Happy Belly Bioinformatics: an open-source resource dedicated to helping biologists utilize bioinformatics

Michael D. Lee[1]

1 Exobiology Branch, NASA Ames Research Center, Moffett Field, CA, USA

Summary

Our civilization's ability to generate data has far surpassed our current, collective capabilities to efficiently access and utilize it. And while the necessary educational infrastructure is trying to adapt and grow as quickly as possible, many graduate students, postdocs, faculty, and other researchers have found ourselves in situations where self-education with little direction is the only option available. This can make it difficult to get through the initial steep learning curve of beginning to use bioinformatics effectively – particularly when this is expected to happen alongside all of our other research and educational responsibilities.

There are many resources available online for learning programming. For example, Codecademy offers free, online lessons focused on learning several different languages from scratch. And more geared towards the sciences, The Carpentries is an international organization largely powered by volunteers designing and maintaining lessons for in-person workshops that are then openly available. However, the plethora of things out there to possibly learn is also overwhelming when starting out, as it can be difficult to know what will be important to us and therefore where to begin. And most available resources are lacking an emphasis or any discussion on the philosophy of being a biologist trying to learn bioinformatics – which I've found to be integral to the process.

Statement of need: Happy Belly Bioinformatics is a resource designed and maintained specifically around the goal of helping biologists be able to learn and apply bioinformatics as needed in our research. It has been directly born out of my personal experience of, like many of us these days, being a biologist thrust into the computational world with little or no guidance/advice/resources in my surrounding lab/department/environment – it is the resource I wish I had. The underlying educational philosophy is that learning a small set of baseline, fundamental skills and concepts that will directly aid in our research is what is important – rather than, for example, sitting down and trying to master a specific programming language without any direct application to our work.

While Happy Belly Bioinformatics is suitable for use in group-educational activities such as workshop-type environments, part of what makes it stand out from other resources is that each lesson contains all of the information an in-person instructor might convey, written out in a personal and conversational form. This often includes personal anecdotes and components/realizations that I found important in the building and shaping of my own mental landscape while navigating the skills and topics being covered. This interspersed conversational text conveys much of the philosophy that is lacking from other resources (e.g. some main points are summarized in Five things I wish I knew when I started getting into bioinformatics). Currently covered areas span foundational skills such as a crash course in working at a Unix-like command line – with 5 separate modules ending with 6 commands worth knowing about and a primer on variables and for loops –

and working in R (e.g. an introduction to R and a more detailed look at indexing in R). Additionally, and also what makes Happy Belly Bioinformatics differ from most other resources, there are highly comprehensive walkthroughs of common applications in current microbial ecology, e.g.: an amplicon/marker-gene tutorial that covers a full workflow from processing through to examples of many standard, initial analyses; walking through a *de novo* genome assembly and assessment; and recovering **M**etagenome-**A**ssembled **G**enomes (MAGs) from metagenomes.

Happy Belly Bioinformatics helps researchers bridge the gap from having little or no current experience to being comfortable with a baseline set of fundamental skills through comprehensive, heavily annotated, and standalone tutorials. This equips us as researchers to then develop our sphere of capabilities in any and all directions as needed, without there any longer being a technical bottleneck on our science.