
AbSciCon
2019

1
00:00:00,790 --> 00:00:07,320
[Music]

2
00:00:12,220 --> 00:00:09,060
[Applause]

3
00:00:15,430 --> 00:00:12,230
we're interested in the role of RNA in

4
00:00:16,930 --> 00:00:15,440
early evolution of life and today I will

5
00:00:19,170 --> 00:00:16,940
tell you a little bit about our

6
00:00:22,540 --> 00:00:19,180
experiments that we've done in that area

7
00:00:24,759 --> 00:00:22,550
specifically about our interest in

8
00:00:26,560 --> 00:00:24,769
determining the relative role of

9
00:00:28,870 --> 00:00:26,570
different evolutionary mechanisms in the

10
00:00:31,509 --> 00:00:28,880
in the in the evolution of functional

11
00:00:34,390 --> 00:00:31,519
RNA and part of the motivation behind

12
00:00:35,770 --> 00:00:34,400
that is the fact that in biology there

13
00:00:39,370 --> 00:00:35,780

are a number of different evolutionary

14
00:00:45,510 --> 00:00:39,380
mechanisms that operate at creating

15
00:00:52,720 --> 00:00:45,520
diversity and creating more functional

16
00:00:54,370 --> 00:00:52,730
polymers or forms and of those only a

17
00:00:57,820 --> 00:00:54,380
number have been investigated by in

18
00:01:00,370 --> 00:00:57,830
vitro evolution studies and primarily

19
00:01:04,149 --> 00:01:00,380
the the dominant mechanism that has been

20
00:01:06,429 --> 00:01:04,159
investigated point mutation and that is

21
00:01:09,670 --> 00:01:06,439
partly driven by the fact that a number

22
00:01:13,480 --> 00:01:09,680
of theoretical studies have suggested

23
00:01:17,260 --> 00:01:13,490
that RNA is of increased complexity or

24
00:01:21,580 --> 00:01:17,270
increased lengths are connected by large

25
00:01:24,010 --> 00:01:21,590
swathes of neutral networks which would

26
00:01:27,220 --> 00:01:24,020
allow them to by point mutation traverse

27
00:01:29,560 --> 00:01:27,230
large large swathes of sequence space

28
00:01:31,540 --> 00:01:29,570
therefore given the ability to

29
00:01:35,500 --> 00:01:31,550
potentially change structures in very

30
00:01:38,020 --> 00:01:35,510
radical ways so the way we study this is

31
00:01:40,450 --> 00:01:38,030
by evolution of like a stripe designs

32
00:01:42,850 --> 00:01:40,460
and we evolved two separate independent

33
00:01:44,530 --> 00:01:42,860
like a star designs one short 120

34
00:01:46,630 --> 00:01:44,540
nucleotides and one long one of 80

35
00:01:48,250 --> 00:01:46,640
nucleotides and a typical in vitro

36
00:01:50,649 --> 00:01:48,260
evolution experiment begins with a

37
00:01:53,740 --> 00:01:50,659
diverse library in this case this is a

38
00:01:57,100 --> 00:01:53,750

20 nucleotide random region that's

39
00:01:59,410 --> 00:01:57,110
flanked by two constant regions and we

40
00:02:02,950 --> 00:01:59,420
transcribed this into RNA and that

41
00:02:04,630 --> 00:02:02,960
incubate with the substrate to allow for

42
00:02:07,390 --> 00:02:04,640
ligation in this case the substrate

43
00:02:12,580 --> 00:02:07,400
brings about it is a constant region

44
00:02:14,619 --> 00:02:12,590
and right and after that after that the

45
00:02:16,449 --> 00:02:14,629
the partitioning reaction separates the

46
00:02:18,789 --> 00:02:16,459
active molecules from the inactive ones

47
00:02:19,980 --> 00:02:18,799
in the case of the first round that of

48
00:02:23,520 --> 00:02:19,990
course is the the

49
00:02:26,160 --> 00:02:23,530
of the inactive molecules dominates and

50
00:02:29,940 --> 00:02:26,170
then the the smaller number of active

51
00:02:31,800 --> 00:02:29,950
ones is then reverse transcribed PCR

52
00:02:35,700 --> 00:02:31,810
amplified and then converted back to DNA

53
00:02:38,550 --> 00:02:35,710
the process continues we do this we did

54
00:02:40,380 --> 00:02:38,560
this for the 20 nucleotide selection and

55
00:02:43,820 --> 00:02:40,390
then for the 80 nucleotide selection in

56
00:02:46,650 --> 00:02:43,830
which case it was 8 rounds of selection

57
00:02:50,120 --> 00:02:46,660
punctuated by neurogenesis that was

58
00:02:53,940 --> 00:02:50,130
after the 5th round so what do we get

59
00:02:56,790 --> 00:02:53,950
the the winners of the 20th selection

60
00:02:59,220 --> 00:02:56,800
are these like aces they have the

61
00:03:02,910 --> 00:02:59,230
ligation Junction that's not base paired

62
00:03:05,400 --> 00:03:02,920
and this stem loop structure the

63
00:03:07,530 --> 00:03:05,410

dominant molecules from the 20 and

64
00:03:09,390 --> 00:03:07,540
selection are how to account them

65
00:03:11,010 --> 00:03:09,400
basically we look at their abundance in

66
00:03:13,500 --> 00:03:11,020
the final population and we can do this

67
00:03:14,940 --> 00:03:13,510
because there's the the representation

68
00:03:17,610 --> 00:03:14,950
of the molecules and the initial pool is

69
00:03:20,670 --> 00:03:17,620
very large so we have high copy number

70
00:03:24,090 --> 00:03:20,680
in the starting pool and so what do

71
00:03:25,860 --> 00:03:24,100
these populations look like this is my

72
00:03:29,280 --> 00:03:25,870
way of representing them

73
00:03:30,870 --> 00:03:29,290
so we clustered them into sequence

74
00:03:33,570 --> 00:03:30,880
networks based on their ability to

75
00:03:36,690 --> 00:03:33,580
connect through point mutation and the

76
00:03:38,850 --> 00:03:36,700
final population looks the way imagine

77
00:03:42,240 --> 00:03:38,860
it's something like this so we have two

78
00:03:44,130 --> 00:03:42,250
dominant networks one is of but very

79
00:03:45,720 --> 00:03:44,140
small in the number of sequences but

80
00:03:47,280 --> 00:03:45,730
they're very abundant so those are the

81
00:03:51,480 --> 00:03:47,290
winners the ones we expect to be very

82
00:03:53,600 --> 00:03:51,490
active and the larger pool of sequences

83
00:03:56,670 --> 00:03:53,610
that are the representative broader

84
00:03:58,230 --> 00:03:56,680
swath of sequence space but are very low

85
00:04:03,560 --> 00:03:58,240
and abundant so we don't expect them to

86
00:04:05,490 --> 00:04:03,570
be very active the the alien population

87
00:04:08,130 --> 00:04:05,500
these are the winners from the alien

88
00:04:10,520 --> 00:04:08,140

population of course they're alien

89
00:04:13,350 --> 00:04:10,530
sequence space is very large

90
00:04:15,900 --> 00:04:13,360
astronomically large and so we have very

91
00:04:21,180 --> 00:04:15,910
sparse seek a very sparsely sampled

92
00:04:23,490 --> 00:04:21,190
sequence space what happens so the the

93
00:04:24,930 --> 00:04:23,500
winners are the most enriched sequences

94
00:04:26,670 --> 00:04:24,940
they're not the most abundant ones but

95
00:04:28,469 --> 00:04:26,680
the most the ones they actually happen

96
00:04:29,760 --> 00:04:28,479
to be also the very abundant ones but

97
00:04:31,560 --> 00:04:29,770
they are the ones that increase in

98
00:04:33,500 --> 00:04:31,570
abundance over different rounds of

99
00:04:35,909 --> 00:04:33,510
selection the most

100
00:04:39,719 --> 00:04:35,919
what is the important thing here is that

101
00:04:42,510 --> 00:04:39,729
we from these two unrelated selections

102
00:04:44,820 --> 00:04:42,520
we get ligases that are the winners in

103
00:04:47,300 --> 00:04:44,830
their respective categories that are

104
00:04:50,600 --> 00:04:47,310
very sequence similar in sequence or

105
00:04:53,969 --> 00:04:50,610
identical in sequence so what we get is

106
00:04:56,730 --> 00:04:53,979
the same identical ligation Junction and

107
00:04:59,580 --> 00:04:56,740
then that the smaller motif is actually

108
00:05:02,120 --> 00:04:59,590
embedded within the larger motif which

109
00:05:06,480 --> 00:05:02,130
suggests that the addition of this top

110
00:05:09,360 --> 00:05:06,490
stem-loop structure could lead to

111
00:05:13,830 --> 00:05:09,370
evolution of higher activity ligases

112
00:05:15,930 --> 00:05:13,840
so we wanted to test this and we tested

113
00:05:18,360 --> 00:05:15,940

their activity so the 20 and ligase is

114
00:05:21,390 --> 00:05:18,370
not not very fast but the addition of

115
00:05:24,090 --> 00:05:21,400
the stem loop adds about thousand fold

116
00:05:27,510 --> 00:05:24,100
in gives a thousandfold improvement in

117
00:05:30,180 --> 00:05:27,520
activity conversely to that the removal

118
00:05:35,180 --> 00:05:30,190
of the of the loop removes about 50 fold

119
00:05:38,879 --> 00:05:35,190
activity and there's a trend there's a

120
00:05:41,339 --> 00:05:38,889
log linear trend in activity versus

121
00:05:45,240 --> 00:05:41,349
length motif so we wanted to test the

122
00:05:48,180 --> 00:05:45,250
extent of this to do this we took two

123
00:05:50,089 --> 00:05:48,190
very well-known like it's ribosomes l1

124
00:05:54,029 --> 00:05:50,099
in class 2 like s drivers and they were

125
00:05:55,830 --> 00:05:54,039
old and have been optimized over time

126
00:05:58,260 --> 00:05:55,840
and we place them in the identical

127
00:06:00,930 --> 00:05:58,270
structural context as the ligase is that

128
00:06:06,000 --> 00:06:00,940
we evolved and what we find is that that

129
00:06:08,520 --> 00:06:06,010
the trend still continues we also did a

130
00:06:10,290 --> 00:06:08,530
bit of a literature search and from the

131
00:06:13,740 --> 00:06:10,300
literature search we find that there is

132
00:06:17,219 --> 00:06:13,750
this activity length trend among both

133
00:06:20,040 --> 00:06:17,229
the Optimas so for all functional RNA so

134
00:06:23,670 --> 00:06:20,050
after words such as gtp Optima and also

135
00:06:25,830 --> 00:06:23,680
other ligases so we have here just a

136
00:06:30,860 --> 00:06:25,840
survey of different Lycus activities

137
00:06:33,510 --> 00:06:30,870
that have been published we wanted to

138
00:06:35,790 --> 00:06:33,520

examine the extent of the strength so if

139
00:06:38,879 --> 00:06:35,800
we continue to grow these ligases can we

140
00:06:42,089 --> 00:06:38,889
obtain even higher levels of activity

141
00:06:45,360 --> 00:06:42,099
and to do this weari selected the most

142
00:06:46,260 --> 00:06:45,370
active I guess so the winner the

143
00:06:51,360 --> 00:06:46,270
internal loop

144
00:06:53,700 --> 00:06:51,370
we mutagenize dit in 40 positions for

145
00:06:57,240 --> 00:06:53,710
one of the trajectories and then added

146
00:07:00,570 --> 00:06:57,250
fully us at 20% preposition so that's

147
00:07:03,029 --> 00:07:00,580
it's quite high in my mutagenesis and

148
00:07:04,830 --> 00:07:03,039
then added 20 fully random looking

149
00:07:07,110 --> 00:07:04,840
nucleotides to the top of the loop and

150
00:07:10,770 --> 00:07:07,120
then for the third trajectory I added 20

151
00:07:14,640 --> 00:07:10,780
fully random sequence to the 3-prime act

152
00:07:17,159 --> 00:07:14,650
so what do we get what we find is that

153
00:07:20,399 --> 00:07:17,169
there is a limit to the strenght so what

154
00:07:22,980 --> 00:07:20,409
we get is tenfold improvement in

155
00:07:24,990 --> 00:07:22,990
activity upon this internal loop motif

156
00:07:27,629 --> 00:07:25,000
and the way that is achieved is by

157
00:07:29,309 --> 00:07:27,639
creating a slightly larger loop we don't

158
00:07:31,770 --> 00:07:29,319
know the details of this mechanism but

159
00:07:35,670 --> 00:07:31,780
that that is what we observed we also

160
00:07:38,100 --> 00:07:35,680
get a very similar activity for the same

161
00:07:40,950 --> 00:07:38,110
motif in place in a larger context from

162
00:07:43,409 --> 00:07:40,960
so from the third of the Rhys elections

163
00:07:46,490 --> 00:07:43,419

we obtained this motif which is the

164
00:07:49,620 --> 00:07:46,500
exact same motif just containing

165
00:07:52,800 --> 00:07:49,630
additional sequence element and that has

166
00:07:55,589 --> 00:07:52,810
identical activity but only slightly

167
00:07:57,270 --> 00:07:55,599
higher amplitude in activity so the the

168
00:08:01,439 --> 00:07:57,280
the fraction of the molecules don't like

169
00:08:02,129 --> 00:08:01,449
it is just slightly higher what is

170
00:08:04,320 --> 00:08:02,139
striking

171
00:08:07,379 --> 00:08:04,330
among these selections is that what we

172
00:08:10,200 --> 00:08:07,389
get is that the the structure of the

173
00:08:11,959 --> 00:08:10,210
core motif does not change so only the

174
00:08:15,749 --> 00:08:11,969
peripheral elements change

175
00:08:18,629 --> 00:08:15,759
it's like recur remodeling there at the

176
00:08:21,510 --> 00:08:18,639
top of the loop but the core sequence

177
00:08:24,050 --> 00:08:21,520
remains completely unchanged and the

178
00:08:29,100 --> 00:08:24,060
addition and then the remodeling of the

179
00:08:30,899 --> 00:08:29,110
the top stem loop results in thousand

180
00:08:35,579 --> 00:08:30,909
five thousand and ten thousandfold

181
00:08:36,990 --> 00:08:35,589
improvement and activity so the

182
00:08:39,990 --> 00:08:37,000
take-home message from this and I kind

183
00:08:42,449 --> 00:08:40,000
of went through fast so I didn't talk

184
00:08:45,960 --> 00:08:42,459
much about the sequence networks and

185
00:08:48,900 --> 00:08:45,970
their connectedness but what we find is

186
00:08:50,940 --> 00:08:48,910
that contrary to the theoretical work

187
00:08:53,069 --> 00:08:50,950
that has been proposed the sequence

188
00:08:55,290 --> 00:08:53,079

networks are disconnected at all

189
00:08:56,579 --> 00:08:55,300
different lengths so for the shortened

190
00:08:59,069 --> 00:08:56,589
ribozymes they're definitely

191
00:09:00,030 --> 00:08:59,079
disconnected but what was slightly

192
00:09:01,890 --> 00:09:00,040
surprising was

193
00:09:04,650 --> 00:09:01,900
for the Reese election of the longer

194
00:09:08,190 --> 00:09:04,660
ones we definitely see even smaller

195
00:09:10,860 --> 00:09:08,200
networks in sequence space so the point

196
00:09:12,660 --> 00:09:10,870
point mutation leads to only limited

197
00:09:15,510 --> 00:09:12,670
optimization potentials so it's very

198
00:09:17,370 --> 00:09:15,520
local optimization sequence insertions

199
00:09:19,800 --> 00:09:17,380
so these large type of sequence

200
00:09:21,480 --> 00:09:19,810
insertions like recombination could lead

201
00:09:24,300 --> 00:09:21,490
to initially large improvements in

202
00:09:27,600 --> 00:09:24,310
activity but then sort of last start to

203
00:09:29,250 --> 00:09:27,610
lag and the the most important

204
00:09:31,200 --> 00:09:29,260
observation here is that the sequence

205
00:09:34,260 --> 00:09:31,210
insertions preserve the core structure

206
00:09:36,390 --> 00:09:34,270
which is important for interpreting the

207
00:09:38,820 --> 00:09:36,400
molecular record today so you can

208
00:09:41,049 --> 00:09:38,830
imagine if if we look at modern

209
00:09:45,139 --> 00:09:41,059
structures today

210
00:09:50,509 --> 00:09:45,149
envisage what the earlier functional

211
00:09:52,249 --> 00:09:50,519
structures would have been yeah so I'd

212
00:09:54,799 --> 00:09:52,259
like to thank people that have worked in

213
00:09:56,689 --> 00:09:54,809

this current and past members of the

214
00:09:59,539 --> 00:09:56,699
digital group Theresa and Alex in

215
00:10:01,970 --> 00:09:59,549
particular our collaborators Andrew and

216
00:10:04,970 --> 00:10:01,980
Chen Yu I'd like to thank our funding

217
00:10:12,480 --> 00:10:04,980
sources and like to thank you for your

218
00:10:17,800 --> 00:10:15,100
Thank You milena we have time for some

219
00:10:32,290 --> 00:10:17,810
questions it looks like if we'll start

220
00:10:38,630 --> 00:10:35,810
No Thank You Anton that's that's a great

221
00:10:42,070 --> 00:10:38,640
question we haven't but we are planning

222
00:10:44,630 --> 00:10:42,080
to follow up on this project with

223
00:10:47,630 --> 00:10:44,640
additional selections and looking more

224
00:10:48,830 --> 00:10:47,640
into detail about the types of sequences

225
00:10:51,590 --> 00:10:48,840
and structures that would promote

226
00:10:53,240 --> 00:10:51,600
sequence insertions because we do think

227
00:10:55,760 --> 00:10:53,250
like your work that that sequence

228
00:10:57,980 --> 00:10:55,770
insertions was a an important mechanism

229
00:11:00,200 --> 00:10:57,990
for increasing function early on and

230
00:11:02,060 --> 00:11:00,210
that's something that particularly is

231
00:11:06,110 --> 00:11:02,070
important for RNA I could have led to

232
00:11:08,690 --> 00:11:06,120
RNA it could be the underlying reason

233
00:11:46,690 --> 00:11:08,700
that RNA is wasn't is an important

234
00:11:46,700 --> 00:11:53,350
[Music]

235
00:11:53,360 --> 00:12:15,900
I mean I can imagine Rhys elections for

236
00:12:21,360 --> 00:12:20,079
no I mean there's so many important

237
00:12:25,710 --> 00:12:21,370
reactions that would have been in

238
00:12:31,630 --> 00:12:28,329

replication type of reactions reactions

239
00:12:33,639 --> 00:12:31,640
important over metabolism I'm not

240
00:12:36,329 --> 00:12:33,649
actually quite sure that I understand

241
00:12:37,480 --> 00:12:36,339
your question exactly so maybe we can

242
00:12:38,920 --> 00:12:37,490
okay

243
00:12:52,700 --> 00:12:38,930
kindness is always going to select for

244
00:13:10,980 --> 00:12:55,610
so sometimes people have modified that

245
00:13:22,670 --> 00:13:14,080
you try to translate that length

246
00:13:26,780 --> 00:13:24,439
right right thank you for that question

247
00:13:29,090 --> 00:13:26,790
it's a good question um right so I

248
00:13:30,710 --> 00:13:29,100
didn't explain this fully but the motif

249
00:13:32,359 --> 00:13:30,720
length that we talked about can be

250
00:13:33,920 --> 00:13:32,369
directly translated into the

251
00:13:36,230 --> 00:13:33,930
informational content so this is not the

252
00:13:42,259 --> 00:13:36,240
exact length of the molecule it is the

253
00:13:44,179 --> 00:13:42,269
motif length so all right so with that I