



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS**  
**DEPARTAMENTO DE BIOLOGIA**  
**GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS**

**JOSUÉ ROBERTO SANTANA GOMES**

**EXPLORANDO AS FRONTEIRAS DA BIOINFORMÁTICA: UMA REFLEXÃO**  
**SOBRE A ÁREA E SEUS PESQUISADORES**

**FORTALEZA**

**2023**

JOSUÉ ROBERTO SANTANA GOMES

EXPLORANDO AS FRONTEIRAS DA BIOINFORMÁTICA: UMA REFLEXÃO SOBRE  
A ÁREA E SEUS PESQUISADORES

Monografia apresentada ao curso de Graduação em Ciências Biológicas do Departamento de Biologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Ciências Biológicas.

Orientador: Prof. Dr. José Roberto Feitosa

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- G614e Gomes, Josué Roberto Santana.  
Explorando as fronteiras da Bioinformática: uma reflexão sobre a área e seus pesquisadores / Josué Roberto Santana Gomes. – 2023.  
50 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Ciências, Curso de Ciências Biológicas, Fortaleza, 2023.  
Orientação: Prof. Dr. José Roberto Feitosa Silva.
1. Biologia computacional. 2. Dados biológicos. 3. Autônarrativa. 4. Mercado de trabalho. I. Título.  
CDD 570
-

JOSUÉ ROBERTO SANTANA GOMES

EXPLORANDO AS FRONTEIRAS DA BIOINFORMÁTICA: UMA REFLEXÃO SOBRE  
A ÁREA E SEUS PESQUISADORES

Monografia apresentada ao curso de Graduação em Ciências Biológicas do Departamento de Biologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de Bacharel em Ciências Biológicas.

Orientador: Prof. Dr. José Roberto Feitosa

Aprovada em: \_\_/\_\_/\_\_\_\_.

BANCA EXAMINADORA

---

Prof. Dr. José Roberto Feitosa Silva (Orientador)  
Departamento de Biologia - UFC

---

Prof. Dr. Bruno Anderson Matias da Rocha  
Departamento de Bioquímica e Biologia Molecular - UFC

---

Prof. Dra. Denise Cavalcante Hissa  
Departamento de Biologia - UFC

Aos meus pais, Roberto e Mônica, e a minha  
irmã, Lídia.

## AGRADECIMENTOS

Primeiramente a Deus, pois acredito que tudo no universo possui uma força motriz que a rege e, dada as eventualidades da vida, reconheço que a mesma é uma arte complexa projetada por um artista e que deve ser sentida e aproveitada.

Agradeço a minha família, principalmente meus pais. O suor e as lágrimas que derramaram desde o meu nascimento até minha maioridade regaram desde cedo o campo de conhecimento que tenho hoje. Sem eles, nada disso seria possível. A minha namorada, que tornou o meu mundo mais vívido nos últimos tempos, me apoiou quando era necessário e esteve do meu lado nos tempos mais sombrios e difíceis.

Aos meus amigos e colegas dentro e fora da faculdade. Amigos que guardo comigo desde o começo da adolescência, sempre curtindo a vida e falando besteiras quando cabia. Amizades que formei dentro da universidade e que levarei para o resto da vida também, sempre chegando a beira do colapso mental com as cadeiras da graduação e comendo muito no almoço do RU.

Ao professor Roberto, meu orientador, que aceitou o desafio desta tarefa quando ela ainda nem era possível. Vejo neste professor um exemplo de docente dotado de humanidade e empatia. São poucos os profissionais que lecionam na graduação que possui essa qualidade tão essencial na compreensão de que o aluno é um humano com suas peculiaridades, e não apenas mais um número de matrícula. Aos professores Bruno e Denise que se dispuseram a compor a banca deste manuscrito.

Agradeço também a todos que, de alguma forma, contribuíram para a realização deste trabalho acadêmico, sejam eles professores, alunos entrevistados, colegas de curso e pessoas com ou sem diploma. Cada ser vivo é essencial em seu ambiente, e com este manuscrito não foi diferente.

Por fim, agradeço ao meu eu interior, por sempre acreditar ser capaz de superar os desafios que apareceram ao longo de minha vida. Finalizando este enorme ciclo, percebo o quanto a força de vontade e perseverança pode transformar a vida de qualquer pessoa.

“Meu cérebro é apenas um receptor, no Universo existe um núcleo a partir do qual obtemos conhecimento, força e inspiração. Eu não penetrei nos segredos deste núcleo, mas eu sei que ele existe.” (Nikola Tesla)

## RESUMO

A Bioinformática surgiu com o propósito de oferecer soluções a questões biológicas utilizando-se do poderio computacional. Dada a sua importância desde o seu surgimento, a biologia computacional tem se tornado cada vez mais importante no meio científico, representando uma poderosa aliada em metodologias de estudo e pesquisa. Sendo um campo relativamente novo comparado a outros campos de estudo, é notório que graduandos tenham contato com o mesmo em suas graduações. Neste trabalho, foi abordado como a Bioinformática funciona, como suas principais metodologias e vertentes se diferem, além de sua crescente relevância. Também contém uma narrativa autobiográfica e entrevistas com alunos da graduação que possuem vivência na área. Por fim, foi realizado durante um período de um mês uma análise do mercado para o setor no Brasil e nos Estados Unidos. O mercado no exterior apresenta muitas vagas para o campo semanalmente, diferente das vagas no Brasil que são, geralmente, restritas ao ambiente de universidades e empresas vinculadas às mesmas.

**Palavras-chave:** biologia computacional; dados biológicos; autonarrativa; mercado de trabalho.

## ABSTRACT

Bioinformatics emerged with the purpose of offering solutions to biological questions using computational power. Given its importance since its inception, computational biology has become increasingly significant in the scientific community, serving as a powerful ally in study and research methodologies. Being a relatively new field compared to others, it is notable for undergraduate students to be exposed to it during their studies. This work addresses how Bioinformatics functions, how its main methodologies and branches differ, and its growing relevance. It also includes an autobiographical narrative and interviews with undergraduate students who have experience in the field. Finally, an analysis of the market for the sector in Brazil and the United States was conducted over a one-month period. The foreign market regularly offers numerous job opportunities in this field, whereas in Brazil, such positions are usually limited to university environments and affiliated companies.

**Keywords:** computational biology; biological data; autobiographical narrative; job market.

## SUMÁRIO

<b>1</b>	<b>UMA BREVE VISÃO SOBRE A BIOLOGIA .....</b>	<b>11</b>
<b>1.1</b>	<b>O que é ser um biólogo? .....</b>	<b>11</b>
<b>1.2</b>	<b>Biologia como ciência autônoma .....</b>	<b>11</b>
<b>1.3</b>	<b>Legislação do biólogo no Brasil .....</b>	<b>13</b>
<b>1.4</b>	<b>Como o biólogo inicia seus estudos? O papel da universidade .....</b>	<b>15</b>
<b>2</b>	<b>BIOINFORMÁTICA, UMA CIÊNCIA PÓS-MODERNA .....</b>	<b>17</b>
<b>2.1</b>	<b>O que é bioinformática? .....</b>	<b>17</b>
<b>2.2</b>	<b>Importância nas últimas décadas .....</b>	<b>20</b>
<b>2.3</b>	<b>Eventos-chave .....</b>	<b>23</b>
<b>2.4</b>	<b>Gráficos de relevância .....</b>	<b>26</b>
<b>3</b>	<b>EXPERIÊNCIAS E VIVÊNCIAS .....</b>	<b>31</b>
<b>3.1</b>	<b>Como o biólogo inicia na bioinformática? Uma narrativa autobiográfica .</b>	<b>32</b>
<b>3.2</b>	<b>Entrevistas com alunos .....</b>	<b>37</b>
<b>3.3</b>	<b>Pontos convergentes .....</b>	<b>43</b>
<b>4</b>	<b>PERSPECTIVAS DO MERCADO ATUAL .....</b>	<b>44</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>47</b>
	<b>REFERÊNCIAS .....</b>	<b>48</b>

## **1 UMA BREVE VISÃO SOBRE A BIOLOGIA**

### **1.1 O que é ser um biólogo?**

Podemos definir a Biologia como a ciência que estuda a vida, desde a sua simplicidade até sua complexidade. Segundo Alberts et al. (2017), a Biologia “é um contraponto entre dois temas: a admirável variedade em particularidades individuais e a admirável constância nos mecanismos fundamentais”. Podemos então entender ela como a ciência que busca entender a diversidade da vida, como os organismos se desenvolvem e se reproduzem, como as células funcionam e como os organismos interagem com seu ambiente. Ela também compõe o estudo da evolução, mudanças genéticas e adaptações que ocorreram e ocorrem com os seres vivos ao passar do tempo a fim de propiciar a sobrevivência das espécies em diferentes condições do ambiente.

Ser um biólogo vai muito além de apenas estudar os seres vivos e suas complexidades únicas. Significa também estar apto a contribuir com o entendimento e compreensão da gama e sofisticação dos organismos vivos. Os biólogos são profissionais que podem seguir nas mais variadas áreas de conhecimento, como genética, ecologia, biotecnologia, microbiologia, entre outras. Podem, por exemplo, estudar como as reações e processos bioquímicos ocorrem dentro e fora das células, como funcionam as interações entre espécies em um dado ecossistema, sejam essas interações intraespecíficas ou interespecíficas, descrever impactos antrópicos no meio ambiente, etc. Ademais, os biólogos podem trabalhar com pesquisas científicas, com o propósito de aprimorar a saúde humana, desenvolver novas tecnologias emergentes, promover a conservação da biodiversidade e até mesmo trabalhar com a divulgação científica. Em suma, ser biólogo significa contribuir para a compreensão e proteção da vida em todas as suas formas e aspectos, como será abordado em um tópico próprio neste trabalho.

### **1.2 Biologia como ciência autônoma**

Ernst Mayr (2004) em seu livro “Biologia, ciência única”, discute a autonomia da Biologia como Ciência em relação à Física e Química. O biólogo aponta em seu texto que “...estava, contudo, igualmente desapontado com a filosofia da ciência tradicional, que era toda ela baseada em lógica, matemática e ciências físicas e que adotara a conclusão de Descartes de que um organismo nada mais é que uma máquina. Esse cartesianismo me

deixava completamente insatisfeito, assim como o saltacionismo...”. O autor usa como argumento a biologia ser uma ciência autônoma, com características únicas e de certa forma independentes, pois seus estudos englobam sistemas vivos e complexos que possuem questões e metodologias próprias, não podendo ser reduzidas a apenas uma simples extensão da química e da física, como comumente são. Para Mayr (2004), a biologia é uma ciência holística pois utiliza-se não somente de lógica em suas questões, mas também uma visão, de certa forma, poética sobre como tudo está interligado, olhando os organismos vivos como seres e sistemas altamente complexos e integrados.

Mayr (2004) argumenta também que a biologia tem um objeto de estudo central e específico, sendo ela a vida, e que tal objeto não pode ser entendido apenas utilizando dos métodos da química e da física. Destaca também a importância da história da evolução, como ela foi impactante para o que entendemos hoje como biologia e como ela moldou o comportamento e características dos seres vivos, bem como também um fator determinante para a compreensão das ciências biológicas. Por fim, este autor também discute como se diferenciam os conceitos entre a biologia e a química e física. Por um lado temos a biologia que lida com organismos e sistemas vivos no geral, enquanto que no outro temos a química e a física que lidam com entidades abstratas e sem vida, como partículas e moléculas. Tal discussão é fundamental para o entendimento da biologia como ciência à parte, tratando exclusivamente da vida e as leis que a regem.

Mediante a isso, Mayr (2004) reflete a Biologia como uma entidade crítica, o que desvincula o entendimento do leitor a considerar a mesma apenas mais uma ciência dentre várias. A autonomia das ciências biológicas se manifesta em diversos aspectos. Podemos tomar como exemplo a enorme diversidade e complexidade de formas de vida existentes, conhecidas ou não, que apresentam características tão únicas que demandam abordagens, ferramentas, técnicas, estudos e metodologias específicas para analisarmos. Essas demandas não podem ser supridas apenas pela física, química ou etc.

Este trabalho tem como objetivo, então, entender como a bioinformática (um dos campos desta grande ciência) funciona, qual sua importância no campo da pesquisa e como o biólogo (ou outro profissional que se interesse) se insere nesse meio. Para isso, foram realizados levantamentos bibliográficos, pesquisas no projeto pedagógico do curso de Ciências Biológicas, estudo de leis e decretos, acompanhado de uma narrativa autobiográfica e entrevistas de outros estudantes da biologia e de outros cursos que trouxeram vivências e discussões de suas experiências estudantis e laboratoriais.

### 1.3 Legislação do biólogo no Brasil

Como toda profissão regulamentada, a de biólogo tem sua legislação própria. Duas leis e um decreto determinam e regularizam a profissão e o exercício da biologia no Brasil, sendo elas: a Lei 6.684/79, a Lei 7.107/82 e o Decreto 88.438/83. A Lei 6.684/79 determina a exigência do registro profissional para o exercício das atividades pelos profissionais biólogos e dispõe sobre os órgãos de fiscalização do exercício de seus profissionais. O CFBio (Conselho Federal de Biologia) se define como uma Autarquia Federal com autonomia administrativa e financeira que normatiza e supervisiona a fiscalização do exercício profissional em todo o território nacional, enquanto que define que os CRBios (Conselhos Regionais de Biologia) possuem como uma de suas atribuições fiscalizar o exercício profissional na área da sua jurisdição.

Com relação a Lei 7.107/82, dispõe sobre o desmembramento entre conselhos de Biomedicina e Biologia. Antes dessa lei, as tratativas administrativas e regionais eram de responsabilidade dos mesmos órgãos. Com a vinda dessa lei deu-se autonomia e reconhecimento para a diferenciação dessas duas profissões, o que representa uma aceitação, por assim dizer, da biologia como ciência autônoma, tema este discutido anteriormente.

Por fim, o Decreto 88.438/83 trata da regulamentação do exercício em si da profissão, definindo suas atribuições, competências e atividades privativas do biólogo, conforme consta em seu Artigo 2º:

Art. 2º O exercício da profissão de Biólogo é privativo dos portadores de diploma:  
 I - devidamente registrado, de bacharel ou licenciado em curso de História Natural, ou de Ciências Biológicas, em todas as suas especialidades ou de licenciado em Ciências, com habilitação em Biologia, expedido por instituição brasileira oficialmente reconhecida;  
 II - expedido por instituições estrangeiras de ensino superior, regularizado na forma da Lei, cujos cursos forem considerados equivalentes aos mencionados no inciso I.

Em seu Artigo 3º, o decreto dispõe sobre as atribuições do profissional:

Art. 3º Sem prejuízo do exercício das mesmas atividades por outros profissionais igualmente habilitados na forma da legislação específica, o Biólogo poderá:  
 I - formular e elaborar estudo, projeto ou pesquisa científica básica e aplicada, nos vários setores da Biologia ou a ela ligados, bem como os que se relacionem à preservação, saneamento e melhoramento do meio ambiente, executando direta ou indiretamente as atividades resultantes desses trabalhos;  
 II - orientar, dirigir, assessorar e prestar consultoria a empresas, fundações, sociedades e associações de classe, entidades autárquicas, privadas ou do Poder Público, no âmbito de sua especialidade;  
 III - realizar perícias, emitir e assinar laudos técnicos e pareceres, de acordo com o currículo efetivamente realizado.

Vale ressaltar a Resolução 227/2010 do Conselho Federal de Biologia (CFBio), que define as três grandes áreas de atuação do biólogo. São elas: Meio Ambiente e Biodiversidade, Saúde e Biotecnologia e Produção. Dentre essas grandes áreas são múltiplas as atividades que o profissional da biologia pode desempenhar. Como o foco do presente trabalho é bioinformática, dada a sua grande importância nos últimos anos não só na comunidade científica mas também difusa em grandes empresas, vale ressaltar que a mesma é reconhecida como uma área de atuação. Como consta em seu Artigo 4º:

Art. 4º São áreas de atuação em Meio Ambiente e Biodiversidade:

Aqüicultura: Gestão e Produção

Arborização Urbana

Auditoria Ambiental

Bioespeleologia

Bioética

Bioinformática

Biomonitoramento

Biorremediação

Controle de Vetores e Pragas

Curadoria e Gestão de Coleções Biológicas, Científicas e Didáticas

Desenvolvimento, Produção e Comercialização de Materiais, Equipamentos e Kits Biológicos

Diagnóstico, Controle e Monitoramento Ambiental

Ecodesign

Ecoturismo

Educação Ambiental

Fiscalização/Vigilância Ambiental

Gestão Ambiental

Gestão de Bancos de Germoplasma

Gestão de Biotérios

Gestão de Jardins Botânicos

Gestão de Jardins Zoológicos

Gestão de Museus

Gestão da Qualidade

Gestão de Recursos Hídricos e Bacias Hidrográficas

Gestão de Recursos Pesqueiros

Gestão e Tratamento de Efluentes e Resíduos

Gestão, Controle e Monitoramento em Ecotoxicologia

Inventário, Manejo e Produção de Espécies da Flora Nativa e Exótica

Inventário, Manejo e Conservação da Vegetação e da Flora

Inventário, Manejo e Comercialização de Microrganismos

Inventário, Manejo e Conservação de Ecossistemas Aquáticos:

Límnicos, Estuarinos e Marinhos

Inventário, Manejo e Conservação do Patrimônio Fossilífero

Inventário, Manejo e Produção de Espécies da Fauna Silvestre Nativa e Exótica

Inventário, Manejo e Conservação da Fauna

Inventário, Manejo, Produção e Comercialização de Fungos

Licenciamento Ambiental

Mecanismos de Desenvolvimento Limpo (MDL)

Microbiologia Ambiental

Mudanças Climáticas

Paisagismo

Perícia Forense Ambiental/Biologia Forense

Planejamento, Criação e Gestão de Unidades de Conservação (UC)/Áreas Protegidas

Responsabilidade Socioambiental  
 Restauração/Recuperação de Áreas Degradadas e Contaminadas  
 Saneamento Ambiental  
 Treinamento e Ensino na Área de Meio Ambiente e Biodiversidade

Em seu Artigo 6º a resolução também dispõe sobre a bioinformática como área de atuação:

Art. 6º São áreas de atuação em Biotecnologia e Produção:  
 Biodegradação  
 Bioética  
 Bioinformática  
 Biologia Molecular  
 Bioprospecção  
 Biorremediação  
 Biossegurança  
 Cultura de Células e Tecidos  
 Desenvolvimento e Produção de Organismos Geneticamente Modificados (OGMs)  
 Desenvolvimento, Produção e Comercialização de Materiais, Equipamentos e Kits Biológicos  
 Engenharia Genética/Bioengenharia  
 Gestão da Qualidade  
 Melhoramento Genético  
 Perícia/Biologia Forense  
 Processos Biológicos de Fermentação e Transformação  
 Treinamento e Ensino em Biotecnologia e Produção.

Embora disposta em resoluções da Biologia, a bioinformática também é escopo de atuação de outros profissionais, como biotecnólogos, biomédicos e pesquisadores de diversas áreas (não necessariamente ligados a biologia) e ramos de atuação que tendem a resolver alguma questão biológica utilizando a bioinformática.

#### **1.4 Como o biólogo inicia seus estudos? O papel da universidade**

Para se entender como o biólogo pode chegar à atuação que a sua profissão possibilita, como aquelas descritas anteriormente na legislação profissional, mais precisamente a área a qual é o foco deste, é preciso entender como se inicia os estudos na biologia. Para ser um biólogo, não somente a vontade de aprender sobre os seres vivos e suas individualidades basta, mas também é necessário cursar uma graduação, para, em seguida, solicitar seu registro profissional junto ao CRBio, como consta no Artigo 1º da Lei 6.684/79, disposto a seguir:

Art. 1º - O exercício da profissão de Biólogo é privativo dos portadores de diploma:  
 I - devidamente registrado, de bacharel ou licenciado em curso de História Natural, ou de Ciências Biológicas, em todas as suas especialidades ou de licenciado em Ciências, com habilitação em Biologia, expedido por instituição brasileira

oficialmente reconhecida;

II - expedido por instituições estrangeiras de ensino superior, regularizado na forma da lei, cujos cursos forem considerados equivalentes aos mencionados no inciso I.

Logo temos em vista que é necessário uma graduação em um dos cursos citados acima (vale ressaltar que o curso de História Natural não é mais ofertado no Brasil). A forma comumente utilizada para entrar em uma instituição de ensino superior é através da nota do Enem, através do SisU. Entretanto, podem haver particularidades se tratando do modo de ingresso de instituição para instituição, principalmente as particulares. A priori, todo curso superior tem uma estrutura curricular com um conjunto de disciplinas ordenadamente divididas a serem abordadas dentro de um espaço de tempo pré-definido, geralmente ao longo de oito semestres. A estrutura curricular é uma estrutura organizacional que define os cursos e disciplinas necessários para completar um programa de estudos em uma instituição educacional. É geralmente definida por um conjunto de diretrizes ou padrões de aprendizagem, que podem ser definidos pelo governo. A instituição que ministra o curso elabora um projeto pedagógico em consonância com as Diretrizes Curriculares Nacionais para os cursos superiores.

O projeto pedagógico é um documento em que é definido todo o escopo de uma dada instituição educacional, sendo eles: objetivos, princípios, metodologias e estratégias de ensino. De acordo com o MEC (1997) é um instrumento de trabalho que deve ser construído e vivenciado pelos membros da comunidade escolar. Entende-se então que é uma forma de orientar na divisão e organização do conteúdo de toda uma área, se tornando um elemento essencial que assegura a educação e a formação integral (MEC, 2018). Neste tópico, a fim de exemplificar, será utilizado a estrutura curricular do curso de bacharelado em Ciências Biológicas da UFC que está disponível no portal da Prograd (Pró-Reitoria de Graduação). A estrutura curricular de licenciatura é bastante voltada para o ensino, vivências e técnicas utilizadas em sala de aula; por este motivo foi escolhido a grade de bacharelado, pois a mesma apresenta mais cadeiras com assuntos voltados a estudos e pesquisas científicas do que a estrutura de licenciatura.

O Projeto Pedagógico do curso de bacharelado em Ciências Biológicas (PROGRAD) define uma carga horária mínima de 3368 horas de cadeiras obrigatórias e 512 horas de cadeiras optativas. O projeto também define que o curso tem oito períodos letivos mínimos e doze períodos letivos máximos (oito semestres no mínimo e doze semestres no máximo). Sendo assim o curso tem uma duração entre 4-6 anos. Em cada período temos os

componentes curriculares obrigatórios. O caráter obrigatório dado aqui se refere a algo que deva ser concluído entre os períodos letivos mínimo e máximo; a disciplina deve ser cursada e concluída dentro o período de 4-6 anos, não necessariamente no período em que ela está disposta na grade curricular do curso. Embora extenso, o projeto do curso de bacharelado e de licenciatura não apresenta uma disciplina de Bioinformática, ou similar, de caráter obrigatório ou optativa.

Faz-se importante apresentar o projeto pedagógico pois o mesmo é pensado para abranger o máximo de campos de estudo possível durante uma graduação, por exemplo. Mesmo que o projeto pedagógico não apresente em sua estrutura curricular certas cadeiras voltadas a bioinformática, há diversas possibilidades dos graduandos terem seu primeiro contato com a área. Podemos citar como formas alternativas: o acesso às disciplinas livres e disciplinas optativas ofertadas por outros cursos de graduação, centros, polos e até mesmo instituições; participação em projetos de Iniciação Científica (IC) que possuem metodologias com aplicações da bioinformática; participação em programas de pós-graduação em Bioinformática e etc.

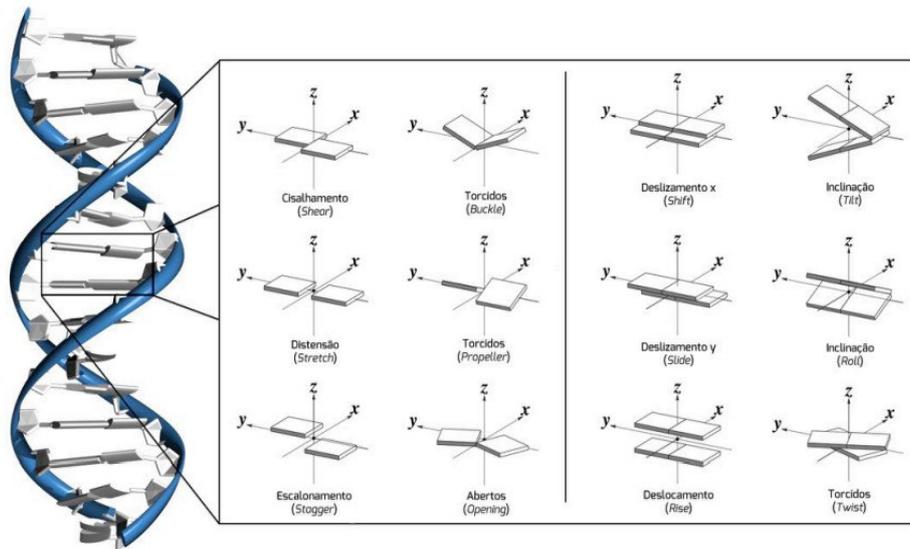
## **2 BIOINFORMÁTICA, UMA CIÊNCIA PÓS-MODERNA**

### **2.1 O que é bioinformática?**

Entrando na temática que é o foco deste trabalho, a bioinformática é uma área interdisciplinar que tem como base o uso de técnicas e algoritmos computacionais para analisar e interpretar dados de natureza biológicos. Esta área surgiu com a necessidade dos biólogos de utilizar, manusear e interpretar a grande quantidade de dados que estão sendo coletados em análises e pesquisas genômicas, somado a suas contrapartes, proteômica e genômica funcional (COHEN, 2004). A proteômica, por exemplo, envolve a aplicação de tecnologias para a identificação e quantificação do conteúdo total de proteínas presentes em uma célula, tecido ou organismo (ASLAM et al., 2016). A metabolômica, em contra-parte, fornece resultados que podem melhorar a compreensão das informações biológicas que estão associados ao metaboloma e gemômica funcional (ABDELNUR, 2011). Edwards et al. (2009) estabelece que “quando se trata de ecologia, neurociência, estudos clínicos ou biologia molecular, os grandes volumes de grandes conjuntos de dados biológicos semanticamente ricos requerem um alto nível de desenvolvimento de software”.

De acordo com Pereira et. al. (2020), a bioinformática utiliza de ferramentas e/ou métodos que envolvem a aplicação de métodos estatísticos, a utilização de algoritmos de machine learning e outras ferramentas computacionais. Os algoritmos de máquina, por exemplo, se provaram ser de grande ajuda para auxiliar cientistas e médicos a lidar com quantidades de dados massivas e a solucionar desafios complexos da biologia (LIBBRECHT et al., 2015; KUHN et al., 2018).

Figura 1 – Parâmetros geométricos empregados como descritores da geometria de ácidos nucleicos.



Fonte: Verli (2014).

Podemos tomar como base, para entendermos melhor essa complexidade, o trabalho de Morozov et al. (2005) que desenvolveu uma função de energia física que utiliza eletrostática, solvatação, ligações de hidrogênio e empacotamentos de átomos para modelar a leitura e a energia conformacional de uma molécula de DNA. É notório o quanto questões biológicas demandam de várias vertentes que passeiam entre a biologia, física e química.

Existe uma vasta gama de ferramentas computacionais que podem ser utilizadas como metodologia em pesquisas da área. Em suma, a ferramenta utilizada depende do tipo de estudo que é feito e quais resultados são relevantes de se obter. Todo ano é desenvolvido novas tecnologias computacionais a fim de refinar os softwares existentes. A Tabela 1 apresenta uma disposição de vinte programas utilizados em pesquisas de bioinformática genômica e estrutural e suas aplicabilidades:

Tabela 1 – Programas amplamente utilizados e suas respectivas funcionalidades.

<b>Software</b>	<b>Funcionalidade</b>
BLAST	Programa e algoritmo que realiza alinhamento de sequências para identificar similaridades entre sequências de DNA ou proteínas
Bowtie	Pacote de software que realiza o alinhamento eficiente de sequências curtas de DNA em genomas de referência
BWA (Burrows-Wheeler Aligner)	Pacote de software que Executa o mapeamento rápido e preciso de sequências de DNA em genomas de referência
SAMtools	Ferramenta para manipulação e análise de arquivos SAM, BAM e CRAM gerados por sequenciadores de DNA
GATK (Genome Analysis Toolkit)	Fornece uma ampla variedade de ferramentas para análise genômica, incluindo variant calling e análise de variantes
Trinity	Montador de transcritos de RNA-Seq que gera sequências de RNA de novo a partir de dados de sequenciamento de RNA
Cufflinks	Ferramenta para análise de transcritos de RNA-Seq, incluindo montagem de transcritos e quantificação de expressão gênica
BEDTools	Conjunto de ferramentas para manipulação e análise de arquivos BED, utilizados para representar regiões genômicas
IGV (Integrative Genomics Viewer)	Visualizador interativo para análise e exploração de dados genômicos, incluindo dados de sequenciamento e anotações genômicas
GROMACS	Pacote de simulação molecular usado para simulações de dinâmica molecular de proteínas, lipídios e ácidos nucleicos
PyMOL	Ferramenta para visualização e análise de estruturas moleculares em 3D
MODELLER	Software para modelagem de proteínas por homologia, permitindo a previsão de estruturas tridimensionais de proteínas
Chimera	Ferramenta para visualização e análise de estruturas moleculares, incluindo proteínas, ácidos nucleicos e complexos macromoleculares
RaxML	Realiza inferência filogenética usando o método de máxima verossimilhança para construir árvores filogenéticas
MEGA (Molecular Evolutionary Genetics Analysis)	Pacote de software para análise e construção de árvores filogenéticas, além de realizar análises evolutivas e genômicas
Rosetta	Software de modelagem e análise, que realiza predição de estruturas de proteínas, design de proteínas e docking molecular
Clustal Omega	Software de alinhamento múltiplo de sequências que compara sequências de proteínas ou DNA para identificar semelhanças e padrões
HMMER	Conjunto de ferramentas para a análise de sequências biológicas usando modelos ocultos de Markov (HMMs), como busca de homologia e anotação funcional
EMBOSS (European Molecular Biology Open Software Suite)	Coleção abrangente de ferramentas para análise de sequências de DNA, RNA e proteínas, incluindo alinhamento, tradução e predição de estrutura
Bioconductor	Ambiente de programação estatística amplamente utilizado para análise de dados genômicos e de expressão gênica, incluindo visualização, modelagem e análises estatísticas.

Fonte: elaborado pelo autor.

A bioinformática, também, combina a biologia molecular, a computação e a estatística para analisar grandes conjuntos de dados biológicos. Para isso, ela utiliza técnicas de análise de dados, modelagem matemática, algoritmos computacionais e programação. Segundo LUSCOMBE et al. (2001), a aplicação de técnicas computacionais para analisar informações associadas com biomoléculas em larga escala, abrange uma ampla gama de áreas temáticas, que vai desde a biologia estrutural e genômica até estudo de expressão genômica em si. Isso, por exemplo, ajuda na identificação de genes e mutações associados a doenças, permite a identificação de proteínas com potencial terapêutico ou até mesmo identificação de genes que podem vir a ser utilizados na produção de biocombustíveis.

Sendo assim, podemos concluir que a bioinformática é um campo de estudo que reúne diversas metodologias e ferramentas de diversas áreas, mas que sempre emprega métodos computacionais para atender hipóteses biológicas, gerar resultados coerentes e estimular o desenvolvimento de tecnologias capazes de atender as mais variadas demandas.

## **2.2 Importância nas últimas décadas**

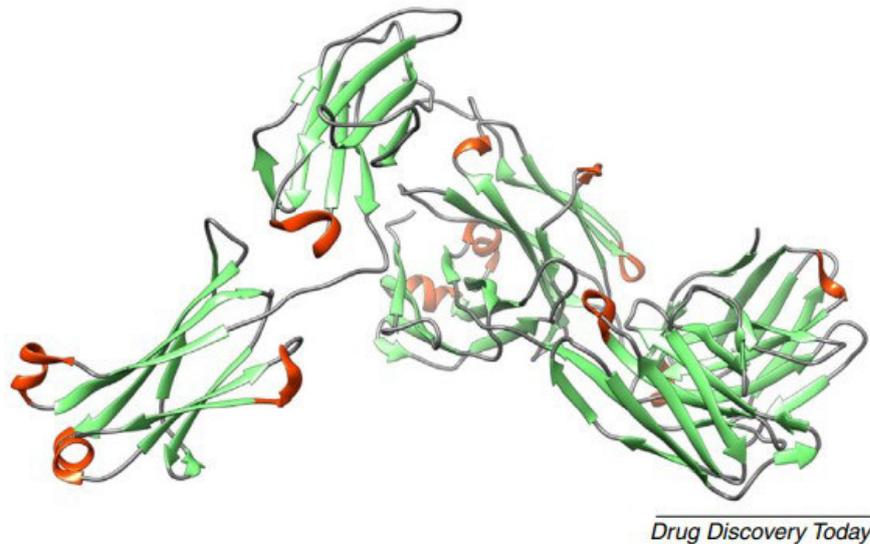
A bioinformática tem sido cada vez mais utilizada para investigar as propriedades físicas e químicas de moléculas biológicas, como proteínas e os próprios genes, por exemplo. A biofísica é uma área da ciência que estuda as propriedades físicas da matéria biológica e está bastante correlacionada com a bioinformática. De acordo com Gerstein et al. (2007), o desenvolvimento de ferramentas computacionais levou a algoritmos capazes de identificar genes com base em suas características físicas e de sequenciamento.

Segundo Zhang e Lai (2011), a bioinformática provê o descobrimento de novas drogas, especialmente com interações proteína-proteína, que advêm do campo de virtual screening que se baseia em docking molecular. A biofísica junto com a bioinformática, portanto, representa um importante aliado para o entendimento das propriedades físicas das mais diversas moléculas, bem como a conformação e estabilidade das mesmas. Na última década, a eficiência de amostragem e a precisão em dados de afinidades das moléculas aumentaram significativamente quando se trata de complexos proteína-proteína (ZHANG e LAI, 2011).

A bioinformática possui um papel fundamental na análise de dados biológicos, o que inclui o estudo e análise de genes e proteínas, o estudo de sequenciamento de DNA e RNA, bem como também a compreensão dos processos metabólicos que ocorrem com os diferentes sistemas das células. De acordo com SU (2006), a bioinformática utiliza

ferramentas de análise de dados e modelagem matemática para estudar a estrutura e a função das moléculas biológicas, bem como suas interações. Por exemplo, a bioinformática pode ser utilizada para conseguir prever a atividade enzimática de certas proteínas utilizando como base sua estrutura tridimensional e dados gerados dos estudos de sequenciamento genético (SU, 2006).

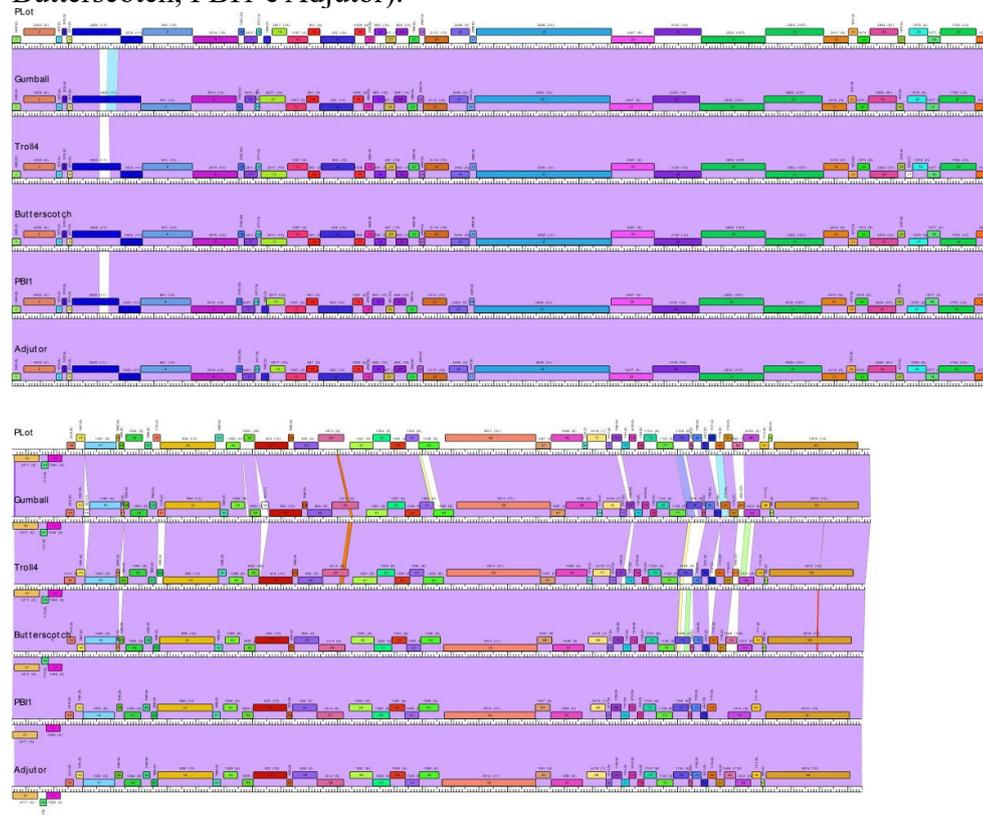
Figura 2 – Estrutura cristalina do pembrolizumab [Protein Data Bank (PDB) ID: 5DK3]. As folhas  $\beta$  são mostradas em verde.



Fonte: Zhan et al. (2016).

De acordo com Schena (1999), a bioinformática representa uma importante ferramenta na análise de grandes conjuntos de dados de expressão gênica e na identificação de padrões da mesma em diferentes tipos de células e tecidos. Isso implica, por exemplo, na utilização desses dados para estudos de mutações cancerígenas e seus padrões. Segundo Singh (2015) a pesquisa computacional em bioinformática visa aprimorar a recuperação, análise e interpretação das informações que estão embutidas nos bancos de dados biológicos contendo as sequências de DNA e proteínas. A bioquímica também está aliada com a bioinformática, haja vista que a análise de sequências de DNA e proteínas permite a identificação de novas enzimas e a caracterização de suas propriedades bioquímicas.

Figura 3 – Mapas do genoma de seis fagos Cluster D (Plot, Gumball, Troll4, Butterscotch, PBI1 e Adjudor).



Fonte: Cresawn et al. (2011).

Como evidencia Singh (2015), a bioinformática pode ser utilizada na identificação de sítios ativos em proteínas e na previsão e estudo de suas propriedades bioquímicas, como a especificidade de substrato e a estabilidade enzimática. Tais estudos são indispensáveis no entendimento das funções biológicas das enzimas, o que leva a um avanço no campo da engenharia enzimática. Portanto, a bioinformática tem se tornado uma área cada vez mais importante nas últimas décadas, impulsionada pelo rápido desenvolvimento da tecnologia de sequenciamento de DNA e RNA e estudo molecular/estrutural.

As inovações que ocorreram ao longo do tempo com as tecnologias de sequenciamento levaram a um aumento exponencial nos rendimentos das pesquisas junto a uma acelerada diminuição no custo por base sequenciada (RAVISHANKAR et al., 2018). De acordo com revisões realizadas (MARDIS (2011); METZKER (2010); PETTERSSON et al. (2009); TUCKER et al. (2009); VOELKERDING et al. (2009)), a velocidade e a precisão do sequenciamento de DNA melhorou exponencialmente nos últimos anos. Tal avanço tecnológico acaba por gerar uma quantidade gigantesca de dados, que, evidentemente, precisam ser processados e analisados para que assim possam fornecer respostas para estudos e pesquisas dos mais variados tipos. Segundo Gerstein et al. (2007), a contagem de genes no

genoma é uma tarefa que demanda em larga escala as capacidades computacionais, dada a quantidade de informações contidas nos genes, sendo desse modo que a ciência computacional vem se tornando, naturalmente, cada vez mais necessária para descrevê-los. Assim, a bioinformática tem se tornando uma ferramenta essencial para a biologia molecular e a bioquímica, auxiliando no avanço da compreensão dos processos biológicos e no desenvolvimento de novas terapias e tecnologias.

A aplicação de métodos computacionais tem se provado uma ferramenta poderosa para a análise de complexos e sistemas, capazes de descrever sistemas com centenas e até mesmo milhares de unidades atômicas. Os avanços tecnológicos no que diz respeito à capacidade computacional para o processamento de dados e o desenvolvimento de métodos como a teoria funcional da densidade tornam cada vez mais viável a aplicação de mecânica quântica na resolução de problemas biológicos (SLIWOSKI et al., 2014). Diversos estudos de diversas naturezas podem ser feitos de forma eficiente através das ferramentas de biologia computacional que possuem alta capacidade de detalhamento de interações de sistemas proteicos (MORAIS et al., 2020), por exemplo. Além do que, estima-se que novas tecnologias deverão ser desenvolvidas no futuro próximo a fim de superar os desafios computacionais que a genômica cada vez mais traz (STEPHENS et al., 2015).

### **2.3 Eventos-chave**

Alguns eventos-chave na ciência foram importantes e marcantes no desenvolvimento da bioinformática, tanto genômica como estrutural, ao longo do tempo. Em 1971 foi estabelecido o Protein Data Bank (PDB), um repositório que possui estruturas tridimensionais de macromoléculas biológicas, como proteínas e ácidos nucleicos (BERMAN et al., 2013). É uma poderosa ferramenta para a bioinformática estrutural, pois, além de possuir as estruturas tridimensionais, possui informações sobre suas propriedades, o que auxilia na modelagem molecular e estudo de interações proteína-proteína e proteína-ligante. O GenBank foi criado em 1982, com o propósito de reunir sequências de nucleotídeos e suas traduções-proteínas, sendo gerenciado e mantido pelo National Center for Biotechnology Information (NCBI). As sequências presentes neste banco são obtidas principalmente por meio de envios de laboratórios individuais e grandes projetos (BENSON et al., 2012). Sua importância se dá pelo compartilhamento mundial de dados genômicos.

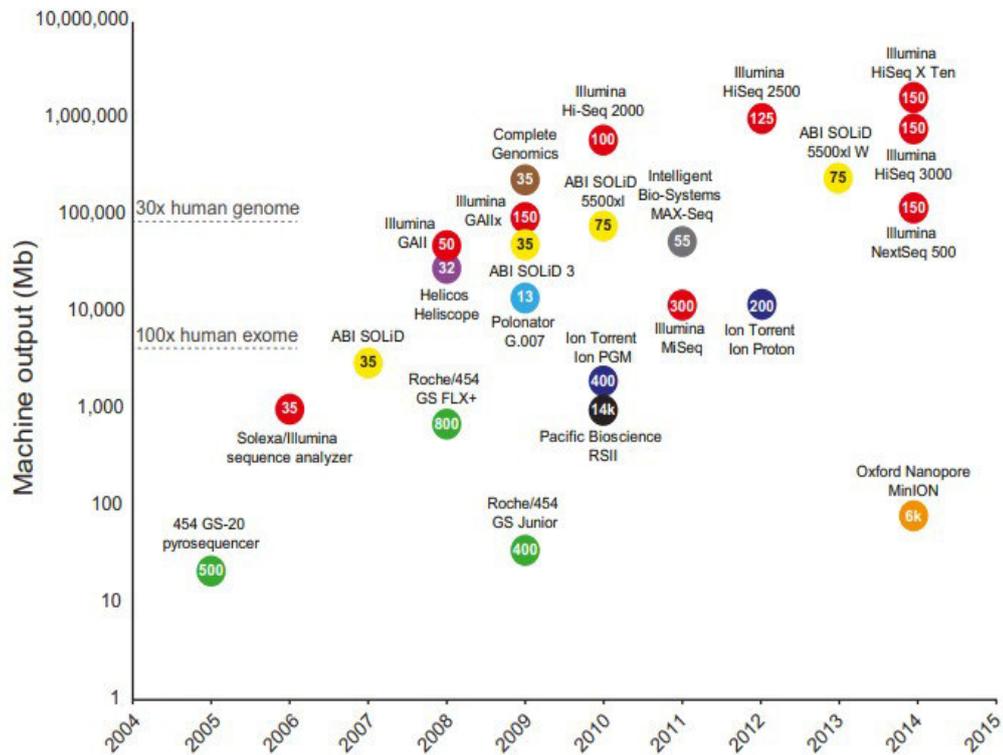
Em 1990 foi desenvolvido o BLAST (Basic Local Alignment Search Tool). Esta ferramenta nada mais é do que um algoritmo bastante utilizado para comparar sequências de

DNA, RNA e proteínas no geral. “Seu mecanismo de funcionamento detecta semelhanças de sequências fracas, mas biologicamente significativas [...]”, Altschul et al. (1990). Com isso, o BLAST permitiu avanços na identificação de sequências homólogas, na caracterização de genes e em estudos filogenéticos. Além disso, na década de 1990 houve um grande e rápido desenvolvimento de softwares e ferramentas de bioinformática estrutural. Podemos tomar como exemplos o PyMOL, Rosetta e o MODELLER. Como afirma Gabriel et al. (2009):

A partir da década de 90, com o advento dos projetos na área de genômica, a bioinformática começou a se destacar pela produção de volumes intensos de informações compartilhadas geradas a partir do sequenciamento das bases nitrogenadas que compõem o ácido desoxirribonucléico DNA. Tais dados biológicos teriam pouca utilidade sem o uso de ferramentas computacionais necessárias para a compreensão da enorme quantidade de dados resultantes. Os benefícios sociais, ambientais e econômicos que prometem os cientistas a partir das pesquisas genômicas apenas serão possíveis com a integração da análise do DNA com as diferentes funções que ele pode determinar no organismo. Sob essa perspectiva, o rápido desenvolvimento de “softwares” ou programas computacionais, que atendam às necessidades específicas dessas pesquisas, é um fator fundamental para o desenvolvimento científico e tecnológico no Brasil.

Essas ferramentas permitiram a visualização, análise e modelagem de estruturas biomoleculares, bem como o desenvolvimento de estudos e avanços em pesquisas que utilizavam esses dados. Temos, então, uma vasta gama de softwares disponíveis, permitindo que cientistas analisem em seus laboratórios seus próprios dados (MOORTHIE et al., 2011). Depois, em 2003, considerado um dos maiores projetos científicos já realizados, o Projeto Genoma Humano (PGH) foi concluído. Esse projeto visava o sequenciamento completo do genoma humano, o que acabou por gerar uma quantidade massiva de dados, que hoje podem ser usados para prever o desenvolvimento de doenças genéticas e para o tratamento do câncer (BUENO, 2009). Após a conclusão do PGH, o Instituto de Pesquisa do Genoma Humano (National Human Genome Research Institute – NHGRI) criou uma iniciativa de 70 milhões de dólares para desenvolvimento de tecnologias de sequenciamento de DNA, o que levou o desenvolvimento de *High-Throughput Sequencing* (HTS), tecnologias de alto rendimento de sequenciamento, ao longo dos próximos dez anos (REUTER, SPACEK e SNYDER, 2015).

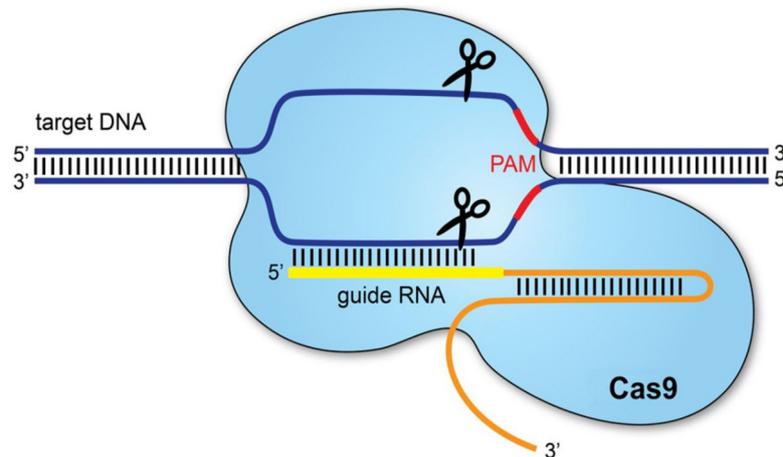
Figura 4 – Linha do tempo e comparação de Instrumentos HTS Comerciais



Fonte: Reuter, Spacek e Snyder (2015).

Dentro deste período, no ano de 2008, surgiu uma nova tecnologia descrita como a segunda geração de sequenciamento: a NGS. Como cita Behjati e Tarpey (2013), a NGS foi uma tecnologia que revolucionou a pesquisa genômica pois levava apenas um único dia para realizar o mapeamento do genoma humano, contrastando com Sanger, a tecnologia anterior que levou mais de uma década para realizar o mesmo sequenciamento. Em 2012 foi descoberto por Emmanuelle Charpentier e Jennifer Doudna o CRISPR-Cas9, uma tecnologia de edição genética que permitia corrigir erros no genoma e ativar ou desativar genes em células e organismo de forma rápida, barata e com relativa facilidade (REDMAN et al., 2016).

Figura 5 – O sistema CRISPR/Cas9



Fonte: REDMAN et al. (2016).

Atualmente, com o advento das IAs (Inteligência Artificial) e *machine learning* há um crescente aumento no desenvolvimento de técnicas voltadas a área de bioinformática. A aplicação dessas técnicas vem impulsionando cada vez mais os avanços no processamento e análise de grandes volumes de dados biológicos, em várias áreas da biologia computacional (COELHO et al., 2016; NEGRETTO, 2016; DIAS et al., 2016; CERRI et al., 2019). Limitou-se aqui apenas alguns eventos que são amplamente conhecidos pela comunidade científica e pelo público em geral. Mas não podemos descartar as demais pesquisas realizadas em seus âmbitos, pois as mesmas possuem a sua própria importância na história da ciência.

## 2.4 Gráficos de relevância

Para mensurar a relevância de um dado tema, é importante observar o fluxo de artigos e dados publicados e gerados, respectivamente, sobre o mesmo. Deste modo, a fim de propiciar uma melhor visão da relevância da Bioinformática nos últimos tempos, foram escolhidos dois mecanismos de busca que apresentam resultados de artigos relacionados a determinado tema. Foram escolhidos o PubMed Trends e o Exaly, por apresentarem gráficos de simples compreensão, relacionando quantidade de artigos publicados com seus respectivos anos de publicação.

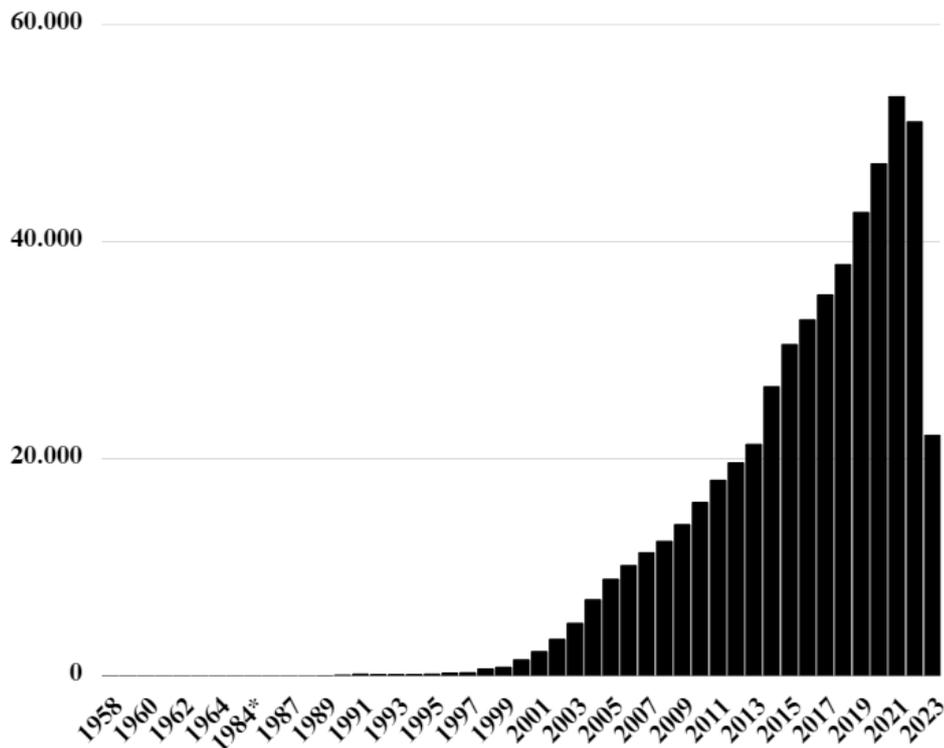
Segundo o PubMed, o mesmo “compreende mais de 35 milhões de citações de literatura biomédica da MEDLINE, revistas de ciências da vida e livros online”. De acordo com o Exaly, o mesmo representa um recurso de banco de dados de literatura acadêmica e dados cientométricos, com serviços de busca do Google Scholar, Web of Science, Publons,

Scopus, Crossref e ResearchGate combinados. Portanto, representam duas ferramentas alinhadas com o propósito deste tópico.

Para reunir estes dados, a metodologia utilizada para realizar as pesquisas foi aplicar a palavra ‘bioinformatics’ (bioinformática em inglês) como palavra-chave e não definir um intervalo de tempo; desse modo foi reunido todos os artigos presentes, desde a publicação mais antiga registrada e o último artigo registrado no banco de dados das ferramentas utilizadas. Logo abaixo estão dispostos os gráficos da relação quantidade x ano de artigos do escopo de Bioinformática do PubMed e Exaly, respectivamente.

Fornecido pelo site PubMed, é apresentado o número de artigos publicados em formato de histograma (Gráfico 1) e uma tabela dispondo o ano e a quantidade de artigos publicados (Tabela 2):

Gráfico 1 – Histograma evidenciando o aumento de artigos publicados relacionados à Bioinformática ao longo dos anos (1958 – dias atuais).



Fonte: PubMed.

Tabela 2 – Relação: ano x nº artigos

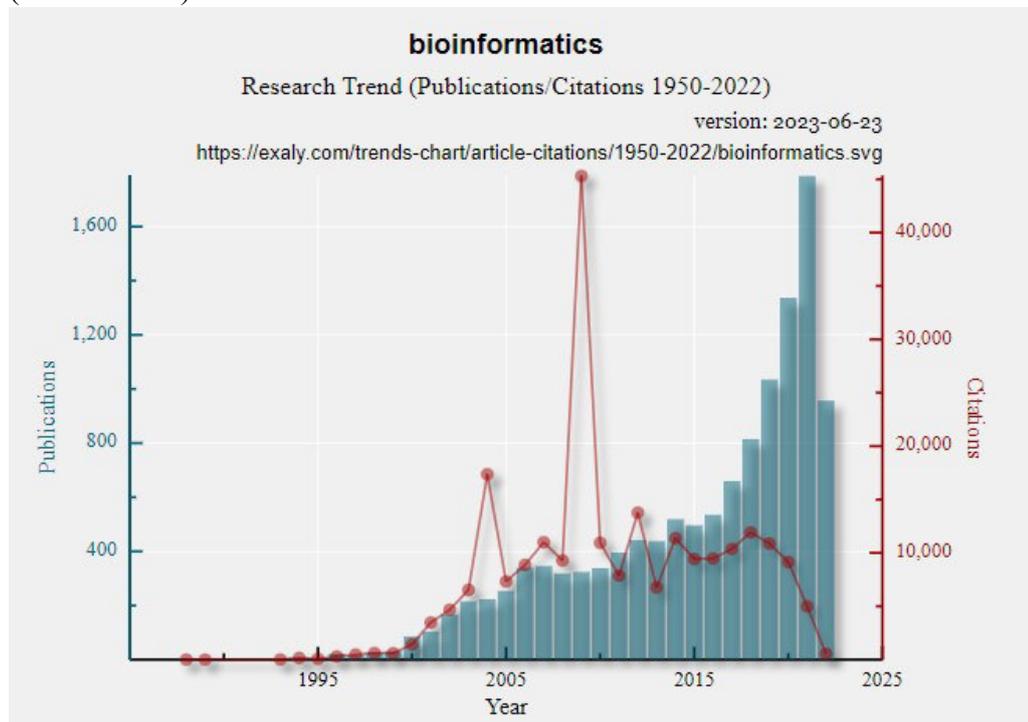
<b>Ano</b>	<b>Artigos</b>	<b>Ano</b>	<b>Artigos</b>
1958	2	2001	2295
1959	1	2002	3409
1960	1	2003	4882
1961	1	2004	7056
1962	1	2005	8945
1963	2	2006	10202
1964	3	2007	11382
1965	1	2008	12447
1984*	1	2009	13974
1986	3	2010	15996
1987	11	2011	18056
1988	12	2012	19648
1989	60	2013	21357
1990	126	2014	26669
1991	196	2015	30542
1992	164	2016	32797
1993	183	2017	35106
1994	166	2018	37912
1995	197	2019	42695
1996	306	2020	47194
1997	350	2021	53391
1998	688	2022	51038
1999	849	2023	22199
2000	1542		

Fonte: PubMed.

O Gráfico 1 e Tabela 2 apresentam um interessante e crescente aumento do número de artigos publicados relacionados à Bioinformática ao longo dos anos. Destaca-se o ano de 1984, por ter apenas um único artigo (BACH et al., 1984) sobre a área publicado e ser o primeiro que referencia um sistema de design para experimentos com clonagem de DNA, o que evidencia o primeiro artigo dentro do PubMed que utiliza-se de bioinformática estrutural, ao contrário de artigos de outrora (1958 - 1983) que utilizavam-se apenas de bioinformática genômica.

Abaixo, retirado do site Exaly, é apresentado o número de artigos publicados em formato de histograma (Gráfico 2) entre os anos de 1950 e 2022:

Gráfico 2 – Histograma evidenciando o aumento de artigos publicados e citações em demais artigos relacionados à Bioinformática ao longo dos anos (1950 – 2022).



Fonte: Exaly.

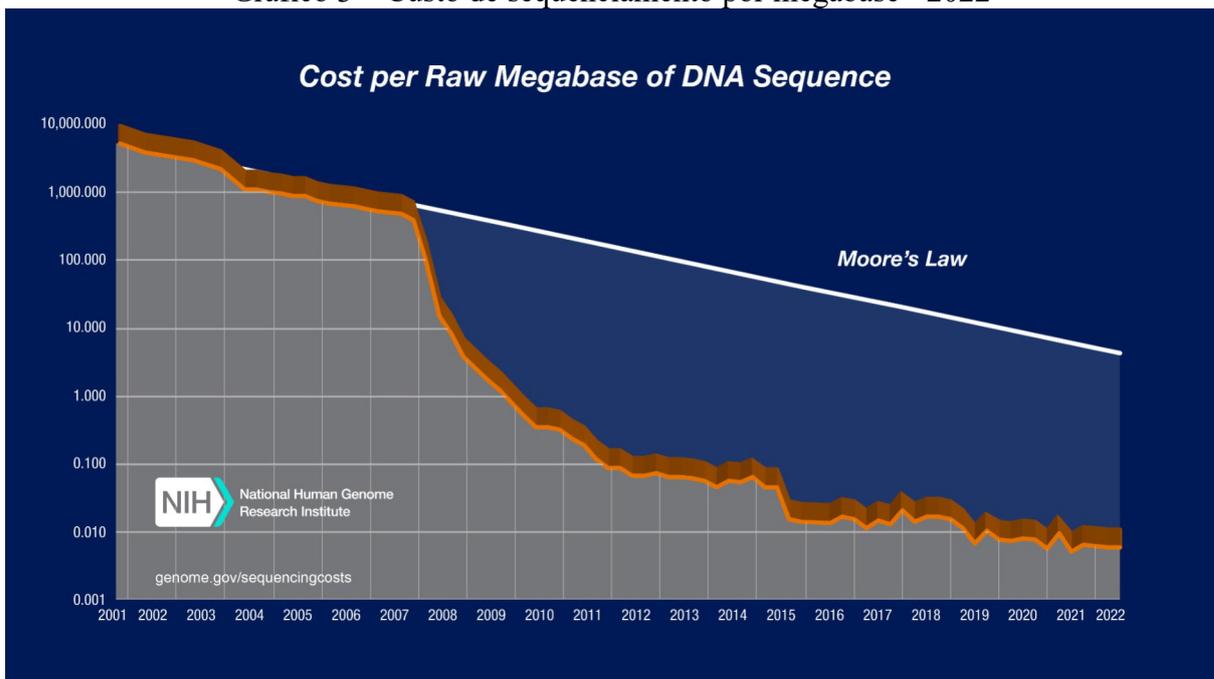
No Gráfico 2 podemos notar três picos de citações nos anos de 2004, 2009 e 2012. Esses picos estão relacionados com alguns eventos-chave descritos no tópico 2.3. No ano de 2004 podemos relacionar o primeiro pico de citação como o ano subsequente ao ano em que o PGH (Projeto Genoma Humano) foi concluído, havendo, então, um crescente aumento de artigos publicados correlacionados ao campo de estudo. Já em 2009 temos o segundo e maior pico de citações, muito vinculado ao fato deste mesmo ano ter sido palco de várias descobertas científicas e realização de estudos na área. Podemos citar, como exemplos, aumento de verbas estatais voltadas para o estudo de células-tronco nos EUA, descobertas envolvendo o mecanismo que torna a enzima telomerase como protetora dos cromossomos, o que levou ao Prêmio Nobel de Medicina do mesmo ano, detalhamento dos ribossomos, que também levou ao Prêmio Nobel de Química do mesmo ano e que ajudou a entender como as bactérias criam resistência e como desenvolver potenciais antibióticos, e a descoberta de um canal que afeta o vírus do HIV através de anticorpos, dentre outras descobertas (FOLHA ONLINE, 2009). Por fim, em 2012 foi concedida a patente de uso do CRISPR-Cas9, capaz de realizar edições genéticas, o que a tornou popular no meio científico naquele ano (BRITO, 2019).

Diante desses dados, podemos concluir que a bioinformática teve um crescente

aumento em sua relevância e demanda ao longo dos anos. Podemos, também, associar esse crescimento exponencial com a Lei de Moore, que afirma que a quantidade de transistores em um circuito integrado dobra a cada dois anos e seu preço diminui, o que leva a um enorme crescimento na capacidade de processamento e acessibilidade do público em geral (MOORE et al., 2006). Em suma: computadores mais potentes e poderosos, que conseguem realizar simulações biológicas mais rápidas, e mais baratos a cada ano, o que oportuniza qualquer laboratório a adquirir um para pesquisas.

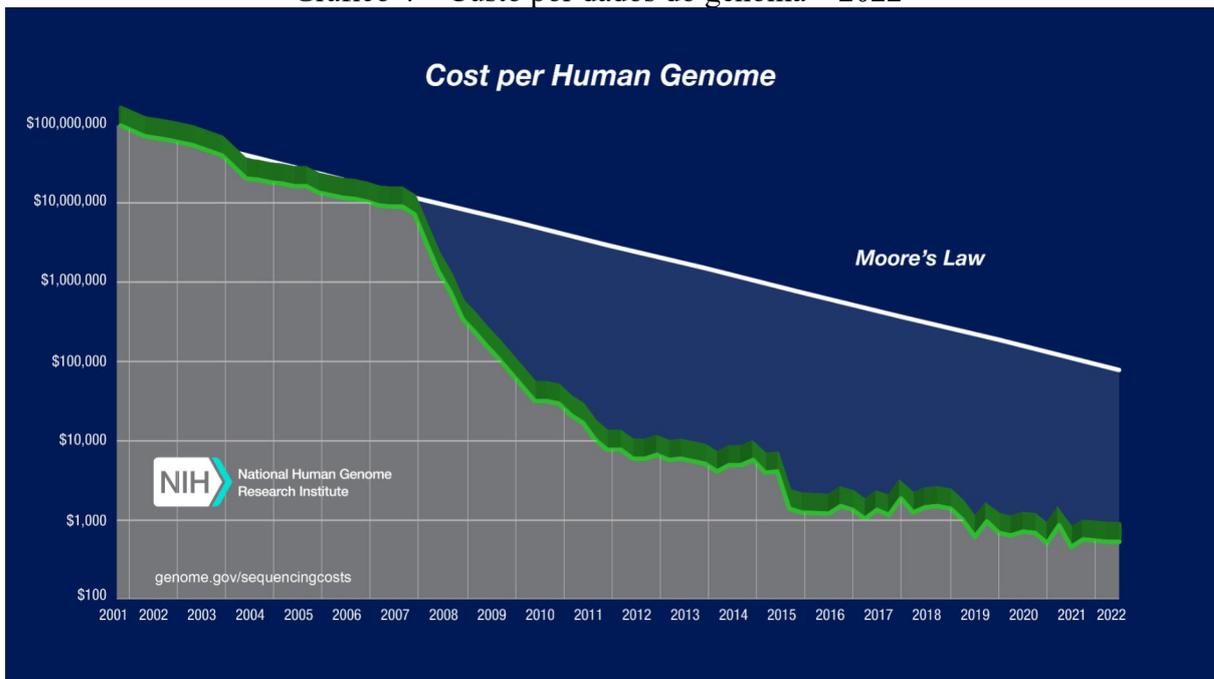
Temos, também, o trabalho de Wetterstrand (2022) que reuniu dados de custos relacionados ao sequenciamento de DNA, em pesquisas realizadas nos centros de sequenciamento financiados pelo instituto ao qual a autora faz parte. Os gráficos 3 e 4 abaixo reúnem dados de custo de 2001 até a data em que este trabalho está sendo escrito, e apresenta uma relação custo x ano:

Gráfico 3 – Custo de sequenciamento por megabase - 2022



Fonte: Wetterstrand (2022).

Gráfico 4 – Custo por dados do genoma – 2022



Fonte: Wetterstrand (2022).

Podemos notar uma queda vertiginosa no custo de sequenciamento de ambas as tecnologias, o que fortalece e corrobora com a ideia proposta por Moore e está relacionado com o surgimento da NGS (anteriormente descrito no tópico 2.3), como pode ser notado no ano de 2008 no Gráfico 3 e 4. Muito se deu, também, por mudanças nos instrumentos utilizados nas pesquisas, o que representou uma rápida evolução das tecnologias envolvidas no sequenciamento de DNA nos últimos anos (WETTERSTRAND, 2022). Edwards et al. (2009) afirma que “esses avanços ainda estão ocorrendo hoje, e agora somos capazes de desenvolver programas mais complexos mais rapidamente do que em qualquer época do passado”.

### 3 EXPERIÊNCIAS E VIVÊNCIAS

Como disposto ao longo deste trabalho, a bioinformática representa uma poderosa ferramenta na ciência. Entretanto, como toda ferramenta, a mesma requer alguém que possa manuseá-la. Conseqüentemente, deixar de fora desta pesquisa o ser pensante por detrás desta ferramenta é, de certa forma, incoerente. Neste tópico, com a intenção de propiciar diferentes visões e vivências dos pesquisadores da área, foi incluída uma narrativa autobiográfica juntamente de entrevistas com alunos de diferentes cursos e caminhos na graduação.

Conforme Lima et al. (2015), a narrativa autobiográfica visa a reconstituição da história e vivência pessoal do próprio autor, possibilitando uma escrita de si e sobre si em seu processo de formação por meio da adoção de dados empíricos coletados por ele mesmo. Batista et al. (2017) defende que fazer uso da entrevista em uma pesquisa para o apanho de informações reflete na busca da compreensão da subjetividade do indivíduo através de seus depoimentos; isso ocorre por levar em consideração o modo como o mesmo observa, vivência e analisa sua história, seu meio social e seu momento. Desta forma, entrevistas, narrativas, depoimentos, etc refletem um interessante método de pesquisa no entendimento em como o homem se comporta e enxerga suas atitudes e opiniões sobre determinado tema.

### **3.1 Como o biólogo inicia na bioinformática? Uma narrativa autobiográfica**

Como citado este trabalho contempla uma narrativa autobiográfica que tenta dar um vislumbre, ou, pelo menos, se esforça em relatar e mostrar um dos diversos caminhos que o biólogo ou outro graduando pode se deparar para entrar na bioinformática. Nesta narrativa tentarei me abster de fatos e detalhes que eu acredito que não sejam tão importantes assim para relatar e para a compreensão do leitor. Entretanto, acredito que exista relatos, acontecimentos e vivências que sejam importantes de serem mencionadas neste trabalho, com a finalidade de dar uma humanização e uma aproximação com o leitor.

Filho de um pai vendedor e de uma professora de ensino infantil, sempre observei em meus pais a vontade de procurar saber sobre no que trabalhava e ensinava, respectivamente. Para meu pai, entender sobre com o que vendia era fundamental para entender as necessidades do cliente e melhor atendê-las. Já para minha mãe, sempre buscou aprender novas técnicas de ensino e aprendizado para melhor orientar seus alunos. Acredito que essas duas características influenciaram muito no tipo de profissional que eu desejei me tornar. Acredito que não só para cientistas, mas para todos os profissionais no geral, entender sobre com o que se trabalha e buscar rigorosamente aprender mais sobre são características fundamentais para o desenvolvimento profissional e para se tornar um profissional exemplar.

Desde criança sempre tive a vontade e a curiosidade de trabalhar com ciência. Lembro-me de sempre ver na televisão no canal TV Cultura programas educativos que ensinavam a fazer experimentos em casa, testar desmontar e montar algo, etc. Porém, como sempre estudei de tudo e procurava saber sobre basicamente várias coisas ao mesmo tempo, acredito que isso me trouxe certas adversidades em questão de escolhas do curso de graduação. Tenho para mim que muito disso se deve ao ensino médio que tive. Estudei em

uma escola profissionalizante, onde o ensino médio é feito junto a algum curso técnico. No meu caso fiz Redes de Computadores, curso que escolhi com base em desde pequeno observar os componentes internos e a placa-mãe e ficar maravilhado com tudo aquilo. Digo que por um lado era bom estudar vários assuntos diferentes ao longo de três anos, mas havia o lado negativo de não haver nada muito aprofundado. Os professores constituem uma parte fundamental no processo de aprendizado e formação do aluno, obviamente; porém, nunca senti realmente que tinha algum que se preocupasse com as escolhas que seus alunos tomariam dali pra frente, ou que tentasse orientar qual rumo tomar. Tenho por mim ser muito inusitado como uma pessoa de dezesseis anos deveria escolher algo com que deveria “trabalhar o resto da vida”. Porém, era a situação em que me encontrava: ao terminar o ensino médio aos dezesseis, não tive a escolha de tirar um tempo para decidir o que gostaria de fazer ou trabalhar. Talvez por pressão dos pais, ou por uma pressão autoimposta, decidi abruptamente fazer o tão temido Enem e com o resultado escolher posteriormente o que fazer. Já havia me preparado nos últimos três anos de, até então, minha curta vida e consegui obter uma média satisfatória (quem decide isso?) para a tão sonhada vaga numa instituição de ensino superior. Acredito que fornecer essa breve introdução minha antes do início da vida acadêmica poderá ajudar no entendimento de certas atitudes e decisões minhas no futuro, principalmente no que se refere a minha experiência com bioinformática.

Até ver a nota, não tinha em mente em que curso entrar. Não havia realmente um curso que me chamasse a atenção por suas peculiaridades ou campos de estudo. Ironicamente, por impulso, resolvi escolher um bacharelado em Química na UFC. Em minhas vivências no ensino médio, as aulas em laboratórios me fascinavam muito; o ambiente de laboratório era um ambiente em que me sentia bastante confortável. Daí a escolha de Química como possível graduação para cursar. Fiquei um tempo em Química e senti que talvez aquilo não fosse o ideal. Talvez pelo conteúdo, ou pelo ambiente em si (primeiro contato com a faculdade), ou por questões de mercado de trabalho; no fundo senti que não gostaria de passar quatro anos naquela graduação e resolvi sair. Após um ano, realizando o mesmo processo de antes, entrei em Biotecnologia, também na UFC, mas novamente não me adaptei bem ao curso. Acredito que tenha sido pelas mesmas questões da minha situação com Química.

Finalmente, decidi que deveria escolher bem o meu próximo curso e tentar de todas as formas me adaptar a ele. Pesquisando um pouco mais a fundo e fazendo testes online de vocação de carreira, descobri que poderia ser uma boa ideia cursar Ciências Biológicas. Julgo ter sido uma boa escolha pela estrutura do curso apresentar os mais diversos assuntos, seria uma possibilidade de me encontrar lá. Prestei novamente o Enem e fui aprovado no

curso de Bacharelado em Ciências Biológicas, novamente na UFC. Quando entrei nessa graduação não fazia ideia de que trabalharia um dia com bioinformática. Honestamente, tentando me recordar, não lembro de saber realmente o que era bioinformática dentro do curso. O contato que tive com a área foi bastante raso, apesar de ver uma pincelada em algumas cadeiras, com: Bioquímica Geral, Genética, Biofísica, Biologia Molecular e Biotecnologia e Evolução.

No ano de 2020 o mundo foi acometido por uma pandemia, causada por um vírus: Coronavírus SARS-CoV-2 ou mais conhecido como COVID-19. Com a pandemia, vieram os tempos de quarentena e aulas na modalidade EaD. Dada a situação, não havia aulas práticas e tudo era resumido apenas a teoria. Apesar disso, alguns professores se empenhavam em tentar trazer, de alguma forma, aulas práticas durante a quarentena. Dentre elas destaco uma simulação de escrita de artigo que o professor de Genética trouxe. A proposta era criar uma situação hipotética, onde precisaríamos escolher um gênero de animal e comparar o genoma de duas espécies diferentes. Escolhi fazer a comparação entre o genoma de duas espécies de golfinhos corriqueiros da região litorânea do nordeste. As ferramentas que utilizei foram: o banco de dados GenBank, que continha o genoma das duas espécies de golfinho, e as ferramentas BLAST e MEGA, que são instrumentos que realizam comparações entre genomas. Esse foi o meu primeiro contato com a Bioinformática.

Após isso, só tornei a trabalhar novamente com bioinformática apenas quando consegui um estágio em um laboratório do Departamento de Física. Baseado em conversas e compartilhamento de experiências com outros alunos da graduação mesma graduação que a minha e de outras graduações, como por exemplo Agronomia e Física, se vê superficialmente sobre a área nas cadeiras de seus respectivos cursos (salvo Biotecnologia que possui uma única cadeira em sua grade curricular voltada a área e Física que possui uma cadeira de Fundamentos de Programação).

Como relatado anteriormente, quando chegou a época de realizar a cadeira de Estágio Supervisionado I procurei um laboratório que trabalhava com Bioinformática, mais precisamente Bioinformática Estrutural. Para mim, a Bioinformática Genômica é igualmente interessante, mas tinha a intenção de trabalhar com docking molecular porque desde pequeno acho extremamente interessante as figuras de moléculas e estruturas biológicas nos livros de Biologia e Química. Consegui encontrar através de um colega dos tempos em que cursava Biotecnologia. Conversei com o mesmo, tirei algumas dúvidas e solicitei o email do professor. Marquei uma reunião com o mesmo, conversamos, acertamos toda a documentação e assim fui aceito e iniciei no laboratório. Quando comecei no laboratório senti certas

dificuldades, muito relacionado à quantidade de informações, teorias e ferramentas que eram utilizadas para realizar as pesquisas. Haviam três-quatro softwares que realizam funções semelhantes, mas que geravam resultados diferentes para cada fase das pesquisas. Em suma, posso categorizar todo o estudo da Bioinformática Estrutural em quatro principais fases: Preparação dos dados necessários para a pesquisa, Docking Molecular, Dinâmica Molecular e Cálculo Quântico. Para o entendimento de cada fase, foi necessário um pouco mais de estudo e pesquisa. Isso porque, fundamentalmente, a Bioinformática Estrutural se utiliza de muitos modelos físico-químicos, além de teorias de mecânica, em especial mecânica clássica e quântica.

Como citei, existem diversos programas para cada fase da pesquisa. A escolha deles vai muito do propósito da pesquisa e da afinidade que o pesquisador tem com o software. Na fase de preparação dos arquivos, é muito importante se atentar pra estrutura da molécula alvo, por exemplo. Diferentes conformações estruturais podem gerar diferentes resultados que não são importantes para o ambiente natural da molécula. Exemplificando, se atentar para entroncamentos tridimensionais da molécula, cargas dos átomos e pH são indispensáveis na preparação do ambiente de simulação do estudo. Nessa fase, utilizei alguns programas como Discovery Studio e o ProteinPrepare para visualizar e preparar o receptor e utilizei o MarvinSketch para realizar o estudo de cargas dos átomos do ligante. No docking, etapa em que é analisado e gerada uma possível conformação de ligação entre um receptor e um ligante, é utilizado alguns softwares para realizar a preparação do ambiente em que ocorrerá a pequena simulação de encaixe. Aqui, utilizei AutoDockTools para preparar os arquivos que serviam de input no software de docking, o AutoDock Vina.

Com os resultados gerados do docking, obtêm-se em média vinte possíveis posições de encaixe entre as duas moléculas. São escolhidas as posições com maior energia negativa de interação, que é calculada através do algoritmo de docking utilizado pelo programa. Quanto mais negativa é a energia, maior a afinidade entre as duas moléculas. Escolhida as prováveis posições, os arquivos gerados servem de input para realizar a dinâmica molecular. Essa dinâmica é uma simulação mais precisa do que o docking, pois simula ambientes biológicos e sistemas biomoleculares, e utiliza-se de Mecânica Clássica. Nela conseguimos definir temperatura e pressão, bem como realizar a mesma em diferentes condições termodinâmicas. Para realizar as dinâmicas moleculares utilizava o GROMACS, um software de código aberto especialmente projetado para essas simulações.

Por fim, com os resultados provenientes da dinâmica molecular era realizado o estudo de cálculo quântico. O cálculo quântico se utiliza principalmente de Mecânica

Quântica, mais precisamente da Teoria Funcional da Densidade (DFT). É realizado uma minimização da energia total do sistema, evidenciando a energia de interação entre determinados átomos do ligante com resíduos do receptor. Para realizar estes cálculos, utilizava o programa Materials 8.0 somado ao DMol3, um pacote que permite calcular a estrutura eletrônica de sistemas moleculares, mas que possui várias outras aplicações.

Portanto esse era basicamente o trabalho que desempenhava no laboratório. Entretanto, há sempre diversidades na vida de qualquer ser humano. Infelizmente, durante este período, meu pai foi acometido de duas trombozes no curto intervalo de tempo, e, dada a situação, tinha que trabalhar para auxiliar financeiramente em casa. Consegui trabalhar como Técnico de TI, já que possuía uma certificação de Técnico em Redes de Computadores. Com isso, vieram as dificuldades em conciliar o novo trabalho com o estágio no laboratório. Apesar disso, me esforçava em atender as demandas pedidas do laboratório sempre quando era possível. Entretanto, senti atritos acontecendo entre minha pessoa e o professor. Muito se dá pelo fato de trabalhar com uma quantidade imensa de informações; a pesquisa que estava desenvolvendo possuía em torno de cinco ligantes diferentes para o receptor, o que gerava cinco resultados diferentes para cada fase do estudo. Dada a enorme quantidade de dados que dispunha, o professor do laboratório, aparentemente, não sabia como dispor dessas informações nos artigos, sempre me pedindo para adicionar mais alguma coisa nas figuras e resultados e depois voltando atrás.

Tudo isso gerou uma tensão minha com o professor, pois, sempre me era pedido algo a mais do que tínhamos acordado em fazer em nossas reuniões, ou vários dados eram descartados para dar lugar a novas formulações da pesquisa. Somada ao estresse que estava passando em minha vida pessoal pela descomunal carga de atividades que estava desempenhando ao longo dos dias, escolhi sair do laboratório para me concentrar e investir meus esforços em uma área crítica da minha vida: a financeira. Optei em não colocar figuras de minha autoria, com a intenção de ilustrar o trabalho que desempenhava no laboratório, neste trabalho, que montei em meu tempo de pesquisas para evitar prováveis problemas no futuro (mesmo que estas figuras estejam presentes em meus relatórios de estágio I e II antes de que qualquer artigo pudesse ser publicado). Também não me sinto à vontade em compartilhar os trabalhos que foram publicados após um ano de pesquisa em que estive presente e fui responsável por 95% dos resultados. Também não fui creditado de forma alguma nesses artigos.

Embora a minha experiência na bioinformática tenha sido um tanto quanto negativa, ainda a vejo como uma área em que gostaria de atuar. Aprendi bastante coisa

durante o tempo em que tive contato com a bioinformática. Conceitos que outrora pareciam impossíveis de se compreender, através da prática, tornaram-se concepções bastante interessantes de se trabalhar. Ainda que minha experiência não tenha sido muito positiva, reconheço que aprendi muito no laboratório e que é uma área fantástica para se atuar, principalmente nos dias de hoje com o rumo que a tecnologia vem tomando. Sendo assim, considero que meu saldo com a Bioinformática é positivo e que meu caso foi apenas um infortúnio da vida.

Inoportunamente, o Brasil não apresenta um mercado tão favorável para esta área, até a data deste presente trabalho, em comparação com o mercado no exterior. Pessoalmente acredito que seja pela falta de altos investimentos em ciência e em tecnologias emergentes, como outros países comumente o fazem. Embora o objetivo deste trabalho seja falar apenas da graduação e possibilidades de inserção no mercado de trabalho, vale ressaltar que existem programas de pós-graduação na área aqui no Brasil que podem fazer a diferença na busca por uma inserção no mercado dentro do país e também no exterior. A título de curiosidade, diversas instituições públicas ao redor do país oferecem programas de Pós-Graduação em Bioinformática em diversos segmentos. Algumas dessas instituições são: USP (Universidade de São Paulo), UFMG (Universidade Federal de Minas Gerais), UFPR (Universidade Federal do Paraná), IOC (Instituto Oswaldo Cruz), dentre outras. Existem, também, diversas instituições particulares que ofertam pós-graduação também em bioinformática em diversas modalidades de ensino.

### **3.2 Entrevistas com alunos**

Aqui, foram adotados alguns critérios para as entrevistas. Os entrevistados, oriundos da bioinformática, foram selecionados através de indicações de professores e alunos que conheciam a área. Um pequeno número de perguntas curtas (cinco) deram liberdade aos entrevistados de responder abertamente as mesmas. As perguntas foram estruturadas de modo a criar uma sutil linha temporal da vivência dos entrevistados, o que facilita na elaboração das respostas e integração entre as mesmas. Por fim, com a finalidade de garantir uma resposta sincera e livre de pesares, a identidade dos entrevistados foram mantidas anônimas.

**1º Ao entrar no curso, você já tinha em mente trabalhar com bioinformática?**

**Aluno 1:** *“Sim, no ensino médio eu fiz um curso técnico de informática e eu queria me tornar um cientista e a área que eu mais me interessava era*

*Biologia, então eu descobri sobre a Bioinformática, que misturava programação com biologia, no terceiro ano e decidi seguir nessa área. Olhando as grades dos cursos da UFC, eu vi que tinha a cadeira de Bioinformática como obrigatória para o curso de Biotecnologia e então entrei no curso.”*

**Aluno 2:** *“Não, eu me formei em Biotecnologia na UFC e quando entrei no curso eu não tinha ideia do que era Bioinformática. Eu fui entender o que era Bioinformática com a disciplina de Bioinformática no meu 5 semestre, mas ainda assim eu não achava que era o que eu queria. No meu último ano do ‘curso foi que eu’ fui fazer um estágio na Fiocruz com Bioinformática e comecei a me encontrar. Eu nunca me identifiquei com o trabalho de laboratório molhado, aí quando entrei na rotina da bioinfo percebi que era com essa área que eu me identificava.”*

**Aluno 3:** *“Não, não tinha ideia de que ia trabalhar com bioinformática; na verdade eu não sabia o que era bioinformática, eu descobri no primeiro semestre mas ainda assim eu não tinha essa ideia de trabalhar com isso né. Eu acabei trabalhando com isso só porque surgiu a oportunidade de bolsa num laboratório e aí acabei indo pela bolsa e gostei do tema.”*

## **2º Como você iniciou na bioinformática?**

**Aluno 1:** *“No primeiro semestre do curso de Biotecnologia eu tentei entrar em um laboratório de Biologia Estrutural na Física em que eles fazem estudos in silico de Modelagem, Docking e Dinâmica Molecular com biomoléculas, mas não fui aceito porque eu não tinha computador pessoal (usava o da minha irmã) e ainda tava muito no começo do curso. Então, no terceiro semestre eu já estava estudando genética e entendendo melhor, eu tentei entrar no laboratório de genômica funcional e bioinformática e consegui.”*

**Aluno 2:** *“Uma amiga da minha turma arranjou um estágio com Bioinformática na Fiocruz, assim como outros amigos, e pareceu uma boa oportunidade, aí eu comecei a me perguntar se essa não era uma área que me interessava. Eu sempre gostei de matemática e queria ir pra uma área que envolvesse ciências exatas. Então acabei aceitando o*

*desafio de mergulhar nessa área no meu último semestre. Eventualmente eu tive que estender a graduação em mais um semestre pra conseguir terminar a monografia. Do ponto de vista de conteúdo, eu comecei estudando linguagens de programação e algoritmos.”*

**Aluno 3:** *“Complementando a primeira com a segunda pergunta, eu comecei a trabalhar com bioinformática antes de ter uma disciplina introdutória sobre o tema né, então quando eu fiz a disciplina eu já tinha muita base de bioinformática. Então na verdade eu descubro a bioinformática no laboratório, então eu tenho um embasamento maior na prática do que no nível teórico, porque acabo aprendendo o teórico depois. Daí iniciei com a bolsa né, trabalhando inicialmente com bioinformática estrutural, que é o que eu continuo trabalhando até hoje. E aí inicialmente eu trabalhei com agrotóxicos e albumina, estudando essa interação, era o tema da bolsa e depois eu acabei abordando outros temas como canabinoides, anticorpos, inibidores de JAK1. Mas o início foi com albumina né, estudando agrotóxicos que tinham potencial para interagir com albumina, como a atrazina por exemplo, que foi o que virou a minha primeira publicação, como primeiro autor.”*

**3º O que foi importante saber para começar na área? Sentiu muita dificuldade no início?**

**Aluno 1:** *“Na área de bioinformática voltada pras ômicas acho que foi importante entender conceitos de Biologia Molecular. Eu já tinha conhecimentos de Python, precisei aprender um pouco de R e de comandos do terminal do Ubuntu. Mas essa parte foi bem simples. No início foi difícil porque um mestrando da UFMG que tava passando só uma semana aqui teve que me ensinar como fazer análise de expressão diferencial nesse curto espaço de tempo. e teve que me ensinar muita coisa: como funciona o sequenciamento, como tratar os dados, quais programas utilizar.”*

**Aluno 2:** *“A primeira coisa que eu tive que aprender ‘foi a mexer’ no terminal do Linux, o que já envolve programação. Essa é a etapa que eu sempre repasso ‘pra’ outros alunos como sendo essencial ‘pra’ começar em bioinformática, porque os softwares normalmente estão disponíveis*

*para linux, muitas vezes também para macOS e poucas vezes para Windows. Eu senti muita dificuldade no início porque eu tive que aprender muitas coisas ao mesmo tempo e a bioinformática que eu trabalhava era muito teórica, a gente estudava modelos de predição de energia livre de proteínas e tudo era novo demais.”*

**Aluno 3:** *“Cara, eu pensei que ia ser mais necessário saber programar mas aí não era tão necessário assim. Foi necessário o embasamento de química e de bioquímica que eu tinha na época, porque envolvia questão de proteína, de protonação, de resíduo, esse tipo de informação, pKa, essas coisas que você vê em química e em bioquímica, e ajuda a entender um pouco, por exemplo, das etapas de preparação das proteínas, também dos tipos de interação, que a gente acaba estudando o tipo interação, se é uma ligação de hidrogênio, se é uma ligação de Van der Waals, se é uma interação hidrofóbica, tudo isso a gente estuda em bioquímica, então isso aí foi importante. A dificuldade foi porque eram muitos softwares, então eu não tinha domínio sobre nenhum deles. A dificuldade maior era manusear a quantidade de ferramentas e softwares para preparar uma simulação, fazer uma modelagem, esse tipo de coisa. Então era meio difícil porque eram muito softwares para dominar, e aí você leva um tempo para saber o que cada software faz: se ele vai ser utilizado para modelagem, qual é utilizado para dinâmica, como utilizar cada um deles de forma correta né para que o protocolo seja feito de modo adequado, porque não pode ser feito de qualquer jeito. E sim, tive dificuldade no início para ir aprendendo né, porque é algo que você vai aprendendo na medida que você vai fazendo e tinham algumas metodologias que eram mais difíceis de entender, inclusive no nível teórico. Por exemplo, você entender como é que funciona a dinâmica molecular, como é que funciona o cálculo baseado em DFT, tem que entender um pouquinho de física. Talvez um pouco mais do que um pouco, um pouquinho né, entender mais do que um pouquinho na verdade para entender exatamente esse tipo de cálculo que é os cálculos que a gente faz né, as simulações dinâmicas e os cálculos quânticos. Os dois são baseados e fundamentados em teorias da física né, então até hoje eu tenho um*

*pouco de dificuldade de explicar 100% esses métodos, mas no início era mais complicado ainda porque era basicamente botar para rodar essas coisas sem entender tanto assim o fundamento por trás dessas técnicas que eram realizadas no softwares.”*

**4º Quais ferramentas você mais utilizou para realizar as pesquisas?**

**Aluno 1:** *“Eu fiz muito script de python no editor de texto padrão do ubuntu, também usei muito o RStudio para fazer a análise dos dados. Para análise de expressão diferencial eu usava o FastQC para analisar qualidade das Reads, Trimmomatic para trimmar as reads com qualidade inferior, TopHat para o alinhamento das Reads, HTseq para obter a contagem bruta. Todos os dados eram retirados do NCBI, e eu usava vários pacotes no R e no Python.”*

**Aluno 2:** *“Desde o meu mestrado, eu tenho me especializado em análise de sequências de nucleotídeos. Nessa área, as ferramentas que eu mais utilizei foram FastQC ‘pra’ analisar a qualidade dos dados de sequenciamento, DADA2 e Phyloseq pra análises de metataxonômica, e Geneious para visualização de sequências.”*

**Aluno 3:** *“Como a minha área é voltada para (bioinformática) estrutural, existem várias ferramentas né, vários softwares. A gente usa softwares para visualizar a proteína, como PyMol, como Discovery Studio, Chimera por exemplo, alguns desses podem ser utilizados para fazer preparação das estruturas de proteínas, como o Discovery e o próprio Quimera. Para fazer os estudos de protonação eu já utilizei o servidor que é baseado no próprio KA, o ProteinPrepare. ‘Pra’ preparar moléculas menores, a gente pode utilizar o Avogadro, pode utilizar o MarvinSketch ‘pra’ ver o estudo de carga. Enfim, tem uma gama de ferramentas né, além dessas que eu falei, mas essas são as que eu mais utilizo pra fazer preparação de estruturas antes de levar, por exemplo, ‘pra’ fazer um docking molecular ou uma dinâmica. Passando da etapa de preparação, geralmente a gente usa o que para fazer um docking, a gente usa o AutoDockTools para fazer a preparação dos arquivos para docking. Pra rodar o docking a gente usa, geralmente eu utilizo o AutoDock Vina que é um software que tem uma eficácia muito boa e o custo computacional dele é bem baixo se comparado com outros, mas*

*também pode ser utilizado AutoDock4, o Haddock, dentre outras ferramentas, existem uma infinidade de ferramentas para fazer docking molecular, tem servidores aí na internet bastante automatizados mas que podem ser utilizados. Mas o que eu mais utilizo é AutoDock Vina e AutoDock4. Na parte de dinâmica molecular, que geralmente a gente usa para refinar um docking ou para fazer uma simulação de um cristal que a gente já possui mas que a gente quer ver novas conformações daquele complexo protéico, ou refinar aquele nosso complexo proteico e tal. Mas aí para fazer dinâmica, já utilizei o NAMD e já utilizei o GROMACS; atualmente eu utilizo mais o GROMACS porque pro poder computacional que a gente dispõe no laboratório o GROMACS é o que melhor se adequa. ‘Pra’ fazer análise, durante muito tempo que eu utilizei o VMD né, que você consegue fazer algumas análises de RMSD e de interação utilizando o VMD. Entretanto, hoje em dia eu não utilizo mais tanto assim o VMD, a não ser em alguns casos muito específicos. Outro método que eu utilizo bastante de cálculo quântico, a gente costuma fazer no Materials 8.0, ele dispõe de vários tipos de pacotes nesse campo da química computacional ele é bastante utilizado e a gente usa um pacote de DMol3 né pra fazer esses cálculos de energia de interação entre proteína-proteína, proteína-ligante, proteína-peptídeos, esse tipo de coisa.”*

**5° Hoje, você vê como uma área em que você gostaria de trabalhar no futuro ou sente que é algo que não seja pra você neste momento?**

**Aluno 1:** *“Eu quero trabalhar com Bioinformática no futuro, mas atualmente eu estou me dedicando para trabalhar na área de ciência de dados mais empresarial para obter estabilidade financeira, porque a maioria das vagas na área de bioinformática requerem muito conhecimento e disponibilidade e não pagam bem.”*

**Aluno 2:** *“Hoje eu vejo a bioinformática como a área que possui mais oportunidades dentro das Ciências Biológicas. Existem inúmeras vantagens práticas em ser um profissional que trabalha com bioinfo, e, por ser novo e considerado algo difícil, o mercado ainda está folgado, com espaço ‘pra’ novos profissionais especializados, e a tendência é só*

*crescer. Essa tendência não é apenas no Brasil, existem oportunidades em diversos lugares, porque as tecnologias de geração e de análise de dados ainda estão se especializando, com diversas barreiras que ainda precisam ser superadas.”*

**Aluno 3:** *“Assim, é uma área que ‘tá’ numa crescente muito interessante né, com o surgimento da Inteligência Artificial né, o boom da Inteligência Artificial ai nos últimos meses, é uma área que vai ser bastante impactada né, a bioinformática, tanto a estrutural quanto a genômica, nos últimos anos a gente tem visto algumas empresas muito grande como a Google e o Facebook investindo pesado nessa parte de predição de proteínas, é um campo que tem crescido bastante e que dá para trabalhar sim, principalmente no exterior. Existem alguns dados mostrando que as indústrias farmacêuticas em outros países, isso ainda não é tão cultural tão forte aqui no Brasil, mas em outros países essas indústrias elas têm contratado bastante profissionais pra atuar nessa área porque elas têm percebido quão poderosas são essas ferramentas nessa área de drug discovery ‘né’, descoberta de novas drogas. É uma área que continua em crescente e essa parte de bioinformática ela tem conseguido bons resultados, boas predições, e tem dado um retorno bom para essas empresas, por isso elas tão contratando. Então é uma área, assim, que eu consigo me enxergar atuando, mas não aqui no Brasil (a não ser como potencial pesquisador), mas fora do Brasil parece ser uma realidade mais provável.”*

### **3.3 Pontos convergentes**

Baseado nas experiências do autor deste trabalho e dos relatos dos entrevistados, apesar de cada uma ser única, podemos traçar pontos convergentes e comuns provenientes das vivências de cada um na área de bioinformática:

- **Difícilmente entram na graduação estando cientes da área:** a bioinformática, apesar de ser um campo tão importante, ainda é muito desconhecida aos olhos do público em geral. A falta de incentivo governamental e desconhecimento geral causam essa falta de difusão (BERZIN e CASTRO, 2021), o que corrobora ao fato de poucos

estudantes entrarem em suas graduações já conscientes da área.

- **Poucas ou nenhuma cadeira voltada a área na graduação e/ou estrutura curricular:** como descrito no tópico 1.4, existem uma quantidade pequena ou até mesmo nenhuma cadeira voltada especialmente à área nas estruturas curriculares dos cursos de graduação, o que acaba por não facilitar um primeiro contato dos alunos com a área.
- **Além dos conhecimentos no campo de estudo, é necessário ter conhecimentos em Biologia Molecular e Biofísica:** a bioinformática, como campo interdisciplinar, se caracteriza como um conjunto de metodologias e conhecimentos biológicos de outras áreas, como a Biologia Molecular e a Biofísica (GHIRINGHELLI e FERNÁNDEZ, 2017). Sendo assim, fundamentos e conhecimentos nessas áreas são essenciais para um maior entendimento das metodologias e funcionamentos envolvidos.
- **Dificuldade no início, devido a enorme quantidade de softwares e dados:** como ilustrado na Tabela 1 e discutido ao longo dos tópicos 2.1 e 2.2, a grande quantidade de ferramentas dispostas para tratar da colossal quantidade de dados torna a adaptabilidade difícil no início de quem está entrando na área.
- **Baixas perspectivas com o mercado de trabalho no Brasil:** como apontado por Berzin e Castro (2021), a área se mantém pouco relevante em território nacional devido ao pouco investimento na educação e na ciência. Embora muito difundido dentro do meio acadêmico e em instituições de pesquisa vinculadas a universidades em geral, o setor é pequeno quando comparado a outros países. No tópico a seguir, o tema será abordado pouco mais a fundo.

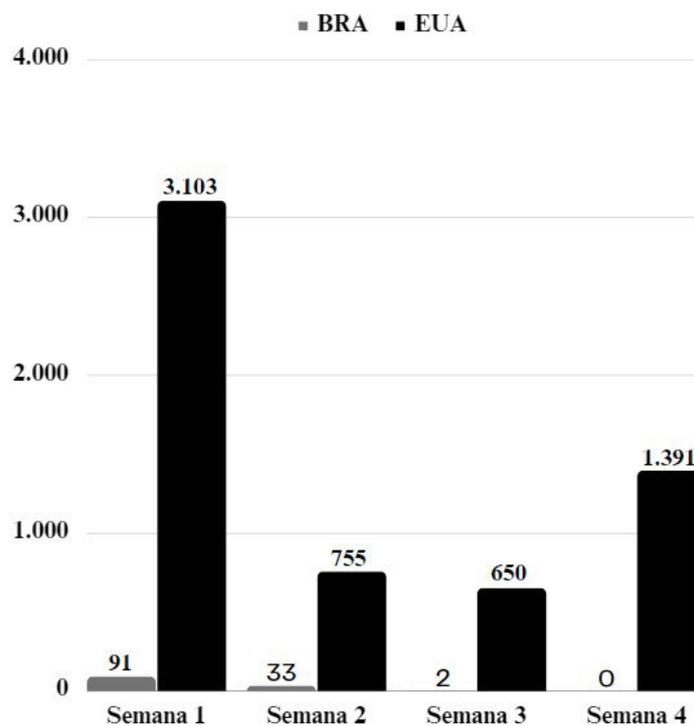
#### 4 PERSPECTIVAS DO MERCADO ATUAL

Com a intenção de trazer uma melhor perspectiva sobre como está o mercado de trabalho de Bioinformática até a data em que este manuscrito está sendo produzido, foi coletado durante um período de quatro semanas as vagas abertas no LinkedIn. Segundo Albrecht et al. (2011), o LinkedIn representa um banco de dados que possui 90 milhões de profissionais de negócios, o que facilita na construção de uma rede de negócios (BRADBURY, 2011). Portanto, a plataforma caracteriza uma poderosa ferramenta na busca e divulgação de vagas de emprego de diversos tipos.

Para uma melhor evidência das vagas em questão, foi adotado dois critérios para o

mecanismo de busca: vagas mais recentes (dentro de um período de uma semana) e localidade. As localidades escolhidas foram o Brasil, por ser o país-alvo dos estudantes que participaram das entrevistas deste trabalho, e os Estados Unidos (EUA), por representar um país com uma crescente demanda por profissionais qualificados na área (HEMMINGER et al., 2005). As palavras de busca foram ‘bioinformática’ para a localidade Brasil e ‘bioinformatics’ para a localidade EUA. O Gráfico 5 representa a disposição das vagas de acordo com cada semana de busca:

Gráfico 5 – Vagas de cada localidade, de acordo com as respectivas semanas.



Fonte: LinkedIn.

Embora o pouco investimento de ciência no país (GHIRINGHELLI e FERNÁNDEZ, 2017), há uma explicação para a enorme discrepância do número de vagas de trabalho para as duas localidades. Culturalmente, vagas aplicadas à ciência nos Estados Unidos são divulgadas no LinkedIn; muito se deve ao fato de haverem instituições públicas, como universidades e centros acadêmicos, e iniciativas privadas que ofertam suas vagas a procura de acadêmicos e/ou profissionais da área na plataforma, diferentemente do Brasil.

Aqui, por outro lado, as instituições de ensino superior tem exercido o papel de fornecer conhecimento científico através de seus programas de pós-graduação, evidenciando uma clara ausência do setor produtivo e/ou privado na Ciência e Tecnologia do país e que

resulta numa grande limitação do sistema (BERTERO, 1994). Em geral, as vagas para ciência no Brasil são bastante concentradas nos polos acadêmicos ao redor do país, o que explica a grande diferença no número de vagas presentes no LinkedIn no Gráfico 5. Outra explicação seria a forma como essas vagas são divulgadas: notoriamente, vagas voltadas a programas de pós-graduação são divulgadas em diários oficiais e editais. Podemos destacar, também, que laboratórios e centros de pesquisa do país também ofertam vagas semelhantemente ao modo dos programas de pós-graduação. É o caso, por exemplo, do LNBio (Laboratório Nacional de Biociências), do CNPEM (Centro Nacional de Pesquisa em Energia e Materiais) e da Embrapa.

Existem, então, ao redor do país várias oportunidades para se trabalhar com pesquisa em bioinformática nas instituições públicas e de pesquisa. Entretanto, é notório que um pequeno número de vagas para o setor de bioinformática no país está sendo divulgado no LinkedIn. Isso pode representar o começo de um crescimento nas vagas de trabalho disponíveis para o setor, principalmente através de empresas privadas e instituições de pesquisa estrangeiras.

## 5 CONSIDERAÇÕES FINAIS

A Bioinformática representa um campo presente dentro da grande área que é a Biologia, estando em constante crescimento desde suas origens e se tornando cada vez mais essencial na atualidade. Ela transita principalmente entre as áreas da Biologia, Química e Computação, mas também se utiliza de conceitos da Física, Estatística e Matemática. Sua importância se dá pela grande quantidade de dados que precisam ser tratados e que geram resultados em pouco tempo. Esses resultados auxiliam profissionais de todos os campos de atuação a resolverem problemas ou a otimizar soluções biológicas.

Sendo uma área tão emergente e importante, é apropriado observar que o mercado no exterior está aquecido. Empresas e universidades investem bastante na área que tem dado bastante retorno, muito vinculado ao avanço tecnológico que os sistemas computacionais sofrem todo ano. Apesar disso, o mercado atual no Brasil ainda está principiando, sendo as vagas vinculadas principalmente a universidades e instituições públicas. Em concordância a isso, temos alunos na graduação de diversos cursos que acabam por trabalhar na área através dos programas de pós-graduação e afins

Levando isso em consideração, a universidade e a coordenação dos cursos em que cabe a área poderiam incorporar mais cadeiras voltadas a mesma em suas estruturas curriculares e direcionar mais o estágio para os alunos que demonstrem interesse. Desse modo, a inserção no mercado de trabalho poderia ser facilitada, haja vista que representa um campo em exponencial crescimento e demanda por pesquisadores e profissionais.

## REFERÊNCIAS

- ABDELNUR, Patrícia Verardi. **Metabolômica e espectrometria de massas**. 2011.
- ALBERTS, Bruce et al. **Biologia molecular da célula**. Artmed Editora, 2017.
- ALBRECHT, W. David et al. **LinkedIn for accounting and business students**. American Journal of Business Education (AJBE), v. 4, n. 10, p. 39-42, 2011.
- ALTSCHUL, Stephen F. et al. **Basic local alignment search tool**. Journal of molecular biology, v. 215, n. 3, p. 403-410, 1990.
- ASLAM, Bilal et al. **Proteomics: technologies and their applications**. Journal of chromatographic science, p. 1-15, 2016.
- BACH, Rene; IWASAKI, Yumi; FRIEDLAND, Peter. **Intelligent computational assistance for experiment design**. 1984.
- BATISTA, Eraldo Carlos; DE MATOS, Luís Alberto Lourenço; NASCIMENTO, Alessandra Bertasi. **A entrevista como técnica de investigação na pesquisa qualitativa**. Revista Interdisciplinar Científica Aplicada, v. 11, n. 3, p. 23-38, 2017.
- BEHJATI, Sam; TARPEY, Patrick S. **What is next generation sequencing?**. Archives of Disease in Childhood-Education and Practice, v. 98, n. 6, p. 236-238, 2013.
- BENSON, Dennis A. et al. **GenBank**. Nucleic acids research, v. 41, n. D1, p. D36-D42, 2012.
- BERMAN, Helen M. et al. **Trendspotting in the protein data bank**. FEBS letters, v. 587, n. 8, p. 1036-1045, 2013.
- BERTERO, Carlos Osmar. **Gestão de Ciência e Tecnologia: uma análise institucional**. PADCT II, 1994.
- BERZIN<sup>1</sup>, Brenda Sanches; CASTRO, Gustavo Pereira. **Bioinformática: Uma visão geral sobre o campo**. 2021.
- BRADBURY, Danny. **Data mining with LinkedIn**. Computer Fraud & Security, v. 2011, n. 10, p. 5-8, 2011.
- BRASIL. **Decreto 88.438, de 28 de junho de 1983**. Dispõe sobre a regulamentação do exercício da profissão de Biólogo, de acordo com a Lei nº 6.684, de 3 de setembro de 1979 e de conformidade com a alteração estabelecida pela Lei nº 7.017 de 30 de agosto de 1982. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto/1980-1989/D88438.htm](http://www.planalto.gov.br/ccivil_03/decreto/1980-1989/D88438.htm). Acesso em: 24 abr. 2023.
- BRASIL. **Lei 6.684, de 3 de setembro de 1979**. Regulamenta as profissões de Biólogo e de Biomédico, cria o Conselho Federal e os Conselhos Regionais de Biologia e Biomedicina, e dá outras providências. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/leis/L6684.htm](http://www.planalto.gov.br/ccivil_03/leis/L6684.htm). Acesso em: 24 abr. 2023.

BRASIL. **Lei 7.107, de 30 de agosto de 1982.** Dispõe sobre o desmembramento dos Conselhos Federal e Regionais de Biomedicina e de Biologia. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/leis/L7107.htm](http://www.planalto.gov.br/ccivil_03/leis/L7107.htm). Acesso em: 24 abr. 2023.

BRITO, S. **Relembre: Os 10 maiores avanços científicos dos anos 2010.** VEJA, 2019. Disponível em: <https://veja.abril.com.br/ciencia/relembre-os-10-maiores-avancos-cientificos-dos-anos-2010>. Acesso em: 16 jul. 2023.

BUENO, Maria Rita Passos. **O projeto genoma humano.** *Revista Bioética*, v. 5, n. 2, 2009.

CERRI, Ricardo et al. **Aprendizado de máquina:** breve introdução e aplicações. *Cadernos de Ciência & Tecnologia*, v. 34, n. 3, p. 297-313, 2019.

CFBio. **Perguntas Frequentes.** Disponível em: <https://cfbio.gov.br/perguntas-frequentes/>. Acesso em: 15 mai. 2023.

CFBio. **Resolução Nº 227, de 18 de setembro de 2010.** Dispõe sobre as atribuições do Biólogo e define as áreas de atuação. Disponível em: [http://www.cfbio.gov.br/files/Resolucoes/Res\\_227.pdf](http://www.cfbio.gov.br/files/Resolucoes/Res_227.pdf). Acesso em: 24 abr. 2023.

COELHO, Micaella et al. **Desenvolvimento de um Framework de Aprendizado de Máquina no Apoio a Gateways Científicos Verdes, Inteligentes e Eficientes:** BioinfoPortal como Caso de Estudo Brasileiro. In: Anais do XXIII Simpósio em Sistemas Computacionais de Alto Desempenho. SBC, 2022. p. 205-216.

COHEN, Jacques. **Bioinformatics—an introduction for computer scientists.** *ACM Computing Surveys (CSUR)*, v. 36, n. 2, p. 122-158, 2004.

CRESAWN, Steven G. et al. **Phamerator:** a bioinformatic tool for comparative bacteriophage genomics. *BMC bioinformatics*, v. 12, n. 1, p. 1-15, 2011.

DIAS, Maria Fernanda Ribeiro; PASCUTTI, Pedro Geraldo; DA SILVA, Manuela Leal. **APRENDIZADO DE MÁQUINA E SUAS APLICAÇÕES EM BIOINFORMÁTICA.** *Semioses*, v. 10, n. 1, p. 23-37, 2016.

EDWARDS, David; STAJICH, Jason; HANSEN, David (Ed.). **Bioinformatics: tools and applications.** Springer science & business media, 2009.

EXALY TRENDS. **Bioinformatics.** Disponível em: <https://exaly.com/trends/bioinformatics/?from=1950&to=2022>. Acesso em: 7 jun. 2023.

FOLHA ONLINE. **Saiba quais foram os principais fatos da ciência em 2009.** Folha de S.Paulo. 2009. Disponível em: <https://m.folha.uol.com.br/ciencia/2009/12/672051-saiba-quais-foram-os-principais-fatos-da-ciencia-em-2009.shtml>. Acesso em: 16 jul. 2023.

GABRIEL, Jane Eyre; DE FIGUEIREDO, Danielle Duarte Lima Germano; DE FARIAS, Rodrigo Pessôa. **A caracterização estrutural e funcional de biomoléculas e as ferramentas de bioinformática.** *Rev. Cient. COOPEX FIP*, Patos, v. 1, p. 1-9, 2009.

GERSTEIN, Mark B. et al. **What is a gene, post-ENCODE?** History and updated definition. *Genome research*, v. 17, n. 6, p. 669-681, 2007.

GHIRINGHELLI, Pablo Daniel; FERNÁNDEZ, Graciela. **La bioinformática y la biología de sistemas**. *Divulgatio. Perfiles académicos de posgrado*, v. 1, n. 02, p. 16-19, 2017.

HEMMINGER, Bradley M.; LOSI, Trish; BAUERS, Anne. **Survey of bioinformatics programs in the United States**. *Journal of the American Society for Information Science and Technology*, v. 56, n. 5, p. 529-537, 2005.

KUHN, Matthias et al. **Finding small somatic structural variants in exome sequencing data: a machine learning approach**. *Computational Statistics*, v. 33, p. 1145-1158, 2018.

LIBBRECHT, Maxwell W.; NOBLE, William Stafford. **Machine learning applications in genetics and genomics**. *Nature Reviews Genetics*, v. 16, n. 6, p. 321-332, 2015.

LIMA, Maria Emília Caixeta de Castro; GERALDI, Corinta Maria Grisolia; GERALDI, João Wanderley. **O trabalho com narrativas na investigação em educação**. *Educação em revista*, v. 31, p. 17-44, 2015.

LUSCOMBE, Nicholas M.; GREENBAUM, Dov; GERSTEIN, Mark. **What is bioinformatics?** An introduction and overview. *Yearbook of medical informatics*, v. 10, n. 01, p. 83-100, 2001.

MARDIS, Elaine R. **A decade's perspective on DNA sequencing technology**. *Nature*, v. 470, n. 7333, p. 198-203, 2011.

MAYR, Ernst. **Biologia, ciência única**. São Paulo: Companhia das Letras, 2004.

METZKER, Michael L. **Sequencing technologies—the next generation**. 2010. *Nature Genetics*.

MINISTÉRIO DA EDUCAÇÃO. **Base Nacional Comum Curricular**. 2018. Disponível em: <http://basenacionalcomum.mec.gov.br/>. Acesso em: 08 maio 2023.

MINISTÉRIO DA EDUCAÇÃO. **PARÂMETROS CURRICULARES NACIONAIS**. Introdução aos parâmetros curriculares nacionais. Brasília: MEC/SEF, 1997. Disponível em: <http://portal.mec.gov.br/seb/arquivos/pdf/livro01.pdf>. Acesso em: 08 maio 2023.

MOORE, Gordon E. et al. **Moore's law at 40**. 2006.

MOORTHIE, Sowmiya; MATTOCKS, Christopher J.; WRIGHT, Caroline F. **Review of massively parallel DNA sequencing technologies**. *The HUGO journal*, v. 5, p. 1-12, 2011.

MORAIS, Pablo A. et al. **The urokinase plasminogen activator binding to its receptor: a quantum biochemistry description within an in/homogeneous dielectric function framework with application to uPA–uPAR peptide inhibitors**. *Physical Chemistry Chemical Physics*, v. 22, n. 6, p. 3570-3583, 2020.

MOROZOV, Alexandre V. et al. **Protein–DNA binding specificity predictions with**

**structural models.** Nucleic acids research, v. 33, n. 18, p. 5781-5798, 2005.

NATIONAL LIBRARY OF MEDICINE. **PubMed.** Disponível em: <https://pubmed.ncbi.nlm.nih.gov/>. Acesso em: 7 jun. 2023.

NEGRETTO, Diego Henrique. **Algoritmos de aprendizado semi-supervisionado baseados em grafos aplicados na bioinformática.** 2016.

PEREIRA, Rute; OLIVEIRA, Jorge; SOUSA, Mário. **Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics.** Journal of clinical medicine, v. 9, n. 1, p. 132, 2020.

PETTERSSON, Erik; LUNDEBERG, Joakim; AHMADIAN, Afshin. **Generations of sequencing technologies.** Genomics, v. 93, n. 2, p. 105-111, 2009.

PROGRAD. **Ciências Biológicas.** Disponível em: <https://prograd.ufc.br/pt/cursos-de-graduacao/ciencias-biologicas-bacharelado-e-licenciatura-fortaleza/>. Acesso em 08 mai. 2023.

RAVISHANKAR, Shashidhar et al. **Next-generation Sequence-analysis Toolkit (NeST): A standardized bioinformatics framework for analyzing Single Nucleotide Polymorphisms in next-generation sequencing data.** bioRxiv, p. 323535, 2018.

REDMAN, Melody et al. **What is CRISPR/Cas9?** Archives of Disease in Childhood-Education and Practice, v. 101, n. 4, p. 213-215, 2016.

REUTER, Jason A.; SPACEK, Damek V.; SNYDER, Michael P. **High-throughput sequencing technologies.** Molecular cell, v. 58, n. 4, p. 586-597, 2015.

SCHENA, Mark (Ed.). **DNA microarrays: a practical approach.** Practical approach series, 1999.

SINGH, Gautam B. **Fundamentals of bioinformatics and computational biology.** Cham: Springer International Publishing, p. 159-170, 2015.

SLIWOSKI, Gregory et al. **Computational methods in drug discovery.** Pharmacological reviews, v. 66, n. 1, p. 334-395, 2014.

STEPHENS, Zachary D. et al. **Big data: astronomical or genetical?.** PLoS biology, v. 13, n. 7, p. e1002195, 2015.

SU, Chen. **Bioinformatics: A Practical Guide to the Analysis of Genes & Proteins,**(third edition). Edited by Andreas D. Baxevanis and BF Francis Ouellette New York: John Wiley & Sons; ISBN: 0471 478784; 540pp.; 2004; \$79.95. 2006.

TUCKER, Tracy; MARRA, Marco; FRIEDMAN, Jan M. **Massively parallel sequencing: the next big thing in genetic medicine.** The American Journal of Human Genetics, v. 85, n. 2, p. 142-154, 2009.

VERLI, Hugo. **Bioinformática: da biologia à flexibilidade molecular.** 2014.

VOELKERDING, Karl V.; DAMES, Shale A.; DURTSCHI, Jacob D. **Next-generation sequencing:** from basic research to diagnostics. *Clinical chemistry*, v. 55, n. 4, p. 641-658, 2009.

WETTERSTRAND, K. A. **DNA Sequencing Costs:** Data from the NHGRI Genome Sequencing Program (GSP). 2022. Disponível em: [www.genome.gov/sequencingcostsdata](http://www.genome.gov/sequencingcostsdata). Acesso em: 08 mai. 2023.

ZHAN, Mei-Miao et al. **From monoclonal antibodies to small molecules:** the development of inhibitors targeting the PD-1/PD-L1 pathway. *Drug Discovery Today*, v. 21, n. 6, p. 1027-1036, 2016.

ZHANG, Changsheng; LAI, Luhua. **Towards structure-based protein drug design.** 2011.