GenomeQuest for IP allows search professionals to comb through the several hundred million sequences found in applications and granted patents. It lets them choose from a number of well-thought search strategies that are designed to handle use cases like prior-art, freedom-to-operate, patent infringement, and competitive intelligence. Search results can be refined within the web interface, and exported to be viewed, printed or shared with others.

**Content is everything**

A search solution is only as good as the content it operates on. Therefore a considerable amount of effort is put into making the GenomeQuest IP database the most comprehensive and up-to-date IP sequence database available.

Every two weeks all IP documents and sequences from patent offices and public databases around the world are collected and combined into a single well-annotated database that is made available to subscribers immediately.

For prior-art searches, a broad set of always-up-to-date biological sequence databases is included in the search options. If this still is not enough, then users can upload their own proprietary sequence databases and annotations into the system.

**Algorithms and Search Strategies**

Different use cases require different search strategies. For example, a freedom-to-operate search with a short oligo-nucleotide sequence is different from a prior art search for the catalytic domain of an industrial enzyme.
GenomeQuest for IP features a diverse set of search algorithms that handle the different use cases with a couple of clicks in the browser. All algorithm parameters can be fine-tuned when needed.

The selection of algorithms includes classic ones like BLAST and motif search – originally designed to find biologically relevant sequence similarity – and more IP-oriented choices like GenePAST and fragment search – non-heuristic algorithms that report the percentage of identity over the entire sequence length.

GenomeQuest handles volume as well. Users can submit up to 250 separate query sequences per search. If needed, our professional service team can setup searches on even larger query sets.

Refining Search Results

With the content and search algorithm issues solved, the next big challenge becomes how to make sense of the large number of alignments a typical search generates.

To overcome this, results are – for example – organized by patent assignee or patent family. This makes it immediately clear how many different companies or IP documents are returned by the search, or – in case the search is done with multiple query sequences – how many query sequences hit a specific patent family or assignee.

All sequence search results can be filtered on alignment properties and annotation coming from the IP.
documents. This includes – but is not limited to – percentage identity, expect value, alignment length, patent assignee, patent family, legal status, kind code, abstract, claims, filing date, and earliest priority date. The intuitive filtering widget makes it easy to focus on the results that really matter.

If the same search has been done at an earlier date, then the new or changed results are flagged for easy retrieval.

**Reporting and sharing**

Results can be exported to several document formats including Word, Excel and BizInk. These reports can be customized, viewed, printed and shared with other people.

The complete, or filtered-down, results can also be shared with other people in your organization directly from the web interface.

**Keyword search**

Search through all available annotation in patent and biological databases using plain text, dates, wild cards, and proximity / phrase searching.

Keyword searches are fine-tuned for specific use cases. For example when searching with a list of patent numbers the patent family information is searched as well. All patent numbers are normalized before the search, applying rules to format patent numbers from different geographical regions and time periods. The combination of search strategy and result presentation is again the key to success here.

**Supporting Features**

Over the last decade, GenomeQuest for IP has become the industry standard for sequence searching. The long list of subscribers includes almost all well-known life-sciences companies, law firms and several major patent offices.

The continuous feedback that this large user base creates has led to an easy-to-use web interface with a large number of helpful features that make it easier to get the job done. For example:

- Database records found using a keyword search can be saved as a personal database for later sequence searching.
- Search results can be organized in folders that can be shared with other people through a couple of clicks.
The workspace can be customized, showing only the information the user is really interested in.

A result set that is being worked on can be saved. This save includes result filtering and organization, check box selection, and sorting.

**Conclusion**

GenomeQuest for IP is a complete sequence search solution that provides all the content, search strategies, result refinement and reporting capabilities required to handle the most common IP search cases.

If you are interested in learning more, then please get in touch with us.

Talk to an expert at 508-599-9181 or email ipsales@genomequest.com

GenomeQuest, Inc.
1700 West Park Drive, Suite 260
Westborough, MA 01581

**Data sources for the GenomeQuest Patent Database.**

- **USPTO.** The GenomeQuest Patent database contains all US patents and published applications with a sequence listing starting from 1980. It is updated using the official daily US Patent and Trademark Office (USPTO) feed. The average time between publication date and availability is 11 days, including 4 days retention at the USPTO itself.

- **EPO / INPADOC.** The International Patent Documentation Center (INPADOC) is a patent collection maintained by the European Patent Office (EPO). It contains patent family and legal status information covering over 70 countries and 40 patent authorities. The GenomeQuest Patent database contains all INPADOC patents and published applications with a sequence listing starting from 1979. It is updated using the official EPO feed. The average time between publication date and availability is 54 days, including 47 days retention at the EPO itself.

- **WIPO / PCT.** The World Intellectual Property Organization (WIPO) maintains the Patent Cooperation Treaty (PCT) database. The GenomeQuest Patent database contains all PCT documents with a sequence listing – electronically filed and paper-only submissions. It is updated using the official WIPO feed. The average time between publication date and availability is 7 days.

**Public databases.** The GenomeQuest Patent database contains all sequences ever published in the patent divisions of Genbank, EMBL and DDBJ sequence databases. Updates are added every two weeks.


**IP document annotation**

The following annotation is available for patent documents.

- **Abstract.** Patent abstract, summary, and related information section in US patents.
- **Application Number.**
- **Application Publication Date.**
- **Claimed SEQ ID NO.** List of claimed sequence ID number in this patent.
- **Claims.** Text of the claims.
- **Date of Entry.** Date the document has been added to the GenomeQuest Patent database.
- **Detailed Legal Status.** Description of the latest kind code and the date it was assigned.
- **Earliest Priority Date.** Publication date of the priority document in the patent family.
- **International Classification.** International, European, and US classification codes.
- **Kind code.** Region-specific code indicating where the application or patent is in the process.
- **Legal Status.** Granted or Application.
- **Number of claims.**
- **Number of sequences.** This includes claimed and non-claimed sequences.
- **Patent Assignee.**
- **Patent Family.** List of all patent numbers in the patent family.
- **Patent Family ID.** Earliest priority number of a patent family.
- **Patent Filing Date.** Date the application has been received by the patent office.
- **Patent Inventors.**
- **Patent Title.**
- **PCT Publication Date.** Date a PCT document entering the US national phase is published by WIPO.
- **PCT Publication Number.**
- **PCT Related Dates.** PCT filing date, 371 date and 102(e) date.

**IP sequence documentation.**

In addition to the annotation listed above, the following is added for sequences.

- **Features.** Sequence features listed in the original sequence listing.
- **Identifier.** Sequence ID, usually a combination of patent number and SEQ ID NO.
- **Molecule Type.**
- **Organism.**
- **Patent SEQ ID NO.** The number used to reference the sequence in the document.
- **Patent Sequence Location.** List of claims that mention the SEQ ID NO.