



# **Faculty of Science**

## **Laboratory Manual**

### **Bioinformatics**

**Bachelor of Biotechnology (Hons.)**

Copyright © 2018

Lincoln University College, Malaysia

All rights reserved

No part of this book can be reproduced or transmitted by any means, electronic or mechanical, including photocopying recording or by any information storage and retrieval system without prior written permission from the publisher.

*Edited By: Dr. Farhat A. Avin*

Published by:

**LINCOLN UNIVERSITY COLLEGE**

Wisma Lincoln, No. 12,14,16 & 18,  
Jalan SS 6/12, Off Jalan Perbandaran  
47301 Petaling, Jaya,  
Selangor Darul Ehsan, Malaysia

Tel.: +603-7806 3478

Fax: +603-7806 3479

Toll Free: 1-300-880-111

E-mail: [lucp@lincoln.edu.my](mailto:lucp@lincoln.edu.my)

[info@lincoln.edu.my](mailto:info@lincoln.edu.my)

Web: [www.lucp.net](http://www.lucp.net)

[www.lincoln.edu.my](http://www.lincoln.edu.my)

ISBN: 978-967-2257-20-2

## Bioinformatics

### Table of Content

Experiments:		Page
1	Laboratory safety rules	1
2	Part I: Revision of chromatograms by Chromas	6
3	Part II: Sequence alignment and amendment by Mega software	9
4	BLAST search by NCBI database	14
5	Design PCR primers and check them for specificity	16
6	Phylogenetic analysis - Phylogenetic tree & bootstrapping	19
7	Analysis of nucleotide polymorphism and molecular data (AMOVA) from aligned DNA sequenced data by DNAsp and Arlequin software	22
8	Network Program: Generate a median-joining haplotype network	33



LINCOLN UNIVERSITY COLLEGE  
FACULTY OF SCIENCE (DEPARTMENT OF BIOTECHNOLOGY)  
LABORATORY SAFETY RULES

---

The following rules must be obeyed by all students in the science laboratory of the faculty. Wilful or repeated in advertent non-compliance may result in dismissal or suspension from the laboratories

---

- **No entry without permission:**
  - Outsiders are not allowed to enter the laboratory without permission.
  - No student is allowed to enter the laboratory unless permission has been given by a laboratory assistant or a lecturer.
  
- **At work in the laboratory:**
  - No experiment may be attempted without the knowledge and permission of a lecturer.
  - Students must wear shoes in the laboratory. Students wearing slippers or sandals are not allowed to work in the laboratory.
  - Lab coat must be worn at all times during practical work in the laboratory.
  - Do not mouth pipette chemicals.
  - Do not eat or smoke in the laboratory.
  - Do not taste any chemicals, including dilute solutions. If any acid or alkali accidentally enters your eyes or mouth, wash immediately with plenty of tap water. Inform your lecturer, and seek medical attention if necessary.
  - Paper should be used to light up the Bunsen burners.
  - Used match sticks, filter papers, and other solid waste must never be thrown into the sinks. They must be thrown into the dustbins provided. Lighted match sticks and smoldering materials must be extinguished with tap water before thrown in to the dustbins.
  - Any equipment broken or damaged must be reported to the laboratory assistant.
  
- **Before leaving the laboratory:**
  - All the equipment and benches must be cleaned at the end of each practical session.
  - Wash hands and arms with soap and water before leaving the laboratory.
  - No student is allowed to take away any chemicals, equipment or other property of the laboratory.

## INTRODUCTION

### 1. The Scientific Method

- Making observations
- Generating hypotheses
- Making predictions
- Designing and carrying out experiments
- Constructing scientific models

### 2. Practical Exercises

To get the most out of the practical exercises, you need to follow carefully the instructions given. These instructions have been designed to provide you with the experience in the following skills:

- Following instructors
- Handling apparatus
- Having due regard for safety
- Making accurate observations
- Recording results in an appropriate form
- Presenting quantitative results
- Drawing conclusions

### 3. Following Instructions

Instructions are provided in the order in which you need to carry them out. We would advise that before carrying out the instructions, you read through the entire exercise. This will help you to remember what you have learned.

Each practical exercise in the book begins with a few lines describing its purpose in most cases the following headings are also used:

- Procedure-numbered steps that need to be carried out.
- For consideration -some questions to help you think carefully about the results you have obtained.
- Materials-a list of the apparatus, chemicals and biological materials you need.

### 4. Handling apparatus

Biologists need to be able to use many different types of apparatus, for example, photometers (to measure water uptake by plants), respirometers (to measure oxygen uptake or carbon dioxide production), Petri dishes (for plating out bacteria and other microorganisms) and the light microscope (to magnify specimens). Many of the practical exercises are designed to help you derive the maximum benefit from a piece of apparatus.

### 5. Having Due Regard for Safety

Surveys have been shown that science laboratories are among the safest places to be. Nevertheless, this is no cause for complacency.

- Always move slowly and carefully in a laboratory.
- Never put your fingers in your mouth or eyes after using chemicals or touching biological specimens until you have washed your hands thoroughly with soap and warm water, and dried them.
- Make sure glass objects (e.g, thermometers, beakers) cannot roll off tables or be knocked onto the floor.
- Wear safety goggles whenever there is a risk of damage to the eyes.

Situations of risk include:

- Heating anything with a Bunsen burner (even heating water has its dangers')
- Handling many liquids, particularly those identified as corrosive, irritant, toxic or harmful
- Handling corrosive or irritant solids
- Some dissection work
- Allow Bunsen burners, tripods, gauzes and beakers to cool down before handling them.
- Never allow your own body fluids (especially blood and saliva) to come into contact with someone else, or theirs into contact with you.
- Keep long hair tied back and do not wear dangly earrings.
- Do not allow electrical equipment to come into contact with water.
- If you are unsure how to carry out a scientific procedure, ask.
- Make sure you understand why you are going to do something before you do it.
- Wear a lab coat when using chemicals or handling any biological specimens.
- Follow exactly agreed procedures with regard to cuts, burns, electric shocks and other accidents (e.g. with chemicals).
- Follow exactly all specific safety instructions given in this book or provided by your teacher for particular practical exercises (e.g. use of gloves, disinfection)

With practice, these procedures should become second nature to you. They will enable you to carry out practical work in safety.

## 6. Making Accurate Observations

In most cases the practical exercise will make it clear what you need to observe, e.g. the time taken for a certain volume of gas to be evolved or the width of a sample cells. Ensure that you know how to use any necessary equipment before starting practical. Think carefully about the precision with which you will make your observations.

## 7. Recording Results in an Appropriate Form

Results can be recorded in various ways. Often it is helpful to record raw data in a table. Most data will be in the form of numbers, i.e. they will be quantitative data (also known as numerical data). However, some data, e.g. flower colour, will be qualitative data.

One form in which some biological findings can be recorded is a drawing. You don't need to be professional artist to make worthwhile biological drawings. If you follow the following guidelines, a drawing can be of considerable biological value:

- Ensure that your completed drawing will cover at least a third of A4 page.
- Plan your drawing so that the various parts are in proportion and will not be drawn too small. Small marks to indicate the length and breadth of the drawing are a great help in planning and a faint outline can be rapidly drawn to show the relative positions of the parts.
- The final drawing should be made with clean, firm lines using a sharp HB pencil and, if needed, a good quality eraser (not a fluid). If important details are too small to be shown in proportion, they can be put in an enlarged drawing at the side of the main drawing.
- Avoid shading and the use of colour unless you are an excellent artist and they really help, for example when drawing soil profiles.
- When drawing structures seen with the naked eye or hand lens, use two lines to delineate such things as blood vessels and petioles. This will help you to indicate the relative widths of such structures.
- When drawing low power plan drawings from the light microscope, do not attempt to draw individual cells-just different tissues.
- When drawing plant cells at high power under the light microscope, use two lines to indicate the width of cell walls, but a single line to indicate a membrane.
- Always put a scale on each drawing.

## 8. Presenting Quantitative Results

Presentation of data is all about using graphs or other visual means to make it easier to see what your results tell you. The following four ways of presenting data are the most frequently used in biology: line graphs, bar charts, histograms and scatter graphs (Figure 1).

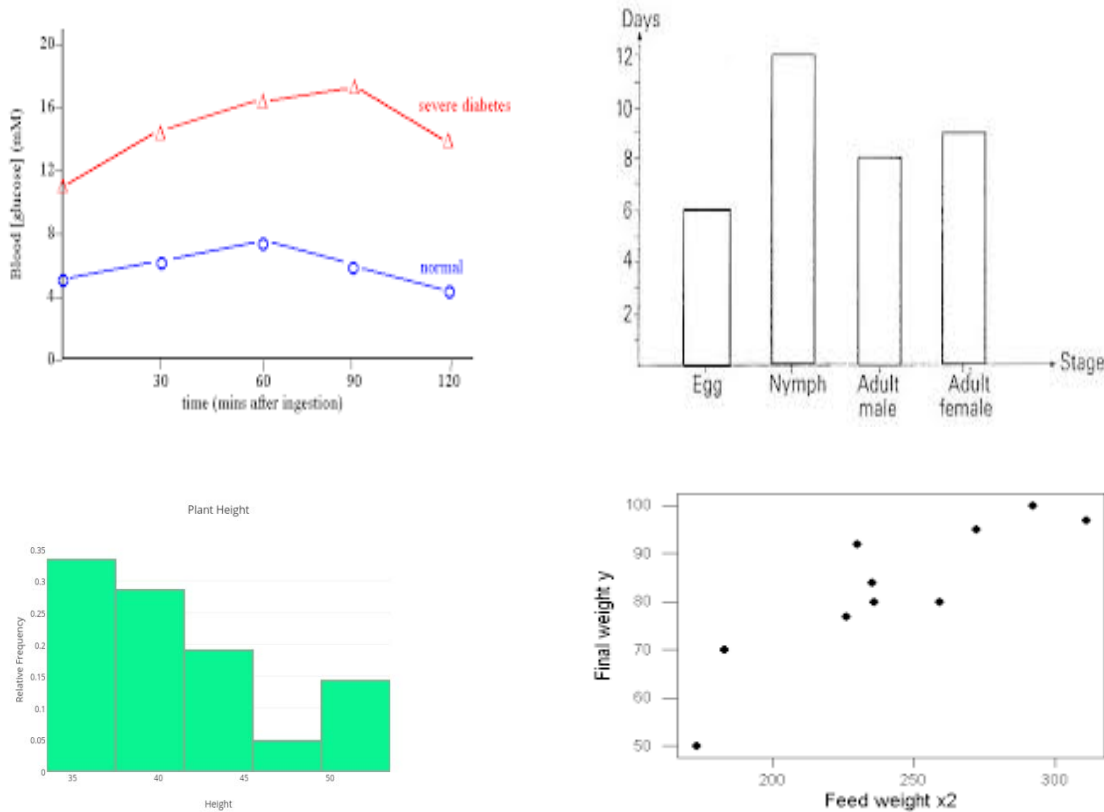


Figure 1: Line graphs, bar charts, histograms and scatter graphs

## 9. Drawing Conclusions

Finally, you will need to draw conclusions. If your practical exercise has involved the testing of a hypothesis, for example that the enzyme pepsin works better at low pH than in neutral or alkaline conditions, your conclusion should indicate whether the hypothesis has been refuted (i.e. shown not to be the case) or supported. Of course, even if your hypothesis has been supported, it doesn't mean that it has been confirmed with 100% certainty- in other words it isn't proved. Science proceeds more by showing that certain ideas are wrong than by showing that others are right (think about that!). Your conclusion might therefore include further ways of testing the original hypothesis, or might raise new possibilities to be investigated.

Often you will only be able to arrive at your conclusions after statistically analysing your data.

## 10. Writing a Scientific Lab Report

### Title

- Communicate the subject investigated in the paper.

### Introduction

- State the hypothesis.
- Give well-defined reasons for making the hypothesis.



- Explain the biological basis of the experiment.
- Cite sources to substantiate background information.
- Explain how the method used will produce information relevant to your hypothesis.
- State a prediction based on your hypothesis. (If the hypothesis is supported, then the results will be.)

### **Materials and Methods**

- Use the appropriate style.
- Give enough detail so the reader could duplicate your experiment
- State the control treatment, replication and standardized variables that were used.

### **Results**

- Summarize the data (do not include raw data).
- Present the data in an appropriate format (table or graph).
- Present tables and figures neatly so they are easily read.
- Label the axes of each graph completely.
- Give units of measurement where appropriate.
- Write a descriptive caption for each table and figure.
- Include a short paragraph pointing out important results but do not interpret the data.

### **Discussion**

- State whether the hypothesis was supported or proven false by the results, or else state that the results were inconclusive.
- Cite specific results that support your conclusions.
- Give the reasoning for your conclusions.
- Demonstrate that you understand the biological meaning of your results.
- Compare the results, with your predictions and explain any unexpected results.
- Compare the results to other research or information available to you.
- Discuss any weaknesses in your experimental design or problems with the execution of the experiment.
- Discuss how you might extend or improve your experiment.

### **Conclusion**

- Restate your conclusion.
- Restate important results.

### **Literature Cited**

- Use the proper citation form in the text.
- Use proper citation form in the Literature Cited section.
- Refer in the text to any source listed in this section.

### **Acknowledgement**

- State any appropriate acknowledgement that you think is necessary.

## Practical 1

**Title:** Part I: Revision of chromatograms by Chromas

### Objectives:

After completing the practical, you will be able:

1. To learn how to revise sequences Chromatograms
2. To apply computer software such as Chromas to observe the Chromatograms

### Introduction:

Chromas is a free trace viewer for simple DNA sequencing projects which do not require assembly of multiple sequences. Chromas has the following features:

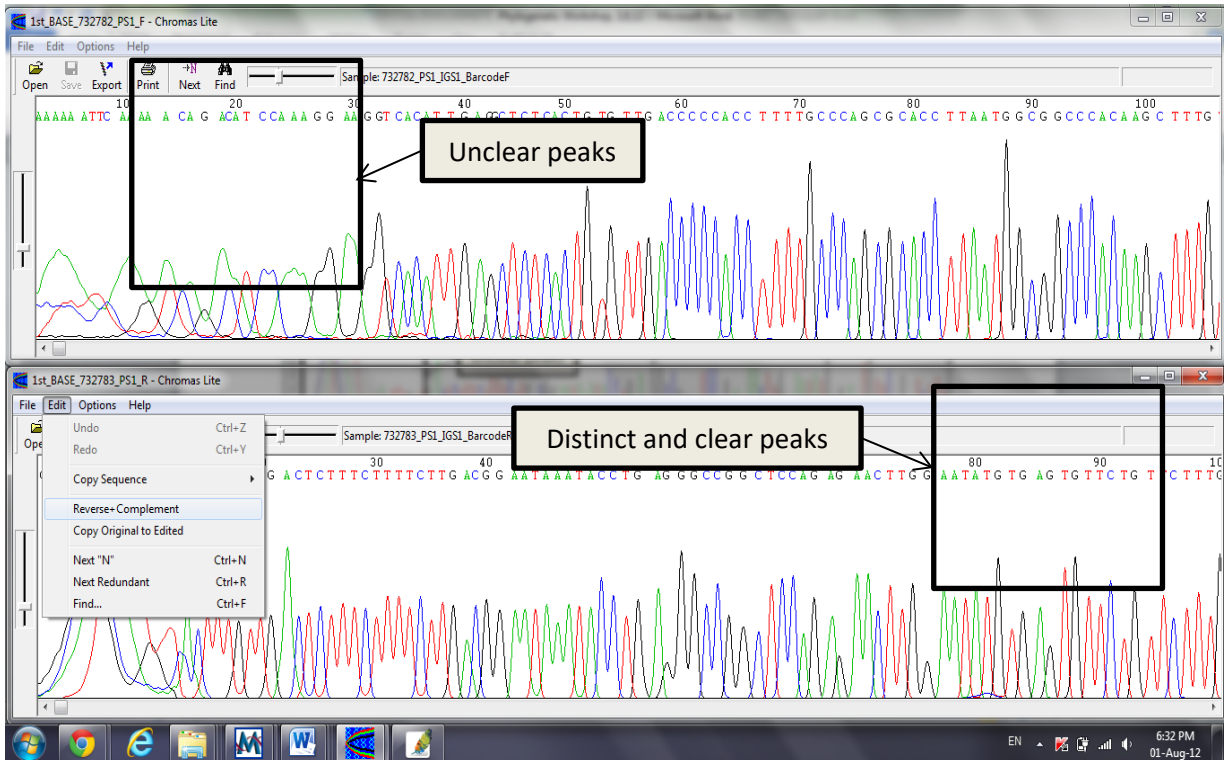
- Opens .ab1 chromatogram files from Applied Biosystems DNA sequencers.
- Opens SCF and ZTR format chromatogram files created by other sequencers or retrieved from databases.
- Save in .scf or Applied Biosystems .ab1 format.
- Print a chromatogram with options to zoom or fit to one page.
- Automatic removal of low-quality sequence (when quality data is available) or vector sequences.
- Export sequences in plain text, FASTA, FASTQ, EMBL, GenBank or GCG formats, or formatted with base numbering for presentation.
- Copy the sequence to the clipboard in plain text, FASTA or FASTQ format for pasting into other applications.
- Reverse & complement the sequence and chromatogram.
- Search for subsequences by exact matching or optimal alignment.
- Display translations in 3 frames along with the sequence.
- Copy an image of a chromatogram section for pasting into documents or presentations.
- Batch processing, including format conversion, sequence export with vector removal, batch printing and batch export of raw data.
- Enhance .ab1 files using the PeakTrace RP component from Nucleics Pty Ltd to improve peak readability and extract more high-quality bases.

### Procedure:

- 1- Open link <http://technelysium.com.au/wp/chromas/>
- 2- Install Chromas software
- 3- Open "sequence" folder
- 4- Choose the sequence 1st\_BASE\_732782\_PS1\_F.ab1 and 1st\_BASE\_732782\_PS1\_R.ab1 and open using Chromas program (sequence template provided).

By opening both forward and reverse sequences and arrange them in one page, we will be able to simply compare the picks:

- For reverse reaction chromatogram, do not forget to reverse complement the sequence



- Both reverse and forward sequences should appear similar. In dissimilar cases, the best peak (either reverse or forward) will be chosen as consensus nucleotide sequence.
- If both forward and reverse peaks appear distinct and clear but different, the code called ambiguous code and indicates with:

**AMBIGUITY CODES:**

M=A/C; R=A/G; W=A/T; S=C/G; Y=C/T; K=G/T; V=A/C/G; H=A/C/T;  
 D=A/G/T; B=C/G/T

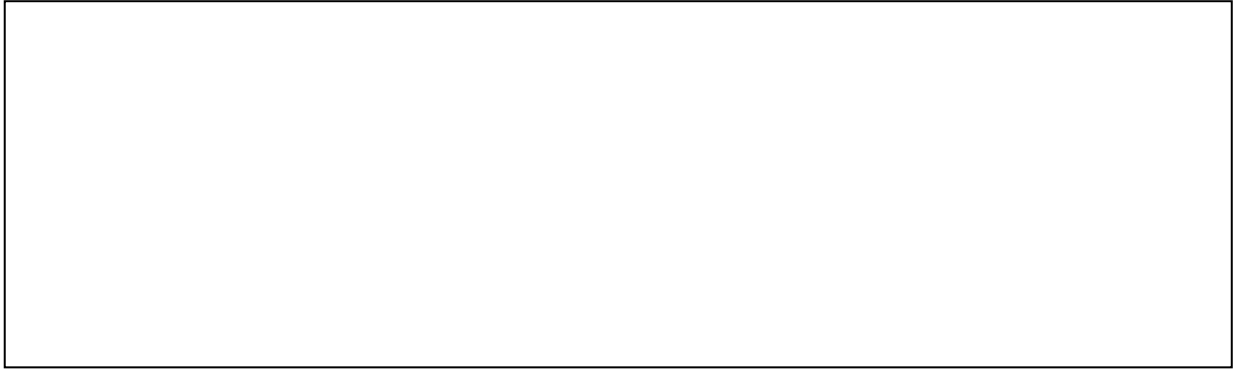
**Questions and Results:**

A) Take a snapshot of your analyzed forward and reversed fragment and paste in the below box:

B) What are the criteria you used for confirming sequence nucleotide?

C) Explain what do you will do if a different type of nucleotide is observed in both forward and reverse sequences in which both are reliable?

D) Take an snapshot of your end points of your forward and reverse sequences and paste in the below box:



- Based on the above picture, describe how to identify distinct and unclear picks and how to make final consensus sequence?

## Title: Part II: Sequence alignment and amendment by Mega software

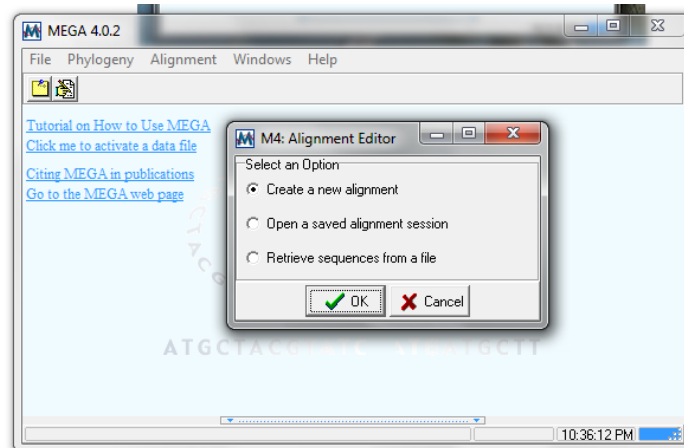
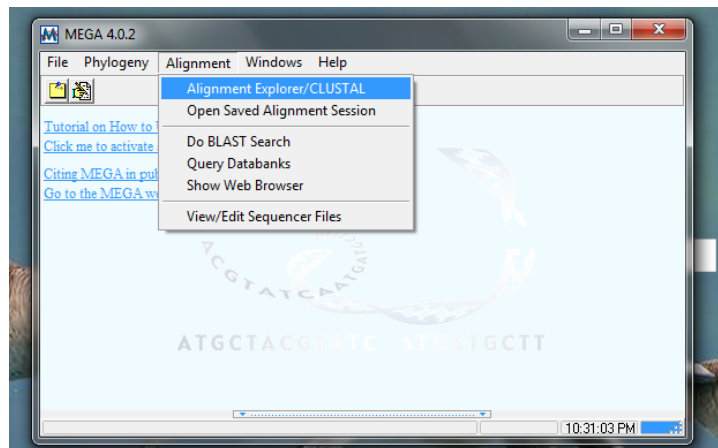
### Objectives:

After completing the practical, you will be able:

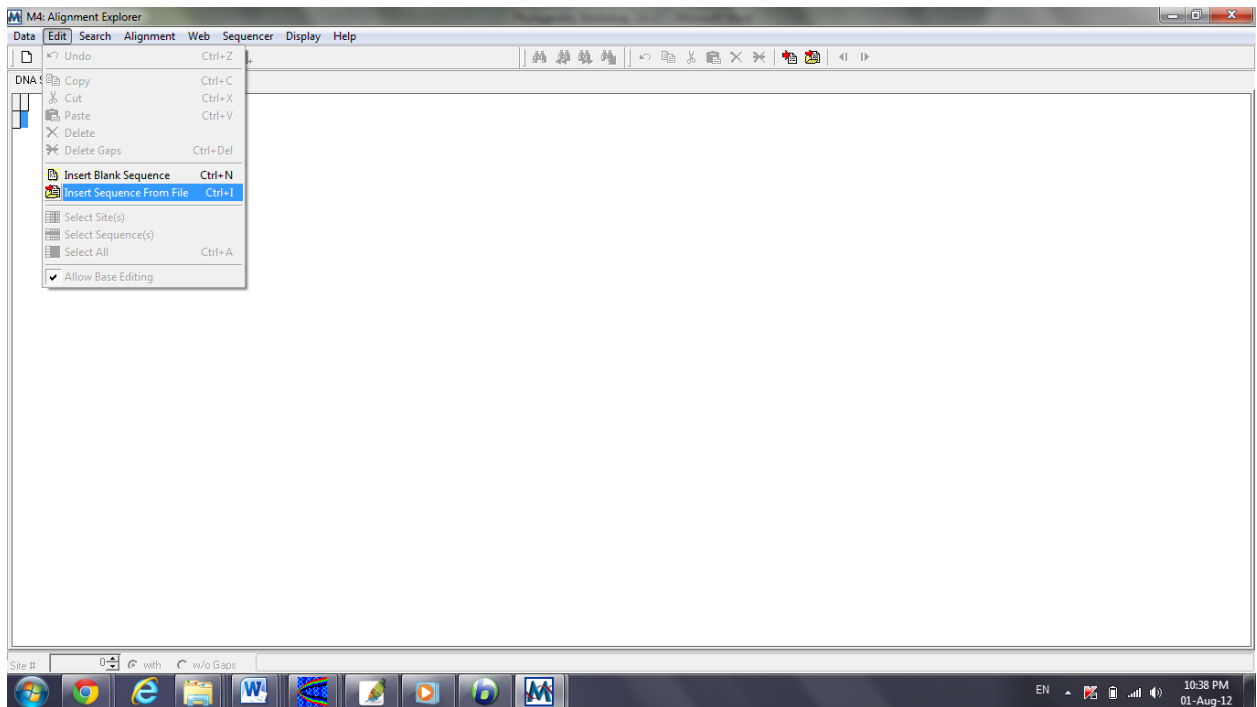
1. To understand the principles of sequence alignment
2. To apply Bioinformatics software such as MEGA to align nucleotide sequences
3. To understand the meanings of terms conserved, variable, parsimony informative and singleton nucleotides and how to calculate.

### Procedure:

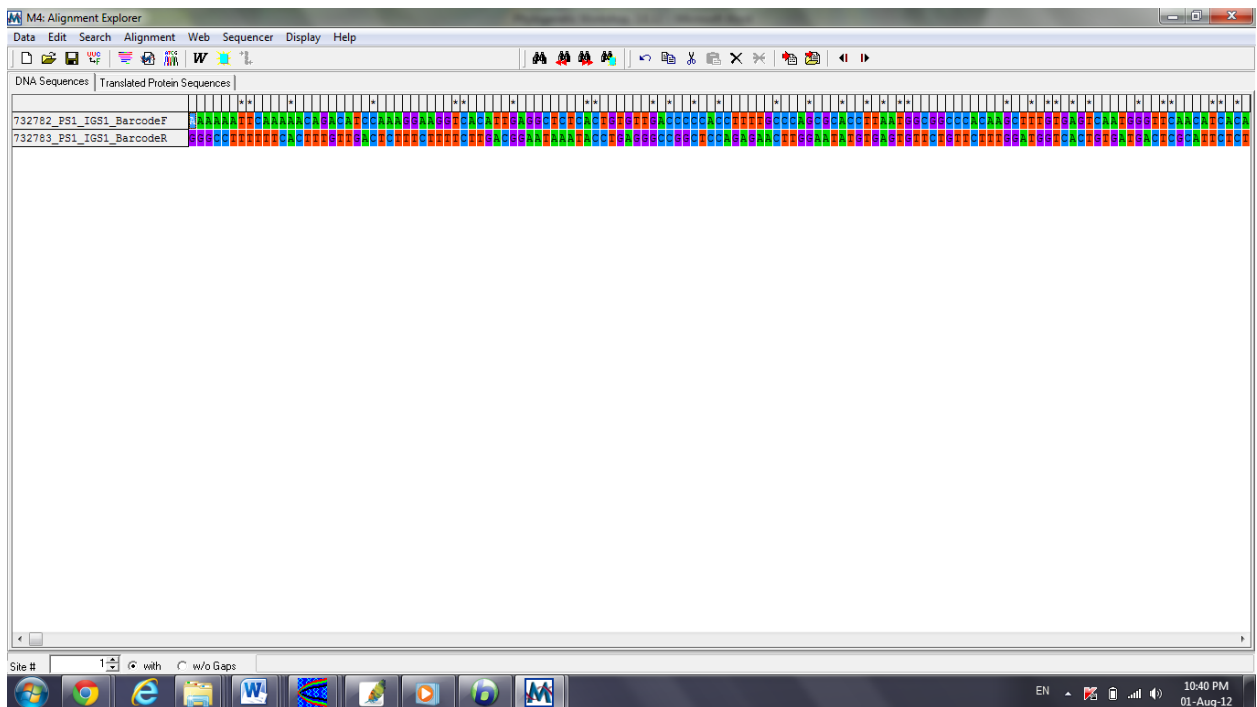
- 1- Open link [http://www.megasoftware.net/download\\_form](http://www.megasoftware.net/download_form)
- 2- Install the latest version of Mega software.
- 3- Open the MEGA program.
- 4- Click on "Alignment".
- 5- Choose "Alignment Explorer/CLUSTAL".
- 6- Choose "create new alignment".
- 7- Build a DNA sequence alignment by selecting YES.



- 8- Click on "Edit" menu and insert sequence from file.

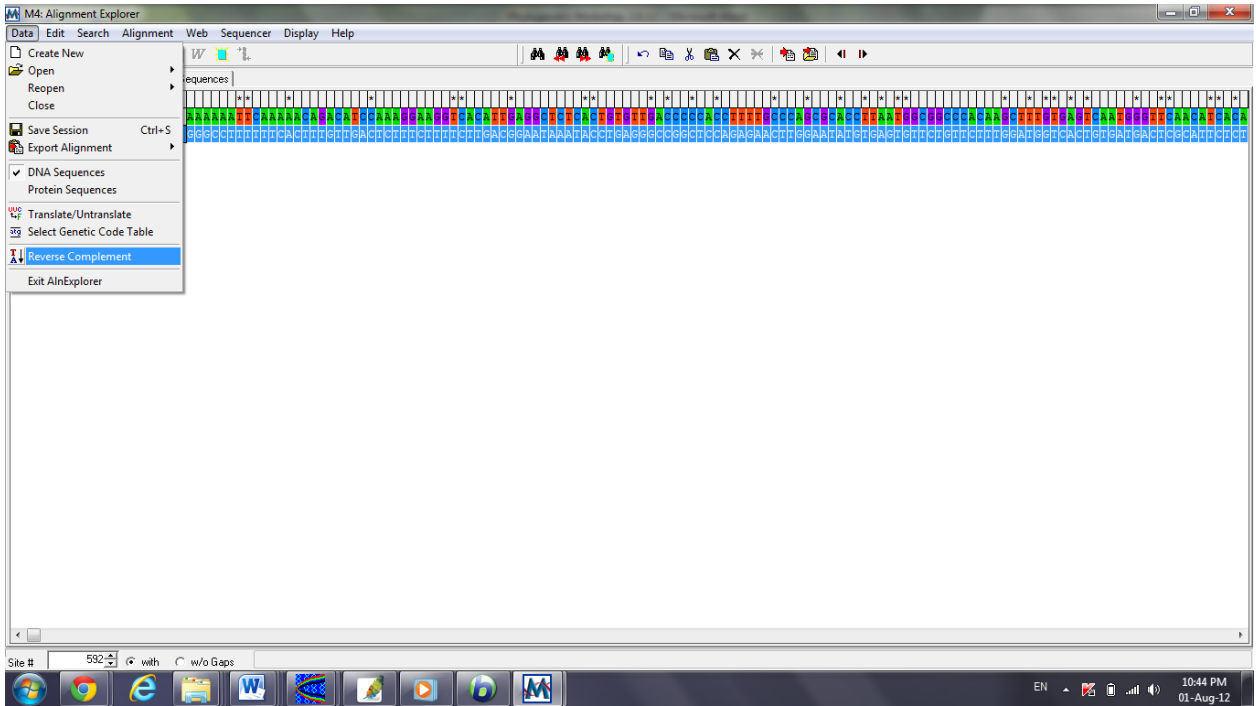


9- Open your sequencing samples from folder "sequence"

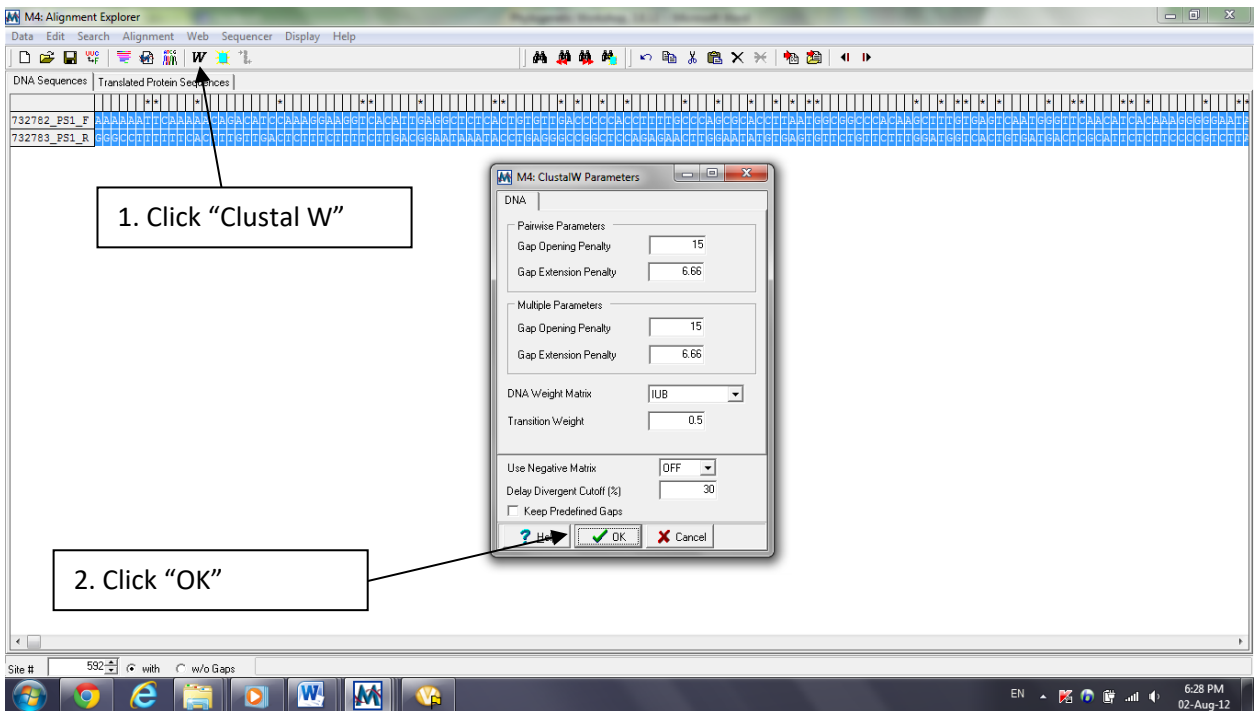


10- Ensure that the Genetic code selected is correct

11- Remember to reverse complement reverse reaction sequences.

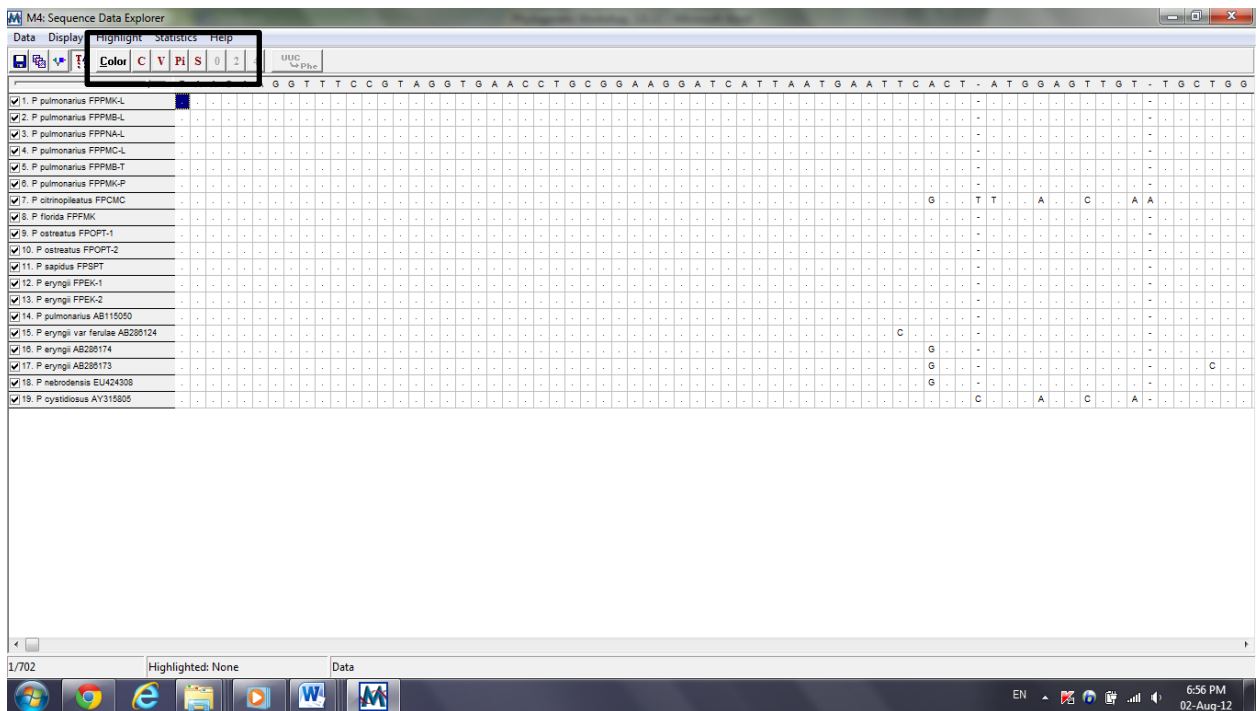


12- Select all and Align by Clustal-w:



- What you do not want to see in your alignment:
  - N characters – remove the sequences/region that has grey N, especially the front part of forward reaction and the end of reverse reaction. Note that the occurrence of N also indicates that the region has quite poor quality sequences.
  - Gap – it is not common to have a gap in mitochondrial DNA

- Note: you can include one full sequence obtained from GENBANK that you have gotten from your BLAST result. This can help you remove the sequence that is not part of your gene of interest. If you are interested in the transition and transversion, you would want to check that there is no stop codon (characterized by \*) in your translated sequence. To correct this, make sure that the first nucleotide is the first frame. You can obtain this information from your BLAST result. The full sequence obtained from Genbank also can guide you,
- Since the correct translation obtained, you can export the DNA sequence as a FASTA or MEGA format. You can even save the current sequence alignment.
- ✓ Open the Mega format file saved earlier from folder sequence (Test.meg).
- ✓ Calculate number of conserved, variable, parsimony info and singleton nucleotides and etc... in your sequence



Close Data Explorer window...



**Exercise:**

- Calculate the number of conserved, variable, parsimony informative and singleton nucleotides and fill in the below table:

	Conserved	Variable	Parsimony informative	Singleton	Total no. sites
Sequences set 1					
Sequences set 2					

- Create and build a sequence alignment for the sequences provided for this practical in bioinformatics folder.

- Align the sequences by cluster W.
- Take a snapshot from the sequences and paste in the box below:

- Based on the obtained picture, how do you differ matched (similar) nucleotide sequences from unmatched/variable sequences?
- Describe how to export your DNA sequences is which they can be used for future studies.
- Is GenBank material required to be included in the alignment? If so, what is the significance? Describe.

## Practical 2

Title: BLAST search by NCBI database

### Objectives:

After completing the practical, you will be able:

1. To understand the basics of BLAST
2. To identify the exact taxonomic position of the provided sequence
3. To use NCBI BLAST search to compare sequences with deposited sequences in database

### Introduction

BLAST (Basic Local Alignment Search Tool) was developed in 1989 at the National Center for Biotechnology Information (NCBI) at the National Institutes of Health (NIH). BLAST uses statistical methods to compare a DNA or protein input sequence (a.k.a. "query sequence") to a database of sequences ("subject sequences") and return those sequences that have a significant level of similarity to the query sequence. Although BLAST is a complex set of programs with many functions and applications and a sophisticated statistical underpinning, its most popular use – using a DNA or protein sequence to query NCBI's GenBank database – is fairly straightforward, thanks to NCBI's BLAST web portal.

Imagine that you have obtained a DNA sequence as part of your research. For example, you may have obtained the DNA as a means of identifying a specimen, or in the process of attempting to clone a gene of interest (e.g., in order to learn more about its function, to produce a protein of interest in large quantities, or to produce a transgenic organism). In order to identify the specimen or confirm that you have obtained the correct DNA, you wish to compare your DNA sequence to a database of known sequences. GenBank is the NIH database of publicly-available nucleotide sequence data (daily exchange of data with the European Molecular Biology Laboratory Nucleotide Sequence Database in Europe and the DNA Data Bank of Japan ensures global coverage and consensus across the major databases), and currently includes over 108 million sequences from over 260,000 named organisms – a search of GenBank is therefore a great place to look for the identity of your unidentified DNA sequence.

The BLAST algorithm calculates similarity scores for *local alignments* (i.e., the most similar regions between 2 sequences) between the query sequence and subject sequences using specific scoring matrices, and returns a table of the best matches ("hits") from the database. The hit table includes several useful pieces of information, including the similarity score, query coverage (percent of the query sequence that overlaps the subject sequence), E-value (see below), and max identity (percent similarity between the query and subject sequences over the length of the coverage area).

The **Expect value (e-value)** of the hit is defined by NCBI as follows:

"The Expect value (E) is a parameter that describes the number of hits one can 'expect' to see by chance when searching a database of a particular size. It decreases exponentially as the Score (S) of the match increases. Essentially, the E value describes the random background noise. For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance. The lower the E-value, or the closer it is to zero, the more 'significant' the match is."

## Exercise

Imagine that you are a researcher involved in the assessment of the spread of *Pleurotus* species from Malaysia. Customs officials have stopped a shipment of a species that show signs of infection, though the fungus is not a previously-known host. Customs has asked you to determine whether the fungi are infected with bacteria or other disease. From the infected material you have obtained a culture that appears to be *Pleurotus*, but the culture is not sporulating so you cannot identify the organism. You therefore decided to extract DNA from the culture, use PCR to amplify the DNA encoding the nuclear ribosomal internal transcribed spacer region (ITS; a region of DNA commonly used for 'DNA barcoding' in fungi and fungal-like organisms), and obtain the sequence of the ITS DNA. The content of the DNA sequence is stored in the file "sequence.txt"

### Procedure:

1. Examine the file "sequence.txt," in the desktop folder containing the course files. The file is in FASTA format. The first line (starting with a greater-than symbol) includes a description of the sequence, and the following lines include the sequence itself.
2. Navigate to the NCBI BLAST website: <http://blast.ncbi.nlm.nih.gov>.
3. Under "Basic BLAST," choose the "nucleotide blast" option, which will take you to the nucleotide BLAST data entry form.
4. Under "Enter Query Sequence," use the "Browse" button (below the large white data entry window) to upload the file "sequence.txt" from the desktop folder containing the course files.
5. Under "Choose Search Set," select the radio button corresponding to "Others (nr, etc.)."
6. Select the blue "BLAST" button in the lower left-hand corner.
7. With the assistance of your instructor, examine the 4 parts of the BLAST output – the header, Graphic Summary, Descriptions, and Alignments. What is the closest match to your "unknown" sequence, and how good is the match? Is it likely that true sequence homology exists (i.e., that the hit didn't occur simply by chance)?

### Results and Evaluation

- Take a snapshot of the blast search and paste in the below box:

- What is the closest match the unknown sequence?
- Enumerate the criteria you used for selecting the best matching sequence and/or for identifying your unknown sequence.
- Was it likely that true sequence homology existed (i.e., that the hit didn't occur simply by chance)? Describe.

### Practical 3

**Title:** Design PCR primers and check them for specificity

#### Objectives:

After completing the practical, you will be able to:

1. To understand the basics of primer designing
2. To apply Bioinformatics software to develop primer pairs

#### Introduction:

Primer-BLAST to alleviate the difficulty in designing target-specific primers. This tool combines BLAST with a global alignment algorithm to ensure a full primer-target alignment and is sensitive enough to detect targets that have a significant number of mismatches to primers. Primer-BLAST allows users to design new target-specific primers in one step as well as to check the specificity of pre-existing primers. Primer-BLAST also supports placing primers based on exon/intron locations and excluding single nucleotide polymorphism (SNP) sites in primers.

#### Procedure:

- Open NCBI primer BLAST page
  - <https://www.ncbi.nlm.nih.gov/tools/primer-blast/>
- Fill up the primer design page:
- Go to the Primer BLAST submission form.
- Enter one or both primer sequences in the Primer Parameters section of the form. If only one primer is available, a template sequence is also required. See "A Target Template Sequence...".
- In the Primer Pair Specificity Checking Parameters section, select the appropriate source Organism and the smallest Database that is likely to contain the target sequence. These settings give the most precise results. For broadest coverage, choose the nr database and do not specify an organism.
- Click the "Get Primers" button to submit the search and retrieve template and specificity information.

#### A TARGET TEMPLATE SEQUENCE OR ACCESSION NUMBER

- Go to the Primer BLAST submission form.
- Enter the target sequence in FASTA format or an accession number of an NCBI nucleotide sequence in the PCR Template section of the form. If the NCBI mRNA reference sequence accession number is used, the tool will automatically design primers that are specific to that splice variant.
- If one or both primer sequences are to be used in the search, enter these in the Primer Parameters section of the form. Primer BLAST performs only a specificity check when a target template and both primers are provided.

- In the Primer Pair Specificity Checking Parameters section, select the appropriate source Organism and the smallest Database that is likely to contain the target sequence. These settings give the most precise results. For broadest coverage, choose the nr database and do not specify an organism.
- Click the "Get Primers" button to submit the search and retrieve specific primer pairs.

**Primer-BLAST**  
A tool for finding specific primers

NCBI Primer-BLAST: Finding primers specific to your PCR template (using Primer3 and BLAST).

**PCR Template**  
Enter accession, gi, or FASTA sequence (A refseq record is preferred)  Clear  
Range: Forward primer From  To  Clear  
Reverse primer   Clear  
Or, upload FASTA file  No file chosen

**Primer Parameters**  
Use my own forward primer (5'→3' on plus strand)  Clear  
Use my own reverse primer (5'→3' on minus strand)  Clear  
PCR product size: Min  Max   
# of primers to return:   
Primer melting temperatures (T<sub>m</sub>): Min  Opt  Max  Max T<sub>m</sub> difference

**Exon/intron selection**  
A refseq mRNA sequence as PCR template input is required for options in the section  
Exon junction span:   
Exon junction match: Exon at 5' side  Exon at 3' side   
Minimal number of bases that must anneal to exons at the 5' or 3' side of the junction  
Intron inclusion:  Primer pair must be separated by at least one intron on the corresponding genomic DNA  
Intron length range: Min  Max

**Primer Pair Specificity Checking Parameters**  
Specificity check:  Enable search for primer pairs specific to the intended PCR template  
Search mode:   
Database:   
Exclusion:  Exclude predicted Refseq transcripts (accession with XM, XR prefix)  Exclude uncultured/environmental sample sequences  
Organism:   
Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy id or select from the suggestion list as you type. [Add more organisms](#)  
Entrez query (optional):   
Primer specificity stringency: Primer must have at least  total mismatches to unintended targets, including at least  mismatches within the last  bps at the 3' end. Ignore targets that have  or more mismatches to the primer.  
Max target size:   
Splice variant handling:  Allow primer to amplify mRNA splice variants (requires refseq mRNA sequence as PCR template input)

Show results in a new window  Use new graphic view

[Advanced parameters](#)

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback on new interface

NCBI | NLM | NIH

### Exercise:

- Use the nucleotide sequence provided (sequence.txt) to design appropriate primers

- Take an snapshot of the generated primers and paste in the box below:

- What is the optimum melting temperature?
- Where should be the location of your designed primers?
- What is the PCR product size?
- What should be the primer specificity stringency?
- What should be the  $T_m$  difference between forward and reverse primers?
- Based on the criteria, paste two best primer pairs in the below table:

	Forward	Reverse
Primer set 1		
Primer set 2		

## Practical 4

Title: Phylogenetic analysis - Phylogenetic tree & bootstrapping

### Objective:

After completing the practical, you will be able:

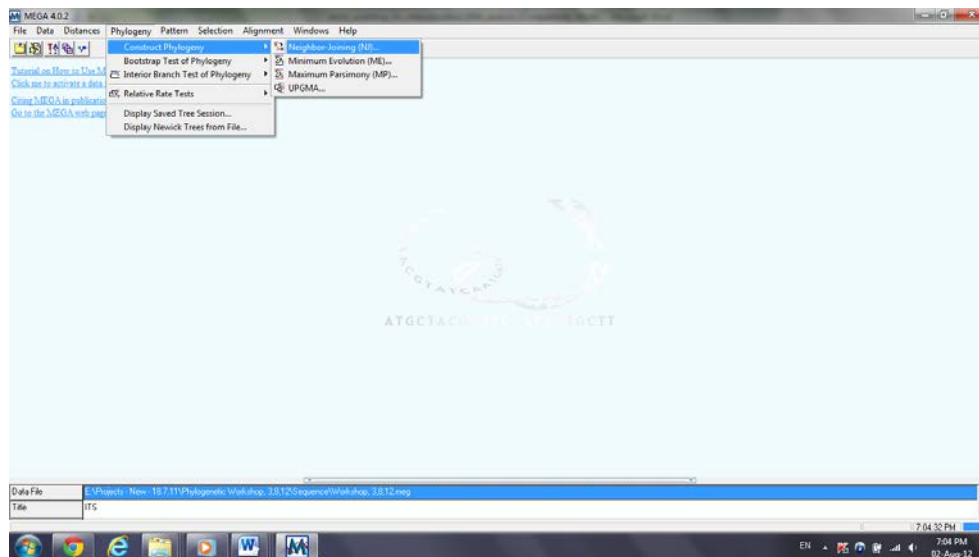
1. To construct phylogenetic tree using MEGA software

### Introduction

A phylogenetic tree is a mathematical structure to show evolutionary relationship between sequences or organism. Actual pattern of the evolutionary relationship that was estimated is called a phylogeny. DNA sequences provide distinct codes that can be used to construct phylogenetic tree.

### Procedure:

- Open MEGA software
- Go to Phylogeny, Construct Phylogeny, then choose one method.

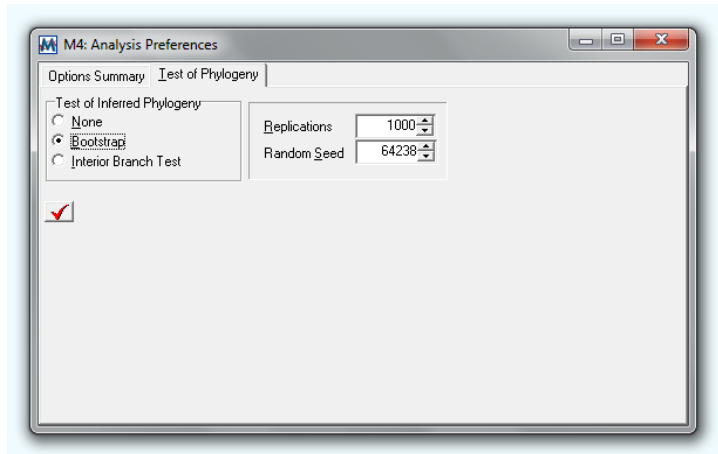


Construct a method eg. neighbor joining and maximum parsimony tree

Neighbor joining: based on distance method (distance = fraction of sites that differ between two sequences in a multiple alignment). It uses a specific calculation method by manipulating the distance matrix.

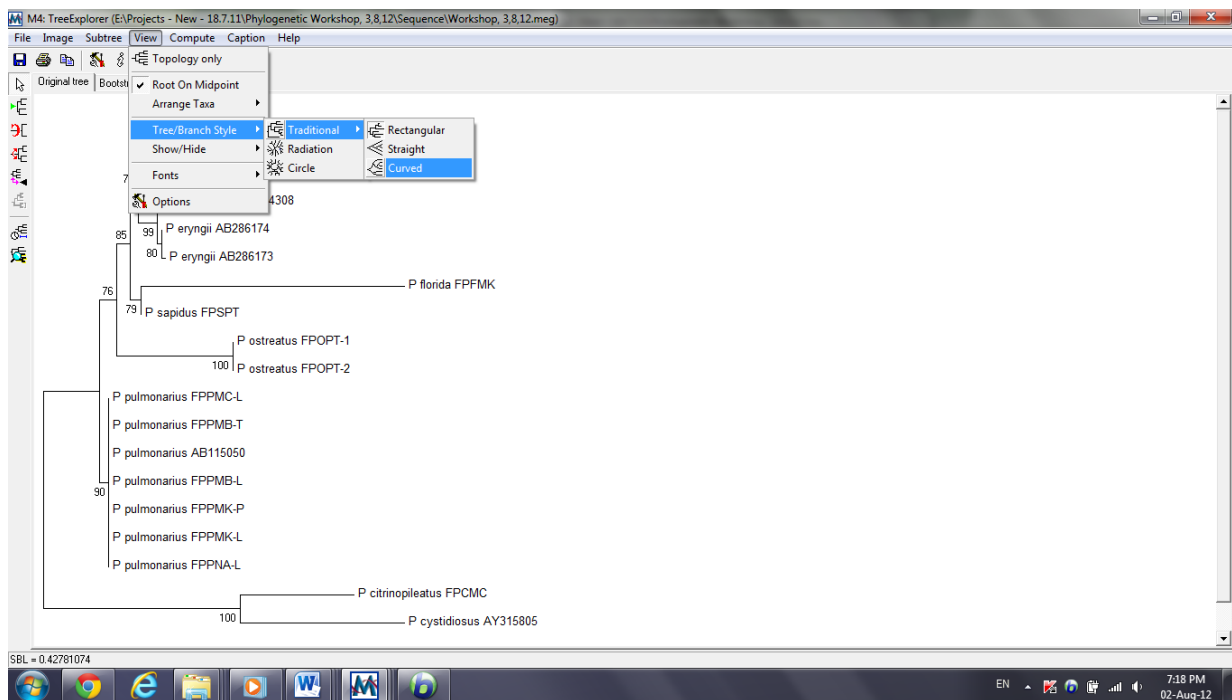
Maximum parsimony: based on the character method. It looks for tree that exhibits the minimum number of changes. There can be more than one tree that has the same number of changes.

- ✓ Make sure to include test of inferred phylogeny (Bootstrap 1000)



➤ Famous example to explain bootstrapping: You have unknown number of green, blue and red marbles in a sack. You take a few times 5 marbles at a time. By repeating this task a few times, (for example for us it is 1000 times) you can infer the number of green, red and blue marbles in the sack

✓ After generating the tree, you can choose the layout of the tree and add caption





## Exercise and Results

- Construct a phylogenetic tree based on UPGMA method using the FASTA file you generated in the first practical session (Bootstrap 1000 replications).
- Paste the tree picture in the box below:



- Give an appropriate title.
- Explain the given phylogenetic tree.
- Draw a flow chart on how to construct a phylogenetic tree.

## Practical 5

Title: Analysis of nucleotide polymorphism and molecular data (AMOVA) from aligned DNA sequenced data by DNAsp and Arlequin software

### Objectives:

After completing the practical, you will be able:

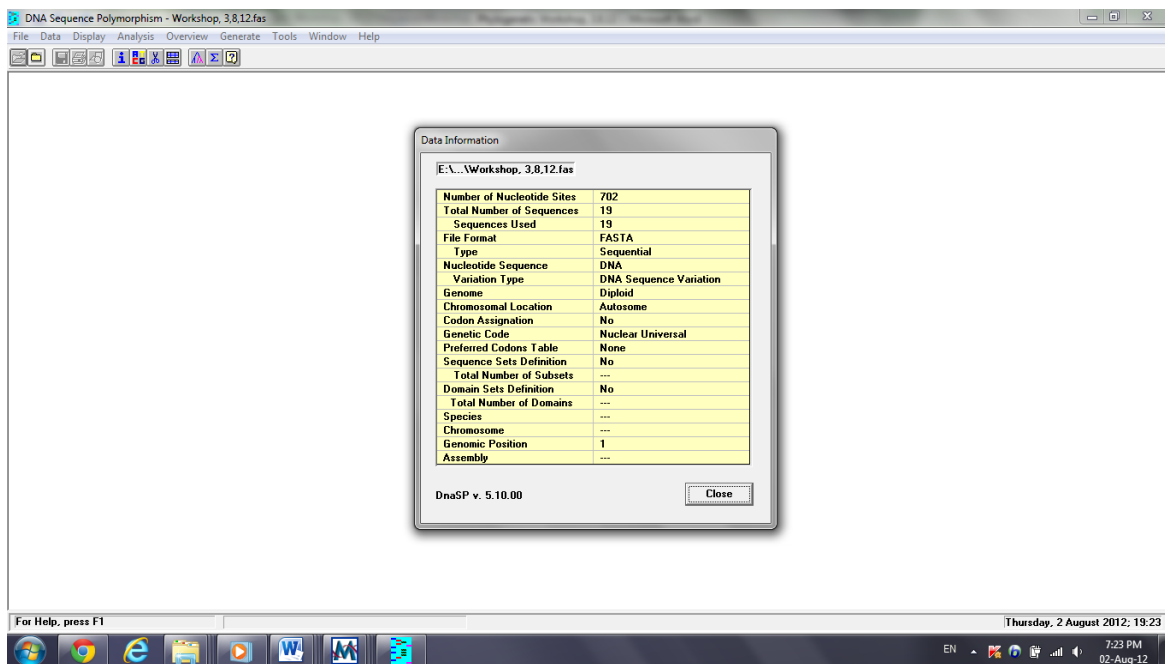
1. To apply Bioinformatics software such as DNAsp and Arlequin to analyze the nucleotide polymorphism and molecular data
2. To learn how to generate haplotype data file and ARP Arlequin input file
3. To learn how to analyze molecular variant (AMOVA) using Arlequin software

### Procedure:

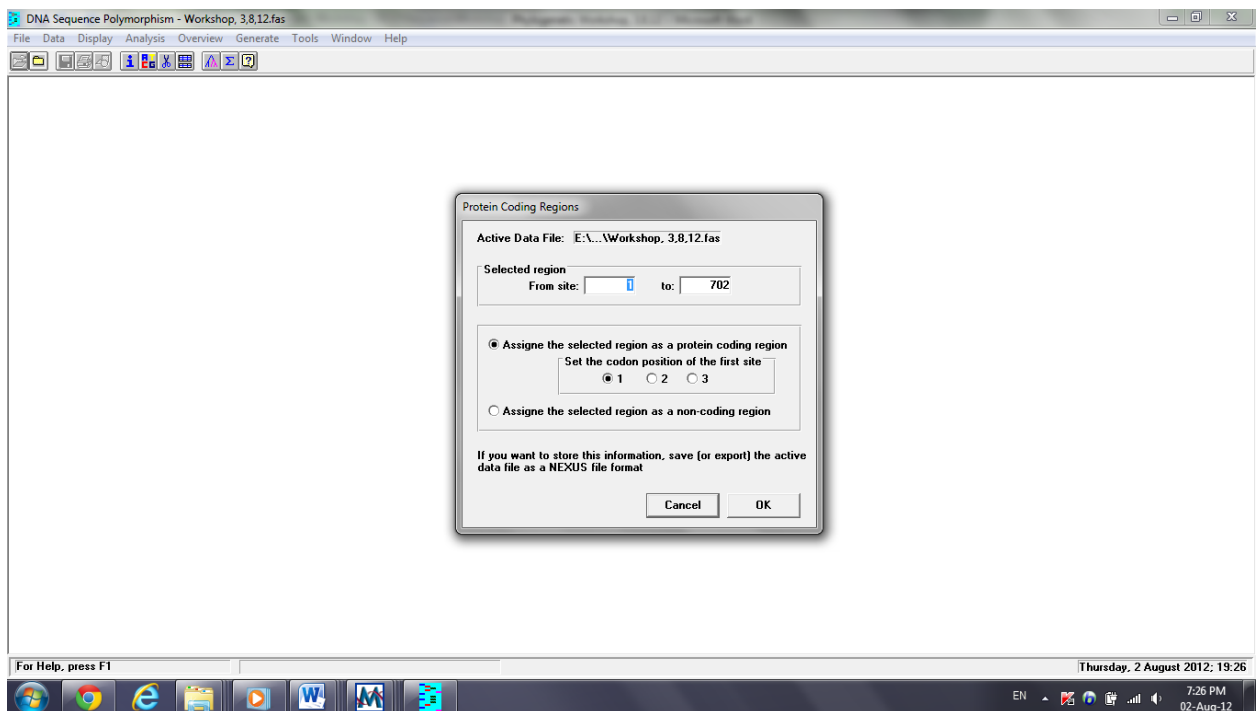
- ✓ Open DNAsp program



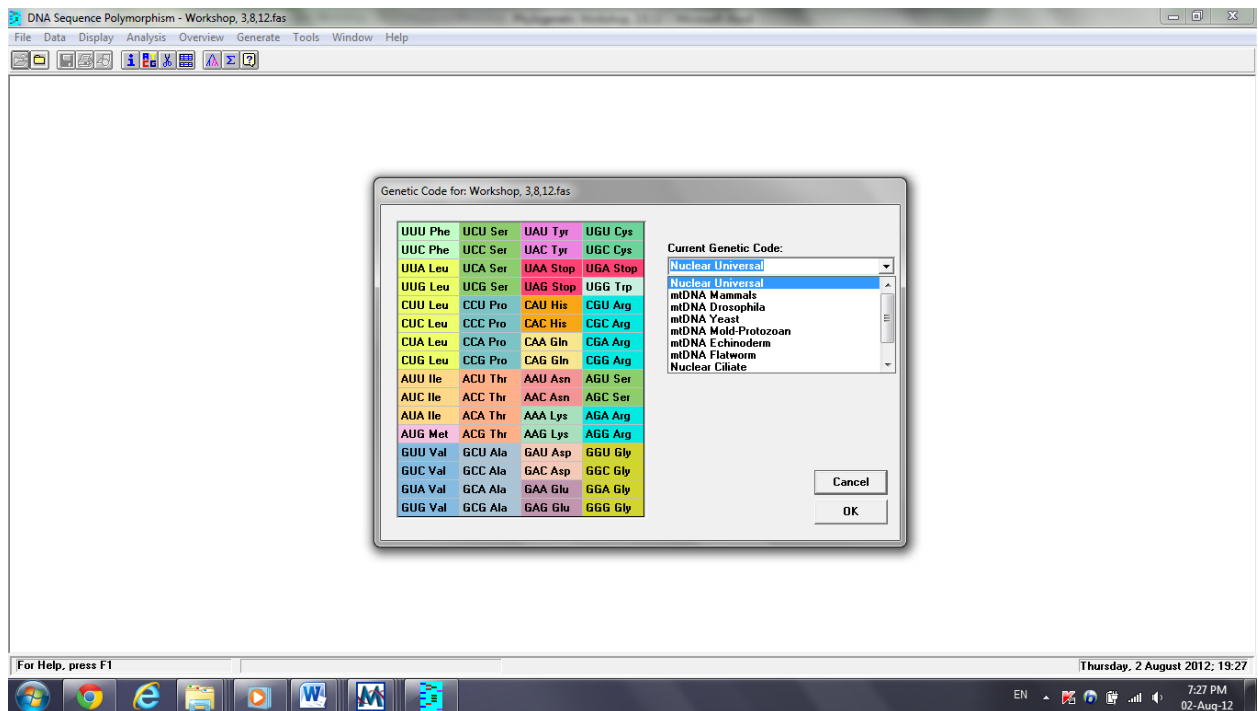
✓ Open the created fasta file:



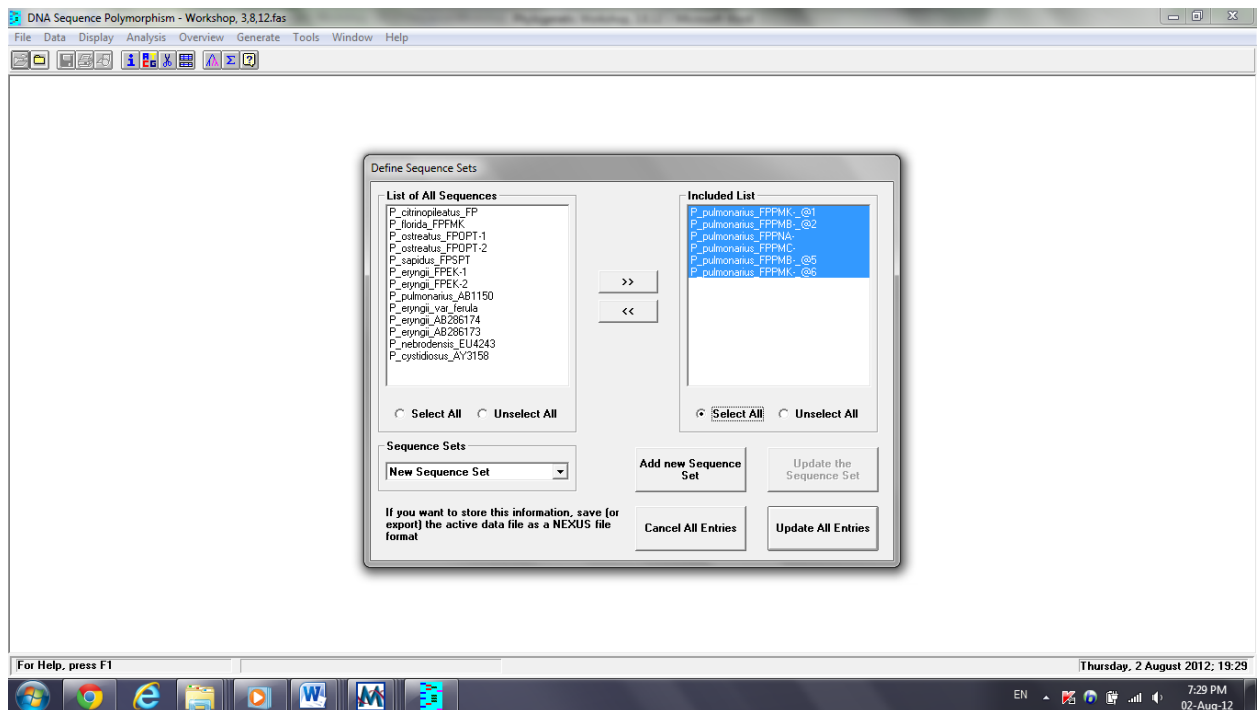
✓ Assign coding region to cover all the sequence



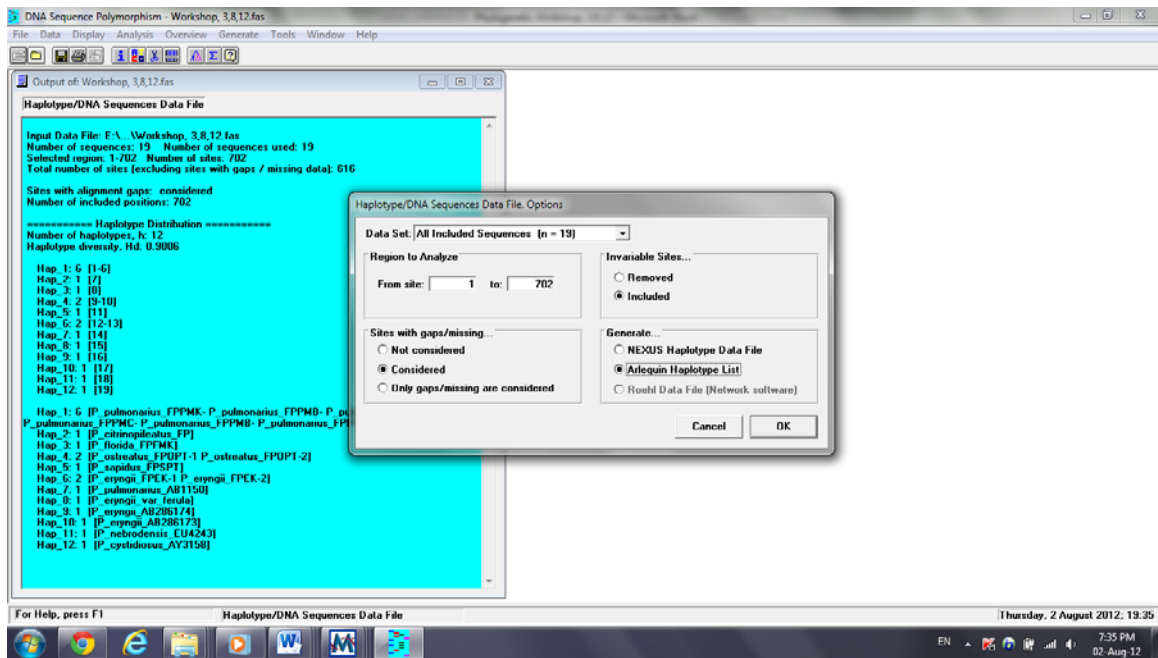
- ✓ Assign the genetic code to "???"



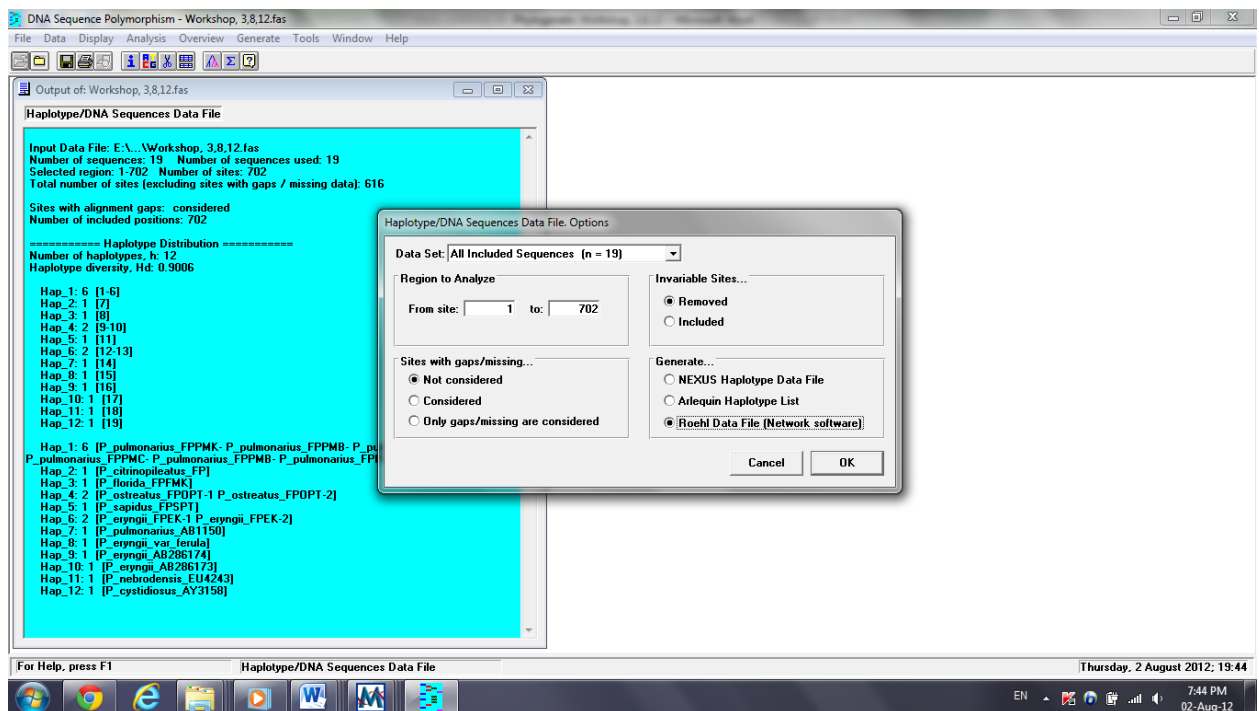
- ✓ Define sequence sets, select each population, press ">>" and choose "Add new sequence set", give it a name. Repeat this task for all the populations and choose "Update all entries"



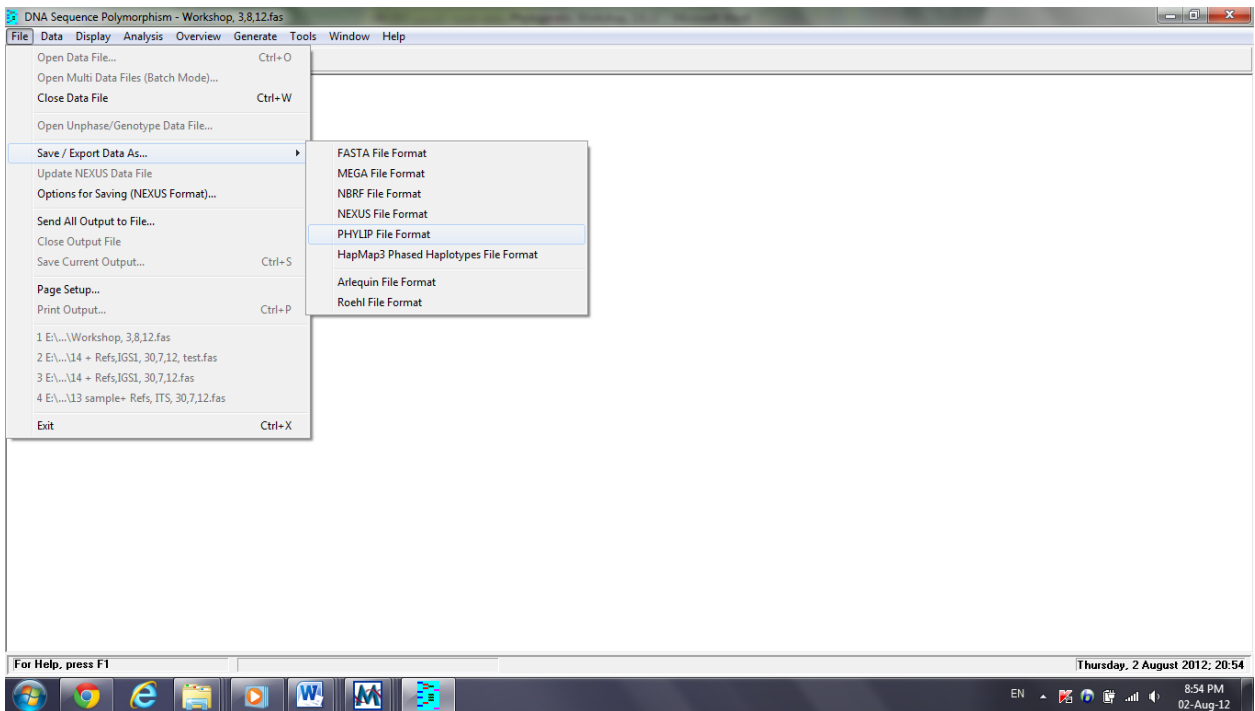
- ✓ Generate haplotype data (Generate ->Haplotype Data file), include invariable sites, considered gaps/missing data, generate "Arlequin Haplotype list". When it prompts to save the file, select cancel (this is because Arlequin cannot really use this file, you have to generate the input file manually, which will be explained later).



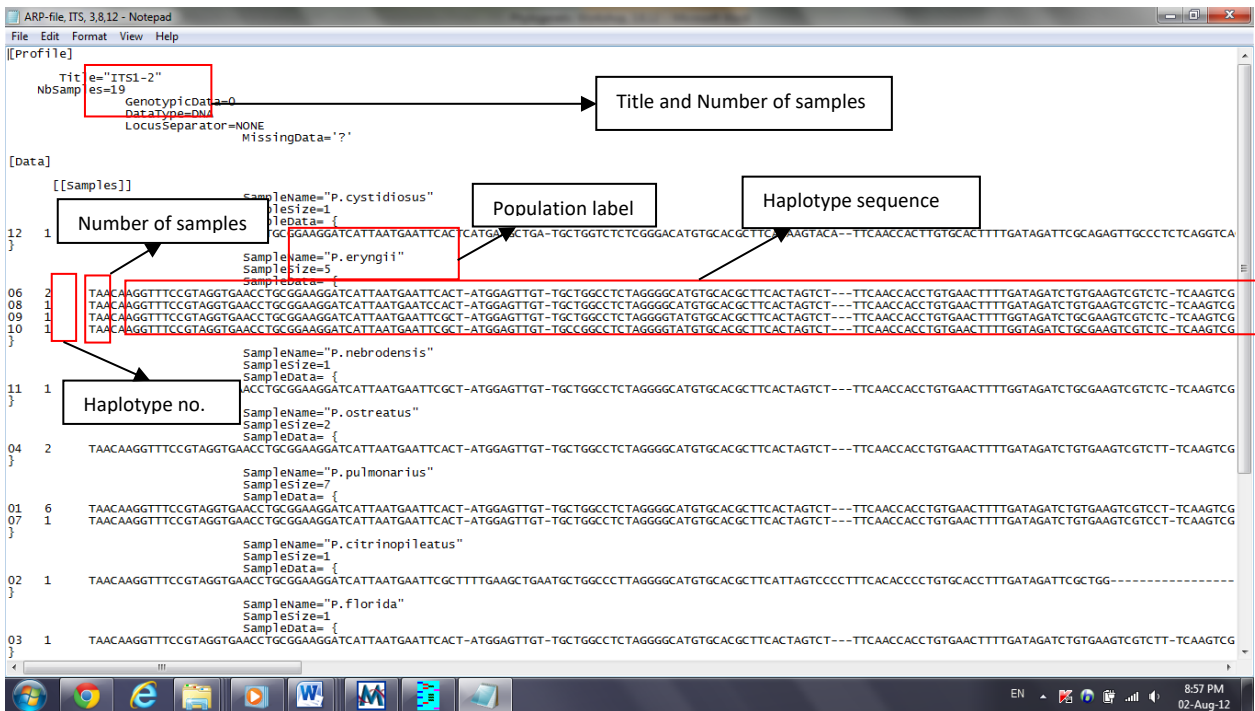
- ✓ Generate Roehl data file (Generate ->Haplotype Data file), exclude invariable sites, and not considered gaps/missing data. Save the exported file which can be used for Network Software.



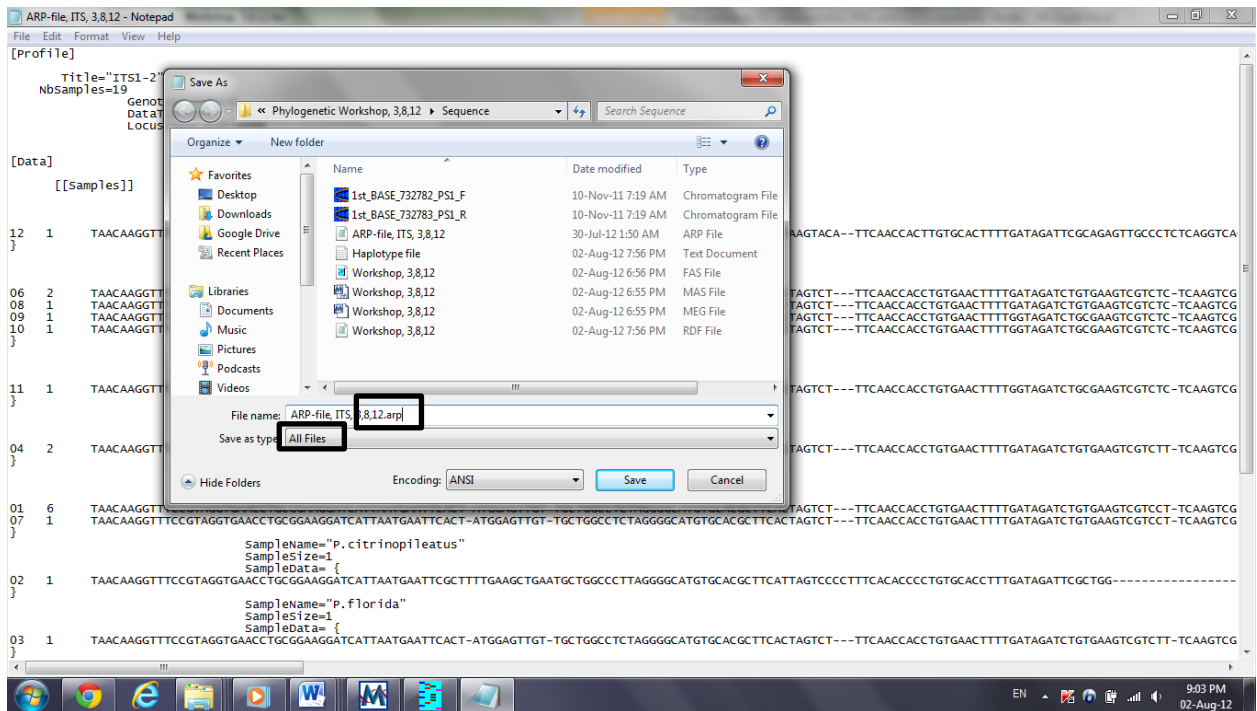
- ✓ Generate a phylip file using DNAsp (save/export as -> phylip format) for TCS genealogy data analysis later



- ✓ Generating manually Arlequin input file, name it as \*.arp. Ensure that you do not include the output sequence

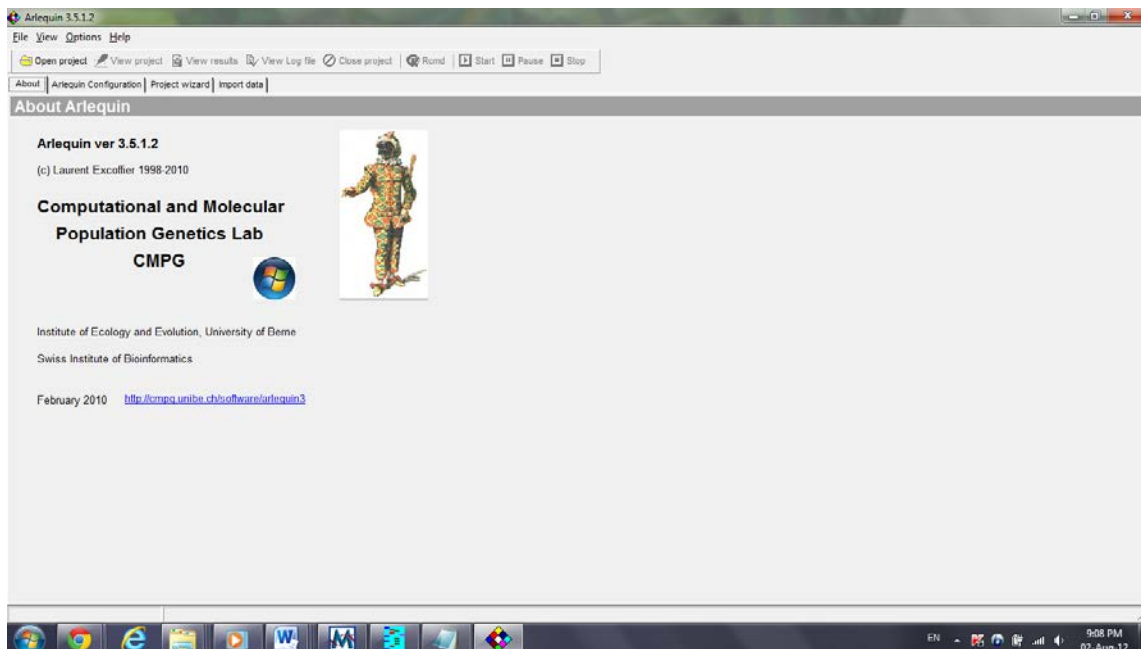


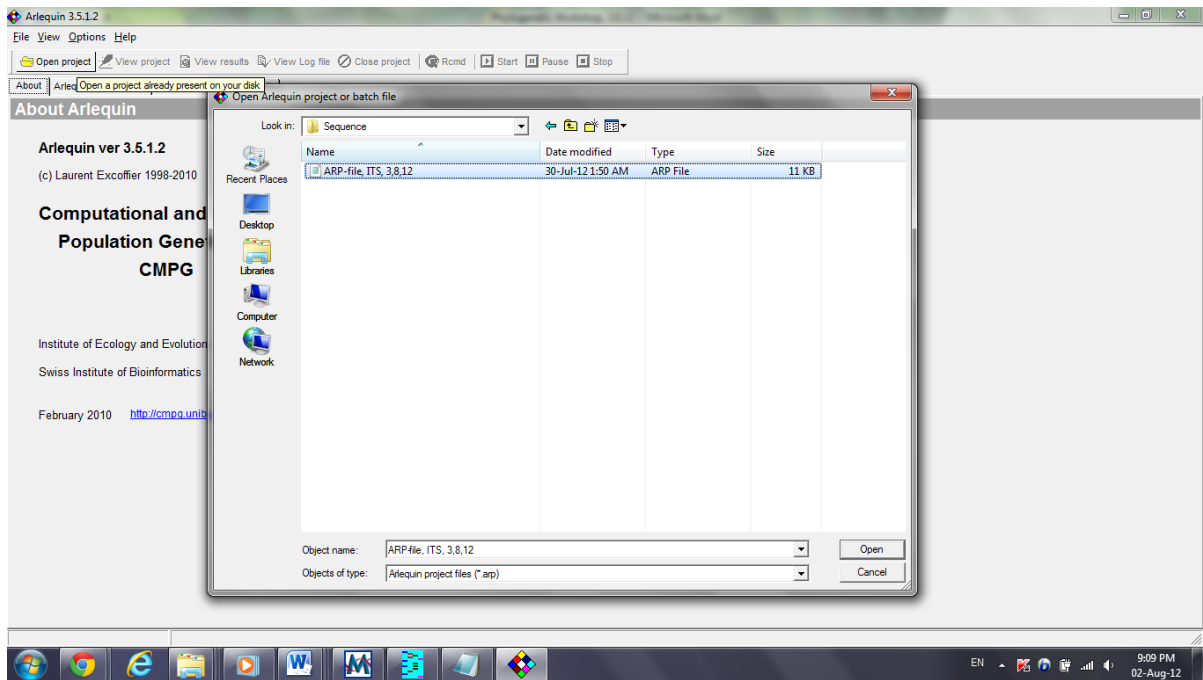
- ✓ Save as \*\*.arp:



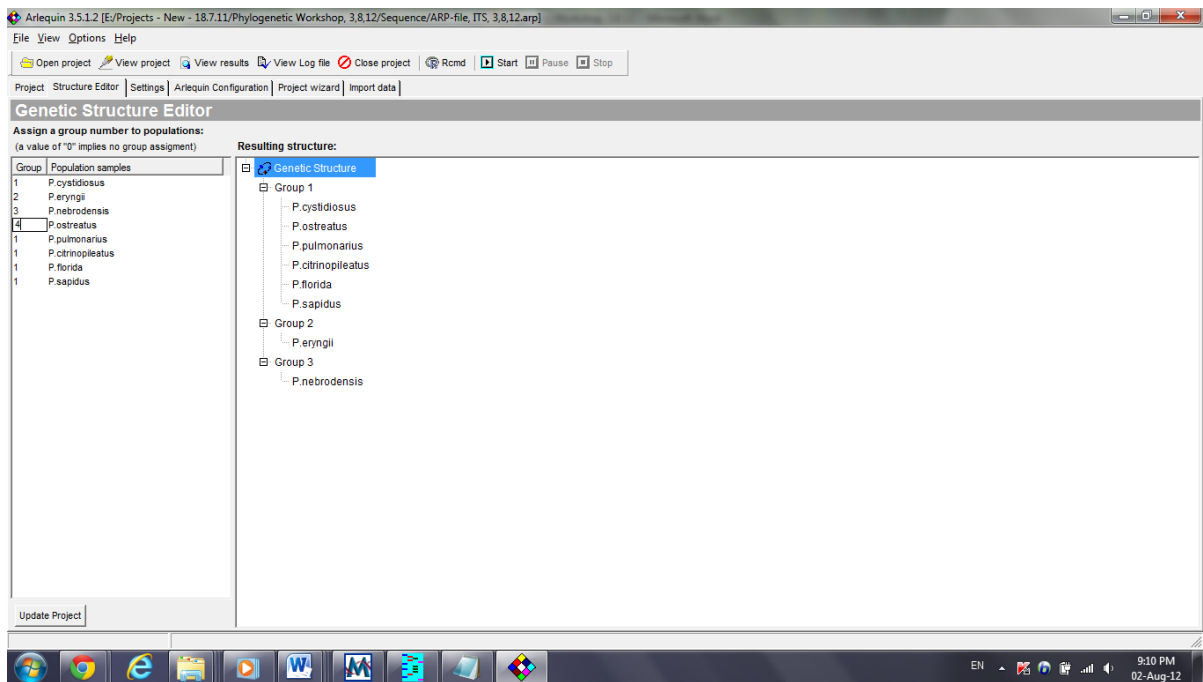
## Arlequin 3.5 program

- ✓ Calculate (AMOVA) using Arlequin
- ✓ Load the \*.arp file to Arlequin (File->open project)



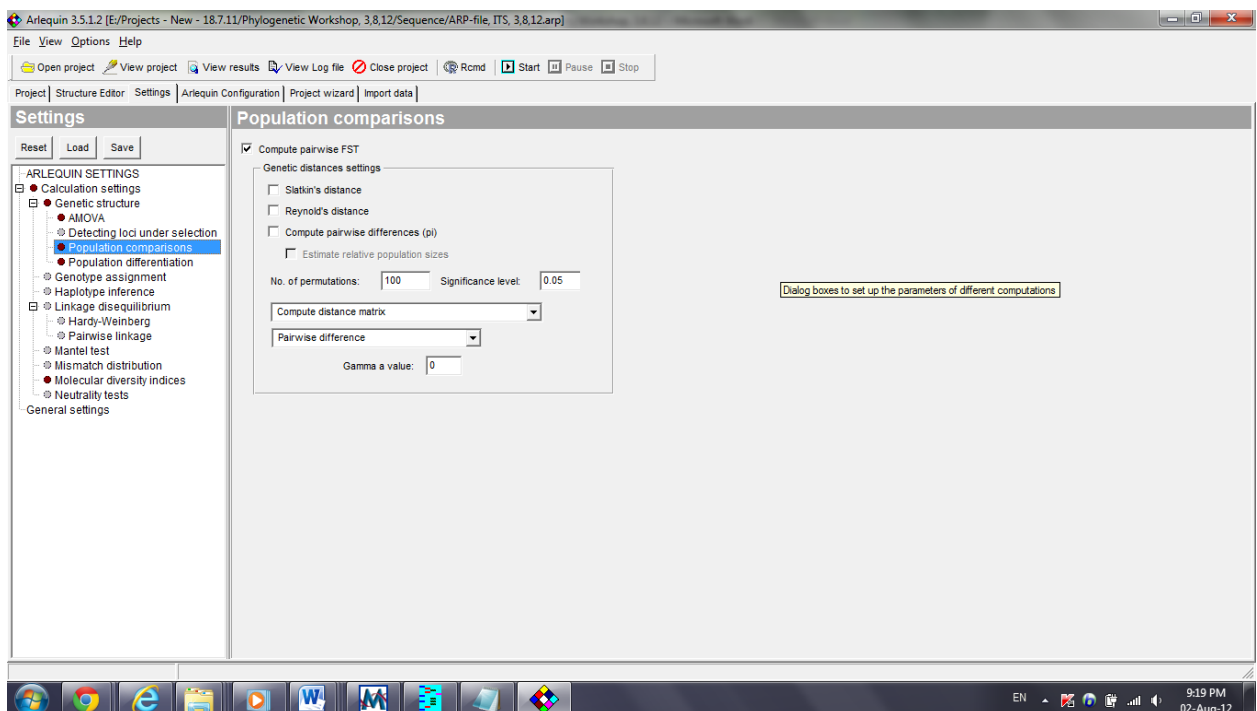
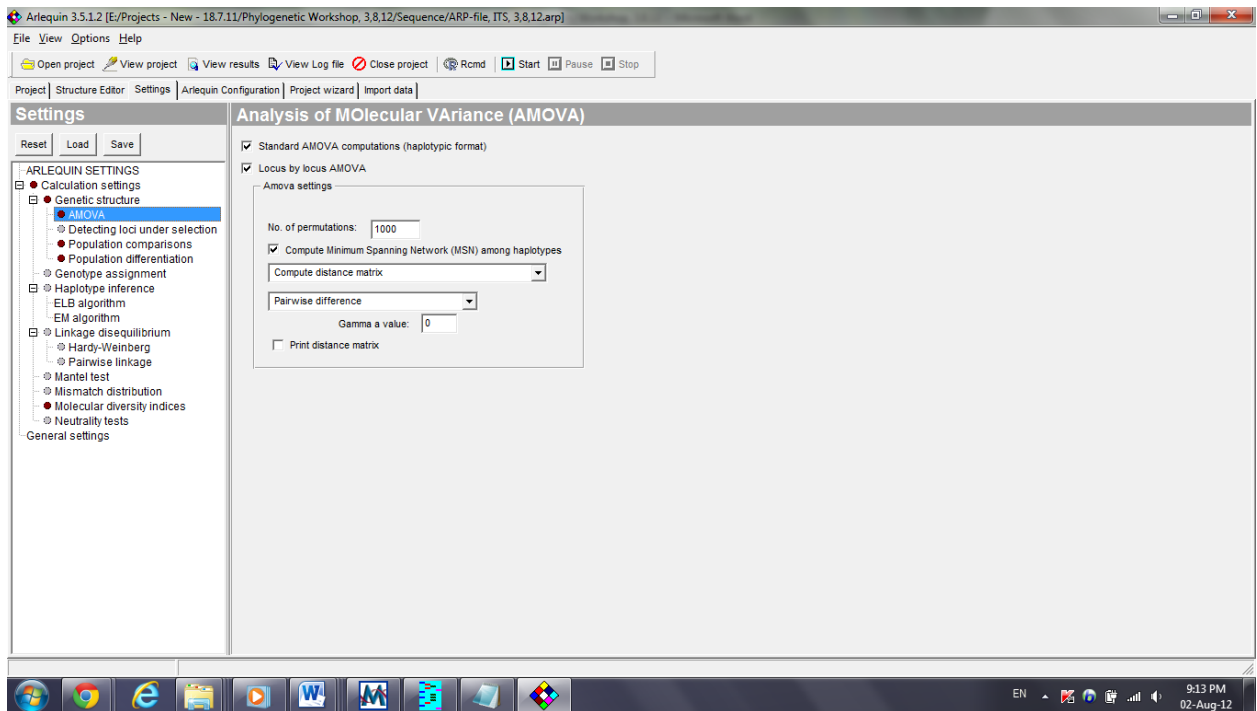


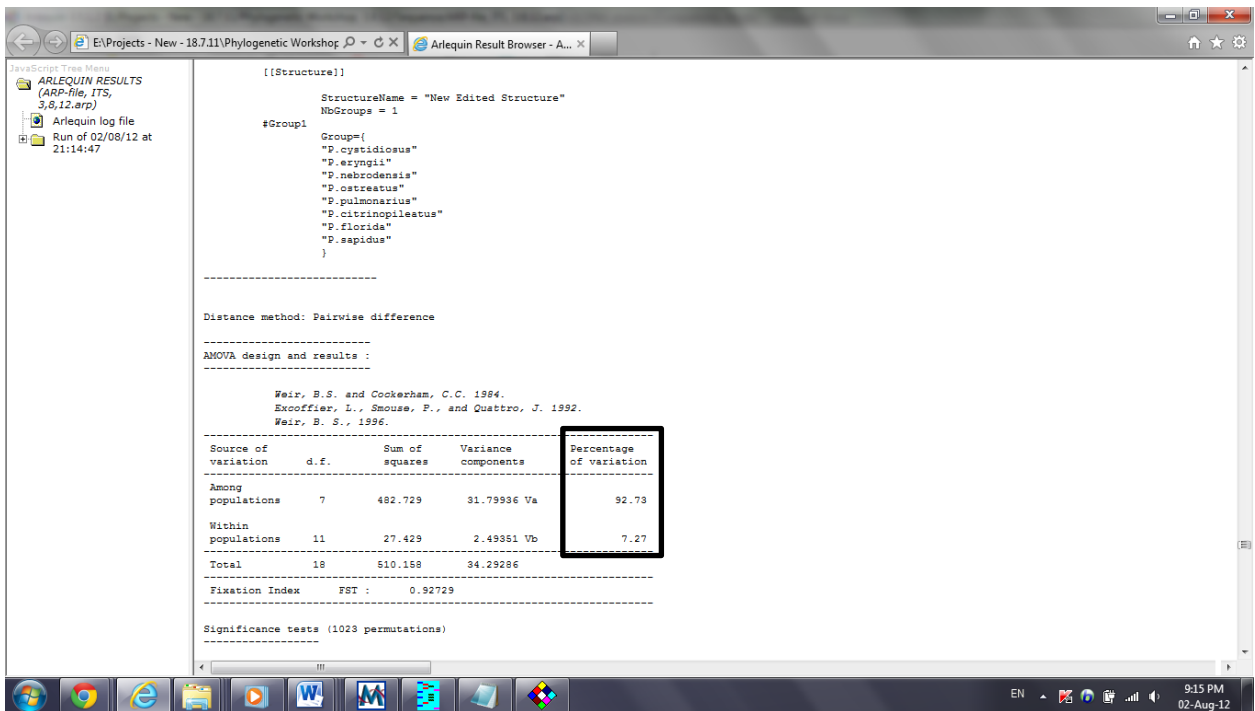
✓ Go to the Structure editor, and edit the Group # to consecutive # (instead of all 1)





- ✓ Go to settings and select AMOVA, Population comparison, Population differentiation and select your preferences





- ✓ Non-differentiation exact  $P$  values
- ✓  $F_{ST}$   $P$  values
- ✓ Exact Test of Sample Differentiation Based on Haplotype Frequencies
- ✓ Etc...

## Exercise & Results

- Generate haplotype data file using the generated FASTA file from first practical session and copy the content of file in the below box:

- Generate manually Arlequin input file (ARP), name it as \*.arp and paste the content of your file in the below box. DO NOT FORGET TO SUBMIT THE SOFT COPY FILE.

- Using Arlequin software, generate AMOVA report for the prepared ARP file.
- Based on the generated report, fill up the below table:

	d.f.	Sum of Squares	Variance Components	Percentage of Variation
Among population				
Within population				

- Generate a table of significant difference (Level 0.05) by using exact *p values* produced in the report. Find and highlight the significant pairs.

--	--

## Practical 6

**Title:** Network Program: Generate a median-joining haplotype network

### Objective:

After completing the practical, you will be able:

1. To generate a median-joining haplotype network

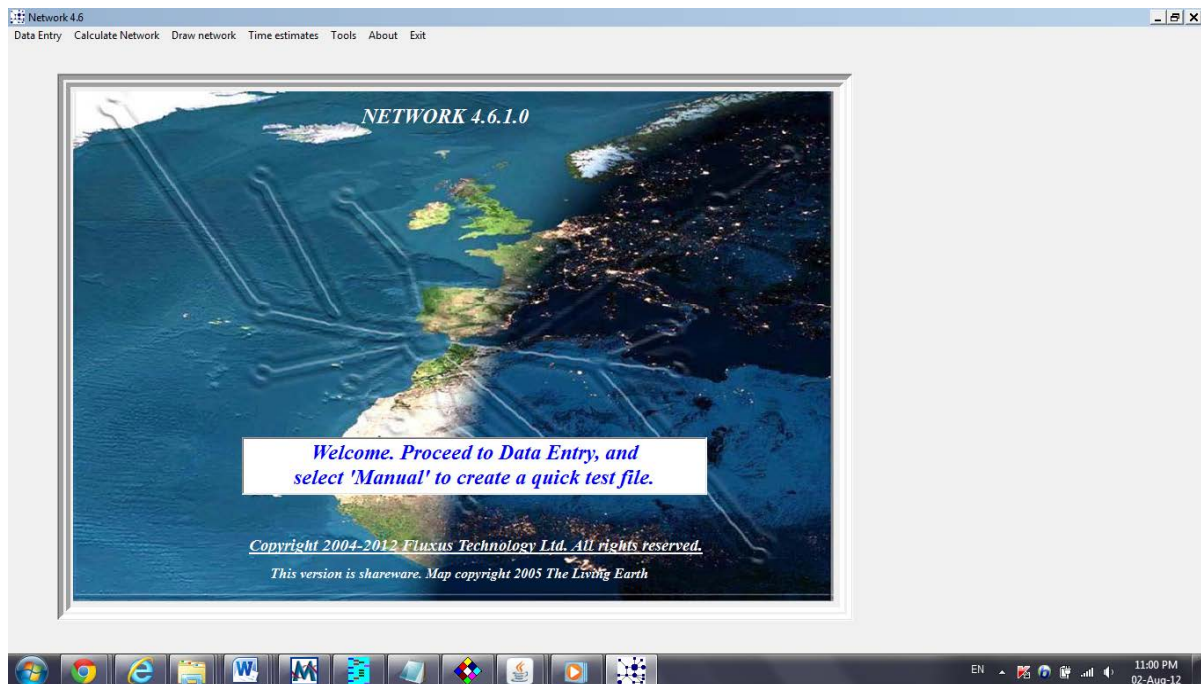
### Introduction:

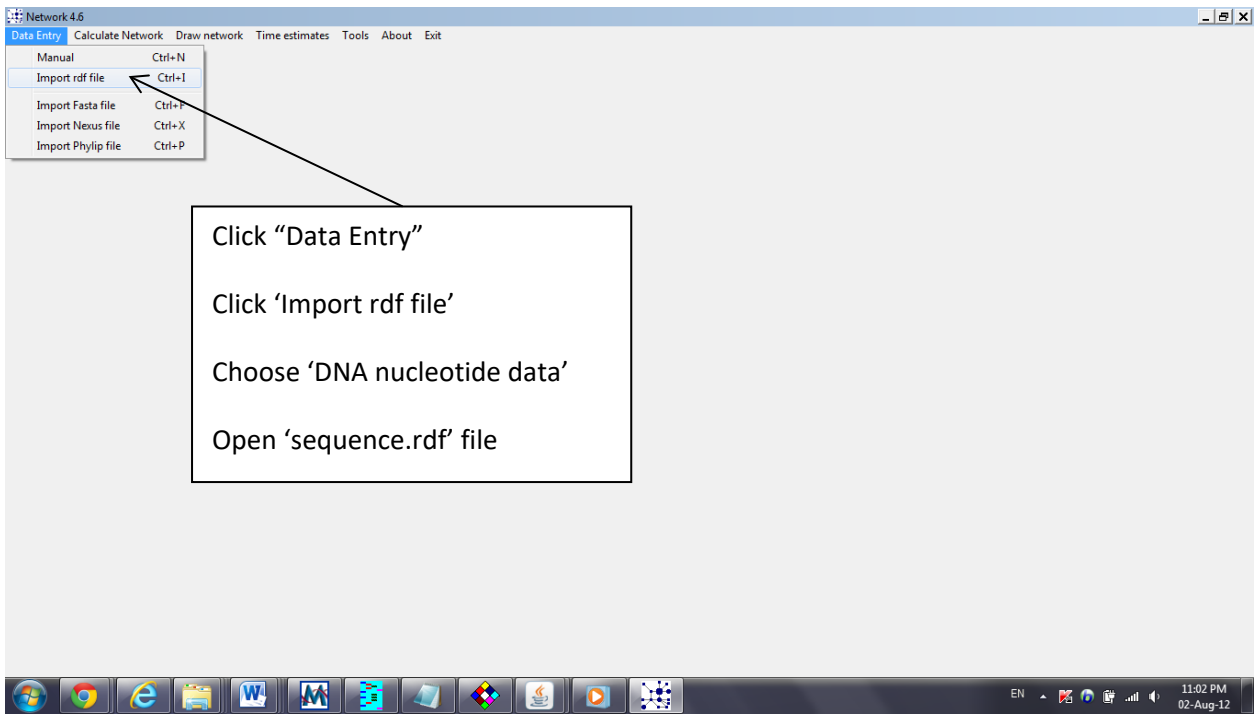
A phylogenetic network or reticulation is any graph used to visualize evolutionary relationships (either abstractly or explicitly) between nucleotide sequences, genes, chromosomes, genomes, or species. They are employed when reticulate events such as hybridization, horizontal gene transfer, recombination, or gene duplication and loss are believed to be involved. They differ from phylogenetic trees by the explicit modeling, by means of the addition of hybrid nodes (nodes with two parents) instead of only tree nodes (nodes with only one parent). Phylogenetic trees are a subset of phylogenetic networks. Phylogenetic networks can be inferred and visualised with software such as Network, SplitsTree and, more recently, Dendroscope. A standard format for representing phylogenetic networks is a variant of Newick format which is extended to support networks as well as trees.

Many kinds and subclasses of phylogenetic networks have been defined based on the biological phenomenon they represent or which data they are built from (hybridization networks, usually built from rooted trees, recombination networks from binary sequences, median networks from a set of splits, optimal realizations and reticulograms from a distance matrix), or restrictions to get computationally tractable problems (galled trees, and their generalizations level-k phylogenetic networks, tree-child or tree-sibling phylogenetic networks).

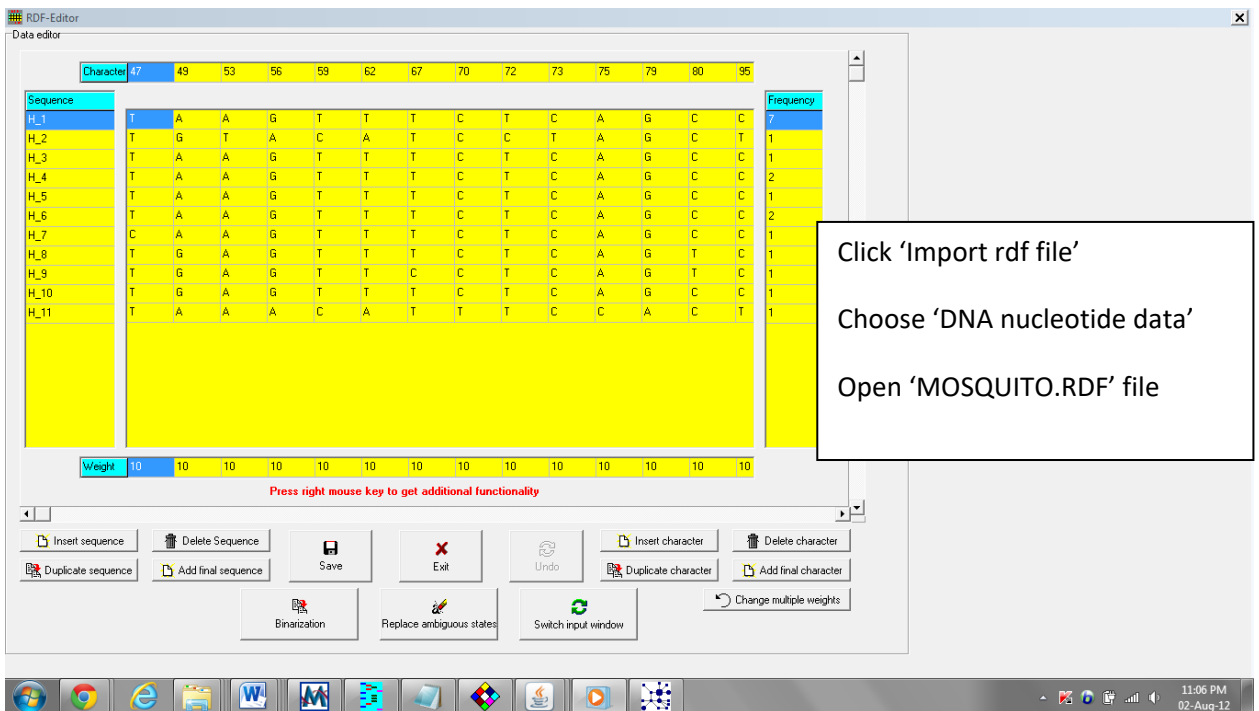
### Instructions:

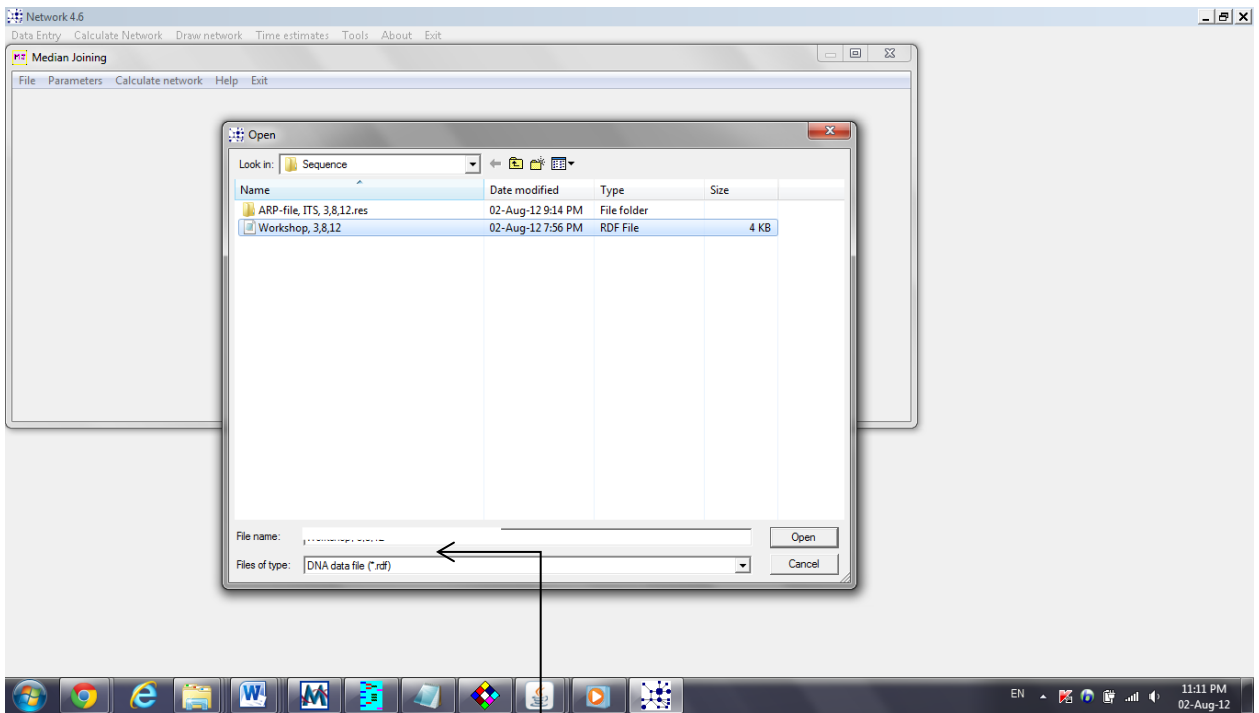
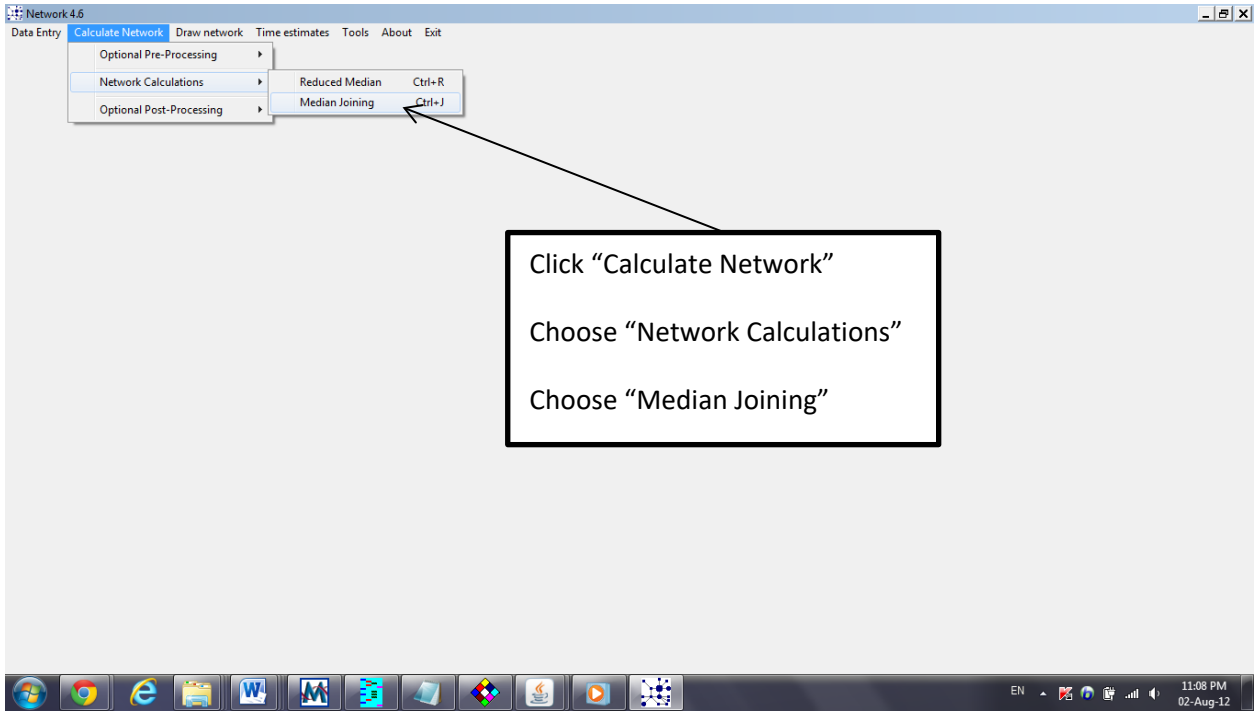
Open Network software from the folder



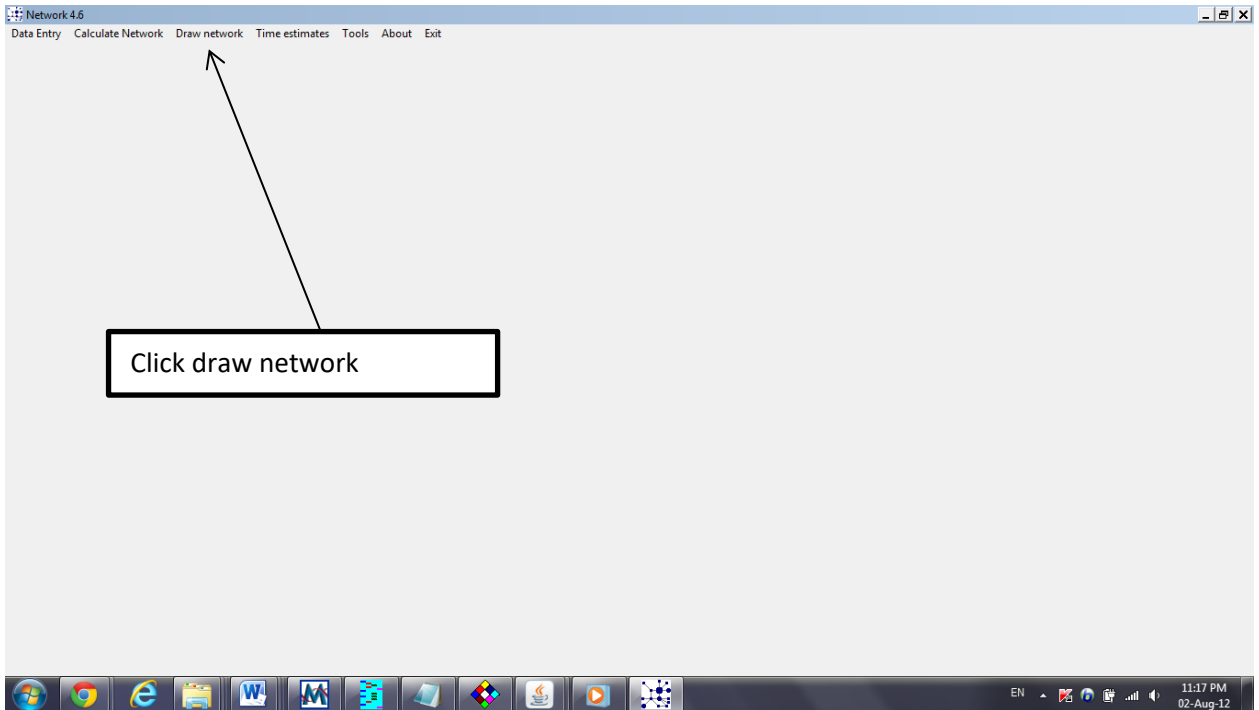
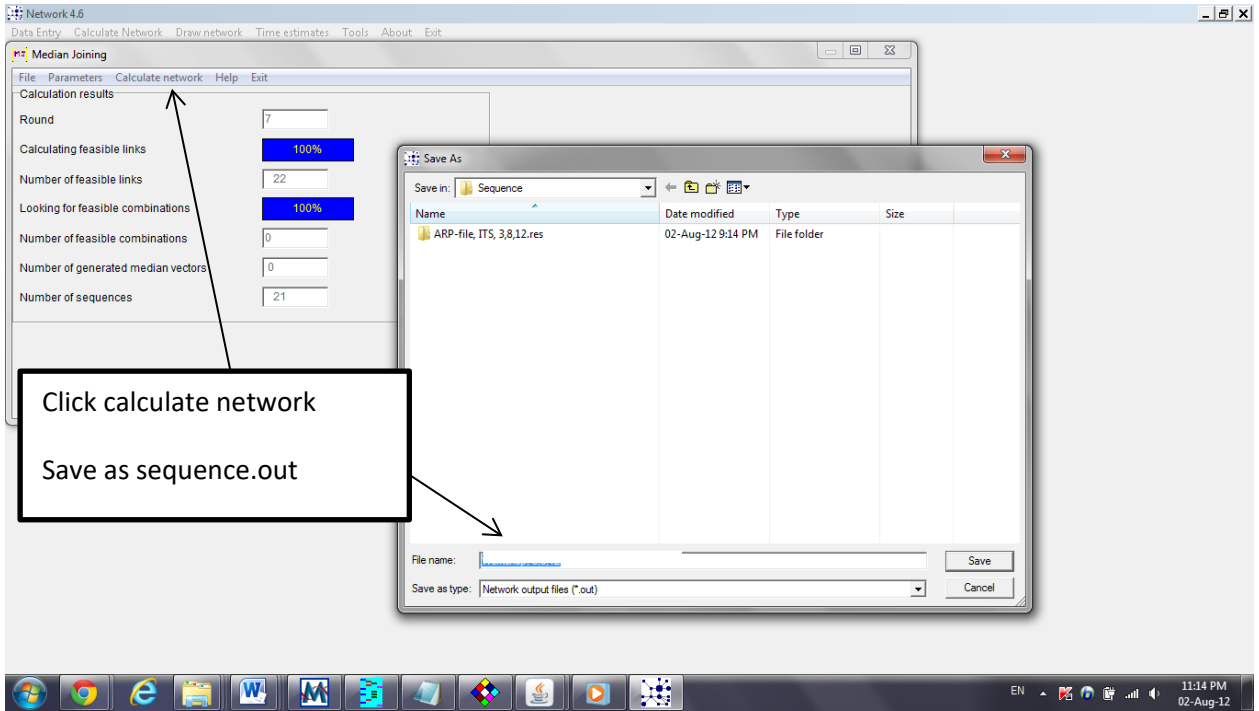


This is the Data Editor and you can see your 'sequence.rdf' file. You can edit if you want, and when finish save it. Otherwise you can just EXIT.

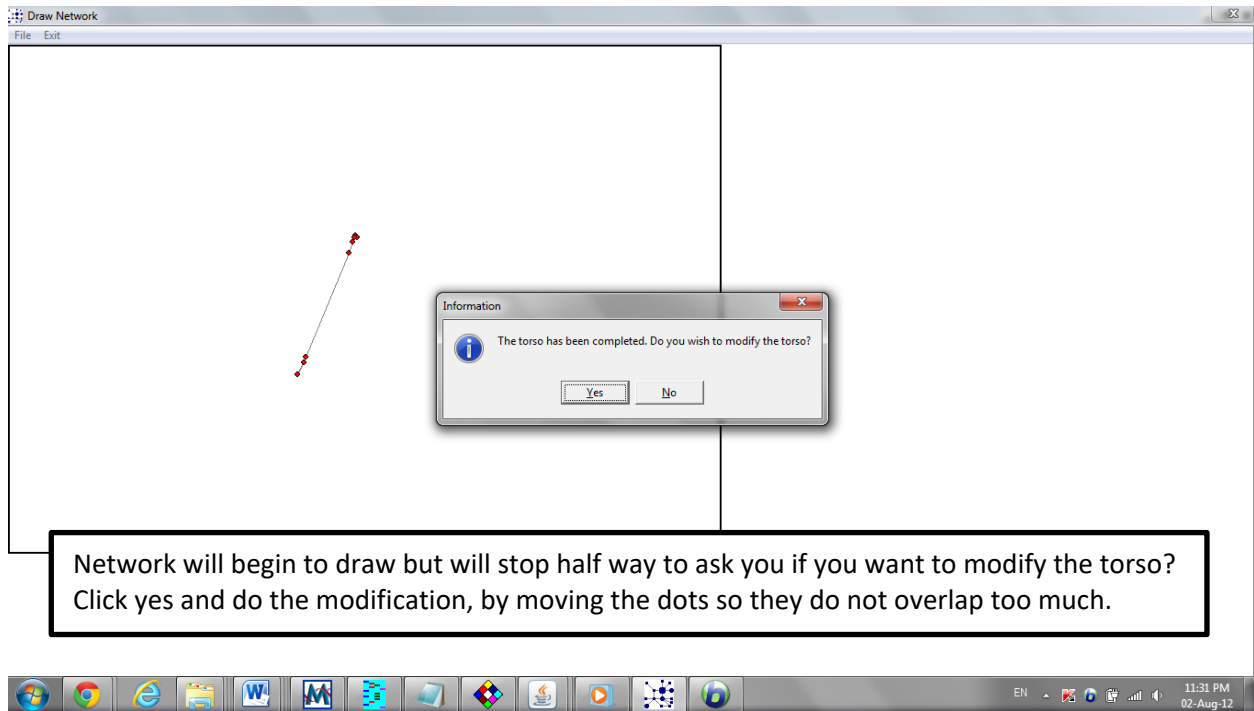
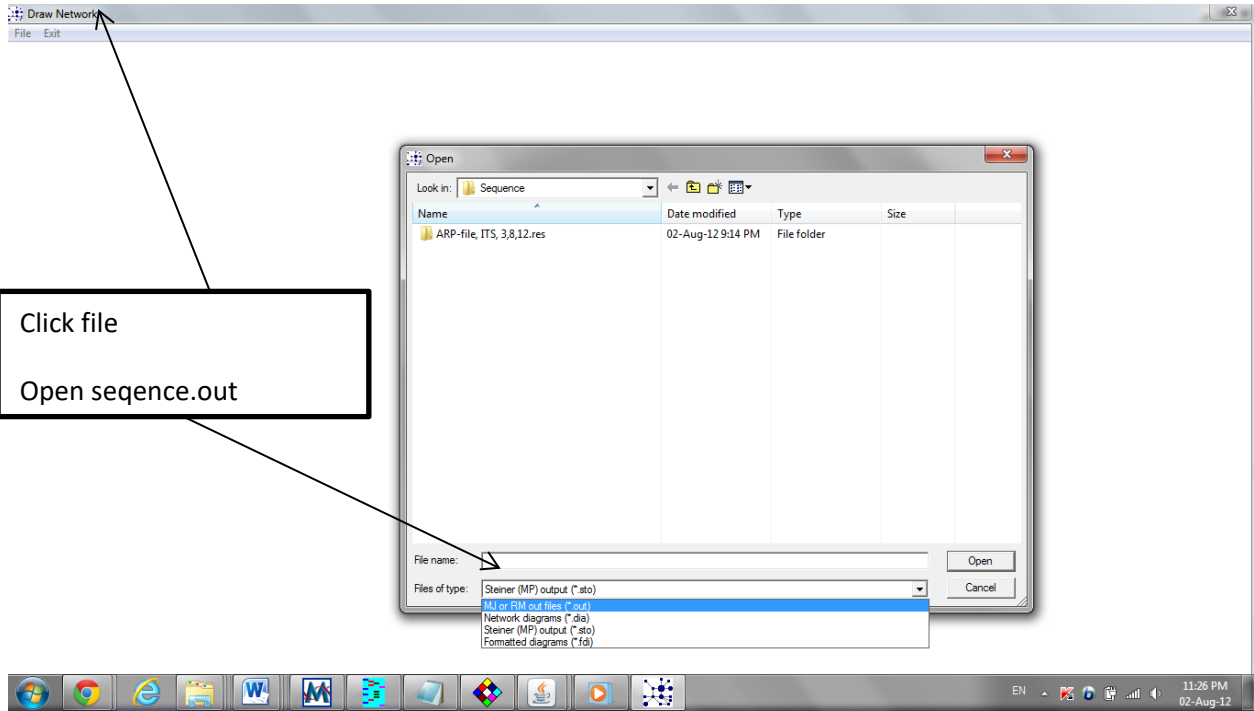




Open 'Sequence.rdf' file







The screenshot shows the 'Draw Network' application window. The main canvas displays a network diagram with several red nodes connected by lines. A text box on the right says 'You can click and drag the dots into new position' with an arrow pointing to one of the red nodes. Below the canvas, there is a control panel with 'Zoom' (+/-) buttons, 'Navigation' buttons (Up, Down, Left, Right), and a 'Screen shift interval (pixels)' set to 30. At the bottom right of the control panel, there are 'Continue' and 'Undo' buttons. A text box on the right says 'When finished, click continue' with an arrow pointing to the 'Continue' button. The Windows taskbar at the bottom shows the time as 11:32 PM on 02-Aug-12.

The screenshot shows the 'Draw Network' application window with a more complex network diagram. The nodes are a mix of red and yellow. A text box on the right says 'You can still move the dots to obtain a desire diagram' and 'Then when finished click finalise' with an arrow pointing to the 'Finalise' button. The control panel at the bottom is the same as in the first screenshot, but the 'Continue' button is replaced by a 'Finalise' button with a green checkmark icon. The Windows taskbar at the bottom shows the time as 11:36 PM on 02-Aug-12.

Draw Network

File Exit

Display parameters

- Nodes proportional to frequencies
- Display only torso
- Display mutated position
- Display node names

Font size: 6

Minimum node size: 4

Maximum node size: 50

Show median vectors

Node size (pixels): 3

? Move Link ? Move Node

? Node Info ? Pie Chart

? Style ? Colour

Statistics

Undo

? Help

Options

Zoom: + -

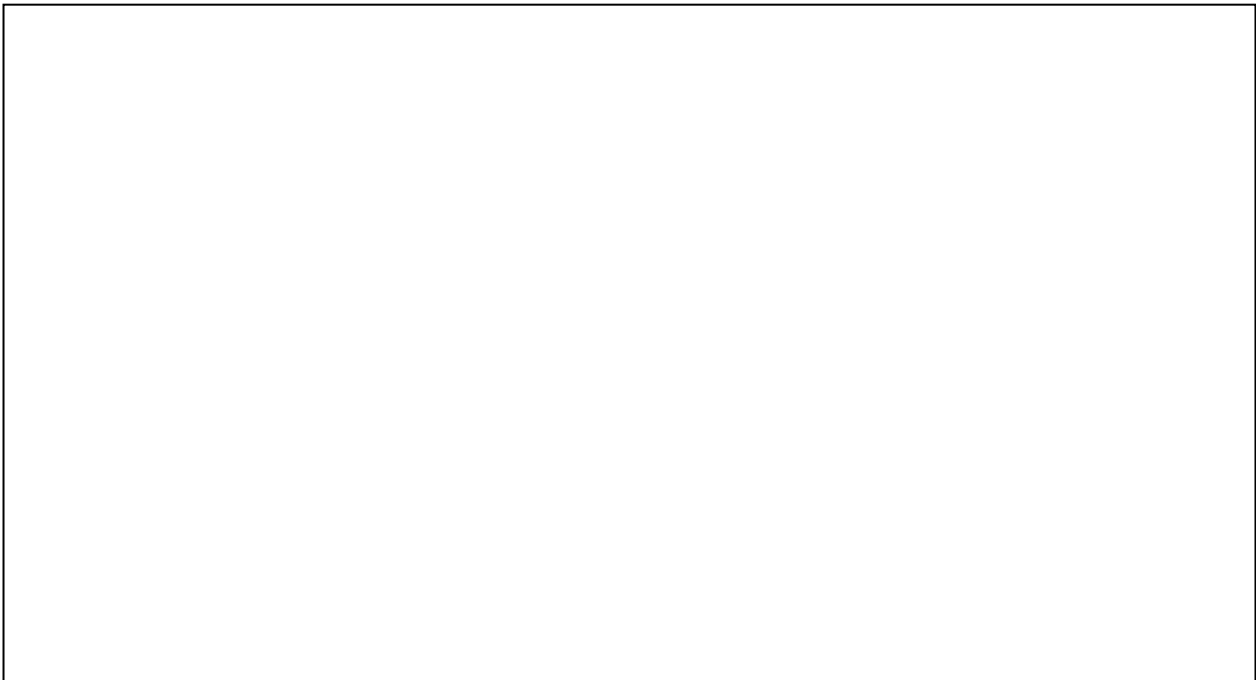
Navigation: Up, Down, Left, Right

Screen shift interval (pixels): 30

EN 11:40 PM 02-Aug-12

## Exercise & Results

- Use the generated sequence.rdf file from earlier practicals to DRAW a network tree based on the constructions given in the current practical.
- Network will begin to draw but will stop half way to ask you if you want to modify the torso? Try to modify the Network by moving the dots so that they should not overlap.
- Use options to draw an appropriate Network tree. Use appropriate font and size, display node names and mutated positions and adjust nodes' size.
- Paste the generated Network tree in the box below:



- ➔ There is abbreviation namely "mv" connects haplotype nodes? Describe what "mv" represents.
- ➔ In some cases, the size of nodes is differed. What is the explanation?

*Notes*

*Notes*





## **LINCOLN UNIVERSITY COLLEGE**

Wisma Lincoln, No. 12, 14, 16 & 18,  
Jalan SS 6/12, 47301 Petaling Jaya,  
Selangor Darul Ehsan, Malaysia.  
Tel.: +603-7806 3478 Fax: +603-7806 3479  
Toll Free: 1-300-880-111  
E-mail: [lucp@lincoln.edu.my](mailto:lucp@lincoln.edu.my)  
Web.: [www.lucp.net](http://www.lucp.net)

ISBN 978-967-2257-20-2



[www.lincoln.edu.my](http://www.lincoln.edu.my) ✽ [www.lucp.net](http://www.lucp.net)