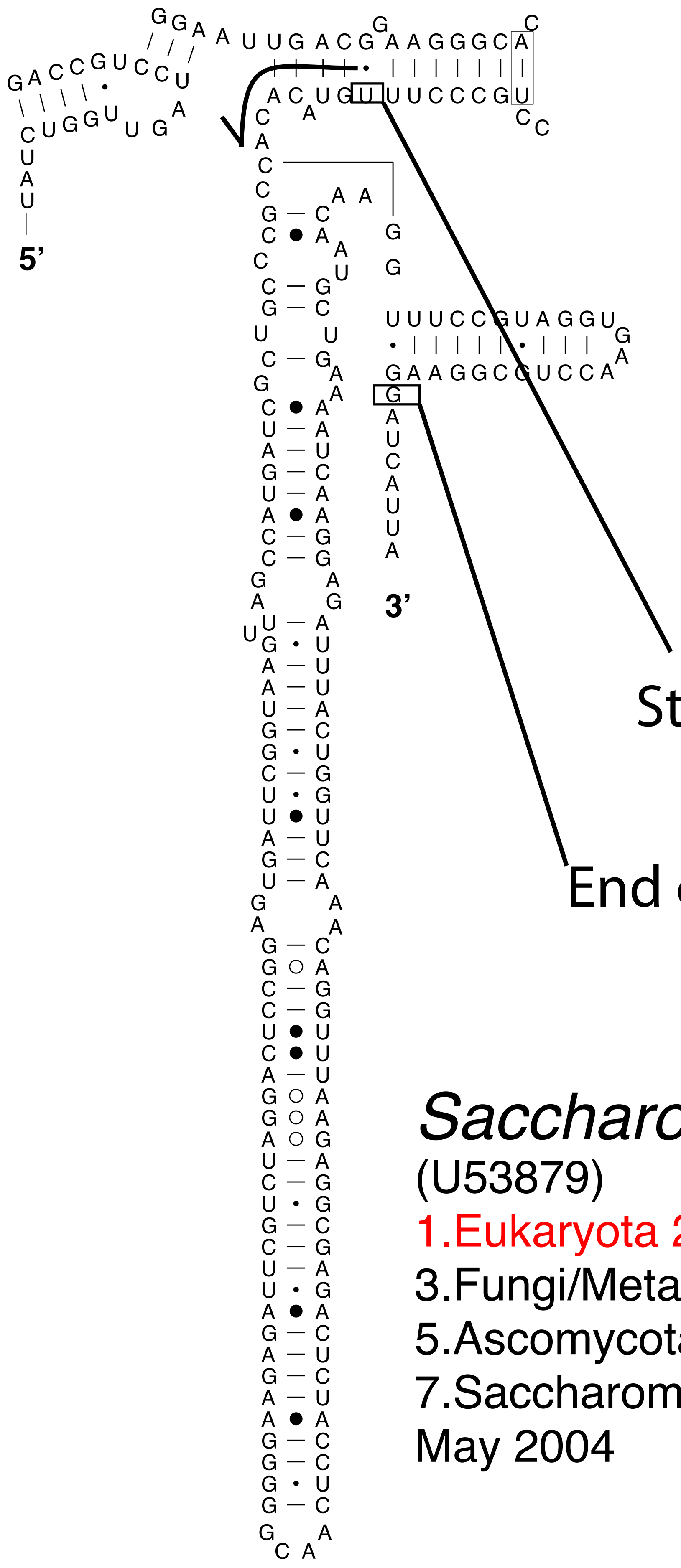
Animating that result

Usearch is great and self-referential (largely)

What is the cost of a 'wrong' sequence?
Chimera
Other artifact

What we did last week:

I handed you a set of __ sequences (drawn from Tara Oceans project)
We took one and used blastn and saw it had a 'good' --
very high identity match across the entire sequence to a Diplonemid.
We then digressed to command line
Look at sequence in 'text editor' WYSIWYG
Then attempt to run mothur
Essentially the next steps were like boiling down gravy -- reducing redundant sequences,
calculating dominant or most common sequences
•••Three window view -- mothur window, list of files in a window of the folder, text editor of results•••
We then could take the list of sequences and run blast on the command line.
If asked, we could summarize the results in different ways, with different levels of detail and specificity.
Started with 65, removed to 35 based on x, y and z.
Determined that Diplonemid was dominant sequence insample (and thus likely in sample of water).
We skipped trimming the sequence based on a reference alignment, chimera removal,
and display of results.



Start of yeast sequence

End of yeast sequence

Saccharomyces cerevisiae

(U53879)

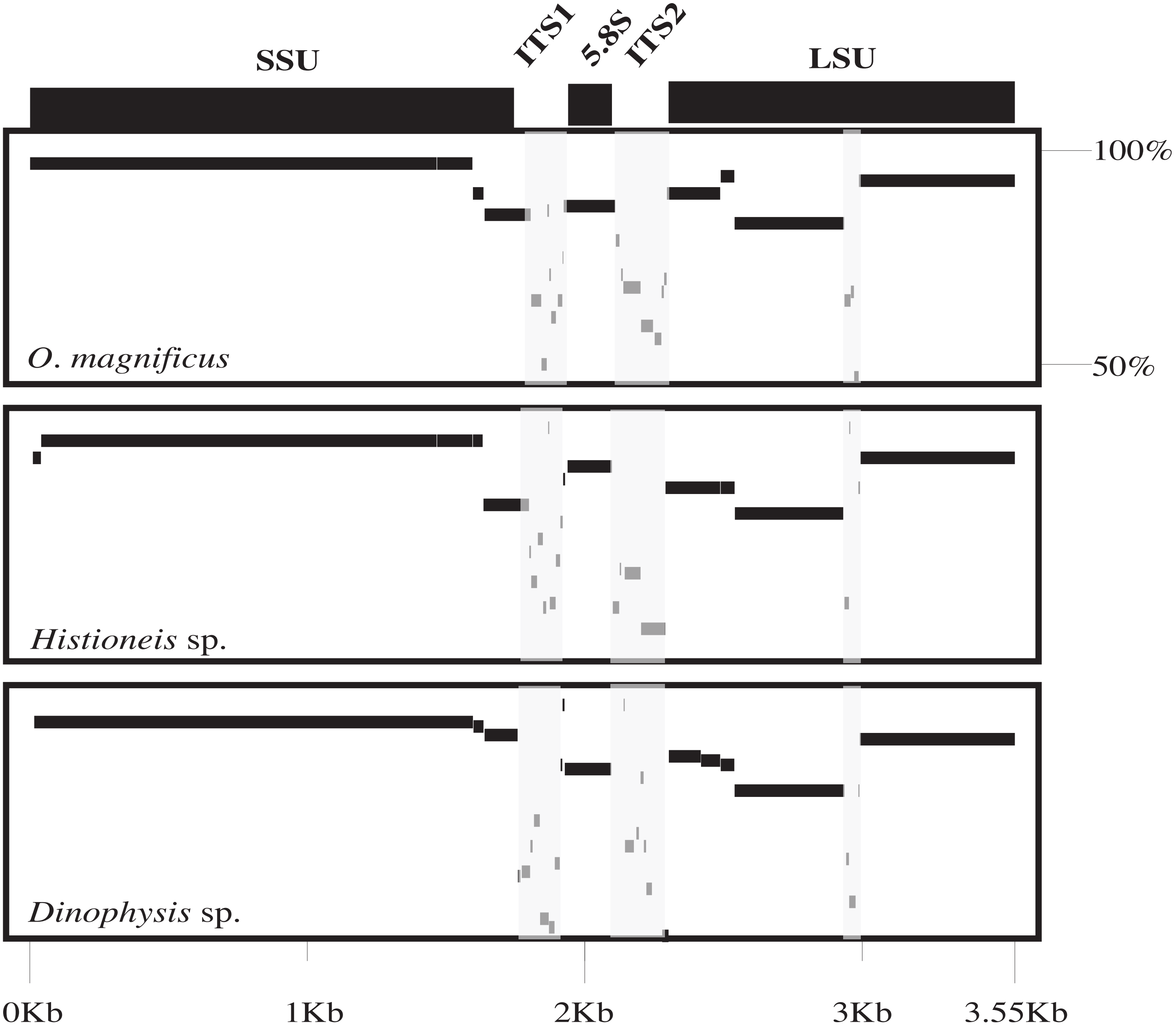
1.Eukaryota 2.eukaryote crown group

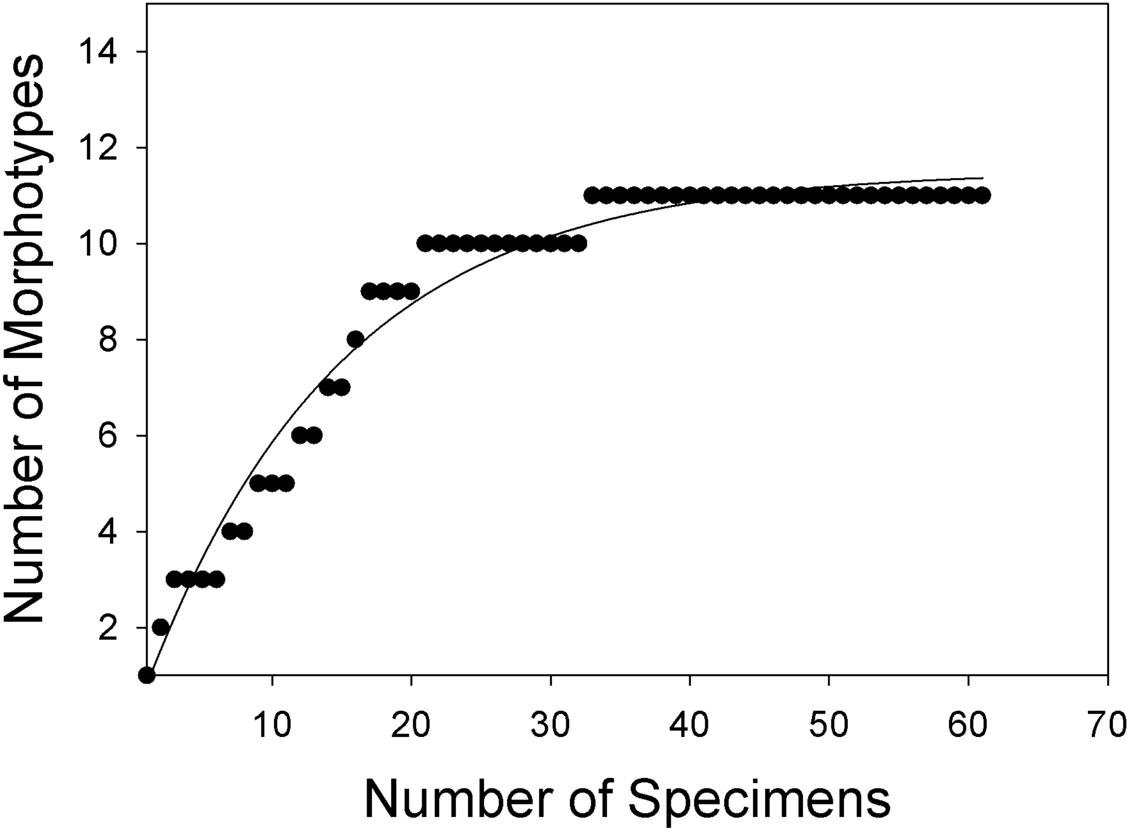
3.Fungi/Metazoa group 4.Fungi

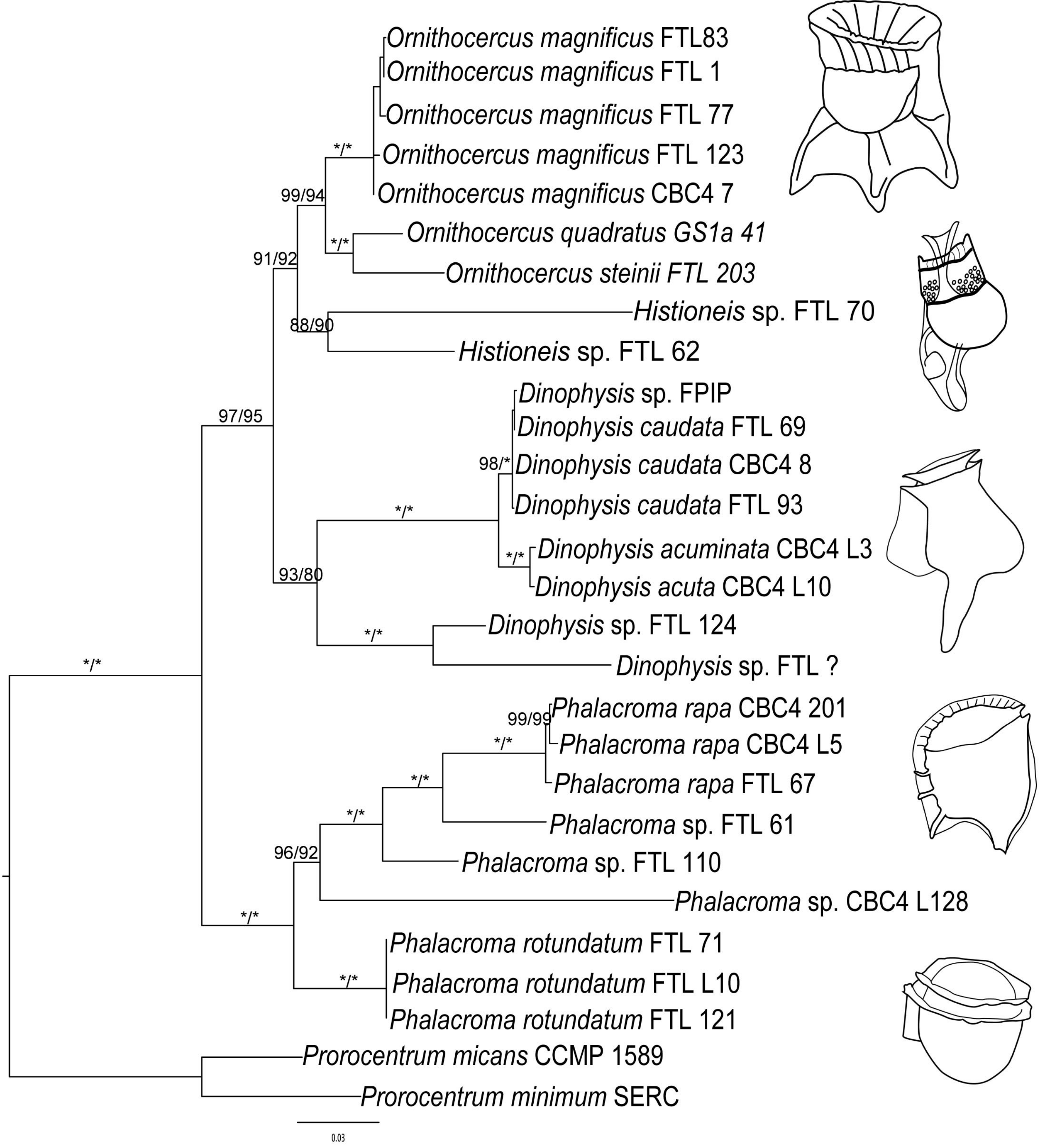
5.Ascomycota 6.Saccharomycetales

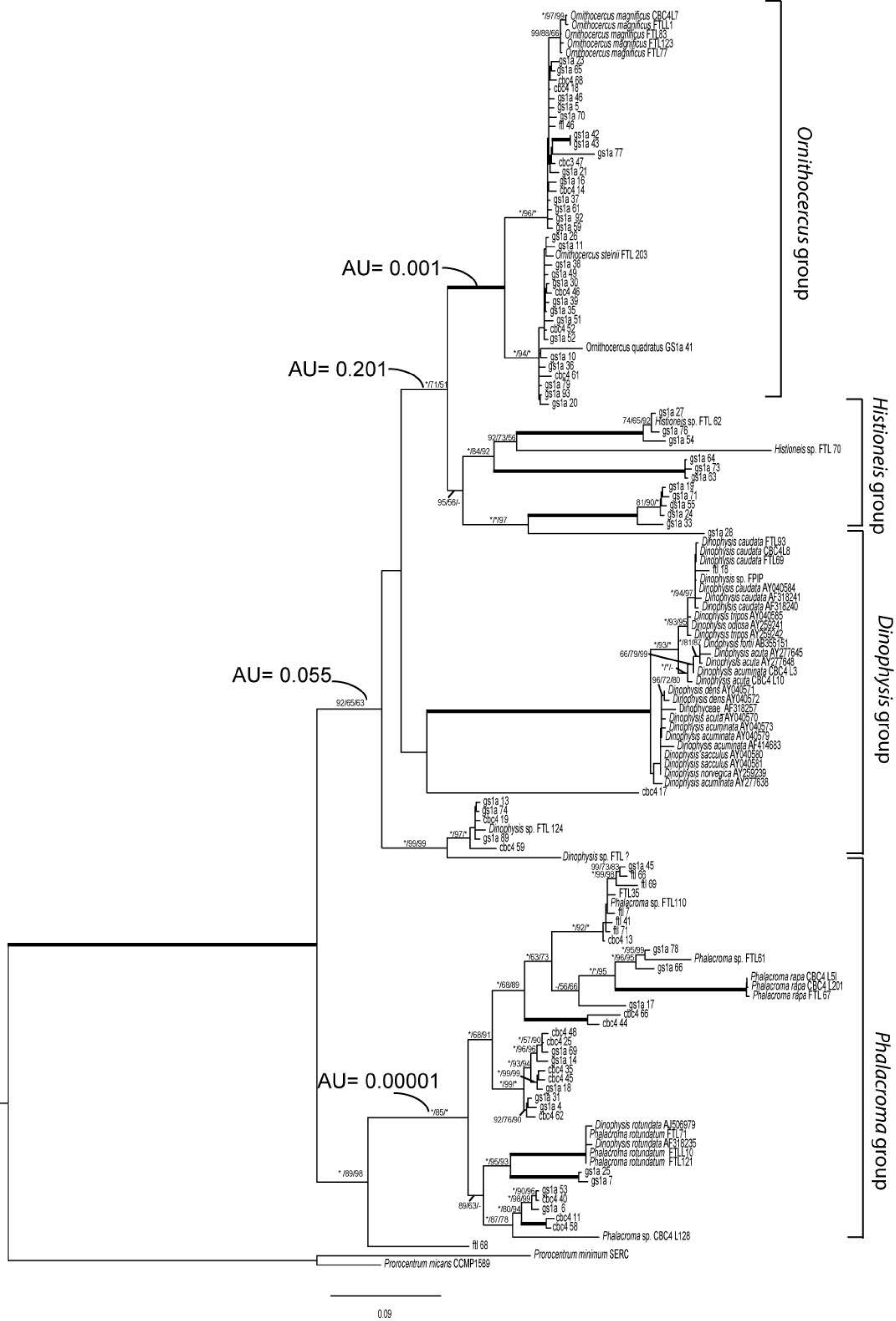
7.Saccharomycetaceae 8.Saccharomyces

May 2004


SSU
ITS1
5.8S
ITS2
LSU
100%
50%
O. magnificus
Histioneis sp.
Dinophysis sp.
0Kb
1Kb
2Kb
3Kb
3.55Kb


Number of Morphotypes
14
12
10
8
6
4
2
10
20
30
40
50
60
70
Number of Specimens


Ornithocercus magnificus FTL83
Ornithocercus magnificus FTL 1
Ornithocercus magnificus FTL 77
Ornithocercus magnificus FTL 123
Ornithocercus magnificus CBC4 7
Ornithocercus quadratus GS1a 41
Ornithocercus steinii FTL 203
Histioneis sp. FTL 70
Histioneis sp. FTL 62
Dinophysis sp. FPIP
Dinophysis caudata FTL 69
Dinophysis caudata CBC4 8
Dinophysis caudata FTL 93
Dinophysis acuminata CBC4 L3
Dinophysis acuta CBC4 L10
Dinophysis sp. FTL 124
Dinophysis sp. FTL ?
Phalacroma rapa CBC4 201
Phalacroma rapa CBC4 L5
Phalacroma rapa FTL 67
Phalacroma sp. FTL 61
Phalacroma sp. FTL 110
Phalacroma sp. CBC4 L128
Phalacroma rotundatum FTL 71
Phalacroma rotundatum FTL L10
Phalacroma rotundatum FTL 121
Prorocentrum micans CCMP 1589
Prorocentrum minimum SERC
/
99/94
/
91/92
88/90
97/95
98/*
/
/
93/80
/
/
99/99
/
/
/
96/92
/
/
0.03


*/97/99
Ornithocercus magnificus CBC4L7
Ornithocercus magnificus FTLL1
99/88/66
Ornithocercus magnificus FTL83
Ornithocercus magnificus FTL123
Ornithocercus magnificus FTL77
gs1a 23
gs1a 65
cbc4 68
cbc4 18
gs1a 46
gs1a 5
gs1a 70
ftl 46
gs1a 42
gs1a 43
gs1a 77
cbc3 47
gs1a 21
gs1a 16
cbc4 14
gs1a 37
gs1a 61
/96/
gs1a 92
gs1a 59
gs1a 26
gs1a 11
Ornithocercus steinii FTL 203
gs1a 38
gs1a 49
gs1a 30
cbc4 46
gs1a 39
gs1a 35
gs1a 51
cbc4 52
gs1a 52
Ornithocercus quadratus GS1a 41
/94/
gs1a 10
gs1a 36
cbc4 61
gs1a 79
gs1a 93
gs1a 20
Ornithocercus group
AU= 0.001
AU= 0.201
*/71/51
74/65/92
gs1a 27
Histioneis sp. FTL 62
gs1a 76
gs1a 54
92/73/56
Histioneis sp. FTL 70
*/84/92
gs1a 64
gs1a 73
gs1a 63
gs1a 19
gs1a 71
81/90/*
gs1a 55
gs1a 24
gs1a 33
95/56/-
//97
gs1a 28
Histioneis group
Dinophysis caudata FTL93
Dinophysis caudata CBC4L8
Dinophysis caudata FTL69
ftl 18
Dinophysis sp. FPIP
*/94/97
Dinophysis caudata AY040584
Dinophysis caudata AF318241
Dinophysis caudata AF318240
*/93/95
Dinophysis tripos AY040585
Dinophysis odiosa AY259241
Dinophysis tripos AY259242
/93/
*/81/87
Dinophysis fortii AB355151
66/79/99
Dinophysis acuta AY277645
Dinophysis acuta AY277648
//-
Dinophysis acuminata CBC4 L3
Dinophysis acuta CBC4 L10
96/72/80
Dinophysis dens AY040571
Dinophysis dens AY040572
Dinophyceae AF318257
Dinophysis acuta AY040570
Dinophysis acuminata AY040573
Dinophysis acuminata AY040579
Dinophysis acuminata AF414683
Dinophysis sacculus AY040580
Dinophysis sacculus AY040581
Dinophysis norvegica AY259239
Dinophysis acuminata AY277638
cbc4 17
Dinophysis group
AU= 0.055
92/65/63
gs1a 13
gs1a 74
cbc4 19
Dinophysis sp. FTL 124
/97/
gs1a 89
*/99/99
cbc4 59
Dinophysis sp. FTL ?
99/73/83
gs1a 45
*/99/98
ftl 66
ftl 69
FTL35
Phalacroma sp. FTL110
ftl 7
ftl 41
/92/
ftl 71
cbc4 13
*/95/99
gs1a 78
Phalacroma sp. FTL61
*/63/73
*/96/95
gs1a 66
//95
Phalacroma rapa CBC4 L5l
*/68/89
-/56/66
Phalacroma rapa CBC4 L201
Phalacroma rapa FTL 67
gs1a 17
cbc4 66
cbc4 44
*/68/91
cbc4 48
*/57/90
cbc4 25
*/96/96
gs1a 69
*/93/94
gs1a 14
*/99/99
cbc4 35
cbc4 45
/99/
gs1a 18
AU= 0.00001
gs1a 31
gs1a 4
/85/
92/76/90
cbc4 62
Dinophysis rotundata AJ506979
Phalacroma rotundatum FTL71
Dinophysis rotundata AF318235
*/95/93
Phalacroma rotundatum FTLL10
Phalacroma rotundatum FTL121
gs1a 25
* /89/98
gs1a 7
*/90/96
gs1a 53
*/98/99
cbc4 40
89/63/-
gs1a 6
*/80/94
cbc4 11
*/87/78
cbc4 58
Phalacroma sp. CBC4 L128
ftl 68
Phalacroma group
Prorocentrum minimum SERC
Prorocentrum micans CCMP1589
0.09