YALE JOURNAL OF BIOLOGY AND MEDICINE 85 (2012), pp.323-329.
Copyright © 2012.

FOCUS: EDUCATING YOURSELF IN BIOINFORMATICS

Thriving in Multidisciplinary Research: Advice for New Bioinformatics Students

Raymond K. Auerbach

Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut

The sciences have seen a large increase in demand for students in bioinformatics and multidisciplinary fields in general. Many new educational programs have been created to satisfy this demand, but navigating these programs requires a non-traditional outlook and emphasizes working in teams of individuals with distinct yet complementary skill sets. Written from the perspective of a current bioinformatics student, this article seeks to offer advice to prospective and current students in bioinformatics regarding what to expect in their educational program, how multidisciplinary fields differ from more traditional paths, and decisions that they will face on the road to becoming successful, productive bioinformaticists.

INTRODUCTION

Since the turn of the century, we have been witnessing a shift in what factors limit research in the biological sciences [1,2]. While experiments were once conducted on a single gene or a single locus at a time, high-throughput techniques such as microarrays and next-generation sequencing now allow us to query entire genomes often in a fraction of the time and study biological systems on multiple scales [3-6]. These techniques have led to the so-called "data deluge" and have shifted scientific research to favor multidisciplinary teams working to explore complex systems by integrating different data types and pooling knowledge [7,8]. Although we have learned a lot from these approaches and will undoubtedly continue to do so, making the most significant contributions requires a different thought process than those associated with classical biology, statistics, mathematics, or

To whom all correspondence should be addressed: Raymond K. Auerbach, Program in Computational Biology and Bioinformatics, Yale University, 266 Whitney Ave., New Haven, CT 06520; Tele: 203-432-5405; Email: raymond.auerbach@yale.edu.

Keywords: bioinformatics, education, multidisciplinary education, multidisciplinary research, bioinformatics education, computational biology

computer science. Our generation of scientists needs to work at the interface of these different fields, or at least be able to converse effectively with others that hail from these backgrounds.

There are several different definitions of the term “bioinformatics.” To some, it is as simple as managing and analyzing data from a multitude of different experiments, whereas to others, it is the design of algorithms and tools to enable such analyses [9]. More generally, I consider bioinformatics to be the interface between statistics, computer science, and biology. The combination of these many different fields is what makes the field so exciting — and also so tough to navigate as a student. This perspective will share some of the insights that I have gleaned over the years. I remind readers that my opinions are exactly that: my opinion. Advice is free, and it is up to the individual to determine what to take from it; however, I do hope my thoughts and experiences will help those considering a career in a multidisciplinary field like bioinformatics as well as those who have already set foot down that path.

So why should you listen to me? Well, ultimately, only you can answer that question, but I will start with some general background. My path to studying bioinformatics has been somewhat unorthodox and has not always been my life’s focus. In fact, when I started college, I had never heard of a bioinformatics degree. I studied several areas that comprise bioinformatics separately, including receiving my BS in Computer Science and an MS in Biology from Northern Arizona University before coming to Yale to study for a PhD in Computational Biology and Bioinformatics. Looking back, this path was probably not the most orthodox but has led me to appreciate members of many different fields, how each are trained to think, and what they can bring to a team environment. As one example, computer scientists and statisticians are very quantitative, expecting an exact solution for most problems and that any problem can be modeled/simplified. Biologists tend to take a more qualitative viewpoint, preferring to learn by observation and realizing that some systems are too complex to simplify given our current knowledge. These viewpoints appear contradictory, but they can be very powerful when leveraged as part of multidisciplinary research. With my collaborators, I have published more than 22 peer-reviewed papers in that time period, learned a lot about many different areas, and have many memories that have shaped me as a bioinformaticist-in-training.

GENERAL CONSIDERATIONS FOR BUDDING BIOINFORMATICISTS

Multidisciplinary fields, bioinformatics among them, require its members to remember one central (and to some, discomforting) fact: You are an expert to everyone yet an expert to no one at the same time. For example, a bioinformaticist is typically trained in at least computer science, statistics, and biology. When working on a multidisciplinary team with members who are classically trained in these disciplines, a biologist will likely view you more as a statistician or a programmer, a programmer will in turn view you more as a statistician or a biologist, and a statistician will peg you as a biologist or a computer scientist. They are all correct. The logical question that follows, and one that I have asked myself many times over the years, is: Why not just have a team with a statistician, a computer scientist, and a biologist? What is the purpose of the bioinformaticist if he or she is not the unquestioned expert in any of these classical disciplines? The answer is that a good bioinformaticist is the glue that can hold a multidisciplinary team together. Although he or she may not be the best pure computer scientist, pure biologist, or pure statistician on the team, a bioinformaticist knows enough about each area to conduct most analyses and to translate for individuals from other disciplines. On some teams, you may need to take what a biologist tells you needs to be done and translate it into a requirements document that a computer scientist will be able to implement, including details that a biologist might take for granted and hence did not mention (e.g., the difference between forward strand vs. reverse strand for DNA sequence analysis). Later, you may be asked to explain the general algorithm implemented by the computer scientist to the biologist so that he or she can explain it in a talk. In this manner, a good bioinformaticist must be knowledgeable in many different, complementary fields simultaneously. This is what can make our field so complex and where bioinformatics education has a high bar to prepare students to meet these challenges.

AN EDUCATION IN BIOINFORMATICS: SO YOU HAVE DECIDED TO TAKE THE PLUNGE

The first question you must answer when undertaking an educational program in bioinformatics is how do you see yourself professionally? Are you a programmer who knows a little biology, a biologist who can write Perl scripts, or another combination entirely? Are you driven by the thrill of discovering the answer to a previously unexplained biological question or are you more excited by designing a new algorithm that can process twice the data in half the time as existing methods? The answer to this question will help determine where you want to focus your education and where you need to improve. If you do not know the exact answer yet, that is fine. The first one or two years of graduate school are about learning what else is out there and trying new things. However, you should have a general idea of what areas excite you before you begin.

Classes

Most bioinformatics programs, including Yale's offering, require a student to take a mix of biology, statistics, and computer science courses. For courses that are not vital to your central interests, learn enough about the material so you can have an intelligent conversation with a domain expert in the area. You will find these conversations necessary and plentiful both inside and outside of the academy. For those topics essential to reaching your career goals, master them. This will mean taking courses, of course, but as with any rapidly changing, technologically driven field, classes typically will only scratch the surface and lag behind what is relevant in practice. Read papers, attend talks, and find other ways to master the skills that interest you. Every bioinformaticist is different and comes with his or her own unique toolset, interests, and specialties. Classes and research while in graduate school are where you will begin to establish your particular mix.

Your Advisor and Your Research Lab

Other aspects of your graduate education that you can easily tailor, at least when you start, are your research lab and your advisor. There are several different paths that graduate students can take. The key to happiness in a graduate program, in my experience, is to find an advisor who is excited by similar things. In this way, meetings and lab discussions will become a joy rather than a chore, your morale will remain high, you will receive support for your endeavors and ideas, and you can focus on learning rather than managing the bureaucracy that unfortunately will always remain a part of graduate training. A happy graduate student is a more productive and will have fewer problems producing high quality work, giving talks, and graduating on time.

In short, how you fit into a lab and an advisor/advisee relationship will have a large effect on your happiness and productivity over the next 5-plus years and, in some cases, may even determine whether you will complete your degree. Do your due diligence and learn the initial ground rules at the start.

Are Two Advisors Better Than One? It Depends

Sometimes finding an advisor who shares all of your passions is not possible, particularly if you find yourself more at the interface of different sub-disciplines that comprise bioinformatics. For example, there are very few individuals who are experts in both computer science and biology. For these students, a joint advising situation should be considered. There are positives and negatives to these arrangements, though. On the positive side, a student who is jointly advised can benefit from complementary areas of expertise. Having advisors in both computer science and biology, for example, will enable a student to work in environments that stress each area and learn skills in both as well as have access to necessary resources (e.g., a computer cluster, data from a new method in your advisor's lab, etc.). The downside is that different advisors also come with different expectations for graduation. A classical biologist might expect a dissertation project to look more like a PhD project in biology, whereas a classical computer scientist will be looking for a project from traditional computer science. These differences can result in delayed graduation and stress for the graduate student, as the student must navigate and satisfy two different sets of expectations simultaneously.

If you decide to go the co-advisor route, it can be very fulfilling, lead to a varied and desirable combination of skills down the road, and enable you to do great things both as a student and after you graduate; however, and I cannot stress this enough, there are considerations that one must take before going down this road. First, make sure that the two advisors that you choose are compatible with each other. Sometimes the whole is less than the sum of its parts and two advisors who are not on the same page will make one life miserable: yours. Second, understand the requirements and expectations that each has of their joint graduate students *before you join their labs*. Discuss possible opportunities for collaboration, prior experience with jointly advising students, and expectations for graduation. Some advisors will only "officially" count papers on which they are the senior author toward your graduation, particularly if a professor is still coming up for tenure. Learn the ground rules of your particular co-advising scenario before you commit and make an informed decision. These requirements will shift during your tenure in a graduate program as your advisor learns more about you and your capabilities, but at least you will learn the initial set of expectations. Another resource that should never be underestimated is talking with other graduate students who are jointly advised. Often, these students will think of questions and scenarios that you have not considered and will offer invaluable advice.

THE ART OF PUBLISHING PAPERS IN A MULTIDISCIPLINARY FIELD

The Typical Models

Now that you have determined your focus and (hopefully) have found a lab/advisor(s) who supports your work, another area of interest to students is how to publish papers as a graduate student in an interdisciplinary field. I have been very fortunate in that I have worked with many great collaborators over the past 10 years and have published a fair number of papers. Publishing papers in bioinformatics can be problematic, as the reward system in biology is still structured toward papers with a sole primary author. For graduation as well as for your CV, often only primary authored publications are weighted heavily; however, primary author bioinformatics papers are also tricky to manage. If a paper features a novel, exciting data set that answers a burning biological question, then typically the biologist will (and in most cases should) get the first position on the author list. There are several things that a bioinformaticist can do to try to manage this. First, a bioinformaticist can concern himself with tool development. In essence, the bioinformaticist designs a novel tool or analysis and applies the new method to interesting biological data. Two different papers can arise from such a project. The first focuses on the biological question and usually features the biologist as the primary author, while the second covers the new algorithm/analysis method with the bioinformaticist taking the lead. Most collaborative projects begin with this two-paper structure in mind. This arrangement at first sounds ideal; however, from experience, these types of projects rarely come to fruition as originally envisioned. Inevitably, either the informatics paper is rolled into the first paper to increase its profile and allow for submission to a better journal or as soon as the biological paper is complete, your focus will need to shift to the next great analysis and the methods paper never gets written.

Another option, and the one that I personally prefer, is to write papers using starred primary authors (e.g., "these authors contributed equally to this work") when multiple authors drive a research project. In many cases, this will be true, as new biological approaches typically require new analysis methods. Before the paper is written, and if you have a good collaborator, one should discuss the authorship plan for a manuscript. If the bulk of the work is analysis, then the bioinformaticist has a claim for starred authorship listed first. Otherwise, starred authorship listed second is perfectly fine. Although PubMed does not track joint authorship, most people hiring in interdisciplinary fields are aware of the challenges of publishing in these areas and will recognize starred authorship papers. Starred authorship also seems to be gaining acceptance among the more classical disciplines.

Although you need primary author papers to graduate, your focus should not start on these types of papers alone. Middle authorship papers are also very valuable for career development, as they illustrate your ability to contribute in a team setting, give you access to resources, and allow you to expand your analytical repertoire sans the pressures that come with primary authorship. Middle author papers are fine and even desirable for a graduate student in his or her first three years, as they allow the student to build a rapport with collaborators, prove their analytical capabilities, and build their publication lists. By year four, however, the student's focus needs to shift to primary author work because these are the papers that will lead to a degree and a shiny new job.

The above discussion has focused on publication models involving both biological and analytical components, but how can one approach the writing process if he or she falls more toward the statistics/computer science region of the spectrum and are not working directly with experimentalists? Possible scenarios include medical informatics students whose research focuses on designing the infrastructure behind electronic health records or biostatistics students who primarily re-analyze biological data sets that are publicly available, among many other examples. Although the above authorship issues usually will not be applicable to you, you should keep the advice in mind to foster successful collaborations with other disciplines.

How to Publish a Paper in Bioinformatics with an Eye Toward Graduation

The first question when designing your own research project is to determine the audience that you want to target. If the result of your research will be a tool or algorithm, then a journal primarily concerned with tools or algorithms (e.g., *Bioinformatics*) will be a better choice than a journal that focuses more on novel biological findings (e.g., *Genome Research*). For the "big three" journals — *Nature*, *Science*, and *Cell* — a paper will need a strong, novel biological component to be competitive. Determine the "hook" that you expect your project to have and the audience that would be the most positive toward your work. Submit your manuscript to a journal that this audience reads and you will avoid a lot of unnecessary hassle in getting your manuscript accepted. Another popular strategy utilized by many professors is to "journal shop" by first submitting a paper to one of the "big three" journals and then apply to a lower tier after each rejection. While this can result in your work being published in a higher profile journal, it also takes a lot of time. Every review will take at least a month, and during this time, your findings can become stale or competitors can catch up and submit a manuscript of their own. One must balance the time and shelf life of one's research against the probability that a higher profile journal will be interested in publishing your work. Aiming high is fine and, in fact, encouraged, but if you have written a manuscript that is only of marginal interest to a general audience and lacks novelty, then submitting to *Nature* is probably not going to be worth your time. Be positive but realistic. The goal of a graduate student who is focused on graduating is to produce one or two strong first author publications. Once you get these, then you can spend any remaining time expanding upon your work and/or laying the foundation for your future Nobel Prize.

Considerations for Publishing Tools, Methods, and Algorithms

Finally, I would also like to give a word of warning to those budding bioinformaticists who are primarily interested in tool design. One of my rotation supervisors had a saying that "pipelines are the enemy of graduation." Designing an informatics pipeline is actually very useful as well as a popular way to publish a true informatics paper, but they can also easily become traps. Remember that any tool, once a paper is published, will be available for use by the general public, and the general public likes to ask questions, make maintenance requests, and expects fast service. If your first major paper is lead authorship on an informatics pipeline, these questions will be filling your mailbox. Make sure that you keep your eye on your goals at all times. Although it is necessary to support your software, graduate students can easily become so involved with maintaining their tool that they do not continue their research to create the next great tool and produce their next manuscript. Remember to keep a balance, and if necessary, discuss the maintenance plan with your advisor after your paper is accepted.

Meetings and Conferences: How to Use Presentation Opportunities to Your Advantage

The last item I will discuss in this perspective involves presentations at meetings and conferences, whether they take the form of a poster or a talk. Personally, I came to Yale abhorring poster presentations, but I have come to recognize their value to a student in multidisciplinary fields. Poster presentations and talks are tough for the same reasons that working as part of a multidisciplinary team is challenging: Everyone has a different expertise/emphasis, and it is your goal to be able to explain the significance of your work to everyone. There have been several books and papers written on how to present your work, so I will avoid most of the common-sense advice that you can find elsewhere. For a bioinformaticist, though, conferences and meetings represent great opportunities to practice the different sales pitches for your research.

As an anecdote that illustrates my above points, I presented a poster at the National Library of Medicine Biomedical Informatics training conference a couple years ago. This environment is generally very friendly and features attendees with interests as varied as biology, text mining, statistics, and theoretical computer science. There were several different methods that I used to pitch my research, and I would always try to find out about the background of my listeners before launching into a pitch. For those with backgrounds in biology, I would focus on the biological contribution of the work, while those hailing from computer science would hear my pitch on how the analytical software was designed, etc., and I would gauge their reactions. Aside from general presentation practice, these opportunities are without parallel for learning the strong and weak points of your work in the eyes of an outside observer and determining which pitch will be the strongest to a journal editor/potential reviewer. As I stressed early in this discussion, bioinformatics represents the glue that holds interdisciplinary teams together and being able to talk with more classical domain experts is a vital part of the job description. Meetings and conferences are the best places to see how far you have come and how far you have yet to travel in meeting this requirement. Relish these opportunities.

CONCLUSION AND OUTLOOK

In closing, I would like to remind the reader that the above opinions are exactly that: my opinions. Everyone must find his or her own approach, and this is what makes bioinformatics such an exciting field. The "prototypical bioinformaticist" does not exist, so do not concern yourself with trying to conform to a model. Determine your passions, pursue them, and remember that everyone on the multidisciplinary team has something invaluable to contribute. This field changes very quickly. Analysis methods will come and go, translational and clinical applications will start to become more prevalent, and there will always be "the next big thing" to learn, but this is what keeps the field so fresh and interesting to us. Remember every so often during those late nights at the bench, at the computer terminal, or in a study group to step back and take stock. We are going to determine the direction of our fledgling field as it continues to mature through paradigm shifts and technology changes. Be responsible stewards for the next generation and strive toward improving the human condition, but never lose sight of why you chose this path. Enjoy the ride!

Acknowledgments: The author would like to thank Dr. Mark Gerstein of Yale University, Dr. Michael Snyder of Stanford University, and Drs. Paul Keim, David Wagner, and Bernard Carey of Northern Arizona University for their advice and support throughout my education. Additionally, the author has been funded by grants from the National Library of Medicine Biomedical Informatics Training Program and the National Institutes of Health.

REFERENCES

1. Frankel F, Reid R. Big data: Distilling meaning from data. Nature. 2008;455(7209):30.
2. Community cleverness required. Nature. 2008;455(7209):1.
3. Schulze A, Downward J. Navigating gene expression using microarrays - a technology review. Nat Cell Biol. 2001;3(8):E190-5.
4. Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. Nat Rev Genet. 2010;11(7):476-86.
5. Shendure J, Ji H. Next-generation DNA sequencing. Nat Biotechnol. 2008;26(10):1135-45.
6. Kohane IS, Butte AJ, Kho A. Microarrays for an Integrative Genomics. Cambridge, MA: MIT Press; 2002.
7. Eddy SR. "Antedisciplinary" Science. PLoS Comput Biol. 2005;1(1):e6.
8. Li J-W, Schmieder R, Ward RM, Delenick J, Olivares EC, Mittelman D. SEQanswers: An open access community for collaboratively decoding genomes. Bioinformatics. 2012; 28(9):1272-3. Epub 2012 Mar 13.
9. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med. 2001;40(4):346-58.