UNIVERSITY OF HELSINKI

Workshop
Clostridium botulinum

8th - 12th December 2008
Helsinki, Finland

WORKSHOP MANUAL

MARIE CURIE ACTIONS

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

MARIE CURIE ACTIONS

Clostridia Workshop 4: Molecular Epidemiological Tools for *Clostridium botulinum*

Venue:	Department of Food and Environmental Hygiene Faculty of Veterinary Medicine University of Helsinki Finland
Organiser:	Miia Lindström
Tutors:	Professor Yukako Fujinaga, Osaka University, Osaka, Japan Dr. Karen Hill, Los Alamos National Laboratory, Los Alamos, New Mexico, USA DVM Päivi Lahti, University of Helsinki, Finland Dr. Miia Lindström, University of Helsinki, Finland Dr. Ute Messelhäusser, Bavarian Health and Food Safety Authority, Oberschleißheim, Germany Dr. Panu Somervuo, University of Helsinki, Finland
Date:	8-12 December 2008

Programme

Please note that the time schedule is prone to change depending on how the laboratory work proceeds!

Monday 8 Dec

Time	Activity	Room	Tutor
18:00-19:00	LECTURE: Introduction to molecular epidemiology of *Clostridium botulinum*	B437	Miia Lindström
19:00-	Welcome snack	Coffee	

Tuesday 9 Dec

Time	Activity	Room	Tutor
9:00-9:15	INFO: Opening and practical info	B437	Miia Lindström
9:15-10:00	LECTURE: Introduction to amplified fragment length polymorphism (AFLP) analysis of *C. botulinum* strains	B437	Karen Hill
10:00-11:15	LAB: AFLP	B439	Karen Hill
	Coffee 15-30 min	Coffee	
11:15-12:15	LECTURE: Introduction to real-time PCR detection of *C. botulinum*	B437	Ute Messelhäusser
12:15-13:45	LAB: Real-time PCR	B439	Ute Messelhäusser
	LAB: AFLP	B439	Karen Hill
13:45-14:30	LUNCH	UniCafe	
14:30-15:15	Group A LAB: Real-time PCR	B439	Ute Messelhäusser
	Group B Coffee	Coffee	
15:15-16:00	Group B LAB: Real-time PCR	B439	Ute Messelhäusser
	Group A Coffee	Coffee	
16:00-17:00	LAB: AFLP	B439	Karen Hill
	Departure to City Centre		
17:00-19:00	Evening meal	UniCafe	
19:00-	Meet-up in Pub Vanhan Kuppila (non-sponsored)		

Wednesday 10 Dec

Time	Activity	Room	Tutor
9:00-9:30	LAB: DNA microarrays	B439	Päivi Lahti
9:30-10:15	LECTURE: Introduction to DNA microarrays	B437	Miia Lindström
10:15-12:45	LAB: AFLP	B439	Karen Hill
	Coffee	Coffee	
	LAB: DNA microarrays	B439	Päivi Lahti
12:45-13:30	LUNCH	UniCafe	
13:30-17:00	LAB: DNA microarrays	B439	Päivi Lahti
	Coffee	Coffee	
	LAB: AFLP	B439	Karen Hill
	Departure to City Centre		
17:00-19:00	Evening meal	UniCafe	
19:00-	Meet-up in Pub Vltava (non-sponsored)		

Thursday 11 Dec

Time	Activity	Room	Tutor
9:00-9:30	LAB: DNA microarrays	B439	Päivi Lahti
9:30-10:15	Group A DEMO: DNA microarray scanning	B339	Panu Somervuo
	Group B Coffee	Coffee	
10:15-11:00	Group B DEMO: DNA microarray scanning	B339	Panu Somervuo
	Group A Coffee	Coffee	
11:00-12:15	Discussions		
12:15-13:00	LUNCH	UniCafe	
13:00-14:30	COMPU: AFLP data analysis	B239	Panu Somervuo Karen Hill
14:30-15:00	Coffee	Coffee	
15:00-15:30	LECTURE: DNA microarray data analysis	B239	Panu Somervuo
15:30-17:30	COMPU: DNA microarray data analysis	B239	Panu Somervuo
	Departure to City Centre		
19:00-	Workshop dinner	Virgin Oil	

Friday 12 Dec

Time	Activity	Room	Tutor
9:00-11:30	COMPU: DNA microarray data analysis	B239	
	Coffee 30 min	Coffee	Panu Somervuo
11:30-12:15	LECTURE: The botulinum neurotoxin	B239	Yukako Fujinaga
12:15-13:15	FINAL DISCUSSION	B239	All tutors
13:15-14:00	LUNCH	UniCafe	
14:00-	Departure		

COMPARATIVE GENOMIC HYBRIDIZATION ANALYSIS OF GROUP I *CLOSTRIDIUM BOTULINUM* STRAINS WITH DNA MICROARRAYS

Päivi Lahti, Miia Lindström and Panu Somervuo

NB: The protocol has been optimized for use with Agilent 8x15K DNA microarray slides. Using other slide types requires optimization of the protocol.

Day 1

1. DNA

1.1 Dilute genomic DNA in 11.5 µl of sterile distilled water (upH_2O) in an amber tube to make a final concentration of 43.5 ng/µl (500 ng DNA in 11.5 µl upH_2O).

NB: The DNA is diluted in advance to the final concentration.

2. DIRECT LABELLING OF DNA

2.1 Add 10 µl of random primer solution into each DNA sample. Heat at 95 °C for 5 min and cool down rapidly on ice.
2.2 Add the following components and incubate at 37 °C for 2 h:

10x dNTP Nucleotide mix	2.5 µl	1.5 µl	*NB: These have been mixed in advance. The mix is in a tube, which is marked with **"Cy3-mix"** or **"Cy5-mix"**.*
Cy3-dCTP or Cy5-dCTP			
Exo-Klenow fragment	0.5 µl		

NB: Keep all tubes on ice during pipetting.

2.3 Add 2.5 µl stop buffer to inactivate Klenow fragment.

NB: Prewarm Buffer EB to 54 °C.

2.4 Purify the DNA mixtures with QIAquick PCR Purification Kit according to kit instructions. In brief:

- 2.4.1 Add 600 ul Buffer **PB** into sample
- 2.4.2 Pipette sample into QIAquick filter tube.
- 2.4.3 Centrifuge at 13 000 rpm for 30-60 s, discard flow-through (DNA is bound to filter).
- 2.4.4 Add 600 ul Buffer PE and centrifuge at 13 000 rpm for 30-60 s, discard flow-through. Repeat this step once with a final centrifugation at 13 000 for 60 s.
- 2.4.5 Transfer the inner filter tube into a clean amber tube. Pipet 44 µl of prewarmed EB buffer onto the filter, incubate at 54 °C for 1 min. Centrifuge at 13 000 rpm for 1 min.
- 2.4.6 Pipette flow-through back onto filter and incubate at 54 °C for 1 min. Centrifuge at 13 000 rpm for 1 min. NB: This step helps to maximise the purified DNA yield.

2.5 Analyse the quantity and quality of labelled DNA.
2.6 Mix 300 ng of Cy3-labelled DNA and 300 ng of Cy5-labelled DNA and add upH_2O to 17.8 µl.

NB: It is important that equal amounts of Cy3-labelled and of Cy5-labelled DNA are used.
NB: If labelling efficiency was weak, concentrate labelled DNA in a concentrator (or diminish the amount of Cy3- and Cy5-labeled DNA from 300 ng for example to 200 ng).

3. HYBRIDIZATION

3.1 Mix 17.8 µl of labelled and combined DNA with 2.2 µl of salmon sperm DNA (1 mg/ml). Denature at 95 ºC for 2 min. Place on ice.
3.2 Add 5 µl of 10x blocking agent.
3.3 Add 25 µl of 2x GE (HI-RPI) Hybridization Buffer. Mix gently avoiding air bubbles and spin down.
3.4 Place a cover slide in a special metal rack so that the rubber seal is on top. Pipette 45 µl of sample mixture into each well.
3.5 Place the microarray slide on top of the cover slip. Note that the array probes are on the same side as the text Agilent ("Agilent against Agilent"). Close the chamber tightly but avoiding breakage of the glass. The air bubble should move freely in each well.
3.6 Hybridize overnight at 65 °C.

Day 2

4. WASHING THE SLIDES

4.1 Prewarm GE Wash Buffer 2 to 37 °C.
4.2 Open the hybridization rack and remove cover slide by gently rinsing in GE Wash Buffer 1.
4.2 Wash in GE Wash Buffer 1 at RT 1 for 1 min.
4.3 Wash in GE Wash Buffer 2 at 37 ºC for 1 min.
4.4 Wash in 100 % acetonitrile at RT for 1 min.
4.5 Dip into Stabilization and Drying Solution at RT for 30 s.
4.6 Dry slide in a minicentrifuge.
4.7 Place slides in a dark box and scan.

NB: Steps 4.4 and 4.5 are optional.

5. SCANNING, IMAGE ANALYSIS AND DATA ANALYSIS

5.1 Scan the slides immediately to avoid degradation of the Cy-dyes.
5.2 Proceed to image and data analysis according to separate instructions.

Protocol for the detection of *C. botulinum* toxin genes with real-time-PCR

Ute Messelhäusser

This protocol describes examples for a singleplex real-time-PCR method for the detection of the BoNT A gene and a multiplex real-time-PCR method for the detection of the BoNT A and B gene or the BoNT E and F gene in combination with an Internal Amplification control (IAC). The methods are based on the TaqMan® technology with using hydrolysis probes.

1. Equipment
   - real-time-PCR-Cycler (TaqMan®)
   - centrifuge
   - pipettes (volumes between 1 µl and 1000 µl)
   - PCR tubes with a volume of 1.5 ml or 2.0 ml
   - thin walled PCR microtubes (0.2 ml/0.5 ml reaction tubes) or multi-well PCR microplates
   -
2. Reagents
   - commercial MasterMix or Multiplex MasterMix (components: dNTPs, Polymerase, $MgCl_2$ and buffer)
   - primers
   - probes
   - plasmid for the IAC

3. Procedure (PCR-Setup)

3.1 Reaction Mix

Dilute primers and probes in nuclease-free water to a concentration of 10 pmol/µl (*the dilution volume for the lyophilised primers and probes to a concentration of 100 pmol/µl is given in the synthesis report of the saling company*). Dispense the primers and probes in suitable aliquots, in order to minimise thawing and re-freezing, and store the aliquots at -20° C until use.

Table 1. real-time-PCR reaction mix for a singleplex PCR

Components	Example for one PCR reaction (25 µl)	Final concentration
commercial MasterMix	12.5 µl	2 x
primers (according to Annex 1)	1.25 µl each	500 µmol
probe (according to Annex 1)	0.5 µl	200 µmol
nuclease-free water	3.5 µl	
Total	20.0 µl	
extracted DNA	5.0 µl	

Table 2. real-time-PCR reaction mix for a multiplex PCR

Components	Example for one PCR reaction (26 µl)	Final concentration
commercial MasterMix	12.5 µl	2 x
primers (according to Annex 1)	0.75 µl each	300 µmol
probe (according to Annex 1)	0.5 µl	200 µmol
Total	20.0 µl	
plasmid (for the IAC)	1.0 µL	1 fg
extracted DNA	5.0 µl	

Prepare the reaction mix for the number of samples you have (*Prepare always reaction mix for two more samples because of loss of MasterMix , when you pipette it into the vessels of the multi-well microplates or thin walled PCR microtubes. For e.g. when you have 10 samples, prepare reaction mix for 12 samples*) and dispense the reaction mix in aliquots into microtubes or multi-well PCR microplates. To each vessel add extracted DNA, nuclease-free water (negative control) or positive control DNA, respectively. Close the reaction vessels and place them into the thermal cycler (*for closing multi-well microplates use a plastic film, but pay attention not to touch the film on the surface*).

3.2 PCR controls

Negative PCR control:
5 µl DNA-free water without any PCR inhibitors is used as a negative control.

Positive PCR control:
5 µl of extracted BoNT/A and BoNT/B DNA mixture is used as positive control.

Amplification control:
For the IAC use 1 µl (concentration of 1 fg/µl) of the pUC 19-plasmid (*The pUC 19 plasmid is normally used as a ladder for gel-based PCR. You can buy the plasmid from different companies in a lyophilised state. Dissolve the plasmid as it is described in the instruction of the company and dilute it to a concentration of 1 fg/µl. You should always use an Internal or an External Amplification Control to exclude inhibitory effects and false-negative results*).

3.3 PCR amplification

Use the following profile for the real-time-PCR (*The profile may need to be modified depending on the cycler being used. Normally you will use a two-step temperature profile for a TaqMan real-time-PCR, but it is also possible to use a three-step temperature profile with a separate elongation step at 72° C after the annealing step)*.

Table 3. Cycling conditions for the real-time-PCR

Step	Temperature	Time	Cycles
Initial Denaturation	95° C	10 min	
Denaturation	95° C	15 s	45 x
Annealing and Elongation	55° C	60 s	

3.4 PCR results

The results obtained, including the controls, should be unambiguous and the controls should yield expected results, otherwise the procedure must be repeated.

The real-time PCR result will be either

a) positive, if a specific PCR product has been detected and all the controls give expected results, or

b) negative within the limits of detection, if a specific PCR product has not been detected, and all controls give expected results.

(*Sometimes you have to correct the threshold line and set it by yourself considering the results of the controls*)

ANNEX

Table 4. Example for primers and probes for the detection of *C. botulinum* toxin genes

Type	Primer	Sequence (5`-3`)
A	CBOT A fw	TCT TAC GCG AAA TGG TTA TGG
	CBOT A re	TGC CTG CAC CTA AAA GAG GA
	CBOT A S	R -TGG TTT TGA GGA GTC ACT TGA A- Q
B	CBOT B fw	GGA GAA GTG GAG CG**X** AAA A
	CBOT B re	TTC CCT TGA TGC AAA ATG AT
	CBOT B S	R -CCT GGG CCA GTT TTA AAT GA-Q
E	CBOT E fw	TCA GCA CCT GGA CTT TCA GA
	CBOT E re	CAT GTT GTT CTA TAT CAC TTG TTC CA
	CBOT E S	R -TCC AAA ATG ATG CTT ATA TAC CAA AA- Q
F	CBOT F fw	ATA CGG GGC TAG GGG AGT TA
	CBOT F re	AAA TCC TGA CCT CCA AAG GTT
	CBOT F S	R -CCG AAA AAC CCA TAA GGC TA- Q
Internal Amplification Control IAC	IAC_pUC_fw	TGT GAA ATA CCG CAC AGA TG
	IAC_pUC_re	AGC TGG CGT AAT AGC GAA G
	IAC_pUC_S	R -GAG AAA ATA CCG CAT CAG GC- Q

X mixed base: A, G
R Reporter dye, for e.g. FAM, HEX or ROX
Q Quencher dye, for e.g. TAMRA or BHQ 1

Table 5. Estimated time of analysis.

Step	Time
preparing the MasterMix	25 min.
pipetting the PCR-setup	25 min.
preparing the real-time-cycler (temperature profile etc.)	20 min.
real-time-PCR run	105 min. (the real-time cycler can run overnight)
analysis of the results and discussion	30 min.

Instructions for using BioNumerics in AFLP data analysis

The following list is a quick cookbook recipe to create new database, include new AFLP data to existing data and perform cluster analysis for comparing strains. In this example, only some aspects of BioNumerics are shown, for those who are curious to learn more, there is both online help and manual available.

I. Create a new database (by copying an existing reference database)

1. Copy **T:\Bionumerics kurssi ELTDK\clostridiumdb** to **R:**
2. Start BioNumerics from start->Programs->BioNumerics and click Create new database:
   - give name to your new database and click Next
   - define Database directory: **R:\clostridiumdb** and click No for question create the required directories? Click Next and Finish
3. In Setup new database window
   - Select Database type: Local database and click Proceed

II. Add new AFLP data

1. Copy directory **T:\Bionumerics kurssi ELTDK\course AFLP data** to **R:**
2. Import new data (File->Import->Import Fingerprint Files from Automated Sequencers…), click Open Curve Converter and OK which opens Curves window:
   - File->Import curves from file… go to **R:\course AFLP data** and select .fsa files with Ctrl pressed and click Open
3. Change sample order using Ctrl up/down
4. Remove bad/water samples with Delete
5. Save data (File->Export curves…) to **R:\clostridiumdb\images** giving a new file name, select Don't reverse and wait…
6. Close BioNumerics (File->Exit) and the program will add the new data

III. Normalization

1. Open your database for analysis
2. In window Files there should be now new data files, numbers 1-5 correspond to different dyes, in our case 3 is for our sample and 5 is for standard. Double-click file 3:
   -click gel image icon (Edit fingerprint data) and choose ABI AFLP Hind-C, Hpy-A
   -click page Strips and select Lanes->Auto search lanes
   -save
4. In window Files, double-click file 5:
   -click gel image icon (Edit fingerprint data) and choose ABI AFLP Hind-C, Hpy-A:
   -click page Strips and select Lanes->Auto search lanes
   -click page Normalization and select References->use all lanes as reference lanes
   -in Normalization page select Normalization->Auto assign and Normalization->Show normalized view
   -save

5. Link sample (file 3) to standard (file 5):
   -open gel images of file 3 and 5 (if windows closed right click file names in Files window and select ->Open experiment file (data)...)
   -In gel window 3, select File->link to reference gel... and give the name of file 5
   -go to Normalization page of window 3 and click show normalized view
   -save

IV. Link new entries to database

1. In main window, select Database->Add new entries... and give the number of new entries
2. Link AFLP data to new database entries by clicking the arrow icon of data item in Files window and drag it to new entry
3. Fill in the information of new entries

V. Clustering analysis

1. Select the strains you want to include in your analysis by clicking space bar (yellow arrow indicates that entry is selected)
2. Click Comparison->create new comparison
3. In a new window, click gel icon in order to see fragments, use right mouse button to select ->Show image
4. Select Clustering->Calculate->Cluster analysis and choose the method (e.g. Pearson correlation and UPGMA)

Amplified fragment length polymorphism (AFLP) analysis of *Clostridium botulinum* strains (ABI PRISM 310)

Riikka Keto-Timonen and Karen Hill

Day 1

1) DNA

1.1 Dilute genomic DNA in sterile distilled water (upH_2O) to make concentration of 17 ng/ul. Working volume is 24 ul.

2) DIGESTION OF DNA WITH TWO DIFFERENT RESTRICTION ENZYMES

2.1 Prepare digestion mixture (10x):

10 reactions	**1 reaction**
OPA (10x) 30 ul	1x
upH_2O 3 ul	
DTT (1.0 M) 1.5 ul	5 mM
BSA (10 mg/ml) 3 ul	0.1 mg/ml
HpyCH4IV (4-cutter) (10 U/ul) 15 ul	15 U
HindIII (6-cutter)(20U/ul) 7.5 ul	15 U
TOTAL 60 ul	6 ul

2.2 Add 6 ul of digestion mixture to each DNA sample. The volume is now 30 ul.
2.3 Incubate at 37 °C for 1 h 30 min.

NB: Keep tubes and samples on ice while pipetting.
NB: Proceed to next step immediately after digestion. For optimal performance of restriction enzymes during ligation, samples must not be frozen between digestion and ligation.

Aim of this step: Total genomic DNA is digested at specific sites by the restriction enzymes Hind and HpyCH4IV. Hind recognizes 5-AAGCTT-3 (6 nucleotides) and HpyCH4IV recognizes 5-ACGT-3 (4 nucleotides) and cleaves the DNA in these regions creating overhanging ends:

5-A’AGCTT-3 and 5-A’CGT-3
3-TTCGA’A-5 3-TGC’A-5

3) LIGATION OF ADAPTERS TO DNA FRAGMENTS

3.1 Prepare adapters in a thermocycler at 95°C for 5 min and cool down at RT for 10 min (“DENADAP” programme).

NB: dsDNA adapters (=linkers) are constructed from two ssDNA oligonucleotides by hybridization.

3.2 Prepare ligation mixture (10x):

10 reactions	1 reaction
OPA (10x) 7.5 ul	1x
upH_2O 42 ul	
ATP (10 mM) 7.5 ul	200 uM
Hind adapter (2 uM) 7.5 ul	0.04 uM
Hpy adapter (20 uM) 7.5 ul	0.4 uM
DTT (1.0 M) 0.4 ul	5 mM
BSA (10 mg/ml) 0.7 ul	0.1 mg/ml
T4 DNA ligaasi (6 U/ul) 1.9 ul	1.1 U
TOTAL 75.0 ul	7.5 ul

3.3 Add 7.5 ul of ligation mixture to each sample. The volume is now 37.5 ul.
3.4 Incubate at 37 °C for 3 h.

NB: Keep tubes and samples on ice.

Aim of this step: Adapters are short pieces of DNA homologous to the overhanging ends created from the restriction enzyme Hind and Hpy digestion sites. The adapters ligate/attach to the ends of the DNA fragments.

4) PRESELECTIVE PCR

4.1 Dilute each digested and ligated sample in upH_2O (1:1) (e.g. 2 ul + 2 ul) in new 0.2 ml eppendorf tube.
4.2 Make a negative control sample (4 ul of upH_2O).
4.3 Prepare pre-PCR reaction mixture (10x):

10 reactions	1 reaction
Amplification core mix 150 ul	
Hind-0 primer (1 uM) 5 ul	25 nM
Hpy-0 primer (5 uM) 5 ul	125 nM
TOTAL 160 ul	16 ul

4.4 Add 16 ul of pre-PCR mixture into 4 ul of each diluted sample.
4.5 Place tubes in a thermocycler and run on “AFLP-1” programme:

1. Initial denaturation 72 ºC 2 min
2. Denaturation 94 ºC 20 s
3. Annealing 56 ºC 30 s
4. Extension 72 ºC 2 min
5. 4 ºC for ever

Steps 2.-4. are repeated 20 times.

NB: Use block heating in thermocycler. Do not use oil.

NB: Place PCR in ROOM TEMPERATURE PCR machine and start PCR program. Do not hot start as this will melt off one strand of the adapter. Starting at room temperature allows the polymerase to fill in the adapter.

NB: After this step you can load 5-10 ul of the reaction on a 1% agarose gel and look for a light smear between 100-500 bp. Moreover, large lane-to-lane intensity differences may lead to intensity differences in the selective amplification. While the

technique is reportedly robust with respect to DNA concentration differences, large differences should be avoided.

Aim of this step: Primers are short sequences of DNA homologous to the adapters. With the adapters ligated to the ends of the DNA fragments (step 3), the primers are now added to exponentially amplify the DNA fragments in a PCR experiment.

Day 2

5) SELECTIVE PCR

5.1 Dilute pre-PCR-treated samples in upH_2O (1:20) (5 ul + 95 ul).
5.2 Prepare selective PCR mixture (10x):

10 reactions	**1 reaction**
Amplification core mix 75 ul	
FAM Hind-C primer (1 uM) 5 ul	50 nM
Hpy-A primer (5 uM) 5 ul	250 nM
TOTAL 85 ul	8.5 ul

5.3 Add 8.5 ul of selective PCR mixture to 1.5 ul of each diluted sample.
5.3 Place tubes in a thermocycler and run on "AFLP-2" programme:

1. Initial denaturation 94 ºC 2 min
2. Denaturation 94 ºC 20 s
3. Annealing 66→56 ºC (first 10 cycles)/ 56 ºC (next 19 cycles) 30 s
4. Extension 72 ºC 2 min
5. Final extension 60 ºC 30 min
6. 4 ºC for ever

Steps 2.-4. are repeated 29 times. The first 10 cycles the annealing temperature is decreased by 1 ºC at each cycle until it is 56 ºC. This is followed by 19 cycles with annealing temperature of 56 ºC.

NB: Use block heating in thermocycler. Do not use oil.

Aim of this step: A second PCR amplification is performed using primer sequences with one additional nucleotide added to the end (Hind-primer+C, Hpy-primer+A). Under high stringency PCR conditions, only those fragments will be amplified that contain that nucleotide next to the adapter. Also Hind-primer+C has a fluorescent dye called FAM attached so that the amplified fragments can be visualized by the laser in the ABI 310 instrument.

6) SAMPLE PREPARATION FOR ELECTROPHORETIC SEPARATION OF FRAGMENTS (ABI PRISM 310)

6.1 Set up ABI 310 with a matrix that will recognize FAM and LIZ dyes.
6.2 Use POP-4 polymer in syringe.
6.3 Prepare a sample sheet of sample names.
6.4 Use local Southern sizing method.
6.5 Prepare loading mixture (~10x):

10 samples	1 sample
HIDI formamide 120 ul	
LIZ Molecular Weight Standard 4 ul	
TOTAL 124 ul	12 ul

6.6 Pipette 12 ul of loading mixture into fresh tube (ABI strips).
6.7 Add 1 ul of selective PCR product (=sample) into each tube.
6.8 Place caps on strips and denature at 95 °C for 2 min ("ABIDEN" programme), cool down at 4 °C.
6.9 Start run (15 kV, 60 °C, 28 min). The running settings are

Inj. secs	5
Inj. kV	15.0
Run kV	15.0
Run °C	60
Run time	28
Module	GS STR POP4 (1 ml) G5v2.md5
Matrix file	GS STR POP4 G5.mtx

Analysis parameters	GS500LIZ (-35-250) test.gsp
Size standard	GS 500LIZ (-25-250) test.szs

NB: Do steps 6.5-6.7 in a laminar flow hood.
NB: Optimal analysis settings and running conditions depend on the analysis performed and the device used, thus optimization and consultation of the device manual are required.
NB: Always wear gloves when handling the ABI machine.

Day 3

7) DATA HANDLING AND ANALYSIS WITH BIONUMERICS

The following list is a quick cookbook recipe to create new database, include new AFLP data to existing data and perform cluster analysis for comparing strains. In this example, only some aspects of BioNumerics are shown, for those who are curious to learn more, there is both online help and manual available.

7.1 Create a new database (by copying an existing reference database)

7.1.1 Copy **T:\Bionumerics kurssi ELTDK\clostridiumdb** to **R:**
7.1.2 Start BioNumerics from start->Programs->BioNumerics and click Create new database:
- give name to your new database and click Next
- define Database directory: **R:\clostridiumdb** and click No for question create the required directories? Click Next and Finish

7.1.3 In Setup new database window
- Select Database type: Local database and click Proceed

7.2 Add new AFLP data

7.2.1 Copy directory **T:\Bionumerics kurssi ELTDK\course AFLP data** to **R:**

7.2.2 Import new data (File->Import->Import Fingerprint Files from Automated Sequencers…), click Open Curve Converter and OK which opens Curves window:
- File->Import curves from file… go to **R:\course AFLP data** and select .fsa files with Ctrl pressed and click Open
7.2.3 Change sample order using Ctrl up/down
7.2.4 Remove bad/water samples with Delete
7.2.5 Save data (File->Export curves…) to **R:\clostridiumdb\images** giving a new file name, select Don't reverse and wait…
7.2.6 Close BioNumerics (File->Exit) and the program will add the new data

7.3 Normalization

7.3.1 Open your database for analysis
7.3.2 In window Files there should be now new data files, numbers 1-5 correspond to different dyes, in our case 3 is for our sample and 5 is for standard. Double-click file 3:
-click gel image icon (Edit fingerprint data) and choose ABI AFLP Hind-C, Hpy-A
-click page Strips and select Lanes->Auto search lanes
-save
7.3.3 In window Files, double-click file 5:
-click gel image icon (Edit fingerprint data) and choose ABI AFLP Hind-C, Hpy-A:
-click page Strips and select Lanes->Auto search lanes
-click page Normalization and select References->use all lanes as reference lanes
-in Normalization page select Normalization->Auto assign and Normalization->Show
normalized view
-save
7.3.4 Link sample (file 3) to standard (file 5):
-open gel images of file 3 and 5 (if windows closed right click file names in Files window and select ->Open experiment file (data)...)
-In gel window 3, select File->link to reference gel... and give the name of file 5
-go to Normalization page of window 3 and click show normalized view
-save

7.4 Link new entries to database

7.4.1 In main window, select Database->Add new entries... and give the number of new entries
7.4.2 Link AFLP data to new database entries by clicking the arrow icon of data item in Files window and drag it to new entry
7.4.3 Fill in the information of new entries

7.5 Clustering analysis

7.5.1 Select the strains you want to include in your analysis by clicking space bar (yellow arrow indicates that entry is selected)
7.5.2 Click Comparison->create new comparison
7.5.3 In a new window, click gel icon in order to see fragments, use right mouse button to select ->Show image
7.5.4 Select Clustering->Calculate->Cluster analysis and choose the method (e.g. Pearson correlation and UPGMA)

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

MARIE CURIE ACTIONS

Microarray Data Analysis

Panu Somervuo

I. R/Bioconductor with limma package

During this exercise, we use software package R for data analysis. R is a programming language which offers statistical tools and graphics for general data analysis. The basic set of R functions can be augmented by additional packages. Bioconductor project provides such packages for bioinformatics related tasks including microarray data analysis. Although some people may find it discouraging that in R commands must be given from keyboard instead of clicking buttons on graphical user interface, R/Bioconductor gives the flexibility which is difficult/impossible to find in any other microarray data analysis program. Because R is free, you can download it also in your home computer. More information about R and Bioconductor:

http://www.r-project.org
http://www.bioconductor.org

Online help

You can access online help and introduction to R from Help menu or using keyboard (note '>' is R prompt which you don't type), e.g.

```
> help(plot)
```

gives information about plot command.

Preliminaries:

1. Copy directory T:\Bionumerics kurssi ELTDK\madata to R:\
2. Start R from start->Programs->R
3. Select limma package from Packages->Load package...
4. Click File->Change dir... and change working directory to R:\

In the following, copy commands after R prompt '>' by writing or using copy and paste from this text document, after each command press Enter

Data import

1. Read file coursemainfo.txt which contains information about hybridizations and store it to variable hyb, read.table is a command to which we give the following parameters: name of file to read, whether the first line of file is a header (TRUE/FALSE), and field delimiter ("\t" stands for tabulator)

```
> hyb <-
read.table("madata/coursemainfo.txt",header=T,sep="\t")
```

2. By writing the name of variable, we can see what it contains:

```
> hyb
```

There are 16 rows and three columns, each row corresponds to one output file from GenePix image analysis program (column file contains the name of .gpr file). Columns cy3 and cy5 tell which samples have been labeled with these dyes.

3. Read the contents of .gpr files and store them to variable rg (we specify that the files are genepix format and in addition to default columns we read also the column "Flags" which contains the spot quality information), parameter path specifies the data directory:

```
> rg <- read.maimages(hyb$file, source="genepix",
other.columns="Flags", path="madata")
```

Among other data fields, rg contains four data matrices: R for Red (cy5) foreground, Rb for Red background, G for Green (cy3) foreground, and Gb for Green background

4. Let's subtract local background intensity from foreground intensity of each spot and store the result in new variable rgb

```
> rgb <- backgroundCorrect(rg,method="half")
```

Visualizing spot intensities

1. Let's plot cy3 signals versus cy5 signals of the first hybridization:

```
> plot(rgb$G[,1],rgb$R[,1])
```

This shows cy5 signal of hybridization 7 vs cy5 of hybridization 12:

```
> plot(rgb$R[,7],rgb$R[,12])
```

The same using intensities in log2-domain:

```
> plot(log2(rgb$R[,7]),log2(rgb$R[,12]))
```

Let's find the indices of spots corresponding to gene ATP-dependent RNA helicase deaD (probe names pCBN_1830_1358 and pCBN_1830_1289):

```
> ind <- grep("pCBN_1830_",rgb$genes$ID)
```

We can highlight the locations of these probes in the previous scatterplot using command:

```
> points(log2(rgb$R[ind,7]),log2(rgb$R[ind,12]),
pch=19,col="red",cex=2)
```

2. Spread of intensity values can be visualized in two ways:

```
> boxplot(log2(rgb$R[,7]))
> plot(density(log2(rgb$R[,7])))
```

Boxplots from all hybridizations can be visualized together (first we create data matrix x which contains both cy5 and cy3 intensities):

```
> x <- log2(cbind(rgb$R, rgb$G))
> colnames(x) <- c(hyb$cy5, hyb$cy3)
> boxplot(x~cols(x),las=2)
```

Outlier removal

1. Let's assume bad hybridizations are in columns 4, 5, and 20 of data matrix x. We can remove them (storing all other columns in variable x2) using command:

```
> x2 <- x[,-c(4,5,20)]
```

Normalization

1. Intensity distributions of all hybridizations can be normalized to be the same using quantile normalization:

```
> qx2 <- normalizeQuantiles(x2)
```

We can visualize the difference:

```
> boxplot(x2~cols(x2), las=2 ,main="Before normalization")
> boxplot(qx2~cols(qx2), las=2, main="After normalization")
```

Gene present/absent threshold

Let's investigate the distribution of normalized data

```
> plot(density(qx2[,1]))
```

When two peaks are visible, we can assume that one of them corresponds to probes with specific and the other to unspecific hybridization. Let's assume log2-intensity value 9 falls between these peaks, we can add this threshold to previous image:

```
> abline(v=9, col="blue")
```

Let's create new data matrix with value 1 for those items with intensity value greater than 9, and 0 otherwise:

```
> dat01 <- matrix(0,nr=nrow(qx2),nc=ncol(qx2))
> i <- which(qx2 > 9)
> dat01[i] = 1
```

Exporting data

Let's save both normalized log2-intensity data and 0-1 data to text file (we add also contents of "probeifno.txt" to these output files):

```
> probeinfo <-
read.table("probeinfo.txt",header=T,sep="\t")
> a <- data.frame(probe=rgb$genes$ID, info=probeinfo,
x=qx2)
> write.table(a,"madatlog2qnorm.txt",row.names=F,sep="\t")
> b <- data.frame(probe=rgb$genes$ID, info=probeinfo,
x=dat01)
> write.table(b,"madat01.txt",row.names=F,sep="\t")
```

R offers many tools for statistical tests, clustering and visualization. In this exercise, however, we do those tasks with MeV program.

II. MultiExperimentViewer MeV

MeV is Java based microarray data analysis program with graphical user interface. It offers data clustering, visualization, and statistical tests. MeV can be downloaded from http://www.tm4.org

1. Start MeV double-clicking T:\Bionumerics kurssi ELTDK\Shortcut to TMEV.bat

Data import and visualization

1. In MultipleArrayViewer window click File->Load Data which opens Expression File Loader window:
   - Select File Loader -> Tab Delimited, Multiple Sample Files
   - Select expression data file (Browse R:\madatlog2qnorm.txt)

- click Single-color Array
- click the upper-leftmost intensity value and Load button

2. Color scale and element sizes can be changed from Display menu

Strain clustering

1. Select Analysis->Clustering->HCL:
   unselect Gene Tree and select only Sample Tree
2. Click HCL in left panel and then HCL Tree

In order to compare these results with 0-1 data, open new window:

3. In MultiExperiment Viewer window, click File->New Multiple Array Viewer
   -File->Load Data
   -Select expression data file (Browse R:\madat01.txt)
4. Repeat the steps above for obtaining hierarchical clustering

Finding genes which distinguish strains

We can apply t-tests for all probes of array. Probes with small p-values indicate differences between the strains.

1. Analysis->Statistics->TTEST, select Between subjects
   -choose two strains at a time, select all hybridizations of the first strain to belong to Group A, all hybridizations of the other strain to Group B, and all other hybridizations to Neither group
   -click OK
2. Click the results from left panel T Tests