

PyMod 2.0

Son Hoang Nguyen - 408779

29.10.2018

Obsah

- ▶ PyMod 2.0
- ▶ Požadavky
- ▶ Instalace
- ▶ BLAST a PSI-BLAST
- ▶ MUSCLE
- ▶ ClustalW a Clustal Omega
- ▶ WebLogo 3
- ▶ ESPript 3.0
- ▶ CAMPO
- ▶ PSIPRED
- ▶ MODELLER a SALIGN
- ▶ DOPE

PyMod 2.0

- ▶ Giacomo Janson, Chengxin Zhang, Maria Giulia Prado, Alessandro Paiardini
- ▶ Kolekce nástrojů pro analýzu proteinových sekvencí a struktur
- ▶ Rozšiřuje PyMOL
- ▶ open-source

Požadavky PyMOD 2.0

- ▶ <https://github.com/pymodproject/pymod/releases>
- ▶ **PyMOL (0.99+)**
 - ▶ Python 2 <https://www.python.org/downloads/>
 - ▶ PWM <https://www.lfd.uci.edu/~gohlke/pythonlibs/#pymol>
 - ▶ Numpy+MKL <https://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy>
- ▶ **BioPython**
<https://www.lfd.uci.edu/~gohlke/pythonlibs/#biopython>
- ▶ **Modeller**
https://salilab.org/modeller/download_installation.html

Instalace

1. Nainstalujete Python 2 - třeba [Python 2.7.15](#) [1.11.2018]
2. Nastavte proměnné prostředí (nepovinné, ale doporučováno)
 1. Pozor pokud máte nainstalováno i Python 3, používejte příkaz pip2.
3. Aktualizujte pip, příkaz: pip2 install --upgrade pip
4. Nainstalujte balíčky Pmw a Numpy+MKL (1.14.6 je stabilnější)
 1. Na semináři bylo řečeno, že lze nainstalovat numpy z repozitáře.
Základní numpy není pymolem podporováno. Musí se použít balík intel-numpy, ale i tak nebude fungovat Pymod.
Proto radši používejte balíky z <https://www.lfd.uci.edu/~gohlke/pythonlibs>.
5. Nainstalujte Pymol
6. Nainstalujte Modeller (nutná registrace, akademická licence je zdarma)

Instalace

7. Stáhněte si pymod a rozbalte zip soubor (v něm bude `install_all.py`)
8. Spust'te pymol a spust'te skript `install_all.py` (menu File → Run...)
9. Restartujte pymol
10. Nyní můžete pracovat s pymodem

BLAST a PSI-BLAST

- ▶ Basic local alignment search tool
- ▶ Position-Specific Iterative BLAST
- ▶ Existuje několik serverů, např:
 - ▶ <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- ▶ Jeden z nejpoužívanějších bioinformatických nástrojů
- ▶ Umožňuje porovnávat a vyhledávat proteinové a nukleotidové sekvence
- ▶ PSI-BLAST umožňuje vyhledávat vzdáleně podobné sekvence
 - ▶ Nejdříve vyhledá podobné sekvence a na ty pak opět aplikuje BLAST

BLAST a PSI-BLAST

- ▶ Heuristický algoritmus
- ▶ E-Value (expectation value)
 - ▶ Nám říká jako je pravděpodobnost, že podobnost sekvencí je dána náhodou.
 - ▶ Čím menší, tím spolehlivější data získáme.

MUSCLE

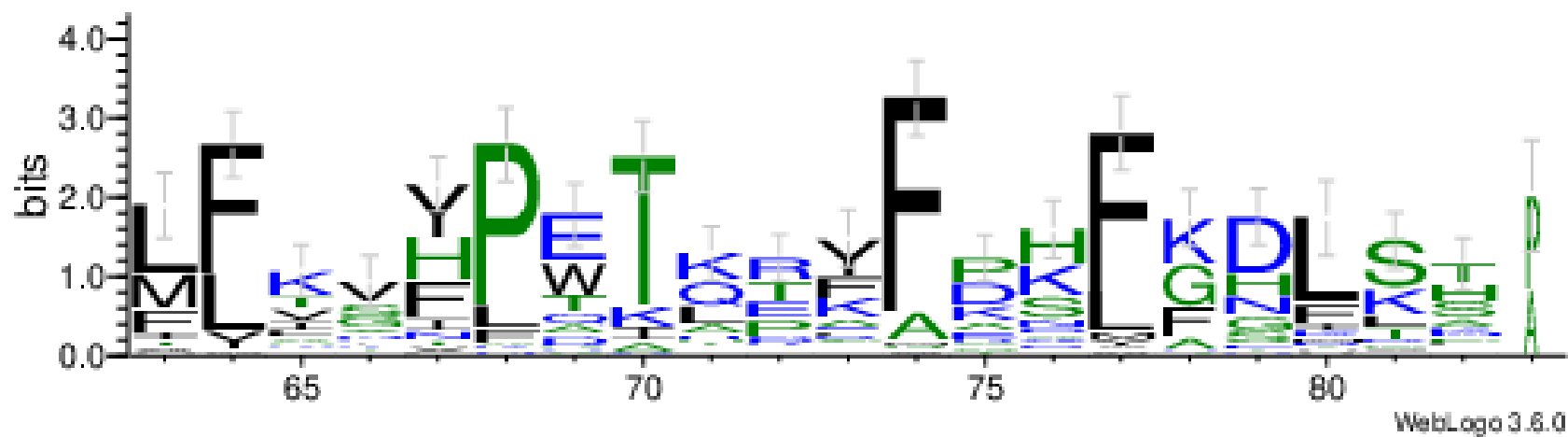
- ▶ **MU**ltiple Sequence Comparison by Log-Expectation
- ▶ <https://www.ebi.ac.uk/Tools/msa/muscle/>
- ▶ Významně rychlejší než Clustal
- ▶ Může ale nemusí dávat lepší výsledky než Clustal

ClustalW a Clustal Omega

- ▶ Mnohonásobné sekvenční příložení
- ▶ <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- ▶ Existuje několik verzí
- ▶ Nejpoužívanější je Clustal Omega
- ▶ Clustal2 je novější implementace ClustalW

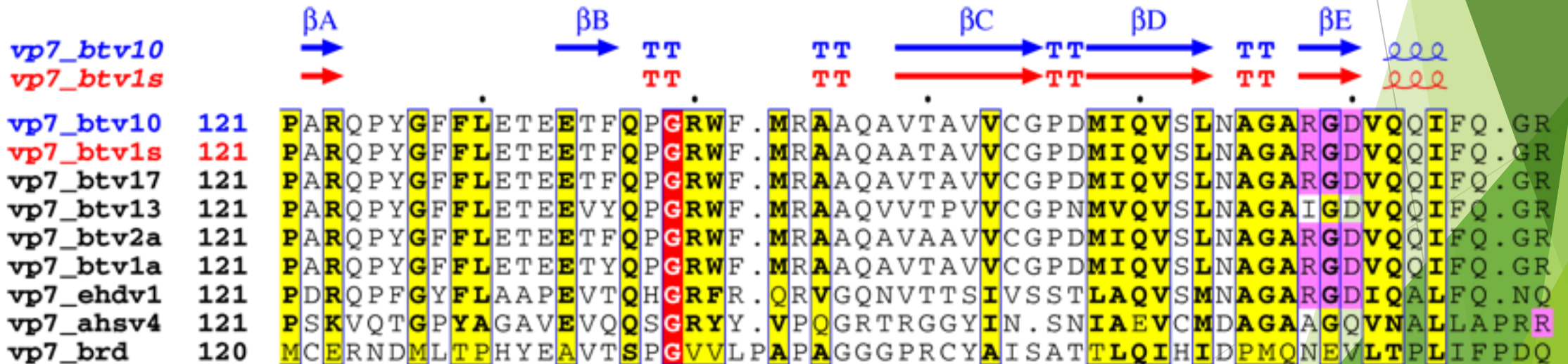
WebLogo 3

- ▶ Grafické zobrazení mnohonásobného zarovnání
- ▶ <http://weblogo.threeplusone.com/>



ESPrript 3.0

- ▶ Easy Sequencing in PostScrip
- ▶ <http://esprript.ibcp.fr/ESPrript/ESPrript/>



CAMPO

- ▶ Analýza konzervovaných oblastí z mnohonásobných příložení
- ▶ <http://schubert.bio.uniroma1.it/campo.html>
- ▶ Součást Pymodu

PSIPRED

- ▶ PSI-blast based secondary structure PREDiction
- ▶ <http://bioinf.cs.ucl.ac.uk/psipred/>
- ▶ Umělá neuronová síť a strojové učení
- ▶ Umožňuje předpovídat sekundární struktury proteinů
- ▶ Přesnost je kolem 80 % a stále se zvyšuje

MODELLER a SALIGN

- ▶ Komerční (akademická licence je zdarma)
- ▶ Homologní modelování - pomocí sekvence, která má experimentálně dokázanou strukturu, vytvoří ze sekvence k ní zarovnané 3d model
- ▶ <https://salilab.org/modeller/>
- ▶ <https://modbase.compbio.ucsf.edu/salign/>
- ▶ Salign je plugin pro Modeller, umožňuje optimalizaci

DOPE

- ▶ Discrete Optimized Protein Energy
- ▶ Umožňuje nám posoudit, jak spolehlivě je část dané sekvence zmodelována
- ▶ Čím menší, tím spolehlivější; smyčky se modelují špatně, a proto mají často vysoké skóre
- ▶ Spočítá se jako funkce (statistický potenciál), která závisí na vzdálenostech jednotlivých neinteragujících atomů v homogenní sféře, jejíž poloměr se spočítá z dat získaných z dokázaných experimentálních struktur
- ▶ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2242414>
- ▶ Součást MODELLERu

Zdroje

- ▶ <https://academic.oup.com/bioinformatics/article/33/3/444/2568608>
- ▶ <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- ▶ <http://www.drive5.com/muscle/>
- ▶ <http://www.clustal.org/>
- ▶ <http://weblogo.threeplusone.com/>
- ▶ <http://esript.ibcp.fr/ESPript/ESPript/>
- ▶ <http://schubert.bio.uniroma1.it/campo.html>
- ▶ <http://bioinf.cs.ucl.ac.uk/psipred/>
- ▶ <https://salilab.org/modeller/>
- ▶ <https://modbase.compbio.ucsf.edu/salign/>