Contribution ID: **158** Type: **not specified**

Biodiversity and HPC: a necessary mutual interest

Tuesday, 20 May 2014 16:50 (15 minutes)

Diversity is one emblematic characteristics of living word. Rooted in Natural History, it has extended with an increasing modelling flavor towards genetic diversity, with coevolution with statistics, and biomolecules, with system biology. Organizing biodiversity data is a challenge as life is not a random assembly of atoms. New sequencing technologies have deeply revolutionized the approach to diversity, as diversity of genomes is an imprint of diversity of organisms. Molecular data can now be produced with high throughput. Many challenges exist for organizing biodiversity data, and here are a few. There exist efficient algorithms for most of the tasks in biodiversity: multiple alignment, phylogenetic inference, clustering, etc⋯which reach a limit, either time or memory, for data produced by NGS. Two research directions coexist: either finding heuristics to speed up time, or to develop scalability. Developing scalability is a challenge, i.e. a common progress in algorithms, codes, and computing architecture. Techniques borrow tools from linear algebra, especially spectral methods, nonlinear optimization, and discrete mathematics, like computations on graphs (finding connected components, cliques, clustering). Several of these methods can be distributed, and we focus as an example on the usefulness to develop aggregative nested clustering on large data sets (between 10^5 and 10^6 specimen), as a way to reconcile Natural History knowledge and molecular phylogenetics.

Presenter: FRANC, Alain (CNRS)

Session Classification: Going beyond grid to enable life science data analysis