

## Kompresia DNA a určovanie príbuznosti organizmov

V súboroch

1. <http://ksvi.mff.cuni.cz/~mraz/bioinf/A.dna>,
2. <http://ksvi.mff.cuni.cz/~mraz/bioinf/B.dna>,
3. <http://ksvi.mff.cuni.cz/~mraz/bioinf/C.dna>,
4. <http://ksvi.mff.cuni.cz/~mraz/bioinf/D.dna>,
5. <http://ksvi.mff.cuni.cz/~mraz/bioinf/E.dna>,
6. <http://ksvi.mff.cuni.cz/~mraz/bioinf/F.dna>,
7. <http://ksvi.mff.cuni.cz/~mraz/bioinf/G.dna>

sú DNA reťazce obsahujúce mitochondriálnu DNA 7 organizmov (orezanú na maximálne 15000 báz, pretože verzia programu GenCompress pre Windows na dlhších reťazcoch “padala”). Určite mieru príbuznosti medzi týmito organizmami.

*Návod:* Použite program GenCompress, ktorý nájdete napríklad na adrese <http://monod.uwaterloo.ca/downloads/gencompress/>.  
Nech

- $\text{Compress}(u)$  – označuje výstup programu GenCompress aplikovaného na reťazec  $u$ .
- $\text{Compress}(u|v)$  – označuje výstup podmienenej kompresie reťazca  $u$  programom GenCompress pri znalosti reťazca  $v$ .

“Vzdialenosť” medzi organizmami  $u$  a  $v$  sa spočíta podľa vzorca:

$$D(u, v) = 1 - \frac{|\text{Compress}(u)| - |\text{Compress}(u|v)|}{|\text{Compress}(uv)|} .$$