Importance of Bioinformatics Databases in Biological Research: A Study Conducted Among Research Scholars in Kerala

Shivani K. T.
Library and Information Associate (Intern)
Indian Institute of Management
Calicut- 673 570, Kerala
shivanikt1993@gmail.com

Dr. T. M. Vasudevan
Professor
Department of Library and Information Science
University of Calicut
Malappuram – 673 635, Kerala
amandev44@yahoo.com

Abstract

The present study attempts to seek the relevance of bioinformatics databases in biological research based on their use among biological research scholars in Kerala. The questionnaire survey was used to collect data from research scholars of Indian Institute of Special Research (IISR), Kerala Agricultural University (KAU), Kerala University of Fisheries and Ocean Studies (KUFOS) and biological research departments of Calicut University. The findings of the study reveal that the majority of the research scholars in the biological field use bioinformatics databases forvarious research purposes, especially for data analysis. Research scholars agree that bioinformatics databases are effective tools to improve their research. Lack of training is the main problem faced by them while accessing and using biological databases.

Keywords: *Bioinformatics, Databases, Bioinformatics databases, Biological databases, Biological research, Research scholars in biology, Kerala.*

1. Introduction

The research activities conducted in various fields give birth to a huge amount of unorganized information leading to information explosion. As a result users find it difficult to access the right information resulting in wasting a lot of time. Therefore, there arises the need to organize this information in a meaningful manner for future research and study. The emergence of databases overcomes this crisis. Databases are the systematically arranged collection of related records. Today databases are used in various fields. The databases used in the biological field are called bioinformatics databases or biological databases.

Identifying sources of information is crucial in today's fast-moving world of biology. Not only does it allow researchers to find information rapidly for their experiments, but it also enables biologists a place where their results can be published (Discala et al., 2000). Biological databases, in fact, serve as the libraries of life science information, collected from scientific experiments, published literature, high throughput experiment technology, and computational analysis. They offer scientists the opportunity to access a wide variety of biological data centrally (Patel,Anjaria&Panchal, 2014).

2. Review of Literature

Many studies are conducted on the importance of bioinformatics databases in biological research. Some recent reviews of related literature are arranged in chronological order.

Cochrane et al. (2015) conducted a study on the update of the International Nucleotide Sequence Database Collaboration (INSDC)'s in 2015. The study shows that there is an increase in the growth rate of data submitted to the INSDC from 2012 to 2015. Patel,Anjaria and Panchal (2014) made a study on the bioinformatics database of some leguminous trees in Anand district of Gujarat state of India. The database contains botanical information of tree species and bioinformatics information with analysis at one platform.Rajoka et al. (2014) studied an interactive bioinformatics database -- Medherb. It would be helpful for researchers in computational analysis, and to direct the experimental investigation towards medicinal herbs, thus leading to discoveries in the area of computational biology. Holton, Vijayakumar and Khaldi (2013) jointly examined the current state of bioinformatics in food and nutritional sciences. The findings reveal that food and nutritional research is greatly enhanced by bioinformatics. James et al. (2013) conducted a study on DIACAN, an

integrated database for anti-diabetic and anti-cancer medicinal plants. It is a value-added database which helps in the process of drug discovery. Zhang et al. (2011) made a study on apple gene function and gene family database. This database provides interactive user utilities and accurate information, thus making it useful for improving productivity in apple research.Barret et al. (2011) examined the role of BioProjectandBioSample databases on the organization of metadata. The study found that the BioProject database facilitates organization and classification of project data submitted to NCBI, EBI and DDBJ databases. BioSample database captures descriptive information about the biological sample investigated in projects. Picket et al. (2011) studied about ViPR (Virus Pathogen Database and Analysis Resource). The powerful suite of resources provided within the ViPR helps the virology research community for tracing the developments in diagnostics and therapeutics fields. Buchen et al. (2011) investigated the role of translational bioinformatics in drug discovery. The study reached to a conclusion that translational bioinformatics has a crucial role in the organization of enormous amount of data generated during the many phases of drug discovery. Amui et al. (2011) conducted a study on Plant-anti venom database of medicinal plants. The creation of this database is a solution for organizing information about medicinal plants with anti-venom properties, thereby effective in assisting research in this area.

3. Objectives of the Study

The main objectives of the study are:

1. To find out the extent of use of bioinformatics databases by research scholars in biology.
2. To determine differences in use and awareness of bioinformatics databases among various institutions in Kerala.
3. To determine differences in using bioinformatics databases in different fields of biological research.
4. To find out the relationship between bioinformatics databases and research development.
5. To determine the level of satisfaction with the use of bioinformatics databases by research scholars.
6. To find out the hindrances faced by research scholars while accessing and using bioinformatics databases.

4. Methodology

The population of the study is confined to four prominent institutionsin Kerala namely Indian Institute of Spices Research (IISR), Calicut, Kerala Agricultural University (KAU), Thrissur, Kerala University of Fisheries and Ocean Studies (KUFOS), Kochi, and Biological Research Departments in Calicut University, Tenhipalam such as Life Sciences, Botany, Zoology, Biotechnology and Health Sciences. A structured questionnaire was used to collect data from the research scholars belonging to these institutions. The investigator distributed 185 questionnaires out of which 133 were received with a response rate of 71.89%.

5. Analysis of Data

The received data were tabulated in Microsoft Excel. The analysis was done with the help of SPSS and excel. Tables and graphs are used to represent the collected data.

5.1 Use of Bioinformatics Databases

Bioinformatics databases are the databases that store biological information in a useful and meaningful manner. The following table shows the use of bioinformatics databases for research purpose by research scholars of four prominent institutions in Kerala.

The analysis shows that large majority of the respondents in IISR (90%), KUFOS (86.4%) and Calicut University (81.6%) use bioinformatics databases for their research work. Majority of the respondents (60%) in KAU also depend on bioinformatics databases for their research work. The overall result shows that a sizeable percentage of the respondents (78.9%) use bioinformatics databases for research purpose. It is also startling to observe that about 21 percent of the research scholars do not make use of the databases in spite of their importance in research.

Chi-Square test is used to compare the use of bioinformatics databases by the research scholars of IISR, KAU, KUFOS, and Calicut University. Test results show that there is a significant difference among differentinstitutions

Table 1

Use of Bioinformatics databases

Opinion	IISR	KAU	KUFOS	CU	Total
Yes	9 (90%)	15 (60%)	19 (86.4%)	62 (81.6%)	105 (78.9%)
No	1 (10%)	10 (40%)	3 (13.6%)	14 (18.4%)	28 (21.1%)

Chi-Square Value = 7.8, d.f=3

in their use of bioinformatics databases for research purposes since p value= 0.66>0.05.

5.2. Use of Bioinformatics Databases in Different Fields of Biological Research

In biology, there are different subfields. The use of bioinformatics databases may vary from one field to another. The table 2 presents the differences in using bioinformatics databases in various biological fields.

It is clear from the above table thatall the respondents in biotechnology use bioinformatics databases for their research. This is followed by molecular biology (92.30%), taxonomy (89.47%), microbiology (84.61%) and genetics (82.35%). At the same time, the share of the respondents from agriculture (65%) and other fields of biology (56.66%) who make use of bioinformatics databases is comparatively less with 65% and 56.66% respectively.

5.3 Awareness of Bioinformatics Databases

Table 3 shows the awareness of different bioinformatics databases among research scholars in biology.

It is evident from the table that all the respondents in IISR are aware of NCBI and GENBANK. In KAU also NCBI (86.7%) and GENBANK (73.3%) are more popular than others. The same phenomenon is visible in KUFOS and Calicut University. PIR appears to be the least popular among the seven databases. Chi-Square test is used to compare the awareness about different bioinformatics databases among research scholars of various institutions. From the Chi-Square test, it is revealed that as far as awareness is concerned, there is association among the various institutions only in the case of EMBL.

5.4 Use of Bioinformatics Databases

There are several databases available in the biological field for different purposes. Table 4 shows the use of different bioinformatics databases among research scholars of various institutions.

Table 2

Use of Bioinformatics databases in different fields of Biological research

Opinion	Gene-tics	Biotech-nology	Agricul-ture	Molecular biology	Micro-biology	Taxo-nomy	Others
Yes	14 (82.35%)	21 (100%)	13 (65%)	12 (92.30%)	11 (84.61%)	17 (89.47%)	17 (56.66%)
No	3 (17.64%)	0 (0%)	7 (35%)	1 (7.69%)	2 (15.38%)	2 (10.52%)	13 (43.33%)

The analysis shows that all the respondents in IISR use NCBI. Though all the respondents from IISR reported to be aware of GENBANK, only one third were found to use it. In KAU, only NCBI and GENBANK were found to be used by more than half of the respondents (53.3%). Another observation is that NCBI has emerged as the most used database in all institutions, though its share of respondents is less in KAU. In other words, the overall result reveals

Table 3

Awareness of Bioinformatics Databases

Institution	NCBI	DDBJ	GENBANK	EMBL	UNIPROT	SWISS-PROT	PIR
IISR	9 (100%)	5 (55.6%)	9 (100%)	6 (66.7%)	6 (66.7%)	6 (66.7%)	4 (44.4%)
KAU	13 (86.7%)	7 (46.7%)	11 (73.3%)	4 (26.7%)	6 (40%)	5 (33.3%)	6 (40%)
KUFOS	18 (94.7%)	14 (73.7%)	17 (89.5%)	14 (73.7%)	12 (63.2%)	14 (73.7%)	9 (47.4%)
CU	60 (96.8%)	37 (59.7%)	56 (90.3%)	44 (71%)	36 (58.1%)	40 (64.5%)	31 (50%)
Chi-Square value	3.213	2.670	4.805	**11.262**	2.436	6.576	0.527
P value	0.360	0.445	0.187	**0.010**	0.487	0.087	0.913

Table 4

Bioinformatics Databases according to Use

Institution	NCBI	DDBJ	GENBANK	EMBL	UNIPROT	SWISSPROT	PIR	OTHERS
IISR	9 (100%)	0 (0%)	3 (33.3%)	1 (11.1%)	2 (22.2%)	1 (11.1%)	0 (0%)	0 (0%)
KAU	8 (53.3%)	3 (20%)	8 (53.3%)	0 (0%)	3 (20%)	2 (13.3%)	0 (0%)	4 (26.7%)
KUFOS	17 (89.5%)	2 (10.5%)	12 (63.2%)	2 (10.5%)	0 (0%)	0 (0%)	1 (5.3%)	0 (0%)
CU	55 (88.7%)	1 (1.6%)	23 (37.1%)	2 (3.2%)	2 (3.2%)	3 (4.8%)	1 (1.6%)	5 (8.1%)
Chi-Square value	**14.164**	**8.98**	4.978	3.265	**10.323**	3.342	1.641	**8.913**
P value	**0.003**	**0.030**	0.173	0.353	**0.016**	0.342	0.650	**0.030**

that the majority of the research scholarsdepend on NCBI for their research work.

Here, the chi-square test is used to compare the use of different bioinformatics databases in various institutions. The P value indicates that chi-square results show a significant relationship between research scholars in different institutionsin the case of NCBI, DDBJ, UNIPROT and OTHERS.

5.5 Purpose of Using Bioinformatics Databases

Research scholars use bioinformatics databases for different purposes. Table 5 shows depicts the purpose of using bioinformatics databases.

It can be seen from the table that a large majority of the respondents (88.88%) in IISRuse bioinformatics databases for data analysis followed by data downloading (77.77%), data interpretation (55.55%) and Phylogeny checking (55.55%). In KAU all the respondents were found to use databases for literature review followed by up-to-date information (93.33%) and data downloading (80%). In KUFOS, the purposes of use are in the order of data analysis (100%), literature search (94.73%), data interpretation (89.47%) and up-to-date information (89.47%). In the departments of Calicut University the purposes are in the order of data analysis (96.77%), literature review (91.93%) and up-to-date information (91.93%). The average of all universities shows that the

Table 5

Purpose of using Bioinformatics databases

Purpose	IISR	KAU	KUFOS	CU	Total
Up-to-date Information	4 (44.44%)	14 (93.33%)	17 (89.47%)	57 (91.93%)	92 (87.61%)
Submit research findings	3 (33.33%)	9 (60%)	14 (73.68%)	43 (69.35%)	69 (65.71%)
Data analysis	8 (88.88%)	8 (53.33%)	19 (100%)	60 (96.77%)	95 (90.47%)
Data Interpretation	5 (55.55%)	6 (40%)	17 (89.47%)	49 (79.03%)	77 (73.33%)
Data sharing	2 (22.22%)	7 (46.66%)	12 (63.15%)	45 (72.58%)	66 (62.85%)
Data downloading	7 (77.77%)	12 (80%)	15 (78.94%)	54 (87.09%)	88 (83.80%)
Phylogeny checking	5 (55.55%)	0 (0%)	11 (57.89%)	35 (56.45%)	51 (48.57%)
Drug designing	1 (11.11%)	0 (0%)	2 (10.52%)	10 (16.12%)	13 (12.38%)
Literature search	3 (33.33%)	15 (100%)	18 (94.73%)	57 (91.93%)	93 (88.57%)
Others	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)

Table 6

Effect of Bioinformatics databases on biological research

Opinion	IISR	KAU	KUFOS	CU	Total
Yes	9 (100%)	14 (93.33%)	19 (100%)	59 (95.16%)	101 (96.19%)
No	0 (0%)	1 (6.66%)	0 (0%)	3 (4.83%)	4 (3.80%)

purposes of using the databases are in the order data analysis (90.47%), literature review (88.57%), up-to-date information (87.61%), data downloading (83.80%) and data interpretation (73.33%). Drug designing (12.38%) is the least cited purpose in all institutions.

5.6 Effect of Bioinformatics Databases on Biological Research

Table 6 shows the opinion of research scholars towards the effectiveness of bioinformatics databases on research in the biological field.

It is evident from the above table that all the respondents in IISR and KUFOS more than 90 percent of other universities think that bioinformatics databases are effective tools to improve their research. Therefore, the overall result shows that 96.19 percent of the respondents are of the view that bioinformatics databases are helpful for various research works conducted in the biological field.

5.7 Comparison of Institutions on Dependence on Bioinformatics Databases

Different institutions may depend on bioinformatics databases for different purposes at different levels. Table 7 shows that the difference in various institutions in their dependence on bioinformatics databases.

A one-way ANOVA was conducted to compare the extent to which the research scholars from different institutions depend on bioinformatics databases for research purpose. The results show that there is no significant difference in the extent to which the research scholars of different institutions depend on bioinformatics databases at 0.05 level of significance for the conditions [F (3,101)=0.771, p=0.513<0.05].

Table 7

Comparison of institutions and dependence on Bioinformatics databases

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
IISR	9	2.2469	.33691	.11230	1.9879	2.5059	1.67	2.78
KAU	15	2.1704	.37998	.09811	1.9599	2.3808	1.44	2.67
KUFOS	19	2.0994	.38826	.08907	1.9123	2.2866	1.44	2.67
CU	62	2.2491	.40921	.05197	2.1452	2.3530	1.11	3.00
Total	105	2.2106	.39492	.03854	2.1342	2.2870	1.11	3.00

ANOVA

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	.363	3	.121	.771	.513
Within Groups	15.857	101	.57		
Total	16.220	104			

Table 8

New trends of research invoked by Bioinformatics databases

New trends of research	IISR	KAU	KUFOS	CU	Total
Production of a massive amount of information	4 (44.44%)	8 (53.33%)	15 (78.94%)	38 (61.29%)	65 (61.9%)
Difficult to evaluate	0 (0%)	2 (13.33%)	5 (26.31%)	6 (9.67%)	13 (12.38%)
Collaborate with colleagues	1 (11.11%)	8 (53.33%)	4 (21.05%)	17 (27.41%)	30 (28.57%)
Share information constraints	4 (44.44%)	8 (53.33%)	1 (5.26%)	26 (41.93%)	39 (37.14%)
Others	0 (0%)	0 (0%)	1 (5.26%)	0 (0%)	1 (0.95%)

5.8 New Trends of Research Invoked by Bioinformatics Databases

Table 8 shows the opinion of the respondents on the ability of bioinformatics databases to keep them abreast of the new trends of research.

The overall result shows that majority of the respondents (61.9%) agree that bioinformatics databases can invoke new research trends by producing/amassing vast amount of information. Therefore, the results reveal that bioinformatics databases enable them to keep abreast of the latest trends of research in the biological field.

5.9 Satisfaction Level of Research Scholars on Bioinformatics Databases

The level of satisfaction on bioinformatics databases may vary from one institution to another or from one research scholar to other. Table 9 shows the level of satisfaction of research scholars on bioinformatics databases.

It is evident from the table thata major chunk of the respondents in IISR (77.77%) is only 'partially satisfied' with the use of bioinformatics databases.While about 60 percent of the respondents of Calicut University is 'partially satisfied' their share is 53.33 percent in Kerala Agricultural University and a little more than 42 percent in KUFOS. Those who are 'fully satisfied' is more in KUFOS (57.89%), their share is the least (22.22%) in IISR. Those who are 'not satisfied' constitute a minuscule. The overall result shows that a little more than 57 percent are 'partially satisfied' with the use of bioinformatics databases. As shown earlier, nearly 21 percent are not in the habit of using databases.

Table 9

Satisfaction level of research scholars on Bioinformatics databases

Level of Satisfaction	IISR	KAU	KUFOS	CU	Total
Fully Satisfied	2 (22.22%)	6 (40%)	11 (57.89%)	24 (38.70%)	43 (40.95%)
Partially Satisfied	7 (77.77%)	8 (53.33%)	8 (42.10%)	37 (59.67%)	60 (57.14%)
Not Satisfied	0 (0%)	1 (6.66%)	0 (0%)	1 (1.61%)	2 (1.90%)
Not Used	1 (10%)	10 (40%)	3 (13.63%)	14 (18.42%)	28 (21.05%)

ANOVA

	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	.611	3	.204	1.291	.282
Within Groups	15.928	101	.158		
Total	16.538	104			

A one-way ANOVA was conducted to compare the satisfaction level of research scholars on the facilities provided by bioinformatics databases. The results show that there is no significant difference in the satisfaction level of research scholars in the use of bioinformatics databases among different institutions at 0.05 level of significance for the conditions [F (3,101)=1.291, p=0.282<0.05].

5.10 Limitations of Using Bioinformatics Databases

Advantages notwithstanding, bioinformatics databases may have their own limitations also. Table 10 shows the problems/ limitations faced by the research scholars while accessing bioinformatics databases.

It can be seen from the table that majority of the respondents in IISR (77.77%), KUFOS (63.15%) and Calicut University (61.29%) think that lack of training is the main problem in making use of the bioinformatics databases. At the same time, two-third of the respondents (66.66%) in KAU think that data restriction in downloading is the major problem. The overall result shows that while 60.95%think that lack of training is the major problem, about 45 percent consider data restriction as the hitch.

6. Major Findings

- Majority of the research scholars in the biological field use bioinformatics databases for research purpose.
- There exists a significant difference between different institutions in their use of bioinformatics databases.
- Use of bioinformatics databases is very high in the field of biotechnology and molecular biology.
- NCBI and GENBANK are the most popular databases among research scholars, and NCBI is the most commonly used one.
- The main purpose of using bioinformatics databases is data analysis followed by literature review.
- Most of the research scholars think that bioinformatics databases are useful tools in improvingbiological research.
- Bioinformatics databases can invoke new trends of research.
- Majority of the research scholars are 'partially satisfied' with the use of bioinformatics databases.
- Lack of training is the major problem encountered by research scholars while accessing bioinformatics databases.

7. Conclusion

In this study, an attempt was made to find out the usefulness of bioinformatics databases in biological

Table 10

Limitations of using Bioinformatics databases

Problems	IISR	KAU	KUFOS	CU	Total
Copyright issues	1 (11.11%)	4 (26.66%)	7 (36.84%)	25 (40.32%)	37 (35.23%)
Lack of training	7 (77.77%)	7 (46.66%)	12 (63.15%)	38 (61.29%)	64 (60.95%)
Data restricted to download	5 (55.55%)	10 (66.66%)	6 (31.57%)	26 (41.93%)	47 (44.76%)
Cost	2 (22.22%)	3 (20%)	2 (10.52%)	19 (30.64%)	26 (24.76%)
Others	0 (0%)	0 (0%)	1 (5.26%)	0 (0%)	1 (0.95%)

research. The results of the study reveal that the majority of the research scholars in the biological fielduse bioinformatics databases for various research activities. At the same time about one fifth of the respondents are not in the habit of using bioinformatics databases. Lack of training is pointed out to be the main problem faced by research scholars while using bioinformatics databases in all institutions. This calls for the research institutions to train the research scholars for the efficient use of bioinformatics databases by conducting training programmes, seminars, workshops, conferences, etc.In other words, there is an urgent need for imparting information literacy training by the information professionals in all the institutions. Research works conducted in the biological field play a crucial role in the creation of a prosperous society. To accomplish this goal researchers need accurate and reliable information expeditiously. It can be achieved with the help of bioinformatics databases which not only provide the latest information but also facilitate submission of research output into it.

References

Amui, S.F. et al. (2011). Plantantivenom: Database of anti-venom medicinal plants. *Electronic Journal of Biotechnology*, *14*(1), 6-7.

Barrett, T. et al.(2011). BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Research*, *40*(D1), D57-D63.

Buchan, N. S. et al.(2011). The role of translational bioinformatics in drug discovery. *Drug Discovery Today*, *16*(9),426-434.

Cochrane, G. et al.(2015).The international nucleotide sequence database collaboration. *Nucleic Acids Research*, *44*(D1), D48-D50

Discala, C. et al. (2000).DBcat: Acatalog of 500 biological databases. *Nucleic Acids Research*, *28*(1), 8-9.

Holton, T. A., Vijayakumar, V., &Khaldi, N. (2013). Bio informatics:Current perspectives and future directions for food andnutritionalresearch facilitated by a Food-Wiki database.*Trends in Food Science & Technology*, *34*(1), 5-17.

James, P. et al. (2013). DIACAN: Integrated database for anti-diabetic and anticancer medicinal plants. *Bioinformation, 9*(18), 941.

Patel, S., Anjaria, K., & Panchal, H. (2014). Bioinformatics databaseof some leguminous trees in Anand district of Gujarat state in India. *ScholarsJournal of Agriculture and Veterinary Sciences* (SJAVS), *1*(2), 50-57.

Pickett, B. E. et al. (2011). ViPR: An open bio informatics database and analysis resource for virology research. *Nucleic Acids Research, 40*(D1), D593-D598

Rajoka, M. I. et al. (2014). Medherb: An interactive Bioinformatics database and analysis resource formedicinally important herbs. *Current Bioinformatics, 9*(1*)*, 23-27.

Zhang, S. et al. (2013). Apple gene function and gene family database: an integrated bioinformatics database for apple research. *Plant Growth Regulation, 70*(2), 199-206.