

# Applying bioinformatic techniques to identify cold adaptive genes in oats

A Masters Dissertation by  
Henrik Thorburn

Supervised by Björn Olsson

# The Oat Project

- EST sequence oats
- Attempt to identify known and novel candidate genes mediating frost tolerance
- Identify cold adaptation pathways
  - Microarray analysis
  - Integration of known regulatory network

(Olsson et al., 2002)

# The Oat Project : Long term goals

- Identify key genes involved in
  - Perception
  - Signaling
  - Induction of cold-adaptation
- Decipher interactions between regulatory pathways

(Olsson et al. 2002)

# Problem description

- Large amount of data
- Largely unstructured data
- No standard method for gene identification
- Important to make use of already acquired knowledge and data
- Easy access to good tools, such as BLAST, FASTA, PROSITE, Pfam and PRINTS

# Problem definition

Can standard bioinformatic tools, applied to EST data, be used to identify genes with a certain functionality ?

Apply tools to data provided by the oat project, in an attempt to identify cold adaptive genes

Evaluate and report the usability of the applied tools

# Hypothesis

The chance that a newly sequenced gene already has known relatives is above 70 %

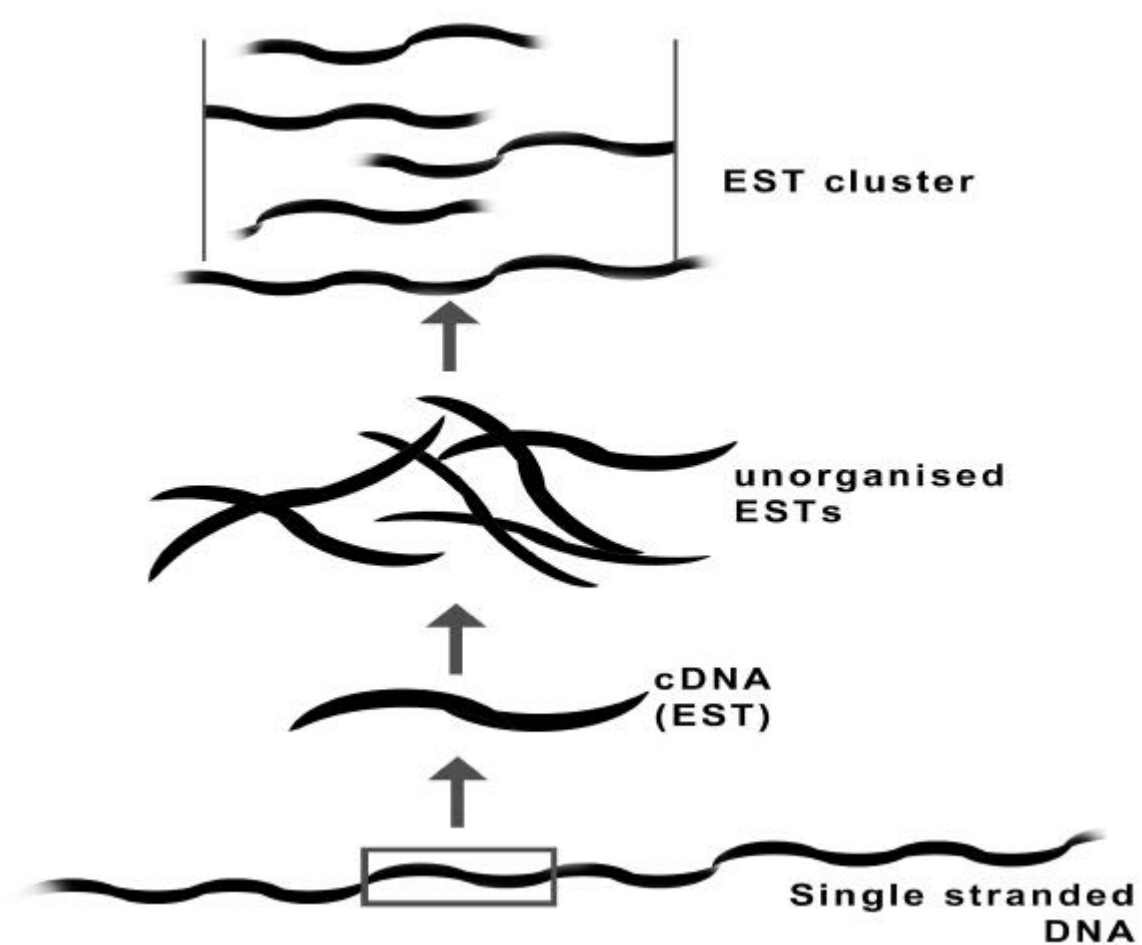
(Bork et al. 1998)

Given such similarities, it could be possible to, through comparative search, identify cold adaptive genes in the oat genome, by looking into other organisms.

# Limitations

- Focus only on bioinformatics
- Only computer science and computer algorithms
- Only look for similarities in the EST data compared to other organisms

# From EST to gene



# Characterisation of genes

- A gene is a DNA sequence, constructed by four bases: A, C, T and G
- Known and unknown genes are stored in large databases
- Similarity search tools, such as BLAST, allows us to search those databases for the occurrence of arbitrary sequences.

# Aims and Objectives

- Aim : Examine whether standard bioinformatic tools can be used to identify genes with a certain functionality
- Objectives:
  - Identify possible reference organisms
  - Perform comparative searches
  - Compare results from various tools
  - Evaluate the examined tools based on the comparison

# Time plan

- February: Preparations, theory grounding, problem definition, literature study.
- March: Set up tools, install databases.  
Test execution and verify results.  
Initial comprehensive searches
- April: Continued tool execution. Analysis and evaluation
- May: Wrap up, draw conclusions.

# BLAST

- Use of filters to avoid low-complexity sequences
- Expect value threshold
- Use iterative BLAST (PSI-BLAST)

# FASTA

- Version of FASTA
- Gap penalties
- Word-length
- Both or just one DNA strand
- Substitution Matrix
- Expect value

# PROSITE

- Scan for
  - Patterns
  - Profiles
  - Rules
- Exclude high probability occurrences
- Splicing variants

# Pfam

- Local search
  - Much more sensitive – if complete domains exist in the query
- Fragment search
  - Less sensitive
  - Able to detect fragments of domains
- Use both (Default) or just one

# PRINTS

- Expect value
- Substitution matrix
- Database to query
- Distance variance

# General issues

- Access of databases:
  - All databases ?
  - Only plant databases ?
- Only best match, or a general threshold ?