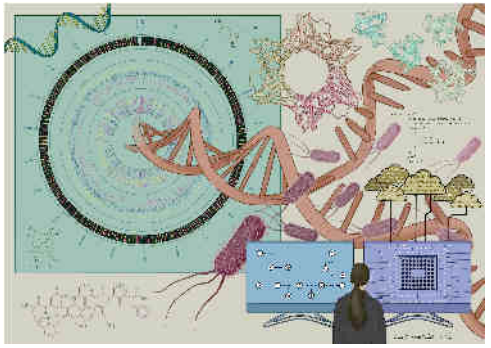
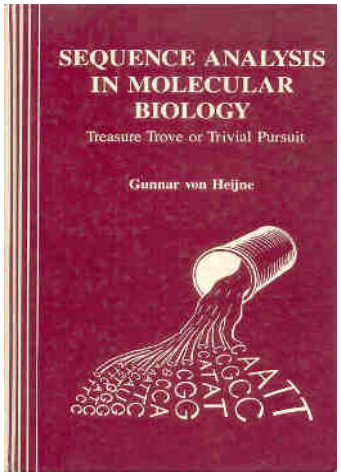
OMICS AND BIOINFORMATICS: STRATEGIC VISION OF THE PAST AND CHALLENGES FOR THE FUTURE

Ana Tereza Vasconcelos

Laboratório Nacional de Computação Científica LNCC/MCTI



1988 – Biofísica UFRJ/LNCC



G. von Heijne (1987)

Prof. Darcy F. de Almeida – UFRJ

Bioinformatics : research *in silico*

(A. Danchin)

54. **VASCONCELOS, A. T. R.**; ALMEIDA, Darcy Fontoura de ; COIMBRA, C. A. . Procaryote DNA damage-inducible SOS system genes detected by nucleotide sequence analyses. In: Academia Brasileira de Ciências, 1994, Rio de Janeiro. An. Academia Brasil Ci. v. 66. p. 127.

14 Dezembro | 15 Dezembro | 16 Dezembro

10:20 - 11:00 — Conferência: Genomics – a Brazilian vocation

Andrew Simpson, Orygen Biotecnologia, São Paulo, SP

11:00 - 12:30 — Painel: Discussões com empreendedores do ramo da genômica e bioinformática

Moderador: Maria Suell Soares Felipe, UnB, Brasília, DF
Gonçalo Amarante Guimarães Pereira, Unicamp, Campinas, SP
João Bosco Oliveira, Genomika Diagnósticos
Luiz Felipe Valter de Oliveira, BiomeHub e Neoprospecta
Michael Sutterfield, Centro de Tecnologia Canavieira, Piracicaba, SP
Paulo Arruda, Unicamp, Campinas, SP

12:30 - 14:30 — Almoço

14:30 - 15:00 — Conferência: Visão geral de CT&I no Brasil

Ildeu de Castro Moreira, Presidente da SBPC, São Paulo, SP

15:30 – 17:00 — Painel: Discussão com líderes de agências e ministros

Moderador: Augusto Schrank, UFRGS, Porto Alegre, RS
Marcelo Marcos Morales - Representante do MCTIC
Marcela Colognesi de Sa - Coordenadora COBRG – CNPq
Marie-Anne Van Sluys – Representante da FAPESP
André Marco de Oliveira Gomes - Representante da FAPERJ
Paulo Sérgio Lacerda Beirão - Diretor da FAPEMIG, Belo Horizonte, MG
Marco Aurélio Krieger - Vice-Presidente da Fiocruz
Carlos Frederico Martins Menck - Coordenador da CAPES/CB1, USP, São Paulo, SP

14 Dezembro | 15 Dezembro | 16 Dezembro

9:30 – 11:00 — Painel: Recursos humanos em genômica e bioinformática

Moderadora: Glória Regina Franco, UFMG, Belo Horizonte, MG
Ana Carolina Ramos Guimarães, Fiocruz, Rio de Janeiro, RJ
Ana Tereza Ribeiro de Vasconcelos, LNCC, Petrópolis, RJ
Glória Regina Franco, UFMG, Belo Horizonte, MG
João Paulo Matos Santos Lima, UFRN, Natal, RN
Alan Mitchell Durham, IME/USP, São Paulo, SP
Pedro Geraldo Pascutti, UFRJ, Rio de Janeiro, RJ

11:00 – 12:30 — Painel: Bioinformática: Oportunidades e desafios

Moderador: Sandro José de Souza, UFRN
Helder Takashi Imoto Nakaya, USP
João Paulo Kitajima, Mendelics Análise Genômica, São Paulo, SP
Paulo Costa Carvalho, Fiocruz, Curitiba, PR
Richard Charles Garrat, USP/São Carlos

12:30 - 14:00 — Almoço

14:00 - 14:30 — Conferência: Percepção em relação à bioinformática e genômica

Luisa Massarani, Fiocruz, Rio de Janeiro, RJ

14:30 – 15:00 — Conferência: A Revolução Genômica na Medicina

Sérgio Danilo Junho Pena, UFMG, Belo Horizonte, MG

15:00 - 16:30 — Painel: O impacto da genômica no cenário de C&T no Brasil

Moderador: Ana Tereza Ribeiro de Vasconcelos, LNCC, Petrópolis, RJ
Ana Maria Benko Iseppon, UFPE, Recife, PE
Marie-Anne Van Sluys, USP, São Paulo, SP
Santuza Maria Ribeiro Teixeira, UFMG, Belo Horizonte, MG

14 Dezembro | 15 Dezembro | 16 Dezembro

9:30 - 11:00 — Painel: Tópicos Avançados em genômica e bioinformática

Moderador: Emmanuel Dias-Neto, AC Camargo Câncer Center, São Paulo, SP
Ândrea Kely Campos Ribeiro dos Santos, UFPA, Belém, PA
Anete Pereira Souza, Unicamp, Campinas, SP
Arnaldo Zaha, UFRGS, Porto Alegre, RS
Célia Maria de Almeida Soares, UFG, Goiânia, GO
Roberto Lins, IAM FIOCRUZ, Recife, PE
Lygia da Veiga Pereira, USP, São Paulo, SP

Painel: Futuro da genômica e bioinformática

Moderador: Lucymara F. Agnez Lima, UFRN, Natal, RN
Eduardo Emrich Soares, Biominas, Belo Horizonte, MG
Elizabeth Pacheco Batista Fontes, UFV, Viçosa, MG
Guilherme Corrêa de Oliveira, ITV, Belém, PA
Renato Santana de Aguiar, UFMG, Belo Horizonte, MG

Encerramento

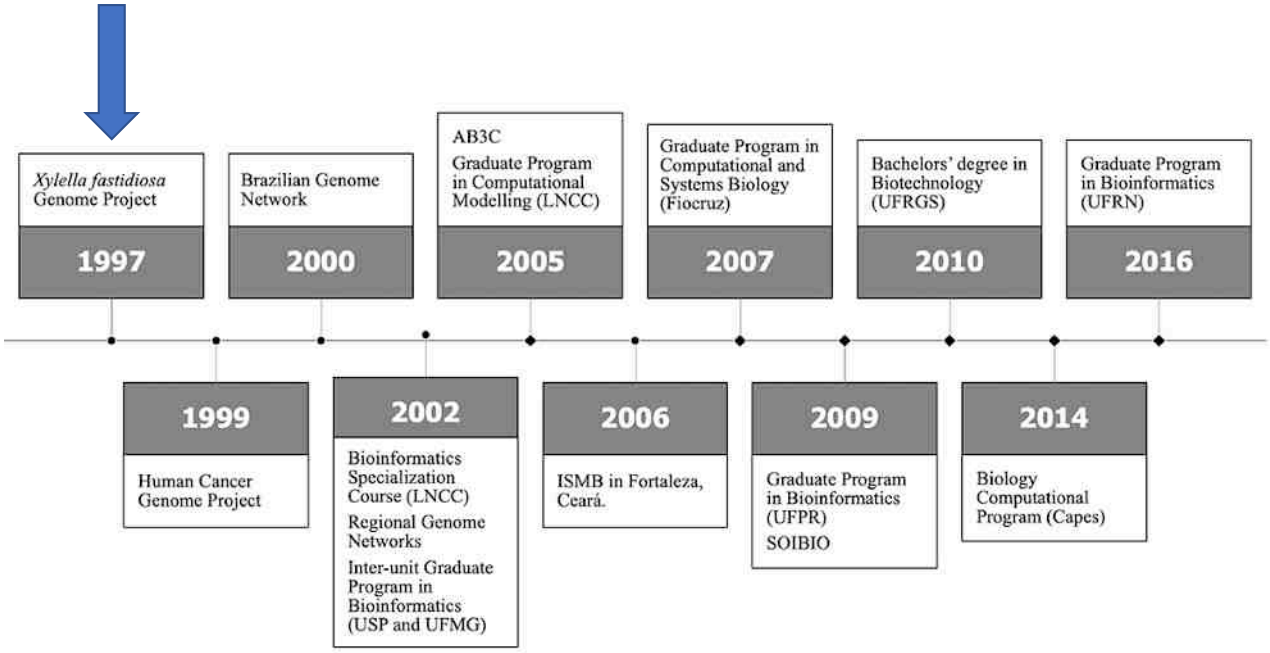
Research Article
Genomics and Bioinformatics

The past, present and future of genomics and bioinformatics: A survey of Brazilian scientists

Mariana Rocha[1], Luisa Massarani[2], Sandro José de Souza[3,4,5] and Ana Tereza R. de Vasconcelos[6]

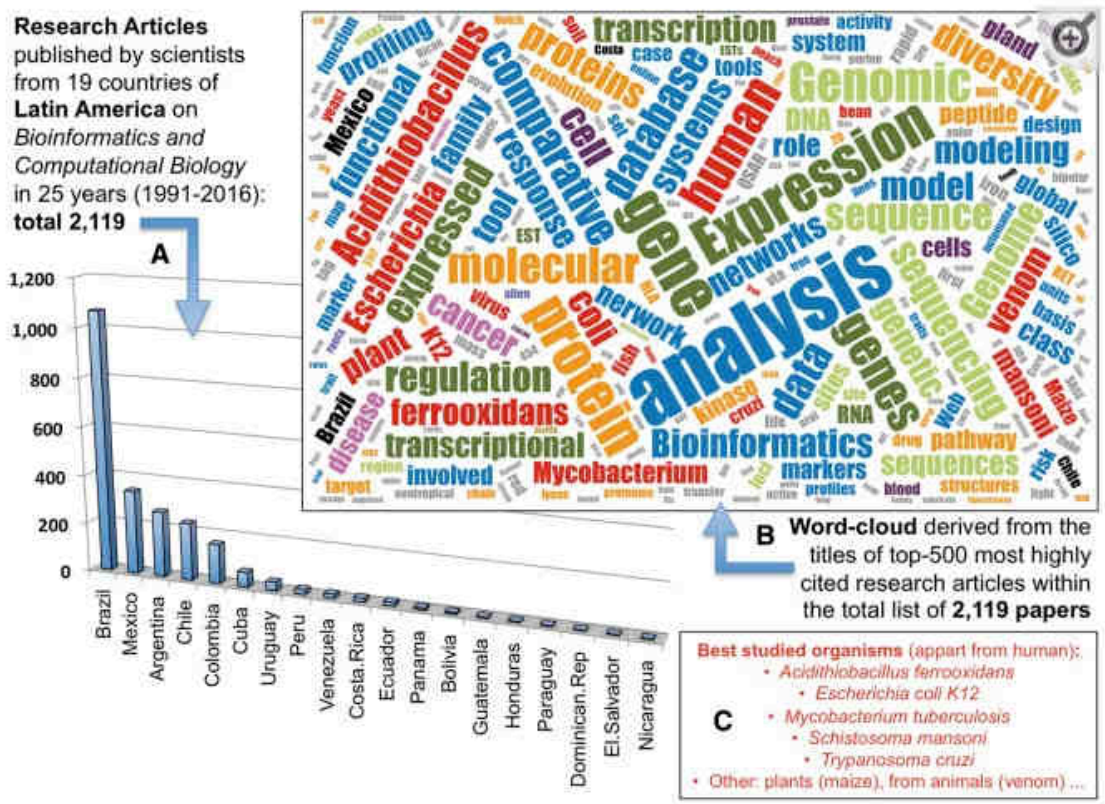
Sub-questions (SQ)	Objectives
SQ1: What is the profile of the researchers currently working in the field?	To collect and evaluate details about genomics and bioinformatics professionals considering demographic features, educational background, and professional experience
SQ 2: What are the perceptions about the present and future? How does it confront the perceptions about the area in other countries?	To collect and evaluate genomics and bioinformatics professionals' thoughts on the current and future situation of the field both in Brazil and abroad.
SQ 3: What are the milestones achieved in the area, and what are the expert wishes for the future?	To collect and evaluate information about the country's main achievements and what should be considered for future work.

https://doi.org/10.1590/1678-4685-GMB-2021-0354



https://doi.org/10.1590/1678-4685-GMB-2021-0354

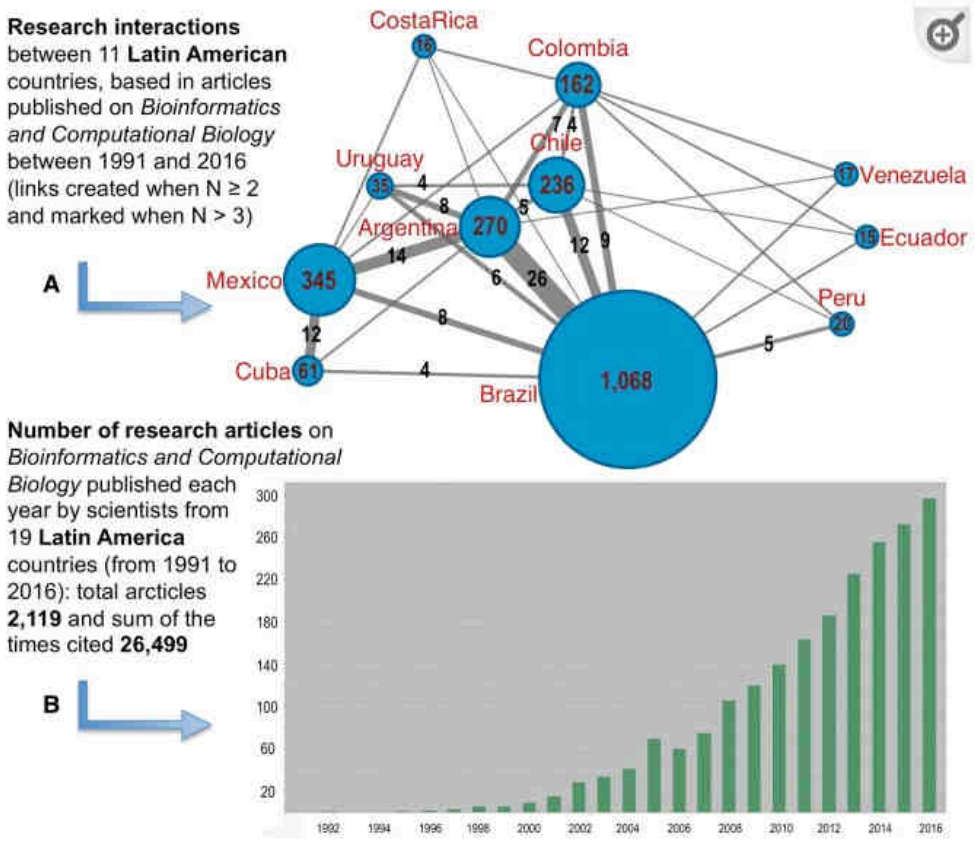
Figure 1



(**A**) Barplot presenting the number of research articles per country published by scientists from 19 countries of LA on bioinformatics and computational biology in 25 years (from 1991 to 2016). The total number found was: 2119. (**B**) Word cloud derived from the titles of top 500 most highly cited research articles within the total list of 2119 papers. (**C**) Selection of some of the organisms studied by scientists of LA.

Brief Bioinform. 2019 Mar; 20(2): 390–397.

Figure 2

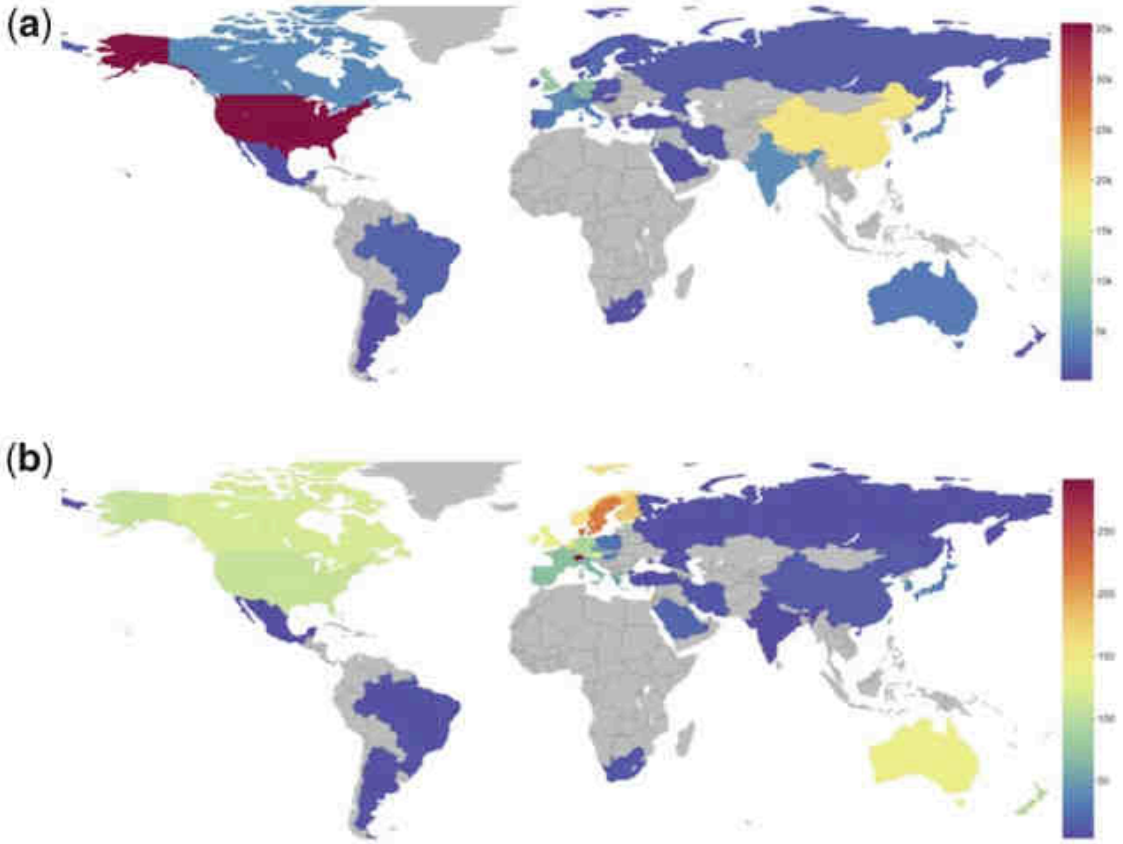


(**A**) Network presenting the research interactions between 11 LA countries, based in articles published on bioinformatics and computational biology between 1991 and 2016 (links created when the number of papers is $N \geq 2$ and marked when $N > 3$). (**B**) Number of research articles on bioinformatics and computational biology published each year by scientists from 19 LA countries (from 1991 to 2016). The total number of articles was 2119 and the total number of times that have been cited: 26 499.

Images in this article

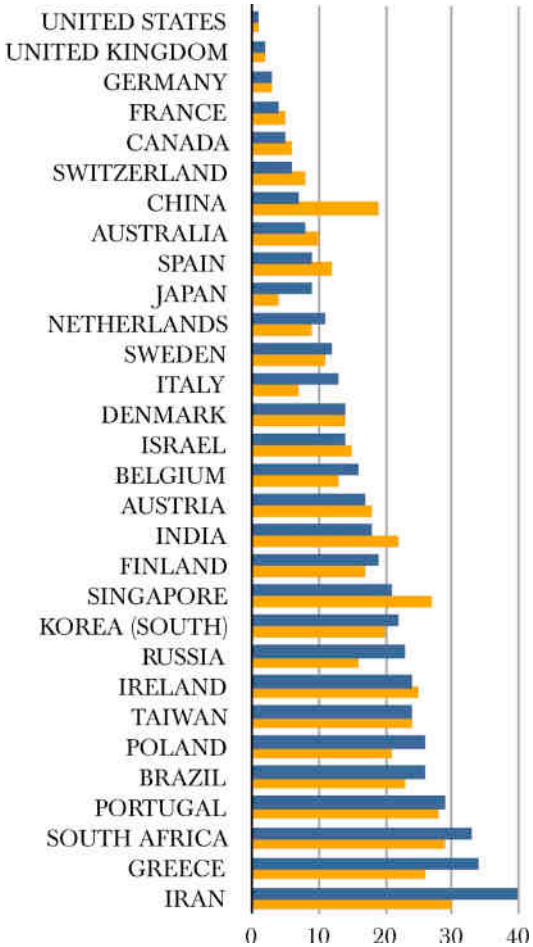
Brief Bioinform. 2019 Mar; 20(2): 390–397.

Fig. 2.



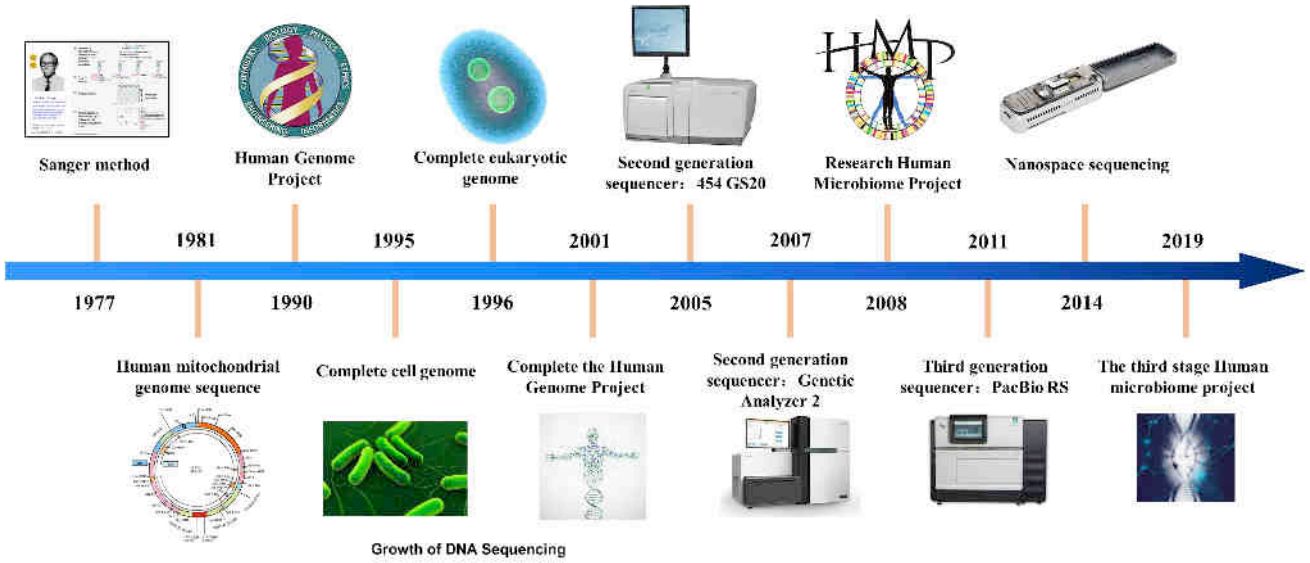
A world map depiction highlighting the top 40 countries in bioinformatics, based on publications output. (**a**) Absolute numbers of bioinformatics publications (scale provided, right), (**b**) relative number of bioinformatics publications per capita (million inhabitants). See Supplementary Table S2 for a full list of 40 countries. Figure generated by Displayr (www.displayr.com)

Bioinformatics. 2020 May 1; 36(9): 2963–2965.



Rank listing of leading countries in bioinformatics and top 1% of highly cited papers.
Bioinformatics h-index rank (blue), top 1% highly cited papers rank (orange)—lower is better.

Bioinformatics. 2020 May 1; 36(9): 2963–2965.



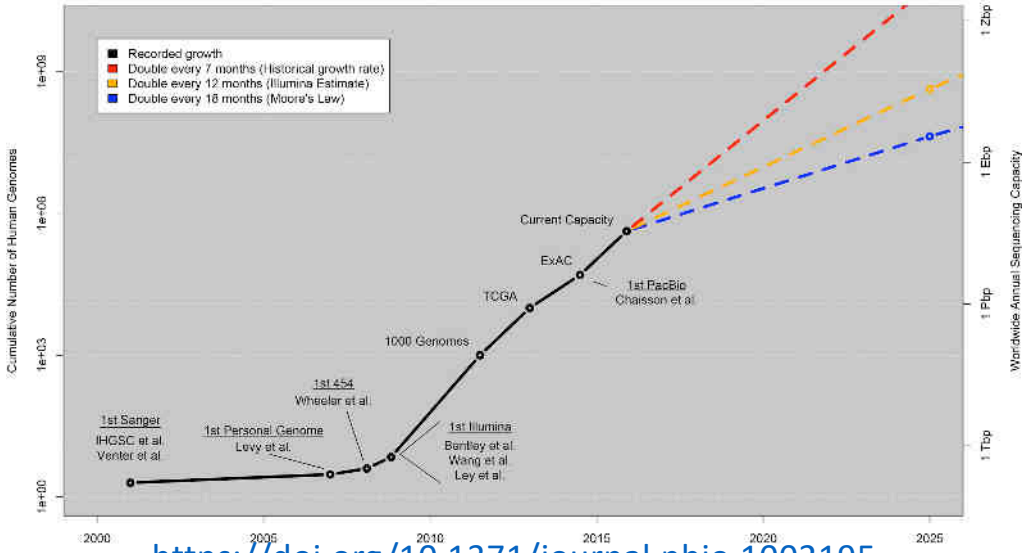
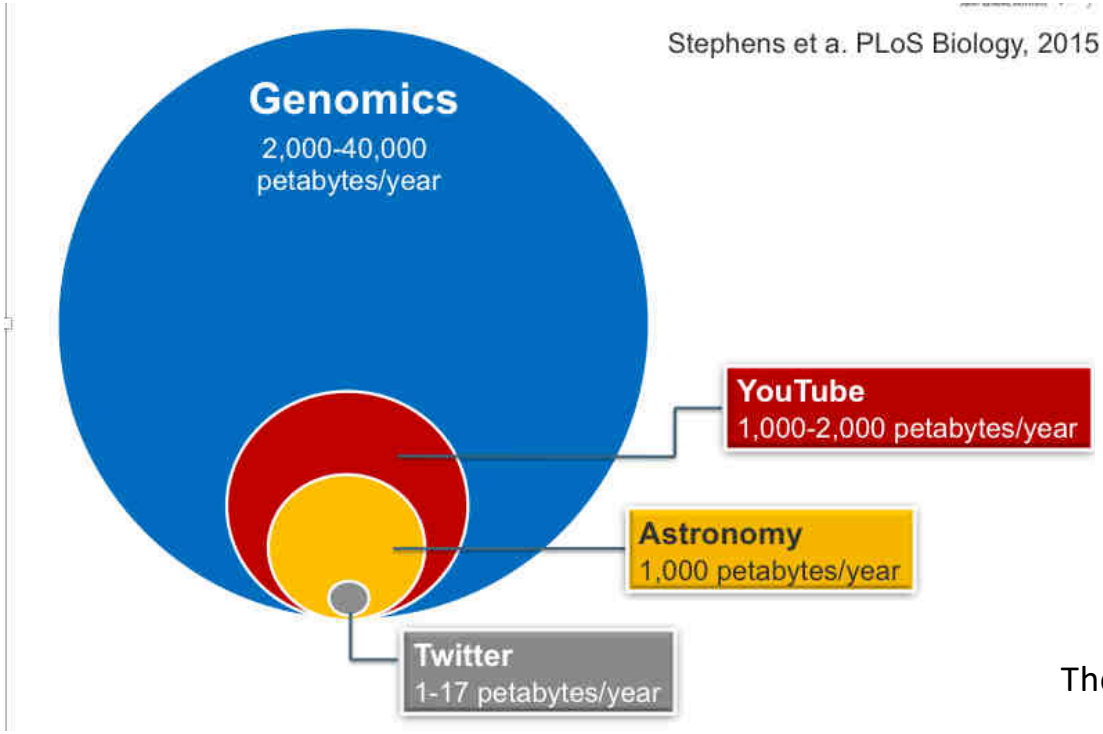
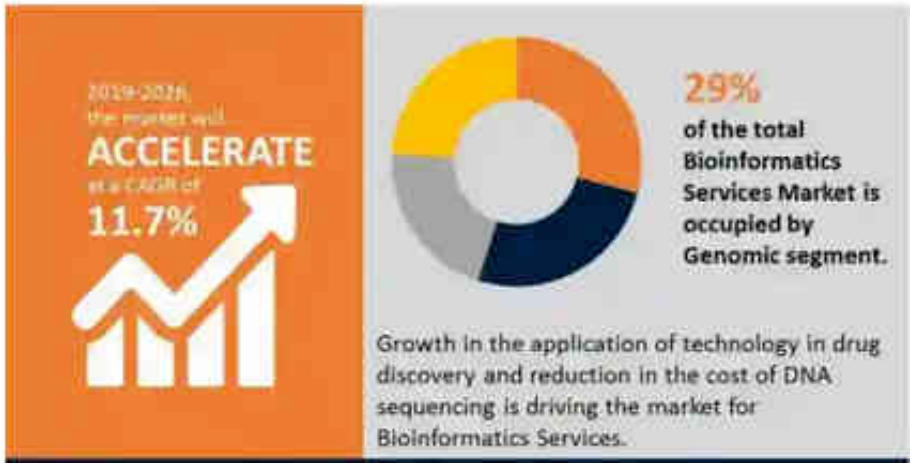


Fig 1. Growth of DNA sequencing.
The plot shows the growth of DNA sequencing both in the total number of human genomes sequenced (left axis) as well as the worldwide annual sequencing capacity (right axis: Tera-basepairs (Tbp), Peta-basepairs (Pbp), Exa-basepairs (Ebp), Zetta-basepairs (Zbps)).

https://doi.org/10.1371/journal.pbio.1002195



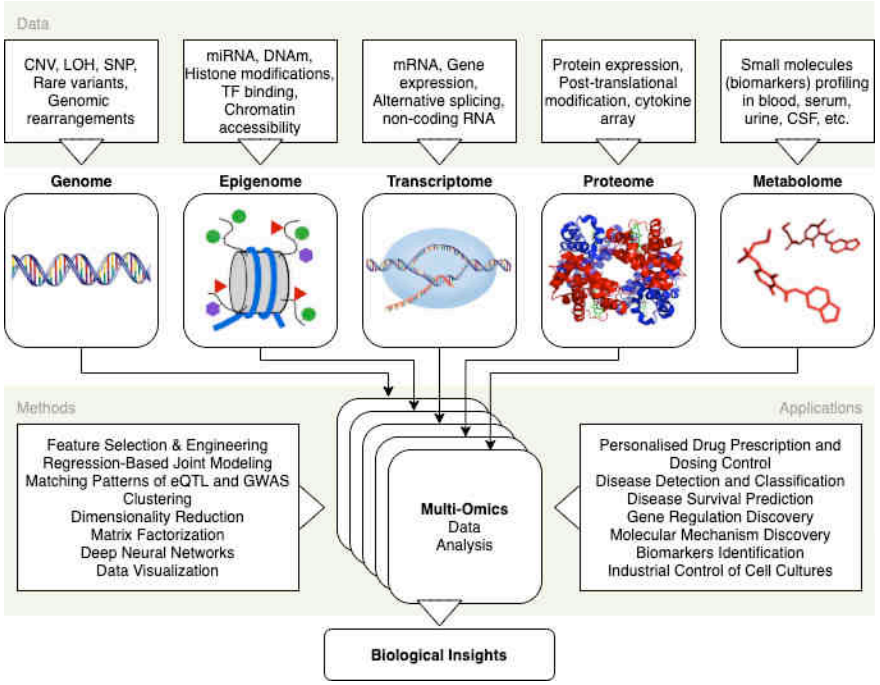


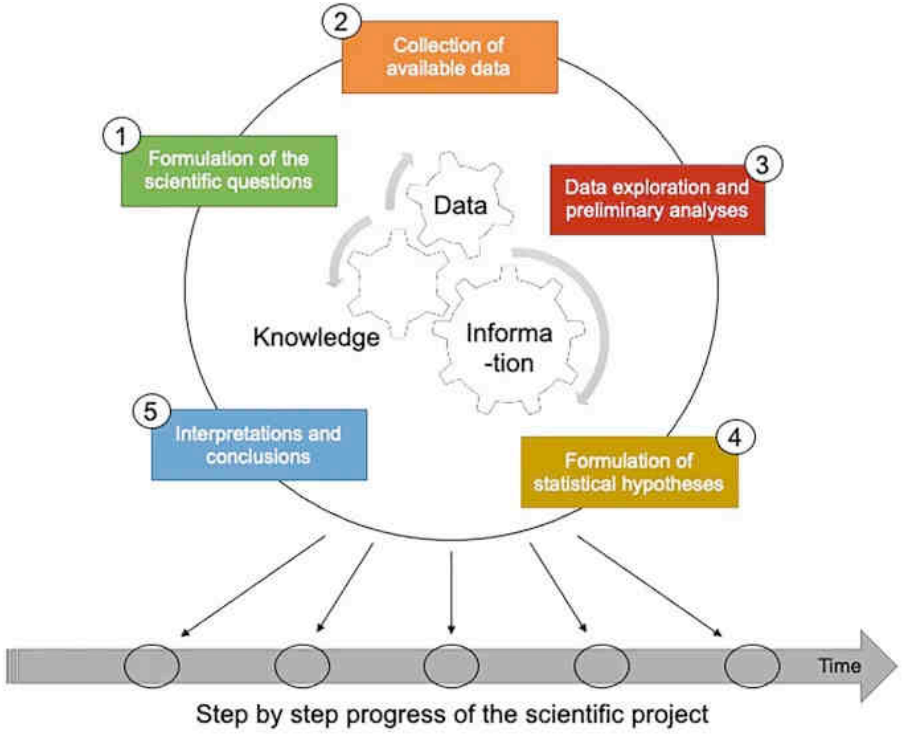
The Global Bioinformatics Services Market is forecast to reach USD 6.44 Billion by 2028.

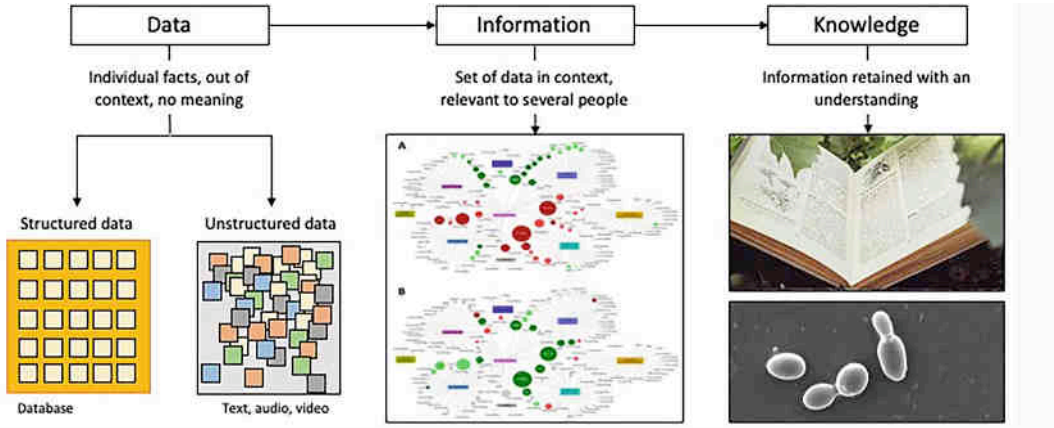
https://www.reportsanddata.com/report-detail/bioinformatics-services-market

"Projecting to the year 2025, we compared genomics with three other major generators of Big Data: astronomy, YouTube, and Twitter. Our estimates show that genomics is a "four-headed beast"—it is either on par with or the most demanding of the domains analyzed here in terms of data acquisition, storage, distribution, and analysis."

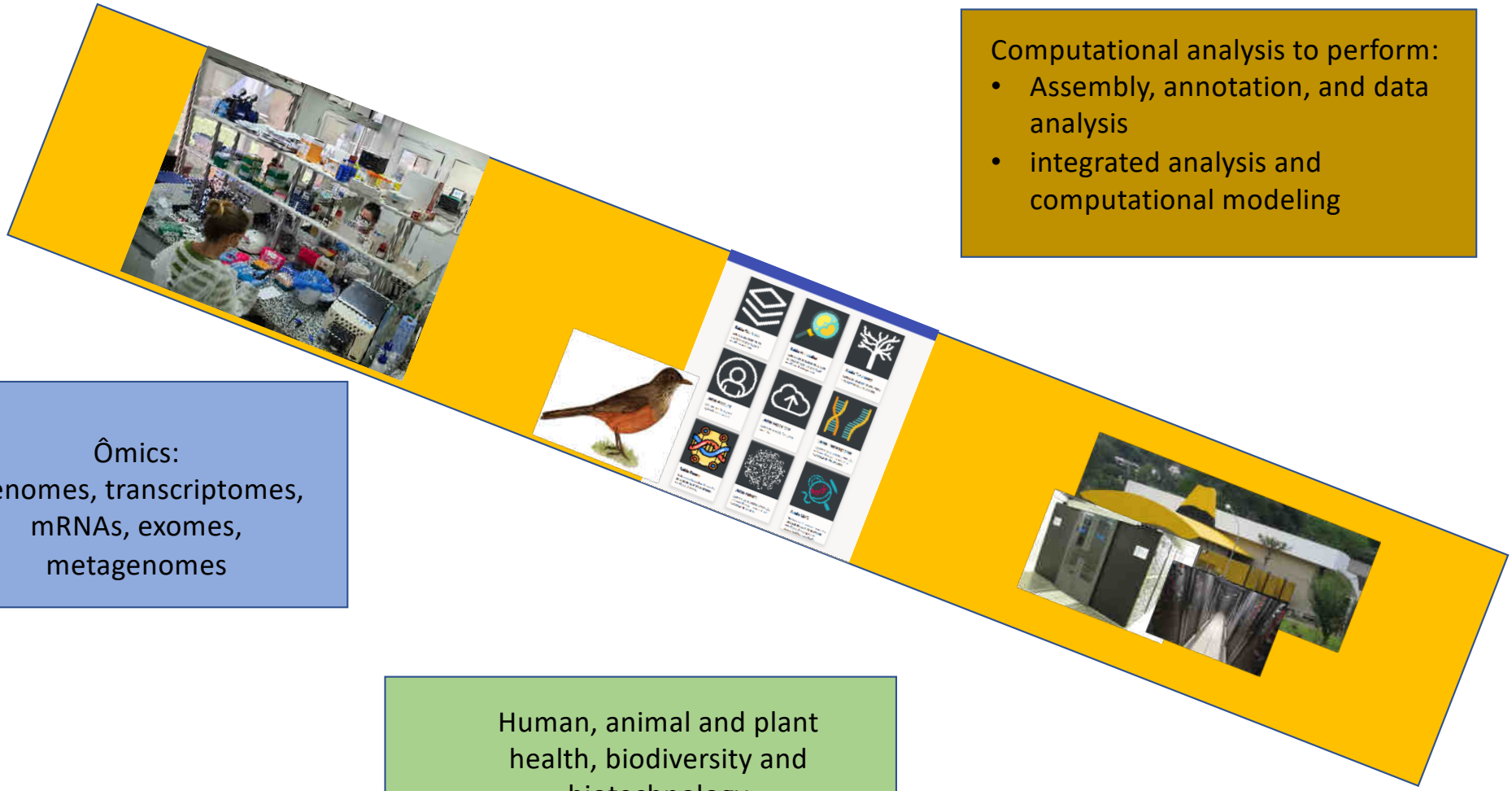
http://www.a2idea.com/omics-analyses







https://doi.org/10.1007/978-1-0716-2257-5_25



Computational analysis to perform:

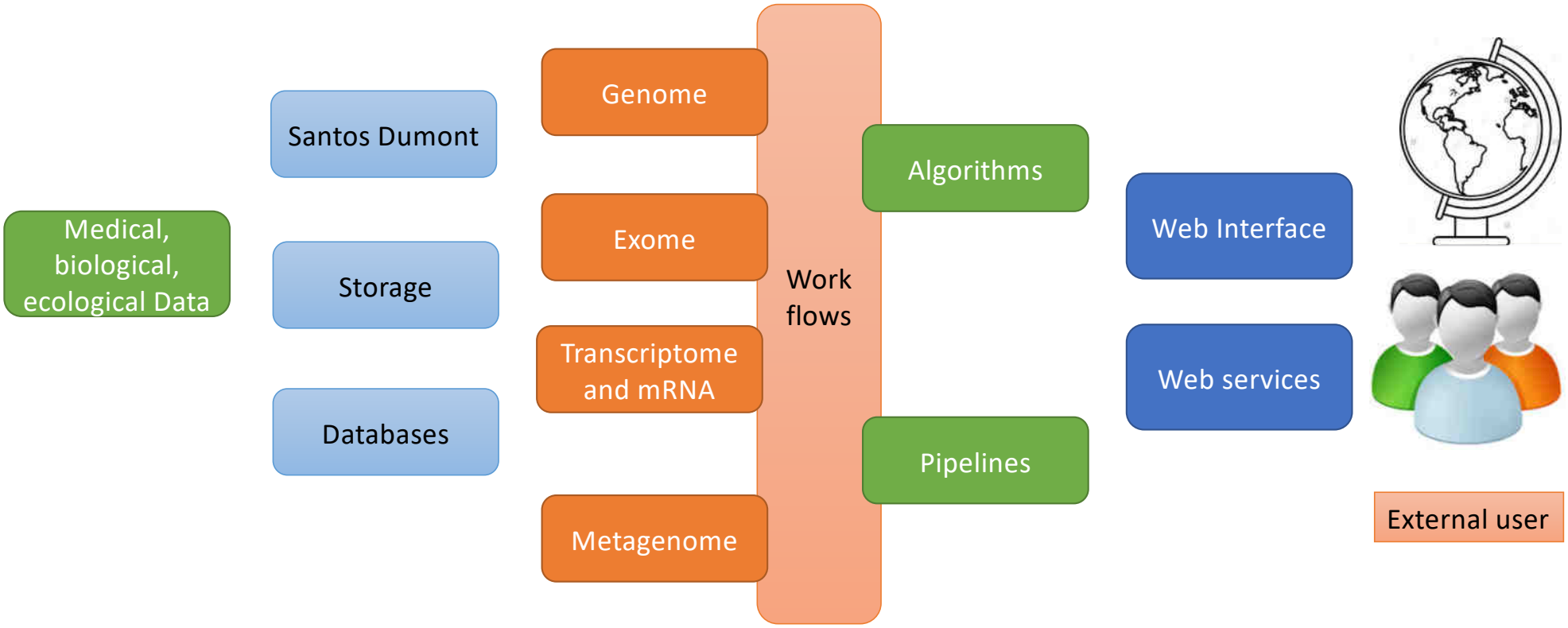
- Assembly, annotation, and data analysis
- integrated analysis and computational modeling

Ômics:
Genomes, transcriptomes, mRNAs, exomes, metagenomes

Human, animal and plant health, biodiversity and biotechnology

Laboratório de BioInformática

Laboratório Nacional de Computação Científica


Medical, biological, ecological Data
Santos Dumont
Storage
Databases
Genome
Exome
Transcriptome and mRNA
Metagenome
Work flows
Algorithms
Pipelines
Web Interface
Web services
External user
DATA
HPC
TOOLS
INTERFACE

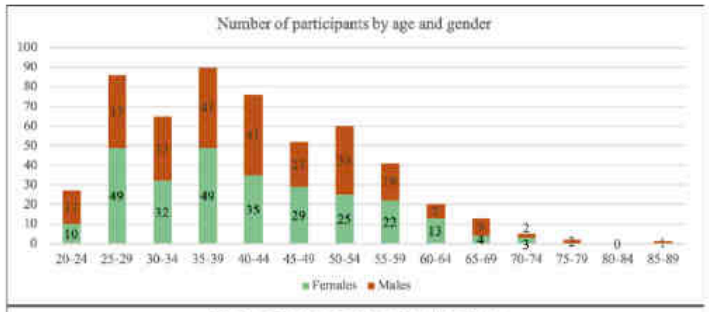
Research Article
Genomics and Bioinformatics

The past, present and future of genomics and bioinformatics: A survey of Brazilian scientists

Mariana Rocha[1], Luisa Massarani[2], Sandro José de Souza[3,4,5] and Ana Tereza R. de Vasconcelos[6]

The data was collected through an online questionnaire containing 25 questions, 19 closed-ended and six open-ended. The first closed-ended question asks how long the respondent is working in the area of genomics and bioinformatics. If the participant selected the option that states s/he is not working in the area, the questionnaire was then ended.

A



B

C

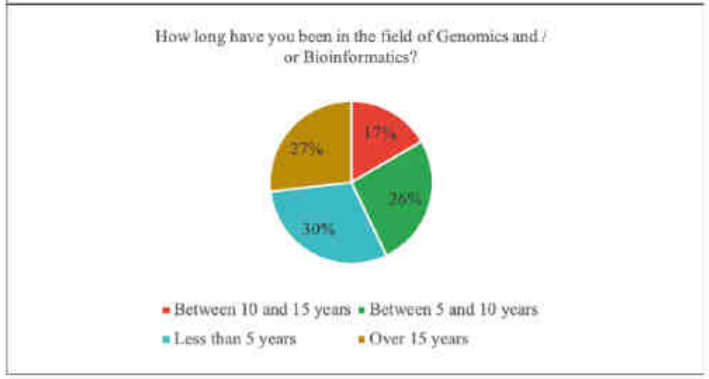


Table 2 - Answer to the question: "Mark in the table below, for all levels of education, if you are attending or have already completed and in case you have already completed what was the year of completion".

Time since completion	Bachelor's degree	Specialisation	Master's degree	PhD	Post-doctorate
Over than 10 years ago	64.5%	18.9%	44.9%	41.0%	22.6%
Between 05 and 10 years	15.5%	2.8%	13.3%	14.6%	12.2%
Up to 05 years	17.2%	3.5%	19.2%	14.4%	12.8%
Over 3 years ago	0.2%	0.0%	0.0%	0.2%	0.0%
Studying	0.9%	1.1%	7.8%	16.1%	10.2%

Figure 2 - Word cloud illustrating the most frequent words occurring in the participants' responses referring to the current situation of the field of genomics and/or bioinformatics in Brazil.

Table 6 - Examples of sentiment analysis classification results performed on the future situation of the field in Brazil.

Sentiment classification	%	Example
Positive	46.0%	Case 38 - Extremely promising, with the emergence of new research and investment groups from the private sector in the creation of new companies specialised in the area.
Neutral	41.0%	Case 453 - I believe that all professionals will have to know the minimum bioinformatics for the development of their research in the near future.
Negative	13.0%	Case 107 - Difficult, resources are still limited and more collaboration is lacking, including data availability

Table 9 - Categories adopted to classify the milestones of genomics and/or bioinformatics in Brazil.

Category	Description	Percentage
Programs and projects	Implementation of programs and projects in the area, such as the Human Genome Project.	87.0%
Sequencing of organisms	Highlights of the main organisms sequenced, such as those relevant to areas like agriculture and health	66.0%
Technology	Development of new technologies, such as sequencers, high-performance computers and servers.	60.0%
Funding	More financial investment on research in the area	4.0%
Education and training	Development of formal education courses (bachelor's and post-graduation degree) and training programs.	13.0%
Industry and institutions	Creation of new institutions and companies /start-ups in the area	15.0%
Techniques	Development of new sequencing techniques.	10.0%

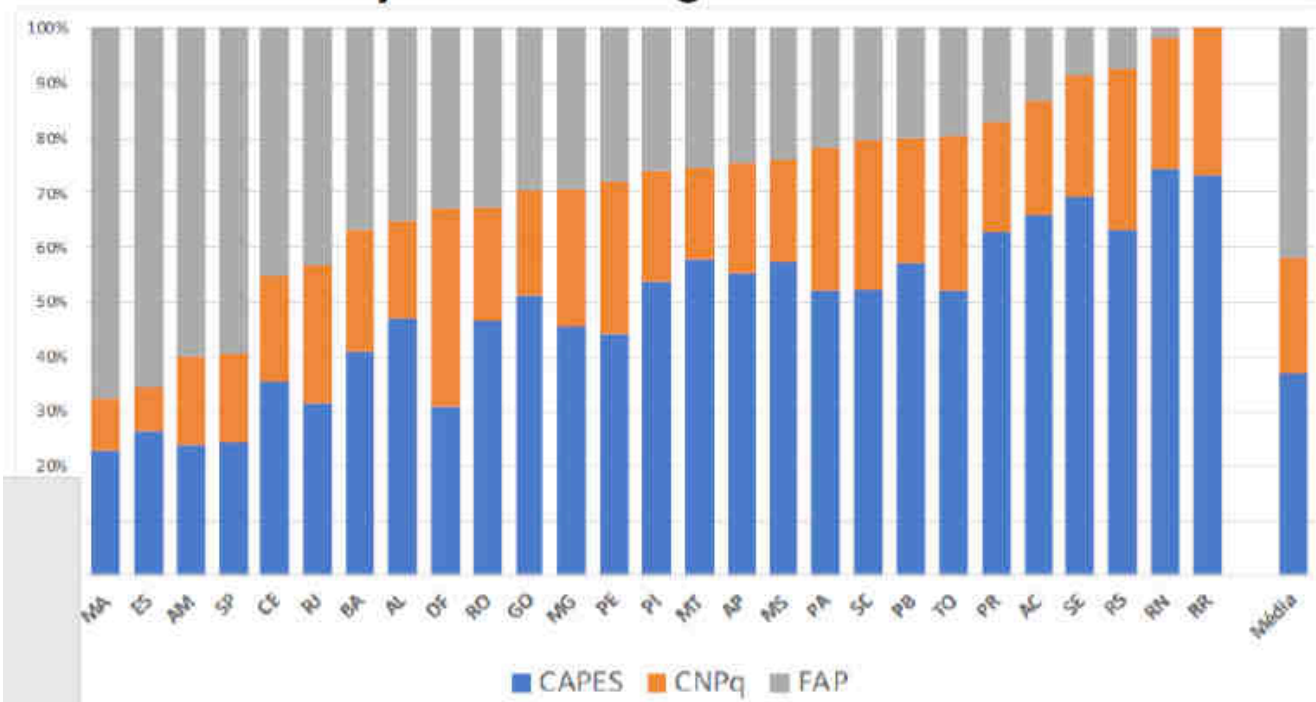
Table 10 - Categories adopted on the classification of the respondents' wish lists.

Category	Description	Percentage
Funding	Responses state the need to funding for research and development in the area. It also includes the reduction of the price of research inputs and equipment, besides investment in the industry providing larger teams and start-ups creation.	56.8%
Formal education	Responses related to the need of implementing new bachelor's, master's and PhD degrees, besides training for specific technologies.	48.2%
Collaboration	States the wishes for more collaboration between research groups, institutions and private companies/industry.	33.3%
Infra-structure	Comments about the need of new equipment researchers would like to have in their hosting university or company, such as supercomputers or research inputs.	32.5%
Career	A higher number of job positions for researchers and professionals in the area, higher salaries etc.	9.6%
Communication	Better communication about the area to the general public, besides including aspects related to the area to the basic school curriculum.	5.7%

	Média 2018-2020	Média 2018-2020	Média 2018-2020			
UF	CNPq	CAPES	FAPs	Total (Soma CAPES+CNPq+FAP)	PIB em 2018 (1.000.000 R$)	% do PIB
AC	R$ 1.379.359,89	R$ 4.370.800,00	R$ 861.205,50	R$ 6.611.365,39	15.331	0,043%
AL	R$ 6.957.728,68	R$ 18.128.400,00	R$ 13.469.500,15	R$ 38.555.628,83	54.413	0,071%
AM	R$ 14.909.830,35	R$ 22.375.200,00	R$ 55.226.347,14	R$ 92.511.377,49	100.109	0,092%
AP	R$ 1.004.960,66	R$ 2.779.600,00	R$ 1.247.455,99	R$ 5.032.016,66	16.795	0,030%
BA	R$ 36.646.609,68	R$ 68.047.200,00	R$ 61.242.543,41	R$ 165.936.353,09	286.240	0,058%
CE	R$ 31.401.020,40	R$ 57.749.200,00	R$ 73.033.453,66	R$ 162.183.674,06	155.904	0,104%
DF	R$ 64.772.071,68	R$ 55.588.400,00	R$ 59.330.797,59	R$ 179.691.269,28	254.817	0,071%
ES	R$ 9.500.469,30	R$ 31.893.600,00	R$ 78.194.349,20	R$ 119.588.418,51	137.020	0,087%
GO	R$ 17.252.596,73	R$ 46.105.600,00	R$ 26.716.504,47	R$ 90.074.701,20	195.682	0,046%
MA	R$ 5.415.352,60	R$ 13.118.800,00	R$ 38.926.358,89	R$ 57.460.511,49	98.179	0,059%
MG	R$ 126.631.095,86	R$ 231.575.200,00	R$ 148.493.413,14	R$ 506.699.709,00	614.876	0,082%
MS	R$ 9.707.091,55	R$ 30.196.400,00	R$ 12.533.147,23	R$ 52.436.638,78	106.969	0,049%
MT	R$ 7.373.514,78	R$ 24.840.400,00	R$ 10.839.015,00	R$ 43.052.929,78	137.443	0,031%
PA	R$ 27.145.510,65	R$ 53.408.000,00	R$ 22.308.096,18	R$ 102.861.606,82	161.350	0,064%
PB	R$ 27.438.942,33	R$ 68.868.400,00	R$ 24.157.379,02	R$ 120.464.721,35	64.374	0,187%
PE	R$ 51.746.265,08	R$ 82.034.000,00	R$ 51.641.429,77	R$ 185.421.694,85	186.352	0,100%
PI	R$ 5.783.512,29	R$ 15.434.400,00	R$ 7.452.378,57	R$ 28.670.290,85	50.378	0,057%
PR	R$ 55.863.983,16	R$ 175.161.600,00	R$ 47.511.898,77	R$ 278.537.481,92	440.029	0,063%
RJ	R$ 222.297.441,18	R$ 273.772.800,00	R$ 375.992.133,33	R$ 872.062.374,51	758.859	0,115%
RN	R$ 16.595.125,62	R$ 50.948.000,00	R$ 1.147.320,00	R$ 68.690.445,62	66.970	0,103%
RO	R$ 2.196.453,31	R$ 4.944.400,00	R$ 3.465.914,96	R$ 10.606.768,28	44.914	0,024%
RR	R$ 984.256,60	R$ 2.708.400,00	R$ -	R$ 3.692.656,60	13.370	0,028%
RS	R$ 120.027.993,00	R$ 256.588.400,00	R$ 29.500.000,00	R$ 406.116.393,00	457.294	0,089%
SC	R$ 51.666.115,22	R$ 98.626.400,00	R$ 38.097.839,27	R$ 188.390.354,49	298.227	0,063%
SE	R$ 7.162.072,75	R$ 22.282.000,00	R$ 2.759.182,31	R$ 32.203.255,06	42.018	0,077%
SP	R$ 375.513.871,46	R$ 576.839.600,00	R$ 1.402.209.850,67	R$ 2.354.563.322,13	2.210.562	0,107%
TO	R$ 3.359.402,19	R$ 6.216.400,00	R$ 2.355.903,74	R$ 11.931.705,93	35.666	0,033%
x-Total	R$ 1.300.732.646,99	R$ 2.294.601.600,00	R$ 2.588.713.417,96	R$ 6.184.047.664,95	7.004.141	0,088%

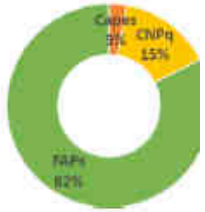
Fonte : Odir Dellagostin

Contribuição de cada agência em cada Estado



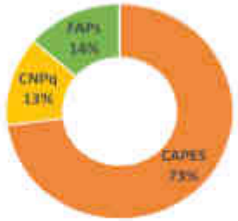
Participação das agências no auxílio à pesquisa - dados de 2020

Agência	Auxílio à Pesquisa em 2020	
Capes	R$	42.495.502,00
CNPq	R$	228.235.575,91
FAPs	R$	1.271.326.649,78

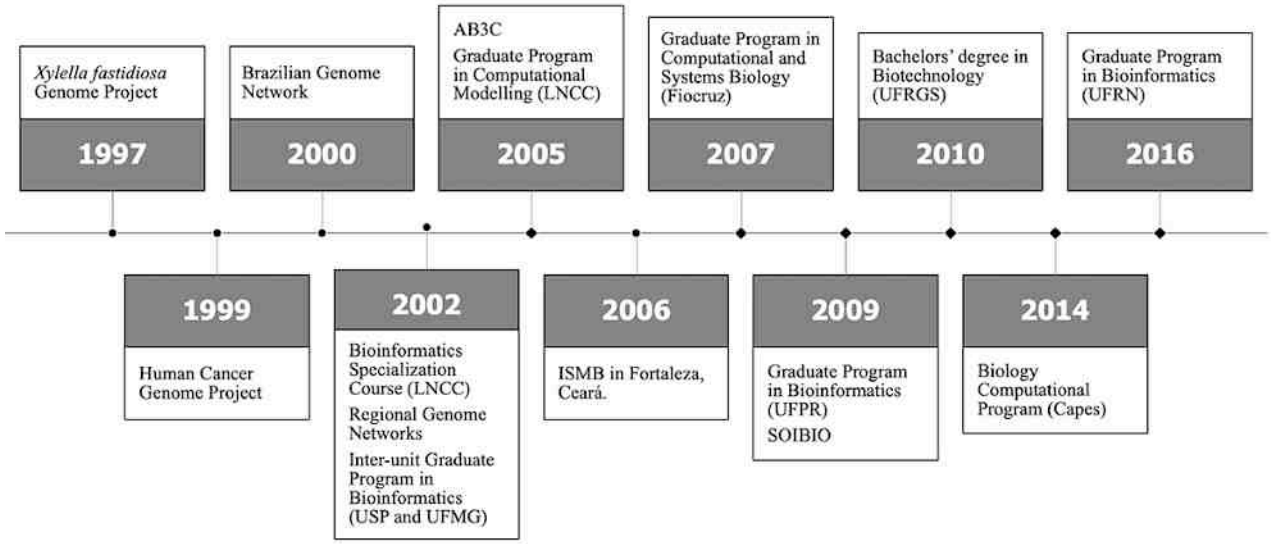


Bolsas de pós-graduação em 2020

Agência	Mestrado	Doutorado
CAPES	43.497	46.105
CNPq	7.906	8.434
FAPs	8.737	8.101
Total	60.140	62.640
Bolsistas	44,2%	50,3%



Obrigado: Odir Dellagostin!


Xylella fastidiosa Genome Project
1997
Brazilian Genome Network
2000
AB3C
Graduate Program in Computational Modelling (LNCC)
2005
Graduate Program in Computational and Systems Biology (Fiocruz)
2007
Bachelors' degree in Biotechnology (UFRGS)
2010
Graduate Program in Bioinformatics (UFRN)
2016
1999
Human Cancer Genome Project
2002
Bioinformatics Specialization Course (LNCC)
Regional Genome Networks
Inter-unit Graduate Program in Bioinformatics (USP and UFMG)
2006
ISMB in Fortaleza, Ceará.
2009
Graduate Program in Bioinformatics (UFPR)
SOIBIO
2014
Biology Computational Program (Capes)


GENOME20+2


Let's
Celebrate


BEM VINDOS
DE VOLTA!

Acknowledgment