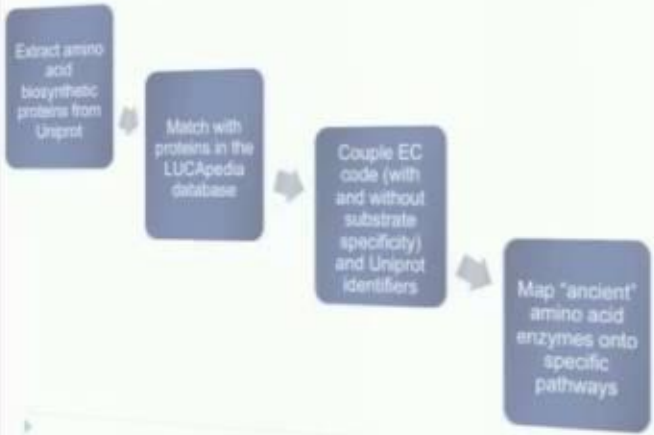
Finding Potential Pathways of Ancient Amino Acid Biosynthesis



1
00:00:11,660 --> 00:00:08,900
yeah my name is Kenny I know it's really

2
00:00:12,890 --> 00:00:11,670
difficult sorry and I guess I'm one of

3
00:00:15,619 --> 00:00:12,900
the people that you don't envy because

4
00:00:20,150 --> 00:00:15,629
I'm going to try to talk about amino

5
00:00:21,769 --> 00:00:20,160
acid biosynthetic pathways and so amino

6
00:00:23,810 --> 00:00:21,779
acids are critical and that they are the

7
00:00:25,910 --> 00:00:23,820
constituents of the proteins which carry

8
00:00:27,320 --> 00:00:25,920
out the functions in a cell and it's

9
00:00:29,810 --> 00:00:27,330
thought that the last Universal common

10
00:00:33,110 --> 00:00:29,820
ancestor or the root of the tree of life

11
00:00:36,560 --> 00:00:33,120
is able to prevail to produce all

12
00:00:38,239 --> 00:00:36,570
currently available amino acids so I'm

13
00:00:40,569 --> 00:00:38,249

interested in characterizing things at

14
00:00:42,500 --> 00:00:40,579
the last Universal common ancestor and

15
00:00:44,059 --> 00:00:42,510
obviously in the context of the

16
00:00:50,000 --> 00:00:44,069
progression of the central dogma or

17
00:00:52,630 --> 00:00:50,010
these other molecular processes so uh so

18
00:00:55,610 --> 00:00:52,640
the first question that I had with this

19
00:00:58,399 --> 00:00:55,620
project is whether the conservation of

20
00:01:00,919 --> 00:00:58,409
proteins associated with amino acid

21
00:01:02,959 --> 00:01:00,929
biosynthesis would be related to a

22
00:01:08,750 --> 00:01:02,969
characteristic of the amino acid like

23
00:01:10,940 --> 00:01:08,760
the complexity of them so I I have a by

24
00:01:14,900 --> 00:01:10,950
automatic approach where I compare

25
00:01:18,260 --> 00:01:14,910
features across species or the whole

26
00:01:20,900 --> 00:01:18,270
tree of life based on either the gene or

27
00:01:23,240 --> 00:01:20,910
protein level and extrapolate this

28
00:01:28,040 --> 00:01:23,250
information on to the last Universal

29
00:01:30,740 --> 00:01:28,050
common ancestor so a super cool a

30
00:01:32,240 --> 00:01:30,750
postdoc in the lab Aaron who seems to be

31
00:01:35,390 --> 00:01:32,250
here in spirit is he got a shout out

32
00:01:38,930 --> 00:01:35,400
earlier to curated this database called

33
00:01:40,910 --> 00:01:38,940
Luca pedia which basically is a

34
00:01:42,380 --> 00:01:40,920
combination of several different data

35
00:01:44,330 --> 00:01:42,390
sets that attempt to characterize

36
00:01:46,160 --> 00:01:44,340
ancient characteristics of proteins

37
00:01:49,430 --> 00:01:46,170
based on these universal type

38
00:01:53,360 --> 00:01:49,440

distributions and they use different

39
00:01:55,100 --> 00:01:53,370
approaches so there's different types of

40
00:01:58,030 --> 00:01:55,110
ways you can approach this obviously you

41
00:02:01,240 --> 00:01:58,040
can do it by gene families by

42
00:02:04,070 --> 00:02:01,250
universally distributed protein motifs

43
00:02:06,500 --> 00:02:04,080
some of these studies use different

44
00:02:09,080 --> 00:02:06,510
structural perspectives as in protein

45
00:02:12,500 --> 00:02:09,090
folds or groups of protein fold

46
00:02:17,059 --> 00:02:12,510
functions and common reactions across

47
00:02:19,370 --> 00:02:17,069
the domains of life so in terms of

48
00:02:24,320 --> 00:02:19,380
complexity one of the reference

49
00:02:26,150 --> 00:02:24,330
I'm using his proxy is the abundance of

50
00:02:28,310 --> 00:02:26,160
certain prebiotic synthesis experiments

51
00:02:30,830 --> 00:02:28,320
so here I just have a table of the

52
00:02:32,450 --> 00:02:30,840
genetic code where I have the codon on

53
00:02:36,020 --> 00:02:32,460
the left and the corresponding amino

54
00:02:37,640 --> 00:02:36,030
acid on the right and the shading just

55
00:02:41,540 --> 00:02:37,650
corresponds with higher abundance in

56
00:02:44,930 --> 00:02:41,550
these experiments so I use a pretty

57
00:02:46,760 --> 00:02:44,940
simple diaphragmatic pipeline I start

58
00:02:48,950 --> 00:02:46,770
with a set of amino acid related

59
00:02:50,210 --> 00:02:48,960
proteins and I match these with things

60
00:02:53,690 --> 00:02:50,220
that I found in at least three of the

61
00:02:56,120 --> 00:02:53,700
studies in the database and I couple

62
00:02:57,650 --> 00:02:56,130
that with some ID matching which i use

63
00:03:01,130 --> 00:02:57,660

the ends on commission code for these

64
00:03:03,290 --> 00:03:01,140
proteins which is just a annotation for

65
00:03:05,690 --> 00:03:03,300
classes of enzyme function where you

66
00:03:08,120 --> 00:03:05,700
have four digits in each digit confers

67
00:03:09,920 --> 00:03:08,130
more specificity and description of the

68
00:03:11,780 --> 00:03:09,930
function of the enzyme and I did this

69
00:03:13,880 --> 00:03:11,790
within without the last digit which

70
00:03:17,450 --> 00:03:13,890
usually describes some substrate

71
00:03:20,300 --> 00:03:17,460
specificity of the enzyme so then I've

72
00:03:23,240 --> 00:03:20,310
mapped these things onto whoa then I met

73
00:03:24,920 --> 00:03:23,250
these things onto specific pathways to

74
00:03:30,320 --> 00:03:24,930
see where things were conserved in pasay

75
00:03:32,150 --> 00:03:30,330
puces so here's a one really cool

76
00:03:34,460 --> 00:03:32,160
pathway and what you're looking at in

77
00:03:36,410 --> 00:03:34,470
red throughout the pathway of the enzyme

78
00:03:39,380 --> 00:03:36,420
the three digit enzyme Commission codes

79
00:03:40,730 --> 00:03:39,390
that are conserved or match three at

80
00:03:43,820 --> 00:03:40,740
least three of these papers and the data

81
00:03:45,950 --> 00:03:43,830
set in the database and it looks a

82
00:03:48,199 --> 00:03:45,960
little daunting at first but if I peel

83
00:03:50,240 --> 00:03:48,209
off things that are not conserved you

84
00:03:52,850 --> 00:03:50,250
can see that there are these functional

85
00:03:56,390 --> 00:03:52,860
cores that seem to be conserved

86
00:03:58,250 --> 00:03:56,400
throughout the pathway so while the main

87
00:04:01,010 --> 00:03:58,260
products in this pathway are sharing

88
00:04:03,199 --> 00:04:01,020

glycine and threonine you can still see

89
00:04:07,150 --> 00:04:03,209
some really close metabolic proximity as

90
00:04:09,680 --> 00:04:07,160
and steps of reactions to get to other

91
00:04:12,260 --> 00:04:09,690
amino acids there's tryptophan over here

92
00:04:14,900 --> 00:04:12,270
since a cytosine then methenamine over

93
00:04:20,060 --> 00:04:14,910
here and other amino acid biosynthetic

94
00:04:23,210 --> 00:04:20,070
pathways as well as metabolism so when I

95
00:04:25,540 --> 00:04:23,220
looked at the main products of this

96
00:04:28,640 --> 00:04:25,550
pathway in in terms of these abundance

97
00:04:31,360 --> 00:04:28,650
proxies there's no direct correlation

98
00:04:32,980 --> 00:04:31,370
between the conservation and the

99
00:04:36,910 --> 00:04:32,990
complexity or

100
00:04:40,600 --> 00:04:36,920
of the amino acid which was at first not

101
00:04:43,060 --> 00:04:40,610
what I expected so another path or that

102
00:04:45,520 --> 00:04:43,070
is probably really representative of the

103
00:04:49,390 --> 00:04:45,530
convergence and not entirely surprising

104
00:04:51,790 --> 00:04:49,400
is valine isoleucine and loosing pathway

105
00:04:54,190 --> 00:04:51,800
in which the pathways are identical

106
00:04:57,850 --> 00:04:54,200
until you get to this last step of this

107
00:04:59,950 --> 00:04:57,860
last reaction so that's probably the

108
00:05:05,260 --> 00:04:59,960
best example of convergent pathways and

109
00:05:07,390 --> 00:05:05,270
given these sets so throughout their

110
00:05:09,700 --> 00:05:07,400
data oh man so throughout the data

111
00:05:12,340 --> 00:05:09,710
there's a there were several nodal

112
00:05:15,460 --> 00:05:12,350
proteins or consistently represented

113
00:05:17,200 --> 00:05:15,470

enzymes in all of these pathways or most

114
00:05:19,090 --> 00:05:17,210
of these pathways and that was

115
00:05:22,900 --> 00:05:19,100
depends synthase alpha chain and searing

116
00:05:24,400 --> 00:05:22,910
methyl ease and right so like I said

117
00:05:26,920 --> 00:05:24,410
their presidents everly these super

118
00:05:29,410 --> 00:05:26,930
pathways and they both um convert

119
00:05:30,850 --> 00:05:29,420
different functions to prevent synthase

120
00:05:33,220 --> 00:05:30,860
obviously the last two sets of

121
00:05:35,170 --> 00:05:33,230
tryptophan biosynthesis and searing

122
00:05:37,390 --> 00:05:35,180
methylase catalyzing the searing the

123
00:05:39,160 --> 00:05:37,400
glycine reaction as well as hydrolysis

124
00:05:41,530 --> 00:05:39,170
of tetrahydrofolate which is just the

125
00:05:44,920 --> 00:05:41,540
common cofactor in amino acid metabolism

126
00:05:47,860 --> 00:05:44,930
as well as nucleotide metabolism so I

127
00:05:49,620 --> 00:05:47,870
tried to find of models of enzyme

128
00:05:52,960 --> 00:05:49,630
evolution that would possibly fit the

129
00:05:54,610 --> 00:05:52,970
data I have and the patchwork model

130
00:05:57,220 --> 00:05:54,620
seems to be a really popular model that

131
00:06:00,340 --> 00:05:57,230
may correspond where you start out with

132
00:06:04,110 --> 00:06:00,350
these initially broadly reactive enzymes

133
00:06:08,110 --> 00:06:04,120
like red green and blue pac-man shape

134
00:06:09,280 --> 00:06:08,120
thing and you they do sir favor a

135
00:06:12,280 --> 00:06:09,290
certain reaction but they're still

136
00:06:14,620 --> 00:06:12,290
broadly reactive and after gene

137
00:06:17,440 --> 00:06:14,630
duplication the least thing and

138
00:06:20,140 --> 00:06:17,450

selection from the environment things

139
00:06:21,910 --> 00:06:20,150
become more specific in their function

140
00:06:24,610 --> 00:06:21,920
so I thought maybe this could help

141
00:06:27,540 --> 00:06:24,620
explain or maybe I'm seeing an artifact

142
00:06:30,610 --> 00:06:27,550
of this with these multifunctional nodes

143
00:06:33,360 --> 00:06:30,620
also another popular theory is a semi

144
00:06:35,800 --> 00:06:33,370
enzymatic theory where you have

145
00:06:39,150 --> 00:06:35,810
reactions that become linked some way

146
00:06:41,440 --> 00:06:39,160
and also it incorporates the use of

147
00:06:46,170 --> 00:06:41,450
spontaneous reactions and development of

148
00:06:48,360 --> 00:06:46,180
these processes so I guess my main

149
00:06:50,610 --> 00:06:48,370
inclusions are that you are able to

150
00:06:51,870 --> 00:06:50,620
identify certain functional chords that

151
00:06:56,129 --> 00:06:51,880
are conserved throughout these pathways

152
00:06:57,960 --> 00:06:56,139
and possibly their support for current

153
00:07:01,830 --> 00:06:57,970
models of enzyme evolution with this

154
00:07:04,290 --> 00:07:01,840
data with these data and I have to

155
00:07:06,150 --> 00:07:04,300
acknowledge my pile or landlubber other

156
00:07:08,760 --> 00:07:06,160
members of the lab do which are in the

157
00:07:10,379 --> 00:07:08,770
audience but not in this picture and of

158
00:07:12,060 --> 00:07:10,389
course Aaron Goldman who is responsible

159
00:07:20,999 --> 00:07:12,070
for curating the database that I worked

160
00:07:27,330 --> 00:07:21,009
so closely with so thank have any

161
00:07:32,610 --> 00:07:27,340
questions for Kinner II oh okay well

162
00:07:35,490 --> 00:07:32,620
then I have one um I don't know a whole

163
00:07:37,230 --> 00:07:35,500

lot about where in the cells amino acid

164
00:07:39,450 --> 00:07:37,240
biosynthesis take place if it's all in

165
00:07:42,960 --> 00:07:39,460
one place or if it's spread out is there

166
00:07:44,790 --> 00:07:42,970
any locational tendencies I realize it

167
00:07:47,879 --> 00:07:44,800
might not be a fair question well I'm

168
00:07:49,620 --> 00:07:47,889
I'm not quite sure I don't and this is

169
00:07:53,580 --> 00:07:49,630
probably personal I don't think of it as

170
00:07:59,430 --> 00:07:53,590
a localized reset of reactions but not