Vol. 18 (2026): Proceedings of 2026 International Conference on Smart Healthcare and Biomedical Engineering (ICSHBE 2026)
E-ISSN: 3105-0522 P-ISSN: 3105-0514
DOI: https://doi.org/10.70267/shbe.2026287293

The Role of GenSAS, BLAST, and UniProt in Bioinformatics for Gene Research

Rumo Bian*

Dalian Polytechnic University, Dalian, China

**Corresponding author: Rumo Bian.*

Abstract

The rapid expansion of genomic and proteomic data has intensified the need for robust bioinformatics tools capable of transforming raw sequences into meaningful biological insights. This review highlights the roles of three foundational platforms—GenSAS, BLAST, and UniProt—in gene research workflows. GenSAS offers comprehensive genome annotation capabilities, particularly for non-model organisms; BLAST enables efficient sequence alignment and homology detection; and UniProt serves as a rich repository of curated protein information. This research analyzes their individual functions, integration potential, and applications across structural annotation, functional prediction, and comparative genomics. Case studies demonstrate their use in diverse research contexts, including plant genome annotation, evolutionary analysis, and protein function characterization. We also discuss future directions, emphasizing the importance of machine learning integration, multi-omics interoperability, and enhanced user accessibility. Together, these tools provide a synergistic framework that continues to drive advances in gene discovery and bioinformatics.

Keywords

bioinformatics, gene function annotation, GenSAS, BLAST, bioinformatics proteomics

1. Introduction

The explosive growth of genomic and proteomic data in the post-genomic era has fundamentally transformed the landscape of biological research. High-throughput sequencing technologies now enable researchers to decode entire genomes within days, resulting in a deluge of raw data. However, the true value of this data lies in its biological interpretation—identifying gene structures, predicting gene functions, and elucidating protein characteristics. This shift from data generation to functional insight has elevated bioinformatics tools to a central role in modern life sciences [1].

Among the vast array of bioinformatics platforms, three tools stand out due to their foundational contributions and widespread adoption: GenSAS (Genome Sequence Annotation Server), BLAST (Basic Local Alignment Search Tool), and UniProt (Universal Protein Resource). Each of these tools addresses a critical stage in the gene analysis pipeline [2]. GenSAS supports structural and functional genome annotation, particularly in non-model organisms. BLAST remains a cornerstone for sequence similarity searching and evolutionary analysis. UniProt, meanwhile, serves as a comprehensive database for protein function, structure, and pathway information [3].

Despite their distinct roles, these tools are often used synergistically in gene research. GenSAS-annotated genes may be further explored via BLAST to determine homologs and evolutionary context, while UniProt links these sequences to protein-level function, aiding in downstream proteomic and biomedical research. Together, they form an integrated framework for transitioning from raw sequence data to biological insight [4-6].

This review aims to examine the capabilities and typical applications of GenSAS, BLAST, and UniProt; analyze their collective impact on structural annotation, functional prediction, and comparative genomics; and explore future trends in tool integration, including the roles of machine learning, multi-omics data fusion, and accessible platform design. Through this, we aim to clarify how these resources continue to shape gene research and support the ongoing advancement of computational biology [7].

2. Application of Bioinformatics in Gene Research

2.1 Bioinformatics Methods for Gene Data

In gene research, especially in the analysis of DNA and RNA sequence data, bioinformatics has come to play an indispensable role. The advent of long-read single-molecule sequencing for both DNA and RNA has created an avalanche of genetic information that demands scalable bioinformatics tools to make sense of the sequences [8]. Genomic tools such as GenSAS, BLAST, and UniProt are integral for the comprehension of gene data used in identifying genes, predicting their functions, or understanding relationships to biological processes, among others. [9]

GenSAS, on the one hand service-based tool suited for eukaryotic genome structural and functional annotation only. GenSAS houses a collection of gene, regulatory element, and other genome feature annotation tools integrated along the following tiers to offer an end-to-end view of genomic content. This is particularly useful for non-model organism projects, where adequate annotation information may be scarce [10,11]. GenSAS empowers researchers with high precision to annotate these genomes, a necessary procedure for decoding the genetic elements underlying different biological functions and processes.

Now BLAST (Basic Local Alignment Search Tool) is a sequence comparison tool that could assist researchers in pinpointing sequences having the same ancestor in a particular database if the query has some similarity to described sequences. It helps to home in on the genes because by finding homologs (similar sequences) of a gene from one organism, researchers can learn more about that gene's function and evolutionary history within others. In summary, BLAST provides a powerful method of aligning sequences and detecting correspondence for the first steps in gene annotation. [12] It is a good way for the researchers to know how that gene has evolved and its functional role, depending on its similarity with other known genes.

UniProt enhances these tools with a widely used database of protein sequences and functional information needed to understand the function of genes at the level of proteins. For proteins, UniProt provides comprehensive information on protein sequences with full descriptions of the function and properties (e.g., structure, interactions). The importance of this repository extends to linking gene sequences (protein-coding regions) with their protein products and elaborating the consequences of genetic changes on predicted function.

2.2 Gene Function Annotation and Prediction

In bioinformatics, gene function annotation is an important task that involves predicting the functions of genes and their products (proteins) in cellular processes. The obvious challenge is that for newly sequenced genomes, many genes will not have known functions. These tools include GenSAS, BLAST, and UniProt (details column) that provide the necessary data to process through genes for fine gene annotation [13].

On the other hand, GenSAS combines gene prediction algorithms with sequence alignments and functional databases to define putative functions of genes. It does this by comparing the gene sequence with libraries of known sequences like Pfam and InterPro, which describe protein families, domains, and functional sites. Genes can infer the possible function of a gene by identifying similar nature to a query gene. Such methods are especially powerful for non-model organisms with limited experimental data.

BLAST makes it possible to predict the function of a gene by finding homologous genes in other species that are likely to have similar functions. BLAST is a method allowing the matching of the sequence of an unknown gene with sequences from other organisms, resulting in predictions on what kind of substances can be produced or processed by this protein based on its similarity to proteins that are known proteins. The reason for this is that genes with related sequences tend to have similar functions, a fundamental idea of comparative genomics. Dr. Ddistone: The work in predicting gene function from sequence similarity has been a cornerstone of rapid annotation of newly sequenced genomes [14,15].

Second, UniProt ensures gene function annotation to end errors in the human genes encoding proteins and provides many types of information on each protein sequence. Final—It is. For example, active sites, In-forma active site motif, Comparative domains, and Post-translational modification. It is crucial to appreciate this detail for the sake of understanding protein function and interaction inside a cell [16]. For instance, identifying the active site of an enzyme will tell us a lot about how it functions as a catalyst (and knowing post-translational modifications can inform us about what regulates its activity). UniProt offers annotations of a wide variety, which provides researchers with valuable insights into how genes function on the molecular level.

Such integration tools enable the researchers to interpret their predictions about a gene considering this web of knowledge, even if they do not have specific experimental results. Especially in large-scale genomic studies, where experimental validation of all genes is not feasible. When considering the predicted functions of or involvement in biological processes filled by genes, bioinformatics provides another manner to save time and effort for scientists. The ability to do this in an automated way and on the annotation of predictive gene function en masse has been a game changer for evolutionary biology as well as functional genomics/systems biology.

2.3 Analysis of the Dataset and Interpretation of Results

Robust tools are needed to analyze the genomic and proteomic datasets that large data volumes will be generated by, giving a greater understanding through a deeper look. Related work GenSAS is a popular frontend with BLAST capabilities and the UniProt search utility, which provide functionalities for data interpretation, so they are very suitable options.

So far, GenSAS supports the full genic annotation: genes, regulators, and other genomic features. This is particularly useful for projects involving newly sequenced genomes, since a genome annotation can provide key insights into what the sequence does or might be doing. The integration of all annotation tools is an advantage because it provides ease, simplicity, and a better experience to cater to more experienced researchers and researchers with less time at hand. Users can employ it in various genomic studies due to its versatility for different types of genomic data, ranging from basic research to biotechnology applications [17].

Having easy access to comparisons between sequences is an essential tool for any molecular biologist; it also allows researchers to identify homologs in other species and provides valuable information. The study of gene-based phylogenies is crucial because it enables the investigation of species relationships among genes and, therefore, the identification of conserved functional elements across species [18]. Comparing sequences among different species provides information about evolution and may help determine how gene functions have been conserved or altered over time. BLAST works efficiently and quickly, making it ideal for use in high-throughput sequencing studies, such as those requiring thousands of sequences for comparison and analysis.

Finally, BLAST is widely used in metagenomics, where it enables identification of microbial species present in environmental samples based on comparison of the DNA sequences generated from such samples to known reference sequences available from. This particular use of BLAST has been critical in propelling our knowledge base forward concerning microbial diversity and the roles microbes play within different ecosystems.

UniProt can be useful as an extension of those resources to accomplish the same analysis and others with protein data. This is of fundamental importance for interpreting genomic and proteomic analyses in which gene sequences must be linked to the proteins that they encode, so as variables around biologicals. The vast amount that can be found in UniProt addresses protein structure, function, and GenSAS interactions, which are important information for molecular biologists researching the biological role of any gene. UniProt connects

the dots between genomic data and GO annotations that allow researchers to move from sequence analysis towards functional understanding of their results by deciphering their relevance.

3. Bioinformatics Case Studies with GenSAS/BLAST and Uniprot

3.1 Annotation of the Plant Novel Genome with Gensas

For example, in a recent study, researchers tried to annotate the genome of a newly discovered plant species that did not have any previous genomic data available. Due to the absence of available annotations, we decided on GenSAS as the main tool in this project. To call genes, identify regulatory elements, and annotate repetitive sequences in the genome of this plant, researchers started by sequencing its whole genome and then used GenSAS. Using GenSAS, an annotation framework that can absorb multiple types of annotations and generate a normalized set from them, the researchers were able to generate a significant improvement in plant genome annotation with the identification of various new genes related to photosynthesis, stress, etc [19]. The case study demonstrates the potential power of GenSAS in annotating non-model organisms, for which conventional annotation approaches might not always deliver their optimal results.

3.2 Identifying Gene Homologs in Evolutionary Biology BLAST

In evolutionary biology, BLAST has become particularly significant in the efforts to uncover what genes do and assess how this function relates between different taxa. In one case, for example, researchers compared genomes from multiple mammalian species using BLAST to find genes responsible for development processes in these organisms. BLAST made it possible to identify the conserved gene families of importance in development (for example, the Hox gene cluster) by aligning genes from various species [20]. This study was able to uncover these functionally conserved genes and the potential regulatory mechanisms that are involved in mediating developmental processes from varying species. The power of BLAST in comparative genomics and its ability to determine the evolutionary relationships among genes are highlighted through this case study [21].

3.3 Uniprot for Protein Function Prediction and Structural Annotation

For example, researchers studying a metabolic pathway in bacteria were able to use UniProt entries as starting points for annotating the proteins involved in that pathway. This, in large part, was made possible by the availability of thousands of protein sequences and related functional annotations from UniProt that facilitated their prediction of functions for many uncharacterized proteins within this pathway. To put this in perspective, the researchers utilized protein-domain and activity data from UniProt to postulate that these proteins might possess enzymatic activities. Furthermore, they used UniProt's connection to structural databases for structure modeling of a related enzyme and obtained insights into their potential roles in the pathway [22]. This case study demonstrates an example of how UniProt can be leveraged to not only aid protein function prediction but also consolidate structural data into our knowledge space [23].

4. Future Directions in Bioinformatics Tools Development

4.1 Progress in Machine Learning for Gene and Protein Annotation

Increasingly, bioinformatics is becoming part of the machine learning universe with tools like GenSAS and BLAST or databases such as UniProt regulated by model-driven data. Machine learning can vastly improve the accuracy and speed of gene and protein annotation by transforming it from a largely manual effort into a mostly automated process. For instance, machine learning can train complex algorithms on large numbers of annotated genomes to predict gene function better than current methods from sequence patterns that have no obvious correlate [24]. These algorithms can be applied in a similar way to predict protein-protein interactions and functional sites with higher accuracy than the predictions provided by databases such as UniProt, which, at large, yield annotations that need improvement [25].

4.2 Multi-omics Data Integration for Comprehensive Biology Insights

Bioinformatics tools can also be extended to integrate multi-omics data (genomic, transcriptomic, proteomic, and metabolomic). This integrative approach would offer a broader perspective and make it possible to study how mutations at the genomic level affect transcriptomes, proteomes, and cell phenotypic functions. Analysis tools like GenSAS, BLAST, and UniProt can be improved to include multi-omics data so that analyses conducted by researchers consider the entangled nature of biological networks [26-28].

4.3 User Accessibility and Interactive Visualization Tools

As bioinformatics tools are used in a wide range of different research environments, we predict that future developments will aim for an even more accessible user experience and improved data visualization capabilities. Such as simplified web interfaces, supportive interactivity to visualize data, and the ability to integrate with cloud-based platforms would enable more users (who may face difficulty in coding) to be able to use tools such as GenSAS/BLAST/UniProt. Better visualization tools would further enable users to function with large and complex sources more interactively, which can aid in the interpretation of research results as well [29,30].

4.4 Expand the Size and Scope of the Databases

As more genomes are sequenced and new protein structures are resolved, the database supporting tools such as UniProt and BLAST will continue to expand. This expansion requires a constant effort to collate and integrate new data, ensuring that these tools remain current and comprehensive. In addition, more advanced algorithms may need to be developed to search and retrieve information from these ever-expanding databases, allowing researchers to find relevant data more efficiently. As bioinformatics becomes more data-intensive, the scalability of these tools will be a key factor in its continued success.

5. Conclusion

GenSAS, BLAST, and UniProt have become indispensable tools in the field of bioinformatics, each making unique contributions to the analysis and interpretation of gene and protein data. GenSAS specializes in structural and functional annotation of eukaryotic genomes, especially those of non-model organisms. BLAST is the basis for sequence comparisons, allowing researchers to identify homologous genes and understand evolutionary relationships between species. UniProt provides a comprehensive library of protein sequences and functional information that is critical for linking genomic data to biological function.

Together, these tools form a powerful toolkit for modern genomics and proteomics research, enabling researchers to make informed predictions about gene and protein function, analyze large data sets, and gain insight into the molecular mechanisms that drive biological processes. As bioinformatics continues to evolve, the integration of advanced technologies such as machine learning, multi-omics data, and enhanced visualization tools will further enhance the capabilities of GenSAS, BLAST, and UniProt to drive discoveries in genetics, molecular biology, and more.

References

[1] Humann JL, Lee T, Ficklin S, Main D. Structural and Functional Annotation of Eukaryotic Genomes with GenSAS. Methods Mol Biol. 2019; 1962:29-51. doi: 10.1007/978-1-4939-9173-0_3.

[2] Ladunga, I. Finding Homologs in Amino Acid Sequences Using Network BLAST Searches. Curr Protoc Bioinformatics. 2017; 59:3.4.1-3.4.24. doi: 10.1002/cpbi.34.

[3] Singh H, Raghava GP. BLAST-based structural annotation of protein residues using the Protein Data Bank. Biol Direct. 2016;11(1):4. doi: 10.1186/s13062-016-0106-9.

[4] Bateman et al. (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research, 51(D1), D523-D531.

[5] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403-10.

[6] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: architecture and applications. BMC Bioinformatics. 2009; 10:421.

[7] Boratyn GM, Schäffer AA, Agarwala R, Altschul SF, Lipman DJ, Madden TL. Domain-enhanced lookup time accelerated BLAST. Biol Direct. 2012; 7:12.

[8] Madden TL. The BLAST Sequence Analysis Tool. In: McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. 2nd edition. Bethesda (MD): National Center for Biotechnology Information (US); 2013.

[9] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12(1):59-60. doi: 10.1038/nmeth.3176.

[10] Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. The NCBI BioSystems database. Nucleic Acids Res. 2010;38(Database issue): D492-6. doi: 10.1093/nar/gkp858.

[11] Ashburner M, Ball CA, Blake JA, et al. Gene ontology: a tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25(1):25-29.

[12] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39(Web Server issue): W29-W37.

[13] Eddy SR. Profile hidden Markov models. Bioinformatics. 1998;14(9):755-763.

[14] Punta M, Coggill PC, Eberhardt RY, et al. The Pfam protein families database. Nucleic Acids Res. 2012;40(Database issue): D290-D301.

[15] Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol. 1999;292(2):195-202.

[16] Hunter S, Jones P, Mitchell A, et al. InterPro in 2011: new developments in the family and domain prediction database. Nucleic Acids Res. 2012;40 (Database issue): D306-D312.

[17] UniProt Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47(D1): D506-D515.

[18] Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017;35(11):1026-1028.

[19] Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. Bioinformatics. 2001;17(3):282-283.

[20] Lee T, Main D. GenSAS: a web-based integrated genome sequence annotation pipeline. Methods Mol Biol. 2019; 1962:33-45. doi: 10.1007/978-1-4939-9173-0_4.

[21] Mount DW. Using BLOSUM in sequence alignment and evolutionary distance estimation. Methods Mol Biol. 2008; 484:575-580.

[22] Lemoine F, Domelevo Entfellner JB, Wilkinson E, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature. 2018;556(7702):452-456.

[23] Sievers F, Wilm A, Dineen D, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7:539. doi: 10.1038/msb.2011.75.

[24] Sayers EW, Barrett T, Benson DA, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2011;39(Database issue): D38-D51.

[25] Thimm O, Bläsing O, Gibon Y, et al. MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 2004;37(6):914-939.

[26] Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841-842.

[27] Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019;47(D1): D607-D613.

[28] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-780.

[29] Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res. 2015;43(W1): W39-W49.

[30] Xie C, Mao X, Huang J, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;39(Web Server issue): W316-W322.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access