How to map mass spectra onto a genome

and a little tail from RNAcode

John Anders

University Leipzig

Leipzig – February, 2020

- The study subject SIHUMI
- The ms-thingy
- Mapping: A simple task?
- "Quality" of data bases
- Mapping as in UCSC
- RNAcode tails

- The study subject SIHUMI
- The ms-thingy
- Mapping: A simple task?
- "Quality" of data bases
- Mapping as in UCSC
- RNAcode tails

- The study subject SIHUMI
- The ms-thingy
- Mapping: A simple task?
- "Quality" of data bases
- Mapping as in UCSC
- RNAcode tails

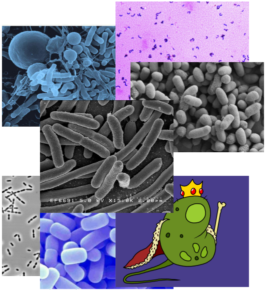
- The study subject SIHUMI
- The ms-thingy
- Mapping: A simple task?
- "Quality" of data bases
- Mapping as in UCSC
- RNAcode tails

- The study subject SIHUMI
- The ms-thingy
- Mapping: A simple task?
- "Quality" of data bases
- Mapping as in UCSC
- RNAcode tails

- The study subject SIHUMI
- The ms-thingy
- Mapping: A simple task?
- "Quality" of data bases
- Mapping as in UCSC
- RNAcode tails

- The study subject SIHUMI
- The ms-thingy
- Mapping: A simple task?
- "Quality" of data bases
- Mapping as in UCSC
- RNAcode tails

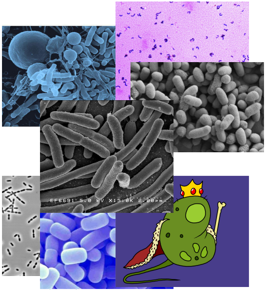
simplified human intestinal microbiota
(SIHUMI)
[Becker 2001]



from indiamart.com

simplified human intestinal microbiota
(SIHUMI)
[Becker 2001]

Anaerostipes caccae
Bacteroides thetaiotaomicron
Bifidobacterium longum
Blautia producta
Clostridium ramosum
Escherichia coli and
Lactobacillus plantarum

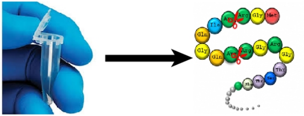


from indiamart.com

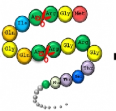
simplified human intestinal microbiota
(SIHUMI)
[Becker 2001]

Anaerostipes caccae
Bacteroides thetaiotaomicron
Bifidobacterium longum
Blautia producta
Clostridium ramosum
Escherichia coli and
Lactobacillus plantarum

Protein

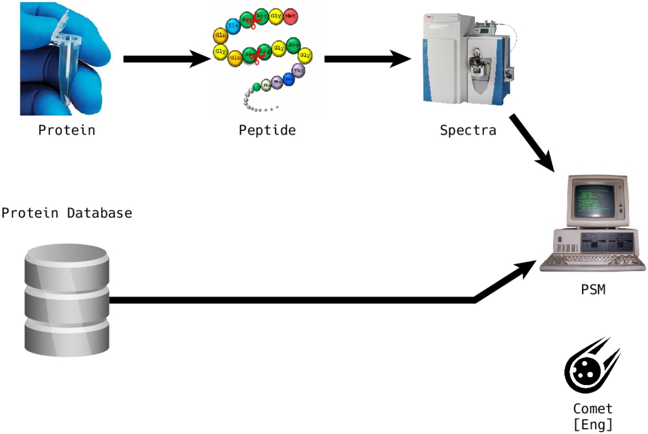

Protein
Peptide

Protein



Peptide

Spectra



Protein

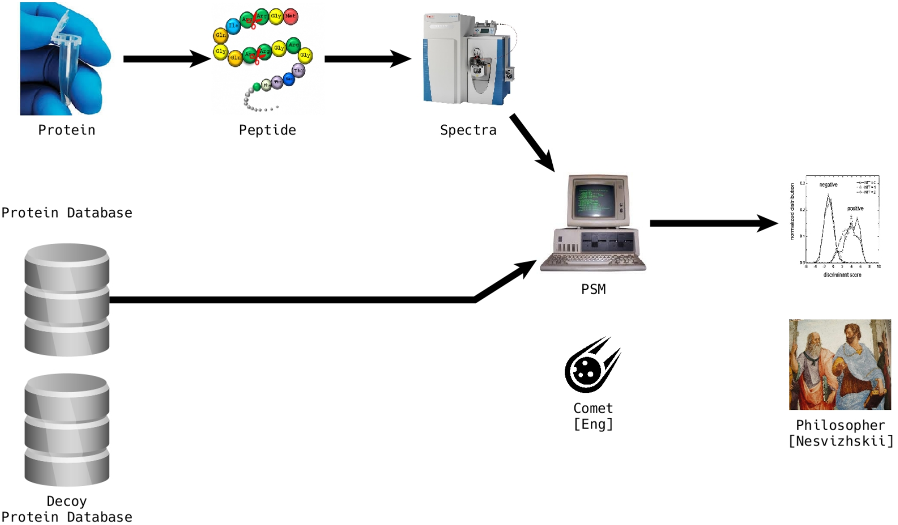
Peptide

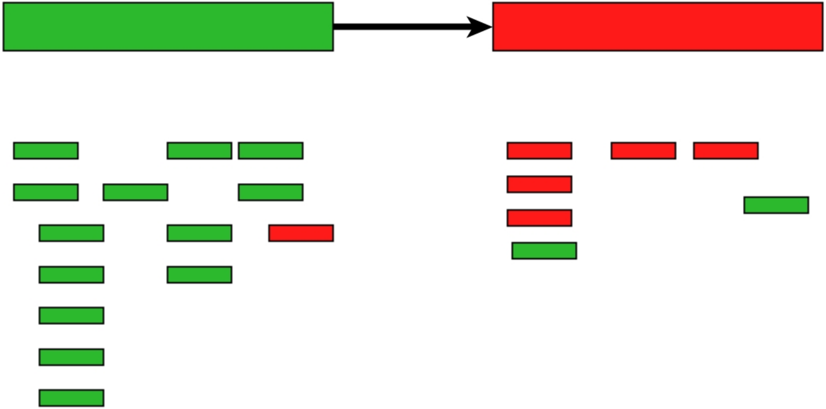
Spectra

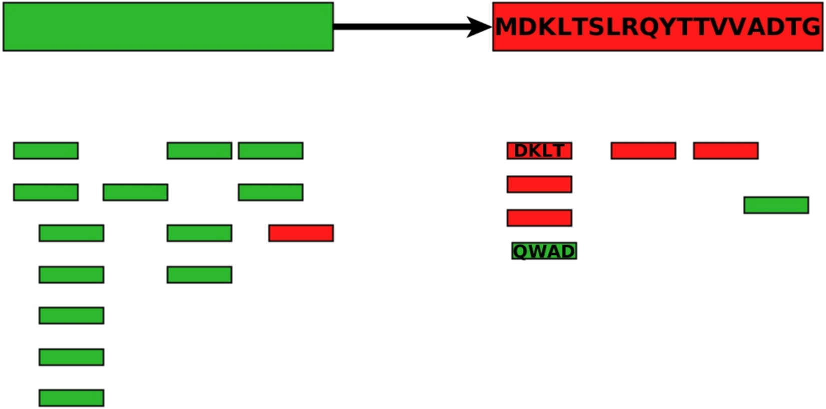
Protein Database

PSM

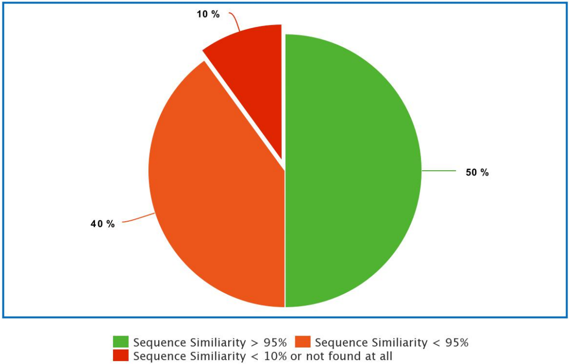
Comet
[Eng]


Protein
Peptide
Spectra
Protein Database
PSM
Decoy
Protein Database
negative
positive
normalized distribution
discriminant score
Comet
[Eng]
Philosopher
[Nesvizhskii]




MDKLTSLRQYTTVVADTG
DKLT
QWAD


UniProt



meta-chart.com

annotated protein Escherichia coli K-12 by NCBI

```
Genome    MFVPFLIMLREGLEAALIVSLIASYLKRTQRG-----RWI-VXCGLACCLPL
          ||||||||||||||||||||||||||||||||-----X||-|XXXX|X||X|
Database  MFVPFLIMLREGLEAALIVSLIASYLKRTQRGMXDCVMWIGVLLAAALCLGL

Genome    ---------------------RCAWAWVSSLTKPPANFRKKNRNCLKVSWRX
          ---------------------XX|XXX|XX||XXXXXX||XX||-X||XXXX
Database  GIFINETTGEFPQKEQELFEGIVAVIAVVILTWMVFWMRKVSRN-VKVQLEQ

Genome    SPWXSL-PGWFSGCAKCRATSKCNWNRQSIAHCSVEIIMAGRWXXWS--FL-
          XXXXX|-X|XXX|X|---------------XXXX|XXXX|-|XXXX|--||-
Database  AVDSALQRGNHHGWA---------------LVMMVFFAVA-REGLESVFFLL

Genome    --PLQGKGWSRSFSCWRHFNKMSGSGRRWVQCSVLLLPWCXASCSTGAVFAS
          --XX|XX|------X|XXXXX|X|XXXXX|--XXX||X|XXXXXXX||X|XX
Database  AAFQQDVG------IWPPLGAMLGLATAVV--LGFLLYWGGIRLNLGAFFKW

Genome    ILVHFLNGPACL---------------------FSSSPQGWQLVPFAHFMKP
          XXXXX|XXX|X|---------------------|XX|XXXXXXXX|XXX|XX
Database  TSLFILFVAAGLAAGAIRAFHEAGLWNHFQEIAFDMSAVLSTHSLFGTLMEG

Genome    DCGTTFRKSPSIXVRCSQLTRCLARXWKGFLAIRKRRASAKSPSGLFISSRR
          XX|XXXXX|X|-----------XXXX|XXX|-|XXXX|X|XX|XXXXXX||.
Database  IFGYQEAPSVS-----------EVAVWFIYL-IPALVAFALPPRAGATASRS

Genome    WWHLLCHHAQGRQRLAPR
          ..................
Database  A.................
```


Window Position
Blautia producta Blautia producta CP039126:1,942,221-1,942,789 (569 bp)
Scale
200 bases
blautia
CP039126:
1,942,250
1,942,300
1,942,350
1,942,400
1,942,450
1,942,500
1,942,550
1,942,600
1,942,650
1,942,700
1,942,750
<---
Annotation
1;1942067-1942286
anno|CP039126|CDS|blautia;CP039126|WP_003574538.1;-1;1942777-1943164
RNAcode
coverage
667
Transcriptome 1
Trans 1
1
6182
Transcriptome 2
Trans 2
4
Mass Spectrometry steffi_6frame_new_SIHUMI
Mass Spectrometry SP3_Nachmessung_C8_6frame_new_SIHUMI
Mass Spectrometry AspN_GelFree_6frame_new_SIHUMI
Mass Spectrometry SP3_Nachmessung_SP3_6frame_new_SIHUMI
Mass Spectrometry method_comparison_InSolution_6frame_new_SIHUMI
Mass Spectrometry AspN_C8_6frame_new_SIHUMI
Mass Spectrometry method_comparison_FASP_6frame_new_SIHUMI
Mass Spectrometry method_comparison_C8_6frame_new_SIHUMI
Mass Spectrometry method_comparison_GelFree_6frame_new_SIHUMI
Mass Spectrometry hannes_6frame_new_SIHUMI

`tl,dr`

$$RNAcode(multi\_nuc\_Seq\_Alig) \rightarrow codingPotential$$

tl,dr

$RNAcode(multi\_nuc\_Seq\_Alig) \rightarrow codingPotential$

Species	Genome Alligment	Blast
anaero	???%	???%
ecoli	???%	???%
ery	???%	???%
blautia	???%	???%
clostri	???%	???%
lacto	???%	???%
bifi	???%	???%
bact	???%	???%
haloferax	???%	???%

Percent of coding genes which can be recalled

tl,dr

$RNAcode(multi\_nuc\_Seq\_Alig) \rightarrow codingPotential$

Species	Genome Alligment	Blast
anaero	69%	???%
ecoli	90%	???%
ery	81%	???%
blautia	78%	???%
clostri	80%	???%
lacto	90%	???%
bifi	87%	???%
bact	79%	???%
haloferax	64%	???%

Percent of coding genes which can be recalled

tl,dr

$$RNAcode(multi\_nuc\_Seq\_Alig) \rightarrow codingPotential$$

Species	Genome Alligment	Blast
anaero	69%	16%
ecoli	90%	18%
ery	81%	17%
blautia	78%	52%
clostri	80%	59%
lacto	90%	???%
bifi	87%	65%
bact	79%	15%
haloferax	64%	48%

Percent of coding genes which can be recalled

Thank you for your attention!