



Aalto-yliopisto  
Perustieteiden  
korkeakoulu

# *Trichoderma reesei* geenisäätelyverkoston ennustaminen

*Oskari Vinko*

*04.11.2013*

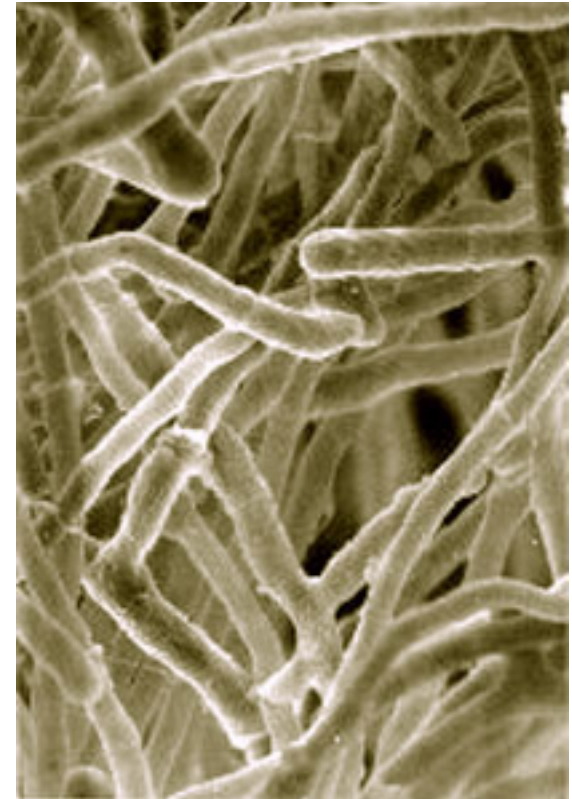
Ohjaaja: *Merja Oja*

Valvoja: *Harri Ehtamo*

Työn saa tallentaa ja julkistaa Aalto-yliopiston avoimilla verkkosivuilla. Muilta osin kaikki oikeudet pidätetään.

# Taustaa

- Sellulaasientsyymien hinnalla on suuri vaikutus biopolttoaineiden tuotantokustannuksiin
- *T. reesei* on teollisuudessa yleisimmin käytetty sellulaasituottaja
- Aikaisempi tutkimus keskittyi ylituottajakantojen kehittämiseen
- Tuoreempi tutkimus pyrkii:
  - Ymmärtämään sellulaasigeenien säätelymekanismeja
  - Tutkimaan ylituottajakantojen hyödyllisten mutaatioiden vaikutusmekanismia



*Trichoderma reesei*

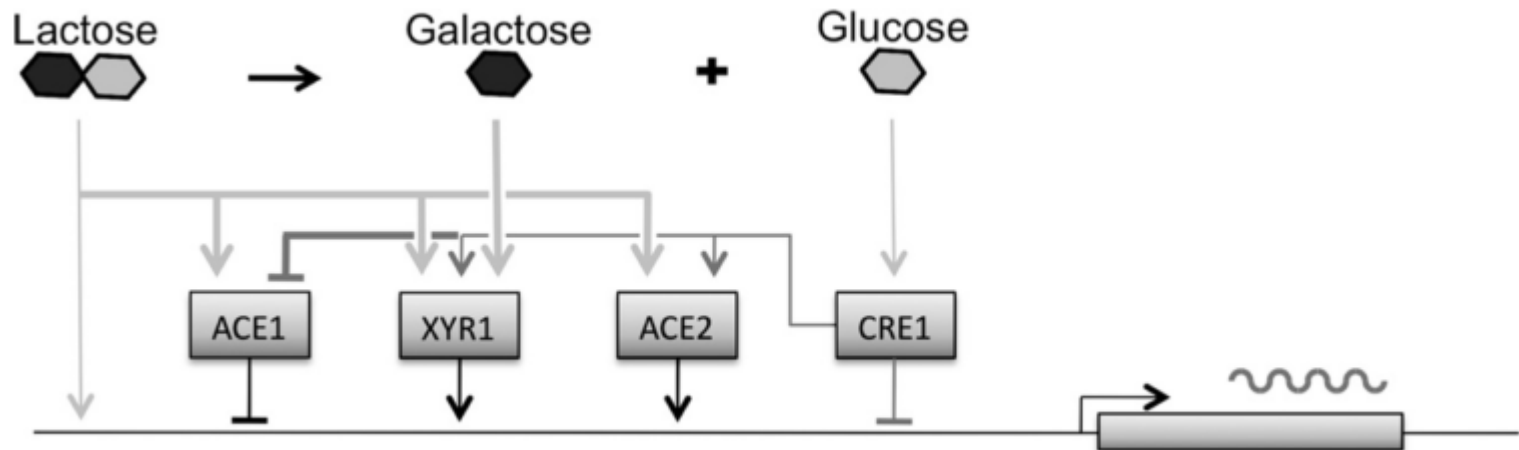
# Taustaa

- Tunnettuja positiivisia säätelijöitä:

- Xyr1
- Ace2
- HAP kompleksi
- Lae1
- Xyl1p

- Tunnettuja negatiivisia säätelijöitä :

- Cre1
- Ace1



Portnoy et al., 2011

Cbh1 ...

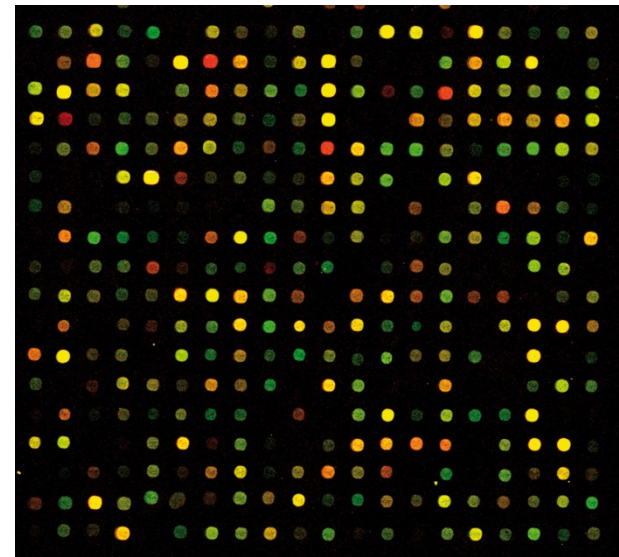
# Tavoite

- Uusien, testattavien säätelijäehdokkaiden muodostus saatavissa olevan transkriptiodatan perusteella
  - Säätelijäehdokkaiden todellinen vaikutus voidaan varmentaa laboriokokeilla
  - Erityisen kiinnostavia geenejä ovat sellulaasisynteesin ja – erityksen säätelijät
- Module networks –menetelmän soveltaminen ja sen arviointi *T. reesein* geeniverkoston säätelyn ennustamisessa

# Transkriptomi

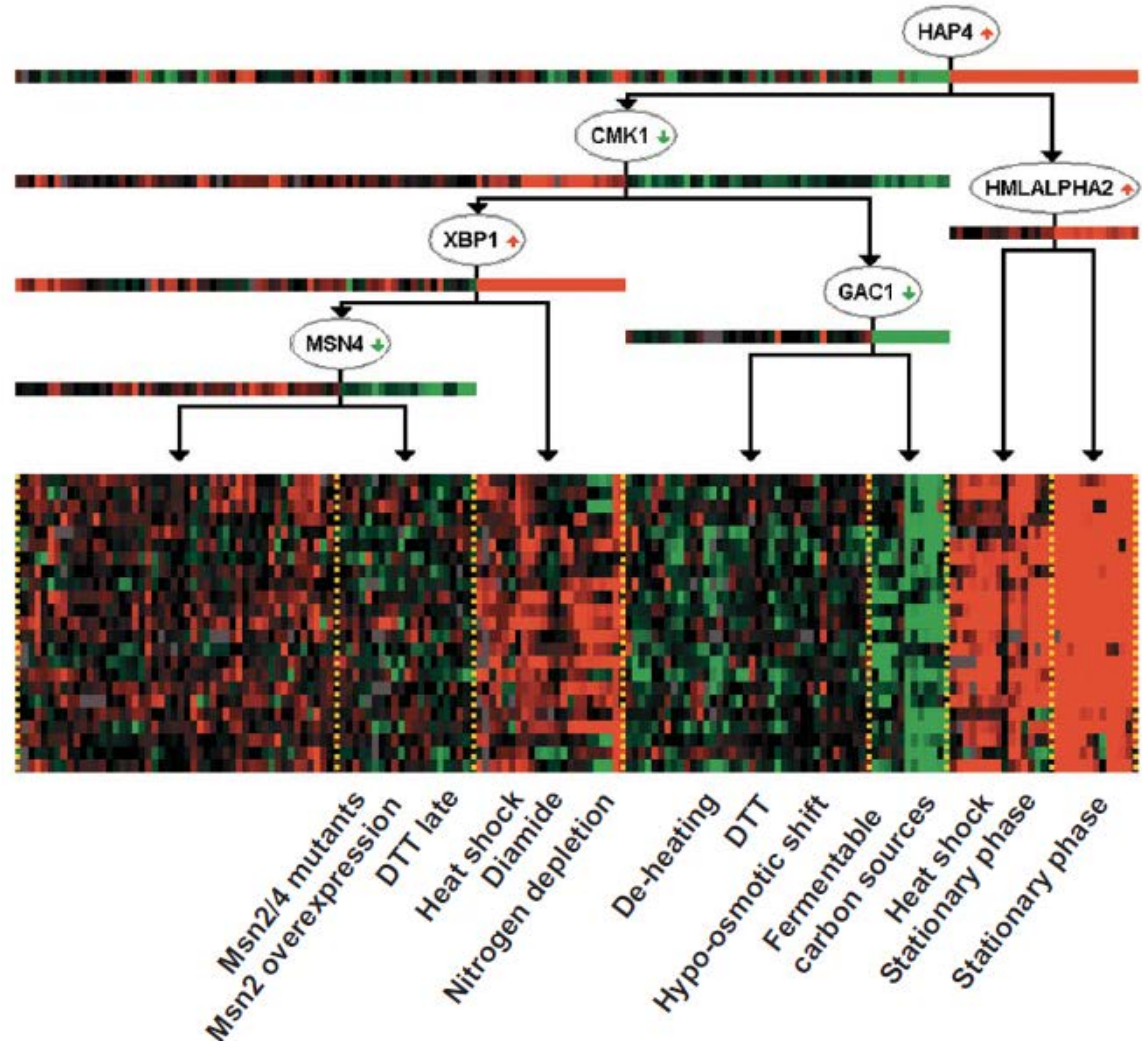
- Kaikkien geenien aktiivisuus yhdellä ajanhetkellä yhdessä organismissa tai populaatiossa
- Tässä tutkielmassa käytetty transkriptomidata oli sarja mittauksia eri olosuhteissa
  - Yhteensä reilut 10 000 geeniä
  - Sarja A sisälsi 66 olosuhdevertailua
    - pH, lämpötila, hiilenlähde yms.
  - Sarja B sisälsi 49 aikavertailua

$$\log_2\left(\frac{A_x}{A_{kontrolli}}\right)$$



# Menetelmä

- Module networks – menetelmä
  - Geenit jaetaan moduleihin, joilla on yksilöllinen säätelyohjelma
  - Geenien aktiivisuus normaalijakautunutta kunkin lehden sisällä
- EM-algoritmia käytettiin modulirakenteen luomiseen
  - Kaksi askelta:
    - Säätelyohjelman muodostaminen
    - Rakenteen muodostaminen
- Algoritmi on toteutettu Genomica-ohjelmassa



# Ongelma

- EM-algoritmissa on bayesiläinen tavoitefunktio
  - Lukuisia muuttujia, joiden vaikutuksesta tavoitefunktioilla on useita paikallisia ääriarvokohtia
    - Tällä hetkellä ei ole olemassa tapaa löytää funktion globaali maksimi
  - Tavoitefunktio suppenee helposti pieniin arvoihin
    - Satunnaisaskel vaimenevalla todennäköisyydellä
      - Hyvin vaihteleva lopputulos samoilla parametreilla

# Ratkaisu

- Algoritmi ajettiin yli 1000 kertaa kummallekin datasetille
- Kunkin modulin säätelyohjelman säätelijät pisteytettiin
- Geenit klusteroitiin pistemäärien perusteella käyttäen hierarkista klusterointia ja k-means-algoritmia
  - Tuloksena geeniryhmiä, joita säätelevät samat geenit
- Kullekin geenille annettiin ”rooli” vastaavien säätelijöiden pisteiden ja pistejakauman perusteella
  - Esim. Yksittäinen vahva säätelijä



# Tulokset

- Löytyi useita korkean pistemäärän regulaattoreita
  - 95 sellulaasisäätelijää
  - Tämän tason päätelmät eivät olisi mahdollisia yksittäisten Genomica-tulosten perusteella
    - Toisto välttämätöntä
- Klusterointimenetelmä vaikutti klusterien rakenteeseen, mutta ei ennustettuihin sellulaasisäätelijöihin
- Aiemmin löydettyjä säätelijöitä ei löytynyt 95 säätelijäehdokkaan joukosta
  - Tämä on kuitenkin perusteltavissa monen puuttuvan säätelygeenin osalta

# Menetelmän rajoitteet

- Analyysi pohjautuu transkriptomidataan
  - Monet geenisäätelyverkoston tasot jäävät menetelmän ulottumattomiin
- Genomica-ohjelman muuttaminen ei ollut mahdollista
- Menetelmä ei pysty ottamaan huomioon ehdollista aktiivisuutta
  - Loogiset rakenteet saattavat hämätä algoritmia
- Menetelmä saattaa valita kaksi kilpailevaa geeniä säätelijöiksi samaan aikaan
  - Jos säätelijät selittävät samaa osaa varianssista, luultavasti vain toinen voi pitää paikkansa

# Päätelmät

- Module networks –menetelmä
    - Kykenee löytämään perustellusti uusia säätelijäehdokkaita
    - EM-algoritmin toisto on tulosten kannalta välttämätöntä
    - Algoritmissa ja menetelmän muissa vaiheissa on myös parannettavaa
  - Hypoteesit
    - Sellulaasisäätelijäehdokkaat ovat lupaavia
    - Korkean pistemäärän saavuttaneet säätelijät ovat hyvin vahvoja
      - Nämä hypoteettiset säätelijät eivät kuitenkaan säätele sellulaasigeenejä
    - Tulokset tarjoavat hyödyllisiä hypoteeseja *T. reesein* genomien kattavasta geenisäätelyverkostosta
-