4-day Module – Introduction to Scientific Computing

Session	Topic
1	Introduction to Linux Environment
2	Linux File System & Basic System Administration
3	Utility Compilation and Installation with *nix
4	Basic File Handling
5	Advanced File Handling I
6	Advanced File Handling II
7	Regular Expressions
8	Basics of Pipeline Development
9	Basics of Shell Scripting
10	Shell loops
11	Command-line Arguments
12	Wrap Up

Course Description:

This intermediate level bioinformatics module provides participants with the essential skills required to be able to work in a Linux/*nix system when performing bioinformatics analyses. The participants will receive instruction in basic command line Linux and system administration, and will be instructed in the compilation and installation of software in the Linux environment. Topics covered include the basics of Linux and file systems, file access, file handling, searching and regular expressions, pipeline development and shell scripting. At the conclusion of the course, participants should 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.

SMART Learning Objective:

By the end of the course, the participants are expected to have proficiency in utilizing the power of Linux's command line in scientific computing and bioinformatics. More specifically, participants will gain demonstrable understanding in the following areas:

- File handling and manipulation and program installation in a Linux/*nix environment
- Access and retrieval of data from NCBI and its components
- Pipeline development & shell scripting

4-day Module – Introduction to Computational Microbial Genomics

Session	Topic
1	Next-Generation Sequencing (NGS)
2	Genome Assembly
3	Gene Prediction
4	Functional Annotation
5	Multi-Locus Sequence Typing (MLST)
6	Average Nucleotide Identity (ANI)
7	Virulence Factor Detection
8	Horizontal Gene Transfer
9	Genome Alignment & Visualization
10	NCBI
11	Kyoto Encyclopedia of Genes and Genomes (KEGG)
12	Other Resources

Course Description:

This intermediate level bioinformatics module provides participants with the opportunity to gain hands on experience working with numerous bioinformatics programs and protocols that are commonly used for the computational analysis of microbial genome sequences. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis of next-generation sequencing (NGS) data. Course material includes discussions and exercises involving 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. Visualization techniques to assist in the analysis and presentation of results will also be demonstrated. Bioinformatics programs for which instruction is provided include webservers, graphical user interface (GUI) and command line tools such as FastQC, Velvet, 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.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency performing basic bioinformatics analyses of microbial data. More specifically, participants will gain demonstrable understanding in the following areas:

- Computational genomics pipeline: genome assembly, gene prediction and functional annotation
- Basic comparative genomics analyses such as MLST, virulence factor detection and phylogenetic reconstruction
- Efficiently access and retrieval of sequences from publically accessible databases

4-day Module – Computational Sequence Analysis

Session	Topic
1	Sequencing Terminologies
2	Genome Mapping
3	Variant Calling
4	Pairwise Sequence Alignment
5	Multiple Sequence Alignment
6	Phylogenetic Construction
7	Genome Alignment
8	Feature Prediction
9	Protein and Structure Representations
10	Quantitative Genome Comparisons
11	Sequence Similarity Searching
12	File Access

Course Description:

This intermediate level bioinformatics module provides participants with the opportunity to gain hands on experience working with numerous commonly used bioinformatics programs and includes descriptions of the theories and protocols used to perform various types of analyses of molecular sequences. The participants will receive instruction in basic command line Linux and gain experience working at the command line throughout the module. Participants will be instructed in the methods and tools 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, predicting features in a genome and protein and structure representation. Bioinformatics programs for which instruction is provided include graphical user interface (GUI), webservers and command line tools such as FastQC, BWA, SAMtools, BCFtools, the EMBOSS suite, MEGA, Mauve, Prodigal, Glimmer, RNAmmer, Jmol, JSpecies and BLAST among others.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency performing basic bioinformatics sequence analysis. More specifically, participants will gain demonstrable understanding in the following areas:

- Sequence storage formats and conversions
- Theory and application of alignment methods
- Alignment of short read sequences to a genome (*i.e.,* sequence reads)
- Alignment of two or more small sequences (*e.g.* genes)
- Genome alignments
- Searching for sequences similar to an input sequence

4-day Module – Introduction to Bioinformatic Databases

Session	Topic
1	Bioinformatics Background & Sequence Formats
2	Introduction to Scientific Computing
3	Basic Local Alignment & Search Tool
4	NCBI
5	Functional Annotation Databases
6	Human Health Related Databases
7	Kyoto Encyclopedia of Genes and Genomes (KEGG)
8	Virulence Factor Detection
9	Metagenomics
10	Multi-Locus Sequence Typing (MLST)
11	Custom Genome Databases
12	Other Analytical Tools

Course Description:

This intermediate level bioinformatics module provides participants with an overview of different database resources available for bioinformatics along with some discussions on implementation of bioinformatics databases and their associated tools. Course material includes an introduction to bioinformatics, next-generation sequencing (NGS) technologies, database searching utilities, access and retrieval of data from various general and specialized databases and the basics of setting up custom built databases. Databases that will be discussed include NCBI, specialized databases for functional annotation, human health related databases, Kyoto Encyclopedia of Genes and Genomes (KEGG), Virulence Factor Database (VFDB), metagenomics databases and PubMLST. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analyses.

SMART Learning Objective:

By the end of the course participants are expected to have proficiency in utilizing bioinformatics databases. More specifically, participants will gain demonstrable understanding in the following areas:

- Differentiation between primary and secondary databases
- Sequence storage formats and conversions
- Applications of databases in specific areas of research
- Access and retrieval of data from databases discussed in this module and their components

4-day Module – Comparative Genomics

Session	Topic
1	Pairwise Sequence Alignment
2	Multiple Sequence Alignment
3	Introduction to Scientific Computing
4	Sequence Based Database Searching
5	Average Nucleotide Identity (ANI)
6	Genome Alignment
7	Molecular Phylogenetics I
8	Molecular Phylogenetics II
9	Multi-Locus Sequence Typing (MLST)
10	Clustering Analysis
11	Hands On Session
12	Other Sequencing Technologies

Course Description:

This intermediate level bioinformatics module provides participants with the commonly used methods and techniques used for comparative analysis of genomic features. Course material includes discussions and exercises on the techniques for aligning molecular sequences, methods for performing genome-scale comparisons and some of the molecular typing tools. The participants will be instructed on aligning a pair of or multiple short sequences, molecular phylogenetics, sequence based database searching, aligning a pair of or multiple genomes, multi-locus sequence typing and gene/protein clustering. Discussions and exercises are included with each of the topics as well as time where the participants can practice the techniques on their own. Tools that are used include graphical user interface (GUI), webservers and command line tools such as the EMBOSS suite, MEGA, BLAST suite, JSpecies and Mauve among others. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in the performance of bioinformatics comparative analyses. More specifically, participants will gain demonstrable understanding in the following areas:

- Sequence alignments and visualizations
- Phylogenetic reconstruction and comparison
- Genome level quantitative comparisons

4-day Module – Genomic Approaches to Molecular Epidemiology and Typing

Session	Topic
1	Phylogenetic Reconstruction
2	Multi-Locus Sequence Typing (MLST)
3	Introduction to Scientific Computing
4	Next-Generation Sequencing (NGS)
5	Genome Assembly
6	Assembly Comparison & Merging
7	Searching for Allele Profiles Using Sequences
8	Custom Typing Schemes
9	Average Nucleotide Identity (ANI)
10	Genome Mapping
11	SNP Calling
12	Other Analytical Tools & Third Generation Sequencing Technologies

Course Description:

This intermediate level bioinformatics module provides participants with methods for the identification and *in silico* typing of pathogens in outbreak control. Course material includes the evolutionary underpinning of MLST based typing methods, the concept of MLST, its current implementation, associated resources and tools, next-generation sequencing (NGS) technologies and their influences on MLST, variant detection and the concept of SNP based typing and the general concepts behind creation of custom typing schemes. Participants will also be instructed on underlying algorithms for the tools covered. Tools that are used include a mix of graphical user interface (GUI), webservers and command line tools such as MEGA, PubMLST, Velvet, SOAPdenovo2, QUAST, CISA, BLAST, JSpecies, BWA, SAMtools, BCFtools and IGV among others. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency to correctly perform *in silico* molecular typing using genomic approaches. More specifically, the participants will gain demonstrable understanding in the following areas:

- Multi-locus sequence typing (MLST) and interpretation of results
- Computational genomics route for MLST
- Searching for allele profiles using local databases
- Genome mapping and SNP calling for SNP-based typing schemes
- Use of existing tools to aid in the development of custom typing schemes

4-day Module – Clinical Metagenomics

Session	Topic
1	Next-Generation Sequencing (NGS)
2	Introduction to Scientific Computing
3	Subtraction Technique I: Genome Mapping
4	Subtraction Technique I: Read Extraction
5	Bioinformatics Databases
6	Sequence Similarity Searching
7	Basics of Phylogenetics
8	BLAST & Diversity Analysis
9	Species Definition and Average Nucleotide Identity (ANI)
10	Genome Assembly
11	K-mer based techniques
12	Other Resources

Course Description:

This intermediate level bioinformatics module provides participants with concepts and methods involved in clinical metagenomics. Course material includes the next-generation sequencing (NGS) technologies, contemporary computational genomics techniques, metagenomics and analytical protocols in clinical projects. Participants will also be instructed on underlying algorithms for the tools covered. Tools that are used include a mix of graphical user interface (GUI), webservers and command line tools such as FastQC, Velvet, SOAPdenovo2, BLAST, JSpecies, BWA, SAMtools, BCFtools, IGV and Kraken among others. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analyses.

SMART Learning Objective:

By the end of the course, participants are expected to have understanding of concepts and protocols involved in clinical metagenomics analysis. More specifically, the participants will gain demonstrable understanding in the following areas:

- Overview of contemporary genomics and protocols involved
- Subtraction methods in metagenomics
- Sample diversity analysis

4-day Module – The National Center for Biotechnology Information (NCBI)

Session	Topic
1	Bioinformatics Background & Sequencing Formats
2	NCBI I
3	NCBI II
4	Introduction to Scientific Computing
5	BLAST I
6	BLAST II
7	NCBI FTP Server
8	Programmatic Access to NCBI I
9	Programmatic Access to NCBI II
10	Hands On Session
11	NCBI Genome Workbench
12	Other NCBI tools/databases

Course Description:

This intermediate level bioinformatics module provides participants with the opportunity to gain hands on experience working with different components of NCBI used for bioinformatics analysis. Course material includes discussions and exercises on NCBI's sequence storage format, tour of different components of NCBI, sequence searching through BLAST and its variants, programmatic access to NCBI, writing custom data analysis pipelines and NCBI's Genome Workbench. The participants will receive instruction in basic command line Linux and gain experience working at the command line to perform bioinformatics analysis.

SMART Learning Objective:

By the end of the course, participants are expected to gain proficiency in the efficient use of and retrieval of data from NCBI. More specifically, participants will gain demonstrable understanding in the following areas:

- Sequence storage formats and conversions
- Understanding different components of NCBI and their applications
- Sequence searching via BLAST, underlying parameters and different variants
- Programmatic access to NCBI and creating custom data analysis pipelines

4-day Module – Perl for Bioinformatics Programming

Session	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	Regular Expressions I
9	CPAN and Some Useful Modules
10	BioPerl – Installation, BioSeq, BioSeqIO
11	BLAST
12	Wrap Up

Course Description:

This intermediate level bioinformatics module provides participants with the opportunity to gain hands on experience working with the Perl programming language, a commonly used tool in the analysis of biological data. Course material includes the basics of programming concepts and Perl, file access, handling and manipulation, searching and regular expressions and CPAN. By the end of the course, participants will have sufficient knowledge to develop their own working Perl scripts and pipelines that can be applied to their work and research. The whole course will be geared towards application of Perl in bioinformatics.

SMART Learning Objective:

By the end of the course, the participants are expected to have demonstrable skill in Perl and creating their own pipelines and scripts. More specifically, participants 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

4-day Module – Next-Generation Sequencing

Session	Topic
1	Sequencing Generations & Sequencing Formats
2	Introduction to Scientific Computing
3	Quality Control
4	Pyrosequencing (Single end)
5	Illumina Sequencing (Single end)
6	Illumina Sequencing (Paired end)
7	Assembly Comparison & Merging
8	Assemblathon (Hands On Session)
9	Genome Alignment & Visualization
10	Genome Mapping
11	SNP Calling
12	Other Sequencing Technologies

Course Description:

This advanced level bioinformatics module provides participants with the opportunity to gain hands on experience in handling and analysis of next-generation sequencing (NGS) data. Course material includes bioinformatics aspects of sequencing basics, 454 sequencing, Illumina single end and paired end sequencing, genome visualization, genome mapping and variant calling. Discussions and exercises will be included with each topic and participants will be provided with adequate time to practice the techniques on their own. The module will involve a mix of graphical user interface (GUI) and command line tools such as FastQC, PRINSEQ, Ray, MIRA, Velvet, SOAPdenovo2, ABySS, QUAST, CISA, Mauve, BRIG/CCT, BWA, SAMtools, BCFtools and IGV among others. The participants will receive instruction in command line Linux and gain experience working at the command line to perform bioinformatics analysis of NGS data.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in correctly performing genome assemblies and reads to genome mapping. More specifically, the participant will gain demonstrable understanding in the following areas:

- Quality assessment and control of sequencing reads
- Assembly of reads from different sequencing platforms (454 and Illumina)
- Assembly comparison and merging
- Genome alignment and visualization
- Genome mapping and SNP calling

4-day Module – Introduction to Computational Genomics

Session	Topic
1	Sequencing Generations & Sequencing Formats
2	Introduction to Scientific Computing
3	Quality Control
4	Illumina Sequencing (Paired end)
5	Assembly Comparison & Merging
6	Assemblathon (Hands On Session)
7	Prokaryotic Gene Prediction
8	Eukaryotic Gene Prediction
9	Sequence Similarity Searching
10	Genome Mapping
11	SNP Calling
12	Other Sequencing Technologies

Course Description:

This advanced level bioinformatics module provides participants with the opportunity to gain hands on experience in genome assembling, gene prediction and reference mapping with Next Generation Sequencing (NGS) data. Course material includes bioinformatics aspects of sequencing basics, Illumina paired-end sequencing, prokaryotic and eukaryotic gene prediction, homology based prediction validation, genome mapping and variant calling. Discussions and exercises will be included with each topic and participants will be provided with adequate time to practice the techniques on their own. The module will involve a mix of graphical user interface (GUI) and command line tools such as FastQC, PRINSEQ, Velvet, SOAPdenovo2, QUAST, CISA, GeneMark Suite, Glimmer, Prodigal, BLAST, BWA, SAMtools, BCFtools and IGV among others. The participants will receive instruction in command line Linux and gain experience working at the command line to perform bioinformatics analysis of NGS data.

SMART Learning Objective:

By the end of the course, participants are expected to have proficiency in understanding and executing the computational genomics protocol with NGS data. More specifically, participants will gain demonstrable understanding in the following areas:

- Computational genomics pipeline: genome assembly, gene prediction
- Quality control and assessment of sequencing reads
- Gene prediction and validation
- Genome mapping and SNP calling

4-day Module – High-Performance and Cluster Computing

Session	Topic
1	Module introduction and Virtual Machines
2	Introduction to Linux Environment
3	Linux File System & Basic System Administration
4	Utility Compilation and Installation with *nix
5	Basic File Handling
6	Advanced File Handling
7	Regular Expressions
8	Version Control and Examples
9	Perl: Introduction
10	CPAN and BioPerl
11	Basics of Parallel Computing: Threading & Parallelization
12	Cloud Computing

Course Description:

This advanced level bioinformatics module provides participants with the opportunity to gain hands on experience in learning about different computational technologies involved in providing high-performance support and cluster computing. Course material includes virtual machine and setting up your own VM instances, Linux and basic administration in Linux, utility installation, file handling, regular expression, version controls, introduction to Perl, threading and cloud computing. Discussions and exercises will be included with each topic and participants will be provided with adequate time to practice the techniques on their own. This module is geared towards helping participants gain substantial exposure to the computer technologies and will involve heavy command line usage throughout the course.

SMART Learning Objective:

By the end of the course, participants are expected to have an understanding of different technological innovation and their applications in a scientific environment. More specifically, participants will gain demonstrable understanding in the following areas:

- Virtual Machines
- Overview of Linux Operating systems
- Version controls
- Exposure to Perl, parallelization and cloud computing and their application to bioinformatics