Tracing Transmission (Science During A Pandemic #3), with Dr. Samantha Lycett
Not Another Science Podcast
August 12th, 2020

Pre-episode disclaimer

[Tom Edwick] Hey guys! Welcome back to the wardrobe. Before we get started on today's episode I just want to say thank you so much for tuning in to the coronavirus miniseries. We've put a lot of work into getting this podcast off the ground and learned a lot in the process. Personally I've learned that not only are wardrobes an excellent clothing storage solution, but also the ideal home recording studio. I also want to extend a massive thank you to the editor of this podcast, Helena. She's way more organised than me, has so many great ideas, and has kept the momentum going when I've not been feeling up to it. Also her editing skills are unrivalled - I'm a ridiculously noisy and heavy breather, but you'd never know listening to the show. I also want to thank Karolina, the president of Edinburgh University Science Media, for giving me this opportunity. When I inquired about the podcast, I did not expect to actually be hosting it. It was safe to say I was terrified, but I've really enjoyed the challenge, and I'm excited for the episodes that we've got in the pipeline for you guys.

Last but not least, we want your feedback! We've created a Google form where you can anonymously roast the podcast for all it's worth - though we'd appreciate some nice stuff too… It only takes 30 seconds to fill in and we would be eternally grateful,

Audio extract from Toy Story (1995) of the aliens saying "We are eternally grateful".

… plus its a chance to let us know if you have any ideas or suggestions for future episodes. Okay - let's get on with the show.

General Introduction

Intro music starts

Hello and welcome to *Not Another Science Podcast?!*, I'm your host, Tom Edwick.

In each episode we explore fascinating themes and ideas, talk to awesome researchers about their work, and find out about the science being done by our very own staff and students here at the University of Edinburgh.

If you'd like to get in touch with a question, suggestion, or if you want to be featured on the podcast, you can reach us on our Facebook page, Edinburgh University Science Media, or at our twitter, @eusci, that's @e-u-s-c-i. You can also drop us an email at eusci.podcast@gmail.com

Specific Introduction

Today we bring to you the fourth and final part of our miniseries on coronavirus, in which we're exploring what it's like to do science during a pandemic, and diving into the incredible response from scientists and students from the University of Edinburgh.

Today, we're speaking with Dr. Sam Lycett, who is a specialist in bioinformatics and computational biology, with a focus on pathogen phylodynamics. Okay, I know what you're thinking - Tom slow down, you just said a lot of big and scary words there, what the heck are you talking about? But don't worry guys, **all will become clear**. Basically she's like a biological detective, using big computers to look for small clues in the DNA of infectious, disease-causing bacteria and viruses, that allows her to track how they spread through the population. She has been working on SARS-CoV-2 since the outbreak, and she told us how you actually go about tracking a virus (hint: it takes a lot of data).

In the last episode we spoke to Jamie Davies, who is a professor of experimental anatomy at the university of

edinburgh, and runs the International Union of Pharmacology's drug database. We spoke about a whole new section of the database that he and his team created as a one-stop shop for potential COVID-19 treatments, and discussed how he thinks science can and should change as a result of the pandemic. To find all other episodes and show notes, head to our website at eusci.org.uk, or subscribe on your platform of choice.

Intro music ends

Main

> **[Dr. Samantha Lycett]** So I'm Sam Lycett, and I am a group leader in pathogen phylodynamics. I work in the division of Infection and Immunity at Roslin Institute.

That's Dr Sam Lycett. When we spoke, she had been swamped with Zoom calls, and as such was quickly becoming a master of the teleconference, including using the carribean Zoom background, which I'm ashamed to admit I didn't realise was fake for far too long.

> **[Dr. Lycett]** So you can see my Caribbean {…}. *laughs*
> **[Tom]** *laughs* Loving it! Have you been doing a lot of Zoom meetings and stuff then?
> **[Dr. Lycett]** Yeah, yeah, I mean you know, I was in one this afternoon, and yesterday, and probably pretty much every day.

Dr Lycett has been busy since the outbreak of SARS-CoV-2, the virus responsible for causing COVID-19. As a computational biologist and expert in pathogen phylodynamics, she was well placed to start studying coronavirus. But WTF is pathogen phylodynamics?

> **[Tom]** So how would you describe pathogen phylodynamics for someone who has no idea about biology, or genomes, DNA, stuff like that?
> **[Dr. Lycett]** So pathogen phylodynamics, it's a computational technique and it uses a combination of

> evolving pathogen sequence data, and uses other epidemiological data at the same time. So it’s a modelling technique.

A pathogen is simply a microorganism that causes disease. Many microorganisms, like the trillions of bacteria lining your gut, are harmless, and many are even beneficial. But not all microorganisms are welcome. You know, like that one time you ate that dodgy chicken kebab. You probably shouldn’t have, but it was late, you’d had a couple of drinks, and the takeaway looked really inviting. Now it's the morning after and you’re paying for it. Not literally, since you paid for it with your credit card the night before, but the vomiting and diarrhea certainly feel pretty literal right now. The cause of these excess bodily fluids? Food poisoning, which is caused by food that has been contaminated with *E. coli,* or *Salmonella,* which are both bacteria, or a virus like norovirus. COVID-19 is caused by a viral pathogen, SARS-CoV-2, or to use its full name, like a parent about to give their child the telling-off of a lifetime, Severe Acute Respiratory Syndrome Coronavirus 2.

So, let’s talk phylodynamics. Phylodynamics is a combination of phylogenetics and dynamics. Phylogenetics is the study of the evolutionary relationships between organisms, in this case between individual viruses, and this is done by looking at their DNA (hence the genetics part of phylogenetics). By comparing how similar the DNA is between two viruses, we can tell if they came from the same place or not, which ultimately allows us to model the transmission of said virus - this is where the dynamics part comes from.

> **[Tom]** How do researchers go about getting this data in the first place? Are they taking DNA samples from viruses and stuff?
> **[Dr. Lycett]** Yes, so in fact there’s several ways that you can get the right kind of DNA. The first thing to say is that there are DNA viruses and RNA viruses.

SARS-CoV-2 is an RNA virus, which just means that its genes are stored as RNA rather than DNA.

A quick genetics crash course before we go any further: DNA and RNA are both long molecules that consist of chains of what are called nucleic acids, and these nucleic acids come in 4 different flavours, or bases, which are adenine, guanine, cytosine, and thymine - biologists just refer to them as A, G, C, and T. Don't worry about remembering the names, all you need to know is that the order of these bases in the DNA or RNA chain, or sequence, is important. Cellular machines read the sequences like a ticker tape, and make things the cell needs based on the order of the bases. This is why you will have heard the term genetic code - we say that different DNA or RNA sequences code for different cellular molecules, or proteins as biologists call them. A stretch of DNA or RNA that codes for a protein is what we call a gene, and the collection of all of an organism's genes is what we call the genome.

RNA and DNA are very similar, and can both be used to store genetic information as genes. The difference is in their structure: DNA consists of two strands of nucleic acids coiled in the classic double-helix, whereas RNA is single-stranded. Most organisms store their genes as DNA, as the double-helix configuration makes it more stable, but some, including certain viruses, store their genes as RNA. Viruses are incredibly simple organisms, without the cellular machinery to make copies of themselves. Their solution? Invade a host's cells and use theirs.

> **[Dr. Lycett]** So viruses have genes, like I said some are DNA viruses and some are RNA viruses. Bacteria also; you can do pathogen phylodynamics on bacteria. Bacteria have DNA. But yeah, what you might do is you might isolate a pathogen directly from the host, and so you might be applying some specific wet lab techniques to extract the pathogen from the blood or a nasal swab or something, and then you would sequence it, and there's a variety of methods of sequencing.

Sequencing is the process of reading an organism's genetic information, i.e. determining the exact sequence of the DNA or RNA bases in its genome. Sequencing technology has come on in leaps and bounds in recent years. The original sequencing machines were weighty beasts, and expensive too. The first

human genome ever sequenced was carried out by the Human Genome Project, which cost them between 500 million and 1 billion US dollars[1]. Now it's possible to sequence DNA on machines the size of a memory stick, though most sequencing these days is done on so-called Next Generation Sequencing machines, and the cost of sequencing a human genome is in the region of a few thousand dollars, as compared to hundreds of millions of dollars.

> **[Dr. Lycett]** There is the so-called Next Generation Sequencing, which uses giant machines at the Sanger Institute to generate sequence data, but there's also a really tiny sequencing machine called an Oxford Nanopore device, and that's about this big, it's very small…

When she said "this big" she was miming the size of a large memory stick - think one of those dongles that you plug into your computer to access the internet.

> **[Dr. Lycett]** And you connect to your laptop via the USB port, and this kind of sucks the DNA through the nanopore device, it makes a little electrical signal which is transformed into the DNA sequence. And this method is also becoming a lot more popular as well because you can use this method in the field. In fact, as well as from samples, from specific hosts, you can also extract virus and pathogen sequences from things like fecal environmental samples as well, and it's possible to do non-specific sequencing, called meta-genomic sequencing as well[2].
> **[Tom]** Amazing, and so once you've kind of, you've got the DNA, and you've sequenced it so you know, you basically have the blueprint, what comes next? How do you use this information to look into the kind of things that you research?
> **[Dr. Lycett]** So the basic principle is that if sequences are similar to each other, then they are related. So for example viruses, they accumulate

[1] https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
[2] https://en.wikipedia.org/wiki/Metagenomics

> mutations over time, and you can use this information to infer who infected who.

So let’s say you’ve got two people, who I’m imaginatively going to call person 1 and person 2. Sorry to those who were rooting for calling them person A and B, but you know I’m just feeling the numerical system today? So, persons 1 and 2 both have coronavirus, and one of them infects person 3. To determine who infected who, we can sequence the viruses from all three people, and see whichever virus from host 1 or 2 is the most similar to the virus from host 3.

Looking for differences between two people can be challenging though, as there may not have been enough mutations in the viral genome to be detectable, and picking up, say, one mutation in a genome made up of thirty thousand bases is extremely difficult. However, over longer time periods the mutations start to build up in the genome, which means we can work out the similarity on larger scales, like between groups of people, or between regions.

> **[Dr. Lycett]** It might not be exactly, like literally which individual infected which other individual, it might be more like which group of individuals infected which other group of individuals. So that might be, which of the one host species infected which of the other host species, or which city infected which other city. I’ll give you an example. Suppose you detected a sequence in host A at some time, and then at some slightly later time, you detected another sequence in host B, and there was maybe only a few mutations different between them. And if you assumed that the virus say, only lasted in the host for a certain length of time anyways, then you might be able to say that A infected B, because the sequences would be similar, and the timings would be appropriate.

In the end though, there are many reasons why it's often hard to attribute infections to particular individuals.

> **[Dr. Lycett]** Of course, there’s a lot of caveats with saying it was exactly A that infected B, timing is one

of them. And the other one though is that you may… Suppose there was another host in between A and B that you didn't measure, well you can't exclude that from happening as well. So you might have similar sequences, and you might say: "Ah well it looks like A infected B," but it might not be a direct infection. But, nevertheless you can still…

[Tom] You can try.

[Dr. Lycett] … you can still say something, and you can talk about the general direction of things. And often that's good enough.

[Tom] And so, is it allowing you to just track the progress of the virus, if you can see it changing a little bit per person, you can say: "This is where it came from, this is where it's going"?

[Dr. Lycett] Absolutely, yes. And this is the principle of pathogen phylodynamics.

So there you have it, now we are all experts in pathogen phylodynamics! ... Okay, maybe not quite experts, but we know a lot more than we did. I told you all would become clear.

[Tom] Obviously, in light of the pandemic, you're now looking at the coronavirus, SARS-CoV-2. What is it specifically that you're working on at the moment?

[Dr. Lycett] So right now, I'm looking at Covid-19 sequences from people. I have also looked at the precursor sequences from bats and pangolins, to people. But right now, I'm looking at sequences from the people because there is a good 10,000 sequences now, globally, and also many thousands from the UK as well, and I can use this to look at the spread between the countries, but also within the regions of one country, and I'm just starting to look at that within the UK as well.

[Tom] What kind of questions are you hoping to answer by looking at this data?

[Dr. Lycett] So it really is: look at the spreading patterns. But also: look to see whether the spreading is accelerating, which hopefully it is now not because of the lockdown, but hopefully, hopefully what we'll see is evidence of the lockdown essentially stopping

> the spread, and stopping new spread, and it just being kind of like old spread dying out. This we really do hope to be able to see from the sequence data.

Pathogen phylodynamics is a powerful tool in a scientist's arsenal when trying to fully understand the complexity of a disease. By tracing a virus' recent evolutionary history through its DNA (or RNA, in the case of SARS-CoV-2), we can discover things that we couldn't otherwise discover using traditional surveillance and contact tracing methods, which are both limited by time and resources[3].

Viral genome sequences are uploaded to online, open-access databases in huge numbers, and scientists like Dr Lycett can utilise them in their research. This requires huge computers to handle the thousands of genomes, which are all tens of thousands of bases long. The sequences are then used to create an evolutionary tree, showing the relationships between the viruses sampled. Think of the tree of life, which starts from a universal common ancestor and goes on to branch off into all the main kingdoms, like the plants, the animals, the fungi, the bacteria, and the archaea[4]. Now imagine a similar tree, but on the much smaller scale of an individual virus species, and the branches of the tree relate to individual viruses or groups of related viruses. Using these so-called phylogenetic trees, we can trace the path of the COVID-19 pandemic.

> **[Dr. Lycett]** So the particular signatures that we'd be looking for are essentially clusters, and cluster growths and cluster shrinkages…

The clusters in this case are the viruses in a particular region, which group together if they are more closely related to each other. If the cluster is shrinking, then that means the virus is dying out, but if it is growing, then the virus is still being transmitted.

---

[3] https://ghrp.biomedcentral.com/articles/10.1186/s41256-017-0034-y

[4] To be more precise, the tree of life is generally considered to have three main branches: bacteria, archaea, and eukaryotes. Eukaryotes (organisms whose cells have a nucleus with a membrane) include animals, fungi, and plants. Viruses do not fit neatly into any of these categories.

[Dr. Lycett] … and a quantity called viral effective population size, because this is showing the number of active infections at one time.

From this information we can also calculate things like the R0 value, which tells us how infectious a disease is, or we can look at the impact that our interventions are having on the spread of the virus.

[Tom] I was interested in your jump from physics to the life sciences and bioinformatics. Because I guess physics is quite mathematical and computational as well. Was that a natural jump, or was that a big leap for you?

[Dr. Lycett] When I came into bioinformatics — this would have been about 2006 — I'd already been doing signal processing in physics for some time, so from the statistical signal processing side, into bioinformatics, the jump is not that big. It depends exactly on the flavour of bioinformatics that you do. Since then, there's a lot more data actually. Around 2006, slightly before, this was when the Next Generation Sequencing was starting. There starts to be a lot more data and there starts to be interesting things you can do with the data. Of course, and since this time, because there's so much data and so many things that can be done, and also how you link the biology of what's happening, it's become a lot more. So now I would say the field of bioinformatics is much richer than it was even when I started.

[Tom] I guess, yeah, that the sequencing is getting cheaper every day, and you have various platforms where people post sequence data just open access, which I think is really cool. I guess, with the pandemic, science has had to move really quickly and there's had to be a lot of international collaboration. What is it like for you, at this time, doing research at this, such a fast pace where things are changing on a daily basis?

[Dr. Lycett] One thing to say is that in terms of international collaborations, even before the pandemic I had quite a few international collaborations anyway,

> so all that's happened is they've just got more. *laughs* So that part wasn't exactly a step change, it's just more, more of it. In terms of doing things very quickly, I guess, as I said, things have moved on anyway, so it is more possible to do things more quickly in any case, and so it's just kind of busier.
> **[Tom]** From your perspective, what would you say is the most important thing that people can be doing individually, and then potentially, what we can be doing as a country, to tackle this virus.
> **[Dr. Lycett]** So on an individual level, of course it's important to follow the government guidelines, of course you should do this. For me personally, I don't especially need to go out to work, you know because I'm computational, so it's kind of fine for me to stay locked down, and I'm very fortunate to have a garden anyway. So this is not a huge issue on a personal front. So I would encourage people that, if you can, stay not-transmitting. I think the most important thing is: don't spread it, right? So if you can do things to not spread it, that's great, and you should do that.

That means wearing a mask, people. I'm personally not a huge fan of being able to smell my own stinky morning breath when I wear my mask to the shop, but it is a sacrifice I am willing to make. Stinky breath = no death, as I always say. Okay I've literally never said that, and let's be honest it doesn't really make sense. Just wear a mask, okay? That's all I'm saying: just wear a mask.

> **[Dr. Lycett]** As a country, I think again, you want to keep the transmission down but particularly you want to be not transmitting to vulnerable groups. So again, things that can be done to protect the vulnerable groups, I think is very important. And I think the final thing, I would say is, although I love going to international conferences and meeting my international collaborators, perhaps at this time, perhaps in the next couple of years, perhaps so much air travel is not quite so necessary.

[Tom] Yeah, have you found it's possible to do a lot of that over the Internet, like we are now, just using platforms like Zoom? How have you found that process?
[Dr. Lycett] Yes, I would say so. I mean, it is much easier with people that you already know, of course. But yes! I'm on a very large, European project, and all of our meetings, even with more than 50 people, have been done by Zoom, and this is okay.
[Tom] Amazing! Well, that's all my questions. I don't know if there's anything else you like to add?
[Dr. Lycett] Yes, I would just like to say to people: just stick there, it will be okay.

Outro

Outro music starts

And with those kind words from Dr Lycett, that is the end of the show and the end of the series! Huge thanks to her and to all of the other guests we have had on the podcast so far. We hope you enjoyed the coronavirus miniseries, and learned a thing or two about doing science during a pandemic. We've had a blast making it, and we will be back in September, so as I've been known to say, keep those ears peeled.

The coronavirus miniseries was an opportunity to get your feedback: what you liked, what you didn't like, and what we could do differently, so please don't hesitate to get in touch. Reminder, there will be a link in the show notes to a feedback form where you can let us know what you thought, and if you have any ideas, suggestions, guests or topics you would like us to feature. The podcast is brought to you by the Edinburgh University Science Magazine. You can find the show notes and the latest issue of the magazine at eusci.org.uk.

This podcast is edited by my partner in crime, Helena Cornu. The awesome podcast cover art was designed by our EUSci co-editor-in-chief, Apple Chew. The intro music is an edited version of Funkorama, and the outro music is an edited version of Funk Game Loop, both by Kevin Macleod.

I’ve been your host, Tom Edwick. Until next time, keep it science.

Outro music ends