GENE ANNOTATION BY USING APOLLO

Spoorthy Nirvani, K. Aparna, T. Sajitha, Ashwini Bhuse, P. S. Deshmukh

Dept. of Biotechnology and Bioinformatics, MGM's College of Computer Science,
Near Airport, Nanded 431606.

ABSTRACT

The process of identifying the locations of coding regions in a gene and determining what those regions do. An annotation (irrespective of the context) is a note added by way of explanation or commentary. Once a genome is sequenced, it needs to be annotated to make sense of it. Assigning identities and functions to sequences within the gene is called gene annotation.

Apollo, a genome annotation viewer and editor, has become a widely used genome annotation and visualization tool for distributed genome annotation projects. When using Apollo for annotation, database updates are carried out by uploading intermediate annotation files into the respective database. This non-direct database upload is laborious and evokes problems of data synchronicity. The well-established inaccuracy of purely computational methods for annotating gene sequences necessitates an interactive tool to allow biological experts to refine these approximations by viewing and independently evaluating the data supporting each annotation. Apollo was developed to meet this need, enabling curators to inspect genome annotations closely and edit them.

Keywords: Gene, Sequence, Annotation, Browser, Editor

Introduction

Annotations describe the genome and transform raw genome sequences into biological information by integrating computational analyses, other biological data and biological expertise. Genome annotation is the process of attaching biological information to sequences. [1] It consists of two main steps:

1. identifying elements on the genome, a process called gene prediction, and

2. Attaching biological information to these elements.

The process of identifying the locations of coding regions in a gene is called gene annotation. Gene annotation is the assignment of some known or predicted biological function or experimental observation to a particular physical region of a gene sequence. A region consists of a contiguous span along a nucleotide sequence from a gene sequence assembly, marking the

beginning and ending position of the annotation on the sequence. Apollo is a genomic annotation viewer and editor, it is freely available at www.bdgp.org/annot/apollo/. The gene of Mus musculus of Glia maturation factor-gamma (Gmfg) is taken into the consideration for doing gene annotation.

The Ensembl database relies on both curated data sources as well as a range of different software tools in their automated genome annotation pipeline.[2]

Beyond Gene Annotation

Once the genes are identified, they need to be categorized in terms of the molecular and cellular organization of the encoded proteins. Such categorization will help in the query of genome databases at the level of specific function of a gene product, the role it plays in cellular processes, and its localization and associations. As more genome databases become available, it is essential that categorization is carried out using a controlled set of shared vocabulary terms to describe the gene products based on current knowledge. The Gene Ontology Consortium is leading the development in this direction by annotating protein sequences from mouse, *Drosophila* and yeast using a standardized ontology. Furthermore, the annotated gene features need to be integrated with other biological knowledge. The Mouse Genome Informatics web site at The Jackson Laboratory provides an illustrative example and provides integrated access to mouse genome, expression and genome sequence databases. The mouse haploid genome is about three billion bases long (3000mb distributed over 20 chromosomes) and therefore equal to the human genome. The annotated genomes used for "completed" genome sequences themselves continue to improve over time as gaps are closed and the sequence assembly becomes more complete. Mapping annotations from one genome assembly version to another is an ongoing effort.

Materials & methods: - For the annotation of the any sequence needed the unannotated sequence. Such unannotated sequences present in the NCBI can be downloaded. We downloaded Glia Maturation Factor – gamma (Gmfg) which was unannotated sequence. By using Apollo (v1.11.3) www.bdgp.org/annot/apollo/ which is annotation tool can annotate the sequence shows features like regulatory sequences protein binding sites of DNA, predicted ORF sequence. It predicts coding and noncoding sequences of the DNA.
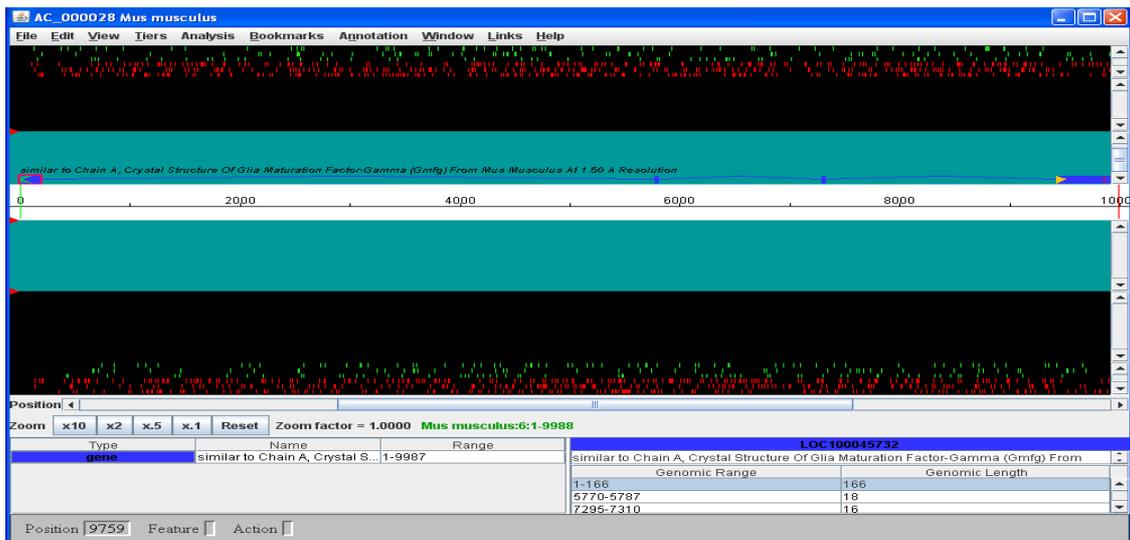
Fig 1: the *mus musculus* gene annotated sequence from 1bp to 9987bp. Green tier indicates start codons & red tiers indicate stop codons.

Result & Discussion:- we can do the different new annotations by using Apollo tool, in this the new annotations are protein binding site annotation, oligonucleotide,gene, point mutation, remarks, regulatory protein, DNA motif annotation has been done.
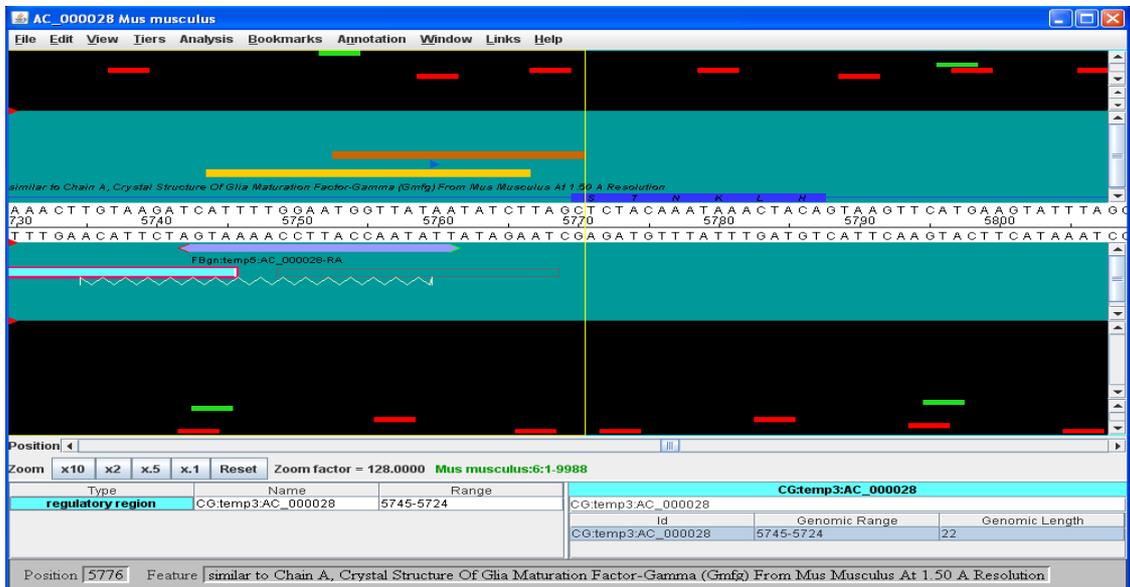

Fig 2: shows the new annotation has done, they are 1) protein binding site 2) oligonucleotide 3) gene 4) point mutation 6) remarks 7) regulatory protein 8) DNA motif...

Table 1:

Type of Annotation	Annotation Id	Range	Length (bp)
Gene annotation	LOG100045732	1bp – 9987bp	9987
Protein binding site	CG:temp4:AC_000028	5744bp – 5766bp	23
Regulatory region	CG:temp3:AC_000028	5745bp – 5724bp	22
Point mutation	CG:temp6:AC_000028	5747bp – 5772bp	26
Oligonucleotide	CG:temp2:AC_000028	5753bp – 5770bp	18
Remark annotation	CG:temp1:AC_000028	5759bp – 5735bp	25
Pseudogene	CG:temp5:AC_000028	5761bp – 5742bp	20
DNA motif	CG:temp7:AC_000028	5768bp -5749bp	20

The Exon Detail Editor (EDE) explains the association of the selected annotated sequences. We can edit the annotated sequences which were positioned at the reverse strand of annotation panel in zoom, in this window it is editing the regulatory region annotation, DNA motif annotation, pseudogene annotation and it can also edits the remark annotation in specific length.
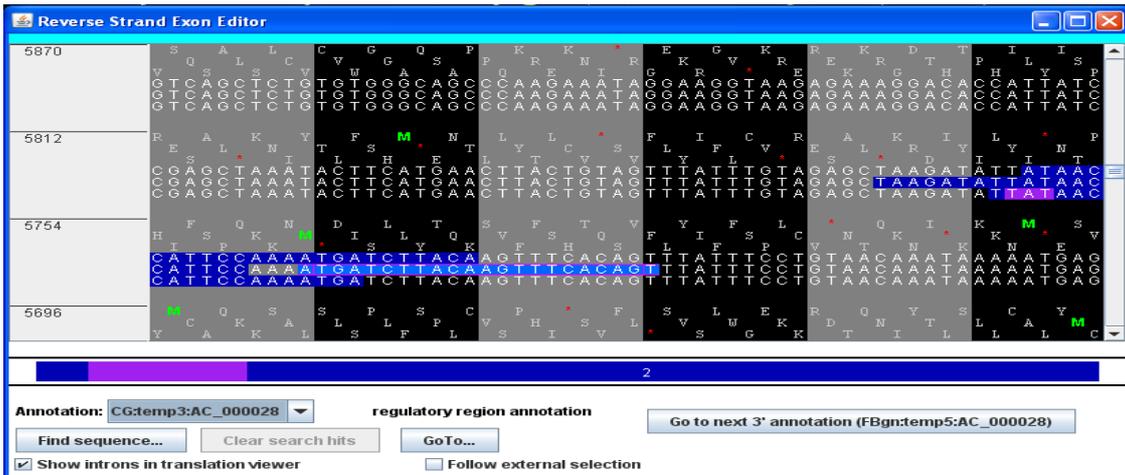
Fig 3: shows the Exon Detail Editor

Table 2:

Annotation Type	Position	Length (bp)
Regulatory region annotation	5745bp – 5724bp	22
Remark annotation	5759bp – 5735bp	25
DNAmotif annotation	5768bp – 5749bp	20
Pseudogene annotation	5761bp – 5741bp	20

We can edit the information about annotation from the "Annotation info" editor; we can also edit the annotation tree through this window it provides an alternate way to explore the annotations, it allows browsing and editing of text information, including comments, associated with annotations.
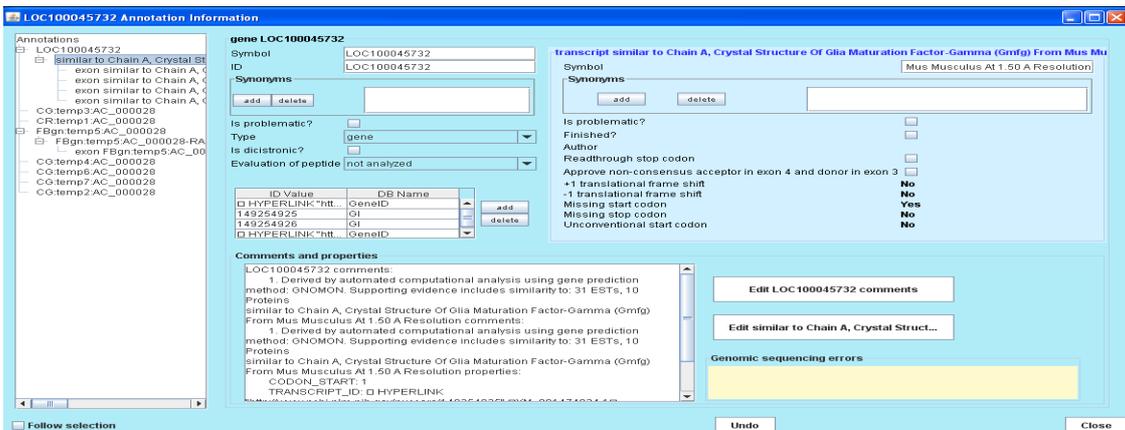

Fig 4: Annotation Info Editor

Conclusion: - In this paper, we have done the annotation of Glia Maturation Factor-Gamma (Gmfg) by using Apollo annotation curation tool. We have generated the different types of annotations in a single gene sequence. The annotations like Regulatory region annotation, gene, protein binding site annotation, oligonucleotide, DNA motif, point mutation etc. we can predict the how much sequence has annotated the position and length of the annotated sequence. The restriction enzymes and restriction sites also can predict by using this Apollo tool.

Acknowledgement: - we thank sincerely to Mr.Deshmukh P.S for his guidance and sugessions, and thank to Mr. Kupteker R.S for his encouragement. Lastly, we want to thank to our team members.

References:-
1. Li, R.; Zhu, H.; Ruan, J.; Qian, W.; Fang, X.; Shi, Z.; Li, Y.; Li, S. *et al.* (Feb 2010). "De novo assembly of human genomes with massively parallel short read sequencing.". *Genome Res* 20 (2): 265-72. doi:10.1101/gr.097261.109. PMID 20019144.
2. Rasmussen, M.; Li, Y.; Lindgreen, S.; Pedersen, JS.; Albrechtsen, A.; Moltke, I.; Metspalu, M.; Metspalu, E. *et al.* (Feb 2010). "Ancient human genome sequence of an extinct Palaeo-Eskimo.". *Nature* 463 (7282): 757-62. doi: 10.1038/nature08835. PMID 20148029.

Web References:-

Apollo tool:www.bdgp.org/annot/apollo/

Gene Sequence: http://www.ncbi.nlm.nih.gov/nuccors/Ac_000028./

Slide share: http://www.sldeshare.net/ntalspangh/how-to-annotate

Annotation: http://en.wikipedia.org/wiki/Annotation

Glia maturation factor (Gmf):http://wiki.org/wiki/Glia-Maturation-Factor