



**Bioinformatics Meeting 21 March 2016 Vienna
University of Innsbruck**

University of Innsbruck :: Institutional Information



University of Innsbruck uibk.ac.at

- founded 1669
- 16 faculties
- 3200 faculty + 1400 staff
- 29000 students (5000 Master + 2900 PhD)
- 100 fields of study
- 270 M€ yearly budget
- 200.000 m² floor area

Faculty of Biology uibk.ac.at/fakultaeten/biologie

- 5 institutes Botany, Microbiology, Molecular Biology, Ecology, Zoology
- 2 research institutes Biomedical Aging Research, Limnology
- 93 faculty + 88 staff
- 2000 students (500 Master + 130 PhD)

Faculty of Chemistry and Pharmacy

Currently 10 groups in emerging BI + HPC initiative
43 persons involved

Research Area Scientific Computing uibk.ac.at/scientific-computing

Integrates all research activity of University of Innsbruck in Information Technology and e-Science.

- 3 research centres: **Computer Science, High Performance Computing, Computational Engineering**
planned: **Bioinformatics**
- 41 participating units (institutes + central IT services)
- Goals: develop synergies by sharing information, knowledge exchange, interdisciplinary research
coordinated procurement of IT infrastructure for HPC

University of Innsbruck :: Hardware

General purpose HPC machines

Compute Clusters LEO 3 + LEO 3e

- Commodity hardware (Intel servers + Infiniband high speed network)
- 162 + 45 nodes, 1944 + 900 cores, 3888 + 3776 GB memory, 61 + 43 TB central mass storage
- 300 users, 22 million core-hours used in 2015

Shared Memory Server MACH – 50% share – joint operation with University of Linz

- Single OS image, well-parallelized + high memory jobs
- Bioinformatics: **high memory** - MetAMOS, Trinity
- 2048 cores, 16 TB memory (total), 58 TB mass storage (uibk)
- 200 users, 8 million core-hours in 2015 (uibk)

VSC 3

- 5,4% share uibk
- 22 users from uibk, 20 million core-hours used in 2015 (large projects only)

OS: Linux (CentOS, SuSE, Scientific Linux)

Software: modules environment + self-installed software

At this time: **no special provisions for bioinformatics** (e.g. database, software, local mass storage)



University of Innsbruck :: Personnel + Infrastructural Setup



Central IT Services

- 100 staff, **5 HPC specialists** (sysadmin, software, analytics)
- run HPC infrastructure, give **support** in porting and running software
- no bioinformatics specialists at this time

Biology

- some teams starting to hire bioinformatics specialists
- planned: visiting professor (teaching)
- need: bioinformatics specialist for all institutes (bridge between IT and science – HRSM?)

Infrastructure

- Servers operated by Central IT Services and VSC
- Campus + Austrian Academic Networks (10 Gb/s ; workstations: 1 Gb/s)
- Use of installations at other institutions in cooperations
- Occasional use of cloud resources (expensive)

University of Innsbruck :: Scientific Use Cases

Examples

Scientific focus

- complex microbial communities in the environment
- parasite-host interactions based on differential gene expression analysis
- fungal biodiversity analysis

Methods

- Usual sequencing platform: Illumina, Ion Torrent, (planned: PacBio or Oxford Nanopore)
- Amplicon sequencing (multiplex sequencing, paired end approach)
- RNA sequencing

Software

- Tuxedo Suite (Bowtie 2, TopHat, Cufflinks), Trinity
- QIIME, Mothur, USEARCH, VSEARCH
- Jellyfish, Kanalyze, MG Rast, (arb)
- R, Python, Perl



University of Innsbruck :: Thoughts and Comments



IT Services + Research Area Scientific Computing perspective

- wish to satisfy needs of **bioinformatics** within integrated **HPC** infrastructure
- **current understanding:**
 - **OK:** standard Linux OS, batch system (SGE/SLURM), central mass storage, modules environment
 - **need:** high memory nodes, local disk storage (+SSD?), database + resource management
 - **questions:** alternative SW deployment (e.g. containers)?
workflow requirements ↔ cluster architecture + machine roles?
- need to replace 5 years old HPC machines LEO 3 + MACH
 - coordinate procurement HRSM (general purpose + bioinformatics)

Biology perspective

- Future developments
 - whole genome sequencing
 - 3rd generation sequencing