AB GRAD CON23

1
00:00:04,230 --> 00:00:10,990
[Music]

2
00:00:14,990 --> 00:00:13,789
hello everyone I'm ready I'm a graduate

3
00:00:16,730 --> 00:00:15,000
student at the Earth Life Science

4
00:00:18,590 --> 00:00:16,740
Institute

5
00:00:20,990 --> 00:00:18,600
um in our study we have tried to

6
00:00:23,150 --> 00:00:21,000
understand scaling of protein function

7
00:00:26,210 --> 00:00:23,160
across the tree of life and the

8
00:00:28,189 --> 00:00:26,220
mechanisms that might have led to the

9
00:00:30,290 --> 00:00:28,199
species diversity on the tree of life

10
00:00:32,690 --> 00:00:30,300
right here

11
00:00:34,850 --> 00:00:32,700
so we study scaling using power loss and

12
00:00:36,590 --> 00:00:34,860
power laws are found everywhere if we

13
00:00:38,389 --> 00:00:36,600

consider the number of Web Hits on a web

14
00:00:40,490 --> 00:00:38,399
page in a given period of time or the

15
00:00:42,650 --> 00:00:40,500
earthquake magnitude in an area over a

16
00:00:45,360 --> 00:00:42,660
given period of time many natural and

17
00:00:47,389 --> 00:00:45,370
man-made processes follow power loss

18
00:00:49,190 --> 00:00:47,399
[Music]

19
00:00:50,930 --> 00:00:49,200
so how can we help

20
00:00:53,270 --> 00:00:50,940
um help us how can this help us

21
00:00:55,189 --> 00:00:53,280
understand some Concepts in biology for

22
00:00:57,170 --> 00:00:55,199
that let's consider a Lego set and

23
00:00:59,750 --> 00:00:57,180
consider the unique pieces in relation

24
00:01:02,869 --> 00:00:59,760
to the total pieces in the Lego set when

25
00:01:06,350 --> 00:01:02,879
we plot this on a log log scale

26
00:01:07,969 --> 00:01:06,360
we see the larger Lego sets use more

27
00:01:09,649 --> 00:01:07,979
Unique Piece types but they

28
00:01:11,090 --> 00:01:09,659
progressively go on using lesser

29
00:01:13,490 --> 00:01:11,100
additional piece types so they're

30
00:01:16,010 --> 00:01:13,500
becoming more efficient which means the

31
00:01:17,929 --> 00:01:16,020
larger sets are using uh the same pieces

32
00:01:20,390 --> 00:01:17,939
the smaller sets are using but in more

33
00:01:22,490 --> 00:01:20,400
efficient and more complex ways so what

34
00:01:24,530 --> 00:01:22,500
we're observing in these plots is a

35
00:01:26,630 --> 00:01:24,540
scaling relationship and when we observe

36
00:01:28,670 --> 00:01:26,640
a scaling relationship we could say that

37
00:01:30,050 --> 00:01:28,680
maybe there's a set of rules that sort

38
00:01:32,570 --> 00:01:30,060

of governing the way something is

39
00:01:36,350 --> 00:01:34,609
so you as a power law equation one

40
00:01:38,630 --> 00:01:36,360
quantity varying as a power law of the

41
00:01:40,550 --> 00:01:38,640
other and when we plot this on a log log

42
00:01:42,890 --> 00:01:40,560
scale we get a straight line with the

43
00:01:44,690 --> 00:01:42,900
slope Alpha so previous Studies have

44
00:01:46,789 --> 00:01:44,700
shown that genes in a specific

45
00:01:48,770 --> 00:01:46,799
functional category scale as a power law

46
00:01:51,050 --> 00:01:48,780
of the total number of genes in a genome

47
00:01:53,510 --> 00:01:51,060
so for example transcription regulation

48
00:01:56,510 --> 00:01:53,520
is almost quadratically scaling which

49
00:01:58,190 --> 00:01:56,520
means if the genome doubles in size the

50
00:01:59,810 --> 00:01:58,200
genes in this specific category are

51
00:02:02,030 --> 00:01:59,820
going to quadruple

52
00:02:04,069 --> 00:02:02,040
so we have tried to include an expanded

53
00:02:06,830 --> 00:02:04,079
taxonomy in our study and for that we

54
00:02:08,690 --> 00:02:06,840
use the eggnog database so after power

55
00:02:10,309 --> 00:02:08,700
of fitting we saw different Trends in

56
00:02:12,110 --> 00:02:10,319
our data for the smaller and the larger

57
00:02:13,970 --> 00:02:12,120
genomes so we carried out piecewise

58
00:02:15,949 --> 00:02:13,980
regression to give Justice to the

59
00:02:17,630 --> 00:02:15,959
different patterns and scaling observed

60
00:02:19,790 --> 00:02:17,640
in the plots

61
00:02:21,530 --> 00:02:19,800
so for example we wanted to capture the

62
00:02:23,930 --> 00:02:21,540
slope variability for the smaller and

63
00:02:25,910 --> 00:02:23,940

the larger genome sizes so for example

64
00:02:27,890 --> 00:02:25,920
in category M which is cell wall and

65
00:02:29,809 --> 00:02:27,900
cell membrane proteins

66
00:02:31,850 --> 00:02:29,819
um so the x-axis has the total protein

67
00:02:33,530 --> 00:02:31,860
annotations and the y-axis has the

68
00:02:36,410 --> 00:02:33,540
category annotations for that specific

69
00:02:38,089 --> 00:02:36,420
category so for the smaller genomes we

70
00:02:39,830 --> 00:02:38,099
can see the proteins are scaling fast

71
00:02:42,410 --> 00:02:39,840
which means as the genome size is

72
00:02:44,270 --> 00:02:42,420
increasing they are incorporating

73
00:02:46,250 --> 00:02:44,280
um more and more proteins faster than

74
00:02:48,110 --> 00:02:46,260
the larger genomes because after a

75
00:02:50,449 --> 00:02:48,120
statistically detected breakpoint the

76
00:02:52,670 --> 00:02:50,459
scaling slows down and we can see a

77
00:02:55,850 --> 00:02:52,680
similar Trend in category tree category

78
00:02:58,670 --> 00:02:55,860
T so we saw this trend in most of the

79
00:03:01,250 --> 00:02:58,680
breakpoints that that were supported in

80
00:03:03,650 --> 00:03:01,260
bacteria interestingly we observed an

81
00:03:05,750 --> 00:03:03,660
opposite Trend in archaea the scaling is

82
00:03:08,750 --> 00:03:05,760
slow in the start but fastens up after

83
00:03:10,369 --> 00:03:08,760
the statistically detected breakpoint so

84
00:03:12,350 --> 00:03:10,379
a lot of categories were common between

85
00:03:14,330 --> 00:03:12,360
archaea and bacteria

86
00:03:15,770 --> 00:03:14,340
um but there were some categories that

87
00:03:17,930 --> 00:03:15,780
were exclusively present either in

88
00:03:19,550 --> 00:03:17,940

archaeon bacteria

89
00:03:20,930 --> 00:03:19,560
um and it's also interesting to observe

90
00:03:22,309 --> 00:03:20,940
these differences in scaling pattern

91
00:03:25,009 --> 00:03:22,319
before and after the break point in

92
00:03:26,990 --> 00:03:25,019
Archaea and bacteria so we thought maybe

93
00:03:29,210 --> 00:03:27,000
these differences in scaling patterns

94
00:03:30,649 --> 00:03:29,220
were caused by phyla specific scaling so

95
00:03:33,290 --> 00:03:30,659
we broke these domains down into

96
00:03:35,210 --> 00:03:33,300
specific phyla and found great variation

97
00:03:37,910 --> 00:03:35,220
in all the phyla for all the categories

98
00:03:40,729 --> 00:03:37,920
so for example we have category H here

99
00:03:42,229 --> 00:03:40,739
which is coenzyme transport proteins

100
00:03:44,449 --> 00:03:42,239
um which also happens to be the most

101
00:03:46,430 --> 00:03:44,459
variable across all the phyla

102
00:03:48,350 --> 00:03:46,440
so we thought maybe this file a specific

103
00:03:49,670 --> 00:03:48,360
scaling is causing the positioning of

104
00:03:51,289 --> 00:03:49,680
the breakpoints that we observed

105
00:03:52,070 --> 00:03:51,299
previously

106
00:03:54,350 --> 00:03:52,080
um

107
00:03:56,149 --> 00:03:54,360
so I place these breakpoints on the

108
00:03:58,430 --> 00:03:56,159
total protein annotations to see if

109
00:04:00,410 --> 00:03:58,440
there's any specific pattern but we can

110
00:04:01,970 --> 00:04:00,420
see these individual phyla are spanning

111
00:04:04,850 --> 00:04:01,980
the breakpoints and there is no specific

112
00:04:06,589 --> 00:04:04,860
preference of for the break the file as

113
00:04:09,470 --> 00:04:06,599

to be present either on either sides of

114
00:04:10,850 --> 00:04:09,480
the breakpoints so maybe taxonomy is not

115
00:04:12,649 --> 00:04:10,860
causing the positioning of these

116
00:04:14,449 --> 00:04:12,659
breakpoints and maybe there are some

117
00:04:15,949 --> 00:04:14,459
other factors like physiological or

118
00:04:18,770 --> 00:04:15,959
environmental factors that are causing

119
00:04:20,569 --> 00:04:18,780
these fake points

120
00:04:23,270 --> 00:04:20,579
so we were also interested in these

121
00:04:25,490 --> 00:04:23,280
groups CPR and d-pan uh so these groups

122
00:04:28,969 --> 00:04:25,500
have extremely small genomes and they

123
00:04:31,310 --> 00:04:28,979
lack um major metabolic pathways uh so

124
00:04:33,770 --> 00:04:31,320
we compared them with um

125
00:04:36,170 --> 00:04:33,780
eukaryotes unicellular eukaryotes and

126
00:04:37,969 --> 00:04:36,180
Asgard alkia so for some categories we

127
00:04:40,129 --> 00:04:37,979
can see the scaling is very similar for

128
00:04:41,570 --> 00:04:40,139
category o but for some categories the

129
00:04:44,090 --> 00:04:41,580
scaling is very different like in

130
00:04:45,710 --> 00:04:44,100
category C which goes on to show there

131
00:04:47,749 --> 00:04:45,720
are different ways in which an organism

132
00:04:49,370 --> 00:04:47,759
can adapt while growing in their genome

133
00:04:50,030 --> 00:04:49,380
sizes

134
00:04:52,189 --> 00:04:50,040
um

135
00:04:54,409 --> 00:04:52,199
I've just discussed a few key results in

136
00:04:56,390 --> 00:04:54,419
my uh talk so if you want to discuss I

137
00:04:58,790 --> 00:04:56,400
would be interested please come by and

138
00:05:03,640 --> 00:04:58,800

stop at panel two for the poster thank