

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 info@lincoln.edu.my Web: www.lucp.net www.lincoln.edu.my

ISBN: 978-967-2257-20-2

Bioinformatics

Table of Content

Experin	nents:	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 safely
- 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 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 safely 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, gauzez 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 safely 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 is 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
- 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.

M.	M4: Alignment Explorer		Automatic Station, 212 Manufacture State	
Dat	ta Edit Search Alignment	Web Sequencer Di	splay Help	
JC) ווֹ Undo	Ctrl+Z	爇 藜 藓 ∽ 陥 耑 麀 × 米 勉 趨 « D	
DN	A Dopy	Ctrl+C		
Г	👗 Cut	Ctrl+X		
	💼 Paste	Ctrl+V		
Г	X Delete			
	✤ Delete Gaps	Ctrl+Del		
	Insert Blank Sequence	Ctrl+N		
	🕍 Insert Sequence From File	e Ctrl+I		
	Select Site(s)			
	Select Sequence(s)			
	Select All	Ctrl+A		
	✓ Allow Base Editing			
H	0.0			
Site	# U∃ © with C	w/o Gaps		
	🗿 🔽 (Ĉ		🥰 📝 📀 🏠 🕅	N 🔺 🔀 🗊II 🕪 10:38 PM

9- Open your sequencing samples from folder "sequence"

M At Alignment Explorer	- 0 - X
Data Edit Search Alignment Web Sequencer Display Help	
□ ☞ 🖬 罕 美 秘 朧 ₩ 兼 1. 単 韓 韓 韓 博 ∽ 暭 ℁ 鳴 × 米 物 20 ◀ ▶	
DNA Sequences Translated Protein Sequences	
	** ***
	GITCAACAICACA
	ACICGCATICICI
	1
	1
	•
Site # 1 🚖 © with ○ w/o Gaps	
🚱 🗿 🥝 📜 🗮 🎽 🚺 🕼 📶 🚱 🚮	()) 10:40 PM 01-Aug-12

- 10- Ensure that the Genetic code selected is correct
- 11- Remember to reverse complement reverse reaction sequences.

M4: Alignment Explorer		x
Data Edit Search Alignment	Web Sequencer Display Help	
Create New	₩ ¥ 1. 🗛 ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ ஜ	
🚔 Open 🕨		
Reopen •		* *
Close		CACA
Save Session Ctrl+S	seeconn in roach ne meachch in christe headean baan bach eaest ceach chagagaach reaan bheas rein christe an each chan gach ceann	стст
🚯 Export Alignment 🔹 🕨		
DNA Sequences		
Protein Sequences		
WC Translate/Untranslate		
3 Select Genetic Code Table		
TI Reverse Complement		
A Neverse complement		
Exit AlnExplorer		
Site # 592 • with	C w/o Gaps	
🚱 💿 🥭	EN 🛪 🔀 🙆 🧔 🕼 EN 🛪 🔀 🗈 and 40 10444	M -12

12- Select all and Align by Clustal-w:

M4: Alignment Explorer	
Data Edit Search Alignment Web Sequencer Display Help	
🗅 🚅 🖬 🛱 🗮 🌚 🎆 W 💥 🏗	桷 쵖 鐃 檜 🖌 🗈 🐰 🏗 🗙 🗮 🗙 🦄 🕼 💶 🕨
DNA Sequences Translated Protein Securitices	
732782_PS1_F AJAJAJATTCAJAJA CAGACATCCAJAGGAJAGGTCACATTGAGGC	TTCALTSIGTEGACCCCACCTTTTGCCAGCGACACGCCCCCCCCCCCC
732783_PS1_R GGGCCTTTTTTCAC TTGTTGACTCTTTCTTTTCTTGACGGAAT	alataccessesecsectocasacatessatatessatetestetettesatestcatessatestcacescatetestccccetete
	M M4: ClustalW Parameters
	DNA
1. Click "Clustal W"	Pairwise Parameters
	Gap Opening Penalty 15
	Gan Evension Penaltu 5.66
	Multiple Parameters
	Gap Opening Penalty 15
	Gan Extension Penaltu 5.65
	DNA Weight Matrix IUB 👻
	Hanston weight 0.3
	Use Negative Matrix DFF 💌
	Delay Divergent Cutoff (%) 30
	Experience Constraints and the constraints and
	2 Her CK X Cancel
2. Click "OK"	
•	ь
Site # 592 🐳 💿 with 🔿 w/o Gaos	
	EN
	EN 🔺 🕅 🗗 📑il 🌵 02-Aug-12

- 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.
- o 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

	🖬 💀 👯 Color C V P	is	0	2	4		UUC	Die -																																											
10. Pathwards FPBAL 0		_			<u> </u>	6	6	T	J T T		c	GT	A	6 0	Т	G	A	A C	C	т	6 0	 6	A 4	G	G	АТ	C	A T	т	A A	т	G A	A	т	r c	A	ст		A .	T G	G	A	G .	гт	G	т	. т	6	C	т	-
22 Publicantas FP0484. 0 <th>1. P. pulmonarius EPPMK-L</th> <th></th> <th>Ι.</th> <th>T</th> <th>T</th> <th></th> <th></th> <th></th> <th></th> <th>-</th> <th>Ť.</th> <th></th> <th>-</th> <th></th> <th>T.T</th> <th></th> <th>T.T</th> <th></th> <th></th> <th></th> <th></th> <th></th> <th></th> <th>Ê</th> <th></th> <th></th> <th></th> <th></th> <th>Ū</th> <th></th> <th></th> <th></th> <th>Ū</th> <th></th> <th></th> <th>Ŭ</th> <th></th> <th>÷</th> <th>Ť</th>	1. P. pulmonarius EPPMK-L		Ι.	T	T					-	Ť.		-														T.T		T.T							Ê					Ū				Ū			Ŭ		÷	Ť
3. Publications FPNAL 1	2. P pulmonarius EPPMB-L		Ľ.		H	÷	Ľ	Ľ.		÷	÷					÷.																																		÷	+
9. Publicity 1 <t< td=""><td>3. P pulmonarius FPPNA-L</td><td></td><td></td><td></td><td>1</td><td>+</td><td></td><td>L.</td><td></td><td>+</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>÷</td><td>t</td></t<>	3. P pulmonarius FPPNA-L				1	+		L.		+																																								÷	t
9. Pation 9. Pation 9. Pation 9	4. P pulmonarius FPPMC-L					+-				+.																																									t
9. Public statis FPGK/P	5. P pulmonarius FPPMB-T					+-				+																																					• .				t
7. Patringingenessis FPCMC 1 <	6. P pulmonarius FPPMK-P																																					-													t
g. P. dotas PEPM(1	7. P citrinopileatus FPCMC					1.				1.																										G		т	т		A		. (s .		A	Α.				t
19 Postestate FPGFT-1 10 Postestate FPGFT-1 10. Postestate FPGFT 10. Po	8. P florida FPFMK					1.				1.																												-													t
10. Postratus FPCF12 . <td< td=""><td>9. P ostreatus FPOPT-1</td><td></td><td></td><td></td><td></td><td>1.</td><td></td><td></td><td></td><td>1.</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>-</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td>• •</td><td></td><td></td><td></td><td>t</td></td<>	9. P ostreatus FPOPT-1					1.				1.																												-									• •				t
11.1 Augusta PSPT . <	10. P ostreatus FPOPT-2					1.				1.																												-													t
12. P syngi FPEX2 .	11. P sapidus FPSPT									1.																												-													t
13.1 Partial 1.1	12. P eryngii FPEK-1									1.																												-									• •				t
1/1. P. proj. 1/1. P. Proj.<	13. P eryngii FPEK-2									1.																												-													t
151 P. Propila A8280174 1 <td>14. P pulmonarius AB115050</td> <td></td> <td>1</td> <td></td> <td>-</td> <td></td> <td>t</td>	14. P pulmonarius AB115050		1																																			-													t
191 Perconstruint 1 <td1< td=""> 1 <td1< td=""></td1<></td1<>	15. P eryngii var ferulae AB288124																																	. (: .			-													T
17. P syngl:A820473 .	16. P eryngii AB288174																																			G		-									•				T
19. P retrostenis EV424300 1	17. P eryngii AB286173																																			G		-												С	
198. P cyntaldosae AV315000 .	18. P nebrodensis EU424308																																			G		-													
	19. P cystidiosus AY315805																							-														С			Α		. (s .		A					T

Close Data Explorer window...

Exercise:

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

	Conserved	Variable	Parsimony informative	Singlton	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: <u>http://blast.ncbi.nlm.nih.gov</u>.
- 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
- o 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.

PCR Template	адеру улих ловы унивности станов ходохи лерник - следовой - средовод - радо станов, унивности станов с
Enter accession, gi, or FASTA	sequence (A refleag record is preferred) 🤬 Clear Range
	Forward primer
Or, upload FASTA file	Choose File No file chosen
Primer Parameters	
Use my own forward primer (5'->3' on plus strand) Use my own reverse primer (5'->2' on minus strand)	e Sinn Cinn
PCR product size	Min Max 70 1000
# of primers to return	10 Min Opt Max Max Tim difference
Primer melting temperatures (Tm)	67 0 60 0 63 0 3 9
Exon/intron selection	A refseq mRNA sequence as PCR template input is required for options in the section 👔
Exon junction span	No preference
Exon junction match	Exon at 5' side Exon at 3' side 7 4 4 Minimal runber of bases that must anneal to exons at the 5' or 3' side of the junction as
Intron inclusion	Primer pair must be separated by at least one infron on the corresponding genomic DNA 😳
Intron length range	Min Max 1000 1000000 @
Primer Pair Specificity Ct	iecking Peremeters
Specificity check	C Enable search for primer pairs specific to the intended PCR template
Search mode	Automatic 🔹 🤬
Database	Refseq mRNA
Exclusion	📄 Exclude predicted Refseq transcripts (accession with XM, XR prefix) 🗐 Exclude uncultured/environmental sample sequences 🌚
Organism	Homo sapiens Enter an organism name (or organism group name such as enterobacteriaceae, rodents), taxonomy (d or select from the suggestion list as you type. 🔬 Add more organisme
Entrez query (optional)	
Primer specificity stringency	Primer must have at least $2 \cdot \overline{2}$ total mismatches to unintended targets, including at least $2 \cdot \overline{2}$ mismatches within the least $5 \cdot \overline{2}$ bins at the 2 and 20
	Ignore targets that have 6 🔻 or more mismatches to the primer.
Max target size	1000
Splice variant handling	auou 💦 😸
Get Primers	□ Show results in a new window 🖉 Use new graphic view 👔

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 Tm 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)

M M4: Analysis Preferences
Options Summary Lest of Phylogeny Test of Inferred Phylogeny Peplications None Beplications Bootstrap Random Seed Interior Branch Test 64238 -

- 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 three based on UPGMA method using the FASTA file you generated in the first practical session (Bootstap 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:

DNA Sequence Polymorphism - Workshop, 3,8,12.fas	Billion Rollington Hardware	that the set of the set	
File Data Display Analysis Overview Generate Tools Window Help			
	Data Information EtcWorkshop, 3,8,12,1ar Number of Nucleotide Sites Total Number of Sequences Sequences Used File Format Type Genome Chromosomal Location Codon Assignation Genetic Code Prefetered Codons Table Sequence Sotta Definition Total Number of Domains Species Chromosome Genomic Position Assembly DnaSP v. 5,10.00	702 19 19 FASTA Sequential DNA Sequence Variation Diploid Autosome No Nuclear Universal None Close	Taurdan 2 August 2012-19-23
			maroady, 2 August 2012, 13.23
🚳 💽 🤅 📜 🔍 📉 🛃			EN 🔺 🎉 🕼 📺 💷 🕂 7:23 PM 🚺

✓ Assign coding region to cover all the sequence

DNA Sequence Polymorphism - Workshop, 3,8,12.fas	Married Rylapont Southy 21.2 Married Rev.	
File Data Display Analysis Overview Generate Tools Window Help		
	Protein Coding Regions Active Data File: E:\\Workshop, 3,8,12.fas Selected region From site: Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the first site Image: Transition of the selected region as a non-coding region If you want to store this information, save (or export) the active data file as a NEXUS file format Cancel DK	Thursday 2 August 2012, 18-26
		Thursday, 2 August 2012; 13:26
		EN 🔺 🌠 🕼 🛱 .ul 🌵 02-Aug-12

✓ Assign the genetic code to "???"

DNA Sequence Polymorphism - Workshop, 3,8,12.fas	Strength and Adapte States 117 Monthly 198	
File Data Display Analysis Overview Generate Tools Window	Help	
	Genetic Code for: Workshop, 3.8.12 fas UUU Phe UCD Ser UAC Tyr UGC Cys UUA Leu UCA Ser UAA Stap UGA Stop UUG Leu UCB Ser UAA Stap UGA Stop UUG Leu UCB Ser UAA Stap UGA Stop UUG Leu UCB Ser UAA Stap UGA Stop CUC Leu CCD Pro CAA Sin CGA Aig CUA Leu CGE Pro CAA Gin CGA Aig AUU IIe ACU Thr AAU San AGG Stop AUU IIe ACU Thr AAU Asan AGG Stop GUA Cae GCA Aig GGA Aig AUU IIe ACU Thr AAU Asan AGG Stop GUI Vai GCC Ala GAA Ga GA Aig GUI Vai GCC Ala GAA Ga GA Gin GGA Aig GUI Vai GCC Ala GAA Ga GA Gin GGA Gin GUI Vai GCC Ala GAA Ga GA Gin GGA Gin GUI Vai GCC Ala GAA Ga GA Gin GGA Gin GUI Vai GCC Ala GAA Gin GGA Gin GUI Vai GCC Ala GAA Gin	
	Thuisday, a	Mugust 2012; 13:27
	N 🛛 🚺	7:27 PM 02-Aug-12

✓ 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"

DNA Sequence Polymorphism - Workshop, 3,8,12.fas	Relation Relation Relation and Real Provider	
File Data Display Analysis Overview Generate Tools Window	Help	
2°		
	Vefine Sequence Sets P. diding Pharks P. exception Pharks	
For Help, press F1		Thursday, 2 August 2012; 19:29
🚳 💿 健 📋 💽 W	N 🤰	名 🍙 🎼nl 🗣 7:29 PM 👘 02-Aug-12

✓ 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).

DNA Sequence Polymorphism - Workshop, 3,8,12.fas	Statement Statements Statements	And Married Real Property lies	
File Data Display Analysis Overview Generate Tools Window	Help		
e iexe Arq			
Output of: Workshop, 3,8,12.fas			
Haplotype/DNA Sequences Data File			
Input Data File: E.\\Wakshop, 3,8,12 fas Number of requences: 19 Number of requences used: 19 Selected region: 1-702 Number of sites: 702 Total number of sites (excluding sites with gaps / missing data): 6 Sites with alignment gaps: considered	6		
Number of included positions: 702	Haplotype/DNA Sequences Data File. Options		
Number of haplotypes, h: 12 Hardenand dware, h: 0,9005	Data Set: All Included Sequences (n = 19)	•	
Hap (1: 61-6] Hap (1: 61-6] Hap (2: 1] [7] Hap (2: 1] [7] Hap (2: 1] [9] Hap (2: 1] [1] Hap (2: 1] [1]	Region to Analyze From site: 1 to: 702 Sites with gaps/missing Not considered @ Considered C Unly gaps/missing are considered	Invariable Sites Removed Included Generate NEXUS Haplotype Data File Alequin Haplotype Litt O Roehl Data File (Network software)	
 Hap, 2: 1 [P. chimopelicatus, FP] Hap, 2: 1 [P. losida, FPWH, P. losida, FPWH-1 Hap, 4: 2 [P. ostradus, FPUPI-1 P. ostradus, FPUPI-2] Hap, 5: 1 [P. sopidus, FPSP, 101100] Hap, 7: 1 [P. polmonasus, JM1100] Hap, 8: 1 [P. polmonasus, JM1101] Hap, 8: 1 [P. polmonasus, JM205174] Hap, 1: 1 [P. netrodenii, EU4240] Hap, 1: 1 [P. cycludiosus, JM3150] 			
For Help, press F1 Haplotype/DNA Sequenc	rs Data File		Thursday, 2 August 2012; 19:35
🚳 💿 ⋵ 🚞 💿 💌 🜆	8 👌		EN 🔺 🐹 🕼 🎯 and 🕕 7-35 PM 02-Aug-12

✓ 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.

DNA Sequence Polymorphism - Workshop, 3,8,12.fas	Reparet States	that the said first the said of the said o	
File Data Display Analysis Overview Generate Tools Window	Help		
Uutput of: Workshop, 3,8,12.fas			
Haplotype/DNA Sequences Data File			
Input Data File: E1Workshop, 3,8,12.fas Number of sequences: 19 Number of sequences used: 19 Selected region: 1-702. Number of sequences used: 19 Selected region: 1-702. Number of ster: 702 Total number of sites (excluding sters with tops / missing data): 61 Sites with alignment gap:: considered Number of included positions: 702 	Haplotype/DNA Sequences Data File. Options Data Set: All Included Sequences (n = 19) Region to Analyze From site: 1 to: 702 Sites with gaps/missing Not considered Considered Only gaps/missing are considered	Invariable Sites Removed Included Generate NEXUS Haplotype Data File Atlequin Haplotype List Rochl Data File (Network software) Cancel DK	
For Help, press F1 Haplotype/DNA Senuence	s Data File		Thursday, 2 August 2012: 19:44
🚳 💿 ⋵ 🚞 💽 💌 🜆			EN 🔺 🌠 🕼 💣nl 🌵 7:44 PM 🚺

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

DNA Sequence Polymorphism - Workshop, 3,8	3,12.fas	Real Property Name	story (a) " Should be "	
File Data Display Analysis Overview Gen	nerate Tool	s Window Help		
Open Data File	Ctrl+O			
Open Multi Data Files (Batch Mode)				
Close Data File	Ctrl+W			
Open Unphase/Genotype Data File				
Save / Export Data As	+	FASTA File Format		
Update NEXUS Data File		MEGA File Format		
Options for Saving (NEXUS Format)		NBRF File Format		
Send All Output to File		NEXUS File Format		
Close Output File		PHYLIP File Format		
Save Current Output	Ctrl+S	HapMap3 Phased Haplotypes File Format		
Page Setup		Arlequin File Format		
Print Output	Ctrl+P	Roehl File Format		
1 E:\\Workshop, 3,8,12.fas				
2 E:\\14 + Refs,IGS1, 30,7,12, test.fas				
3 E:\\14 + Refs,IGS1, 30,7,12.fas				
4 E:\\13 sample+ Refs, ITS, 30,7,12.fas				
Exit	Ctrl+X			
For Help, press F1				Thursday, 2 August 2012; 20:54
🚳 💿 🙆 📋 🖉		l 脉 🧕		EN 🔺 🌠 🍙 🛱nl 🌵 8:54 PM 🚺

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

RP-file, ITS, 3,8,12 - Notepad	
File Edit Format View Help	
[[Profile]	<u>^</u>
Title="ITSI-2" Nbsamp es-19 Genotypicpata=0 DotatopeemM LocusSeparator=NONE MissingData='?'	
[Data]	
[[Samples]]	
12 1 Number of samples	TTGATAGATTCGCAGAGTTGCCCTCTCAGGTCA
3 Sample kame="p.eryngii" Sample kize=5	E
06 2 TAXEARGETTECGTAGGTGAAECTGECGGAAGATCATTAATGAATTCACT-ATGGAGTTGT-TECTGGECTCTAGGGGCATGTGCACGTTCACTAGTCA-TTCAACCACTGTGAACTTT 10 1 TAXEARGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTATAGAATTCACT-ATGGAGTTGT-TECTGGECTCTAGGGGGTATGCACGTTCACTAGTCA-TTGACCACCTGTGAACTTT 10 1 TAXEARGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTATAGAATTCCACT-ATGGGAGTTGT-TGCGGCCTCTAGGGGGTATGCACGCTTCACTAGTCTTTCAACCACCTGTGAACTTT 10 1 TAXEARGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAATGAATTCGCT-ATGGAGTTGT-TGCGGCCTCTAGGGGGTATGCACGCTTCACTAGTCTTTCAACCACCTGTGAACTTT 10 1 TAXEARGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAATGAATTCGCT-ATGGAGTTGT-TGCGGCCTCTAGGGGGTATGCACGCTTCACTAGTCTTCCAACCACCTGTGAACTTT	TTGATAGATCTGTGAAGTCGTCTC-TCAAGTCG TTGATAGATCTGTGAAGTCGTCTC-TCAAGTCG TTGGTAGATCTGCGAAGTCGTCTC-TCAAGTCG TTGGTAGATCTGCGAAGTCGTCTC-TCAAGTCG
sampleName="P.nebrodensis" SampleSize=1 Haplotype no. SampleSize=2, SampleSize=2, SampleSize=4 AccreceGaAGGATCATTAATGAATTCGCT-ATGGAGTTGT-TGCTGGCCTCTAGGGGCATGTGCACGCTTCACTAGTCTTTCAACCACCTGTGAACTTT SampleSize=2,	TTGGTAGATCTGCGAAGTCGTCTC-TCAAGTCG
Sampleuata= { Sa	TTGATAGATCTGTGAAGTCGTCTT-TCAAGTCG
Samp lestance +, pu monarius Samp lestance Samp lestance 01 6 TAACAAGGTTTCCGTAGGTGAACCTCGGAAGGATCATTAATGAATTCACT-ATGGAGTTGT-TGCTGGCCTCTAGGGGCATGTGCACGCTTCACTAGTCTTTCAACCACCTGTGAACTTT 07 1 TAACAAGGTTTCCGTAGGTGAACCTCGGAAGGATCATTAATGAATTCACT-ATGGAGTTGT-TGCTGGCCTCTAGGGGCATGTGCACGCTTCACTAGTCTTTCAACCACCTGTGAACTTT 1 TAACAAGGTTTCCGTAGGTGAACCTCGGAAGGATCATTAATGAATTCACT-ATGGAGTTGT-TGCTGGCCTCTAGGGGCATGTGCACGCTTCACTAGTCTTTCAACCACCTGTGAACTTT 3 Samp lestize-1 Samp lestize-1 Samp lestize-1 Samp lestize-1	TTGATAGATCTGTGAAGTCGTCCT-TCAAGTCG TTGATAGATCTGTGAAGTCGTCCT-TCAAGTCG
02 1 TAACAAGGTTTCCGTAGGTGAACCTGCGGAGGATCATTAATGAATTCGCTTTTGAAGCTGAATGCTGGCCCTTAGGGGCATGTGCACGCTTCATTAGTCCCCTTTCACACCCCCTGGCACCTT } samplename="P.florida" samplename="f" samplename="f"	TTGATAGATTCGCTGG
03 1 TAACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTAATGAATTCACT-ATGGAGTTGT-TGCTGGCCTCTAGGGGCATGTGCACGCTTCACTAGTCTTTCAACCACCTGTGAACTT	TTGATAGATCTGTGAAGTCGTCTT-TCAAGTCG
	- F
	EN 🔺 🌠 🕼 🛱 ant 🌵 8:57 PM 02-Aug-12

✓ Save as **.arp:

<u> </u>	ARP-file, IT	rs, 3,8,12 - Notepa	d marine and an and					J
File	Edit F	ormat View He	elp					
[Pr	ofile]							-
	Tit NbSam	tle="ITS1-2" ples=19	Save As			×		
	nooun	Genot	O V K Phy	ogenetic Workshop, 3.8.12	Search Seauen	ce O		
		Locus						
	_		Organize 🔻 New	folder		······································		
[Da	taj		☆ Favorites	▲ Name	Date modified	Туре		
	[[58	amples]]	E Desktop	1st BASE 732782 PS1 F	10-Nov-11 7:19 AM	Chromatogram File		
			Downloads	1st_BASE_732783_PS1_R	10-Nov-11 7:19 AM	Chromatogram File		
12	1	TAACAAGGTT	👠 Google Drive	ARP-file, ITS, 3,8,12	30-Jul-12 1:50 AM	ARP File	AAGTACATTCAACCACTTGTGCACTTTTGATAGATTCGCAGAGTTGCCCTCTCAGGTCA	
}			🔛 Recent Places	Haplotype file	02-Aug-12 7:56 PM	Text Document		
				Workshop, 3,8,12	02-Aug-12 6:56 PM	FAS File		
06	2	TAACAAGGTT	Calibraries	Workshop, 3,8,12	02-Aug-12 6:55 PM	MAS File	TAGTCTTTCAACCACCTGTGAACTTTTGATAGATCTGTGAAGTCGTCTC-TCAAGTCG	
09	1	TAACAAGGTT	Documents	Workshop, 3,8,12	02-Aug-12 6:55 PM	MEG File	TAGTCTTTCAACCACCTGTGAACTTTTGGTAGATCTGTGAAGTCGTCTC-TCAAGTCG	
10	1	TAACAAGGTT	Music	Workshop, 3,8,12	02-Aug-12 7:56 PM	RDF File	TAGTCTTTCAACCACCTGTGAACTTTTGGTAGATCTGCGAAGTCGTCTC-TCAAGTCG	
1			Pictures					
			Podcasts					
11	1	TAACAAGGTT	Videos	÷ 1		•	TAGTCTTTCAACCACCTGTGAACTTTTGGTAGATCTGCGAAGTCGTCTC-TCAAGTCG	
-			File name:	ARP-file, ITS, 8,8,12.arp		-		
~	2		Save as type	All Files		•		
}	2	TAACAAGGTT					TAGTCTTICAACCACCTGTGAACTTTTGATAGATCTGTGAAGTCGTCTT-TCAAGTCG	
			Hide Folders	Encoding: ANSI	 Save 	Cancel		ĩ
01	6	TAACAAGGTT						
07	ĭ	TAACAAGGTTT	TCCGTAGGTGAACCTGC	GAAGGATCATTAATGAATTCACT-ATGGAGTT	GT-TGCTGGCCTCTAGGGGG	ATGTGCACGCTTCA	CTAGTCTTTCAACCACCTGTGAACTTTTGATAGATCTGTGAAGTCGTCCT-TCAAGTCG	
3			Sample	Name="P.citrinopileatus"				
			Samples Samples	Size=1 Data= {				
02	1	TAACAAGGTTT	TCCGTAGGTGAACCTGC	GAAGGATCATTAATGAATTCGCTTTTGAAGCT	GAATGCTGGCCCTTAGGGGG	ATGTGCACGCTTCA	ITTAGTCCCCTTTCACACCCCTGTGCACCTTTGATAGATTCGCTGG	
3			Sample	Name="P.florida"				
			Samples	Size=1 Data= {				
03	1	TAACAAGGTTT	ICCGTAGGTGAACCTGC	GGAAGGATCATTAATGAATTCACT-ATGGAGTT	GT-TGCTGGCCTCTAGGGGG	ATGTGCACGCTTCA	CTAGTCTTTCAACCACCTGTGAACTTTTGATAGATCTGTGAAGTCGTCTT-TCAAGTCG	
4							•	
							EN , 📴 🧥 🛱 🚽 🖌 9:03 PM	ľ
	🥑 🔼						02-Aug-12	

Arlequin 3.5 program

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



Arlequin 3.5.1.2	-		and the second se	-	the Manholman 1817	Manual Review	-	-		- 0 X
<u>File View Options H</u> elp										
🔄 Open project 🖉 View project 📓 View	v results 🛛 🖓 View	Log file 🖉 Close	project 🕼 Rcmd 💽 Sta	art 😐	Pause 🔳 Stop					
About Arleo Open a project already present of	on your disk		. El.	_			-	-x-)		
About Arlequin	Copen Anequi	i project or batcr	The	_						
	Look in:	Sequence		•	🗢 🗈 💣 🎫					
Arlequin ver 3.5.1.2	C.	Name	*		Date modified	Туре	Size			
(c) Laurent Excoffier 1998-2010	Recent Places	ARP-file, IT	5, 3,8,12		30-Jul-12 1:50 AM	ARP File	11 KB			
Computational and	Desktop									
Population Gene										
CMPG	Libraries									
	Computer									
Institute of Ecology and Evolution										
institute of Ecology and Erolation	Network									
Swiss Institute of Bioinformatics										
E L 0040 http://empg.upib										
February 2010 http://mbg.unio										
		Ohlantanna						0.00		
		Object name.	Annie, 113, 3,0,12					Cread		
		objects or type:	Anequin project tiles (".arp)				•	Cance		
📀 💿 🙆 🚞		M .							EN 🔺 🔀 🕼 🕻	9:09 PM II () 02-Aug-12

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

Arlequin 3.5.1.2 [E:/Projects - New - 18.7.11)	/Phylogenetic Workshop, 3,8,12/Sequence/ARP-file, ITS, 3,8,12.arp]	- 0 X						
Eile View Options Help								
😑 Open project 🖉 View project 🧕 View re	🔄 Open project 🖉 View project 👰 View results 🖏 View Log file 🔗 Close project 🚳 Rcmd 🕨 Start III Pause III Stop							
Project Structure Editor Settings Arlequin Con	figuration Project wizard Import data							
Genetic Structure Editor								
Assign a group number to populations:								
(a value of "0" implies no group assigment)	Resulting structure:							
Group Poysteliasus 1 Poysteliasus 2 Penyngil 3 Ponetrodensis 4 Postreatus 1 Poutoneatus 1 Potinopietus 1 Potinopietus 1 Potinopietus 1 Potinopietus 1 Pinorda 1 Pinorda 1 Pisapidus	Councile Structure Corrup 1 Council C							
Update Project								
		11.						
🚱 📀 🥭 🚞	🖸 🖳 📶 🧯 🕢 🚸 🛛 🕬 🕬	nl 🗣 9:10 PM 02-Aug-12						

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

Arlequin 3.5.1.2 [E:/Projects - New - 18.7.]	11/Phylogenetic Workshop, 3.8,12/Sequence/ARP-file, ITS, 3.8,12.arp]	- 0 X				
File View Options Help						
😑 Open project 🖉 View project 🛯 🙀 View	results 🖏 View Log file 🖉 Close project 🌚 Rand 💽 Start 🔟 Pause 🔳 Stop					
Project Structure Editor Settings Arlequin C	onfiguration Project wizard Import data					
Settings	Analysis of MOlecular VAriance (AMOVA)					
Reset Load Save ARLEOUIN SETTINGS Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings Calculation settings	✓ Standard AMOVA computations (haplotypic format) ✓ Locus by locus AMOVA Amova settings No. of permutations: © Compute Minimum Spanning Network (MSN) among haplotypes Compute distance matrix ♥ Pairwise difference ♥ Print distance matrix					
		11				
🚳 🧿 🙆 🚞		9:13 PM 02-Aug-12				

Arlequin 3.5.1.2 [E:/Projects - New - 18.7.1	11/Phylogenetic Workshop, 3,8,12/Sequence/ARP-file, ITS, 3,8,12.arp]	- 0 X
<u>File View Options H</u> elp		
😑 Open project 🖉 View project 🗔 View	results 🖏 View Log file 🔗 Close project 🕲 Rcmd 🗈 Start 💷 Pause 🔳 Stop	
Project Structure Editor Settings Arlequin Co	onfiguration Project wizard Import data	
Settings	Population comparisons	
Reset Load Save ARLEQUIN SETTINGS Calculation settings Calculation settings Centeric structure AMOVA Detecting loci under selection Population difference Boulation settings Cenotype assignment Hapiotype inference Divise inkage Mantel test Minisch distribution Molecular diversity indices Reneral settings	Compute pairwise FST Genetic distances settings Reynold's distance Compute parwise differences (p) Estimate relative population sizes No. of permutations: 100 Significance levet 0.05 Compute distance matrix Pairwise difference Gamma a value: 0	
		9:19 PM
		02-Aug-12

Contract of Contract of State	No. of Concession, Name of Street, or other othe	
E\Projects - New - 1	18.7.11\Phylogenetic Workshop 🔎 🕆 🖒 🏈 Arlequin Result Browser - A 🗙	n 🛪 🌣
ALL-SCIPT Tree Menu ALL-SCIPT Tree Menu (ARP-file, ITS, 3,6,12,erp) ALL-SCIPT ALL-	<pre>([Structure]) StructureName = "New Edited Structure" NbDroups = 1 fGroup1 Group=("P.csyngii" "P.setresture" "P.cstresture" "P.cstresture" "P.cstresture" "P.cstresture" "P.cstresture" "P.cstresture" "P.sapidur" "P.sapidur" "P.sapidur" }</pre>	
	Source of Sum of Variance Percentage variation d.f. squares components of variation	
	Among populations 7 402.729 31.79936 Va 92.73	
	Within populations 11 27.429 2.49351 Vb 7.27	(2)
	Total 18 510.158 34.29286	
	Fixation Index FST : 0.92729 	
		Þ
	🗧 🖸 🖳 🚮 🤰 🕢 🗇	EN 🔺 🌠 🕼 🛱 📶 🌵 02-Aug-12

- ✓ Non-differentiation exact *P* values
- ✓ FST *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.
- o 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



🐺 Network 4.6			_ <u>_</u> X
Data Entry Calculate Network Dra	w network Time estimates Tools About Exit		
Manual Ctri+N			
Import Fasta file Ctrl+P	\mathbf{k}		
Import Nexus file Ctrl+P			
import right me can ri			
	Click "Data Entry"		
	· · · · · · · · · · · · · · · · · ·		
	Click 'Import rdf file'		
	Choose 'DNA nucleotide data'		
	Open sequence.rdf file		
📀 💿 🙆 [🗧 🖳 🚺 🎽 🖉 💽	X	EN 🔺 🔀 🕼 🛱 .nll 🌵 11:02 PM 🛛 02-Aug-12

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.

🗰 RDF-Edit	or															×
Data editor																
	Character	47	<mark>49</mark>	53	56	59	62	67	70	72	73	75	79	80	95	
Sequence	e															Frequency
H_1			A	A	G	Т	Т	т	С	T	C	A	G	С	С	7
H_2		Т	G	т	A	С	A	т	С	С	т	A	G	С	Т	1
H_3		Т	A	A	G	Т	Т	Т	С	T	С	A	G	С	С	1
H_4		Т	A	A	G	Т	Т	т	С	T	С	A	G	С	С	2
H_5		т	A	A	G	т	т	т	С	т	С	A	G	С	С	1
H_6		т	A	A	G	т	т	т	С	т	С	A	G	С	С	2
H_7		С	A	A	G	т	т	т	С	т	С	A	G	С	С	
H_8		T	G	A	G	т	т	т	С	т	С	A	G	Т	С	Click Import rdf file
H_9		Т	G	A	G	т	т	С	С	т	С	A	G	т	С	1
H_10		T	G	A	G	т	т	т	С	т	С	A	G	С	С	1
H_11		Т	A	A	A	С	A	T	T	T	С	С	A	С	T	Choose 'DNA nucleotide data'
																Open 'MOSQUITO.RDF' file
	Weight	10	10	10	10	10	10	10	10	10	10	10	10	10	10	
					Pres	s right ma	use key	to get add	litional fu	unctional	ity					
•												-				
🔤 🛅 Inse	ert sequence		👚 Dele	te Sequen	ce	8			×				🎽 Insert c	haracter		Delete character
🖳 Dupli	cate sequenc	e	🖰 Add fi	inal sequer	nce	Save		E	xit		Undo	E R	Duplicate	character	Ľ	Add final character
	Binarization Image multiple weights Switch input window															
@	0	2			N					> (<u></u>	0	X			- 🄀 🗘 🔐 .ati (↓ 11:06 PM 02-Aug-12

Networ	k 4.6				_ 8 ×
Data Entry	Calculate Network Draw network Optional Pre-Processing Network Calculations Optional Post-Processing	 Time estimates Tools Reduced Mediar Median Joining 	About Exit		
				Click "Calculate Network" Choose "Network Calculations" Choose "Median Joining"	
9	O (2) (2) (2) (2) (2) (2) (2) (2) (2) (2)	😬 🚾 📴		 Sector 10 (1998) 	EN 🔺 🌠 🍘 📴iil 🗣 11:08 PM 02-Aug-12

🔆 Network 4.6								_ 8 X
Data Entry Calculate Network Draw netw	vork Time estimates Tools About Exit							
Median Joining						Л		
File Parameters Calculate network H	elp Exit							
(11th Open				×			
	pa, open							
	Look in: 🔒 Sequence 💌	← 🗈 💣 💷▼						
	Name	Date modified	Туре	Size				
	ARP-file, ITS, 3,8,12.res	02-Aug-12 9:14 PM	File folder					
	Workshop, 3,8,12	02-Aug-12 7:56 PM	RDF File	4 KB				
	File name:				Open			
	\leftarrow							
	Files of type: DNA data file (*.rdf)			<u> </u>	Cancel			
	<u></u>							
	🛛 🖳 👫 📑 🖉 🔺	🔈 [🖌 🚺				EN .	• 🔀 🕼 🗑 🖬 🕕	11:11 PM
								02-Aug-12
		10		~				
	0	pen 'Sequ	ience.rdť	tile				
		•						

🙀 Network 4.6		_ 8 ×
Data Entry Calculate Network Draw network Time estimates Tools Ab	About Exit	
File Parameters Calculate network Help Exit		
Calculation results Round 7		
Calculating feasible links 100%	🐨 Save As	
Number of feasible links 22	Save in: 🕌 Sequence 🗸 🔶 🖆 📰 🗸	
Looking for feasible combinations 100%	Name Date modified Type Size	
Number of feasible combinations	ARP-file, ITS, 3,8,12.res 02-Aug-12.9:14 PM File folder	
Number of generated median vectors		
Number of sequences		
Click calculate network	1	
Save as sequence.out		
	File name: Save	
	Save as type: Network output files ("out)	
🚳 👩 🈂 🚞 🔣 🚮	e 🖉 🕹 💽 💥 💿 🖄 e 🕫 🖉	11:14 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 fornt 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?







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 Web.: www.lucp.net

ISBN 978-967-2257-20-2



www.lincoln.edu.my * www.lucp.net