The Berkeley Phylogenomics Group research team at the University of California, Berkeley, receives an NSF EAGER grant to develop a graph database and research environment for Big Data challenges in genomics, building on the Era7 Bioinformatics' Bio4j platform

The Sjölander Lab at the University of California, Berkeley, has recently been awarded a 250K US dollars EAGER grant from the National Science Foundation to build a graph database for Big Data challenges in genomics. Naturally, they're building on Bio4j.

The project "EAGER: Towards a self-organizing map and hyper-dimensional information network for the human genome" aims to create a graph database of genome and proteome data for the human genome and related species to allow biologists and computational biologists to mine the information in gene family trees, biological networks and other graph data that cannot be represented effectively in relational databases. For these goals, they will develop on top of the pioneering graph based bioinformatics platform Bio4j.

"We are excited to see how Bio4j is used by top research groups to build cutting-edge bioinformatics solutions" said Eduardo Pareja, CEO of Era7 Bioinformatics. "To reach an even broader user base, we are pleased to announce that we now provide versions for both Neo4j and Titan graph databases, for which we have developed another layer of abstraction for the domain model using Blueprints."

"EAGER stands for 'Early-concept Grants for Exploratory Research'', explained Professor Kimmen Sjölander. "NSF awards these grants to 'support exploratory work in its early stages on untested, but potentially transformative, research ideas or approaches'. My lab's focus is on machine learning methods for Big Data challenges in biology, particularly for graphical data such as gene trees, networks, pathways and protein structures. The limitations of relational database technologies for graph data, particularly BIG graph data, restrict scientists' ability to get any real information from that data. When we decided to switch to a graph database, we did a lot of research into the options. When we found out about Bio4j, we knew we'd found our solution. The Bio4j team has made our development tasks so much easier, and we look forward to a long and fruitful collaboration in this open-source project."

About The Berkeley Phylogenomics Group:

The Berkeley Phylogenomics Group (http://phylogenomics.berkeley.edu), led by Professor Kimmen Sjölander at the University of California, Berkeley, develop machine learning methods, web servers and the PhyloFacts databases for Big Data challenges in genomics. The University of California, Berkeley, is the most highly ranked public university in the United States, with 22 Nobel Laureates over its history (and 8 current Nobel Laureates).

About Bio4j:

Bio4j (www.bio4j.com), is an open source bioinformatics graph-based DB including most data available in UniProt KB (SwissProt + Trembl), Gene Ontology (GO), UniRef (50,90,100), RefSeq, NCBI taxonomy, and Expasy Enzyme DB. The current version (0.9) has 1.216.993.547 relationships, 190.625.351 nodes and 584.436.429 properties. Bio4j is a project of Oh no sequences! (www.ohnosequences.com) the research group of Era7 Bioinformatics.

About Era7 Bioinformatics:

Era7 Bioinformatics, www.era7bioinformatics.com, is a Next Generation Sequencing service provider that specializes in microbial genomics and biological networks data analysis. Its business model is based on offering complete and integrated solutions, from project design to deep bioinformatics analysis and reporting. Era7 Bioinformatics is based in Research, Cloud Computing and Open Source.

Contacts:

Professor Kimmen Sjolander
Department of Bioengineering
Berkeley Phylogenomics Group
University of California, Berkeley
kimmen@berkeley.edu
Phone: (+1) 510-642-9932

Eduardo Pareja
CEO
Era7 Bioinformatics
epareja@era7.com
Phone: 617-401-3340