ISSN (P) : 2615 - 2584
ISSN (E) : 2615 - 3343

BIOINFORMATICS, BETWEEN NECESSITY AND CHALLENGES: AN INDONESIAN PERSPECTIVE

Kholis Abdurachim Audah
Department of Biomedical Engineering and Academic Research and Community Services, Swiss German University
E-mail: kholis.audah@sgu.ac.id

Abstract
Biological Information is a priceless asset for many purposes. From Indonesian perspective, this become more important due to the fact that Indonesia is one of the richest country in the world in terms of biodiversity both in the ocean and on the land. In addition, Indonesia is also a country with fourth largest population in the world. All these natural resources can be very powerful if they are well managed and optimally utilized. A large collection of data (Big Data) can be obtained from different levels of organization of organisms. These data can then be processed, analysed and interpreted to become important biological information that eventually being utilized for the betterment of mankind. In order to obtain such big data and turns the data into a meaningful information, scientists require a set of tools such as Bioinformatics. Bioinformatics is a interdisciplinary field of studies that encompasses different area of sciences which includes Biological Sciences, Medical Sciences, Computer Sciences, Mathematics and Statistics. This opens up a wide range of collaborative works, from field to bench involving both "wet" and "dry" laboratory. In this article, the necessity of implementing bioinformatics and challenges are discussed from Indonesian point of view.

Keywords: b*ig data*, *biodiversity, biological information, genomics*

Introduction

Biological information can be utilized for many purposes. These include but not limited to food and energy security, plant and environment protection, drug discovery and medicines. The information can be obtained from different levels of organism classification or organization, from the smallest to the biggest, from molecular to population level. Bioinformatics play a role in processing biological data at the molecular level originated either from microbes, plants, animals or human. The data will then be analyzed and interpreted into meaningful information. In the other words, the object of bioinformatics its self is organism at the molecular level; that is genome (genomics) and its derivatives such as transcriptomics, proteomics, lipidomics and metabolomics. The more organisms studied, the more data collected and the more information obtained.

Bioinformatics as an Interdisciplinary Science

Bioinformatics had been defined in different versions. All these definitions emphasizing the link between biological data with techniques for information storage, distribution and analysis to support multiple areas of scientific research, including biomedicine (Lesk, 2013). The works of bioinformatics had been initiated in the mid 1940s when Fred Sanger worked on the sequence of insulin by using a set of chemical and enzymatic techniques (Attwood, T.K, 2011). As the results of his phenomenon works, protein and DNA sequencing methods, Sanger was awarded with two Noble Laureattes in the year 1958 and 1980, respectively. Bioinformatics encompasses different area of sciences which

ISSN (P) : 2615 - 2584
ISSN (E) : 2615 - 3343

includes Biological Sciences, Medical Sciences, Computer Sciences, Mathematics and Statistics as depicted in Figure 1. Being an interdiscipline in nature, implementing bioinformatics in research or in education is not an easy task, since it requires different expertise that makes it nearly impossible to be done by single individual. It means that a collaborative work is a must in implementing this emerging field of study.

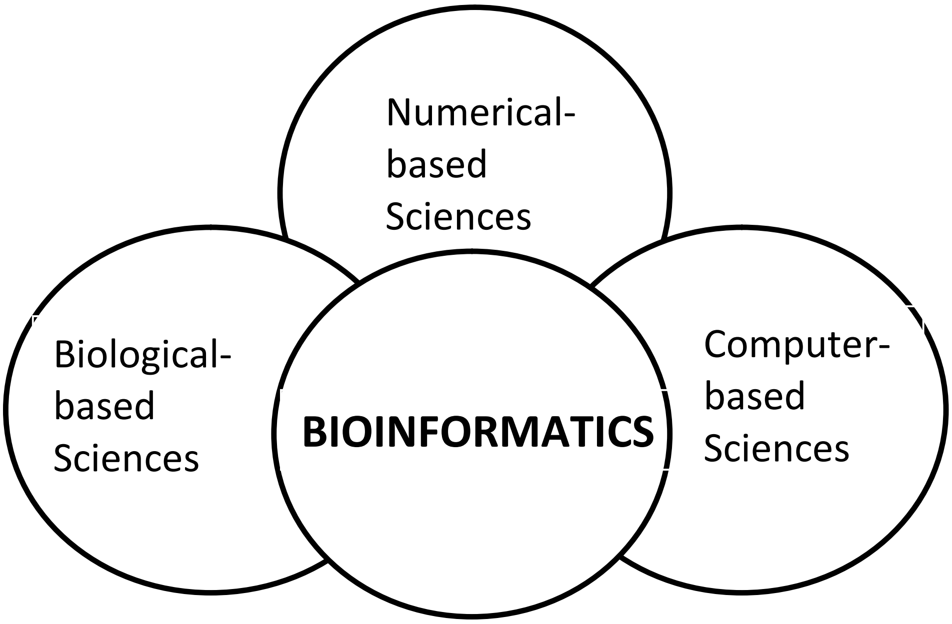

Figure 1. Interconnection of bioinformatics with other major sciences

Despite the very powerful tool bioinformatics could be, the implementation of it still heavily depend upon high-throughput data-generating experiments, including genomics, proteomics and other emerging branches of omics. Generation of data can only be achieved through extensive researches employing various techniques in molecular biology such as DNA, RNA and protein sequencing, DNA or protein arrays and diverse types of chromatography techniques. High-throughput data allow scientist to do different types of experiments such as high-throughput screening (HTS) in drug discovery process (Quinn, 2012). As many as 100,000 compounds can be tested only in few days to obtained desired molecules (hits) (Butler, 2005). HTS can also be performed by using computational approach prior to experiments in laboratory.

Biodiversity as invaluable source of bioinformatics

Currently, there are about 391,000 species of vascular plants currently known to science, of which about 369,000 species (or 94 percent) are flowering plants, according to a report by the Royal Botanic Gardens, Kew, in the United Kingdom (State of the World's Plants, 2017). Out of this number, only about 31,000 species have at least one documented use. These include uses for food, medicine, recreation, genes, poisons, animal feed, and building material (Figure 2) (State of the World's Plants, 2017).

Indonesia is well known as one of the richest country in the world in biodiversity both in the ocean and on the land as shown in Table 1 (UNEP, 2004). A full list of countries with their respective biodiversity is listed in Appendix. Based on this facts, biodiversity-based research is very suitable for Indonesia. Therefore, the implementation of bioinformatics in Indonesia becomes extremely important to make the most out of these natural resources.

ISSN (P) : 2615 - 2584
ISSN (E) : 2615 - 3343

Data used for bioinformatics is usually considered as big data. For example, data sets generated from a genomics study employing Next Generation Sequencing (NGS) platforms can be 20 gigabytes of data per sample for hundreds of samples with downstream analyses generating 100's of gigabytes of data per sample (May, 2014). A massive sets of data can be stored as invaluable databases that can be utilized for different purposes through data mining activities and analyses generating various meaningful information.

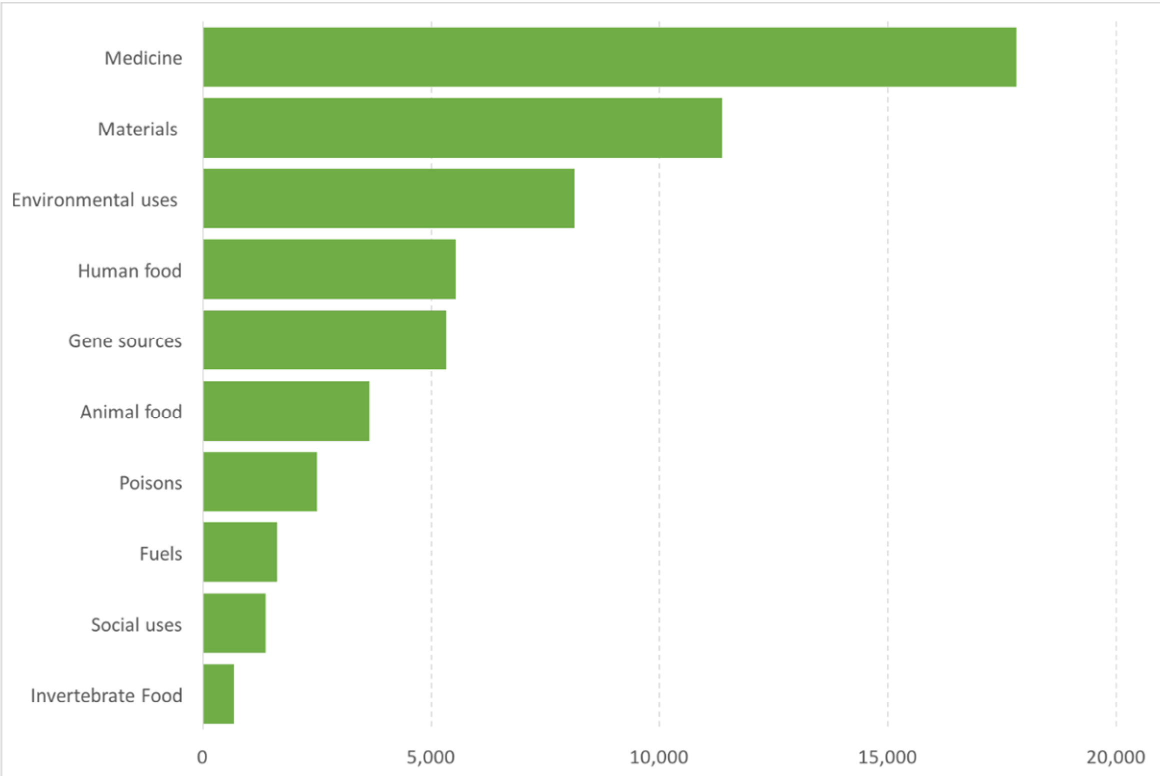

Figure 2. Top documented uses of plant species. Graphic: Mongabay; Source: Royal Botanic Gardens Kew: State of the World's Plants, 2017.

Tabel 1. Top Ten Countries with Their Respective Biodiversity in Numbers

Country	Birds	Amphibians	Mammals	Reptiles	Fish	Vascularplants
Brazil	1753	1022	648	807	4521	56215
Colombia	1826	771	442	601	2053	51220
China	1240	411	551	478	3330	32200
Indonesia	**1615**	**347**	**670**	**728**	**4682**	**29375**
Mexico	1081	377	523	916	2602	26071
South Africa	755	128	297	447	2059	23420
Venezuela	1364	360	363	405	1709	21073
United States	844	300	440	530	3067	19473
India	1180	390	412	689	2465	18664
Ecuador	1588	539	372	444	1098	19362

Source: Plant data is from the World Conservation Monitoring Centre of the United Nations Environment Programme (UNEP-WCMC), 2004. Species Data. Fish: Fishbase; Birds: Birdlife International; Amphibians: AmphibiaWeb; Mammals: IUCN; Reptiles: the Reptile Database.

ISSN (P) : 2615 - 2584
ISSN (E) : 2615 - 3343

Bioinformatics as a necessity for Indonesia

By taking into consideration of the great numbers of biodiversity, the massive data generated and the power of bioinformatics as a tool in producing a diverse biological information, it is necessary for Indonesia to take a quick and comprehensive action in implementing bioinformatics. This is to ensure that all potential Indonesia has is optimally utilized for Indonesian people first before anybody else or to the least extent it can be utilized as bargaining power for mutual cooperation with countries with more advance in science and technology.By implementing bioinformatics, many genetic diseases either in plants, animal even human can be revealed and possibly cured.

The challenges in implementing bioinformatics in Indonesia

Bioinformatics as a tool require sets of data to be analysed acquired from high-throughput data-generating experiments which unfortunately come with high cost. It will take quite a lot of investment, effort and time in order to generate significant amount of data.Bioinformatics is relatively new introduced in Indonesia so that the importance of it is still not well noticed. As the consequence, there is no sense of urgency in generating high-throughput data as the object of bioinformatics, because there is no adequate infrastructure and facilities to generate big data. This is an ironical situation considering that Indonesia has great potential in terms of biodiversity as the object of bioinformatics research.Therefore, the government should take the lead of this noble effort so it can be accelerated at significant pace.

Necessary actions to be taken

Some necessary actions need to be taken to fulfill the needs of implementing bioinformatics and other supporting activities. These actions are very important to make sure that Indonesian natural resources are optimally utilized to solve current and future problems in different aspects of life. These actions include but not limited to 1) Establishment of Indonesia Genomic Institute as center of excellent for generating high-throughput data; 2) National natural product library and or bio bank; and 3) National Biological Information System which is important for collecting and integrating biological databases or as data bank.

Acknowledgement

Author would like to thank Swiss German University for financial support for publication of this manuscript.

References

Attwood, T.K., A. Gisel, N-E. Eriksson and E. Bongcam-Rudloff, Concepts, Historical Milestones and the Central Place of Bioinformatics in Modern Biology: A European Perspective in INTECH, Mahdavi, M. A, 2011 (Ed.), Computer and Information Science » Numerical Analysis and Scientific Computing » "Bioinformatics - Trends and Methodologies", ISBN 978-953-307-282-1., Published: November 2, 2011 under CC BY 3.0 license, DOI: 10.5772/23535

Butler, M.S., 2005, Nat. Prod. Rep. 22:162

May, M, Jun. 13, 2014, Big biological impacts from big data, Science Magazine, DOI:10.1126/science.opms.p1400086, Retrieved on February 14, 2018

Lesk, A. M, 2013, "Bioinformatics", Encyclopaedia Britannica. Retrieved on February 12, 2018

Quinn, 2012, in Chemical Genomics, Edited by Haian Fu, Emory University School of Medicine, Cambridge University Press

State of the World's Plants 2017, Royal Botanical Garden Report, 2016. Retrieved on December 21, 2017

UNEP, 2004, Plant data is from the World Conservation Monitoring Centre of the United Nations Environment Programme Species Data. Fish: Fishbase; Birds: Birdlife International; Amphibians: AmphibiaWeb; Mammals: IUCN; Reptiles: the Reptile Database. Retrieved February 12, 2018

Appendix

Tabel 1. Ten Top Countries with Their Respective Biodiversity in Numbers

Country	Birds	Amphibians	Mammals	Reptiles	Fish	Vascular plants
Brazil	1753	1022	648	807	4521	56215
Colombia	1826	771	442	601	2053	51220
China	1240	411	551	478	3330	32200
Indonesia	**1615**	**347**	**670**	**728**	**4682**	**29375**
Mexico	1081	377	523	916	2602	26071
South Africa	755	128	297	447	2059	23420
Venezuela	1364	360	363	405	1709	21073
United States	844	300	440	530	3067	19473
India	1180	390	412	689	2465	18664
Ecuador	1588	539	372	444	1098	19362
Australia	705	242	349	1038	4869	15638
Peru	1804	572	467	480	1544	17144
Bolivia	1427	244	363	310	392	17367
Malaysia	702	260	336	484	1930	15500
Papua New Guinea	717	367	269	273	2807	11544
Thailand	925	140	311	434	2123	11625
Costa Rica	856	202	227	263	1106	12119
Viet Nam	823	225	287	463	2411	10500
DR Congo	1087	244	430	294	1480	11007
Tanzania	1052	204	359	359	1750	10008
Philippines	569	113	190	207	3291	8931
Russia	256	33	300	89	948	11400
Panama	879	214	246	272	1399	9915
Argentina	998	172	374	437	1002	9372
Madagascar	244	311	228	414	1147	9505
Cameroon	874	218	335	283	1042	8260
Guatemala	706	163	220	262	896	8681
Japan	439	79	144	98	4020	5565
Turkey	390	37	149	148	790	8650
Laos	690	98	215	170	563	8286
Nicaragua	675	71	203	198	1071	7590
Myanmar	1013	91	294	307	1040	7000

ISSN (P) : 2615 - 2584
ISSN (E) : 2615 - 3343

Iran	471	21	186	332	632	8000
Kenya	1034	114	376	280	1045	6506
Paraguay	691	79	165	184	258	7851
Guyana	789	135	235	185	982	6409
Mozambique	671	87	236	222	1752	5692
Nepal	809	45	182	147	291	6973
Gabon	606	96	182	138	774	6651
Uganda	987	60	319	175	264	6500
Cuba	317	68	65	175	1099	6522
Ethiopia	806	65	271	238	170	6603
Honduras	697	122	212	270	1017	5680
Congo	606	75	197	344	771	6000
French Guiana	707	108	207	166	955	5625
Angola	905	99	285	245	895	5185
Suriname	698	113	210	58	1041	5018
Brunei	411	12	138	28	487	6000
Nigeria	854	115	285	199	772	4715
Chile	441	63	143	150	766	5,284
Cambodia	509	50	164	186	933	5000
Italy	352	48	113	58	618	5599
Kazakhstan	440	14	155	50	115	6000
Pakistan	603	22	191	208	707	4,950
Dominican Republic	246	49	49	124	527	5657
Bangladesh	596	35	139	130	557	5,000
Spain	378	37	115	69	791	5,050
Zambia	726	87	234	185	401	4,747
Bhutan	609	7	99	62	49	5468
Haiti	258	58	49	123	526	5,242
Greece	346	24	101	70	609	4,992
Portugal	308	21	90	42	546	5,050
France	354	42	123	53	629	4,630
Ukraine	317	21	112	22	239	5,100
Zimbabwe	623	59	200	190	150	4,440
Ghana	681	86	257	181	687	3725
Guinea	629	74	223	595	979	3000
CÃ´te d'Ivoire	666	80	252	153	671	3660
Tajikistan	325	5	68	48	3	5,000
Uzbekistan	348	2	83	59	85	4,800
Sri Lanka	375	121	117	224	1100	3314
Malawi	630	77	189	125	464	3,765
Canada	516	46	202	54	1097	3270
Croatia	308	16	95	40	398	4,288
Georgia	280	13	102	154	130	4,350

Source: Plant data is from the World Conservation Monitoring Centre of the United Nations Environment Programme (UNEP-WCMC), 2004. Species Data. Fish: Fishbase; Birds: Birdlife International; Amphibians: AmphibiaWeb; Mammals: IUCN; Reptiles: the Reptile Database.