

IN THE SPOTLIGHT

GQ Life Sciences: searching for sequence IP

The products made by GQ Life Sciences, based in Boston, make IP-related sequence searching easier for scientists, patent searchers and practitioners, as senior product manager Ellen Sherin tells *LSIPR*.

In September 2015, life science patent search company GenomeQuest changed its name to GQ Life Sciences (GQ) as it evolved to a two-product company. GQ has clients across a range of sectors—it works with the majority of large pharmaceutical, agrichemical and seed companies.

GQ has around 20 dedicated employees including additional staffing from its parent company Apteon, which has over 1,700 software professionals and support staff.

GQ's flagship product, GenomeQuest, was first released in 2004. GenomeQuest comprises a web-based interface and a collection of sequence databases. To give a sense of the scale of the information available, over 815,000 patents are indexed in the GQ-Pat database—containing almost 369 million sequences—making certain that searches for patent nucleic acid and protein sequences are as comprehensive as possible. The sheer size of the GQ-Pat database makes it stand out from the crowd. This simple interface also allows simultaneous searching of sequences from public databases, such as Genbank, EMBL and SwissProt.

Although other sequence databases are available, according to senior product manager Ellen Sherin, they are a fraction of the size.

GenBank, one of GQ's closest competitors in size (but not in focus), is a genetic sequence database that collects all publicly available DNA sequences and has around 200 million sequences; however, the majority of these are not IP-related sequences.

How does a database this size get built? As Sherin explains, information from patent applications filed around the world is extracted and made available to GQ's customers. The database is continuously updated with datastreams from patent offices around the world, including those in the US, Europe, China, Brazil, India, and the World Intellectual Property Organization (WIPO).

The first step is using a collection of computer programs to extract information from the standard "Sequence Listing (ST25)" file. In most jurisdictions, when patent applications are filed at the local patent office, the inventor must put all sequences into a formatted list, according to the WIPO ST.25 listing specification.

But this doesn't always happen—sequences may be found anywhere in a patent application,

either supplementing the ones in the attached ST25 document, or, especially in older documents, embedded within the text of the patent document. Sequences can be included in figures and tables as well. So, the next step is for proprietary algorithms to identify documents which potentially contain these "embedded sequences" and this set is manually curated to ensure all these sequences are identified and added to the GQ-Pat database. From there, the sequence data is combined and enriched with the textual and bibliographic data from their patents, and the resulting combination of sequence and text becomes GQ-Pat.

Always looking to improve its GQ-Pat database, GQ is currently in negotiations with China's State Intellectual Property Office publishing centre to receive information directly from the office, rather than manually extracting it from the applications.

Researchers can enter one or more sequences into GQ's sequence search interface and find all the patents that contain these sequences or similar ones. They can also choose to search non-patent databases and process all results simultaneously. Researchers can also have questions such as "who filed gene patents on CRISPR/Cas9 before 2014?" and "is there any prior art citing this specific sequence?" answered at the click of a button through the combination of text and sequence searching in the GenomeQuest interface, Sherin says.

GQ offers sequence searchers a choice of four different algorithms: BLAST, Fragment, MOTIF and GenePast, which, according to Sherin, is unlike the sequence comparison algorithms with which companies may be familiar.

"It's completely unique," she says, adding that the algorithm was published in 2002 and is specifically geared towards IP searching and the questions that are relevant to IP. Unlike BLAST searches, GenePast is suited to answering IP-related questions that are typically phrased as: "Find all sequences in the database that are 70% or more identical to my query sequence." Its strength is that it looks for matches across the entire sequence, rather than just finding a small region and building from there, as local alignment algorithms like BLAST do.

LifeQuest

GQ launched its second product, LifeQuest, in 2015, leveraging its 11 years of skill and experience with life science patents.



“WE WANT ALL OUR CUSTOMERS TO BE ABLE TO ANALYSE RESULTS AND GET ANSWERS TO THEIR QUESTIONS.”

According to GQ, LifeQuest is “for every life science researcher who wants to search the patent domain”.

And unlike conventional or free search tools, says GQ, LifeQuest understands the biology behind the query, essentially providing complete results without false-positive hits from non-life sciences patents.

LifeQuest works by using life sciences ontologies to enable users to search the global life sciences patent domain with ideas rather than just text words.

Complementarity

LifeQuest and GenomeQuest are key sources for scientists for research, as well as when they’re performing patent searches, explains Sherin.

“As a scientist, why not search the biggest database there is? If you just search Genbank’s public database, you’re still missing a lot of sequences, meaning you don’t have all the information,” she says.

“Searching a database as large as GQ-Pat provides additional fundamental information about your sequence, just from the knowledge standpoint alone. And as for patents, don’t you want to know if you’re at IP-risk from the very beginning? Why waste your time and money working on a sequence if there are a dozen patents claiming it?” she questions.

Conducting a quick IP search in the early days of research can save a lot of potentially wasted research money, while also focusing efforts on particular sequences, Sherin adds.

GenomeQuest’s straightforward interface makes it relatively simple for someone who is not a search professional to do these preliminary searches. Of course, as always, in the later stages of research, a professional freedom to operate search and analysis by a patent professional is advisable.

“No responsible company would want to infringe on anybody else’s patent, it would just be foolish,” she explains.

These searches also present opportunities to focus research in IP white space—where there is no IP coverage, people can develop and market, says Sherin.

For more than ten years, she was a customer of GQ, so she understands the benefits of using its search functions.

Sherin adds that the company works with many of biotech companies, law firms, and patent offices around the world.

A never-ending effort

A never-ending effort to increase the quality of the database and a focus on clients are vital ingredients to the success of the company.

The firm has launched a new interface for its results browser, currently in the beta stages, a design Sherin helped develop when she was a customer.

“The new browser makes analysing search results even more intuitive, meaning that customers can use GenomeQuest even if they are not IP search professionals. We want all our customers to be able to analyse results and get answers to their questions,” she says.

Also in the pipeline are two new modules: one on sequence variation and one an antibody search.

The sequence variation module allows users to see a landscape of variations of their sequence and look for IP white space, while the antibody search module identifies sequences containing multiple CDRs (complementarity-determining regions).

That’s all GQ will give away for now, but Sherin assures *LSIPR* that GQ is constantly looking for the next step and that there are many more projects to come.

GQ can also help with understanding the IP around “high-value inventions” such as CRISPR/Cas9 technology, which Sherin expects will be fought over for a long time.

GQ’s products can help scour patent information for single point mutations in biological sequences. For example, in just a few minutes, GQ’s database can retrieve the patents that describe Cas9 isoforms carrying

one or more of a list of six mutations of interest. This same capability is applicable to other areas of industry that focus on mutations or variations in sequences, such as industrial biotechnology and immunology.

GQ’s development efforts reflect the constantly developing and growing landscape of biological sequences in industry. In the 1980s, when Sherin worked for a Fortune 500 chemical company, she talked to a senior scientist about using such microbial strains to try to produce chemicals. She was told it was not practical and too expensive. But when you look at the field today, genetically-engineered microbial strains are many biotechnology companies’ crown jewels.

The amount of data searching, the tools that are used and the power of those tools is astounding, says Sherin, explaining that this is a direct result of the role of sequences in IP.

“With patents come money, and with money comes the drive to index the sequences, to record them, to have databases and to make them searchable,” she explains.

As an increasing number of companies turn to protecting their sequence IP, this fortifies the growth.

The US Supreme Court’s decision in *Association for Molecular Pathology v Myriad Genetics* that said isolated DNA is not patent-eligible has also had an impact.

“There’s been more of a shift towards claiming sequences in different ways, as a direct result of this decision,” she says.

Overall, it’s clear that being able to search for sequences is a vital skill for researchers and patent professionals, and GQ knows that to stay ahead, it must continuously develop and enhance its databases and search tools. ■



Ellen Sherin is senior product manager at GQ Life Sciences. She is a registered US patent agent, and currently serves as the product manager for GenomeQuest and LifeQuest. Before her appointment at GQ Life Sciences, Sherin worked for a Fortune 500 company for over 35 years. She can be contacted at: ellen.sherin@aptean.com