

Abstract

Background: It is well established that raw sequence data from a sequencing project is of no value unless it is annotated with structural and functional information. As annotation relies on the available tools and knowledge at the time of annotating, with time genomes need re-annotating to reflect improvements in our knowledge and available tools. There is huge scope for annotation work as next generation sequencing has vastly increased the sequencing capacity and many genome sequencing projects are being undertaken. The traditional methods of automatic annotation and manual annotation are flawed by their low accuracy and high cost respectively. The community annotation projects have sprung up to take the middle ground between the above. But the science community is slow to recognize its role in genome annotation. The main concern is the qualifications and competencies of the annotators and lack of validation of results in current systems. Though there are many free and open source platforms available, lack of support and ability to customize makes bespoke platform development worthwhile.

Methods: A web based platform was developed for community genome annotation. The platform was developed on Java 2 Enterprise Edition platform with Java Server pages delivering the dynamic web content. User authentication and authorization was implemented to secure the platform. A relational database was used to store data. A dynamic coordinate system was implemented to allow sequence revisions without affecting the annotations already done. A training module was implemented to provide a targeted collection of educational material for the users to access. All annotation works were linked to the evidence as well as the annotators. Annotations can be forwarded to designated curators to access the quality before making the annotations available to the public. The annotations were linked to the www.srilankangenome.net site for visualization. To increase the productivity of the annotation work, frequently used functionalities of bioinformatics tools were provided through external servers.

Conclusion: The platform was implemented within the local area network of the bioinformatics laboratory at the Faculty of Medicine, Colombo. The desired functionality of a multiuser environment was achieved. Piloting for the project is planned to annotate chromosome 21 of the Sri Lankan Genome.