

Bioinformatika

pro PrfUK 2006

Jiří Vondrášek

Ústav organické chemie a biochemie

vondrasek@uochb.cas.cz



Jan Pačes

Ústav molekulární genetiky

hpaces@img.cas.cz



<http://bio.img.cas.cz/PrfUK2006>

syllabus

Úterý, 16:30 CH5

L1 Úvod do bioinformatiky	10. říjen 2006
L2 Biologické databáze (<i>Pačes</i>)	17. říjen
L6 Vlastnosti proteinů (<i>Vondrášek</i>)	24. říjen ?
L7 Sekundární motivy a jejich predikce (<i>Vondrášek</i>)	31. říjen
L8 3D struktury, strukturní alignment (<i>Vondrášek</i>)	7. listopad
L9 Protein folding (<i>Vondrášek</i>)	14. listopad
L3 Alignment I (<i>Pačes</i>)	21. listopad
L4 Alignment II (<i>Pačes</i>)	28. listopad
L5 Predikce genů, fylogenetická analýza (<i>Pačes</i>)	5. prosinec ?
L10 Molekulární docking, drug design (<i>Vondrášek</i>)	12. prosinec
L11 Statistický aparát bioinformatiky (<i>Vondrášek</i>)	18. prosinec

co je bioinformatika?



- ▶ Home
- ▶ Word of the Day
- ▶ Word Games
- ▶ Word for the Wise
- ▶ Books and CDs
- ▶ Company Info
- ▶ Customer Service
- ▶ Network Options
- ▶ Language Zone
- ▶ The Lighter Side
- ▶ Site Map



Merriam-Webster's
Unabridged
Now on CD-ROM!

Merriam-Webster's SPELLING HELP



Powered by Franklin Electronic Publishers

Download the new
Merriam-Webster
Toolbar and look up
words anywhere on the
Web — it's free!

The word you've entered isn't in the dictionary. Click on a spelling suggestion below or try again using the Dictionary search box to the right.

Suggestions for **bioinformatics**:

1. [bioenergetics](#)
2. [bioenergetic](#)
3. [Bonapartism](#)
4. [biomathematics](#)
5. [beneficiaries](#)
6. [biometrical](#)
7. [Bonapartisms](#)
8. [Bonapartist](#)
9. [benefactor](#)
10. [biometeorology](#)

Dictionary

Thesaurus

Put a
Merriam-Webster dictionary
on your Palm Pilot!

Get the **Word of the Day** e-mailed every morning. It's

CAMBRIDGE | Dictionaries Online

Search Contact us



...teaching tips,
articles and interviews



Search Results

bioinformatics was not found in the *Cambridge International Dictionary of English*

Did you spell it correctly? Here are some alternatives:

- bombard
- bombarded
- bombardier
- bombardiers
- bombarding
- bombardment
- bombardments
- bombards
- bomber
- bombers
- bonfire
- bonfires
- bumper
- bumpers
- bumpier

Additional Hypertext Webster Gateway Looku

Enter word here:

Exact Approx

Hypertext Webster Gateway Error

No definition found for "bioinformatics"
You may wish to try an alternative spelling (change what you had in the box)

bioinformatika

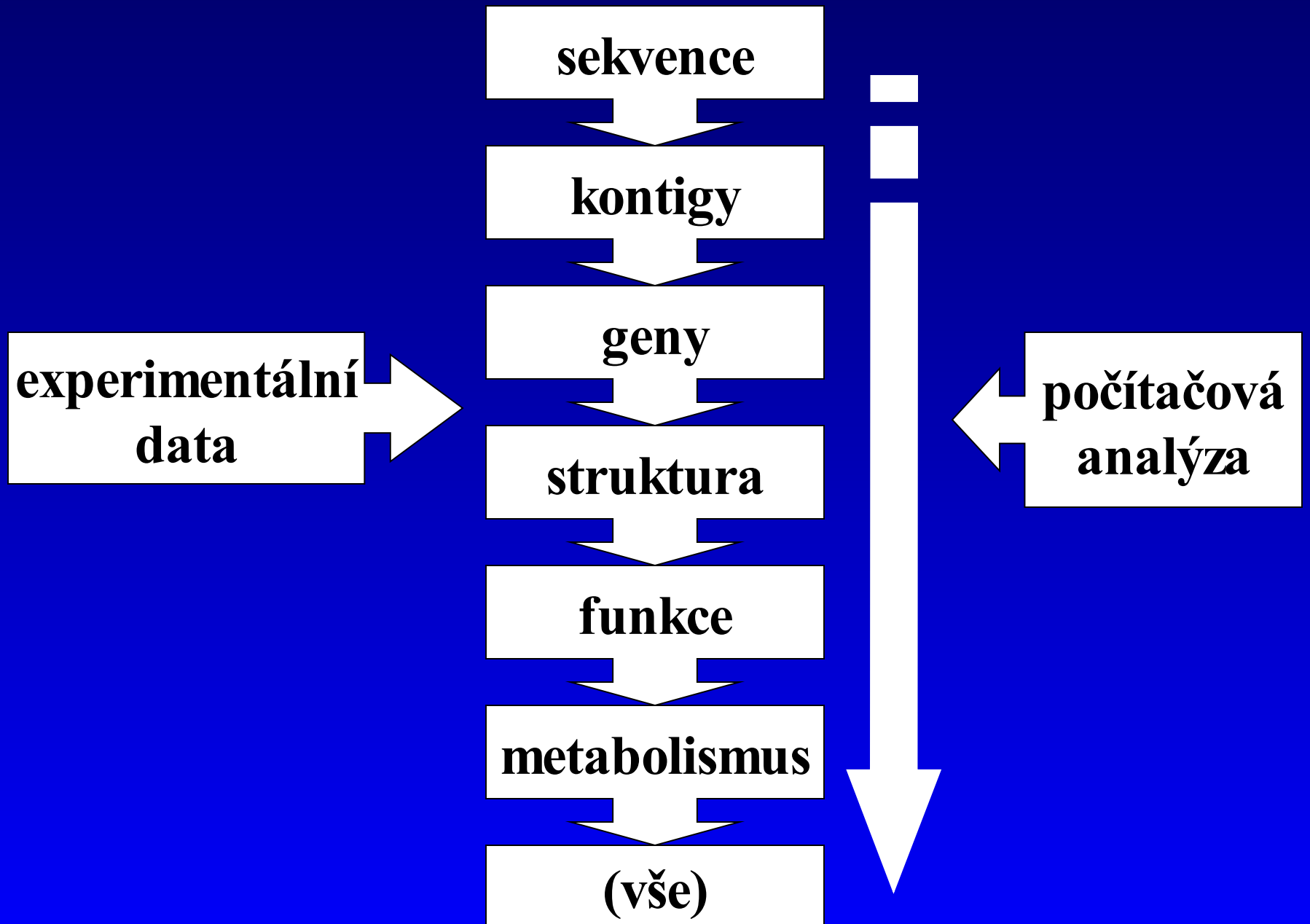
Informatika nad biologickými molekulami (daty).

Bioinformatika extrahuje molekulární informační systém pro molekulární biologii.

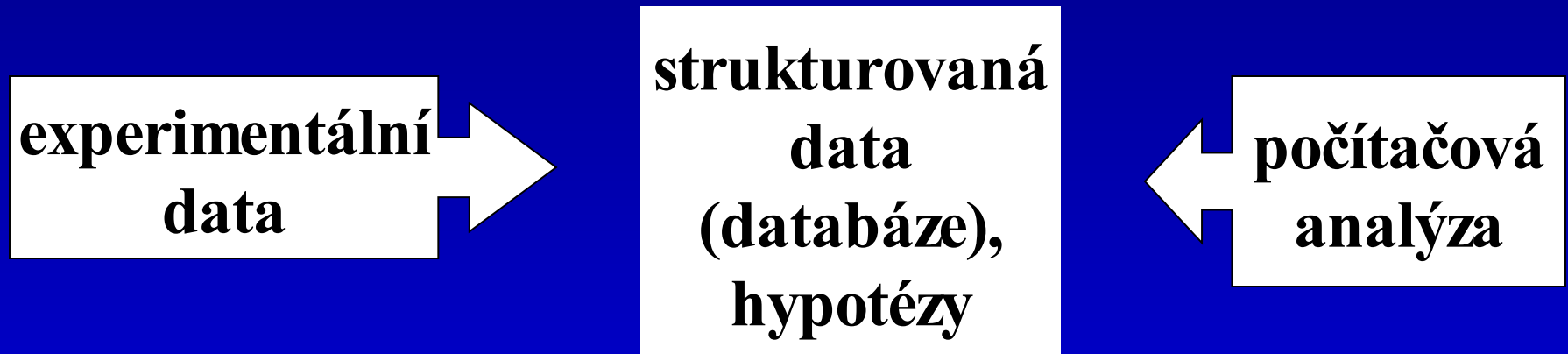
Bioinformatika je konceptualizovaná molekulární biologie (ve smyslu fyzikálně chemickém) na niž je aplikována informatika (odvozená od matematické informatiky a statistiky).

Aplikace: teorie
biotechnologie
farmacie
medicína
genetické inženýrství

bioinformatika



bioinformatika



velikosti genomů



Mycoplasma genitalium

0.58 Mbp



Escherichia coli

4.6 Mbp



Saccharomyces cerevisiae 16 chr.

11.2 Mbp



Arabidopsis thaliana 5 chr.

115.4 Mbp



Drosophila melanogaster 5 chr.

~137.0 Mbp

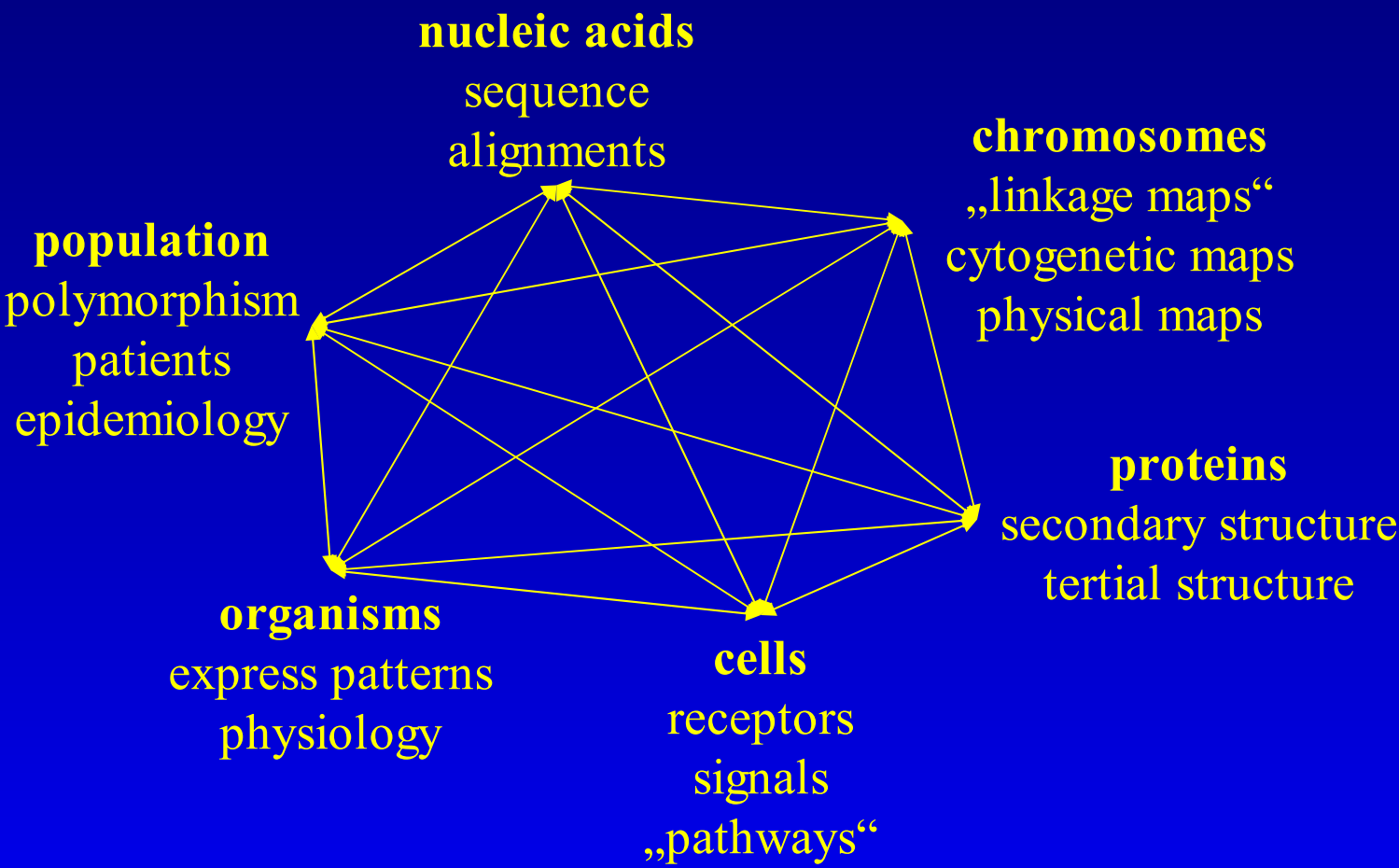


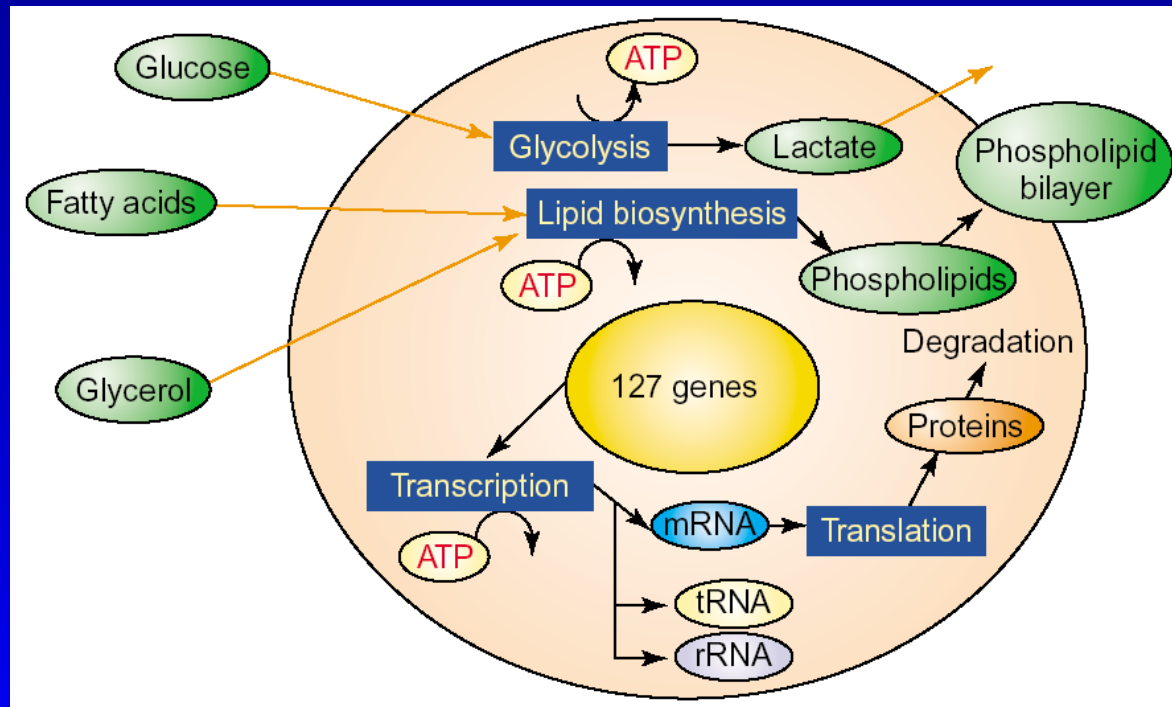
Homo sapiens

24 chr.

~ 3.3 Gbp

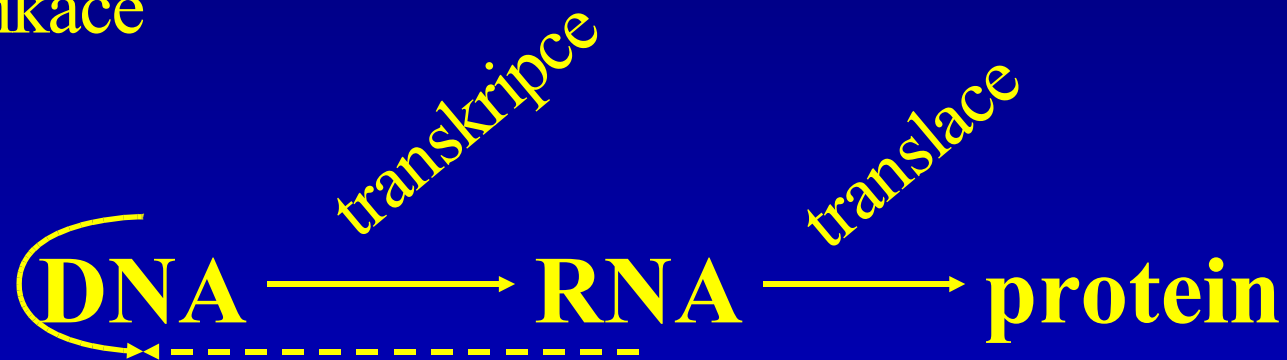
complexity





centrální dogma molekulární genetiky

replikace



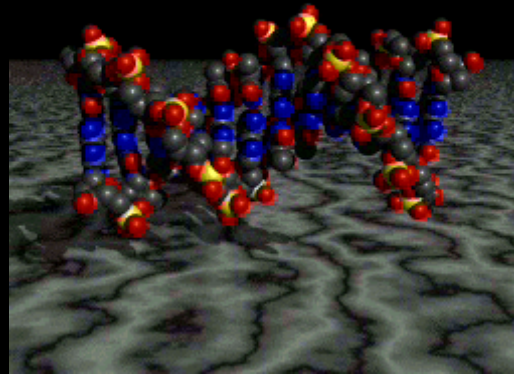
reverzní
transkripce

informace \longrightarrow funkce

**evoluční vztahy
mezi geny a
organizmy**

geny

funkce



proteiny

struktura

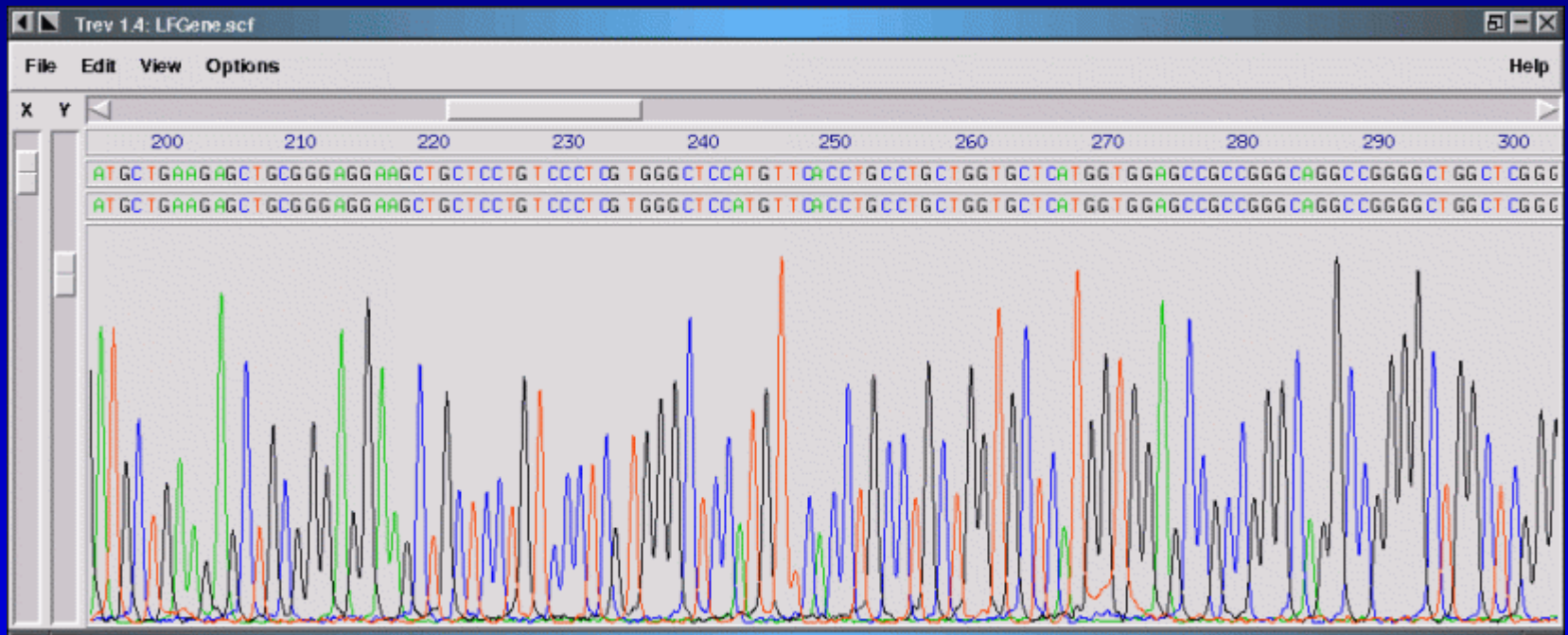
modelový příklad

Genome project of *Rhodopseudomonas palustris*

Sekvenace a charakterizace 5kb regionu.

(výsledky diplomové práce Jany Prejdové)

sekvence



sekvence

The screenshot displays a bioinformatics software interface with several windows:

- Top Left Window:** A window titled `.show_templates0: xb61h12.s1 #38` showing a sequence alignment with blue arrows and a vertical line. It includes zoom controls (+10%, +50%, zoom out) and a `crosshairs` checkbox.
- Top Right Window:** A `Contig Selector` window showing a horizontal bar with tick marks and zoom controls.
- Middle Left Panel:** A list of files including `xb54g5.s1RCF`, `xb58e7.s1`, `xb60g1.s1`, `xb54a3.s1`, `xb62c2.s1`, `xb64d11.s1`, and `xb60b9.s1`. It also shows a `CONSENSUS` section and a `Strands` section.
- Middle Right Panel:** A `Contig Editor` window for `38 xb61h12.s1` showing a sequence editor with a text area containing DNA sequence and a `Consensus` line. The sequence is: `CTATTAGAAAA*CACCCCC*TTGCTCCTGTTGTGCAATTTTTGTTTTAAGTTTT`. The consensus is: `CTATTAGAAAA*CACCCCC*TTGCTCCTGTTGTGCAATTTTTGTTTTAAGTTTT`. The editor includes navigation and editing tools.
- Bottom Section:** Multiple chromatogram plots showing signal intensity for different bases (A, C, G, T) across various positions (e.g., 180, 190, 20, 110, 100, 90, 80, 70, 390, 100, 60, 70, 80, 90, 100). Each plot includes a `Quit` button.

sekvence

>jana (4797 nt)

GAATTCGCCGCGGGGCTGCGCATCACCGATGCCGCCACCATCGAGATCGTTCGAGATGGTACTGGCCGGCTCGATCAACAAGCAGCTCGTCCGGCTACATCA
ACGAAGCGGGCGGCAAGGCCGTCGGCTGTGCGGCAAGGACGGCAACATGGTGTCCGCCACCAAGGCGACGCGCACCATGGTTCGATCCGGATTTCGCGGAT
CGAAGAGGTGATCGACCTCGGTTTTCGTCGGCGAGCCGGAGAAGGTCGACCTCACCCCTGCTCAACCAGCTGATCGGCCACGAGTTGATCCCGGTGCTGGCG
CCGCTGGCGACCTCCGCGTCCGGCCAGACCTTCAACGTCAATGCCGACACCTTTGACAGTTCGCGGTTGCCGGTTCGCTGCGGGCCAAGCGCTGCTGCTGC
TGACCGACGTGCCGGGCGTGTTCGACCAGAACAAGAAGCTGATCCCCGAACGTGTCGATCAAGGATGCCCGCAAGCTGATCGCAGACGGCACCATCTCGGG
CGGCATGATCCCCAAGGTCGAGACCTGCATCTACGCGCTCGAACAGGGCGTTCGAAGGCGTTCGATCCCTCGACGGCAAGGTCGCGCACGAGTGTGCTGC
GAATTTGTTACCAACCAGGGCACCGGCACGCTGATCCACAAGTGTGCGGAGGCTGCGGGCACAACATCCGTTCATGGCCGGGCTCGTCCCGGCCATCCAG
TCTTTCCGGCGGTTTTTCTCAGCAAGACGTGGATGCCCGGCACAAGGCCGGGCATGACGGGGTGGAGATCGCGCGCCCTCGCCGCCATTGTCACCACCCTC
GCCCTCACCTCCGCCGCCACGCCGACCTCAAGCTCTGCAACCGCATGAGCTACGTGGTTCGAGACGGCGATCGGGGTCGATTCCAACGGCACCACCGCCT
CGCGCGGATGGCTGCGGATTGATCCGGCGCAATGCCGGTTCGTGGTTCGAAGGCGCGCTCAACGCCGACCGCATCATGCTGAATGCCCGCGCGCTGGCGGT
GTACGGCGTCTCGCCGCTGCCGCGAAGCGCACTGACCGGCTGTGCATTGCCGAAGACAATTTTCGTCATCGCCCGCGCGGGCAATGCCCGGGCGGCCAA
ACGCTCGCCGCCTTACCGAGATCAAGCCACCGACACCGAGGACGGCAACAAGATCGCTTATCTGGCGGAAGACTCCGGCTACGACGACGAACAGGCCA
AACTCGCCGCGATCCAGCGGCTGCTGGTGTGATCGCCGGTTACGACGCCTCGCCGATCGACGGCGTTCGACGGCCCGAAGACGCAGGCCGCGCTGTCCGCCCT
CCTCAAGAGCCGAGGCTGAAGCCCGAGATCGTTCGATGCGCCGGATTTCTTCGACGTGATGATCAAGGCAGTGCAGCAGCCGTCCGGCAGCGGGCTGACC
TGGTGC AACGACACCAAGTACAAGATCATGGCGGCCGTCGGCGAAGACGACGGCAAGACTGTCACCAGCCGCGGCTGGTACGGTGTTCGCCCGGCCAAT
GCCTGCGCCCCGACCTCGGGCGACAGCCGAAGCGGGTGTTCAGCTTCGCCGAAGCGGTTCGACGGCAGCGGCAGGCCGGTGAACATCAAGGGCCGTGCGCT
GAACTGGGGCGGCGGCGTGCAGCTGTGCACGCGTGCAGCAAGTTCGAGATCGGGCAGCAAGGCGATTGCGCGGGCGCGGGCCTCGCCGCCACCGGCTTC
GCCGCCGTCGATCTCAGTAGCGGCAAGACATTGAGGTTGTCCGCCCATGATGCAGCTCGGCAAACGCGGCTTCGATCACGTCGAGACCTGGGTGTTTCGA
TCTCGACAACACGCTGTACCCGCATCACCTCAACCTATGGCAGCAGGTCGATGCGCGGATCCGCGACTTCGTCGCCGACTGGCTGAAGGTTTCGCCGGAA
GAAGCCTTCGATCCAGAAGGATTACTACAAGCGCTACGGCACCACGATGCGCGGGATGATGACCGAGCACGGCGTTCACGCCGACGACTACCTGGCTT
ATGTCCACGCCATCGACCATTCGCCGCTGCAGCCGAATCCGGCGATGGGCGATGCGATCGAGCGACTGCCGGGCGCAAGCTGATCCTGACCAACGGCTC
GACCGCCCATGCGGGCAAGGTGCTGGAGCGGCTCGGCATCGGCCATCATTTTCGAGGCGGTGTTTCGACATCATTCGCGCCGACCTCGAGCCGAAGCCGGCG
CCGCAGACCTACCGCCGTTTTTCTCGATCGCCATGGTGTGACCCGGCCCGCGCCGCGATGTTTCGAAGACCTCGCCCGCAACCTCACCGTTCGCCACACG
TCGGCATGACCACCGTGTGGTGGTGCCTGACGATAGCCAGGACGTGGTCCGCGAAGATTGGGAGCTTGAAGGCCGCGACGCCGCCACGTCGATCACGT
GACTGATGATTTGACAGGGTTCTTGGGGAAGCTGAGTTTCGCTGTAGGCCGGGACGCCTCCCAAGCGTCAATCGTCATCGCCGCCGGATGCAAGGCGGCT
AGGTATTGCGGAGCGCTCGCGATCTTCGTCCAATGCCCTGGGATACTGGATCGCCCGGACGAGCCGGGCGACGAGTTGAAGAGAGATGACGTGGCGTC
ACCACATCCCCGCGTCATCGCCCGCGCAGGCGGGCGATGACTTGGCGGACGGGGCGGGCGCCTTGACTCCGACCCGGCGAATCCGGACAACACTCCGCA
AAACTCTCCCTGAAATCAGCCTCCCAAGGACCCGTCGATGCCGCTCACCGCCCTGGAATCTACCATCAACGCCGCTTTTCGACGCGCGCGACACCGTTACC
GCGGCGACGCAGGGCGAGATTTCGTCAGGCCGTCGAGGATGCGCTCGATCTGCTCGACCAGGGCAAGGTGCGGGTGGCGGGCGCGACGACTCCGGCGCCT
GGACGGTCAATCAGTGGCTGAAGAAAGCAGTGTGCTGTGCTTCCGGCTCAACGACATGGGCGTGTGATCGCCGGCGGCCCGGGCGGCCAACTGGTGGGA
CAAGGTGCCGTCGAAGTTCGAGGGCTGGGGTGAGAACCCTTCGCGAGGGCCGGCTTCGCGCCGTCGCCGGGCGGATCGTTCGCGGCTCGGCCTTTATCGC
CAAGACGCGGTACTGATCCGTCCTTCGTC AATCTCGGCGCTTACGTCGATGAAAGCACCATGGTTCGAACACCTGGGCGACCGTCGGCTCCTGCGCCAGA
TCGGCAAGCGCGTGCACATCTCCGGCGGTGCCGGCATCGGCGGCGTGTGTCGAGCCGCTGCAGGCCGGCCCGGTGATCATCGAGGACGACTGCTTCATCGG
CGCCCGCTCCGAAGTCGCCGAAGGCGTGTGCTGCGCAAGGGTTCGGTGTGCGGATGGGCGTTTTCTTCGCGCCCTCGACCAAGATCGTCGACCGCGAG
ACCGCGAAATCTTCGTCGGCGAAGTCCCGGAATATGCCGTGCTGGTCCCGGCACCCTGCCCGCAAGCCGATGAAGAA . .

všeobecná analýza

Co lze v DNA najít?

strukturní a organizační elementy

evoluční vztahy

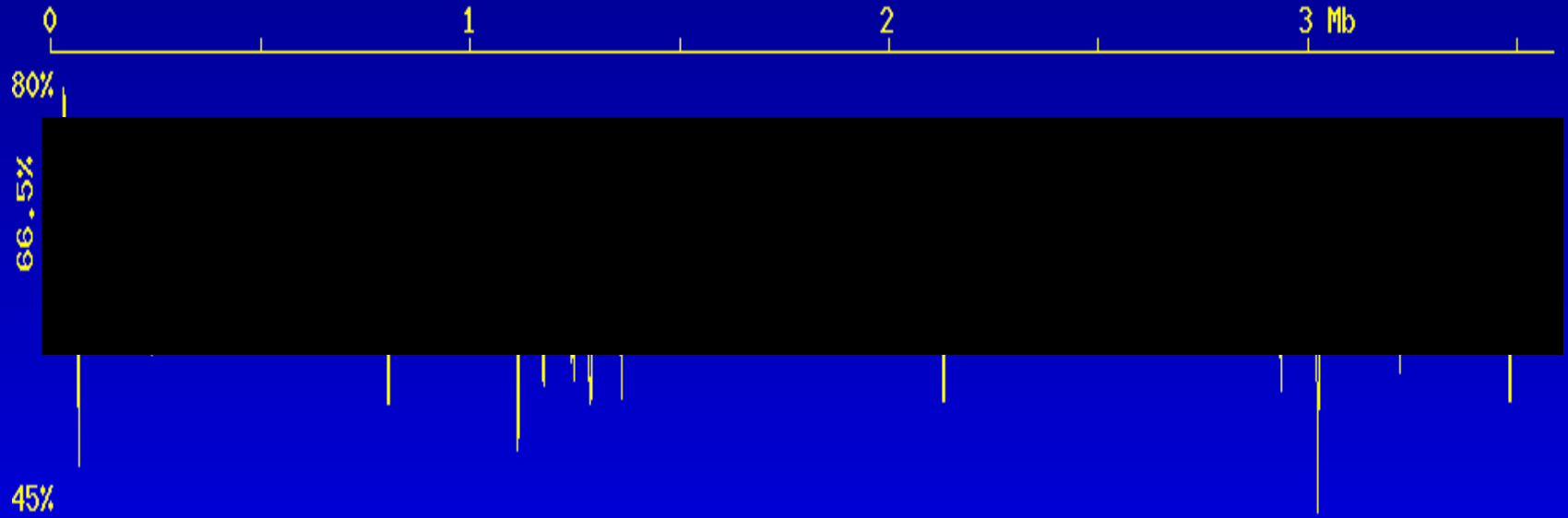
geny

promotory a další řídicí elementy

„cizí“ DNA

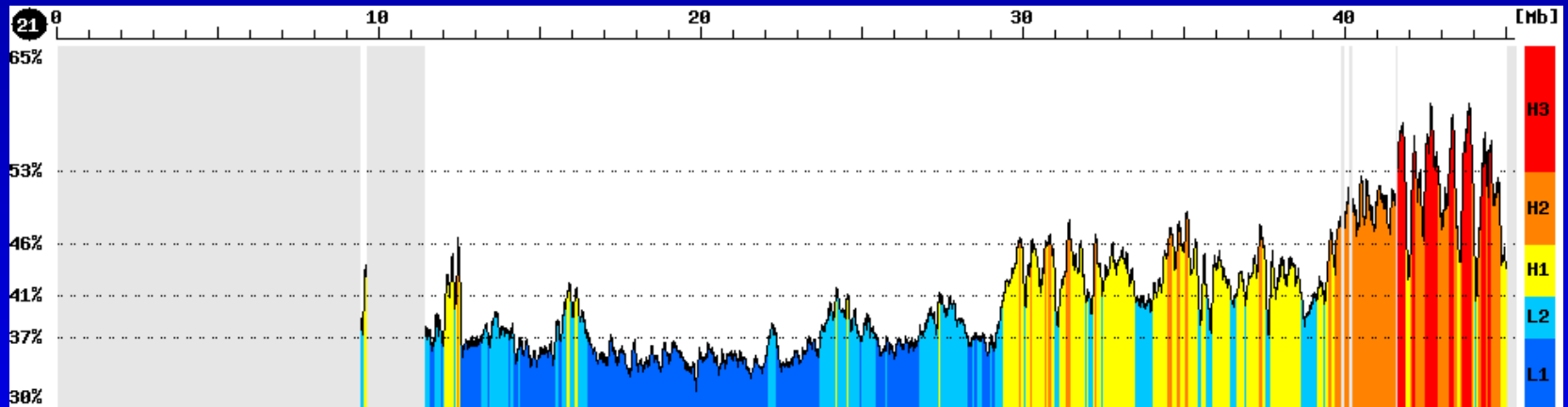
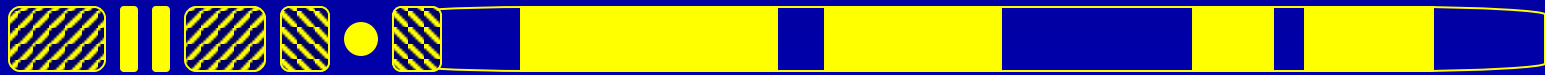
všeobecná analýza

Rhodobacter capsulatus, GC profil

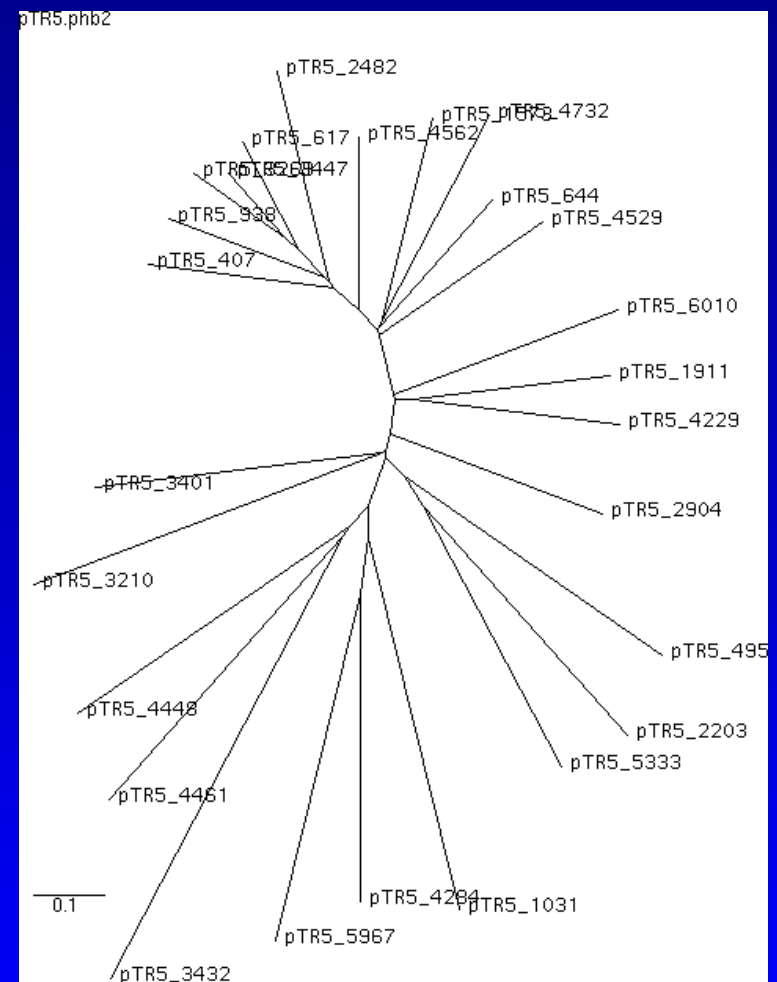
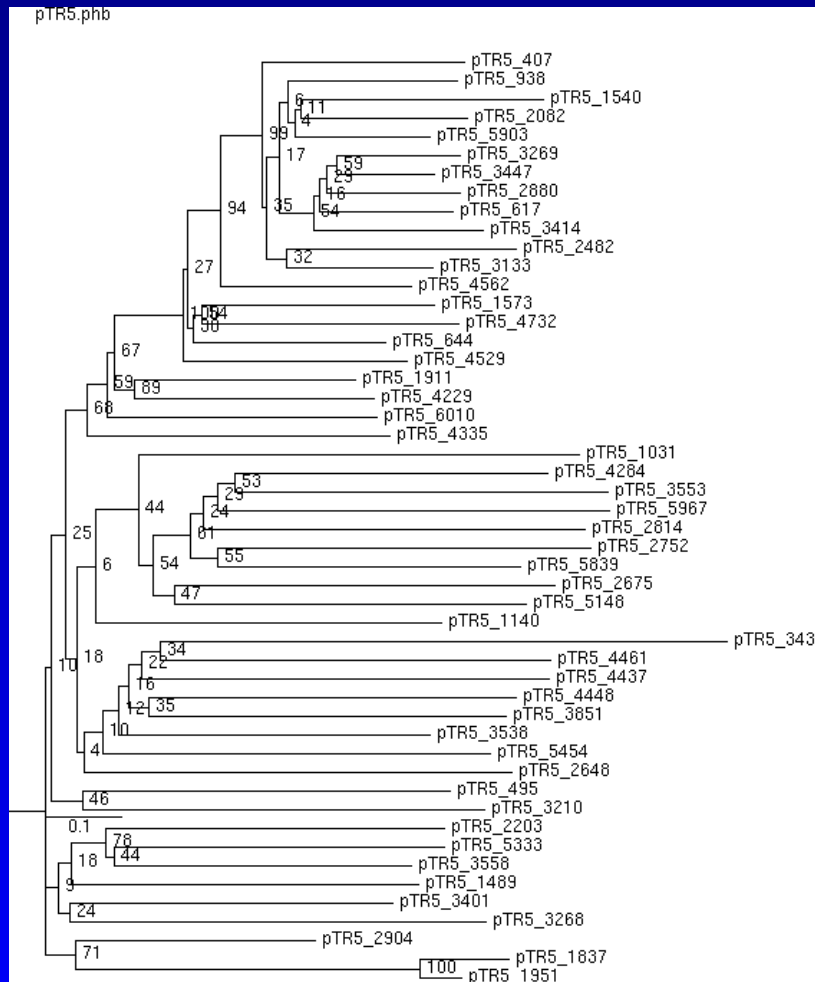


všeobecná analýza

Homo sapiens, chromozóm 21, GC profil



Evoluční strom pTR5 rodiny lidských endogenních retrovirů



geny

Jak najít geny?

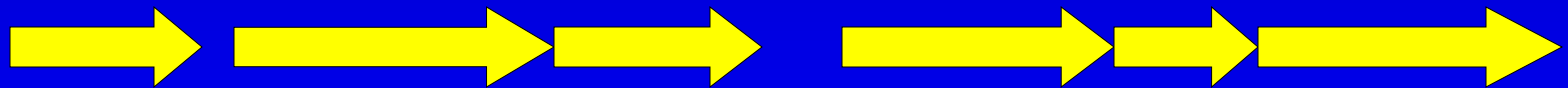
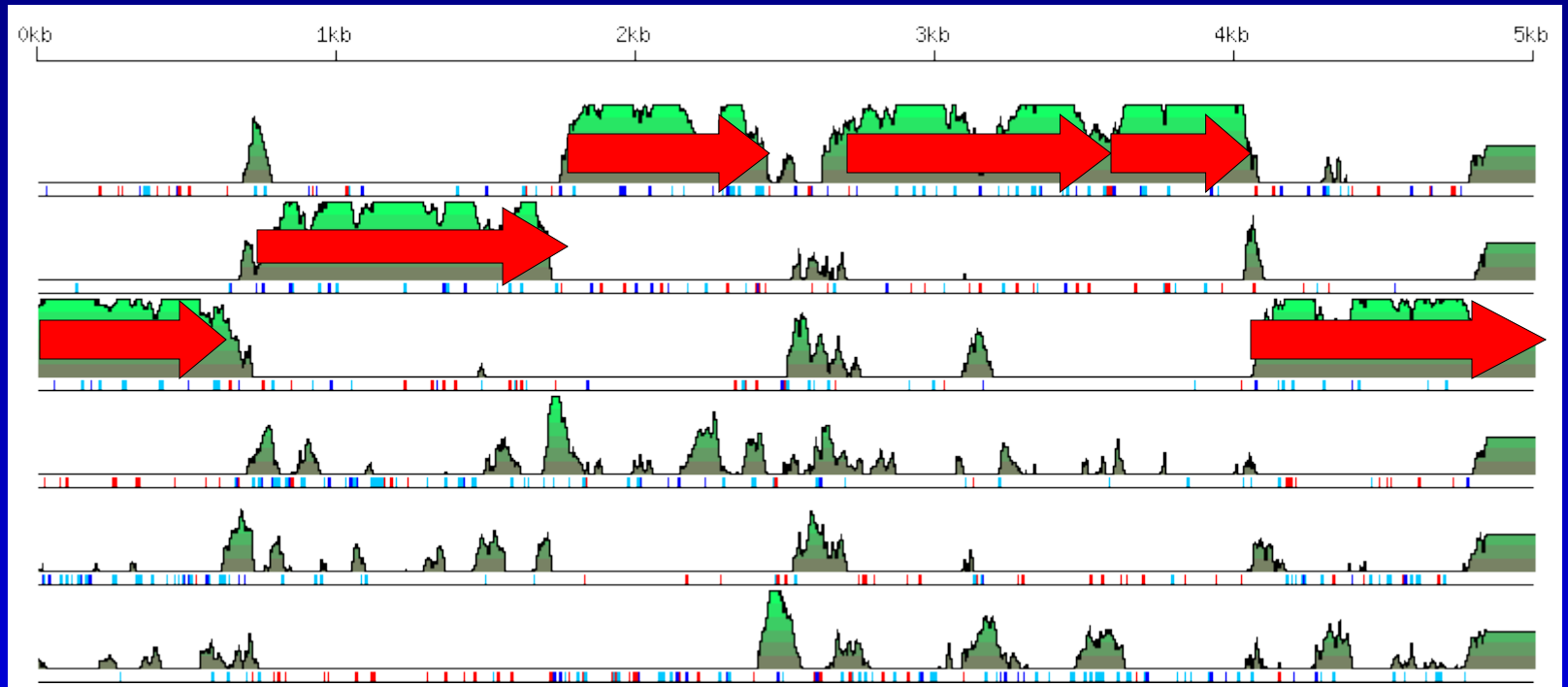
Leucin

Rhodobacter capsulatus

Escherichia coli

antikodón	počet	%	%
CUA	3	<1	4
CUC	119	16	9
CUG	458	60	52
CUU	157	20	10
UUA	0	0	11
UUG	27	3	13

geny



Homo sapiens, chromozóm 21, predikce genů

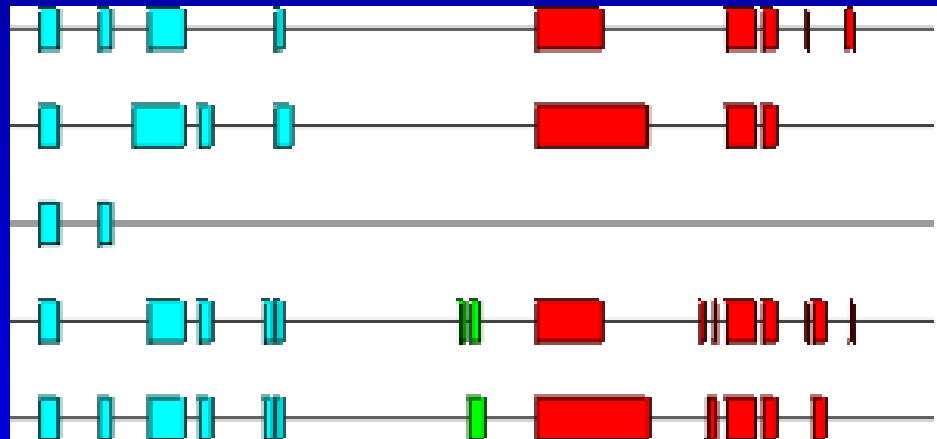
Sanger

Ch21 (*in Nature*)

cDNA

GENESCAN

EXOFISH



Jaké proteiny geny kódují?

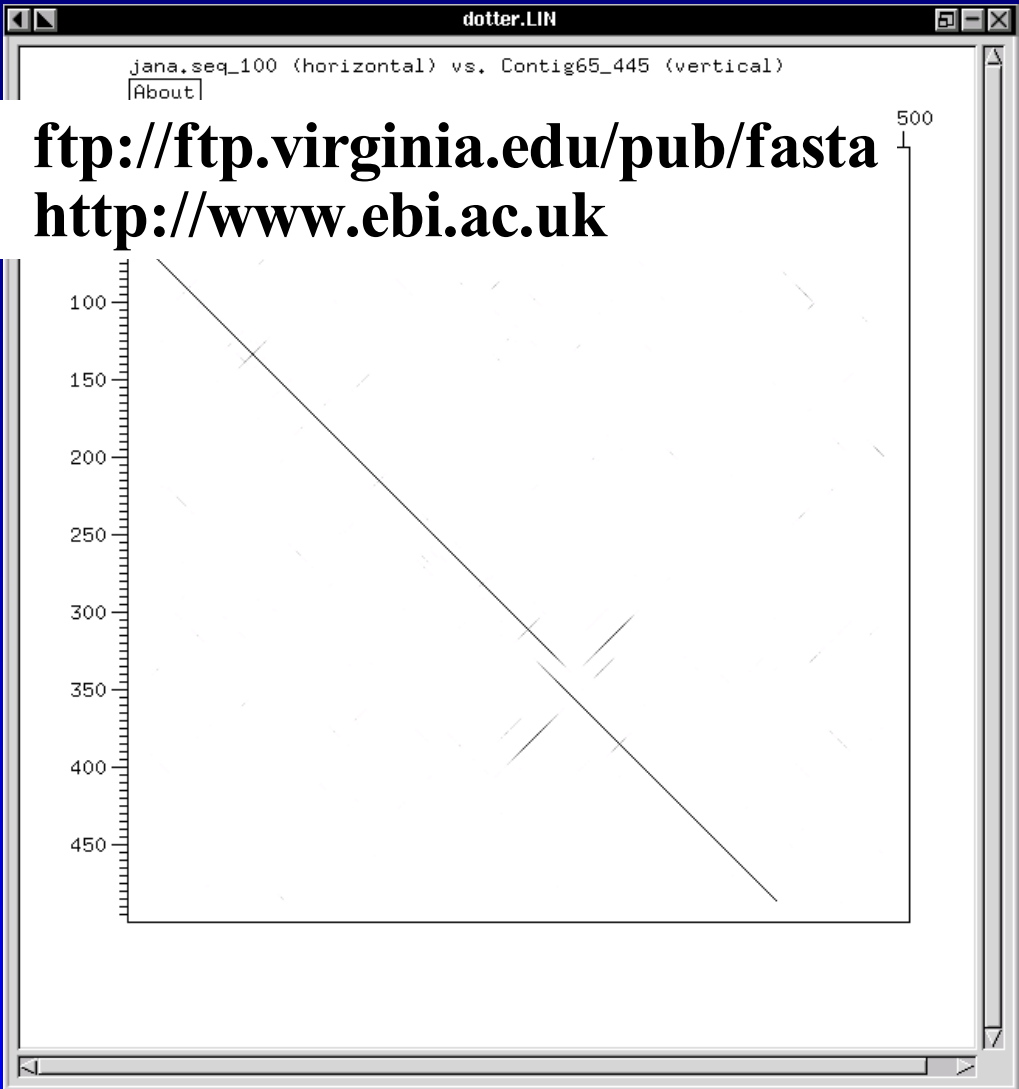
alignment



1:1

Dot plot

**SSEARCH
BLITZ ...**



1:n

**FASTA
BLAST**

n:n

**PSI-BLAST
HMMER**

n

**ClustalW
MultAlign**

alignment



1:1

Dot plot

SSEARCH
BLITZ

1:n

FASTA
BLAST

n:n

PSI-BLAST
HMMER

n

ClustalW
MultAlign

Netscape: BLAST Search Results

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop

Location: <http://www.ncbi.nlm.nih.gov/blast/blast.cgi?RID=955977286-251> What's Related

Database: Non-redundant SwissProt sequences
85,973 sequences; 30,868,576 total letters

If you have any problems or questions with the results of this search please refer to the [BLAST FAQs](#)

[Distribution of 116 Blast Hits on the Query Sequence](#)

Mouse-over to show defline and scores. Click to show alignments

Color Key for Alignment Scores

<40	40-50	50-80	80-200	>200
-----	-------	-------	--------	------

Sequences producing significant alignments:

Accession	Description	Score (bits)	E Value
sp Q9ZEX2 DAPD_BORPE	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLA...	272	3e-72
sp P03946 DAPD_ECOLI	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLA...	267	1e-70
sp P56220 DAPD_MYCBO	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLA...	265	6e-70
sp P45284 DAPD_HAEIN	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLA...	265	8e-70
sp P41396 DAPD_ACTPL	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLA...	258	6e-68
sp Q92DX0 DAPD_RICPR	2,3,4,5-TETRAHYDROPYRIDINE-2-CARBOXYLA...	241	1e-62
sp P31852 TABB_PSESZ	TABB PROTEIN	239	3e-62
sp P44514 DAPD_HAEIN	SUCCINYL-DIAMINOPIPELATE DESUCCINYLA...	192	5e-48
sp P24176 DAPD_ECOLI	SUCCINYL-DIAMINOPIPELATE DESUCCINYLA...	189	3e-47
sp P73326 ARGB_SYNY3	ACETYLGUTAMATE KINASE (NAG KINASE) (A...	183	2e-45
sp Q60362 ARGB_METJA	PROBABLE ACETYLGUTAMATE KINASE (NAG K...	176	4e-43
sp P31595 ARGB_PORPU	ACETYLGUTAMATE KINASE (NAG KINASE) (A...	174	1e-42
sp Q20988 ARGB_ARCFU	PROBABLE ACETYLGUTAMATE KINASE (NAG K...	171	6e-42
sp Q9X244 ARGB_THENA	ACETYLGUTAMATE KINASE (NAG KINASE) (A...	169	7e-41
sp Q26285 ARGB_METTH	PROBABLE ACETYLGUTAMATE KINASE (NAG K...	168	1e-40
sp Q67848 ARGB_AQUAE	ACETYLGUTAMATE KINASE (NAG KINASE) (A...	168	3e-38
sp P94989 ARGB_MYCTU	ACETYLGUTAMATE KINASE (NAG KINASE) (A...	123	2e-27

alignment



1:1

Dot plot

SSEARCH
BLITZ

1:n

FASTA
BLAST

n:n

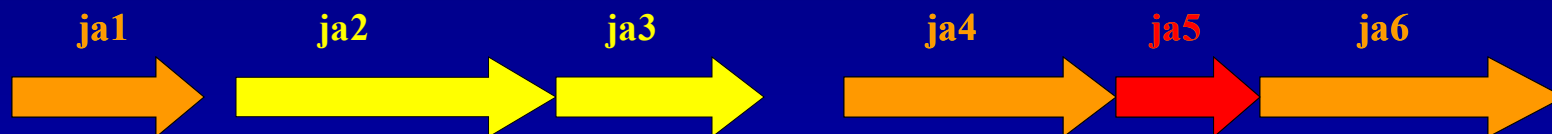
PSI-BLAST <http://ncbi.nlm.nih.gov>
HMMER

n

ClustalW
MultAlign

Jakou mají geny v buňce funkci?

funkce



ja1 ACETYLGLUTAMATE KINASE EC 2.7.2.8

ja2

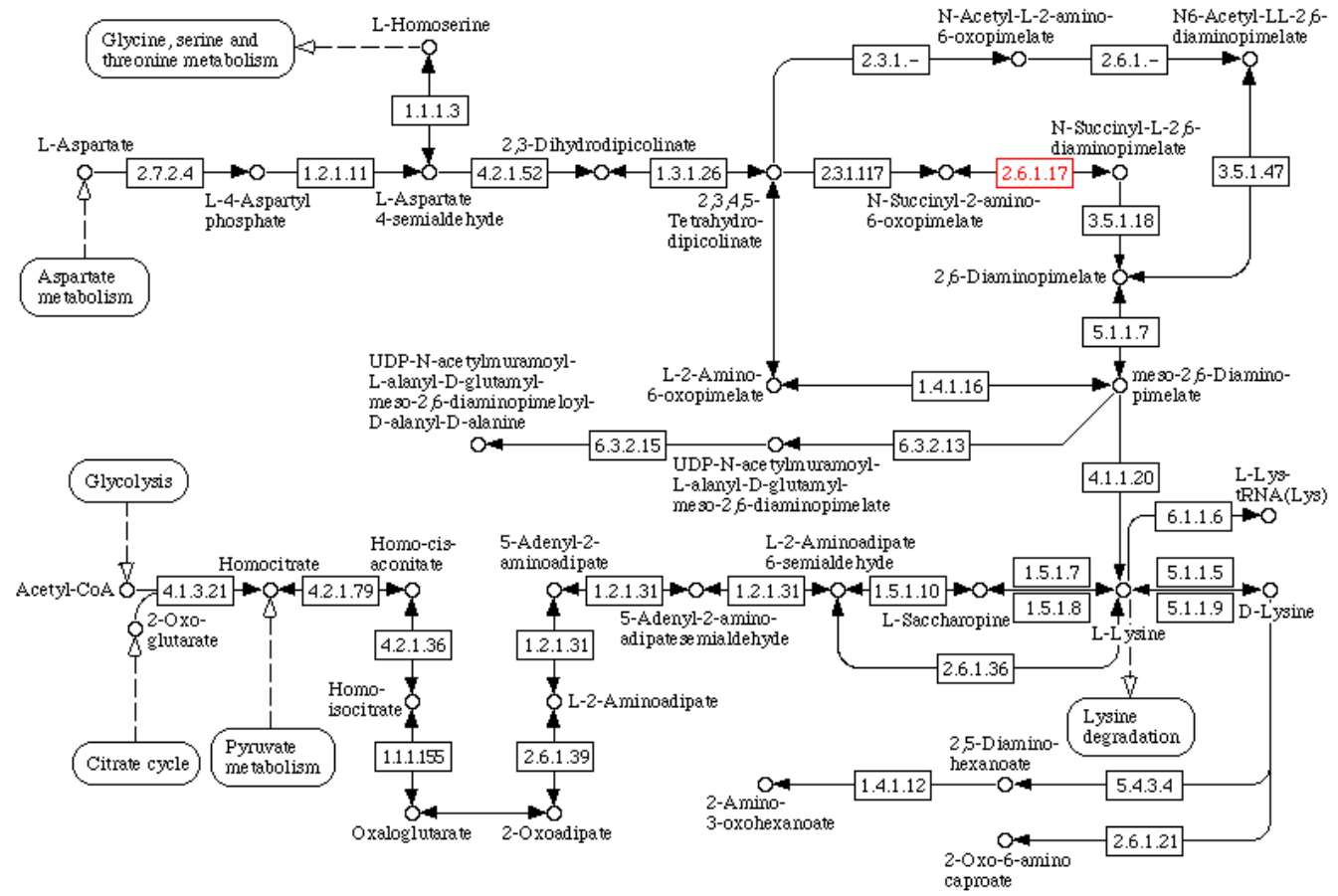
ja3

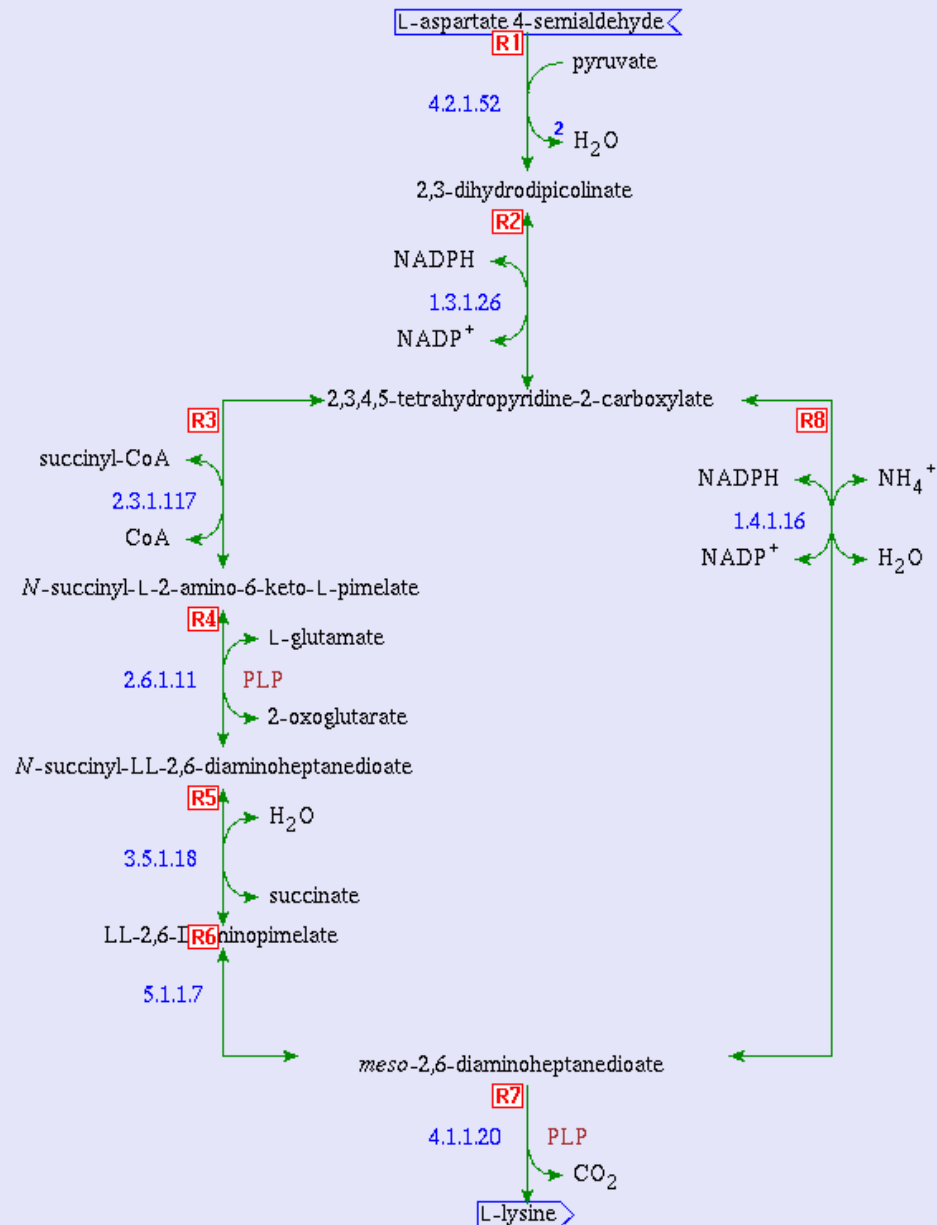
ja4 TETRAHYDRODIPICOLINATE EC 2.3.1.117
N-SUCCINYLTRANSFERASE

ja5

ja6 SUCCINYL-DIAMINOPIMELATE EC 3.5.1.18
DESUCCINYLAASE

LYSINE BIOSYNTHESIS





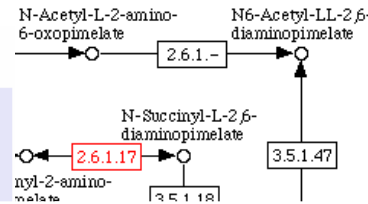
funkce



- | | | |
|------------|---|--------------|
| ja4 | TETRAHYDRODIPICOLINATE
N-SUCCINYLTRANSFERASE | EC 2.3.1.117 |
| ja5 | ACETYLORNITHINE
TRANSAMINASE | EC 2.6.1.11 |
| ja6 | SUCCINYL-DIAMINOPIMELATE
DESUCCINYLAASE | EC 3.5.1.18 |

LYSINE BIOSYNTHESIS

L-Homoserine

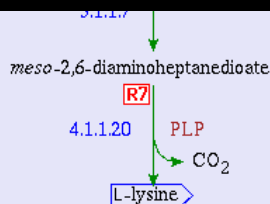


Cytosol

L-aspartate 4-semialdehyde

Rhodopseudomonas palustris může syntetizovat aminokyselinu lysine biochemickou dráhou přes enzym EC 2.6.1.17.

NADPH
13.1
NADP⁺



- Methionine Transport
 - ATP-driven methionine Transport
 - Glutathione Dependent Methionine Transport
- Methionine Anabolism
 - Homocysteine to Methionine
 - Adenosylhomocysteine to Methionine
 - Homoserine to Methionine
 - Cystathionine gamma-lyase reaction
 - Methionine Anabolism (via 1.8.4.5)
 - Adenosylmethionine Biosynthesis
 - S'-adenosylmethioninamine Biosynthesis
 - Methionine S'-methyltransferase reaction
- Ornithine Metabolism
 - Ornithine Transport
 - ATP-driven ornithine Transport

synthesis)

metabolism)