Semester Course – Computational Microbial Genomics

Class	Topic	Lab/Homework
1	Course Introduction and Basics	Readings on NGS technologies, etc.
2	Introduction to Linux I	Basic Linux exercises
3	Introduction to Linux II	More Linux exercises
4	Introduction to NCBI	Entry search and download
5	Sequence Data Storage, CG pipeline and QA/QC	QC exercises
6	Genome Assembly I	Assembly tools
7	Genome Assembly II	Comparison and merging tools
8	Pairwise Sequence Alignment	NW/SW, dot plot
9	Sequence Comparison and BLAST	BLAST based exercises
10	Average Nucleotide Identity	ANIb, MUMmer
11	Genome Alignment	Mauve
12	Gene Prediction I	Gene prediction
13	Gene Prediction II	Merging and formats
14	Functional Annotation	Annotating predicted genes
15	Pathway analysis and KEGG	KAAS
16	Phylogenetic Reconstruction	Gene tree construction
17	Multi-Sequence Sequence Typing	MLST based analysis
18	Detecting virulence factors	VF prediction and visualization

Description and SMART learning objective:

This bioinformatics course provides trainees experience with the commonly used bioinformatics programs and protocols for the computational analysis of microbial genome sequences. Topics covered include discussions and exercises on Next Generation Sequencing (NGS) read quality assessment, genome assembly, gene/feature prediction, functional annotation, multi-locus sequence typing, average nucleotide identity, virulence factor detection, horizontal gene transfer detection, genome alignment and visualization and microbial genomics databases. Many of these topics also include visualization techniques to assist in result analysis and presentation. The course will involve basic Linux command line usage, and sufficient information will be provided to the trainees in the first few classes of the course. Tools that are used include command line tools, GUIs and webservers such as FastQC, Velvet, QUAST, Prodigal, RNAmmer, BLAST, SignalP, Pfam, InterPro, MEGA, JSpecies, TM4 MeV, Alien_hunter and Mauve. Discussions and tours of microbial genomics databases include NCBI and KEGG.

Learning objective: By the end of the course, the trainees are expected to have proficiency to correctly perform basic bioinformatics sequence analysis. More specifically, the trainee will have demonstrable understanding in the following areas of bioinformatics:

- Sequence storage formats and conversions
- Accessing and retrieving data from NCBI and its components
- Computational genomics pipeline: genome assembly, gene prediction and functional annotation
- Basic comparative genomics analyses such as MLST, virulence factor detection and phylogenetic reconstruction

Semester Course – Computational Sequence Analysis

Class	Topic	Lab/Homework
1	Course Introduction and Basics	Readings on NGS technologies, etc.
2	Introduction to Linux I	Basic Linux exercises
3	Introduction to Linux II	More Linux exercises
4	Advanced Linux Operations	AWK and Regular expressions
5	Introduction to NCBI	Entry search and download
6	Programmatic Access to NCBI I	Eutilities
7	Programmatic Access to NCBI II	Edirect
8	Kyoto Encyclopedia of Genes and Genomes	Pathway exercises
9	Pairwise Sequence Alignment	NW/SW, dot plot
10	Multiple Sequence Alignment	MSA Exercises
11	Sequence Comparison and BLAST	BLAST based exercises
12	Phylogenetic Reconstruction	Gene tree construction
13	Average Nucleotide Identity	ANIb, MUMmer
14	Genome Alignment	Mauve
15	Sequence Data Storage, CG pipeline and QA/QC	QC exercises
16	Reads to Genome Mapping	Sequence Mapping Exercises
17	Variant calling	Variant calling exercises
18	Variant filtering and comparison	VCF based operations

This bioinformatics course provides trainees with the skills and knowledge necessary for various methods of analysis of molecular sequences. Topics covered include access and retrieval of sequence data, tools and methods for aligning short read sequences to a genome, calling variants, aligning a pair of or multiple short sequences, the basics of molecular phylogenetics, utilizing sequence alignment for database searching, aligning a pair of or multiple genomes and rearrangement analysis. Advanced Linux command line usage is involved and the first part of the course is dedicated to cover this. Tools that are used include GUIs, webservers and command line tools such as FastQC, BWA, SAMtools, BCFtools, EMBOSS suite, MEGA, Mauve and BLAST among others.

Learning objective: By the end of the course, the trainees are expected to have proficiency to correctly perform basic bioinformatics sequence analysis. More specifically, the trainee will have demonstrable understanding in the following areas of bioinformatics:

- Sequence storage formats and conversions
- Accessing and retrieving data from NCBI and its components
- Computational genome comparison
- Variant calling and VCF based operations

Semester Course – Linux for Bioinformatics

Class	Topic
1	Introduction to Linux Environment
2	Linux File System
3	Basic System Administration
4	Accessing and retrieving files from Linux
5	Introduction to Bioinformatics Databases
6	Programmatic Access to NCBI I
7	Programmatic Access to NCBI II
8	Utility Compilation and Installation with *nix
9	Basic File Handling
10	Advanced File Handling I
11	Advanced File Handling II
12	Regular Expressions
13	Exam
14	Basic of Pipeline Development
15	Basics of Shell Scripting
16	Shell loops
17	Command-line Arguments
18	SNP Calling Pipeline

NOTE: Each lecture will have accompanying lab sessions that pertain to the materials covered in the session.

This bioinformatics course provides trainees with the skills to be able to productively work in a Linux/*nix system for Bioinformatics analysis. Topics covered include the basics of Linux and file systems, file access, file and data retrieval from various sources, file handling, searching and regular expressions, pipeline development and shell scripting. By the end of the course, trainees will have sufficient knowledge to develop their own working pipelines and shell scripts that can be applied to their research. The whole course will be geared towards application of Linux in Bioinformatics.

Learning objective: By the end of the course, the trainees are expected to have demonstrable skill in Linux and creating their own pipelines and scripts. More specifically, the trainee will have understanding in the following areas of Linux and Bioinformatics:

- File handling and manipulation
- Accessing and retrieving data from NCBI and its components
- Pipeline development & shell scripting

Semester Course – Perl for Bioinformatics

Class	Topic
1	Introduction to Programming Concepts and Perl
2	Sequences and Data Structure
3	More Data Manipulation Operations
4	Loops and Control
5	Files and Devices
6	Subroutines and Modules
7	Command-line Arguments and Other Utilities
8	Mid-term Exam I
9	Regular Expressions I
10	Regular Expressions II
11	CPAN and Some Useful Modules
12	BioPerl – Installation, BioSeq, BioSeqIO
13	BLAST
14	Mid-term Exam II
15	Object Oriented Programming (OOPs) Paradigm
16	Creating Modules
17	Threads and Parallelization
18	Class Project

NOTE: Each lecture will have accompanying lab sessions that pertain to the materials covered in the session.

This bioinformatics course provides trainees with programming skills in Perl for Bioinformatics. Topics covered include the basics of programming concepts and Perl, file access, handling and manipulation, searching and regular expressions, CPAN, OOPs and basics of writing parallelized scripts. By the end of the course, trainees will have sufficient knowledge to develop their own working scripts and pipelines that can be applied to their work and research. The whole course will be geared towards application of Perl in Bioinformatics.

Learning objective: By the end of the course, the trainees are expected to have demonstrable skill in Perl and creating their own pipelines and scripts. More specifically, the trainee will have understanding in the following areas of Perl and Bioinformatics:

- Programming paradigms
- Basic data structures and related operations
- File handling and manipulation
- Regular expressions
- Scripts and pipeline development