Plant-based vaccine production has become a viable alternative to other manufacturing methods, as it often involves fewer purification steps, is less expensive than traditional production, and yields ambient temperature-stable molecules [1]. This is particularly important for less developed regions of the world, such as Africa, where viruses like Ebola have devastated populations [2]. Because they have such high potential for providing low-cost immediate production, plant-produced vaccines hold immense promise to help the areas that need them the most [3]. Recent studies have demonstrated the feasibility of expressing an Ebola Immune Complex (EIC), coupling the Ebola glycoprotein GP1 to a humanized monoclonal antibody in tobacco plants [4]. Preliminary clinical trials for virus protection are showing promise in animal model systems [5].

In-house R&D genome sequencing of one Ebola strain identified a potential protein epitope that could be coupled with an immunoglobulin and expressed in plants, for purification and usage as a vaccine to protect against this devastating filovirus. FTO for each step of the process had to be ensured, and patentability for potential inventive aspects needed to be evaluated.

**Definition of the query parameters**

For assessing nucleotide and protein sequence similarities for the EIC protein the R&D group identified, GenePAST was the first algorithm used, as it allows percent identity cutoffs to be specified in the query step. The results were easy to analyze versus typical claim language in the technology space. If particular motifs or fragments of sequences were identified that were commonly referenced in patent claims, refinement searches using the fragment algorithm were run. The fragment algorithm is particularly suited to smaller sequence lengths that have higher percent identity across local regions.

Both the nucleotide and corresponding polypeptide sequences were searched, against all databases that may be relevant (patent and non-patent). GenomeQuest allowed the selection of any one or multiple subject databases, as well as the option to create a custom database (such as in-house genomic sequencing collections) against which to search. Coverage of every important patent authority ensured comprehensive results. Both the GQ Gold+ collection (ST.25 listings from the US, EPO, WIPO, JP, KR) and the Platinum collection (non-ST.25 listings as well as extended legal status, normalized patent assignee, unique family sequence IDs for US, EPO, WIPO, JP, KR, AT, AU, BE, CA, CH, DE, ES, FR, GB, LU, NL, NO, TW, with in-country documents for CN, BR, IN, RU) were selected.

In this case, as the prior art potentially overlapped many different data sources such as plants (vaccine expression), humans (disease target), animals (model studies), viruses (Ebola), as well as likely hits in high throughput sequence collections, the first query run was against every database that GQ offers – encompassing more sequences than any other search platform in the world.

**Identification of different patentability and FTO considerations with meta data**

Before assessing sequence hits, it was possible to view a snapshot of patent vs. non-patent results, earliest priority dates, and other key details. Metadata, such as earliest instance of the sequence in a patent document as well as in literature, informed further investigations. GenomeQuest revealed in a summary table that the EIC sequence had a non-patent disclosure nearly a decade prior to a hit against a US patent, indicating that the composition and methods of use would have different patentability and FTO considerations. From a single query, it was straightforward to then assess the patent and non-patent literature, either simultaneously or separately.

**Organization of results with user-defined requirements**

GenomeQuest provided results that were sortable, filterable, and views were customizable, with many options for displaying and ordering fields. Key information was also displayed in graphics, for quick visual assessment. Hyperlinks to legal authority original

---

GenomeQuest provided results that were sortable, filterable, and views were customizable, with many options for displaying and organizing results with user-defined requirements. Query, it was straightforward to then assess the patent and non-patent literature, either and methods of use would have different patentability and FTO considerations. From a single disclosure nearly a decade prior to a hit against a US patent, indicating that the composition GenomeQuest revealed in a summary table that the EIC sequence had a non-patent sequence in a patent document as well as in literature, informed further investigations. Results, earliest priority dates, and other key details. Metadata, such as earliest instance of the results, against every database that GQ offers – encompassing more sequences than any other search platform in the world. (non-ST.25 listings as well as extended legal status, normalized patent assignee, unique family sequence IDs for US, EPO, WIPO, JP, KR, ensured comprehensive results. Both the GQ Gold+ collection (ST.25 listings from the US, EPO, WIPO, JP, KR) and the Platinum collection databases are updated. Audit trails were kept of every search parameter, execution, and result set. Instant alerts were also be set up to identify newly published sequences that would impact existing queries. Sharing between GQ users, with customizable privileges, was made included complete and accurate information. In both analysis and decision making, there was no need to execute redundant searches, and no need to guess which search parameters another searcher might have used. Attorneys rendering opinions were assured that the dataset from which their decisions were made included complete and accurate information. Although some outside counsel and search firms offer GenomeQuest searching services, it was important to the company to have GQ in-house for accessibility of queries and results, as well as continued re-analysis at different time points and as the underlying databases are updated. Audit trails were kept of every search parameter, execution, and result set. Instant alerts were also be set up to identify newly published sequences that would impact existing queries. Sharing between GQ users, with customizable privileges, meant that group projects were streamlined and seamless. Everyone was working with the same information, in one place. GenomeQuest was the only solution that could provide all of the content required for a comprehensive analysis and all of the features needed for reporting and decision making. No other IP sequence platform approached the scope, accuracy, and flexibility of GenomeQuest, and thus was the only sequence search tool that was needed to conduct FTO and patentability analyses.