

## Lessons learned from implementing a national infrastructure in Sweden for storage and analysis of next-generation sequencing data

Martin Dahlö 

SciLifeLab, Stockholm, Sweden

### Motivation and Objectives

As sequencing get cheaper more and more researchers turn to this technology for answers to their questions. The large amounts of generated data will have to be stored somewhere, and the tools to analyse it will have to be updated constantly. UPPNEX tries to solve these problems through high performance computing, large scale and high availability storage, an extensive bioinformatics software suite, up-to-date reference genomes and annotations and a support function with systems and application experts. The software maintained on the computers are mostly the popular open source applications used by the international bioinformatics community, and comprises functions such as alignment, de novo assembly, SNP calling, methylation and RNA-seq analysis.

There are over 300 separate projects at UPPNEX which all belong to different research projects, such as the sequencing of the flycatcher and Norwegian spruce.

The focus of the poster is on the technical and organizational implementations at UPPNEX to handle its task.

### Methods

UPPNEX consists of half a cluster and storage system, and the other half is owned by Uppsala University's high performance computing (HPC) centre, UPPMAX. This sharing of resources has made it possible to compare the typical system usage by bioinformatics projects and the other more general HPC users. The attributes we have been measuring the last 2 years are storage, core hour usage and the size of the booked resources. We have summarized and plotted these attributes and compared the bioinformatics projects to the other general HPC projects (see Figure 1). We have also kept record over how many processor cores each job has used and compared this with non-bioinformatics projects.

### Results and Discussion

The number of bioinformatics projects at UPPNEX, and their individual storage usage is increasing everyday and the trend shows no signs of decreasing. The amount of core hours used by UPPNEX is not increasing as much as the storage. Looking at the core hour usage during an average day, it is clear that UPPNEX users are clearly more prone to run interactive sessions and shorter jobs than the other HPC users.

The decision to pool our resources with the existing HPC centre in Uppsala (UPPMAX) gave us a running start since the systems experts already had a lot of the infrastructure in place.

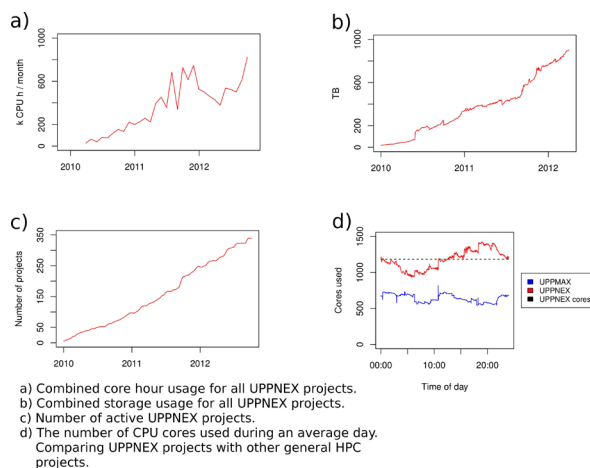


Figure 1. a) Combined core hour usage for all UPPNEX projects. b) Combined storage usage for all UPPNEX projects. c) The number of active UPPNEX projects. d) The number of cores used during an average day. Comparing UPPNEX projects (red) with other general HPC projects (blue).

### Acknowledgements

The UPPNEX project is funded by a KAW grant, and maintenance is carried out by the system experts at UPPMAX at Uppsala University.