



Article:

Kenneth Chappell, Bruno Francou, Christophe Habib, Thomas Huby, Marco Leoni, Aurélien Cottin, Florian Nadal, Eric Adnet, Eric Paoli, Christophe Oliveira, Céline Verstuyft, Anne Davit-Spraul, Pauline Gaignard, Elise Lebigot, Jean-Charles Duclos-Vallee, Jacques Young, Peter Kamenicky, David Adams, Andoni Echaniz-Laguna, Emmanuel Gonzales, Claire Bouvattier, Agnes Linglart, Véronique Picard, Emilie Bergoin, Emmanuel Jacquemin, Anne Guiochon-Mantel, Alexis Proust, Jérôme Bouligand

Galaxy Is a Suitable Bioinformatics Platform for the Molecular Diagnosis of Human Genetic Disorders Using High-Throughput Sequencing Data Analysis: Five Years of Experience in a Clinical Laboratory

Clin Chem 2022; 68:2 313-321 <https://doi.org/10.1093/clinchem/hvab220>

Guest: Dr. Jérôme Bouligand is a molecular and cellular biologist specializing in molecular endocrinology and pharmacology, and responsible for high-throughput sequencing at Bicêtre Hospital in France. Kenneth Chappell is a Ph.D. student at Paris-Saclay University working in the field of pharmacogenetics.

Bob Barrett: This is a podcast from *Clinical Chemistry* sponsored by the Department of Laboratory Medicine at Boston Children's Hospital. I'm Bob Barrett.

Over the past few years, the genetic assessment of individuals with inherited diseases has increasingly used massively parallel high-throughput sequencing, which has allowed for the simultaneous analysis of thousands of genes. Even for targeted gene sequencing in clinical applications, such sequencing generates large volumes of complex data, and bioinformatics has become a necessary component in any clinical laboratory using them.

Some laboratories outsource this crucial analytical step, but that can incur significant additional costs and sometimes to an uncertain transfer of responsibility. To help solve this problem researchers developed Galaxy, a system for the integration of genomic sequences, their alignments, and functional annotation. This is an open-source Bioinformatics platform that facilitates access to Bioinformatic tools. through an intuitive interface. And there have been constant improvements and successful evolution since the initiation of the project in 2005. However, up to now, the use of galaxy has been primarily in research applications, but how applicable is this system to medical laboratories where accurate test results are crucial to patient care, and these laboratories must meet certain accreditation and regulatory oversight?

A paper appearing in the February 2022 issue of *Clinical Chemistry*, examined the use of Galaxy in the Paris public hospital system to help address that very question. We are pleased to welcome two authors of that paper. They are Jérôme Bouligand, a molecular and cellular biologist

specializing in molecular endocrinology and pharmacology, and responsible for high-throughput sequencing at Bicêtre Hospital in France. He is joined by Kenneth Chappell, a Ph.D. student at Paris-Saclay University working in the field of pharmacogenetics.

And we'll start off with you Dr. Bouligand. What exactly is high-throughput sequencing and what advantages does it offer to clinical laboratories?

Jérôme Bouligand: Okay, so high-throughput sequencing appeared in diagnostic laboratories in the mid of 2010 and as they gradually revolutionized the diagnosis of hereditary disease. It drastically reduces the time and cost of analysis compared to the old Sanger sequencing technique as the parallel analysis of hundreds of genes and dozens of patients is possible with this method. In our laboratory, we have developed numerous applications ranging from rare disease, such as hypogonadotropic hypogonadism to common disease such as depression.

Bob Barrett: So, Kenneth, let's go to you now. What are the typical applications in your clinical laboratory near Paris?

Kenneth Chappell: As Dr. Bouligand mentioned, we have rather different applications for rare and common diseases. In our lab, we analyze many different pathologies in connection with the various clinical services in our hospital, which are themselves associated with reference centers at the national level that are tasked with overseeing the care of patients with rare diseases.

Since 2015, our lab has sequenced more than 10,000 patients with rare diseases in a wide range of fields, including neurology, hepatology, and endocrinology as well as having applications in the field of pharmacology and pharmacogenetics, particularly in the field of psychiatry.

Bob Barrett: There must be challenges to treat and interpret the massive amounts of data generated by high-throughput sequencing. How do you deal with that?

Kenneth Chappell: Yes. So that is definitely the main and probably the biggest challenge that we've seen. The lab is the processing of the massive amounts of complex data that are generated through high-throughput sequencing. It requires a certain degree of expertise in bioinformatics that is essential in order to implement relevant and validate protocols. But not all the data needs to be kept. So it's important to be able to distinguish between those data that can be kept and those that we can purge and remove from our analysis.

Bob Barrett: Doctor, let me ask you this. Why did you choose the Galaxy platform for a high-throughput sequencing data analysis for your laboratory?

Jérôme Bouligand: Yes. Our project started at the end of 2014. When our lab received an Illumina MiSeq benchtop sequencer. Our project budget, at the time a concern and pushed us to choose an open source of approach. We were able to recruit only one bioinformatician at that time, and most of the biologists in their lab were not trained in bioinformatics. So we chose Galaxy which allows our biologists to easily access the bioinformatics needed.

We were impressed by the fully open and scalable nature of Galaxy, which allows for us to quickly evolve the pipeline according to the development of our lab. The commercial software solutions that we tested at the time were entirely closed and did not permit the level of scalability, nor the opportunity to learn about self-teach bioinformatics which Galaxy offers. The quality with which Galaxy was deployed for diagnosis in the hospital environment was made possible for the equipment of bioinformatician and the numerous exchanges. We add with the bioinformatics unit of the hospital and the university while we had doubts about the viability of this approach at first, we were quickly convinced and impressed by the many possibilities that Galaxy offered.

In the end, it was quite easy to install Galaxy locally in a secure hospital environment which we require to process sensitive medical data. For this installation, we were able to benefit from all the development and support of online by the Galaxy developer team, which had already made many powerful tools available in the Galaxy tool shed, such as, the Board Institute Genome Analysis Toolkit. Some colleagues predicted that we would have problems using Galaxy in the long term, but this was not the case and we do not regret our decision to go with Galaxy. It was a good decision.

Bob Barrett: Well, thank you, Doctor. Ken, what were the advantages and limitations that you encountered with Galaxy, and would you recommend it to a colleague or to listeners of this podcast?

Kenneth Chappell: So again, the greatest limitation that we encountered so far is related to data management storage. It's really up to the user to rigorously and regularly purge the intermediate data that's not necessary. Fortunately, with Galaxy that is made entirely possible with tools such as BioBlend, which has been a critical tool for us in our own lab.

As far as advantages go, Galaxy is tremendously scalable in the context of diagnosis. And for quality assurance reasons, we decided to freeze our analytical pipeline once it was validated, but when necessary we know that thanks to Galaxy

we can return to develop and evolve our pipelines easily and comfortably. Additionally, for non-bioinformatic users such as biologists and technicians within our lab, they have greatly appreciated the intuitive interface that is offered by Galaxy. And for those who are more bioinformatically inclined, they've also appreciated again the scalability of Galaxy, but also the option to manage data more behind the scenes using the command line interface.

Specifically, again, with the BioBlend software. We definitely would strongly recommend Galaxy to colleagues and others, especially for those who like us are in a mid-sized diagnostic lab who specialize more in targeted exome type analysis. For us, has been a reliable and robust approach. And we believe and hope that our experience with Galaxy between 2015 to 2020 as was reported in *Clinical Chemistry* demonstrates this reliability and robustness.

Bob Barrett: Finally, Dr. Bouligand, what is your overall view of using Galaxy in your clinical laboratory?

Jérôme Bouligand: Yes, Galaxy offers many prospects for implementing massive data processing applications in many fields of medical biology. Not just in the field of molecular genetics. This is case for example in the field of computational chemistry which has many tools available on Galaxy such as software for protein compound docking. There are up-to-date 177 repositories in this category located within Galaxy. We also believe that Galaxy is an exceptional interface for teaching bioinformatics. We have used Galaxy to teach bioinformatic as it applies to processing high-throughput sequencing data within the frame works of our university for a few years now. Thus far, students have provided great feedback from this process so it's very pleasant to work with Galaxy.

Bob Barrett: That was Dr. Jérôme Bouligand from Bicêtre Hospital in France and he was joined by Kenneth Chappell from Paris-Saclay University. They have been our guests in this podcast on high-throughput sequencing in clinical laboratories. Their paper describing their use of the Galaxy platform appears in the February 2022 issue of *Clinical Chemistry*. I am Bob Barrett. Thanks for listening.