2 Mapeamento de competências em bioinformática: panorama geral dos profissionais das áreas ômicas no Brasil

Autores 2.1

Letícia Marcorin

Revisão: Luana Luiza Bastos, Ana Carolina Silva Bulla

Resumo 2.1

O campo da bioinformática avançou muito desde seu início, na década de 1950, com *Margaret Dayhoff* que hoje é considerada a mãe da bioinformática, e atualmente ela é aplicada em diversas áreas de estudo da biologia. No entanto, o ensino da disciplina não acompanhou os avanços da área, especialmente em países em desenvolvimento, como o Brasil. Em uma pesquisa realizada no ano de 2021 por uma aluna do curso de Análise e Desenvolvimento de Sistemas da FATEC de Ribeirão Preto, que coletou informações de bioinformatas das áreas ômicas, foi possível constatar que os cursos de formação das áreas biológicas e de saúde não estão preparados para instruir os alunos em bioinformática. Somente 27,5% dos participantes declararam ter disciplinas em bioinformática oferecidas em seus cursos de formação, geralmente na modalidade não obrigatória. Há uma crescente demanda por bioinformatas e, por isso, é importante entendermos como se dá a formação desses profissionais para adequação dos cursos de formação que abastecem esse mercado. Nesse artigo será descrito um panorama geral da formação dos profissionais brasileiros que atuam nas áreas ômicas.

2.1 Um breve histórico da bioinformática

Apesar do termo ter se popularizado com os avanços da área genômica, a bioinformática surgiu com o estudo da estrutura de proteínas, desenvolvido durante a década de 1950, por Margaret Dayhoff, considerada a mãe da bioinformática [1][2].

Dayhoff foi pioneira na aplicação de métodos computacionais em biologia e, na virada dos anos 1960, em colaboração com *Robert S. Ledley*, desenvolveu o COMPROTEIN [3], o primeiro programa de computador capaz de elucidar a sequência de aminoácidos de uma proteína. Esse programa, escrito em FORTRAN

e cartões perfurados, pode ser considerado o primeiro *software* de reconstrução de sequências de novo. [4].

Somente em 1977 foi desenvolvida a primeira técnica de sequenciamento de DNA (ácido desoxirribonucleico, do inglês *deoxyribonucleic acid*) pelo laboratório de *Frederick Sanger* [5]. Apesar de ter passado por adaptações, essa metodologia é utilizada até hoje. O primeiro programa para análise computacional de dados de sequenciamento Sanger foi desenvolvido dois anos depois, por *Roger Staden* [6], e ainda está disponível para uso em http://staden.sourceforge.net/.

Graças aos avanços na área da computação, que acompanharam a demanda por máquinas mais baratas e *user-friendly*, várias universidades e grupos de pesquisa adquiriram seus próprios computadores. Assim, na década de 1980 surgiram diversos pacotes de programas, de distribuição livre, para análise de dados de sequenciamento de DNA, RNA (ácido ribonucleico, do inglês *ribonucleic acid*) e proteínas [1].

Finalmente, os avanços nas áreas de biologia molecular e tecnologia da computação levaram ao desenvolvimento de tecnologias de sequenciamento em larga escala, a partir de 1990. Em 1991 foi iniciado o Projeto Genoma Humano [7][8], que sequenciou o primeiro genoma humano completo, em 2003, com um investimento aproximado de 2,7 bilhões de dólares. Atualmente é possível sequenciar um genoma humano completo por aproximadamente mil dólares em menos de uma semana [9][10]. Assim, não surpreende que o volume de dados nas áreas ômicas (*e.g.*, genômica, epigenômica, transcriptômica, proteômica e metabolômica) tenha atingido a casa dos exabytes [1][9], sendo indispensável o uso de ferramentas de bioinformática para sua análise.

2.2 O ensino da bioinformática nos cursos de nível superior

A bioinformática foi inicialmente definida por duas publicações de *Paulien Hogeweg* e *Ben Hesper*, como o "estudo de processos informativos em sistemas biológicos" [11][12]. Os pesquisadores não descrevem a bioinformática somente

como um pacote de serviços providos pela área da computação às áreas biológicas, mas sim como a ciência que visa compreender o funcionamento de sistemas biológicos em todos os seus níveis (molecular, bioquímico e comportamental) por meio da análise de padrões e modelagem [13]. Assim, a crescente implementação de métodos computacionais nas áreas biológicas e da saúde demanda profissionais capacitados para o uso desses métodos: os bioinformatas.

Entre 2014 e 2018, a Sociedade Internacional de Biologia Computacional (ISCB, do inglês *International Society of Computational Biology*) conduziu estudos sobre as competências esperadas de um bioinformata em centros de bioinformática localizados na América do Norte, Europa e Ásia [14]. Os diretores desses centros relatam que a contratação de profissionais bacharelados (sem pós-graduação) é menos comum, pois esses profissionais não possuem um conhecimento interdisciplinar em biologia, estatística e computação. A partir desses estudos, a ISCB também descreveu 16 competências cujos níveis de experiência determinam três categorias de bioinformatas: usuário, cientista e engenheiro (Figura 2.1) [14][15].

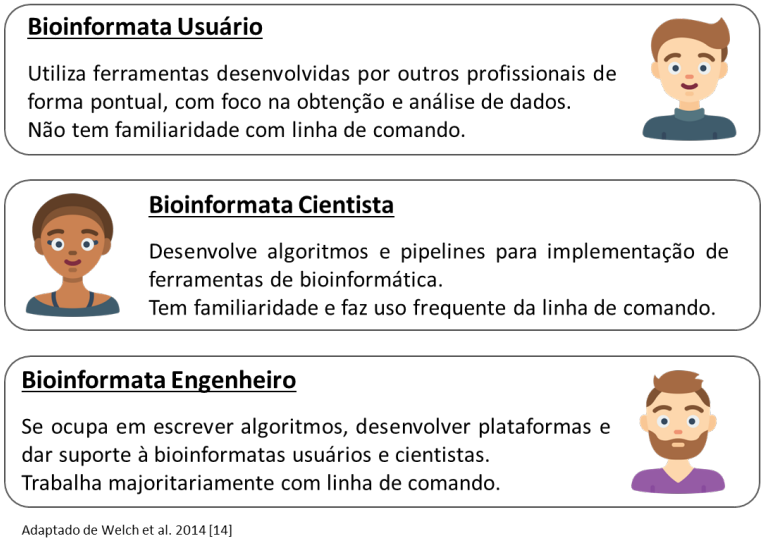

Figura 2.1: Descrição breve dos perfis de bioinformatas usuários, cientistas e engenheiros segundo estudo conduzido pela ISCB. Fonte: adaptado de [14].

Apesar de terem papéis diferentes, todos esses profissionais devem ser capazes de, pelo menos, compreender o funcionamento dos algoritmos para interpretar de forma correta os resultados obtidos a partir deles [14][16]. O estudo foi, posteriormente, utilizado para reformular e criar novos cursos de formação em nível superior, pós-graduação e treinamentos em instituições de ensino na África, Estados Unidos, Europa, Reino Unido e Austrália [15].

No Brasil, estudos como esse são escassos, focando em práticas pontuais de ensino em bioinformática. Em 2018, Mota publicou em sua monografia os resultados da inserção da bioinformática, na disciplina de Genética Molecular do curso de Ciências Biológicas da Universidade Federal do Sergipe [17]. O autor relata que os alunos reconhecem a importância da bioinformática para o entendimento mais compreensivo da genética, além de reportarem a necessidade desse tipo de disciplina para que se preparem melhor para o mercado de trabalho.

Resultados semelhantes foram obtidos com a aplicação do Curso de Introdução à Bioinformática inserido na programação da IX Semana da Biologia na Universidade Estadual do Ceará, no ano de 2010 [18]. Posteriormente, esse estudo foi ampliado e aplicado nas disciplinas de Biologia Molecular e Genética Geral do curso de Ciências Biológicas da Universidade Estadual do Ceará. Nesse estudo, os alunos demonstraram ter conhecimento dos termos básicos, porém desconhecem a maioria das técnicas de bioinformática [19].

Outro estudo, desenvolvido por pesquisadores do programa de Pós-Graduação em Genética e Melhoramento de Plantas da Universidade Federal de Goiás, também avaliou a aplicação de cursos de bioinformática no curso de ciências biológicas da universidade [20]. Os alunos participaram de aulas teóricas e, opcionalmente, aulas práticas utilizando a Plataforma Rosalind (http://rosalind.info/problems/locations/). Os autores viram um desempenho melhor nas avaliações dos alunos que fizeram ambos os cursos teóricos e práticos.

Os estudos citados anteriormente mostram uma defasagem na formação dos alunos matriculados nos cursos de ciências biológicas quanto à aplicação das técnicas de bioinformática que, provavelmente, se reflete no restante das

instituições de ensino. No entanto, é necessário ter um panorama mais compreensivo desses alunos e profissionais em âmbito nacional.

2.3 Panorama geral da formação dos profissionais das áreas ômicas no Brasil

Em 2021, um trabalho do curso de Análise e Desenvolvimento de Sistemas da FATEC de Ribeirão Preto, conduziu um estudo sobre a formação dos profissionais que atuam nas áreas ômicas (genômica, transcriptômica, proteômica, epigenômica e metabolômica) no Brasil [21]. Para isso foi distribuído um formulário para preenchimento online, contendo perguntas fechadas sobre a formação desses profissionais, desde os cursos de graduação à especializações e cursos complementares. O formulário também questionou a facilidade do uso das ferramentas de bioinformática pelos respondentes quando iniciaram suas atividades nesse campo de atuação, e como foi o seu progresso até o momento do preenchimento do formulário. Por fim, os participantes também foram classificados segundo as categorias de bioinformatas definidas pela ISCB. O formulário foi preenchido de forma anônima pelos participantes.

No total foram coletadas respostas de 80 participantes de 36 Instituições de Ensino (IE) das cinco regiões do país, sendo que 75% delas vieram de profissionais da região sudeste. Quanto à formação básica, 55% dos participantes têm formação em cursos das áreas de Ciências Biológicas, cerca de 26% é formado em cursos de Ciências da Saúde e o restante têm formação em Ciências Exatas (Figura 2.2A). Mais de 90% declararam ter cursado ou estar cursando mestrado ou doutorado (Figura 2.2B), o que já dá um indício de que a formação básica oferecida pelos cursos de graduação não é suficiente para a atuação nas áreas ômicas, muito provavelmente por falta de instrução em bioinformática.

Quanto à importância da bioinformática para o desenvolvimento das ciências biológicas, todos os voluntários, independente da formação, concordam que a bioinformática é essencial. No entanto, somente 27,5% dos voluntários tiveram disciplinas de bioinformática oferecidas em seus cursos de graduação. Dentre

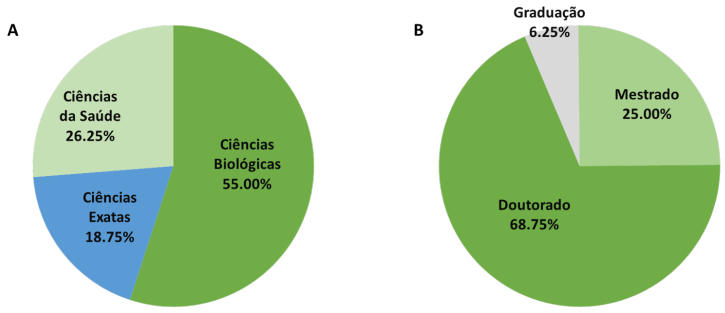

Figura 2.2: Distribuição dos participantes que preencheram o formulário conforme a formação inicial (A) e pós-graduação (B). Fonte: próprio autor.

esses casos, 95,5% dos alunos optaram por cursá-las, mesmo quando não obrigatórias. Dentre os que que não tiveram disciplinas em bioinformática oferecidas em seus cursos, 15,3% cursaram em outros programas.

Adicionalmente, 97,5% dos voluntários concordam que os cursos das áreas de ciências biológicas devem oferecer disciplinas em bioinformática, e 87,5% concordam que devem também oferecer disciplinas de programação básica (Figura 2.3). Esses números mostram que há uma clara demanda por ensino de bioinformática já na graduação. Quanto ao uso de ferramentas de bioinformática, 75% declarou que não seria possível desenvolver seus trabalhos sem o uso dessas ferramentas, e somente 2,5% declarou não utilizar qualquer ferramenta.

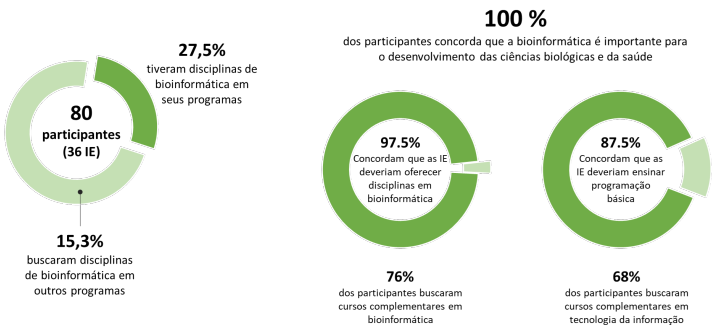

Figura 2.3: Infográfico mostrando que os profissionais participantes da pesquisa concordam que a bioinformática é importante para o avanço das ciências biológicas e da saúde, e que há uma grande demanda de ensino dessa ciência já na graduação. Abreviações: IE = Instituição de Ensino. Fonte: próprio autor.

Os voluntários formados nas áreas de Ciências Exatas indicaram menor dificuldade com o uso das ferramentas. Eles também parecem aprender ou se adaptar ao uso desses recursos com mais facilidade, sendo o único grupo a não declarar grandes dificuldades com o uso da bioinformática depois de algum tempo desenvolvendo seus projetos. Esses voluntários também são os que mais declararam já ter desenvolvido seus próprios *scripts* e *pipelines* (80%), sendo que os demais, das áreas de ciências biológicas e da saúde, apresentam uma média de 54,7% de profissionais com esse nível de conforto com as ferramentas.

Os participantes também foram categorizados nas três categorias de bioinformatas (usuário, cientista e engenheiro) com base nas suas classificações em cada uma das 16 competências definidas pela ISCB. Do total de respondentes, 11 (13,75%) não obtiveram classificação mínima e, segundo os critérios da ISCB, não são considerados bioinformatas. Desses, 91% têm formação em áreas biológicas e de saúde. Dentre os que atingiram pontuação suficiente, e somente 15% foram classificados como engenheiros, 25% foram classificados como cientistas e 46,3% como usuários. Na Figura 4 é possível notar que dentre os participantes com formação nas áreas exatas, cerca de 50% têm proficiência em bioinformática, sendo classificados como cientistas e engenheiros. Dentre os com formação em ciências biológicas e da saúde, essa taxa é menor, e a proporção de participantes que não tem qualquer proficiência em bioinformática é pelo menos o dobro da vista no grupo dos formados em áreas exatas.

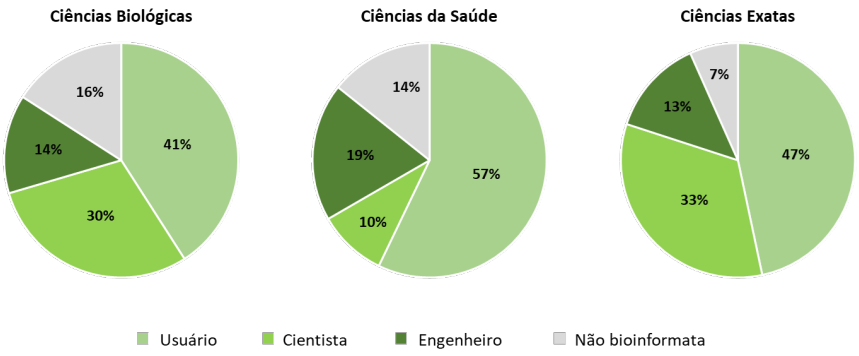

Figura 2.4: Distribuição das classes de bioinformata dentre os participantes com formação em ciências biológicas, da saúde e exatas. Fonte: próprio autor.

Perspectivas futuras: como você pode ajudar a mudar esse cenário A partir dos dados obtidos nesse estudo, que foi apresentado como Trabalho de Conclusão de Curso (TCC) na FATEC de Ribeirão Preto, é possível notar que há uma demanda por formação mais compreensiva em bioinformática já na graduação, principalmente nos cursos das áreas biológicas e de saúde, que formam a grande maioria dos futuros bioinformatas do país.

O objetivo deste estudo foi mostrar a defasagem na formação dos profissionais pelos cursos de graduação disponíveis atualmente nas universidades brasileiras e impulsionar a melhoria dessas grades de ensino. No entanto, o número de respostas coletadas ainda é pequeno para que o estudo atinja esse objetivo. Por isso, convido todos os bioinformatas a participarem dessa pesquisa para termos uma visão mais compreensiva da formação desses profissionais no nosso país. As respostas são anônimas.

Bloco 2.1

Este artigo está disponível em https://bioinfo.com.br/mapeamento-de-competencias-em-bioinformatica-panorama-geral-dos-profissionais-das-areas-omicas-no-brasil/

2.4 Referências

[1] Gauthier, J., Vincent, A. T., Charette, S. J. & Derome, N. A brief history of bioinformatics. Brief. Bioinform. 20, 1981–1996 (2019).

[2] Hagen, J. B. The origins of bioinformatics. Nat. Rev. Genet. 1, 231–236 (2000).

[3] Dayhoff, M. O. & Ledley, R. S. Comprotein: a computer program to aid primary protein structure determination. in Fall Joint Computer Conference 4–6 (Proceedings of the December, 1962).

[4] Hersh, R. T., Eck, R. V. & Dayhoff, M. O. Atlas of Protein Sequence and Structure, 1966. Syst. Zool. 16, 262 (1967).

[5] Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. 74, 5463–5467 (1977).

[6] Staden, R. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 6, 2601–2610 (1979).

[7] Venter, J. C. et al. The Sequence of the Human Genome. Science (80-.). 291, 1304–1351 (2001).

[8] Lander, E. S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860–921 (2001).

[9] Kumar, K. R., Cowley, M. J. & Davis, R. L. Next-Generation Sequencing and Emerging Technologies. Semin. Thromb. Hemost. 45, 661–673 (2019).

[10] Wetterstrand, K. A. DNA Sequencing Costs. NHGRI Genome Sequencing Program (GSP) www.genome.gov/sequencingcostsdata (2020).

[11] Hogeweg, P. & Hesper, B. Interactive instruction on population interactions. Comput. Biol. Med. 8, 319–327 (1978).

[12] Hogeweg, P. Simulating the growth of cellular forms. Simulation 31, 90–96 (1978).

[13] Hogeweg, P. The Roots of Bioinformatics in Theoretical Biology. PLoS Comput. Biol. 7, e1002021 (2011).

[14] Welch, L. et al. Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. PLoS Comput. Biol. 10, e1003496 (2014).

[15] Mulder, N. et al. The development and application of bioinformatics core competencies to improve bioinformatics training and education. PLOS Comput. Biol. 14, e1005772 (2018).

[16] McClatchy, S., Bass, K. M., Gatti, D. M., Moylan, A. & Churchill, G. Nine quick tips for efficient bioinformatics curriculum development and training. PLOS Comput. Biol. 16, e1008007 (2020).

[17] Mota, E. S. BIOINFORMÁTICA NO ENSINO DE GENÉTICA PARA O CURSO DE GRADUAÇÃO DE CIÊNCIAS BIOLÓGICAS SOB METODOLOGIA ATIVA. CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE (Universidade Federal de Sergipe, 2018).

[18] Ribeiro Junior, H. L. et al. Abordagem Prática de Bioinformática em Evento Acadêmico-Científico na Cidade de Fortaleza-Ce. Rev. Bras. Ensino Ciência e Tecnol. 4, (2011).

[19] Ribeiro Junior, H. L., Oliveira, R. T. G. de & Ceccatto, V. M. Bioinformática como recurso pedagógico para o curso de ciências biológicas na Universidade Estadual do Ceará – UECE – Fortaleza, Estado do Ceará. Acta Sci. Educ. 34, (2012).

[20] Nunes, R., Barbosa de Almeida Júnior, E., Pessoa Pinto de Menezes, I. & Malafaia, G. Learning nucleic acids solving by bioinformatics problems. Biochem. Mol. Biol. Educ. 43, 377–383 (2015).

[21] Marcorin, L., Pereira, M. A. A. Mapeamento de competências em bioinformática: panorama geral dos profissionais das áreas ômicas no Brasil. V WorkTec – Workshop de Tecnologia da FATEC Ribeirão Preto. (2022).