UPPSALA
UNIVERSITET

An algorithm for simultaneous search for multiple QTL

(An example of what use scientific computing can be to QTL mapping)

Kajsa Ljungberg, Sverker Holmgren and Örjan Carlborg

Complex Trait Consortium meeting
July 2003

UPPSALA
UNIVERSITET

People

Kajsa Ljungberg and Sverker Holmgren
Dept. of Scientific Computing, Uppsala University

Örjan Carlborg
Roslin Institute, Edinburgh

Leif Andersson
Dept. of Animal Breeding and Genetics
Swedish University of Agricultural Sciences, Uppsala

Martina Persson
Dept. of Mathematical Statistics, Uppsala University

UPPSALA
UNIVERSITET

Goal

To develop efficient algorithms for simultaneous search for multiple QTL.

So far experimental crosses, QTL mapping based on genomic search, mainly ordinary least squares ("Haley-Knott regression"), but also ML approaches.

Efficient algorithms for quantitative trait loci mapping problems.
K. Ljungberg, S. Holmgren and Ö. Carlborg. *Journal of computational biology*, Vol 9, pp. 793-804, 2002.

An optimization algorithm for simultaneous search for multiple QTL.
K. Ljungberg, S. Holmgren and Ö. Carlborg. *In preparation.*

UPPSALA
UNIVERSITET

Current work

We do *not* develop new mapping methods, but make standard methods faster.

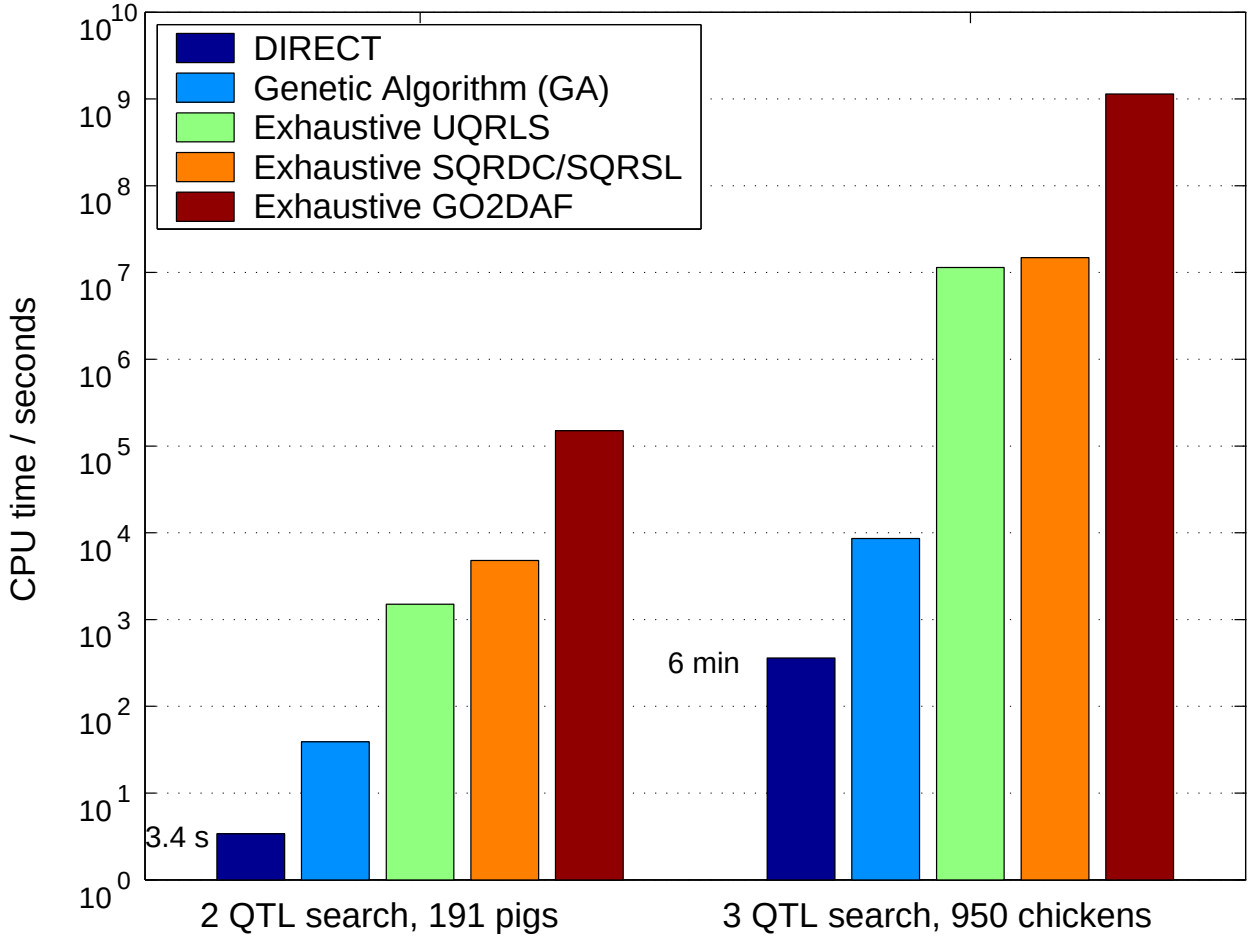
A time-consuming part of standard analysis is the exhaustive search, i.e. stepping through the genome calculating the test statistic at every position looking for the highest peak.

We calculate the same test statistic, but develop faster methods of finding the location of the highest peak without wasting time in less interesting regions.

The algorithm we use is called DIRECT. We have compared it with a genetic optimization algorithm (GA) and with exhaustive search.

UPPSALA
UNIVERSITET

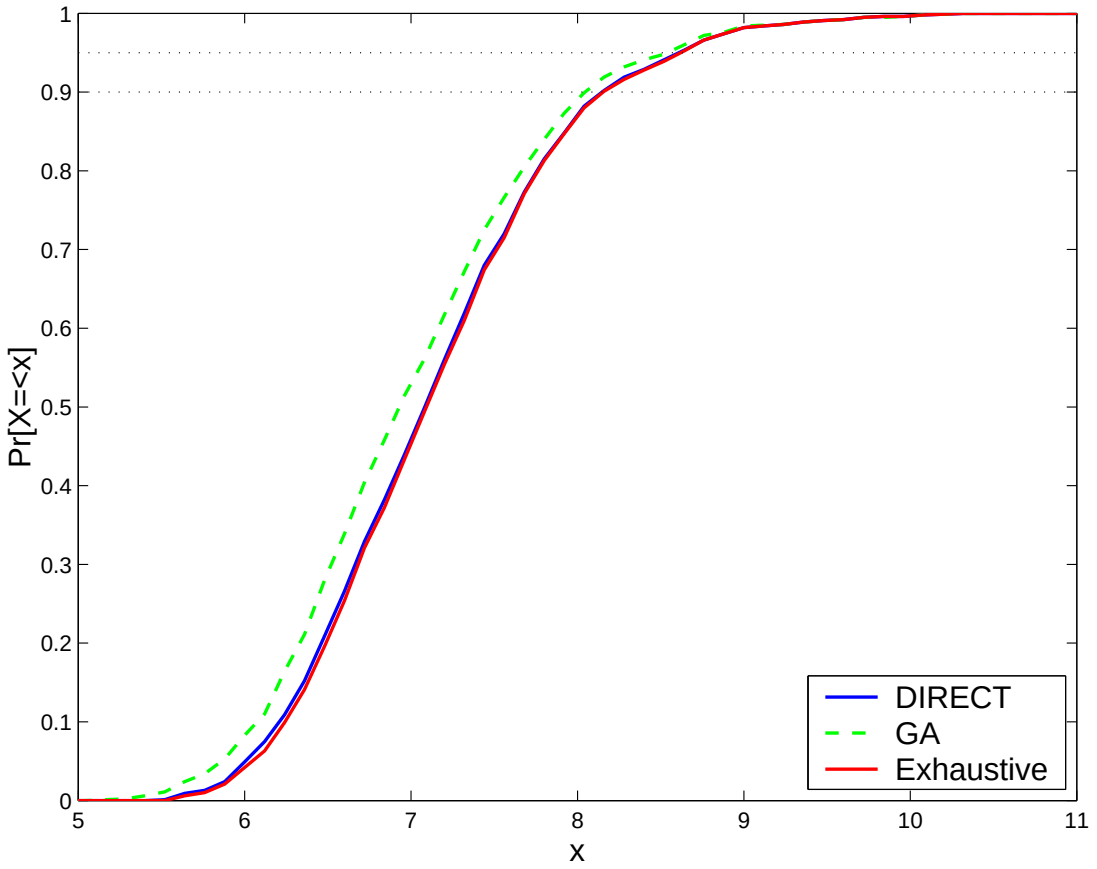
Results: CPU time



(10^9 s ≈ 31 years, 10^7 s ≈ 4 months, 10^5 s ≈ 28 hours.)

UPPSALA UNIVERSITET

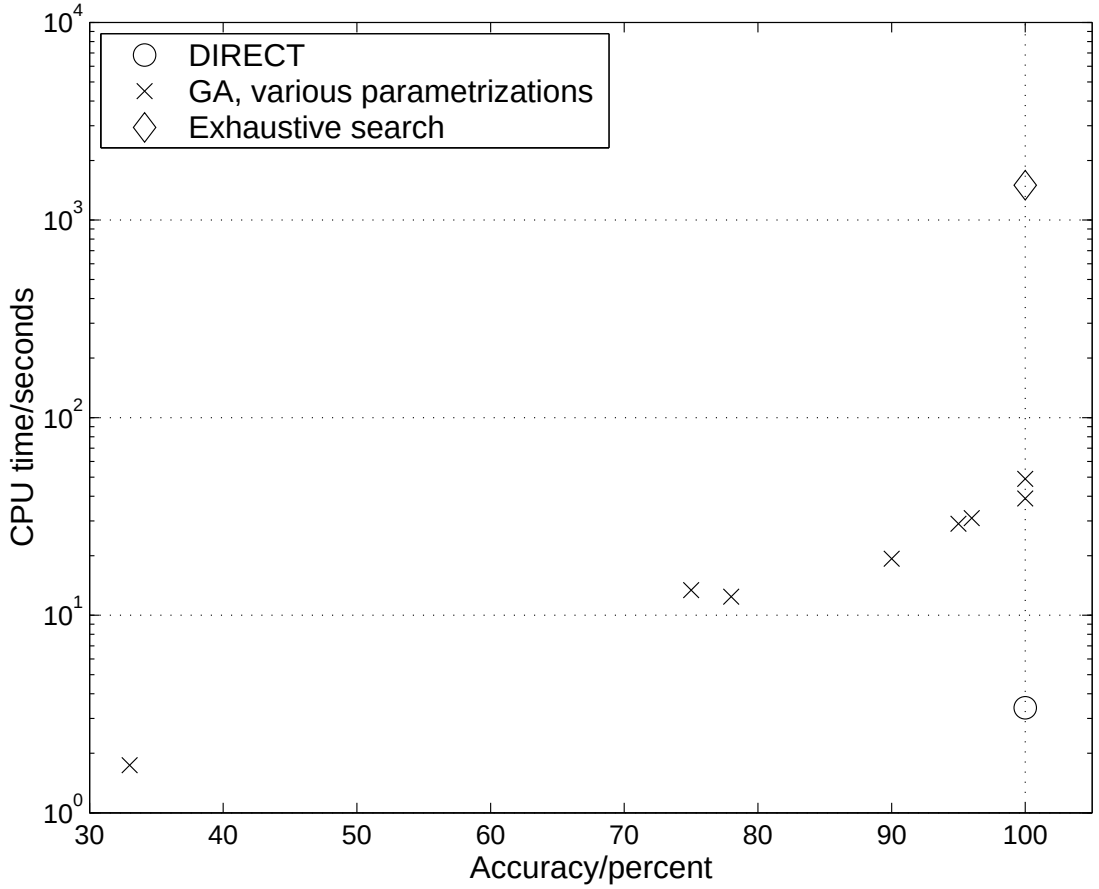
Results: Empirical significance thresholds



CDF for an empirical distribution from randomization testing

UPPSALA UNIVERSITET

Why is a new algorithm needed?



Exhaustive search is too slow. Genetic algorithm not adapted to QTL mapping problem, results sensitive to settings of many parameters.

UPPSALA
UNIVERSITET

What difference does new algorithm make?

Possible to do new things:

- Compare forward selection with simultaneous search.
- Look for epistatic interactions between QTL that lack significant marginal effects.
- Simultaneous search in up to at least 5 dimensions (4D and 5D not implemented yet).

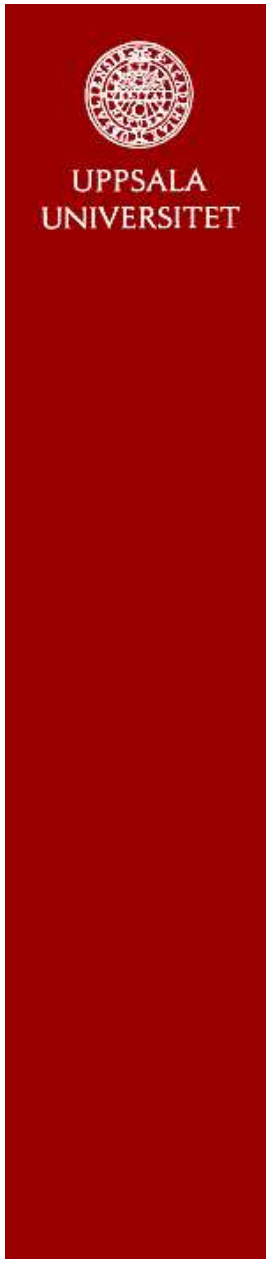
UPPSALA
UNIVERSITET

About the algorithm

DIRECT (DIviding RECTangles)

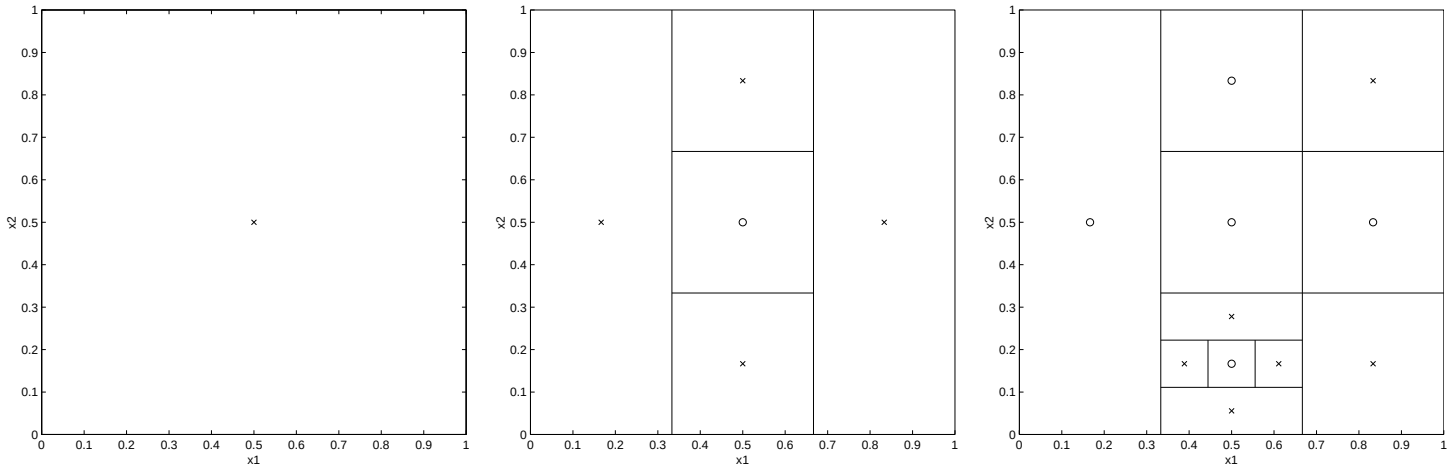
Reference:
Lipschitzian Optimization Without the Lipschitz Constant
D. R. Jones, C. D. Perttunen, and B. E. Stuckman *Journal of optimization theory and application*, 79:157-181, 1993.

DIRECT is designed for optimization of Lipschitz continuous functions. Lipschitz continuity means that the slope of the function is limited by some (unknown) constant K everywhere.



General principle of DIRECT

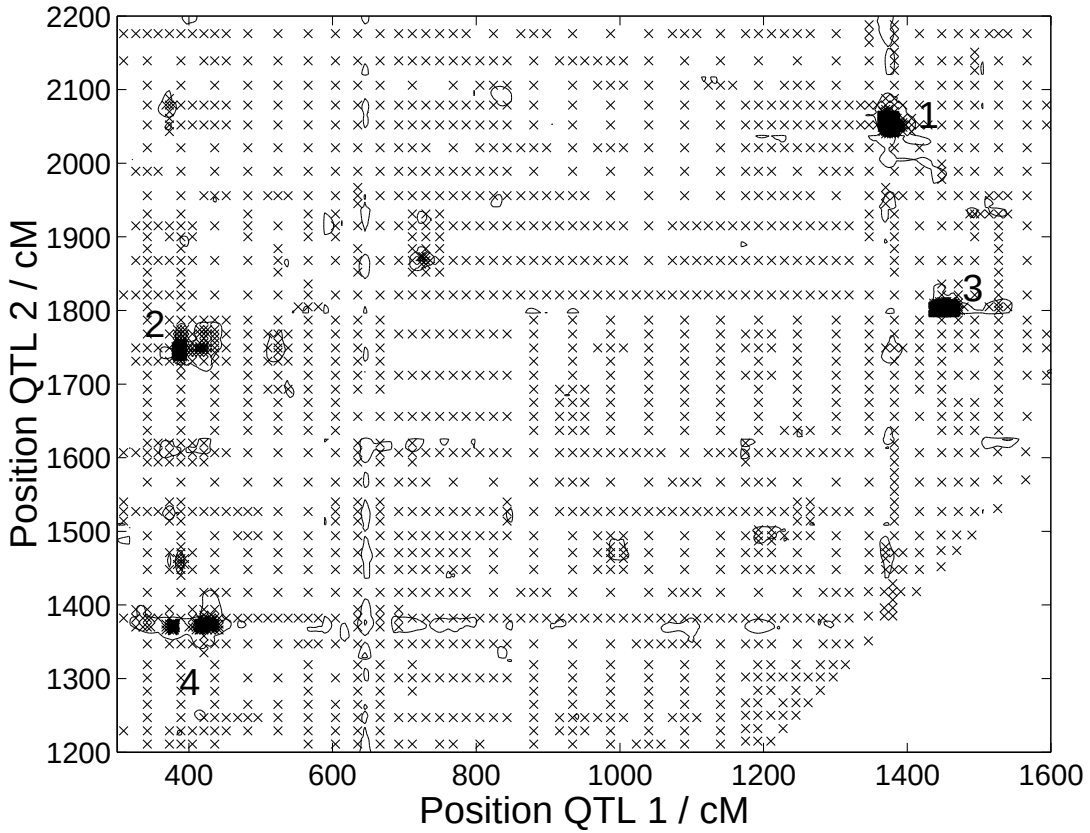
1. Assume search space is divided into rectangles.
2. Calculate function value at the center of each one.
3. Select all rectangles which, for some K, are the most promising.
4. Divide all selected rectangles into smaller ones. Go to 1.



Needs modifications to fit QTL search problem.

UPPSALA UNIVERSITET

Efficient use of function evaluations



DIRECT covers the whole search space and concentrates search around all the highest peaks.

UPPSALA
UNIVERSITET

Why does DIRECT work well for QTL mapping?

The objective function values at the markers is essentially determined by the phenotype means marker genotype classes. The change in function value is determined by the number of individuals (and their phenotypic values) that change genotype class. Genetic distance is a measure of the number of changes in genotype class, thus a small genetic distance means that there is a limit on the possible change in function value. Local Lipschitz continuity.

UPPSALA
UNIVERSITET

Summary

- DIRECT gives accurate results.
- CPU time reduced by many orders of magnitude, making new types of analyses possible.
- There might be other computational problems where dramatic speed-ups are possible if suitable library routines and/or specially adapted algorithms are used!

UPPSALA
UNIVERSITET

End of talk.