The impact of Bioinformatics

Computational approaches often go as unrecognized CO_2 producers

95kg CO2 equivalents[1] to simulate the Tobacco Mosaic Virus for 100ns

9 Billion hours[2] of computation in 2020 alone

The Impact

Let's at first throw some numbers around. While an RNA read alignment might "just" produce approximately 0.5 kg of CO2 equivalents, a Genome assembly already causes ≈12kg. Still, these numbers depend on which tools are used, for example, a Genome-wide association studies (GWAS) can release between 4.7kg to 17kg of CO2 equivalents[1]. And finally, some phylogenetic analyses might cause about 3000kg of CO2 equivalents to be produced. To leave biology, the biggest astronomical observatory in France produce about 4,2 Million tons of CO2 equivalents every year just measuring research purposes[3].

The Quantification

Until recently, not many numbers were available to estimate the environmental impact of bioinformatics. It really came down to a few researchers to publish their approximations.

It is not easy to calcaute these footprints. As we have seen, the tools you use cause big variances. However, here a few other factors that influence the amount of CO2 equivalents produced:

- What hardware do you use? (e.g. CPUs vs GPUs)
- What amount of memory is required for your process?
- The runninng time of your experiment

If you want to assess your own footprint, you can use some tools such as the Green Algorithms calculator or if you have multiple jobs run in parallel you can check out an high performance computing calculator. All of that is also nicely discussed right here[4] as well.

What else to consider

One process but many factors that play a role

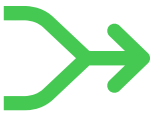
The Location Matters

Since cloud computing is big these days, one might assume that it really doesn't matter where you do your research. However, this remains a false belief. You location can be decisive because where the facility or data center is matters. The question is whether their electricity is derived from coal or, for example, hydroelectric plants. Indeed, the carbon footprint one leaves behind can be up to 11-fold higher depending on how where you do your studies[3].

Furthermore, do not forget that bigger data centers might be able to run more efficiently than smaller ones. Not to say that upscaling is solving the problem but how the data center is run could reduce its footprint by about 34%[1] (or optimally more).

Finally, there are such considerations as at what time you run your experiments. Are servers already busy or is vacancy left? You could use time slots overnight if there are no jobs processed otherwise. Finally, the "conditions" of your location matter. What temperatures are outside? It should not come as a surprise that cooling requires energy as well.

More Than Just 0s And 1s



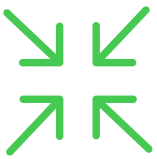
There is something else you should keep in mind. Running your experiment is only one source of environmental footprint. Another one is their production. Just consider how many miles their pieces travel around the world before being put together and sent to you. Especially for small devices, such as your smartphone, these impacts can be much bigger than the electricity they use. Furthermore, all the metals and rare earths are certainly not easy to isolate. No wonder that some companies want to mine in space soon …

Something for everyone

Computation is relevant to everyone

Tips for You

Isn't the environmental impact of computational applications not just important for these dry-lab people that drink coffee all day and hang around this git-hub website?



At first, do not pretend you would not drink coffee all day ;) But there are few things we all can take away from talking about this topic:

We can spread the word to our dear collaborators and raise awareness about this issue!

Can you optimize running times? Can you run parts of your code to make it failproof before running it in its entirety? Are you aware of the service/center you use? And what tools do you make use of?

Of course, it might often not be feasible to change software in the middle of a project. However, when people know that this is an important factor, they might check first for their next project.

B) We all store our data in clouds. Whether it is your Email, your microscopy pictures, lab book or favorite papers. However, that could take up to 1 000 000x times more energy than keeping them on a hard drive (yes, these are many zeros!). Some estimates claim that about 1.5% of the entire carbon footprint worldwide is derived from cloud storage

C) A single google search might take as much energy as necessary to light a bulb for around 25 seconds. And the new AI language models like ChatGPT take significantly more electricity. Just the training of GPT3 might have required 700,000 liters of freshwater to cool its operations. Thus, use these tools carefully!

1) The Carbon Footprint of Bioinformatics - 10.1093/molbev/msac034
2) Green Algorithms: Quantifying the Carbon Footprint of Computation - 10.1002/advs.202100707
3) A comprehensive assessment of the carbon footprint of an astronomical institute - 10.1038/s41550-022-01771-3
4) Carbon footprint estimation for computational research - 10.1038/s43586-023-00202-5

READVANCE

www.ReAdvance.com

We are an initiative **run by researchers and students**. It is the passion for a greener future that keeps us going. We want to **share the most exciting and best solutions** with you!
Join our family or follow us on social media to find out more!

I like sustainability in the lab!

Join the green family

Let me find out more!

The green family offers for you

- Many more ideas & protocols
- An international network
- Sustainability in the industry