1. What is GATB?

Motivation
NGS technologies produce terabytes of data. Efficient and fast NGS algorithms are essential to analyze them.

Objective
The Genome Assembly Tool Box (GATB)
- is an open-source software
- provides an easy way to develop efficient and fast NGS tools
- is based on data structure with a very low memory footprint
- allows complex genomes to be processed on desktop computers

3. Compact de Bruijn graph data structure

The core data structure of GATB is a de Bruijn graph that encodes the main information from the sequencing reads.

Strength of GATB
GATB makes this graph compact by using a Bloom filter (a space efficient probabilistic data structure) and by using a CFP additional structure that avoids false positive answers from the Bloom filter due to its probabilistic nature.

5. GATB helps you as a NGS user

GATB’s de Bruijn graph: a basis for families of tools
- Data error correction
- Assembly
- Biological motif detection

Several tools based on GATB are already available:
- Bloocoo: K-mer spectrum based read error corrector for large datasets
- Minia: Short read assembler based on a de Bruijn graph. Results are of similar contiguity and accuracy to other de Bruijn assemblers (e.g. Velvet)
- DiscoSNP: Discover Single Nucleotide Polymorphism (SNP) from non-assembled reads
- TakeABreak: Detects inversion breakpoints without a reference genome by looking for fixed size topological patterns in the de Bruijn graph

6. GATB helps you as a NGS developer

The GATB C++ library gives you the opportunity to quickly develop new NGS tools that fit your needs.

Major facts about the GATB C++ library
- Object Oriented Design
- Simple and powerful graph API
- Simple and powerful multithreading model
- HDFS usage for data storage
- Fully documented with numerous code samples
- Complete test suite

License & Web Site
GATB is released under the GNU Affero General Public License.
Proprietary licencing for software editors or services providers is currently being studied.

For more details on GATB:
http://gatb.inria.fr

Published Work
R. Chikhi, G. Rizk: Space-efficient and exact de Bruijn graph representation based on a Bloom filter. Algorithms for Molecular Biology 2013, 8:22
G. Salikhov, R. Chikhi, D. Lavenier: Minia on Raspberry Pi, assembling a 100 Mbp genome on a Credit Card Sized Computer. Poster at the JOBIM conference, 2013 Jul 1-4 (Toulouse) Best poster award.

Partners
GATB is a project of INRIA. It benefits from the support of its members.