Como começar na Bioinformática?

By Lucas Miguel

8 de junho de 2021

Como começar na Bioinformática?

Lucas Miguel Carvalho

Revisão: Diego Mariano

BIOINFO – Revista Brasileira de Bioinformática. Edição #01. Julho, 2021.

DOI: [10.51780/978-6-599-275326-01](https://doi.org/10.51780/978-6-599-275326-01)

O conteúdo e imagens do texto "Como começar na bioinformática?" foram previamente publicados em totalidade em Terabytes of life [Blogs de Ciência da Unicamp]. [acesso: 26 de maio de 2021].

A Bioinformática é uma ciência que vem crescendo ao longo dos anos, o que faz com que a adesão profissional nesta área seja altamente necessária. Mas caso você esteja iniciando na área, por onde se deve começar? Como dar o pontapé inicial? Como criar uma formação multidisciplinar e buscar aprimoramento? Esse é o objetivo deste capítulo.

Um perfil de profissional multidisciplinar é o desejado para um(a) bioinformata e, para isso, deve-se criar sempre o hábito de se aventurar por novos horizontes e não hesitar. A primeira característica principal de um(a) bioinformata é a curiosidade. Quantas vezes você já não se encontrou em uma barreira lógica e teve que buscar soluções? Essa habilidade de buscar soluções em diferentes locais e maneiras faz com que este tipo de profissional também seja muito cobiçado no mercado. Independente da sua formação, seja ela vinda de exatas ou biológicas, existem dois passos

essenciais para uma progressão inicial em Bioinformática que te ajudarão a se inteirar nos assuntos das discussões em grupos. Eles são:

1. Ler a introdução de artigos científicos
2. Ler e testar programas e pipelines do Material e Métodos

Estas duas ações te ajudarão muito a crescer. O passo (1) sempre trás citações de outros artigos da área de estudo que te embasarão no tema. As introduções de artigos de análise clássicas em bioinformática ou de artigos mais aplicados, que mesmo assim utilizam o ferramental da área, sempre proporcionam uma cascata de conhecimento gigantesca. Por sua vez, o passo (2) acaba sendo essencial para que você inicie um primeiro contato com os programas utilizados em sua área de estudo e esteja sempre atento, principalmente aos parâmetros utilizados nos programas de bioinformática, que são essenciais para reprodução.

Como mencionado, o perfil de um (a) bioinformata é bastante heterogêneo. Dependendo do perfil inicial do profissional, há algumas diferenças em suas abordagens de início, e são elas que irão ser abordadas ao longo deste capítulo. Não deixem de acompanhar as dicas de leitura, vídeos, cursos e afins no final deste texto.

Antes de começar, aconselho a todos lerem um pouco mais sobre a história da Bioinformática em um ótimo texto elaborado por Jeff Gauthier e colaboradores [1], que se intitula "*A brief history of bioinformatics*". Nesse texto você entende o surgimento da Bioinformática há mais de 50 anos atrás, e toda a sua evolução até os dias de hoje.

Perfil de Ciências Exatas

Os iniciantes no ramo da bioinformática oriundos de cursos de exatas, como Ciência da Computação, Matemática, Física ou Engenharias, tendem a ter mais dificuldade em temas relacionados à biologia. Mas, esses profissionais devem usar a capacidade de já ter tido contato com algoritmos e linguagens de programação para se aperfeiçoar. As dicas são as seguintes:

Introdução de artigos. Como mencionado, ler as introduções de vários artigos da área de estudo de interesse te traz uma grande bagagem. Inclusive, busque outros artigos citados nas introduções. Um dos principais desafios e motivações para alguém das exatas é saber onde aplicar seu conhecimento na área de biológicas. Os programas de bioinformática, por exemplo, introduzem o problema que querem solucionar, e acabam por motivar e exemplificar tais problemas.

Leiam revisões. As revisões ajudam e muito a entender os estudos atuais das áreas de pesquisas. Pergunte ao seu orientador ou supervisor sobre revisões atuais, se não, busque no Google Scholar por palavras-chave relacionadas a sua pesquisa.

Se aperfeiçoe em linguagens de programação. Chega um momento que estudantes e profissionais de exatas terão que desenvolver pacotes ou software de bioinformática, e necessitarão de muita especialidade, principalmente para paralelização e web. Além disso, integrar resultados e realizar pipelines é essencial, logo, precisam de habilidades nestes aspectos. Na bioinformática existe uma alta tendência na utilização do Python e R para análises mais complexas, mas não que Perl, C++ ou MATLAB deixem de contribuir com pacotes e programas sensacionais.

Aprenda gerenciamento de processos e servidores Linux. Esta dica cabe tanto para os dois perfis discutidos neste capítulo. Saber gerenciar processos em um servidor é essencial para análises mais robustas e que necessitem de paralelização. Quem nunca teve que reiniciar um servidor, verificar uma fila de processos, observar os verboses, utilizar GPUs, alterar permissões e matar o processo baleia do amiguinho?

Seminários e livros de genética e biologia molecular. Importante destacar que, na imersão nos temas de bioinformática, é necessário entender o que acontece dentro da célula. Livros de Genética e Biologia Molecular trazem conceitos básicos e importantes sobre o porquê executar tal ferramenta em determinado aspecto. Por exemplo, a diferença no tratamento de montadores de genomas e transcriptomas devido ao tipo de dado que cada um recebe como entrada. Os seminários em si são essenciais para conhecer rotinas de pesquisa na área de genética, além de promover uma imersão com o grupo. Meu orientador do doutorado sempre dizia que os “Seminários são o momento em que se tem a interação do laboratório e integração de habilidades, logo, eles são essenciais”.

Leia os artigos dos programas que você utiliza. Uma das grandes habilidades que vejo em grandes profissionais da área é tentar ensinar as aplicações de forma mais clara. Entender o processo por trás do software, e não apenas apertar um botão, faz com que você consiga interpretar e explicar ele com mais facilidade. Vindo da área de exatas, a todo o momento você, estudante ou profissional, teve contato com manuais e afins. Logo, a tendência é saber

lidar com manuais e artigos técnicos, assim, tente entendê-los mais a fundo e repasse o conhecimento.

Pergunte “por quês” e “ondes”. Por que realizar tal análise? Por que o mecanismo é este? Por que não executar um pipeline assim? Por que o experimento é desta maneira? Onde queremos chegar? Onde posso começar? Onde posso buscar? Se questionar quando não entender os procedimentos biológicos te trás uma maior bagagem sobre o que está acontecendo a sua volta e onde você pode ajudar. Não tenha medo.

Realize cursos de bioinformática. Existem muitos cursos de bioinformática, canais do YouTube, blogs, entre outros, que tratam sobre assuntos de bioinformática e trazem visões diferentes de problemas abordados por diversas áreas. Um curso de bioinformática pode te trazer tanta bagagem, que você mudará suas visões a respeito da área. Os problemas que podem ser resolvidos através da bioinformática são inúmeros, e isso é o que a faz ser tão essencial.

Não deixe os algoritmos morrerem. Algoritmos estão em toda bioinformática e biologia computacional, e são uma pitada de elegância nesta área. Pessoas vindas da computação e afins se esbarram com eles a todo o momento, logo, tragam eles para seu dia a dia. Análise de algoritmos e heurísticas são disciplinas importantes para se elaborar e solucionar problemas complexos da área e não merecem ser deixados de lado! Se transforme em um(a) biologista computacional de sucesso.

Resolva problemas do Rosalind. Resolver problemas iniciais na bioinformática é o primeiro grande passo. Então, sugiro que gaste um tempo resolvendo problemas do Rosalind, que é uma plataforma única para aprender bioinformática e programação através da resolução de problemas práticos. Você pode acessar através do link: <http://rosalind.info/problems/list-view/>.

Aprenda a gerenciar banco de dados. Esse requisito inclui bancos de dados relacionais tradicionais que são à base do SQL (por exemplo, MySQL, MariaDB, SQL Server e Oracle). Você também deve conhecer os bancos de dados NoSQL que não são relacionais, distribuídos, de código aberto e escalonáveis horizontalmente (por exemplo, MongoDB). Por fim, existem bancos de dados públicos (por exemplo, TCGA) e bancos de dados de análise de big data (por exemplo, Vertica) sobre os quais você deve aprender.

Perfil de Ciências Biológicas

Você, pesquisador (a) ou estudante, oriundo de cursos de biológicas, como Biologia, Medicina, Biomedicina e afins, necessita de uma atenção maior em adquirir habilidades computacionais. As dicas que deixo para vocês são:

Aprenda Lógica Computacional. Você, teoricamente, nunca teve contato com lógica computacional ao longo da graduação, logo, chegou o momento dela ser introduzida a você. Aprender a elaborar um algoritmo, e conhecer as ferramentas para solucioná-lo e processá-lo são as principais necessidades aqui. Com esta habilidade você poderá abordar problemas mais complexos, utilizando a lógica para resolvê-los.

Aprenda uma linguagem de programação. Saber programar é a habilidade que te fornece autonomia na bioinformática. Não deixe seu medo de informática te bloquear, siga em frente! A programação vai te ajudar a estruturar algoritmos e solucionar problemas. Aposto que será a melhor porta a ser aberta em toda sua vida! Muitas escolas inclusive estão colocando programação como grade básica, já que ela estará presente em toda sociedade nos próximos anos. Se quiser começar, sugiro o Python, Perl, C++ ou R.

Realize cursos de bioinformática. Existem muitos cursos de bioinformática, canais do YouTube, blogs, entre outros, que tratam sobre assuntos de bioinformática e trazem visões diferentes de problemas abordados por diversas áreas. Um curso de bioinformática pode te trazer tanta bagagem, que você mudará suas visões a respeito da área. Os problemas que podem ser resolvidos através da bioinformática são inúmeros, e isso é o que a faz ser tão essencial.

Leia revisões. As revisões ajudam muito a entender os estudos atuais das áreas de pesquisas. Pergunte ao seu orientador sobre revisões atuais, se não, busque no Google Scholar por palavras chaves. Revisões mais profundas de análises e pipelines auxiliam demais quem não está acostumado a utilizar programas para análises de bioinfo.

Aumente sua capacidade biológica sobre sua pesquisa. Você tem habilidades em biologia que alguém vindo de exatas não tem, então, use isto a seu favor. Leia bastante sobre o tema, e busque a dica da leitura das introduções, no mais, seja curioso (a).

Aprenda Linux. Os sistemas computacionais na bioinformática são, em sua imensa maioria, todos vinculados a alguma distribuição do GNU. Seja ele Ubuntu, Fedora, CentOS, entre outros. Você criará uma habilidade realmente

necessária para manipular dados biológicos. Seja parte do time do MKDIR. Não economize no tempo gasto nesta dica. Não se esqueça de aprender Bash. Uma grande diferença é se aventurar e acabar se esbarrando em dificuldades. Mas, primeiro, tente se aventurar! Vários pipelines na bioinformática já estão pré-estabelecidos, e basta você os executar para realizar sua análise. Mas aqui vai uma dica: não seja a pessoa que aperta somente o botão. Sempre busque saber o que está acontecendo em cada parte e o porquê!

Resolva problemas do Rosalind. Resolver problemas iniciais na bioinformática é o primeiro grande passo. Então sugiro que gaste um tempo resolvendo problemas do Rosalind, uma plataforma única para aprender bioinformática e programação através da resolução de problemas. Mas antes não se esqueça dos passos anteriores! Você pode acessar através do link: <http://rosalind.info/problems/list-view/>.

Aprenda algoritmos a fundo. Algoritmos são ferramentas essenciais para se equilibrar com problemas complexos na bioinformática. Ter conhecimentos em construção de algoritmos pode te fazer um pesquisador (a) mais consolidado (a) em saber criticar e comparar programas e metodologias. Existem livros diversos de análise de algoritmos. Então, comece a lê-los.

Aprenda Estatística. Essa dica serve para os dois perfis em questão deste capítulo. Saber aplicar a estatística correta em seus dados é essencial, e softwares podem te ajudar neste momento, como o SPSS e o SAS, além das análises em R.

Habilidades gerais

O profissional em Bioinformática pode atuar em várias frentes, principalmente nas análises em diferentes dados ômicos. Mas quais as principais habilidades esperadas por um (a) bioinformata? A partir de alguns artigos da área, pude resumir tais habilidades [2, 3].

Habilidades em bioinformática

- Ferramentas de alinhamento de sequência, como BLAST ou Bowtie;
- O *Genome Analysis Toolkit* (GATK);
- Software para sequenciamento NGS, *microarray*, qPCR e análise de dados;
- Ferramentas para lidar com dados de sequenciamento, como o samtools;
- Para obter conjuntos de dados genéticos, use uma ferramenta como o Ensembl;
- Ferramentas para sistemas de busca de banco de dados, como o Entrez;
- Pipeline de análise de dados de NGS.

Mineração de dados e aprendizado de máquina

Técnicas de aprendizado de máquina, como clusterização hierárquica, agrupamentos, árvore de regressão e árvores de decisão, também são úteis.

Habilidades em áreas generalizadas

Existem habilidades adicionais importantes, como ser multitarefa, possuir independência, boas habilidades de comunicação, ser curioso, possuir raciocínio analítico e habilidades gerenciais.

Conhecimento em biologia

Esse requisito varia de acordo com o seu domínio de estudo ou o trabalho específico ao qual você está se candidatando. Você provavelmente precisará de conhecimentos sobre biologia molecular, genética, biologia do câncer e/ou biologia moderna.

Habilidades estatísticas

- Sistemas de software estatístico como SPSS e SAS.
- Como fazer análises estatísticas com Python ou R.

Habilidades de programação

- Uma ou mais destas linguagens de programação: R, Perl, Python, Java e Matlab.
- Ferramentas e bibliotecas de aprendizado de máquina, como Mllib e Scikit-Learn em Python, são muito úteis para aprender.

Dicas de cursos, livros e afins

Cursos

Uma das principais maneiras de agregar conhecimento em bioinformática, seja teórico ou prático, é através de cursos. Dentre eles, destaco:

1. EMBL-EBI treinamentos

2. Biologia Molecular

3. Perl – Tutorial

4. Aprenda Perl em 21 dias

5. Curso em vídeo – Python

6. Cursos de R

7. Cursos de R – Video

8. Machine Learning

9. CookBook em R

10. Real Python – Cursos

11. Scripts em Shell – Linux

12. Curso de Linux COMPLETO

13. Curso verão Bioinfo – USP

14. Curso LaCTAD Bioinfo – UNICAMP

15. Cursos de Bioinfo em geral

16. Curso de curta duração – UFRN

Canais YouTube

Muitas vezes, através de uma explicação mais detalhada e visual, o aprendizado se torna mais consolidado. Muitos canais de YouTube, sendo a maioria de fora do país, tentam explicar a Bioinformática e temas relacionados de uma maneira mais simples. Dentre eles, destaco:

1. *Online Bioinfo (conteúdo em português)* (Profa. Dra. Raquel Minardi – UFMG)

2. *MIT OpenCourseWare – MIT 7.91J Foundations of Computational and Systems Biology playlist*

3. *Manolis Kellis – MIT CompBio playlist*

4. *StatQuest with Josh Starmer*

5. *Bioinformatics Algorithms: An Active Learning Approach*

Livros

Assim como muitos preferem ler livros em PDF, também existe o time daqueles que preferem o livro em papel. Há muitos livros em Bioinformática lançados, incluindo aqueles focados em atualizações de pipelines, mas também existem aqueles que descrevem de conceitos básicos até avançados. Para vocês, deixo estas opções:

1. Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2002). Algoritmos: teoria e prática. *Editora Campus*, 2, 296.
2. Claverie, J. M., & Notredame, C. (2006). *Bioinformatics for dummies*. John Wiley & Sons.
3. Lesk, A. M. (2008). *Introdução à bioinformática*. Artmed.
4. Compeau, P., & Pevzner, P. A. (2018). *Bioinformatics Algorithms: An Active Learning Approach*. La Jolla, CA: Active Learning Publishers.
5. Siddesh, G. M. (2020). *Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools, and Applications*. Springer Nature.
6. Yamagishi, M. E. B. (2017). *Mathematical Grammar of Biology*. Springer International Publishing.
7. Setubal, J. C., Meidanis, J., & ... Setubal-Meidanis. (1997). *Introduction to computational molecular biology* (No. 04; QH506, S4.). Boston: PWS Pub.
8. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. (2017) *Biologia Molecular da Célula*. 6a ed. Artmed, Porto Alegre. 1464p.
9. Griffiths, A.J.F.; Wessler, S.R.; Carroll, S.B.; Doebley, J. (2016) *Introdução à Genética*. 11a ed. Guanabara Koogan, Rio de Janeiro. 780p.
10. Lehninger, A; Nelson, D.L.; Cox, M.M. (2018) *Princípios de bioquímica de Lehninger*. 7a ed. Artmed, Porto Alegre. 1312p.
11. Material didático da UFRGS de Biologia Molecular. Disponível em: <http://www.ufrgs.br/depbiot/discipl/linksHBF/MatDidat.htm>
12. Judith L. Gersting (1995). *Fundamentos matemáticos para Ciência da Computação*. 3ª edição. Disponível em: <https://www.cin.ufpe.br/~dmd/inf101/biblio/FMCCJK.pdf>
13. Mariano, Diego & Barroso, J.R.P.M. & Correia, Thiago & Melo-Minardi, Raquel. (2015). *Introdução à Programação para Bioinformática com Biopython*. Disponível em: <https://diegomariano.com/introducao-a-programacao-para-bioinformatica-com-biopython/>
14. Bioinformática: da Biologia à Flexibilidade Moleculares. Hugo Verli e colaboradores (2014). Disponível em: <https://www.ufrgs.br/bioinfo/ebook/>.

Referências

- [1] Gauthier, J., Vincent, A. T., Charette, S. J., & Derome, N. (2019). A brief history of bioinformatics. *Briefings in bioinformatics*, 20(6), 1981–1996.
- [2] Welch, L., Lewitter, F., Schwartz, R., Brooksbank, C., Radivojac, P., Gaeta, B., & Schneider, M. V. (2014). Bioinformatics curriculum guidelines: toward a definition of core competencies. *PLoS Comput Biol*, 10(3), e1003496.
- [3] Wu, H., & Palani, A. (2015, October). Bioinformatics curriculum development and skill sets for bioinformaticians. In *2015 IEEE Frontiers in Education Conference (FIE)* (pp. 1–7). IEEE.
- [4] Carvalho, L.M. *Como começar na Bioinformática?*. Teraytes of Life [Blogs de Ciência da Unicamp]. Campinas, 21 de julho. 2020. [acesso: 26 de 05 de 2021]. Disponível em: <https://www.blogs.unicamp.br/tb-of-life/2020/07/21/como-comecar-na-bioinformatica>.