International Journal of General Medicine

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Awareness and Predictors of the Use of Bioinformatics in Genome Research in Saudi Arabia

Lamya Alomair[1,2], Mostafa A Abolfotouh[2,3]

[1]AI and Bioinformatics Department, King Abdullah International Medical Research Center (KAIMRC), Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia; [2]King Saud Bin-Abdulaziz University for Health Sciences (KSAU-HS), Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia; [3]Research Training and Development Section, King Abdullah International Medical Research Center (KAIMRC), Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia

Correspondence: Mostafa A Abolfotouh, Research Training and Development Section, King Abdullah International Medical Research Center (KAIMRC), Ministry of National Guard-Health Affairs, P.O. Box 3660, Mail Code 1515 (KAIMRC), Riyadh, 11481, Saudi Arabia, Tel +966 11 429-4460; +966 503659204, Fax +966 11429-4440, Email mabolfotouh@gmail.com

Background: With the advances in genomics research, many countries still need more bioinformatics skills. This study aimed to assess the levels of awareness of bioinformatics and predictors of its use in genomics research among scientists in Saudi Arabia.

Methods: In a cross-sectional survey, 309 scientists of different biological and biomedical specialties were subjected to a previously validated e-questionnaire to collect data on (1) Knowledge about bioinformatics programming languages and tools, (2) Attitude toward acceptance of bioinformatics resources in genome-related research, and (3) The pattern of information-seeking to online bioinformatics resources. Logistic regression analysis was applied to identify the predictors of using bioinformatics in research. Significance was set at $p<0.05$.

Results: More than one-half (248, 56.4%) of all scientists reported a lack of bioinformatics knowledge. Most participants had a neutral attitude toward bioinformatics (295, 95.4%). The barriers facing acceptance of bioinformatics tools reported were; lack of training (210, 67.9%), insufficient support (180, 58.2%), and complexity of software (138, 44.6%). The limited experience was reported in; having one or more bioinformatics tools (98, 31.7%), using a supercomputer in their research inside (44, 14.2%) and outside Saudi Arabia (55, 17.8%), the need for developing a program to solve a biological problem (129, 41.7%), working in one or more fields of bioinformatics (93, 30.1%), using web applications (112, 36.2%), and using programming languages (102, 33.0%). Significant predictors of conducting genomics research were; younger scientists ($p=0.039$), Ph.D. education ($p=0.003$), more than five years of experience ($p<0.05$), previous training ($p<0.001$), and higher bioinformatics knowledge scores ($p<0.001$).

Conclusion: The study revealed a short knowledge, a neutral attitude, a lack of resources, and limited use of bioinformatics resources in genomics research. Education and training during each education level and during the job is recommended. Cloud-based resources may help scientists do research using publicly available Omics data. Further studies are necessary to evaluate collaboration among bioinformatics software developers and biologists.

Keywords: big data, artificial intelligence, cloud computing, precision medicine, HGP, precision medicine, genomics, SDAIA

Background

Bioinformatics is a new discipline that merges biology, computer science, and information technology. It involves developing and integrating techniques, such as applied mathematics, statistics, computer science, chemistry, and biochemistry, to solve biological problems. Bioinformatics researchers develop and use computing tools to gain more about the life and survival of organisms in terms of; genetic background and molecular structure.[1] Bioinformatics manifested itself as an important research field, where the importance of this field came from its ability to mine large biological databases for relevant information. It is a base for any successful research in many different fields of life sciences such as molecular biology, drug discovery, biotechnology etc.[2]

Many advanced bioinformatics methods and approaches are available. Therefore, various "omics" research, such as genomics, transcriptomics, epigenomics, proteomics, interatomes, metabolomics, and many others, can be applied in

Received: 1 June 2023
Accepted: 3 August 2023
Published: 11 August 2023

biological and biomedical research.[2,3] Many bioinformatics training and workshops in this field have been conducted over the years, and all these efforts have rapidly changed the face of biology.[3,4] The new era of precision medicine and the need to provide the right treatments to suitable patients at the right time has created an environment for bioinformatics using Big data.[5,6] Moreover, The Human Genome Project (HGP) has created the need for new scientists of different disciplines, such as; computer sciences, biology, and mathematics, to contribute to genomic research.[7,8]

The shortage of bioinformatics skills is reported in many countries, including the United Kingdom (U.K.), France, Germany, Switzerland, Japan, and Australia.[9] In the UK, an approximately 60% rise in staff numbers is required in the bioscience academic sector, and nearly 45% rise in large pharmaceutical companies. In addition, there needs to be more research on biologists' awareness and help-seeking behavior, and limited publications are available in the literature directly related to bioinformatics in this area.[10,11] In Saudi Arabia, the results of a survey of 179 scientists showed that 93% agreed on the vital role of bioinformatics software tools, 53% obtained training in an education period, 11% received training during their job, and 48% were working in Gene expression analysis, protein expression analysis, and mutation analysis.[12] The Saudi Data & Artificial Intelligence Authority (SDAIA) was established in August 2019 by a Royal Decree to facilitate the transaction and help achieve Vision 2030 goals and reach its full potential.[13] It drives the national agenda for data and artificial intelligence to elevate the Kingdom as a global leader in the elite league of data-driven economies. This requires unifying national efforts and enabling initiatives related to data and AI to reach optimal utilization.[14]

A study of biologists' knowledge and perception of bioinformatics and how they access and use online bioinformatics resources is necessary.[1] Thus, this study aimed to assess the level of bioinformatics literacy among scientists in Saudi Arabia through the following: (1) assessment of the levels of knowledge regarding bioinformatics tools, (2) assessment of the attitude towards the use of bioinformatics tools in research, and barriers for accepting such tools, (3) determination of the predictors of use of bioinformatics tools in research, and (4) description of their previous experiences in using bioinformatics tools in their institutions, and pattern of such use.

Methods

Study Design and Setting

A cross-sectional study was conducted as an online survey using emails for scientists of different specialties.

Study Subjects

BSc degree-holders of different academic backgrounds (biology, pharmacy, medicine, computer science, statistics, etc.) and degrees (BSc, diploma, master, Ph.D.), whose job description is a scientist, and were available in Saudi Arabia during the time of the survey, in different biological and biomedical fields in universities, research institutes, hospitals, and private sectors, made the target of the study. An online questionnaire was developed via Survey Monkey, (https://www.surveymonkey.com/r/8K688DH) restricted to one participation per unique internet protocol (IP) address (accessed on 10 October 2022).

Sample Size and Sampling Technique

Assuming a prevalence of 7% lack of awareness of the importance of bioinformatics. Among scientists,[12] a confidence level of 95%, and a 3% margin of error, the estimated sample size was 278 scientists. To compensate for an average 50% expected non-response to the e-questionnaire and incomplete data collection, a total sample of 600 scientists of different educational grades were targeted, using an equal allocation method of sampling, via emails. Those who responded with complete questionnaires were 309 scientists.

Data Collection

A previously validated data collection tool based on the previous literature was used,[11,15] in English, with some modifications. The questionnaire was designed to allow for clear skip patterns for contingency questions and was broken into logical sections when possible. Test–retest reliability over a 2-week period was estimated ($r = 0.85$, $p< 0.001$). The scale demonstrated an adequate Cronbach's internal consistency of 0.83, and was considered adequate. Construct validity of the checklist was assessed using expert opinion, and the final version was approved accordingly.

The questionnaire is composed of four parts to gather descriptive data regarding the following:

1. Knowledge about bioinformatics programming languages and tools such as; genomic sequence analysis, sequence databases, RNA structure prediction, etc. Bioinformatics knowledge was assessed as reported by scientists using 11 statements, responded by "poor, fair, average, good, and excellent." The knowledge score was calculated as 1 point for poor and 5 points for excellent. The total score for each participant was calculated by summing scores for all responses, and the overall level of knowledge was assessed by adding scores for all responses. A total score ranged from 11 to 55 points. Percent score was calculated, and the level of knowledge was categorized for each participant, based on a previous study,[16] into satisfactory (>75%) and unsatisfactory (≤75%).
2. Attitude towards using bioinformatics resources in genome-related research in Saudi Arabia, creating bioinformatics jobs, centers, training programs, and teaching bioinformatics in high schools, and the degree of availability of bioinformatics resources in Saudi Arabia. Perceptions towards Problems and barriers that scientists face regarding online bioinformatics resources, such as; software complexity and reliability, knowledge gap or learning time, insufficient support, etc., were also assessed.13 Attitude statements were responded to by "strongly agree", "agree", "not sure", "disagree", or "strongly disagree." Agreeing with the five statements number "1, 2, 6, 9, 10" and disagreeing with the six statements number "3, 4, 5, 7, 8, 11" imply a positive attitude. A score can be calculated for each respondent using a scale of 1 to 5, with 5 = strongly agree. The opposite score was applied for negative attitude statements, with 1 = strongly agree and 5 = strongly disagree. A total score ranged from 13 to 65 points. Percent mean score was calculated, a positive attitude was considered for a percent score of 75% or more points, and a negative attitude was considered for the percent score of less than 50% points or less; otherwise, a neutral attitude was considered.[16]
3. The pattern of information-seeking by scientists who work on genome-related research regarding online bioinformatics resources. For example, user profiles such as frequency of use of online bioinformatics resources, types of tools used in Saudi Arabia, development of a program to solve bioinformatics problems, and prior training on how to use bioinformatics tools.
4. The use of bioinformatics in research. Scientists were asked to use bioinformatics tools in their research, the frequency of use, and their behavior when facing a problem that needs bioinformatics tools, the use of supercomputers in research. Data on variables that may predict the use of bioinformatics in research, such as; gender, academic background [biology, medicine, pharmacy, computer science, statistics, mathematics, others], academic degree [BSc, Applied Master science, Thesis-based Master science, Ph.D.], and years of experience, were collected.

Data Analysis

SPSS software Ver. 27 was used for data entry and analysis.[17] Descriptive statistics such as mean score and standard deviation, frequency and percentages of all independent variables (age, gender, educational grade, etc.) were used. A scoring system was applied to assess both knowledge and attitudes of scientists, and their percentage mean scores were calculated, then transformed into qualitative data. To identify the significant predictors of scientists' conduction of research using bioinformatics tools, logistic regression analysis was applied. Significance was set at a p-value <0.05.

Ethical Considerations

Participation in this study was voluntary; the participants were asked if they agreed to participate and were assured their responses would remain anonymous. Each participant received an electronic informed consent form, which had a short description of the study and asked potential subjects for their consent to participate. The subject's privacy and confidentiality were assured, no identifiers were collected, and all data were kept in a secure place within the Ministry of National Guard-Health Affairs (MNG-HA) premises, both hard and soft copies. The MNG-HA's Institutional Review Board (IRB) approved this study [Ref. #NRC21R/193/04].

Results

Of all scientists surveyed (n=309), two-thirds were females (64.7%), and medicine was the background for one-third (35.6%), one-half (51.4%) had their Bachelor's degree, and one-third (32.4%) reported having 15 years or more work experience. Table 1.

Knowledge About Bioinformatics Programming Languages and Tools

More than one-half (56.4%) of all scientists reported an unsatisfactory level of knowledge about bioinformatics, and only a few of them (12% to 22%) reported good or excellent knowledge of different fields of bioinformatics. The percent mean score of knowledge was 70.7% (unsatisfactory level), Table 2.

Attitudes Towards Using Bioinformatics Resources in Research

Table 3 shows that most of the scientists agreed/strongly agreed on the barriers to accepting bioinformatics, such as; the lack of training (67.9%), insufficient support (58.2%), and complexity of software (44.6%). The majority of scientists were in favor of the creation of bioinformatics jobs (77.7%) and Bioinformatics Centers (78.7%) and the integration of Bioinformatics in high school biology courses (70.2%) in Saudi Arabia. On the other hand, some of them considered bioinformatics a standalone discipline (44.3%), Arabization of bioinformatics training is not essential (40.5%), and 22.3% reported no need for bioinformatics training in Saudi Arabia. Nearly all participants (94.4%) reported a neutral attitude to bioinformatics.

Education Required for a Bioinformatics Career

Table 4 shows the education required for a career in bioinformatics, as reported by scientists. Apart from the need for a college degree in bioinformatics as a requirement to become a bioinformatician (61.5%), a biology degree (64.7%) or an I.T. degree (45.3%) was reported as an effective alternative way to become a bioinformatician after bioinformatics training. Biology ranked first as the specialty required for a promising career in bioinformatics (66.3%), followed by medicine (49.8%), computer sciences (48.9%), and life sciences (44.3%). A Bachelor's degree in bioinformatics and bioinformatics course was the suggested bioinformatics education by 34.6% and 30.4% of scientists, respectively.

Table 1 Personal Characteristics of Scientists

Characteristics	No. (%)
Gender:	
Male	109(35.3)
Female	200(64.7)
Academic background@	
Biology	43(13.9)
Medicine	100(32.4)
Pharmacy	7(2.3)
Computer Science	14(4.5)
Statistics	2(0.6)
Others (more than one specialty)	143(46.3)
Education (the highest achieved degree)@	
BS	159(51.5)
Master	60(19.4)
PhD	90 (29.1)
Years of work experience (years)	
1–5	59 (19.1)
6–10	73 (23.6)
11–15	77 (24.9)
>15 years	100 (32.4)

Note: @Categories are mutually exclusive.

Table 2 Levels of Knowledge About Different Fields of Bioinformatics

Fields of Bioinformatics	Poor n (%)	Fair n (%)	Average n (%)	Good n (%)	Excellent n (%)
1. Human Genome Project	199(38.5)	56(18.1)	67(21.7)	42(13.6)	25(8.1)
2. Genomic Sequence Analysis	127(41.1)	51(16.5)	61(19.7)	43(13.9)	27(8.7)
3. Sequence Databases	137(44.3)	54(17.5)	51(16.5)	43(13.9)	24(7.8)
4. Phylogenetics Prediction	169(45.7)	54(17.5)	50(16.2)	23(7.4)	13(4.2)
5. Prediction of RNA structure	171(55.3)	41(13.3)	60(19.4)	26(8.4)	11(3.6)
6. Gene Prediction and regulation	152(49.2)	61(19.7)	53(17.2)	29(9.4)	14(4.5)
7. Protein classification	139(45.0)	58(18.8)	54(17.5)	36(11.7)	22(7.1)
8. Protein structure prediction	160(51.8)	52(16.8)	51(16.5)	31(10.0)	15(4.9)
9. Genome analysis	141(45.6)	66(21.6)	52(16.8)	27(8.7)	23(7.4)
10. Microarray analysis	162(52.4)	63(20.4)	42(13.6)	25(8.1)	17(5.5)
11. Bioinformatics programming languages	176(57.0)	50(16.2)	39(12.6)	30(9.7)	14(4.5)
Percentage mean score (%)	70.7 [unsatisfactory]				
Overall Knowledge	**No.**	**%**			
Satisfactory	192	43.6			
Unsatisfactory	248	56.4			

Table 3 Response of Scientists on Attitudinal Statements on Using Bioinformatics

Statements	SA	AG	NS	DA	SD
1. There is need to create bioinformatics jobs in KSA.	117(37.9)	123(39.8)	58(18.8)	6(1.9)	5(1.6)
2. We must have Bioinformatics Centers in KSA.	122(39.5)	121(39.2)	57(18.4)	4(1.3)	5(1.6)
3. There is no pressing need to teach bioinformatics in the KSA institutes.*	17(5.5)	52(16.8)	96(31.1)	72(23.3)	72(23.3)
4. Software complexity and reliability were the main barriers for accepting bioinformatics.*	19(6.1)	119(38.5)	143(46.3)	21(6.8)	7(2.3)
5. Bioinformatics is a standalone discipline.*	35(11.3)	102(33.0)	122(39.5)	38(12.3)	12(3.9)
6. Teaching bioinformatics in Arabic language is essential.	18(5.8)	98(31.7)	111(35.9)	60(19.4)	22(7.1)
7. Lack of training in bioinformatics use is the main barrier for accepting bioinformatics tools.*	57(18.4)	153(49.5)	88(28.5)	10(3.2)	1(0.3)
8. Arabization of the bioinformatics training materials is not essential.*	24(7.8)	101(32.7)	132(42.7)	41(13.3)	11(3.6)
9. Knowledge gap in using bioinformatics tools has negative impacts on biomedical research.	60(19.4)	142(46.0)	92(29.8)	9(2.9)	6(1.9)
10. There is a need to introduce the bioinformatics concept in KSA high school biology course.	69(22.3)	148(47.9)	71(23.0)	15(4.9)	6(1.9)
11. Insufficient support is the main barrier for accepting bioinformatics tools.*	48(15.5)	132(42.7)	113(36.6)	12(3.9)	4(1.3)
Percent mean score (%):	63.4±6.0 [neutral]				
Overall attitude	**No.**	**%**			
Positive (no.,%)	11	3.6			
Neutral (no., %)	295	95.4			
Negative (no., %)	3	1.0			

Note: *Negative attitude statement.
Abbreviations: SA, strongly agree; AG, agree; NS, not sure; DA, disagree; SD, strongly disagree.

The Use of Bioinformatics in Research

Table 5 also shows that less than one-third of scientists said their institutes have one or more licensed bioinformatics tools (31.7%). Less than one-half of scientists reported using bioinformatics tools in their research (131, 42.4%). Most of those used these tools rarely (68, 51.9%). We asked the participants what they would do if their research needed a bioinformatics tool; one-third (33.7) said they learned how to use it, and 47.6% said they hired or included a bioinformatician. A small percentage of the scientists who use bioinformatic tools reported using a supercomputer in

Table 4 Required Education for a Career in Bioinformatics

Required Education	No. (%)
What is the most effective way to become a bioinformatician?@	
1. IT training courses to be bioinformatician	140(45.3)
2. Biology training courses to be a bioinformatician	200(64.7)
3. College degree in bioinformatics is a must	190(61.5)
4. Other	27(8.7)
What level of bioinformatics education should be introduced in Saudi Arabia institutes?@	
1. Bachelor degree in Bioinformatics	107(34.6)
2. Applied Master science degree in Bioinformatics	30(9.7)
3. Thesis-based Master science degree in Bioinformatics	46(14.9)
4. PhD degree in Bioinformatics	22(7.1)
5. Teach Bioinformatics courses as part of different Discipline curricula (ie IT, Biology, medicine…etc.)	94(30.4)
6. Other	10(3.2)
Bioinformatics a good career for the student to graduate from which specialty?@	
1. Medicine	154(49.8)
2. Pharmacist	90(29.1)
3. Computer Science	151(48.9)
4. Mathematics	72(23.3)
5. Statistics	121(39.2)
6. Any life science specialty (including Biology)	205(66.3)
7. Others	22(7.1)

Note: @Categories are not mutually exclusive.

Table 5 Previous Experience with Bioinformatics

Statement	No (%)
Having any licensed bioinformatics tools in your institute (n=309)	
Yes	98(31.7)
No	63(20.4)
I do not know	148(47.9)
Conduction of research using bioinformatics tools (n=309)	
Yes	131(42.4)
No	178(57.6)
Frequency of conduction of research using bioinformatics tools (n=131)	
Rarely (1–39%)	68(51.9)
Sometimes (40–79%)	40(30.5)
Always (80–100%)	23(17.6)
Conduction of a research that used super computer capability in Saudi Arabia (n=131) [Yes]	44(33.6)
Conduction of a research that used super computer capability outside Saudi Arabia (n=131) [Yes]	55(42.0)
Need to develop a program to solve biological problems by you and/or your team? (n=309) [Yes]	129 (41.7)
Approach if your research needed to use a bioinformatics tool you are not familiar with (n=131):@	
Learn the tool myself (college course/ self online training/reading books/on-job training)	104(79.3)
Hire and/or include bioinformatician to do the job for me	69(52.6)
Current work in any field of bioinformatics (n=309)	
No	216(69.9)
Yes	93 (30.1)
Which field of bioinformatics do you currently work in (n=93)@	
Genome/Sequence analysis	51(54.8)
Gene expression	35(37.6)
Others (Phylogenetics, Structural Bioinformatics, Genetics and population analysis, Systems Biology, Data and Text Mining, Databases and Ontologies, Molecular Dynamic and simulation)	37 (39.7)

(*Continued*)

Table 5 (Continued).

Statement	No (%)
Bioinformatics Do you regularly use bioinformatics web applications (n=309)	
Yes	112 (36.2)
No	197(63.8)
Which web application do you currently use? (n=112)@	
NCBI-Blast	58(51.8)
Swiss-model	17(15.2)
Others (Hugo gene database, Expasy proteomics server, String database)	43(38.4)
Programming languages you regularly use regarding biological problem (n=309)	
Yes	102(33.0)
No	207(67.0)
Which language programmes do you currently use? (n=102)@	
Java	25(24.5)
Python	41(40.2)
R	34(33.3)
Matlab	13(12.7)
Others (Perl, Ruby, Lisp, Smalltalk, C/C++)	34(33.3)

Notes: Figures in the table are only for individuals whose data are available. @Categories are not mutually exclusive.

their research inside (33.6%) and outside (42.0%) Saudi Arabia. The need for developing a program to solve the biological problem was reported by 41.7% of all participants.

Table 5 shows that only 30.1% of participants reported working in one or more fields of bioinformatics. The fields of Bioinformatics Currently used by those participants were; genome /sequence analysis (51,54.8%) and gene expression (35, 37.6%). More than one-third of all participants (39.7%) reported they were currently working in other fields such as; Phylogenetics, Structural Bioinformatics, Genetics and population analysis, Systems Biology, Data and Text Mining, Databases, and Ontologies, and Molecular Dynamics and simulation. Only 36.2% reported using bioinformatics web applications. The most regularly used web applications reported by those participants were; NCBI-Blast (51.8%) and Swiss-model (15.2%). Other applications, such as; the Hugo gene database, Expasy proteomics server, and String database, were reported by one-third of those participants (38.4%) of those participants. Programming languages were currently used by one-third of all participants (33.0%). The most regularly used programming languages reported by those participants were Python (40.2%), R (33.3%), Java (24.5%), and Matlab (12.7%). Other used languages such as; Perl, Ruby, Lisp, Smalltalk, and C/C++ were reported by one-third (33.3%) of those participants.

Figure 1 shows the frequency of use of different web applications among scientists whose current field of bioinformatics is genome/sequence analysis. NCBI-Blast was used by two-thirds of those scientists, followed by the Swiss (15%) and the String database (12.5%). Less than one-third (30%) reported using other web applications.

Figure 2 shows that knowledge and attitude levels on bioinformatics were significantly associated with conducting research in bioinformatics. The percent mean score of knowledge was significantly higher among those who conducted research than those who did not (53.2% versus 32.9%, $p<0.001$). Likewise, the percent mean score of attitude was significantly higher among those who conducted research than those who did not (64.3% versus 62.7%, $p=0.023$).

Figure 3 shows that levels of knowledge of bioinformatics were significantly associated with the conduction of training in bioinformatics. The percent mean score of knowledge was significantly higher among those who reported conducting training than those who did not (45.1% versus 32.4%, $p=0.011$). However, the percent mean score of attitude was not significantly higher among those who reported conducting training than those who did not (63.7% versus 62.4%, $p=0.10$).

Predictors of Conducting Research Using Bioinformatics

Table 6 shows the logistic regression analysis of conducting research among scientists by some variables. Scientists' conduction of research using bioinformatics was significantly associated with age, education, experience, previous

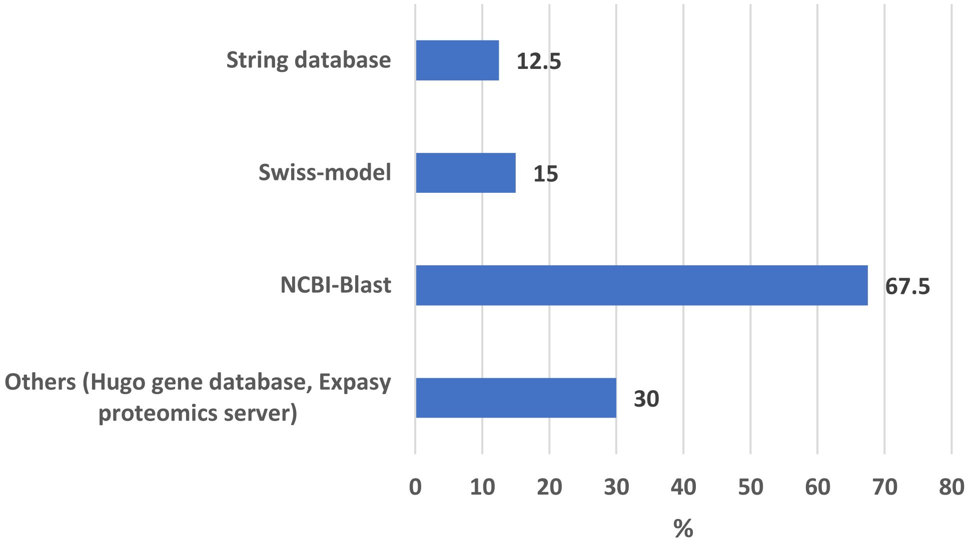

Figure 1 The frequency of use of different web applications by scientists whose current field of bioinformatics is Genome/Sequence analysis.

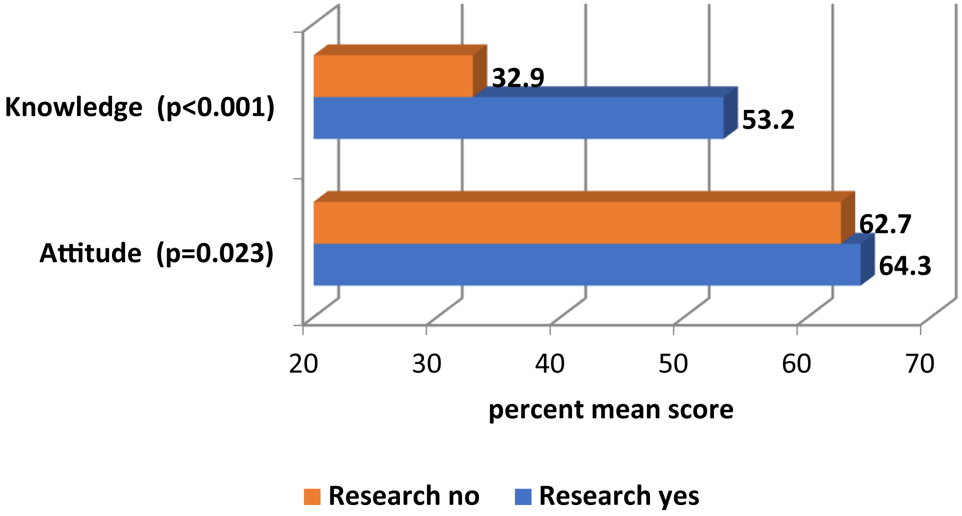

Figure 2 Association of knowledge and attitude percent mean scores with conduction of research in bioinformatics among scientists.

training, and bioinformatics knowledge. Younger scientists ($p=0.039$), those with Ph.D. education (($p=0.003$), those with more than five years of experience ($p<0.05$), those who reported conducting training ($p<0.001$), and those with higher knowledge scores about bioinformatics ($p<0.001$) were more likely to conduct research using bioinformatics. Those who reported attending the training were nearly five times more likely to conduct research than those who did not (OR=3.7, $p<0.001$). Those with >15 years of experience were 3.4 times more likely to conduct research than those with <5 years of experience (OR=3.4, $p=0.048$). Those holding Ph.D. education degrees were 3.4 times more likely to conduct research than those holding Bachelor's degrees (OR=3.4, $p=0.003$).

Discussion

With the advances in genomics research, the shortage of bioinformatics awareness is still limited in many countries. Several studies have been conducted on how computers for information handling were used to identify the bioinformatics practices conducted by biologists.[11,18] The results showed that the use of bioinformatics tools differed based on the field

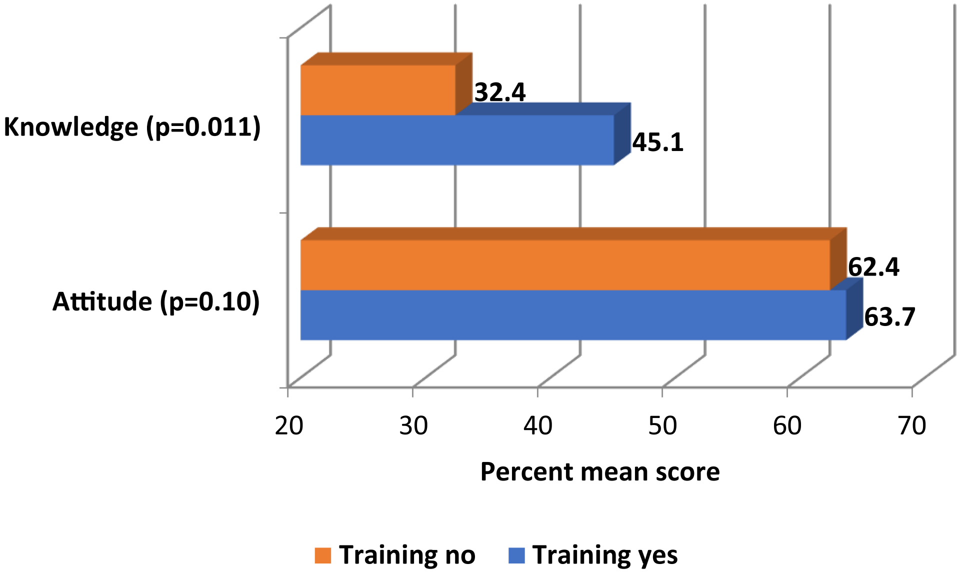

Figure 3 Association of knowledge and attitude percent mean scores with conduction of training in bioinformatics among scientists.

and the institution involved, where the senior researchers in biology were actually more active than junior researchers, as information providers and recipients, possibly because of less pressure on senior staff time and less financial restrictions.[10] In our study, participants' conduction of research using bioinformatics was significantly associated with age, background education, years of experience, previous training, and level of bioinformatics knowledge. Younger

Table 6 Logistic Regression Analysis of Conducting Research Using Bioinformatics Among Scientists by Some Variables

Variables	*B*	*S.E.*	p-value	OR	95% C.I.	
					Lower	Upper
Gender [Male@ vs female]	0.492	0.338	0.146	1.636	0.843	3.176
Age1 (25 years)@			0.143			
Age (35-years)	−0.865	0.419	0.039*	0.421	0.185	0.957
Age (45- years)	−1.312	0.635	0.039*	0.269	0.078	0.935
Age (55–64 years)	−0.938	0.801	0.242	0.392	0.082	1.880
Background (Others)@			0.112			
Background (Biology)	0.786	0.495	0.112	2.196	0.833	5.789
Background (Medicine)	−0.420	0.357	0.239	0.657	0.327	1.322
Background (Computer sciences)	0.586	0.841	0.486	1.797	0.346	9.333
Education (Bachelor)@			0.007*			
Education (Master)	0.104	0.423	0.806	1.109	0.484	2.541
Education (PhD)	1.224	0.417	0.003*	3.401	1.503	7.694
Experience (1–5 years)@			0.123			
Experience (6–10 years)	0.985	0.473	0.037*	2.679	1.060	6.771
Experience (11–15 years)	1.054	0.513	0.040*	2.870	1.050	7.848
Experience (>15 years)	1.227	0.620	0.048*	3.412	1.012	11.506
Knowledge (score)	0.029	0.008	<0.001*	1.030	1.015	1.045
Attitude (score)	0.035	0.026	0.178	1.036	0.984	1.091
Previous training [yes@ vs no]	1.316	0.321	<0.001*	3.727	1.988	6.990
Constant	−5.023	1.777	0.005	0.007		

Notes: @Reference category, *Statistically significant association.
Abbreviations: OR, odds ratio; CI, confidence interval; *B*, beta coefficient of determination; *SE*, standard error.

scientists, those with Ph.D. education, those with more than five years of experience, those who reported conducting training, and those with higher scores of knowledge about bioinformatics were more likely to conduct research using bioinformatics. Knowledge of bioinformatics tools should be given to the students. Bioinformatics can be included in their curriculum. This might improve their research skills in the future.[1]

Bioinformatics education and practice are still in their infancy, and efforts have been made recently to introduce bioinformatics modules into the biological sciences curriculum of universities.[1] In our study, the need for a college degree in bioinformatics as a requirement to become a bioinformatician was reported by two-thirds of the participants. A biology degree or I.T. degree was reported as an effective alternative way to become a bioinformatician after bioinformatics training. The environment is suitable for computer scientists to share in and support bioinformatics programs.[19] In our study, life science specialties, including biology, ranked first as the specialty required for a promising career in bioinformatics, followed by medicine and computer sciences. A Bachelor's degree in bioinformatics and bioinformatics course was the suggested bioinformatics education by nearly one-third of scientists. In the US, bioinformatics education has been instructed at the graduate level.[20–22] However, little effort has been made to integrate it into undergraduate biology curricula due to a lack of training and instruction and a lack of student interest and preparation.[23]

Our study showed that less than one-third of scientists said their institutes have one or more bioinformatics tools. Less than one-half of scientists reported using bioinformatics tools in their research, and half of them reported using these tools rarely. In a previous study on the effect of hands-on training workshops,[24] most the interviewees utilized public databases such as PubMed and Online Mendelian Inheritance in Man (OMIM). Many also used sequence analysis tools. A previous survey conducted in Saudi Arabia showed that the majority of respondents found bioinformatics tools necessary in modern biology; one-half of respondents claimed that they were first introduced to these tools in the education period, one-half said that sequence analysis and structure analysis were the main areas of implementation of bioinformatics tools.[12] In the present study, the scientists of the most frequent field currently work were; genome/ sequence analysis, followed gene expression. One-third of all participants reported they were currently working in other areas such as; Phylogenetics, Structural Bioinformatics, Genetics and population analysis, Systems Biology, Data and Text Mining, Databases and Ontologies, and Molecular dynamics and simulation.

Perceptions of the acceptance of bioinformatics tools are important as predictors of the actual use of such tools.[25,26] In our study, most of the scientists agreed/strongly agreed on the barriers to accepting bioinformatics, such as; the lack of training, insufficient support and complexity of software, creation of bioinformatics jobs and bioinformatics centers, and integration of bioinformatics in the high school biology courses, in Saudi Arabia. In the present study, the percent mean score of attitude was significantly higher among those who conducted research than those who did not. However, the total attitude score was not a significant predictor of conducting research using bioinformatics. To manipulate a large volume of data, you must be empowered with more knowledge of and interest in bioinformatics-related disciplines such as; computer science, mathematics, and biology or rely upon a team of trained personnel in these disciplines to help you do so.[27]

In a previous study, the difficulty of interpreting results was one of the main barriers to accepting bioinformatics tools, and poor documentation quality available with these tools was another barrier.[12] In our study, only one-third of scientists reported using bioinformatics web applications. The participants' most regularly used web applications were; NCBI-Blast and Swiss-model. NCBI-Blast was used by two-thirds of scientists whose current field of bioinformatics is genome/ sequence analysis, followed by the Swiss and the String database. Less than one-third of those scientists reported using other web applications such as; the Hugo gene database, Expasy proteomics server, and String database (Figure 1). The most regularly used programming languages reported by the participants were Python, R, Java, and Matlab. In a qualitative study of the acceptance of bioinformatics tools,[15] of the reasons reported for taking the training workshops in bioinformatics were; to fulfill a specific job need and because of the positive effect of training on their decision to use software tools.

Many training programs, such as; short courses and senior fellowships, need to be introduced or increased to address the skills gap in bioinformatics.[9] The present study showed that more than one-half of all scientists reported an unsatisfactory level of bioinformatics knowledge, and the majority of them reported poor or fair knowledge of different fields of bioinformatics. This is expected given that almost half the participants only had a BSc level of education. Those with higher knowledge scores about bioinformatics in the present study were more likely to conduct research in bioinformatics. Research and structural predictions of biomolecules can be made with ease with the help of bioinformatics, and also know that bioinformatics helps in learning molecular biology with ease.[1] In our study, although most of the participants are aware of the importance of bioinformatics, almost two-thirds of them were not aware of the various databases and tools employed in bioinformatics in their institutes, and one-half did not know if they have any bioinformatics tools in their institute, more than one half did not participate in any research using bioinformatics tools. For those who needed to use a bioinformatics tool in research, only one-third reported self-learning the tool, while half stated hiring a bioinformatician to do the job.

The impact of hands-on training workshops was investigated. They positively impacted the primer design tool's perceived ease of use (PEOU), decreased the PEOU of the microarray data analysis tools, and the intention to use both types of software decreased after hands-on training.[28] In our study, those who reported attending the training were nearly five times more likely to conduct research than those who did not. Previous bioinformatics training was a significant predictor of using bioinformatics tools in research.

Healthcare institutions are well equipped with information technology to support the clinical environment and billing, but not the research environment of Big data that requires a unique environment to store, handle, combine, curate, and analyze large volumes of data.[6] Some aspects of science (for example, wet labs) need significant resources to establish and maintain.[29] The present study showed a limited percentage of scientists who reported using a supercomputer in their research inside and outside Saudi Arabia. Meanwhile, only one-fifth reported participation in research using grid computing. Less than one-half of the participants reported the need to develop a program to solve biological problems. Nowadays, available low-cost computing allows scientists with sufficient training and cloud access to develop novel computational methods.[29]

Strengths and Limitations

This study may offer novel results which would help introduce important information to the literature on the level of bioinformatics literacy among scientists, and its impact on research and development in Saudi Arabia, during the era of precision medicine, especially in genomic medicine. It may act as a pilot study to other ones from similar countries. However, this study has some limitations: First, although we tried to contact different scientific institutions in the country, the scientists who participated in the study might not represent all Saudi Arabian scientists regarding their academic background and biological and biomedical fields; almost half the participants only had a BSc level of education. Thus, the study might be subjected to selection bias. The study was based upon a questionnaire that might subject the study to a recall bias, especially since the assessment of knowledge about the fields of bioinformatics was based upon their reported self-assessment. The information is from an e-questionnaire, which may also potentially affect the neutrality or validation of the information.

Moreover, the study used a previously validated tool of data collection that may not cover some of the new fields or a new category of bioinformatics such as; functional genomics, epigenetics, single-cell omics, etc. The cross-section study design does not guarantee the cause-and-effect relationship between the conduction of research using bioinformatics as an outcome and the predictors, such as age, gender, academic background, and specialty.

Conclusion

The study revealed a need for more knowledge and experience in bioinformatics, a neutral attitude by scientists in Saudi Arabia, limited resources, and a modest contribution of bioinformatics to developments in biological science education and research. Bioinformatics education and training, in each education level and during the job, is recommended. Cloud-based resources can help scientists to train, research, and do scientific discoveries using publicly available omics data.

Abbreviations

HGP, Human Genome Project; UK, United Kingdom; MNG-HA, Ministry of National Guard-Health Affairs; IRB, Institutional Review Board; SDAIA, Saudi Data & Artificial Intelligence Authority; OMIM, Online Mendelian Inheritance in Man; PEOU, perceived ease of use; IT, information technology.

Data Sharing Statement

Most of the data supporting our findings is contained within the manuscript, and all others, excluding identifying/confidential patient data, will be shared upon request by contacting the corresponding author [Mostafa Abolfotouh mabolfotouh@gmail.com].

Ethics Approval and Consent to Participate

This research was approved by the Institutional Review Board (IRB) of The Ministry of National Guard-Health Affairs, Riyadh, Saudi Arabia (Ref.# NRC21R/193/04). Participation in this study was voluntary. Informed consent for study participation was obtained from all subjects. Scientists were assured in the informed consent that their responses would remain anonymous. The purpose of the study was explained, and all issues regarding confidentiality and privacy were assured and protected at all times. The study was performed following the Declaration of Helsinki.

Acknowledgments

This study was initiated by King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia. Special thanks are due to His Excellency Dr. Tariq bin Abdullah Al-Sheddi, President of the National Data Management Office (NDMO), Saudi Data & AI Authority (SDAIA), Saudi Arabia, for his support during the conduction of the study. The final draft of the manuscript was English language edited by Macmillan Science Communication.

Author Contributions

Both authors made substantial contributions to conception and design, acquisition of data, or analysis and interpretation of data; took part in drafting the article or revising it critically for important intellectual content; agreed to submit to the current journal; gave final approval of the version to be published; and agree to be accountable for all aspects of the work.

Disclosure

The authors declare that they have no competing interests.

References

1. Bharathi R, Gayathri R, Vishnu Priya V. Awareness on the application of bioinformatics tools among college – a survey. *Drug Invent Today*. 2019;11(4):1.
2. Fenstermacher D. Introduction to bioinformatics. *J Am Soc Inf Sci*. 2005;56(5):440–446. doi:10.1002/asi.20133
3. Micheel CM, Nass SJ, Omenn GS; Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care Services; Board on Health Sciences Policy; Institute of Medicine. *Evolution of Translational Omics: Lessons Learned and the Path Forward*. Washington (DC): National Academies Press (US); 2012.
4. Fatumo SA, Adoga MP, Nashiru O. Computational biology and bioinformatics in Nigeria. *PLoS Comput Biol*. 2014;10:e1003516. doi:10.1371/journal.pcbi.1003516
5. Ashley EA. Towards precision medicine. *Nat Rev Genet*. 2016;17:507–522. doi:10.1038/nrg.2016.86
6. Hulsen T, Jamuar SS, Moody AR, et al. From big data to precision medicine. *Front Med*. 2019;6:34. doi:10.3389/fmed.2019.00034
7. Lewis J, Bartlett A, Atkinson P. Hidden in the middle: culture, value and reward in bioinformatics. *Minerva*. 2016;54:471–490. doi:10.1007/s11024-016-9304-y
8. Francis S, Patrinos A, Jordan E, Chakravarti A, Gesteland R, Walters L. New goals for the U.S. Human genome project: 1998–2003. *Science*. 1998;282(5389):682–689. doi:10.1126/science.282.5389.682
9. MacLean M, Miles C. Swift action needed to close the skills gap in bioinformatics. *Nature*. 1999;401(6748):10. doi:10.1038/43269
10. Dihui L. Biologists' information seeking behavior with online bioinformatics resources for genome research. A Master's paper for the M.S. in I. S degree; 2003:44.
11. Stevens R, Goble C, Baker P, Brass A. A classification of tasks in bioinformatics. *Bioinformatics*. 2001;17(2):180–188. doi:10.1093/bioinformatics/17.2.180
12. Ahmadi T, Al-Senaidy AM. Software tools in bioinformatics: a survey on the importance and issues faced in implementation. *Glob Eng Technol Rev*. 2013;3:6–10.
13. Memish ZA, Altuwaijri MM, Almoeen AH, Enani SM. The Saudi Data & Artificial Intelligence Authority (SDAIA) vision: leading the kingdom's journey toward global leadership. *J Epidemiol Glob Health*. 2021;11(2):140–142. PMID: 33876596; PMCID: PMC8242113. doi:10.2991/jegh.k.210405.001
14. SDAIA. SDAIA and Vision 2030; 2023. Available from: https://sdaia.gov.sa/en/SDAIA/SdaiaStrategies/Pages/sdaiaAnd2030Vision.aspx. Accessed July 19, 2023.

15. Shachak A, Shuval K, Fine S. Barriers and enablers to the acceptance of bioinformatics tools: a qualitative study. *J Med Libr Assoc*. 2007;95 (4):454. doi:10.3163/1536-5050.95.4.454
16. Aldebasi B, Alhassan AI, Al-Nasser S, Abolfotouh MA. Level of awareness of Saudi medical students of the internet-based health-related information seeking and developing to support health services. *BMC Med Inform Decis Mak*. 2020;20(1):209. PMID: 32883284; PMCID: PMC7469907. doi:10.1186/s12911-020-01233-8
17. IBM Corp. IBM SPSS statistics for windows, version 27.0. Armonk, NY: IBM Corp; 2020. Available from: https://www.ibm.com/support/pages/downloading-ibm-spss-statistics-27. Accessed August 9, 2023.
18. Bukhari AA, Meadows AJ. The use of information technology by scientists in British and Saudi Arabian universities: a comparative study. *J Inf Sci*. 1992;18(5):409–415. doi:10.1177/016555159502100
19. Dyer BL, Dyer BD. Bioinformatics and computing curricula 2001 why computer science is well positioned in a post-genomic world. *ACM SIGCSE Bull*. 2004;36:64–67. doi:10.1145/1041624.1041659
20. Hack C, Kendall G. Bioinformatics: current practice and future challenges for life science education. *Biochem Mol Biol Educ*. 2005;33(2):82–85. doi:10.1002/bmb.2005.494033022424
21. Karikari TK, Quansah E, Mohamed WMY. Developing expertise in bioinformatics for biomedical research in Africa. *Applied & Translational Genomics*. 2015;6:31–34. doi:10.1016/j.atg.2015.10.002
22. Kulkarni-Kale U, Sawant S, Chavan V. Bioinformatics education in India. *Brief Bioinform*. 2010;11(6):616–625. doi:10.1093/bib/bbq027
23. Williams JJ, Drew JC, Galindo-Gonzalez S, et al. Barriers to integration of bioinformatics into undergraduate life sciences education: a national study of US life sciences faculty uncover significant barriers to integrating bioinformatics into undergraduate instruction. *PLoS One*. 2019;14(11): e0224288. PMID: 31738797; PMCID: PMC6860448. doi:10.1371/journal.pone.0224288
24. Shachak A. *The Effect of Training Methods on Perceptions and Intended Use of Bioinformatics Tools* [doctoral dissertation]. Ramat-Gan, Israel: Department of Information Science, Bar-Ilan University; 2005.
25. Rogers EM. *Diffusion of Innovations*. 4th ed. New York, NY: Free Press; 1995:14.
26. Moore GC, Benbasat I. Development of an instrument to measure the perceptions of adopting an information technology innovation. *Inf Syst Res*. 1991;2(3):192–222. doi:10.1287/isre.2.3.192
27. Campbell P. Don't leave the biology out of bioinformatics. *Nature*. 1999;401(6751):321. doi:10.1038/43764
28. Reference and User Services Association, American Library Association. Guidelines for information services. Chicago, IL: The American Association; 2000. Available from: http://www.ala.org/ala/rusa/rusaprotools/referenceguide/guidelinesinformation.html. Accessed January 25, 2021.
29. Mangul S, Martin LS, Langmead B, et al. How bioinformatics and open data can boost basic science in countries and universities with limited resources. *Nat Biotechnol*. 2019;37:324–326. doi:10.1038/s41587-019-0053-y