



HPC/Exascale
Centre of
Excellence in
Personalised
Medicine

PATC course

Introduction to HPC for Life Scientists

31 January - 02 February 2022



BioExcel and the PerMedCoE projects have received funding from the European Union's Horizon 2020 research and innovation programme under the grant agreements 823830 & 675728 and 951773, respectively.



Funding



Horizon 2020
European Union Funding
for Research & Innovation



HPC/Exascale
Centre of
Excellence in
Personalised
Medicine

HMMER

Sequence Alignment Practical

PATC course: Introduction to HPC for Life Scientists, 31 January - 02 February 2022

Partners



Funding



Horizon 2020
European Union Funding
for Research & Innovation

Reusing this material



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en_US

This means you are free to copy and redistribute the material and adapt and build on the material under the following terms: You must give appropriate credit, provide a link to the license and indicate if changes were made. If you adapt or build on the material you must distribute your work under the same license as the original.

Aims

- Gain experience using an HPC system
 - Running
 - Using a batch system
- Run a real bioinformatics software package in parallel
 - Run on different numbers of cores
 - Measure time taken (elapsed **wall clock time**, **cpu time**)
 - Observe how performance scales
 - Relate performance to how software solves problem in parallel

Sequence Alignment

- Use HMMER to search (query) a protein database for similarity-based matches with a given protein sequence:

Alignments for each domain:

== domain 1 score: 2966.8 bits; conditional E-value: 0

```
sp|P01024|C03_HUMAN 1 mgptsgpsllllllthlplalgsmpysiitpnilrleseetmvleahdaaggdvpvtvtvhdfpgkklvlssektvltpatnhmgnvtf 88
m ptsgpsllllll lp+alg+pmys+itpnilrleseet+vleah qg + v+vtvhdfp+kk vls+e t l +++ vt+
C03_BOVIN 1 MKPTSGPSLLLLLLASLPMALGNPMYSMITPNILRLESEETVVLEAHGGGTIQSVTVHDFPAKKQVLSNENTQLNSNNGYLSTVTI 88
67***** PP
```

Alignments for each domain:

== domain 1 score: 2956.0 bits; conditional E-value: 0

```
sp|P01024|C03_HUMAN 1 mgptsgpsllllllthlplalgsmpysiitpnilrleseetmvleahdaaggdvpvtvtvhdfpgkklvlssektvltpatnhmgnvtf 88
mg tsgp llllllt lplalg p+y+iitpn+lrlesee +vleah+ qgd+ v+vtvhdfp+k+ vlsse t l a n+++ v +
C03_PIG 1 MGSTSGPRLLLLLLTSPLALGDPIYTIITPNVLRLESEEMVVLEAHGQGDIVSVTVHDFPAKRQVLSSETTTLNNANNYLSTVNI 88
89***** PP
```

Alignments for each domain:

== domain 1 score: 1343.7 bits; conditional E-value: 0

```
sp|P01024|C03_HUMAN 938 mnktvavrtldperlgregvqkedippadlsdqvptdesetrillqgtpvaqmtedavdaerlkhlivtpsgcgeqnmigmtptviav 1025
mnktvavrtldpe+lg+ gvqke+ip ad+sdqvp teset+illqgtpvaqmteda+d erlkhlivt sgcgeqnm+mt tviav
C03_RABIT 1 MNKTVAVRTLDPENLGQGGVQKEEIPSAIDISDQVPGTESETKILLQGTPVAQMTEDAIDGERLKHIVTSGSGCQEQNMIAMTHTVIAV 88
8***** PP
```

phmmer

- Alternative to BLAST-style sequence alignment
- Generates Hidden Markov Model profile for query sequence
- Three-stage filter for probabilistic alignment scoring against target database
- Returns top scoring matches, subject to chosen thresholds

phmmer – parallel execution

- Two parallel execution modes (mutually exclusive):
 - Threads (“pthreads”) - restricted to single node
 - Message passing (MPI) - can span many nodes
- Both use task farm / work queue model

Practical

- Log on to MareNostrum4
- Run HMMER on one core on login node
- Run HMMER in parallel on compute node using batch system
 - Multithreaded execution
- Record runtimes for different numbers of cores
- Evaluate parallel performance
- See instruction sheet