

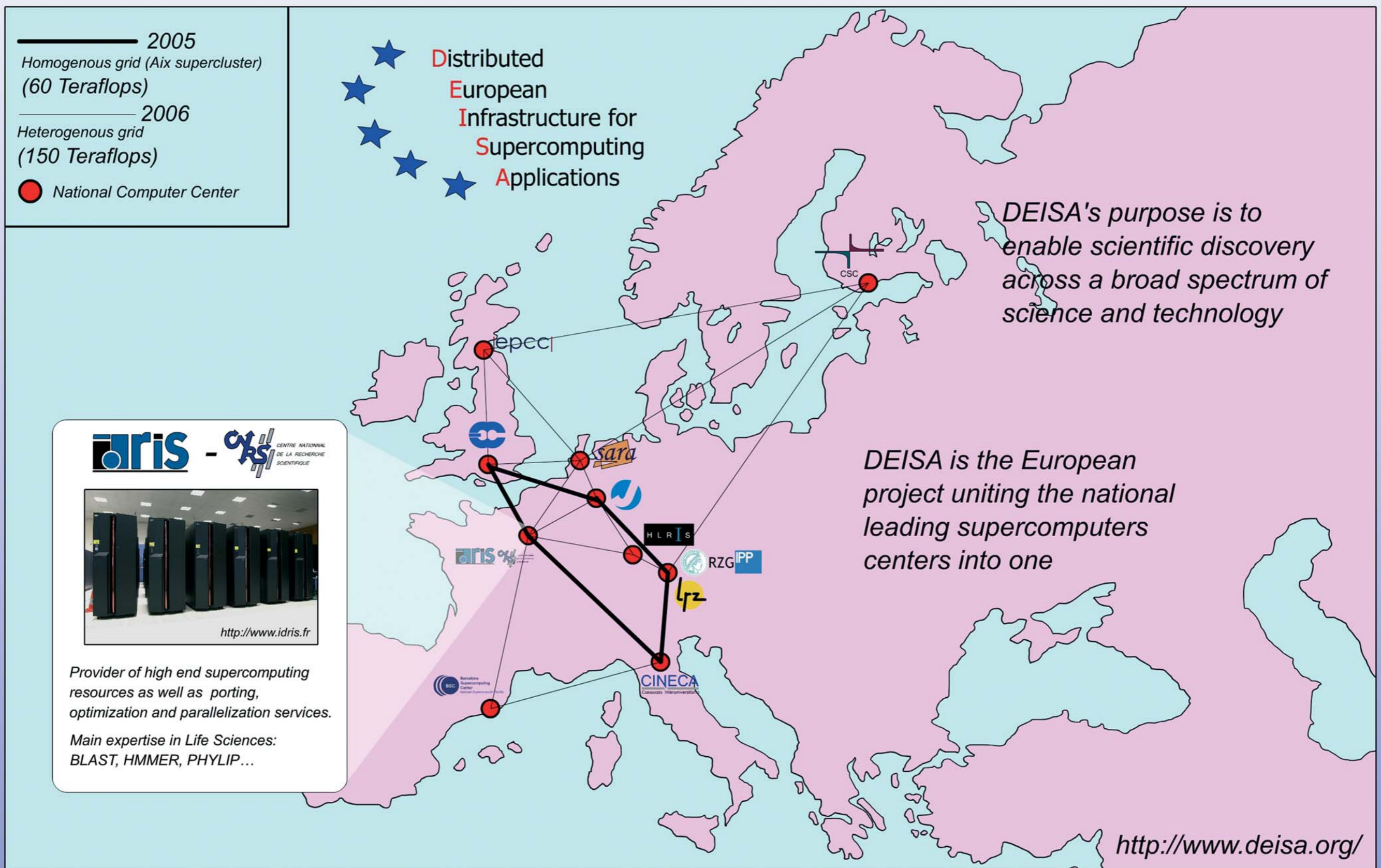
Optimization of the NCBI-BLAST code for High-throughput *in silico* comparative genomics in DEISA project

Olivier Glorieux¹, Marc Ferré², Anne Fouilloux¹, Isabelle Dupays¹, Denis Raux¹, Dominique Lavenier³, Yves Malthiery², Pascal Reynier², Yves Tourmen²

¹Institut du Développement et des Ressources en Informatique Scientifique, CNRS, Bât 506, BP 167, 91403 ORSAY Cedex, France

²Laboratoire de Biochimie et Biologie moléculaire, INSERM U694, CHU Angers, 4 rue Larrey, 49033 ANGERS, France

³Institut de Recherches en Informatique et systèmes aléatoires, Campus Universitaire de Beaulieu, Avenue du Général Leclerc, 35042 RENNES Cedex, France



Challenging problem proposed to DEISA by INSERM U694

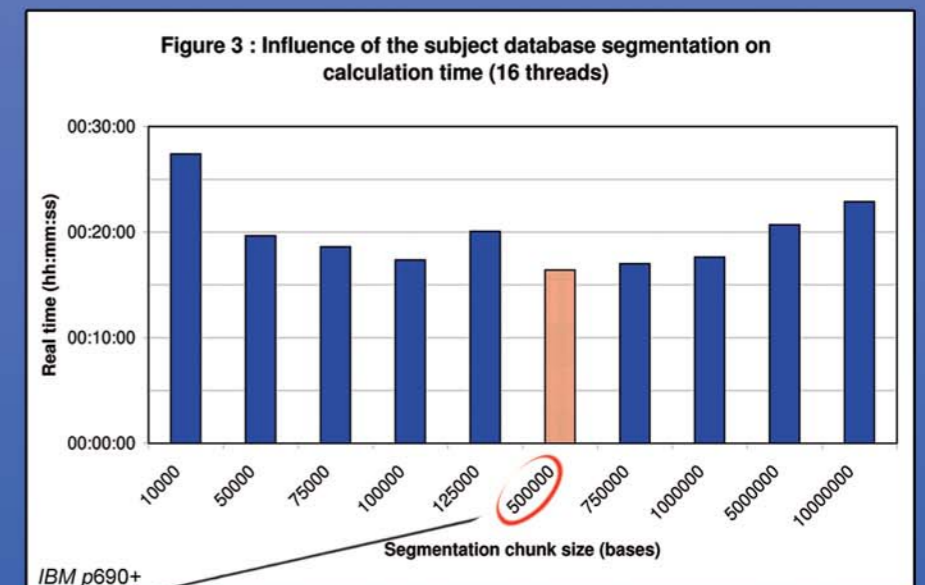
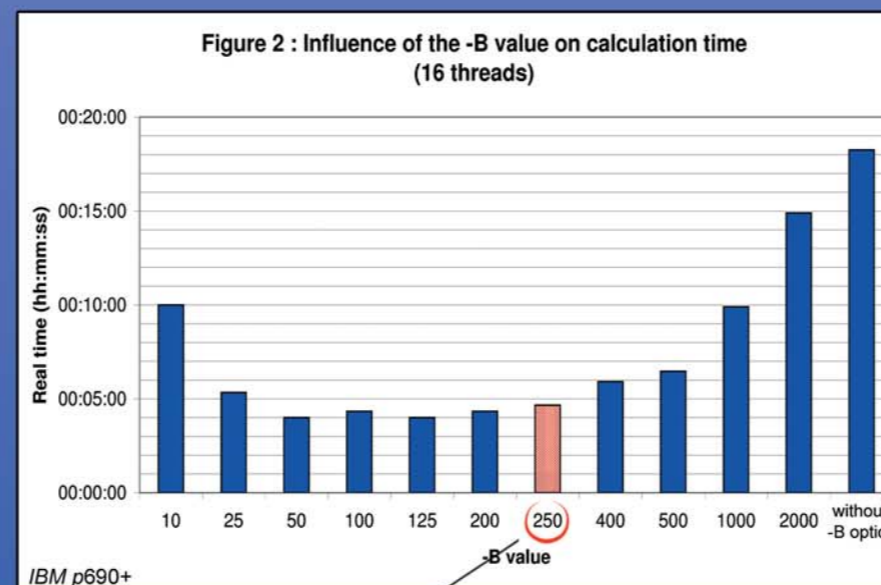
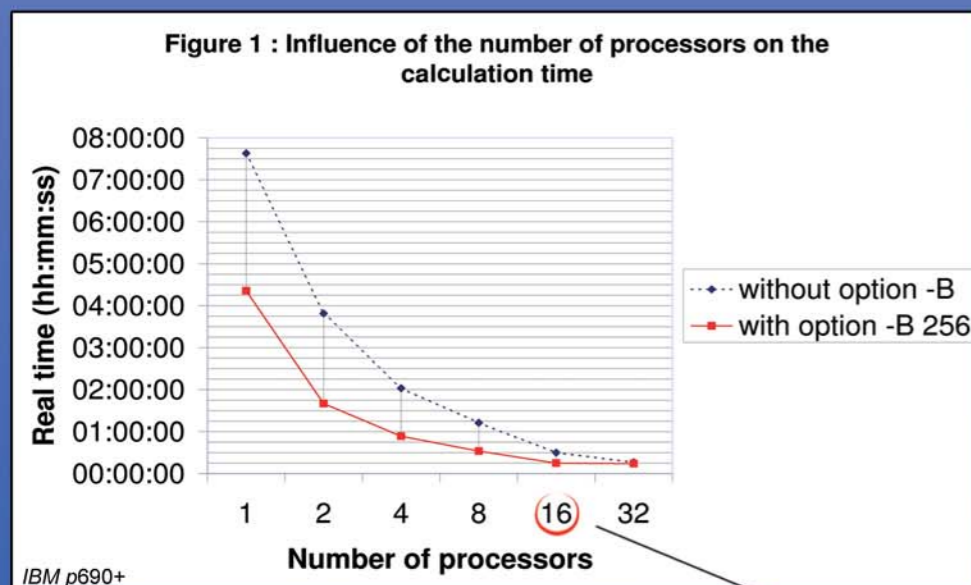
600,000 procaryotic proteins VS the human genome
(0.26 Gbases) (3 Gbases)

=> requires to conduct a very large NCBI's tblastn

Estimated Elapsed Time : 540 days on a Sun Fire 280R (750-MHz/4Go RAM)

Final Elapsed Time : 8 days on IDRIS IBM p690+ (1.7-GHz/128Go RAM)

IDRIS approach : optimization through parameter setting and code modifications



`blastall -p tblastn -i AllBact001.faa -a 16 -B 256 -d Hg17 -o AllBact001.out`

The BLAST -B option : concatenates a given number of queries in order to obtain the optimal query length for BLAST

The estimated time has been dramatically reduced with the supercomputer of IDRIS, and would be even shorter with the grid of DEISA. Contact bioinfo@idris.fr to access DEISA high computing facilities. DEISA grants support to innovative and promising projects.