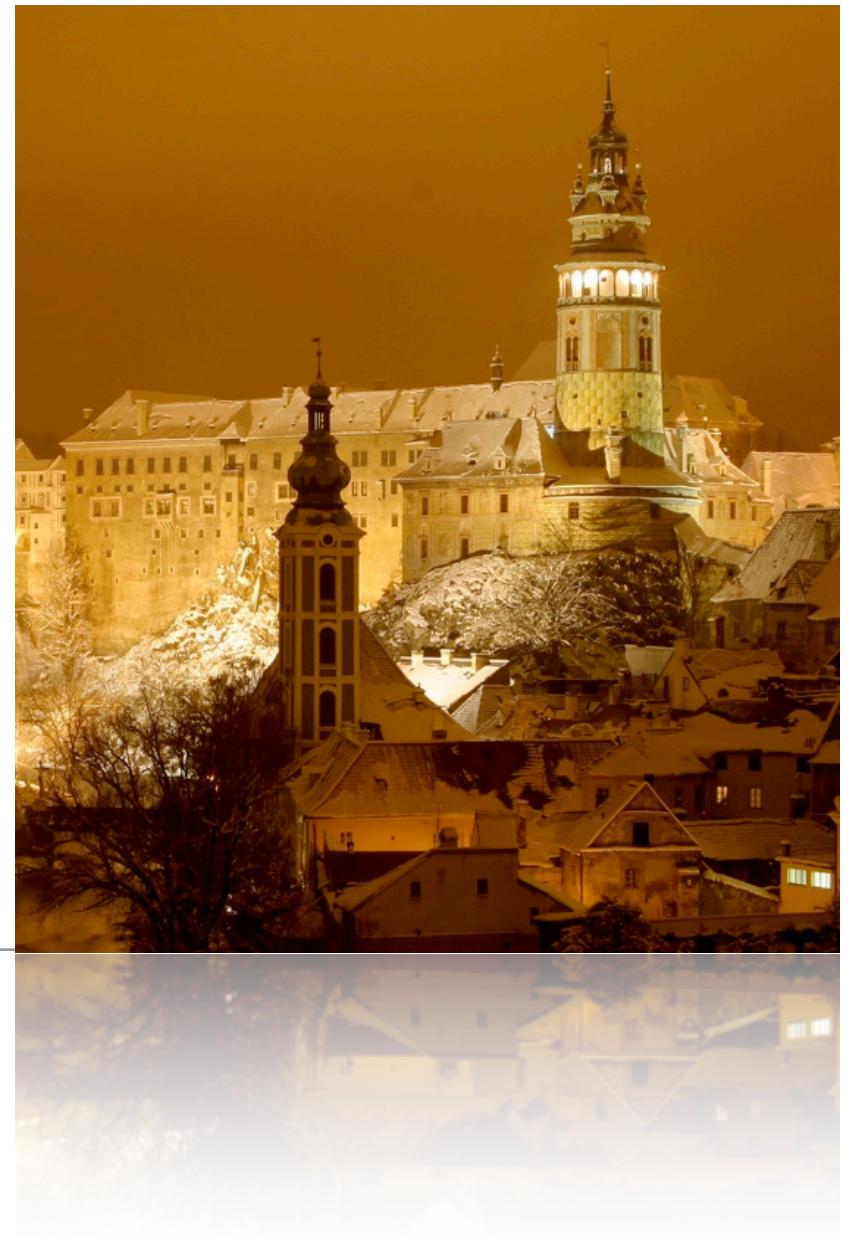# workshop on genomics

europe
2012

objectives of this presentation

- provide some information about the setting and logistics
- provide some background information about the Workshop
- help you establish an appropriate set of expectations

frequently asked questions

- Where is the restroom?
- Coffee and internet?
  - Next door, lodging, House of Prelate (computer lab)
- Where do I eat?
  - Workshop Google Map
- Where is the pub?
- Where is the computer laboratory?
  - walk after this session
- How long will the materials on the web site be available?

who

you

- proportions (n = 72)

  0.52 graduate students

  0.34 postdoctoral scholars

  0.07 faculty/principal investigators

  0.06 other

- broad international representation (n=28)

  Sweden (18), Germany (10), USA (6), Czech Republic (5), Canada (4), Spain (4), Switzerland (4), United Kingdom (4), Mexico (2), Belgium (1), Colombia (1), Estonia (1), Finland (1), India (1), Iran (1), Netherlands (1), Norway (1), Poland (1), Portugal (1), Russia (1), Thailand (1)

research interests

comparative genomics 7 (14)
genetic diagnostics 2
de novo genome assembly 11 (4)
metagenomics 4 (2)
population genomics 12 (3)
RNA seq 3 (2)
SNPs 7 (6)
transcriptomes 13 (5)
phylogenomics (3)

→

comparative genomics
genome assembly
transcriptomics
metagenomics
evolutionary and population genomics
detection of structural variation

participant presentations

- two stages (tomorrow and Wednesday)

  - small group discussions (~1.5 hours)

  - participant presentations (~1.5 hours)

- goals:

  - get to know each other

  - learn about the other participant research interests and projects

  - identify common strategies, issues and problems within sub-disciplines of the genomic sciences

instructors and co-directors

Adam Bazinet   Dag Ahren   Lionel Guy   Daniel McDonald   Tejashwari Meerupati

Scott Handley   Karin Rengefors   Naiara Rodríguez-Ezpeleta

staff research expertise

grid computing, molecular evolution and phylogenetics, RNA-seq and transcriptomics, metagenomics, and bioinformatics algorithms, web interfaces, programming, systems and workflows

comparative genomics, genomic ecology, protein family evolution and targeted metagenomics. He is primarily working with fungi but also on other microorganisms in soil ecosystems as well as a hot spring in the Yellowstone national park

evolution of host-pathogen relationships, with a specific focus on the host-adaptation systems in bacteria. Comparative genomics, genus-wide study of blood-borne pathogenic *Bartonella* spp.

staff research expertise

metagenomics. applying general evolutionary principles to well characterized phylogenies, such as the Greengenes 16s phylogeny in order to predict functions and traits of uncharacterized organisms

next generation sequence data analysis - assembly and annotation of the fungal genomes, comparative genomics studies,evolution of protein families and phylogenetics

staff research expertise

speciation and biogeography in freshwater and marine phytoplankton. Investigating genetic diversity, population genetic structure, dispersal barriers, and regions experiencing selection

using phylogenomics to resolve deep eukaryotic and bacterial relationships; high throughput sequence data analysis for differential expression/methylation analyses. using genomics to solve marine environment and fisheries related questions
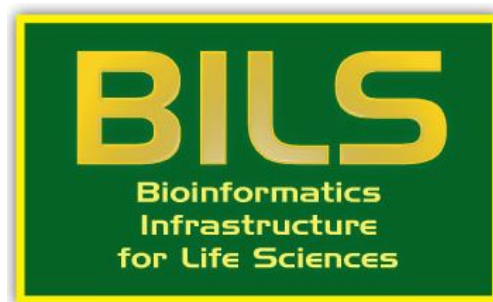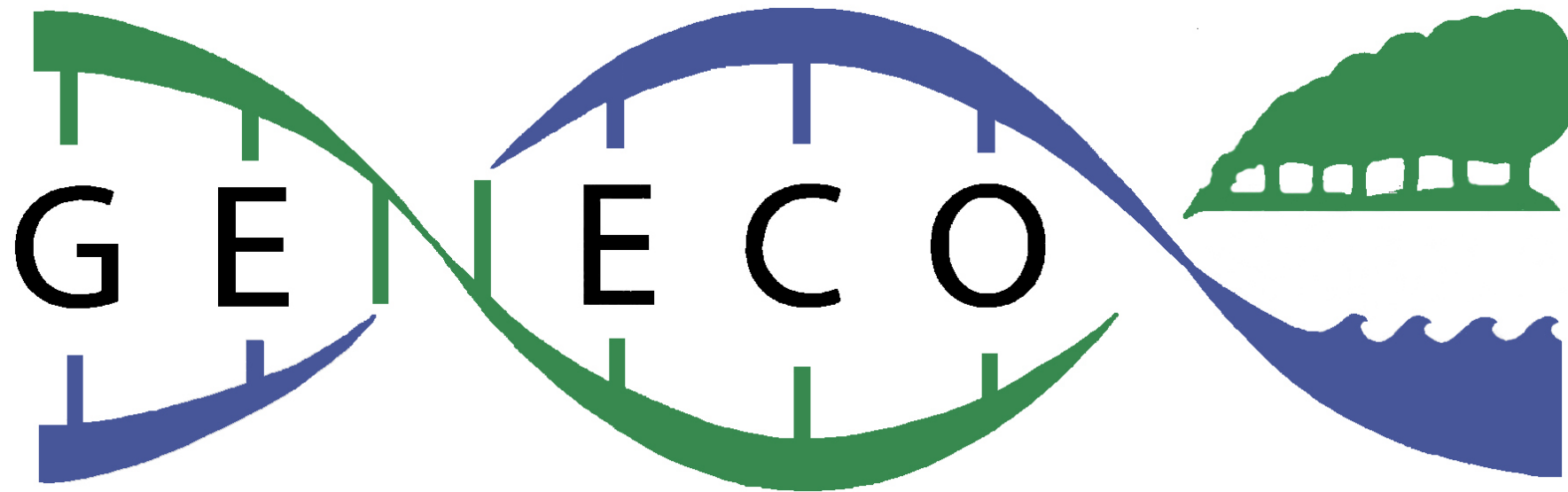

metagenomics, particularly virome characterization and novel pathogen discovery in inflammatory bowel disease. identification of therapeutic immune modulatory molecules in uncharacterized micorbial species

faculty

partners

Graduate Research School in Genomic Ecology

LUND UNIVERSITY, SWEDEN

- Interdisciplinary graduate school for PhD students in Genomic Ecology

- Genomic Ecology - interface between ecology & evolution, molecular biology and genomics

- GENECO is funded (2008-2012) by the Swedish Research Council (VR) to Lund University

GENECO Program

Trainees

- 90 past and current PhD students
- Universities in Sweden, Norway, and Finland
- Activities: Annual Winter & Summer Meetings, PhD courses, Mentor programs, travel grants


www.geneco.se

Courses

- Understanding population genetics
- Analyzing population genetic data
- AFLP for population genetics
- Fermentors, HPLC & experimental evolution
- Introduction to molecular ecology
- Quantitative PCR
- Introduction to using R
- Sequencing to assess microbial diversity: high throughput sequencing and amplicon sequencing
- Microarray analyses in non-model organisms
- Evolutionary processes
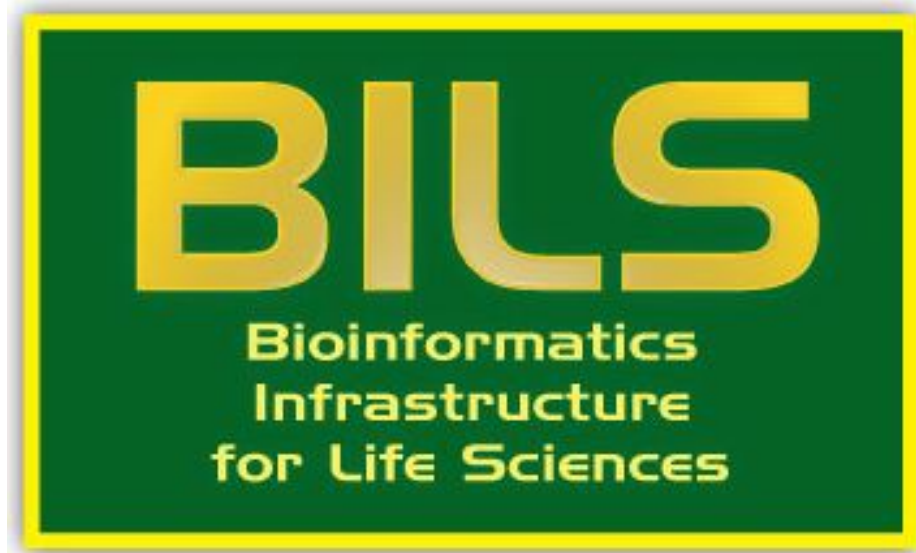
- Distributed national research infrastructure supported by the Swedish Research Council
- BILS provides bioinformatics support in various projects, e.g. in large-scale sequencing, proteomics, systems biology, metabolomics, metagenomics
- Bioinformatics network
- Nodes at each of the 6 large university cities
- Offerings:
  - annual workshop
  - training
  - Bioinformatics computation and data storage
- Current staff of 11 people – will during 2012 expand to 26 persons
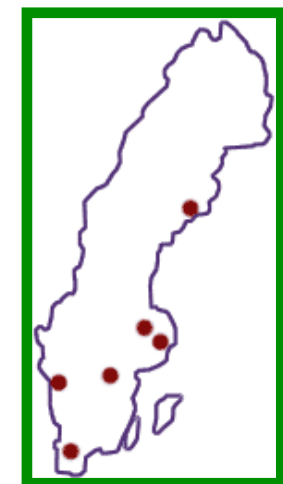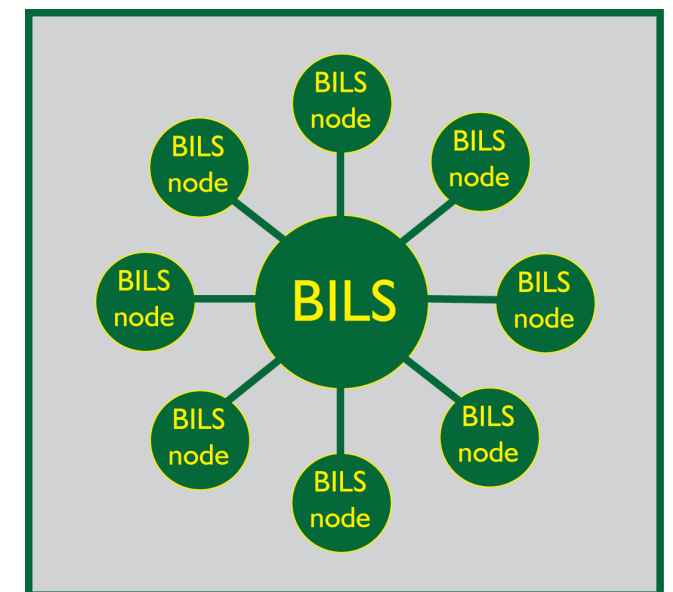
what

basis

Workshop on Molecular Evolution

- focused on molecular phylogenetics and population genetics

- Woods Hole, Marine Biological Laboratory

- Centers for Disease Control and Prevention

- Smithsonian Institution

- Workshop on Comparative Genomics

  ‣ Czech Republic - January 2011

  ‣ Fort Collins, USA - July 2011

  ‣ Smithsonian Institution - November 2011

Lessons

- immersive and intensive

- workshop not a symposium

- isolated, idylic, easy access, infrastructure

curriculum design
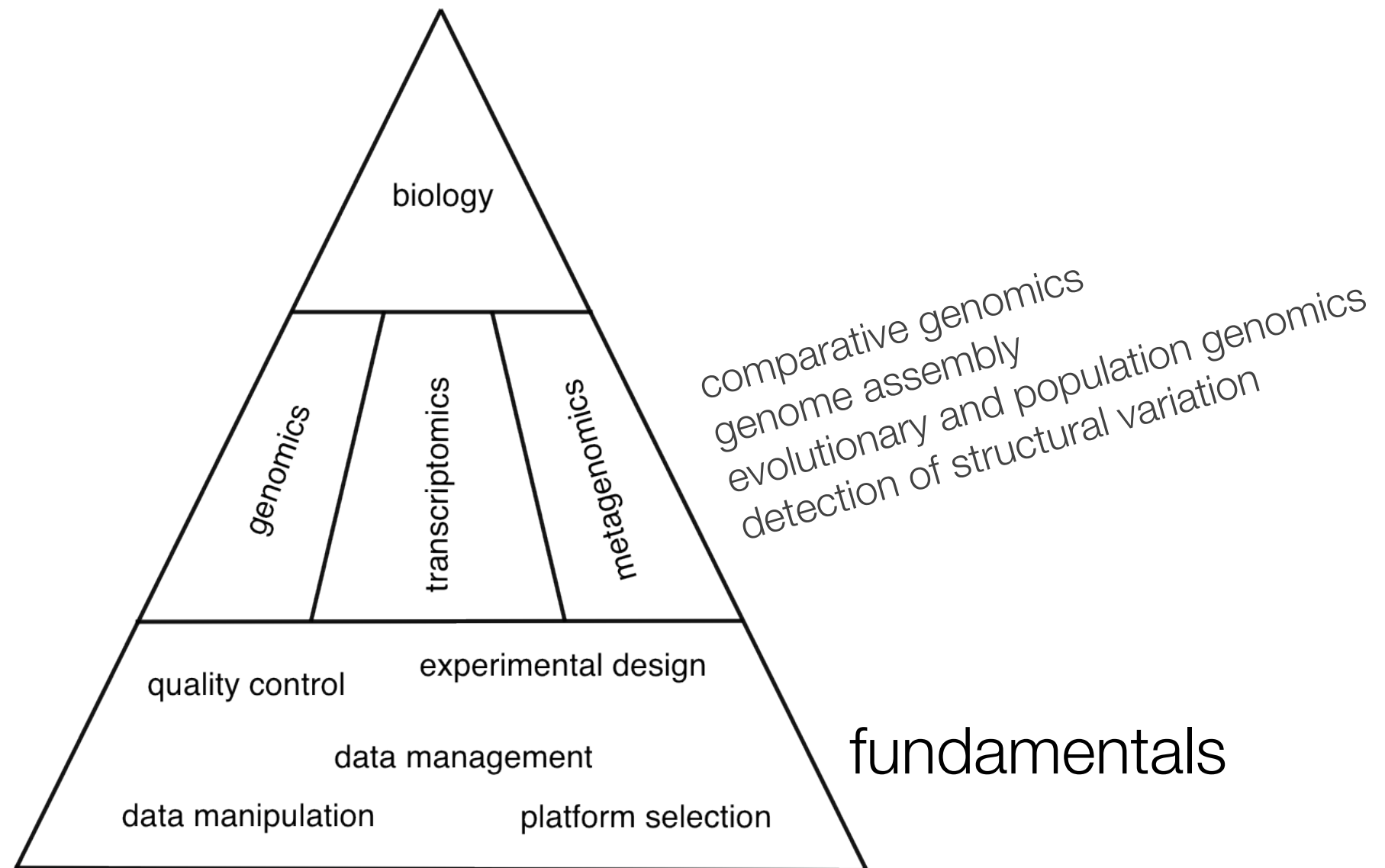
- broad range of biological disciplines and levels of experience

schedule

- Workshop website
  - http://evomics.org

Week 1

Monday

Evolution of mammalian tissue transcriptomes *Laboratory Introduction*

Tuesday

Genomics study design *Using the command line*

Participant presentations I

Wednesday

QA & QC of sequence data *Clinic I*

Participant presentations II

Thursday

Assembly

Synteny alignments

Friday

More assembly *Group photo*

Town/Brewery Tour

Saturday

Transcriptomics *Free time or party*

Week 2

Monday
NGS and non-model organisms   *Stacks*

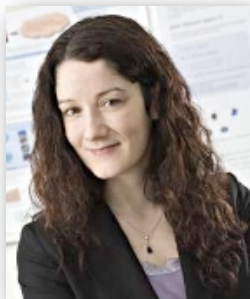*Clinic II*

Tuesday
Read mapping

Variant calling

Wednesday
Metagenomics   *Taxonomic assignment*

Thursday
Data workflows and pipelines   *Galaxy*

Friday
Evolutionary genomics

Pathogenomics

Saturday
Departure

why

need

Instrument	Run Time	Millions of reads/run	Reagent cost per run	Reagent cost/MB	Data file Sizes (GB)
3730xl capillary	2 hours	0.000096	~$100	~$1500	0.03
Illumina HiSeq 2000	8 days	1000	~$20,000	$0.10	1200

bioinformatic landscape

As of 8 January 2012 SEQwiki contains pages for 530 (373 last year) bioinformatics applications related to genomic analysis!
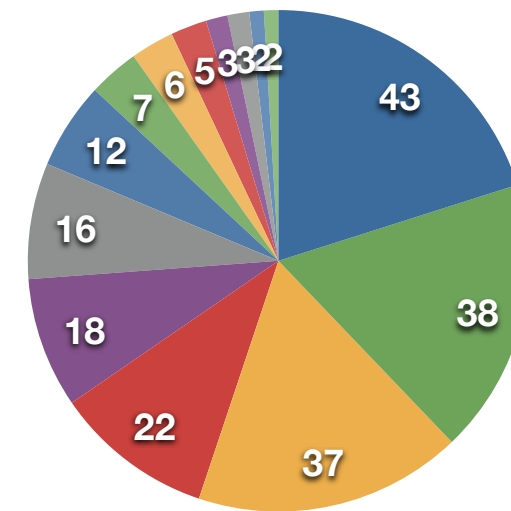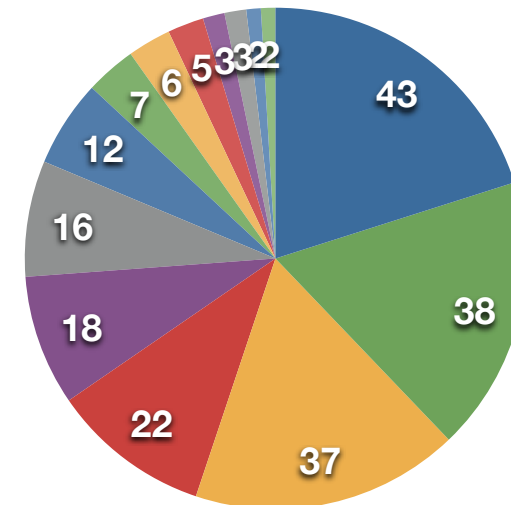

- RNA-Seq
- SNP discovery
- ChIP-Seq
- Metagenomics
- Copy number estimation
- DNA Methylation
- Whole Genome Resequencing

- Genomics
- De-novo assembly
- Structural variants
- Epigenomics
- InDel discovery
- Alignment
- Mapping

current challenge

+

Legend:
- RNA-Seq
- SNP discovery
- ChIP-Seq
- Metagenomics
- Copy number estimation
- DNA Methylation
- Whole Genome Resequencing
- Genomics
- De-novo assembly
- Structural variants
- Epigenomics
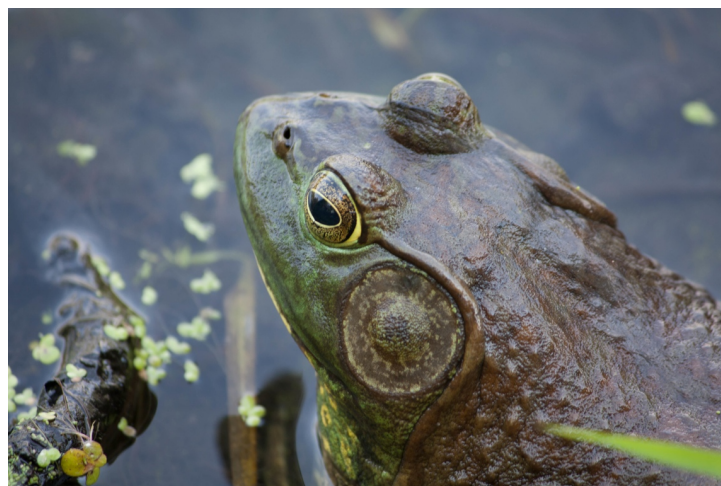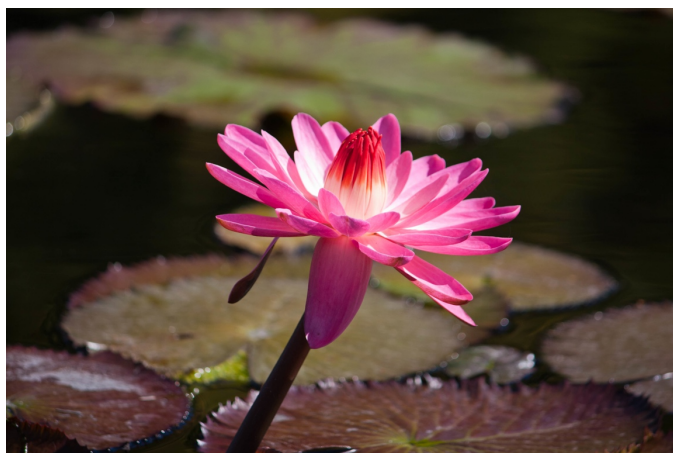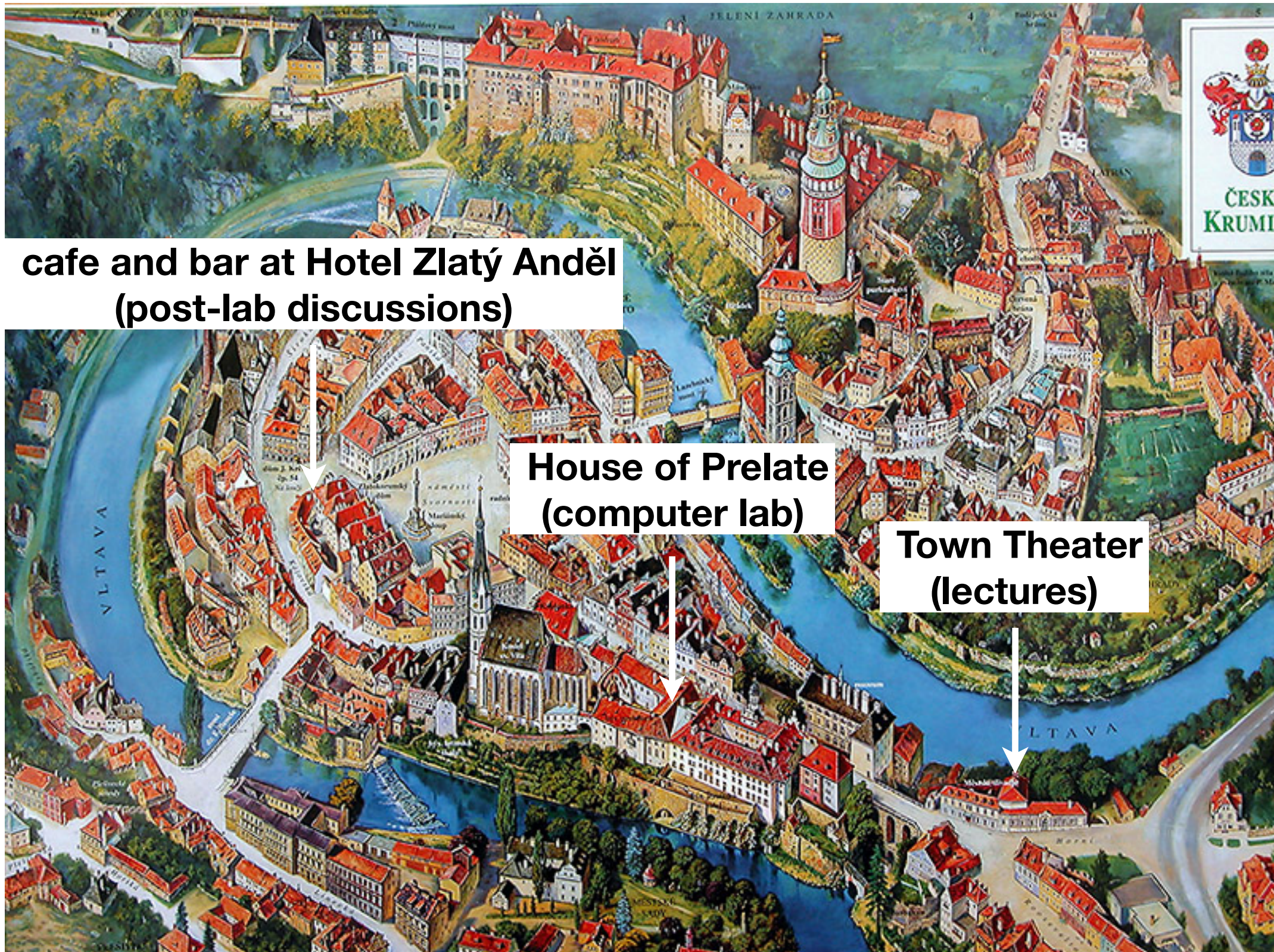- InDel discovery
- Alignment
- Mapping

where

here

cafe and bar at Hotel Zlatý Anděl (post-lab discussions)

House of Prelate (computer lab)

Town Theater (lectures)

why Český Krumlov

important considerations

- relatively inexpensive
- relatively isolated
- idyllic
- conveniently and centrally located in Europe
- appropriate infrastructure

expectations

realistic expectations

- exhaustion
  - please try to relax and drink responsibly
- computing environment
  - highly specialized customized Linux distribution
  - multiple solutions for multiple OS's
  - you may not get everything installed or finished every exercise during the allotted session
- best practices
  - work in pairs or small groups
  - open labs
  - all material is available after you leave

discussion opportunities

- the Workshop provides extensive opportunities for discussions
    - with faculty
    - with teaching assistants
    - with other participants
- these opportunities include
    - lecture and lab breaks
    - meals
    - at the cafe and bar after lab
- you are well advised to take advantage of these opportunities
- faculty arrival and departure dates are on the web site

discussion opportunities

- the Workshop provides extensive opportunities for discussions
  - with faculty
  - with instructors
  - with other participants
- these opportunities include
  - lecture and lab breaks
  - meals
  - at the cafe and bar after lab
- you are well advised to take advantage of these opportunities
- faculty arrival and departure dates are on the web site

feedback

- please provide feedback
  - topics covered
  - constructive. how to improve in future years
- end of Workshop survey

computer lab

House of Prelate

- dates back to second half of the 14th century
- a wing of the building held a brewery from 1596 until 1865
- significant renovations in 1576, 1624, 1652, 1768, 1897-1902, 1924, and last half of 20th century
- break area and coat room
- fully wired and wireless internet
- power distribution

general lab policy issues

- no food or drink inside building
- wipe feet before entering
- coats in coat room
- do not touch walls
  - the painter is unavailable for repainting (he died in 1791)
- remove your things at end of each lab session
- please wear your name tag!

post-lab discussions

- an important feature of the Workshop
- Music Bar behind Hotel Zlatý Anděl
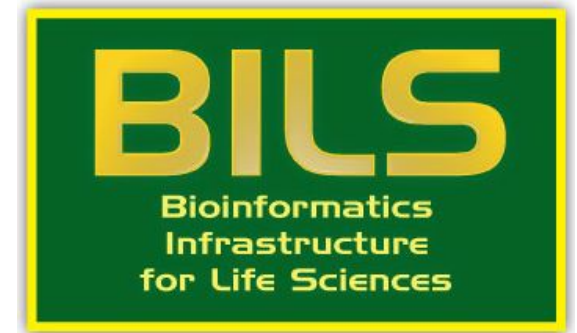- it is not about the drinking, it is about the discussions

t-shirt competition

- http://evomics.org/registration-form/t-shirt-competition/
- the top two designs will receive a copy of the book "Bioinformatics for High Throughput Sequencing" edited by Naiara Rodríguez-Ezpeleta
- package of souvenirs and specialties from around Cesky Krumlov
- fame, fortune! your design will be worn by people from around the world for years to come

thanks

**Graduate Research School in Genomic Ecology
Lund, Sweden**

Adam Bazinet   Dag Ahren   Lionel Guy   Daniel McDonald   Tejashwari Meerupati

Scott Handley   Karin Rengefors   Naiara Rodríguez-Ezpeleta