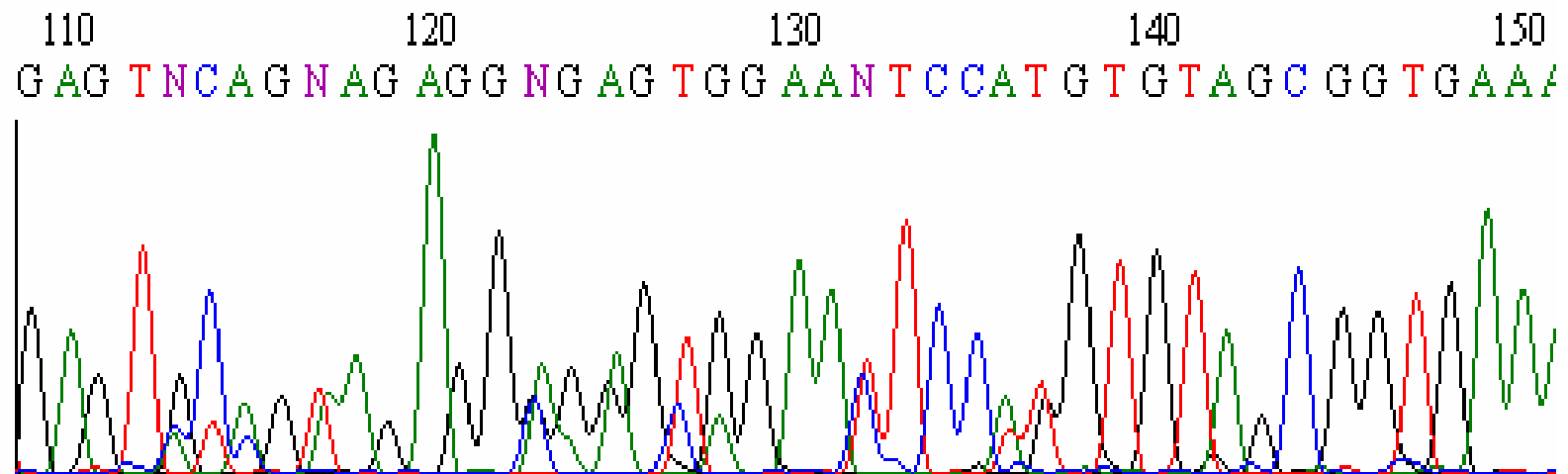


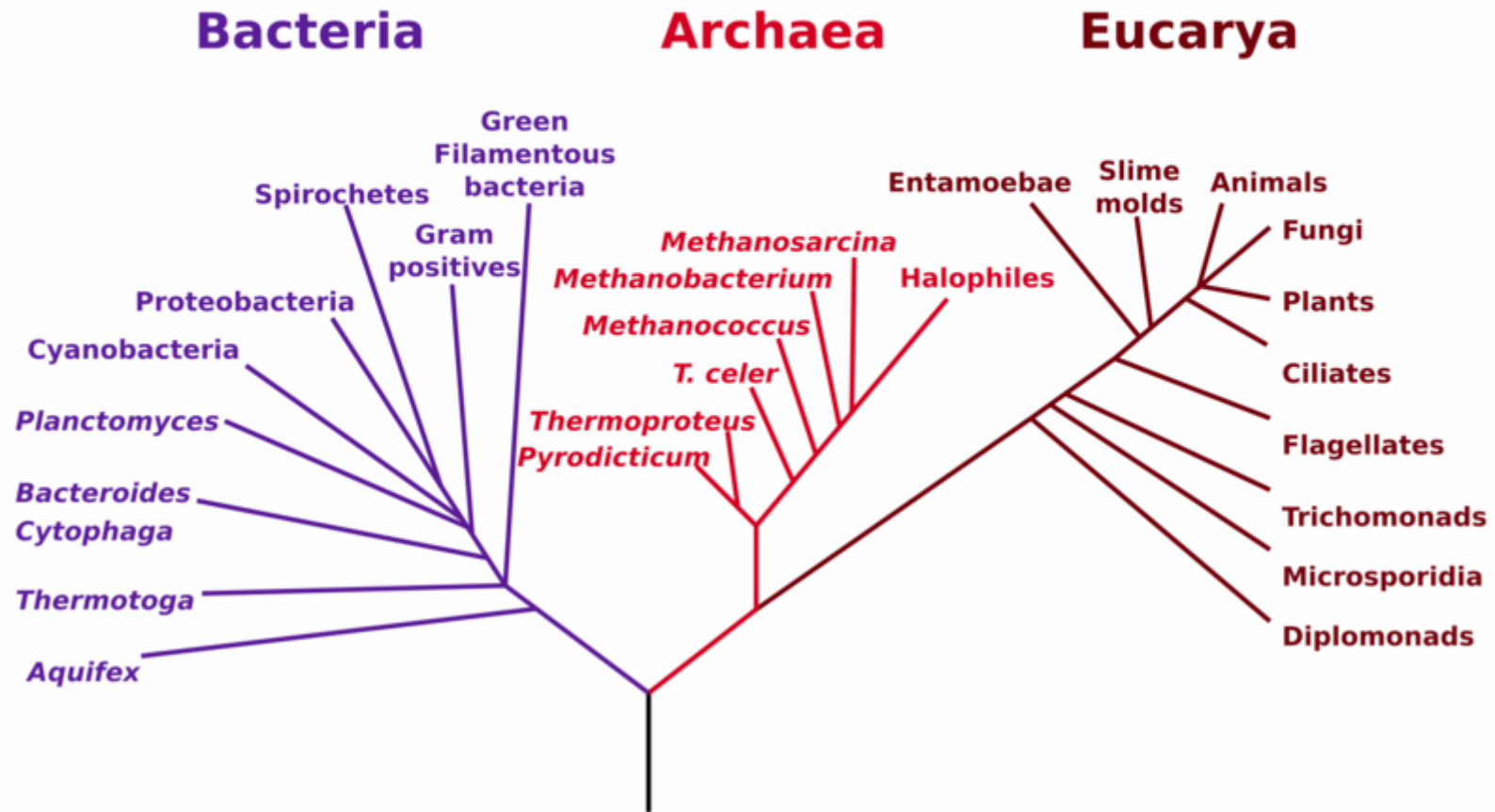
Multivariat kurveoppløsning av blandede DNA sekvensspektra



Bakteriesamfunn

- Kun en liten del av den totale bakteriediversitet har blitt isolert som axenisk kultur (<5%)
- Mange bakteriearter er kjent kun fra sekvenser av miljøkloner
- Den vanligste markøren for denne type analyse er genet for 16S rRNA
- Genet er konserverte i alle bakteriegrupper
- Genet består av variable og konserverte områder, slik at det er egnet for analyse av fylogeni og diversitet

Phylogenetic Tree of Life



Standardmetoder for bakteriesamfunnsanalyse

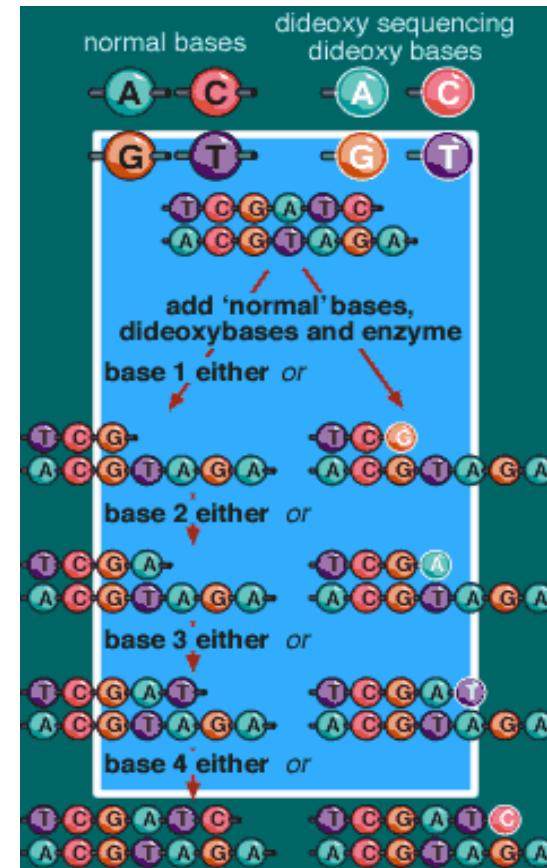
- Fingerprinting av populasjoner: tRFLP, DGGE, TGGE
- Klonbiblioteker
- Kulturbaserte metoder
- (DNA microarrays)
- Disse metodene har begrensninger når det gjelder oppløsning, kvantitative egenskaper, tidsforbruk og kostnader

Målsetninger

- Utvikle en relativt hurtig og billig metode for kvantitativ analyse av blandede bakteriepopulasjoner
- Følge vekstprosesser gjennom tidsserier og gi en biologisk tolkning av data
- Bruke spektra produsert av sekvenseringsapparatet GA3100 (ABI) for måling av sammensetning i blandede 16S rDNA-prøver

DNA sekvensering

- Vanligvis utført v.h.a. Sanger-metoden (kjedeterminering)
- DNA-fragmentet som skal sekvenseres amplifiseres i sykler med en blanding av dNTPer og ddNTPer
- ddNTPene er merket med fluorokromer for deteksjon av fragmenter
- Reaksjonen produserer en blanding av attenuerte fragmenter i alle tenkelige lengder



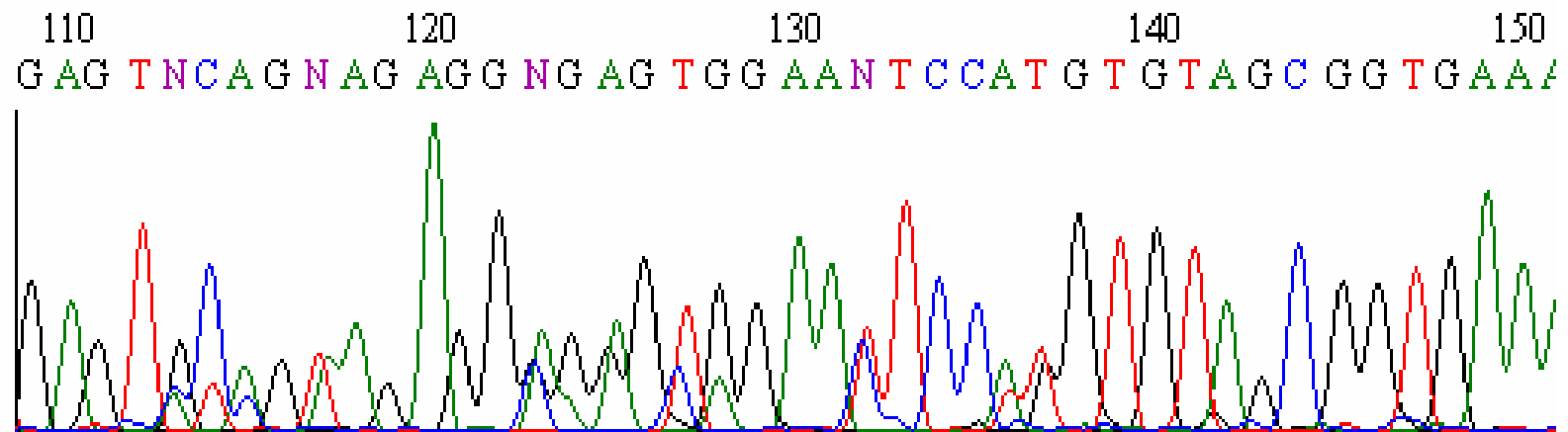
GA3100



You can place two 96- or 384-well plates on the autosampler at once. When the sample plate is in place, you can automatically inject, run, and analyze samples according to user-defined analysis parameters.

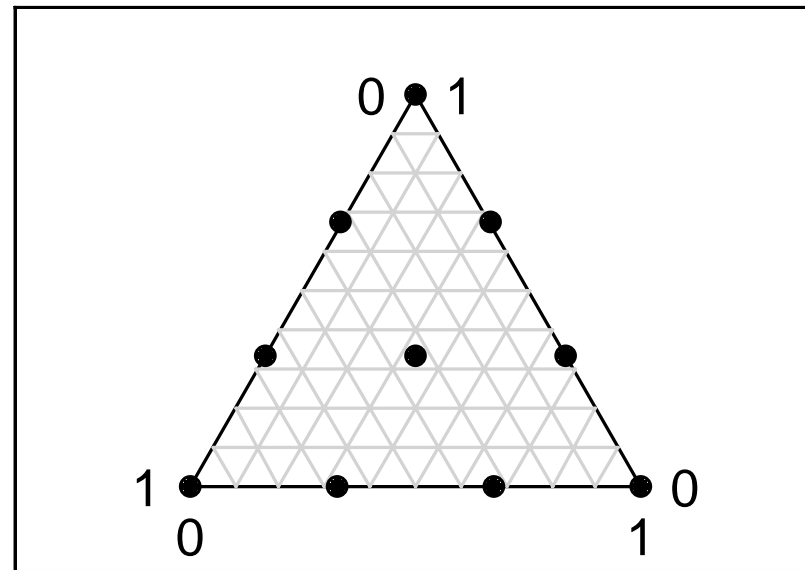
- Merkede DNA-fragmenter lastes på apparatet i arrayplater
- De separeres v.h.a. kapillærelektroforese ved høy spenning i smale silikonrør fylt med en separasjonsmatrix
- Fragmentenes vandringshastighet mot anoden er omvendt proporsjonal med fragmentenes størrelse
- I enden av kapillæret sitter en argonlaser, en serie med speil, et prisme og et CCD-kamera
- De fire fluorokromene som ble brukt ved innmerking emitterer ved ulike bølgelengder som igjen gir ulik brytning i prismet
- Når merkede fragmenter passerer laseren eksiteres fluorokromene og den emitterte energien registreres av CCD kameraet parallelt i fire kanaler

Blandet DNA sekvensspektrum



Analyse

- Kalibrering av kvantitativ respons i GA3100 ved Partial Least Squares (PLS) regresjon
- Kalibrering ble utført for 7 arter valgt fra et kyllingcecum-isolat klonbibliotek
- Kalibreringsdata besto av 80 blandinger og de 7 rensekvensene i et simplex lattice design
- Blandingene ble sekvensert og numeriske data tilsvarende 52 baseposisjoner (~2600 fluorescensmålinger) ble ekstrahert fra hvert spektrum



Preprosessering

- Små ulikheter i molekylære migrasjonsrater forårsaker ujevn spektrallengde
- Spektra justeres til identisk lengde ved kubisk spline interpolering
- Hvert spektrum består av 4 parallelle serier
- Før regresjonsanalyse konkatneres spektrene
- Hvert spektrum normaliseres etter gjennomsnitt for å gjøre modellen robust mot konsentrasjonsvariasjoner og ulik effektivitet i innmerking

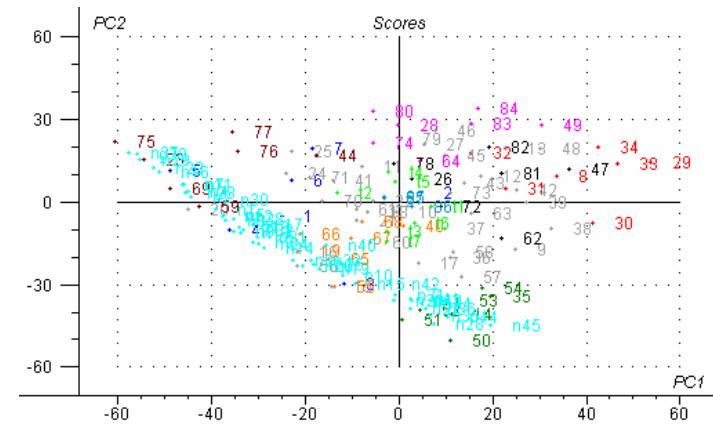
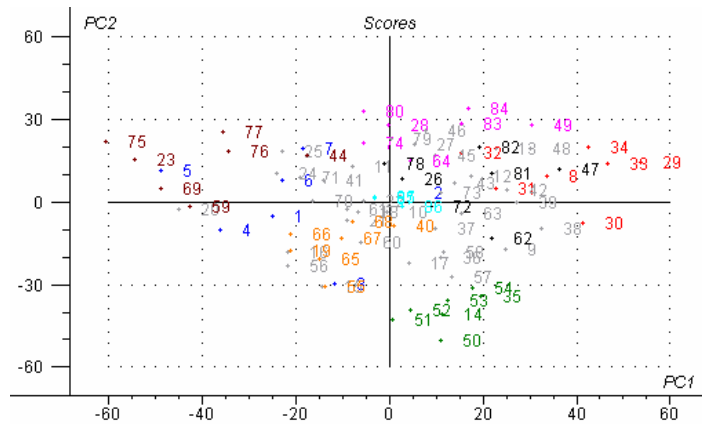
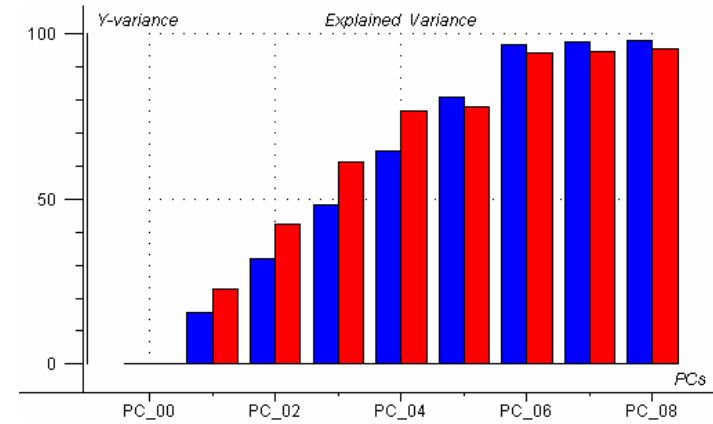
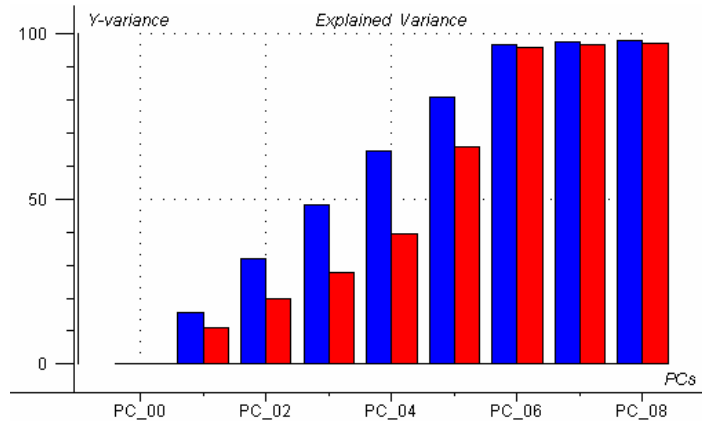
PLS regresjon

- En dekomposisjon av spektralmatrisen \mathbf{X} : $\mathbf{X}=\mathbf{TP}'+\mathbf{E}$
- Responsmatrisen kan da modelleres som: $\mathbf{Y}=\mathbf{TQ}'+\mathbf{F}$
- Den første vektor av scores, \mathbf{T} , finnes som matriseproduktet $\mathbf{X}\mathbf{w}_1$, hvor \mathbf{w}_1 er en vektor av loading weights, i.e. den første egenvektoren til matriseproduktet $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}$
- Den første loading, \mathbf{p}_1 , finnes ved regresjon av \mathbf{X} mot \mathbf{t}_1
- Neste loading weight, \mathbf{w}_2 , finnes ved den samme egenvektordekomposisjon, men med \mathbf{X} og \mathbf{Y} deflatert ved subtraksjon av hhv. $\mathbf{t}_1\mathbf{p}_1'$ og $\mathbf{t}_1\mathbf{q}_1'$, hvor \mathbf{t}_1 er første vektor av scores, \mathbf{p}_1 er første vektor av \mathbf{X} -loadings og \mathbf{q}_1 er første vektor av \mathbf{Y} -loadings.
- Nye scores og loadings beregnes, og denne prosessen gjentas til det ønskede antall komponenter er funnet
- Matrisen av regresjonskoeffisienter, \mathbf{B} , for de endelig lineære prediktorfunksjonene finnes som løsning på ligningen $\mathbf{B}=\mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}\mathbf{Q}$

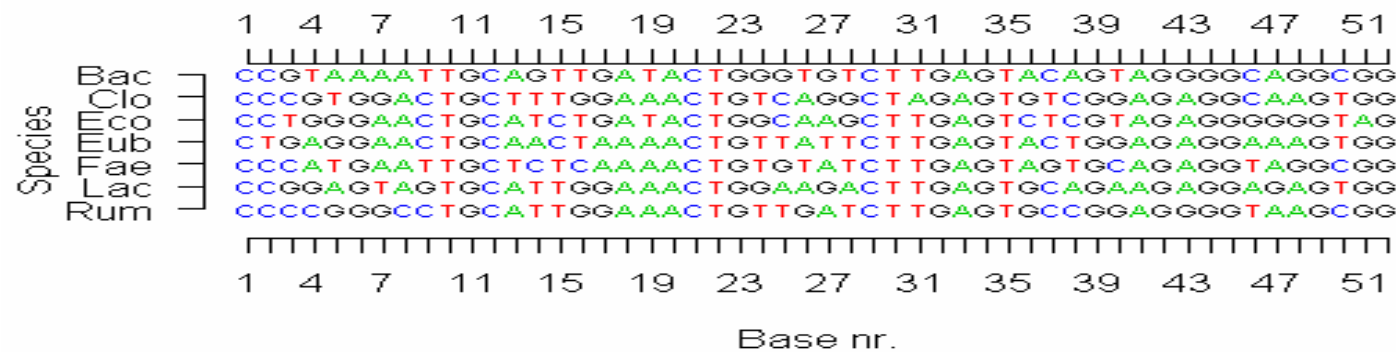
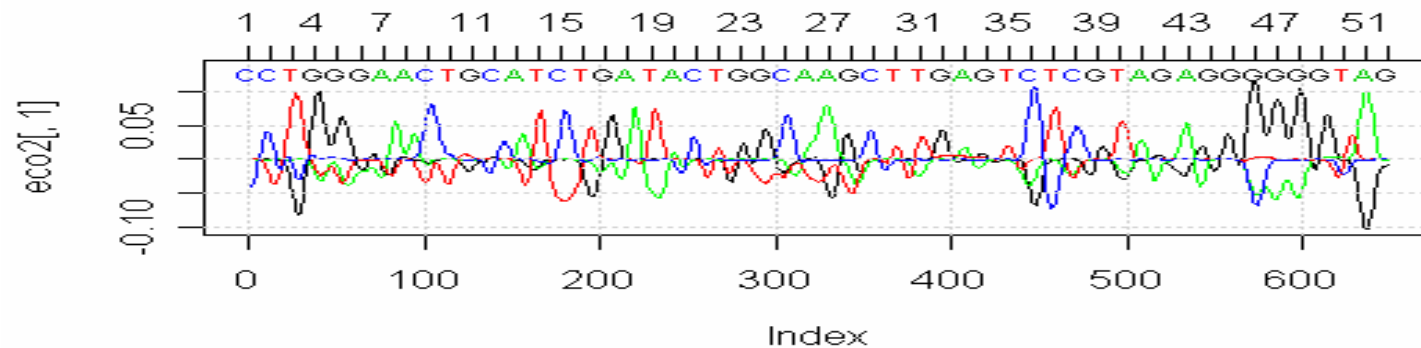
Modellering

- PLS2 regressjon av responsmatrise mot spektralmatrise
- Responsmatrise har dimensjon (87,7)
- Spektralmatrise har dimensjon (87,2596)
- Resultatet er 7 lineære prediktorfunksjoner av typen: $y_i = x_{i1}b_{i1} + \dots + x_{i2596}b_{i2596}$
- Kryssvalidering
- Testsett-validering

Modell

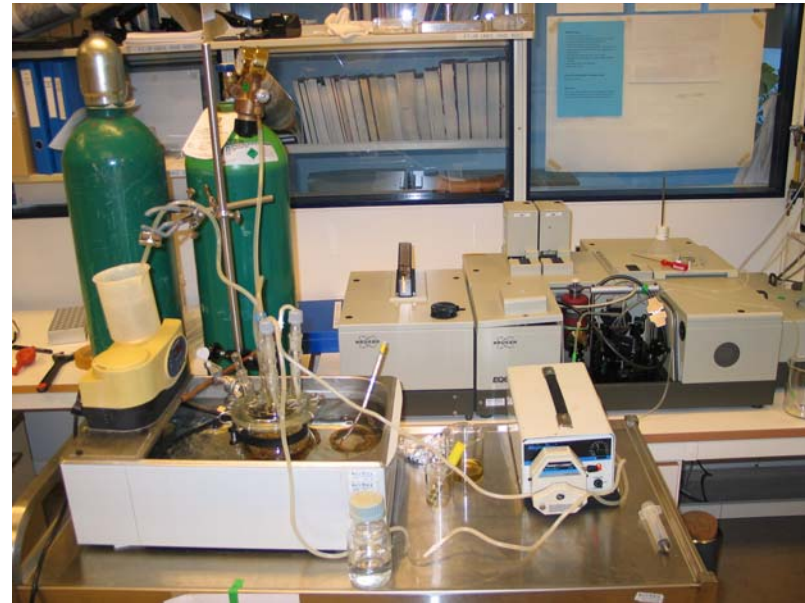


Regresjonskoeffisienter og alignment

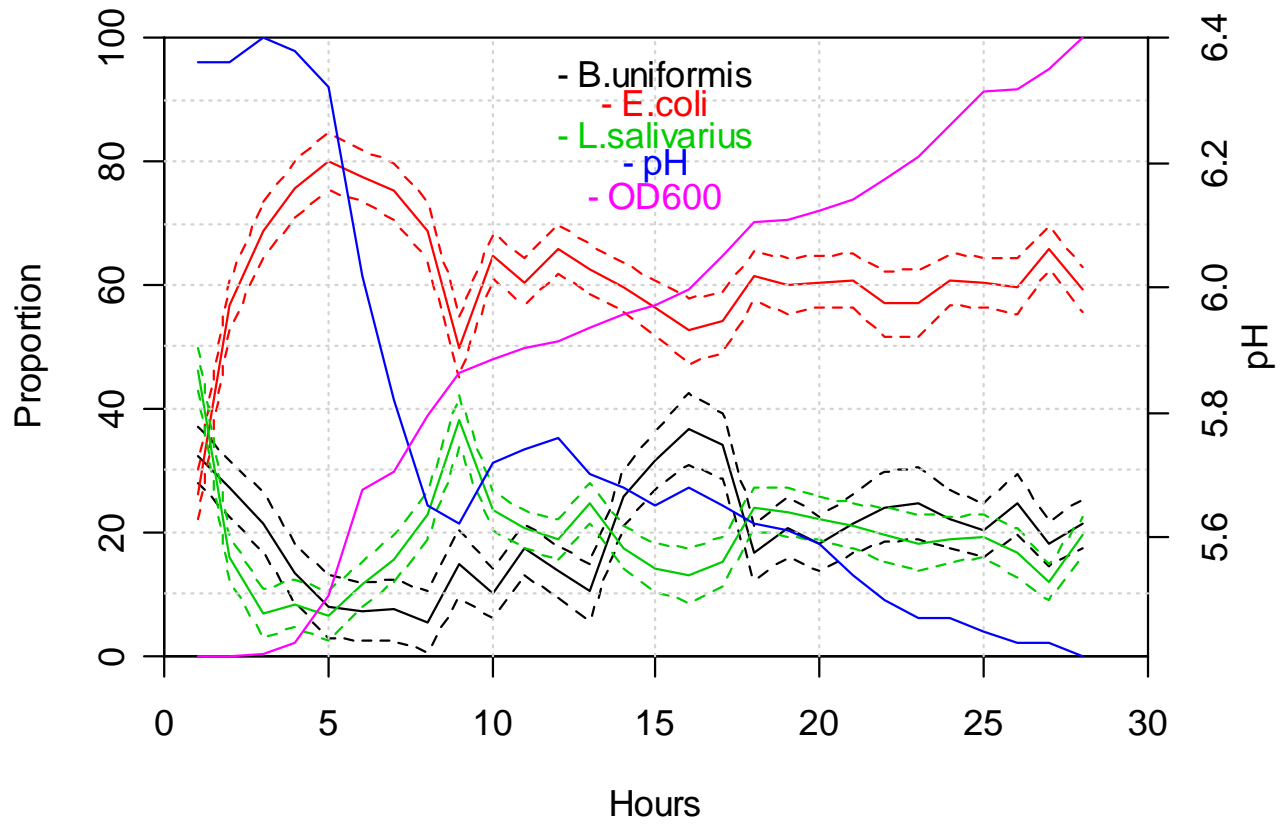


Et eksperiment

- *E.coli*, *L.salivarius* og *B.uniformis* dyrkes sammen i en anaerob fermentor
- Celleprøver tas hver time
- DNA ekstraheres og sekvenseres
- Spektraldata analyseres v.h.a. regresjonsmodell



Resultat

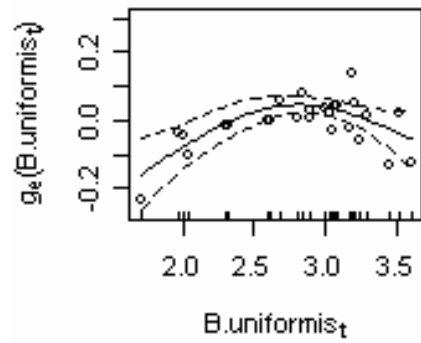
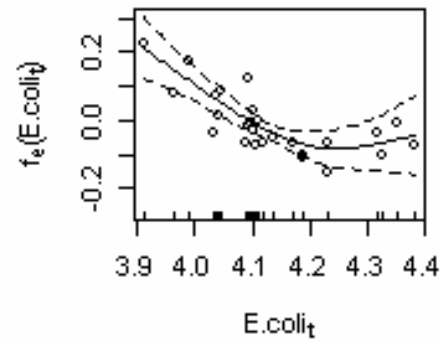
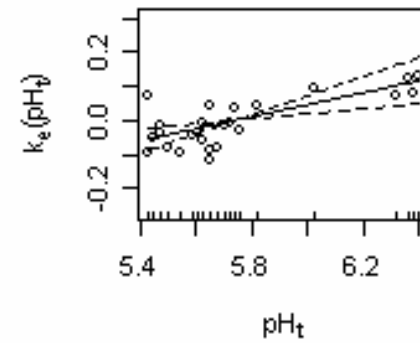
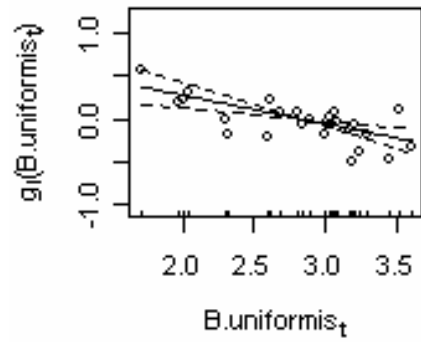
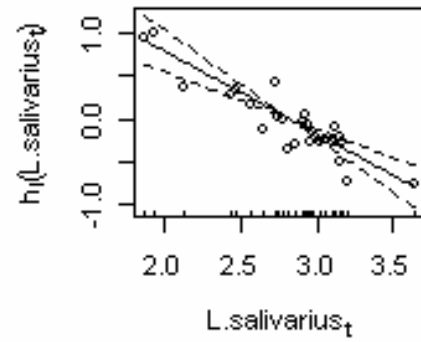
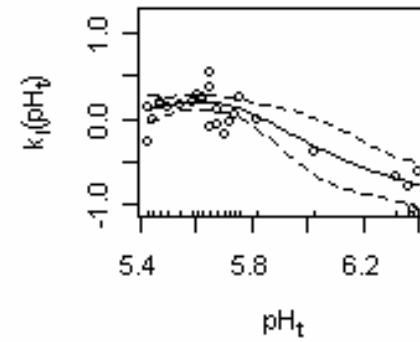
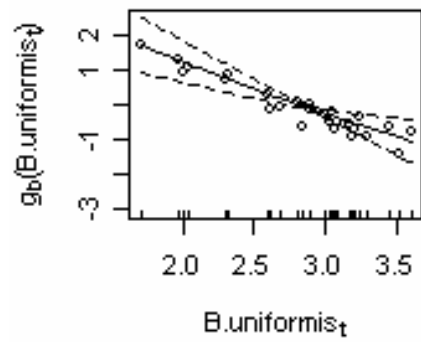
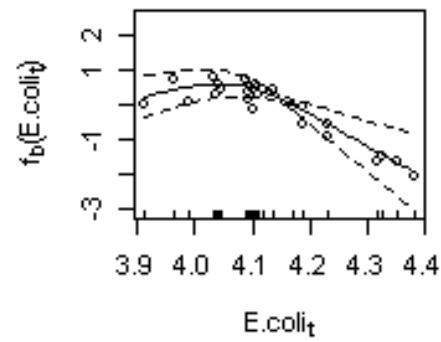


Tidsserieanalyse

- Tidsseriene fra forrige slide brukes til økologisk modellering
- For hver tidsserie tilpasses en Generalisert Additiv Modell (GAM) for å finne sammenheng mellom én arts vekst og de andre variablene
- GAM tilpasser glatte additive ikke-parametriske funksjoner til hver kovariat slik at hver funksjon er en vektet sum av basis cubic spline funksjoner
- Den generelle modellen i tilfellet *E.coli*-vekst er som følger:

$$\Delta E_t = b_e + f_e(E_t) + g_e(B_t) + h_e(L_t) + k_e(\text{pH}_t) + \varepsilon_t$$

hvor ΔE_t er endring i relativ mengde *E.coli* per tidsenhet beregnet som differansen mellom to logaritmiske mengder målt i tilsluttende tidsintervaller (i.e. $\log(E_{t+1}) - \log(E_t)$). f_e , g_e , h_e og k_e er ikke-parametriske glatte funksjoner som estimerer effekt av hhv. *E.coli* (i.e. selvregulering), *B.uniformis*, *L.salivarius* og pH. b_e er intercept og ε_t er et støyledd.

E.coli growth**E.coli growth****E.coli growth****L.salivarius growth****L.salivarius growth****L.salivarius growth****B.uniformis growth****B.uniformis growth****B.uniformis growth**