



1  
00:00:00,790 --> 00:00:07,320

[Music]

2  
00:00:11,840 --> 00:00:09,070

[Applause]

3  
00:00:14,660 --> 00:00:11,850

thank you for the invitation today I'm

4  
00:00:17,600 --> 00:00:14,670

going to tell you about a project that's

5  
00:00:19,310 --> 00:00:17,610

a collaboration between my lab including

6  
00:00:22,160 --> 00:00:19,320

Selena Blanco some of you have met her

7  
00:00:26,269 --> 00:00:22,170

here she's here at apps icon as well as

8  
00:00:28,030 --> 00:00:26,279

Robert Pascal's lab and we have also

9  
00:00:30,229 --> 00:00:28,040

essential input from Ellie Muller and

10  
00:00:32,210 --> 00:00:30,239

Jerry Joyce so I'm going to tell you

11  
00:00:34,910 --> 00:00:32,220

about this project which was published a

12  
00:00:39,319 --> 00:00:37,220

okay so we are interested in the RNA

13  
00:00:42,560 --> 00:00:39,329

world and in evolution of the RNA world

14

00:00:44,450 --> 00:00:42,570

so for those of you who haven't thought

15

00:00:46,460 --> 00:00:44,460

about the concept of Fitness landscapes

16

00:00:49,790 --> 00:00:46,470

before let me just explain to you what

17

00:00:52,220 --> 00:00:49,800

we mean so suppose you have a list of

18

00:00:55,580 --> 00:00:52,230

all the possible sequences so 4 to the N

19

00:00:58,819 --> 00:00:55,590

in our case n is 2121 random nucleotides

20

00:01:00,980 --> 00:00:58,829

you write them down in some fashion C on

21

00:01:02,900 --> 00:01:00,990

a sequence coordinate if you know their

22

00:01:04,969 --> 00:01:02,910

Fitness let if their ribozymes maybe you

23

00:01:07,490 --> 00:01:04,979

know their catalytic activity then you

24

00:01:09,920 --> 00:01:07,500

can write down this function in sequence

25

00:01:12,319 --> 00:01:09,930

space and this function is known as the

26  
00:01:14,480 --> 00:01:12,329  
fitness landscape on this landscape the

27  
00:01:16,700 --> 00:01:14,490  
peaks correspond to families of active

28  
00:01:20,059 --> 00:01:16,710  
molecules these troughs correspond to

29  
00:01:22,370 --> 00:01:20,069  
valleys of inactive molecules so natural

30  
00:01:25,580 --> 00:01:22,380  
selection can be thought of rigorously

31  
00:01:27,830 --> 00:01:25,590  
as a random walk over this landscape so

32  
00:01:29,719 --> 00:01:27,840  
a population will start out in some area

33  
00:01:32,929 --> 00:01:29,729  
of sequence space exploring the local

34  
00:01:34,760 --> 00:01:32,939  
area through mutations if a mutant

35  
00:01:36,679 --> 00:01:34,770  
arises it has higher Fitness than the

36  
00:01:39,739 --> 00:01:36,689  
population as a whole moves upwards and

37  
00:01:42,379 --> 00:01:39,749  
so the population kind of climbs up the

38  
00:01:44,989 --> 00:01:42,389

fitness landscape so we think of it as a

39

00:01:48,830 --> 00:01:44,999

random walk with a bias toward climbing

40

00:01:50,359 --> 00:01:48,840

hills so it's looks it looks nice on

41

00:01:51,980 --> 00:01:50,369

this cartoon but we actually have very

42

00:01:54,349 --> 00:01:51,990

little idea of what real Fitness

43

00:01:56,599 --> 00:01:54,359

landscapes look like and we have some

44

00:01:58,760 --> 00:01:56,609

ideas from work exploring relatively

45

00:02:00,440 --> 00:01:58,770

local areas of Fitness landscapes

46

00:02:01,969 --> 00:02:00,450

looking around a known ribozyme for

47

00:02:05,149 --> 00:02:01,979

example in the evolutionary pathways

48

00:02:07,399 --> 00:02:05,159

around a single known ribozyme but what

49

00:02:09,260 --> 00:02:07,409

my lab is interested in doing is mapping

50

00:02:11,210 --> 00:02:09,270

this entire fitness landscape for the

51  
00:02:14,500 --> 00:02:11,220  
entire space so of course we're limited

52  
00:02:17,270 --> 00:02:14,510  
by experimental

53  
00:02:20,149 --> 00:02:17,280  
sizes so we can only look at relatively

54  
00:02:23,539 --> 00:02:20,159  
short sequences fortunately for RNA even

55  
00:02:25,729 --> 00:02:23,549  
relatively short sequences 21 MERS or at

56  
00:02:28,240 --> 00:02:25,739  
least sequences with a random region of

57  
00:02:30,830 --> 00:02:28,250  
21 in length can still be functional so

58  
00:02:34,339 --> 00:02:30,840  
our experimental approach is based on in

59  
00:02:37,699 --> 00:02:34,349  
vitro evolution so we take a pool of

60  
00:02:40,160 --> 00:02:37,709  
random sequences many copies of each

61  
00:02:43,399 --> 00:02:40,170  
possible sequence and then we subject

62  
00:02:45,140 --> 00:02:43,409  
this to a biochemical selection for the

63  
00:02:47,479 --> 00:02:45,150

ribozyme activity we're interested in

64

00:02:50,059 --> 00:02:47,489

the ideas that will reduce the frequency

65

00:02:52,039 --> 00:02:50,069

of sequences that lack this activity and

66

00:02:54,440 --> 00:02:52,049

enriched for sequences that have this

67

00:02:58,069 --> 00:02:54,450

activity and in this way we can pick out

68

00:03:00,400 --> 00:02:58,079

the active sequences okay what are we

69

00:03:02,780 --> 00:03:00,410

interested in in terms of activities

70

00:03:05,930 --> 00:03:02,790

we're curious about the genetic code

71

00:03:08,720 --> 00:03:05,940

like many of us here and there's a lot

72

00:03:10,430 --> 00:03:08,730

of focus on the ribosome and justly so

73

00:03:13,009 --> 00:03:10,440

but the other part of the genetic code

74

00:03:15,589 --> 00:03:13,019

is the hooking up of the proper amino

75

00:03:18,500 --> 00:03:15,599

acids to the property RNAs and sometimes

76

00:03:20,300 --> 00:03:18,510

that's called the second genetic code so

77

00:03:23,030 --> 00:03:20,310

in modern biology this is done by

78

00:03:26,360 --> 00:03:23,040

synthesis the aminoacyl trna synthetases

79

00:03:28,280 --> 00:03:26,370

michael yaris many years ago showed that

80

00:03:31,900 --> 00:03:28,290

it's possible to find ribozymes that can

81

00:03:36,379 --> 00:03:31,910

catalyze this reaction so ribozymes that

82

00:03:38,629 --> 00:03:36,389

attached amino acids to RNA however

83

00:03:39,979 --> 00:03:38,639

before we jumped into the selection we

84

00:03:43,159 --> 00:03:39,989

noticed that there are some problems

85

00:03:46,809 --> 00:03:43,169

which are experimentally difficult to

86

00:03:49,400 --> 00:03:46,819

deal with this amino acyl adenylate the

87

00:03:52,280 --> 00:03:49,410

reactive substrate is highly reactive in

88

00:03:54,349 --> 00:03:52,290

water so it's not thought to be

89

00:03:55,789 --> 00:03:54,359

prebiotic lee plausible because of this

90

00:03:57,849 --> 00:03:55,799

and it's also from a practical

91

00:04:00,559 --> 00:03:57,859

standpoint difficult to use in the lab

92

00:04:03,229 --> 00:04:00,569

so we look to our organic chemistry

93

00:04:05,150 --> 00:04:03,239

colleagues coa lu working in Robert

94

00:04:07,789 --> 00:04:05,160

Pascal's Lau lab at the University of

95

00:04:11,330 --> 00:04:07,799

Montpellier and they had come up with

96

00:04:13,909 --> 00:04:11,340

this form of activated amino acids the

97

00:04:15,619 --> 00:04:13,919

five for H ox as Alone's these are

98

00:04:19,159 --> 00:04:15,629

related to what might be more familiar

99

00:04:20,800 --> 00:04:19,169

to you the and carboxy on hydrates you

100

00:04:23,360 --> 00:04:20,810

can see looking at the structure the

101  
00:04:25,640 --> 00:04:23,370  
part that will become the amine and the

102  
00:04:25,990 --> 00:04:25,650  
carboxylate and the side chain of the

103  
00:04:28,930 --> 00:04:26,000  
amino

104  
00:04:31,660 --> 00:04:28,940  
said this pending grip here is

105  
00:04:33,550 --> 00:04:31,670  
convenient because we we ziwei can put a

106  
00:04:38,050 --> 00:04:33,560  
biotin on there and that way we have

107  
00:04:40,330 --> 00:04:38,060  
this capture handle for reactions ziwei

108  
00:04:41,950 --> 00:04:40,340  
had also noticed that these ox as loans

109  
00:04:45,280 --> 00:04:41,960  
which are pre radically plausible

110  
00:04:48,490 --> 00:04:45,290  
they're formed from simulated volcanic

111  
00:04:50,890 --> 00:04:48,500  
mixtures these ox as loans do react

112  
00:04:54,360 --> 00:04:50,900  
slowly with nucleotides so in this case

113  
00:04:57,850 --> 00:04:54,370

I'm showing a modified tyrosine which is

114

00:04:59,950 --> 00:04:57,860

very slowly reactive with nucleotides so

115

00:05:01,050 --> 00:04:59,960

it's a perfect setup for a selection we

116

00:05:04,360 --> 00:05:01,060

know the reaction is thermodynamically

117

00:05:09,580 --> 00:05:04,370

downhill we want to find a ribozyme that

118

00:05:11,470 --> 00:05:09,590

catalyzes this reaction so the selection

119

00:05:14,080 --> 00:05:11,480

scheme that was devised by a Pressman in

120

00:05:16,390 --> 00:05:14,090

my lab started with the DNA pool

121

00:05:16,810 --> 00:05:16,400

covering sequence space transcribed into

122

00:05:21,340 --> 00:05:16,820

RNA

123

00:05:24,580 --> 00:05:21,350

we're interested in but a few of them

124

00:05:26,650 --> 00:05:24,590

will react with our substrate if they

125

00:05:27,700 --> 00:05:26,660

react they become biotinylated by the

126

00:05:30,640 --> 00:05:27,710

way this is biotin

127

00:05:32,530 --> 00:05:30,650

so these biotin elated RNAs can then be

128

00:05:35,530 --> 00:05:32,540

captured by streptavidin bead pull down

129

00:05:37,810 --> 00:05:35,540

and then we can amplify these by rt-pcr

130

00:05:43,570 --> 00:05:37,820

do high-throughput sequencing follow the

131

00:05:47,230 --> 00:05:43,580

fate of these sequences over time so

132

00:05:49,120 --> 00:05:47,240

that process gives you a list of active

133

00:05:51,190 --> 00:05:49,130

sequences so think of this as a list of

134

00:05:53,110 --> 00:05:51,200

ribozyme sequences which you've kind of

135

00:05:56,230 --> 00:05:53,120

culled from the space of all possible

136

00:05:57,670 --> 00:05:56,240

sequences then our next step that what

137

00:05:59,860 --> 00:05:57,680

we really want to do is associate a

138

00:06:02,320 --> 00:05:59,870

catalytic activity a rate constant let's

139

00:06:04,420 --> 00:06:02,330

say with each of those sequences so to

140

00:06:07,180 --> 00:06:04,430

do this we came up with a method with

141

00:06:09,610 --> 00:06:07,190

input from Willie and Jerry called

142

00:06:12,730 --> 00:06:09,620

kinetic sequencing the idea is that we

143

00:06:15,280 --> 00:06:12,740

take a RNA pool which has medium

144

00:06:17,530 --> 00:06:15,290

diversity it's not so diverse that we

145

00:06:19,630 --> 00:06:17,540

can't get much information about any

146

00:06:21,760 --> 00:06:19,640

particular sequence but if the pool has

147

00:06:23,740 --> 00:06:21,770

not converged so much also to the point

148

00:06:26,050 --> 00:06:23,750

where we only can look at a small number

149

00:06:28,750 --> 00:06:26,060

of sequences so it's a medium diversity

150

00:06:29,980 --> 00:06:28,760

pool we split it into several elec watts

151

00:06:32,980 --> 00:06:29,990

and react it with different

152

00:06:34,300 --> 00:06:32,990

concentrations of our substrate you

153

00:06:36,240 --> 00:06:34,310

could also think of doing this with

154

00:06:38,650 --> 00:06:36,250

different time points but substrate

155

00:06:39,800 --> 00:06:38,660

concentrations were easier for us for

156

00:06:42,740 --> 00:06:39,810

technical reasons

157

00:06:45,080 --> 00:06:42,750

and then we react them and capture them

158

00:06:47,990 --> 00:06:45,090

capture the reacted molecules sequence

159

00:06:50,750 --> 00:06:48,000

those captured molecules and then that

160

00:06:53,240 --> 00:06:50,760

gives us basically four points along

161

00:06:55,310 --> 00:06:53,250

this kinetic curve for different

162

00:06:57,070 --> 00:06:55,320

concentrations and that allows us to

163

00:06:59,510 --> 00:06:57,080

calculate the rate constant for these

164

00:07:00,980 --> 00:06:59,520

sequences and depending on how deeply

165

00:07:03,500 --> 00:07:00,990

you sequence you could get information

166

00:07:07,310 --> 00:07:03,510

for thousands to perhaps hundreds of

167

00:07:09,520 --> 00:07:07,320

thousands of molecules this kinetics

168

00:07:12,409 --> 00:07:09,530

sequencing scheme works pretty well at

169

00:07:14,840 --> 00:07:12,419

predicting the activity here's a

170

00:07:18,200 --> 00:07:14,850

correlation between the measurement by

171

00:07:21,560 --> 00:07:18,210

KC kinetic sequencing and a traditional

172

00:07:23,150 --> 00:07:21,570

gel shift assay okay

173

00:07:24,800 --> 00:07:23,160

so now we have this list of ribozyme

174

00:07:27,230 --> 00:07:24,810

sequences and their associated rate

175

00:07:28,880 --> 00:07:27,240

constants what we want to do is answer

176

00:07:31,760 --> 00:07:28,890

some interesting questions about this

177

00:07:33,890 --> 00:07:31,770

fitness landscape one interesting

178

00:07:35,330 --> 00:07:33,900

question is how smooth is this landscape

179

00:07:38,510 --> 00:07:35,340

and the reason why this is interesting

180

00:07:39,920 --> 00:07:38,520

is if you had a very smooth landscape

181

00:07:42,680 --> 00:07:39,930

for this kind of an extreme example a

182

00:07:44,330 --> 00:07:42,690

single peak very smooth landscape then

183

00:07:47,000 --> 00:07:44,340

you can imagine no matter where you

184

00:07:51,710 --> 00:07:47,010

start on this landscape you can the

185

00:07:54,500 --> 00:07:51,720

population can feel the the peak so you

186

00:07:56,750 --> 00:07:54,510

can do a very smooth walk up to the top

187

00:07:59,690 --> 00:07:56,760

and optimize find the global optimum

188

00:08:01,430 --> 00:07:59,700

over sequence space on the other hand if

189

00:08:04,219 --> 00:08:01,440

you have a highly rugged landscape like

190

00:08:05,930 --> 00:08:04,229

this then kind of depending on where you

191

00:08:08,779 --> 00:08:05,940

start in the sequence pace you might

192

00:08:11,420 --> 00:08:08,789

make a local climb to what you think is

193

00:08:13,820 --> 00:08:11,430

the top of ap or what is the top of a

194

00:08:16,370 --> 00:08:13,830

peak but it you you've missed the global

195

00:08:18,409 --> 00:08:16,380

optimum so we were very interested in

196

00:08:21,200 --> 00:08:18,419

understanding these potential

197

00:08:23,659 --> 00:08:21,210

evolutionary pathways here's an example

198

00:08:28,310 --> 00:08:23,669

of the type of pathways that we find so

199

00:08:30,350 --> 00:08:28,320

over here is one motif motif one of the

200

00:08:33,440 --> 00:08:30,360

ribozymes that we found ribozyme 1b

201  
00:08:37,219 --> 00:08:33,450  
marked in blue over here motif 2 over

202  
00:08:40,190 --> 00:08:37,229  
here you can see motif sequence 2.1 is

203  
00:08:42,589 --> 00:08:40,200  
the global optimum of the landscape so

204  
00:08:44,570 --> 00:08:42,599  
what we can observe is that looking at

205  
00:08:46,760 --> 00:08:44,580  
the activity of these sequences which we

206  
00:08:48,920 --> 00:08:46,770  
measured by kasich and the evolutionary

207  
00:08:51,590 --> 00:08:48,930  
distances we can find pathways that

208  
00:08:53,780 --> 00:08:51,600  
connect relatively related sequences you

209  
00:08:57,380 --> 00:08:53,790  
could call this motif 1b

210  
00:08:59,390 --> 00:08:57,390  
and one a local Optima which are

211  
00:09:02,810 --> 00:08:59,400  
connected by reasonable evolutionary

212  
00:09:05,900 --> 00:09:02,820  
pathways but going from this motif one

213  
00:09:08,900 --> 00:09:05,910

area to motif two you have to go through

214

00:09:10,550 --> 00:09:08,910

the large valley of sequences that have

215

00:09:13,600 --> 00:09:10,560

basically no activity or at least

216

00:09:16,910 --> 00:09:13,610

baseline activity so this tells us that

217

00:09:19,640 --> 00:09:16,920

only nearby Peaks are connected by

218

00:09:22,850 --> 00:09:19,650

viable pathways if we zoom out to the

219

00:09:24,680 --> 00:09:22,860

entire landscape so I'm not showing all

220

00:09:25,910 --> 00:09:24,690

the different ribozymes that we found

221

00:09:27,980 --> 00:09:25,920

there would be thousands of points if

222

00:09:31,430 --> 00:09:27,990

that were the case but what I'm showing

223

00:09:33,620 --> 00:09:31,440

you is the top motifs that we found

224

00:09:36,590 --> 00:09:33,630

which are shown by these big circles and

225

00:09:41,960 --> 00:09:36,600

the best five pathways that we found

226  
00:09:45,200 --> 00:09:41,970  
between these major ribozyme centers so

227  
00:09:46,000 --> 00:09:45,210  
what you can see is sequence 2.1 high

228  
00:09:51,830 --> 00:09:46,010  
activity

229  
00:09:54,440 --> 00:09:51,840  
kind of low-lying plateau I would say of

230  
00:09:56,210 --> 00:09:54,450  
motif one where there might be some

231  
00:09:57,920 --> 00:09:56,220  
reasonable interconnections but you have

232  
00:10:00,500 --> 00:09:57,930  
to look a little bit hard for them but

233  
00:10:02,570 --> 00:10:00,510  
they exist but motif two in particular

234  
00:10:05,060 --> 00:10:02,580  
is kind of cut off from the rest of this

235  
00:10:07,040 --> 00:10:05,070  
landscape by multiple mutations these

236  
00:10:10,040 --> 00:10:07,050  
dotted lines indicate multiple mutations

237  
00:10:13,580 --> 00:10:10,050  
required multiple mutations required at

238  
00:10:18,290 --> 00:10:13,590

essentially no activity to traverse this

239

00:10:20,150 --> 00:10:18,300

part of the landscape so with that I'd

240

00:10:23,090 --> 00:10:20,160

like to just close by acknowledging the

241

00:10:25,610 --> 00:10:23,100

people involved in this work Abe who led

242

00:10:27,830 --> 00:10:25,620

this project and a celiac who

243

00:10:30,530 --> 00:10:27,840

contributed to a recognized analysis I

244

00:10:32,870 --> 00:10:30,540

didn't have time to talk about and Evan

245

00:10:34,700 --> 00:10:32,880

can contribute to some experimental

246

00:10:37,100 --> 00:10:34,710

validation which I also didn't have time

247

00:10:39,590 --> 00:10:37,110

to talk about and I'd like to again take

248

00:10:45,090 --> 00:10:39,600

our collaborators ziwei and Ribera and

249

00:10:56,519 --> 00:10:49,350

I think we have time for one question

250

00:10:58,889 --> 00:10:56,529

that was a really interesting talk thank

251  
00:11:00,389 --> 00:10:58,899  
you I just wondered in in terms of

252  
00:11:02,040 --> 00:11:00,399  
building up her evolutionary landscape

253  
00:11:03,840 --> 00:11:02,050  
whether you're going to build in looking

254  
00:11:05,910 --> 00:11:03,850  
at the possible role of mobile genetic

255  
00:11:08,220 --> 00:11:05,920  
elements which could potentially cause a

256  
00:11:11,009 --> 00:11:08,230  
discontinuity in evolution that could

257  
00:11:13,680 --> 00:11:11,019  
bridge some of those gaps okay yes so we

258  
00:11:16,620 --> 00:11:13,690  
completely I really oversimplified this

259  
00:11:20,280 --> 00:11:16,630  
as picture saying that we're just

260  
00:11:22,920 --> 00:11:20,290  
looking at the step by step mutations we

261  
00:11:25,590 --> 00:11:22,930  
have in our pathfinding algorithm you

262  
00:11:27,540 --> 00:11:25,600  
can allow larger jumps in sequence space

263  
00:11:29,939 --> 00:11:27,550

and you get pretty much the same picture

264

00:11:34,050 --> 00:11:29,949

if you allow up to four mutations but if

265

00:11:35,730 --> 00:11:34,060

you allow quite large gaps then you can

266

00:11:40,650 --> 00:11:35,740

start to see kind of an it completely