ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

DEVELOPMENT AND EVALUATION OF AN INTERACTIVE WEB PORTAL FOR DESIGNED PRIMERS AND MOLECULAR DISEASE MARKERS

Beatrice A. Ongadi[1,2], Johnson Kinyua[2], Luna Kamau[3], Daniel Kariuki[2] Damaris Matoke[3], Francis Kimani[3]

1. ICT Department, Kenya Medical Research Institute
2. Department of Biochemistry, Jomo Kenyatta University of Agriculture and Technology
3. Centre for Biotechnology Research and Development, Kenya Medical Research Institute

Manuscript Info

Manuscript History:

Received: 12 January 2015
Final Accepted: 22 February 2015
Published Online: March 2015

Key words:

Primer/Molecular Marker/Repository/interactive/portal

Corresponding Author

Beatrice A. Ongadi

Abstract

Background

Bioinformatics being an interdisciplinary field that develops and improves on methods for storing, retrieving, organizing and analyzing biological data has a major activity in developing software tools to generate useful biological knowledge. Interactive web portal for various primers and molecular disease markers will bring together the now scattered information from diverse sources into a single user friendly interface in a uniform and organized format, with links to other related web resources and or tools.

Objectives

To develop and evaluate an interactive web portal for designed primers and molecular disease markers.

Methods

To achieve this; standard system development life cycle, different programming languages and utilities including; PHP, MYSQL, HTML, CSS, Apache web server, Javascript and AJAX are used to develop an interactive web portal and repository that allows users to sign up, upload and cite relevant proposals, search for and use already designed primers and molecular disease markers.

Results

A Web based growing repository and portal of designed primers and molecular disease markers.

Conclusion

Notably a database that acts as a repository of different designed primers is still unrepresented hence posing a challenge to scientists in their research work; despite the availability of various primer design tools. Currently, researchers would have to either design their own primers or alternatively search through papers and rely on primers designed by other scientists without relevant endorsement by knowledgeable experts. This repository aims at availing relevant authenticated primer details without starting design work from start.

I. INTRODUCTION

Every living organism is dependent on three types of macromolecules Nucleic acids, Proteins and Lipids. DNA is an informational macromolecule that encodes the complete set of instructions that are required to assemble, maintain, and reproduce every living organism. Proteins are responsible for catalyzing the myriad biochemical reactions that are required to provide food and energy for every organism, and to carry out all of the additional functions of any given organism, for example movement, neural function, vision, or photosynthesis. RNA is multifunctional, single stranded but its primary responsibility is to make proteins, according to the instructions encoded within a cell's DNA.

The Central Dogma of molecular biology states that the information in biological systems only flows one way from DNA to RNA to proteins where DNA acts as the template for both its own replication and for the transcription to RNA and, with subsequent maturation to mRNA, which then serves as a template for translation into a protein. This theory was developed by Francis Crick, one of the co-discoverers of the structure of DNA, in 1958, and made famous by its publication in 'Nature' in 1970[1]. Study on DNA has played a key role in the advancement of science and medicine with applications ranging from diagnosis, drug discovery to vaccine development.

1.1 Statement of the Problem

Access, efficiency and accuracy of already designed primers continue to be a major impediment for researchers and their research work. While different web based design tools have made the primer design process a little easier and faster than ever, the challenge experienced is that each and every time the design work has to begin from scratch. Currently, researchers would have to either design their own primers or alternatively search through papers and rely on primers designed by other scientists without relevant endorsement by knowledgeable experts. This in itself may result into inaccurate results due to inaccurately reported or written information. The fact that constantly large quantities of information generated from different bioinformatics design tools and research projects are not well-maintained a lot of time is spent through irregular literature searches that may give misleading outputs.

This study seeks to develop and evaluate an interactive web portal for designed primers and molecular disease markers to facilitate handling different aspects of diseases. It will make it easy and faster for different scientists to access primers and molecular markers for their work. Moreover the portal is destined to have a group of competent molecular biologists to curate all the uploaded details and a unique accession number tagged to each; for ease of reference in a single user interface.

The provision of information in one single place in a more organized format for ease of access and retrieval is worth an investment on time and resources. An example of an equivalent repository that has been of great benefit to biologist is GEO (Genome Expression Omnibus), which is a public functional genomics data repository supporting Minimum Information About a Microarray Experiment (MIAME) compliant data submissions. GEO provides a flexible and open design that facilitates submission, storage and retrieval of heterogeneous data sets from high-throughput gene expression and genomic hybridization experiments. GEO was not intended to replace in-house gene expression databases that benefit from coherent data sets, and which are constructed to facilitate a particular analytic method, but rather complement these by acting as a tertiary, central data distribution hub [2]. Relevant tools are provided to help users query and download experiments and curated gene expression profiles. The same benefits will be realized from the repository that is developed in this project.

1.2 Research Objective

The main objective of this work is to develop and evaluate an interactive web portal for designed primers and molecular disease markers. The specific objectives in this regard are:

i. To carry out requirement analysis for the development of the primer web portal.
ii. To design and create a database for the primer web portal.
iii. To carry out a performance evaluation of the primer web portal using online questionnaire.

II. MATERIALS AND METHODS

This was an exclusive study involving two main activities namely, design and development as well as a two level evaluation principle that covered a period not less than twelve months as follows:

2.1 Design and Development

Design and development of this web portal involved the following main stages that conformed to the standard system development life cycle; Concept development, Planning, Requirement analysis, Software design, Software development, Integration and Testing, Implementation, and finally Maintenance of the end product.

Figure: 1.1: System Development Life Cycle***(http://satyamtech.com/service-methodology/)***

2.2 System concept development

The concept was conceived after realizing the need of having a database for designed primers and molecular disease markers. This in real sense had never been developed and shared world over. Literature review and oral interviews revealed that molecular biologists relied on manual search through papers and using unauthenticated markers or alternatively starting design work from scratch. This has accumulated into inaccuracy; a lot of time spent and cost inefficiency.

2.3 Planning and Requirement Analysis for the Primer Web Portal

This involved both functional requirements and non-functional requirements determination. During this process a combination of three main activities was used as follows:

- **Eliciting requirements**: This involved talking to stakeholders who are molecular biologist as well as students who have experience in molecular work in order to gather the kind of information they wished to get in one page at a click of a button relating to primers and molecular disease markers which may have a positive impact in their day to day work.
- **Analyzing requirements**: With reference to relevant literature it was determined whether the stated requirements were clear, complete, consistent and unambiguous, and resolved any deceptive conflicts that arose.
- **Recording requirements**: The requirements were documented in table as follows:

PRIMER	**MICROSATELLITE**	**SNPs**
Part A	**PART A**	**Part A**
Abstract Title	Abstract Title	Abstract

		Title
Abstract	Abstract	Abstract
Disease Category	Vector	Disease Category
Disease	Chromosome (This can be many for one vector)	Disease
Gene	Marker (Can be many for one chromosome)	Gene
Primer Forward	Cytological locations (Every marker has a location)	Primer Forward
Primer Reverse	Cumulative Distances (Every marker has a distance)	Primer Reverse
Primer Alias	Primer Forward	Primer Alias
Fragment Size	Primer Reverse	Fragment Size
Annealing Temperature	Annealing Temperature	Annealing Temperature
Strain	Repeats	Strain
PART B	Allele Size	**PART B**
Authors Name	**PART B**	Authors Name
Phone	Authors Name	Phone
Email	Phone	Email
Institution	Email	Institution
Any Other Information	Institution	Any Other Information
PART C	Any Other Information	**PART C**
Date of Publication	**PART C**	Publication Date
Name of Journal	Date of Publication /Name of Journal	Name of Journal

Table 1.1: Detailed Requirement record.

2.4 Program Design for the Primer Web Portal

The portal was initially prototyped using plain text and interactivity between the major components designed and their functionality determined. This involved information design of the database and the relevant inter-links.

2.5 Design Testing for the Primer Web portal

The designed system was subjected to simple objective verses output evaluation to ensure that all aspects of the portal requirement were captured. A black-box testing was used where the functionality of each simple application was subjected to test without minding its internal structures or workings. This preliminary testing helped to reduce cost and time spent in the final testing stages.

2.6 Program Development of the Primer Web portal

Four main activities formed part of this crucial stage of the portal development. Different programming languages and utilities were applied depending on their functionalities and strengths. Customization was also used in one area

or another to avoid re-inventing the wheel in development. The initial primers and molecular disease markers that were used in this stage were drawn from open literature including books and download and their respective authors cited as per regulations.

2.7 Resource Selection

Resources for the development of this web portal, being Hardware (HP Probook 4330s) pre-installed with Windows 7 Professional and development related softwares including; MySQL, PHP, XHTML, AJAX in good working condition with updated antivirus was prepared and availed.

2.8 Web page Development

The web page was created using PHP embedded with HTML to bring out every aspect of its dynamism, maximum execution and relevant form outputs were achieved.

2.9 Creation of Database

The databases was modeled and created using entity relationship diagrams and represented as a normalized database in MySQL. PHPMyAdmin has been was used to handle the administration of MySQL over the World Wide Web during hosting. HTML used more on the design of the template and its tag used together with CSS to design the user interface.

Algorithms were developed for each function, based on actions to be performed. They were also designed to provide uniform access to heterogeneous autonomous data sources. This involved transforming the algorithm into programming language constructs.

World Wide Web Program Management of the Primer Portal

Apache Web Server was used to serve both static content and dynamic web pages on the World Wide Web. The design of Apache reduces latency and increases throughput, thus insuring consistency and reliability.

3.0 Evaluation of the Primer Web Portal

Overall testing and evaluation of the Web Portal was carried out using three main steps:-

System Testing and Requirement Validation

System component interaction was accurately tested by random entry of data into the system or irregular search of data from the portal and output recorded in a table. Requirement analysis was also carried out after developing the tool. This was to ensure that each of the functionality has been incorporated into the system and works as expected. This testing was boosted by the availability of an online questionnaire that can be accessed voluntarily by prospective users.

Pre-user evaluation and acceptance testing of the Primer Web Portal

The tool was demonstrated and initially pre-evaluated using an online questionnaire in KEMRI Center of Biotechnology Research and Development, ICT Department as well as JKUAT Department of Biochemistry. This was facilitated by restricting initial access to the questionnaire to identified group of users only.

User Evaluation of the Primer Web Portal

A questionnaire was used in evaluating the tool. Data collected was then used to determine the efficiency and effectiveness of referencing already designed primers as opposed to fresh primer design at every instance. Views and suggestions by the users were and are still being used to improve the tool.

Variables

Two variables were evident in this work; they included the use of primer web portal with ease which was independent and the actual contribution to efficiency and effectiveness which was dependent.

DataManagement

The findings from the Likert scale were linked to efficiency and effectiveness of the primer portal. Data was entered into a computer using Ms Excel form. The final data was stored in the following formats: Hard copy; Soft copy (Hard disk, Flash Disk) and in the cloud. Mode or the most frequent response was used as a measure of central tendency. This made the survey results much easier for interpretation. Analysis of variance techniques (Kruskal Wallis Test) was used to analyze responses among the five groups of respondents namely Students, Early Career Researchers, Mid-Career Researchers, Established Researchers and others was used. To simplify the survey data further the four response categories were combined into two nominal categories as "agree' and "disagree" and Chi square test used for analysis.

Confidentiality

The study used data collected through online questionnaire from respondents who voluntarily accepted to participate. To guarantee total confidentiality during collection, analysis and storage, individual accounts were created to log into system and each log session was controlled. Unique activation link was sent to private email addresses provided by the user during sign up and password generation cretaria monitored to ensure total security. Anonymity observed by ensuring that respondents did not have to enter their details in the questionnaire whilst each entry automatically generated a unique identity (ID Number) which did not have the portal users' details. Portal access hitches were reported by email to system administrator.

Intellectual Property Issues

All related Intellectual property issues were channeled through the guidance ofKEMRI Intellectual Property Rights and Technology Transfer Policy. The patenting intended to cover the portal as a tool of bringing together primer sequences and markers being an invention of the investigators from both KEMRI and JKUAT who shall share rights and obligations relating to this invention.

III. RESULTS

The end result of this work is an interactive web portal for designed primers and molecular disease markers also known as "PrimaBank" with the following web pages and different functionalities.

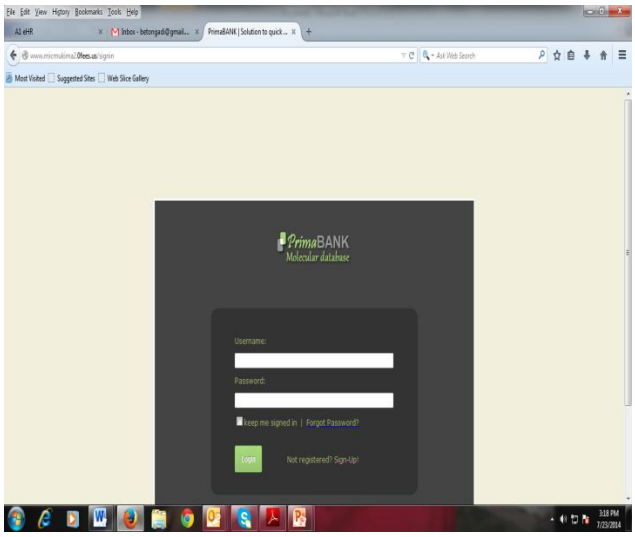

Figure: 1.2: Sign up and Log in Page: Page used by new users as initial step for sign up and also used for Log in procedure.

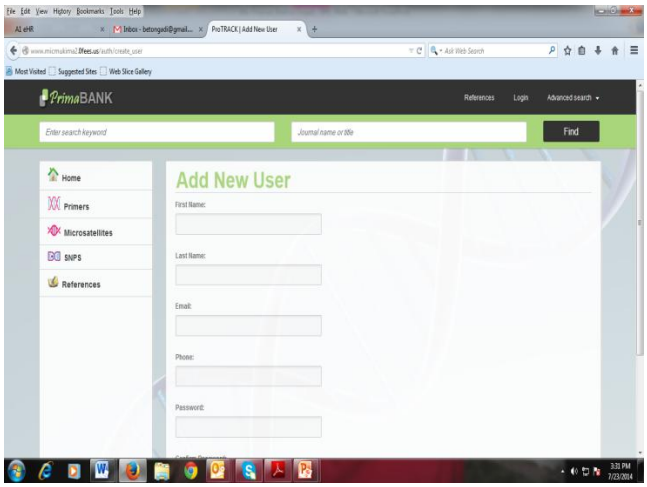

Figure: 1.3: Sign up and Log in Page: Page used by new users to key in their credential to allow portal access

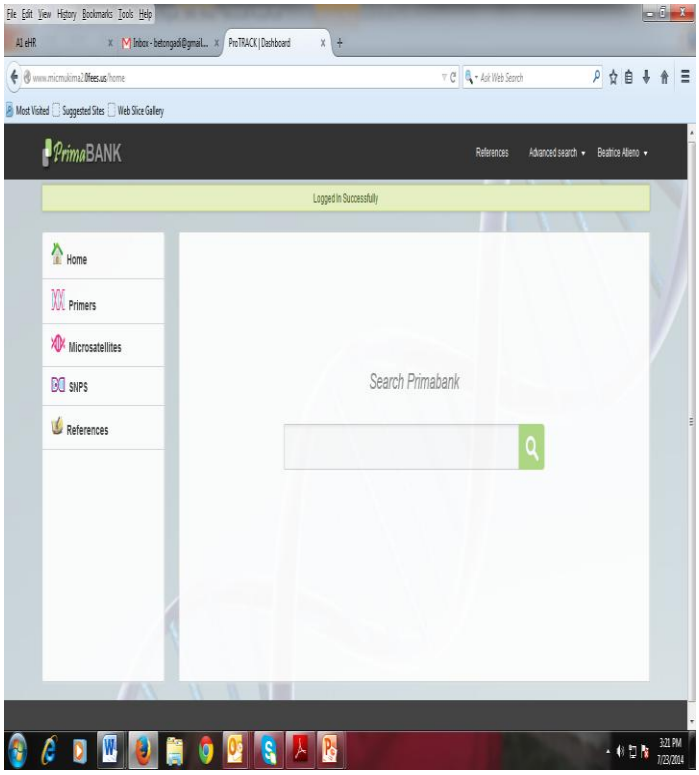

Figure: 1.4: Search Page: Page used to search for primers and molecular disease markers.
(Search for Primers, Microsatellites and SNPs)

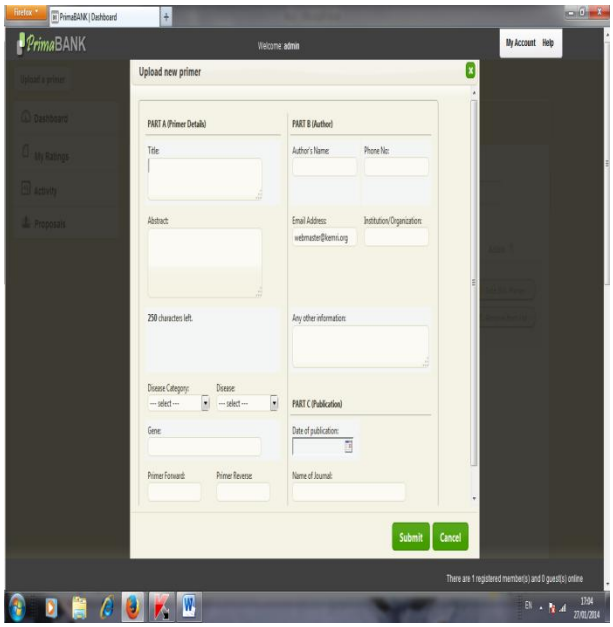

Figure: 1.5: Uploading Page: Page used by molecular biologists to add a primer/ disease marker and submit to the database.

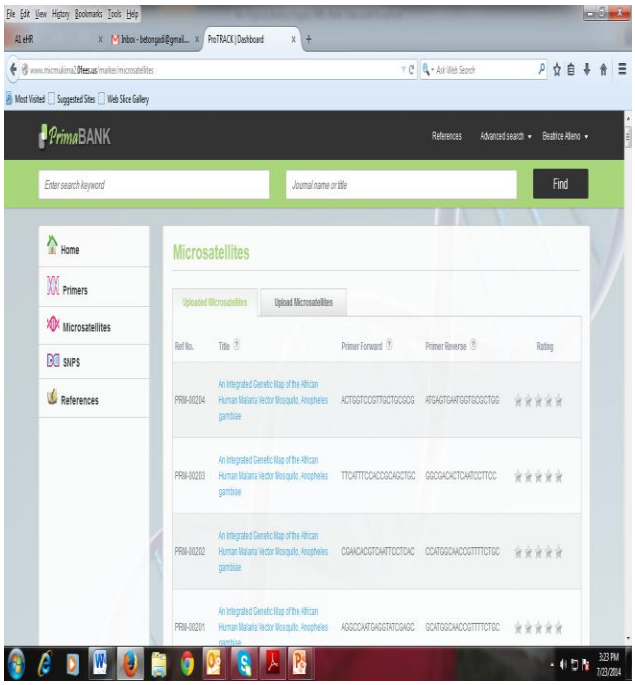

Figure: 1.6: Search Results: Typical search result for microsatellites in general.

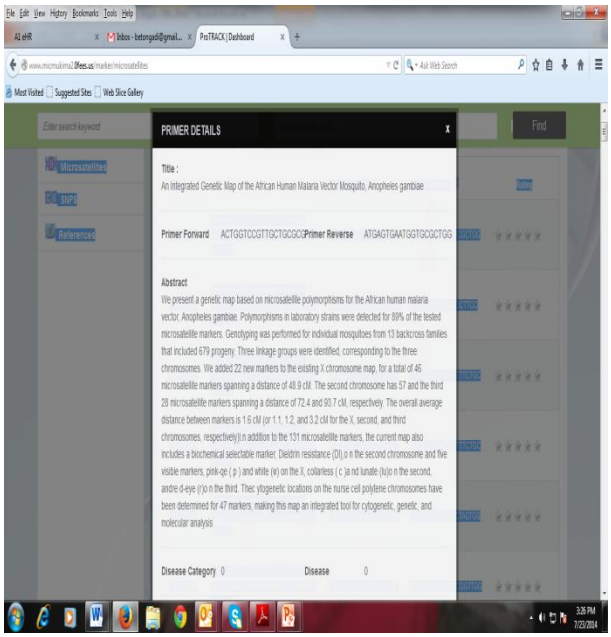

Figure: 1.7: Search Results: Typical search result for a particular microsatellite.

IV. DISCUSSION

MySQL (My Structured Query Language) which is an open source Relational Database Management System has been used to develop this interactive web portal for designed primers and molecular disease markers. This has been achieved since the portal is capable of running on virtually every platform including Linux, Windows, UNIX among others upcoming platforms. In this particular scenario data is stored in separate tables rather than one big store and a logical model with objects such as databases, tables, views, rows and columns offers a flexible programming environment. The reason for this is to speed up access and boost data security. The disadvantage with centralized data storage is that when there is a CIA (Confidentiality, Integrity, and Accessibility) breach, all data is vulnerable. The security of data stored in MySQL has been achieved through six main procedures outlined here as: removing wildcards in the grant tables, using secure passwords, disabling remote access, encrypting client server transmission as well as actively monitoring the MySQL access logs.

To restrict and standardize the search and find process of the web portal; eXtensible Markup Language (XHTML) has been used effectively to ensure that only information meeting the user's search criterion is displayed in a user friendly format. The hypertext functionality of this language helps the user to follow an anchored link to the content whereas the universality allows virtually any computer and platform to read the web pages. The dynamism of the said web pages has been achieved in this work by the use of PHP (Hypertext Preprocessor).In order for the rapid exchange of data with the server and updating of different web pages without unnecessary redundancy. AJAX (Asynchronous Javascript) has been effectively applied in this work

The usage of the web portal is dependent on the guidelines in the documentation also developed together with this system. To gain access this tool one needs to first register as a user and get authentication from systems administrator. The account is then activated via the provided email address. After which a user is now able to add, search and or curate a given primer or any molecular disease marker. (For guidelines on sign up process reference is made to the User Guide attached online). Information gathered during sign up (Name, Username, Email, Phone, and Password) is stored in a database and is referenced any other time a user tries to log in. This is the first security measure the tool observes to ensure that only authenticated persons are able to use the tool. The login details also help in establishment of sessions for the given user so that the interface and functionalities is unique to the user. PHP sessions are therefore applied here. However identified curators have a much higher priority over any other user since their role includes but not limited to authenticating the information entered for further reference by the potential users. All session are kept in a log for reference.

The portal categorizes the molecular markers into three major different groups; Primer, Microsatellite, SNPs. The details captured within the portal include but are not limited to Accession Number (Uniquely issued after curating with the following prefixes: PRM for Primers, MST for Microsatellites and SNP for SNPs), Protocol Title, Principal Investigator, Forward/Reverse primer, Fragment size, Strain, Annealing Temperature, Disease, Category Disease, gene, primer alias, Author details, Date of Publication, name of journal.
Upon successful log in; one is able to search for required details based on the following search criteria: Abstract Title, Disease category, Disease, Gene, Primer Alias, Author and Forward/Reverse Primer. The result that will be received after any search will indicate as either curated or not curated for validity and further referencing. The tool also provides links to other relevant pages.

V. CONCLUSION AND RECOMMENDATION

The interactive web portal for designed primers and molecular diseases is a far much easier way of accessing already designed primers and other molecular markers; this has been evident from the positive responses that are being received from different users. It is recommendable that molecular biologist be encouraged to upload their marker details without hesitation to allow for portal growth.

VI. REFERENCES

[1] Crick F., (1970)Central dogma of molecular biology. *Nature*227 (5258): 561–3.
[2] Edgar R., Domrachev M., Lash A.E., (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*., 30: 207–210. doi: 10.1093/nar/30.1.207.