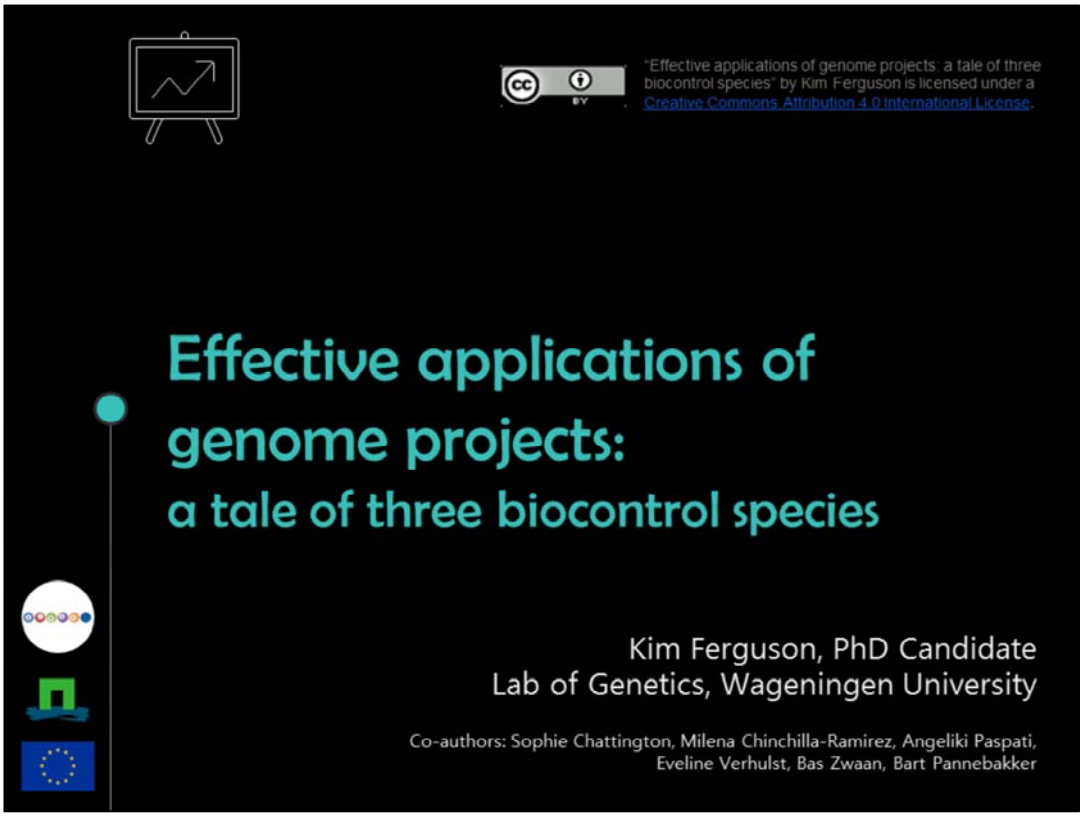
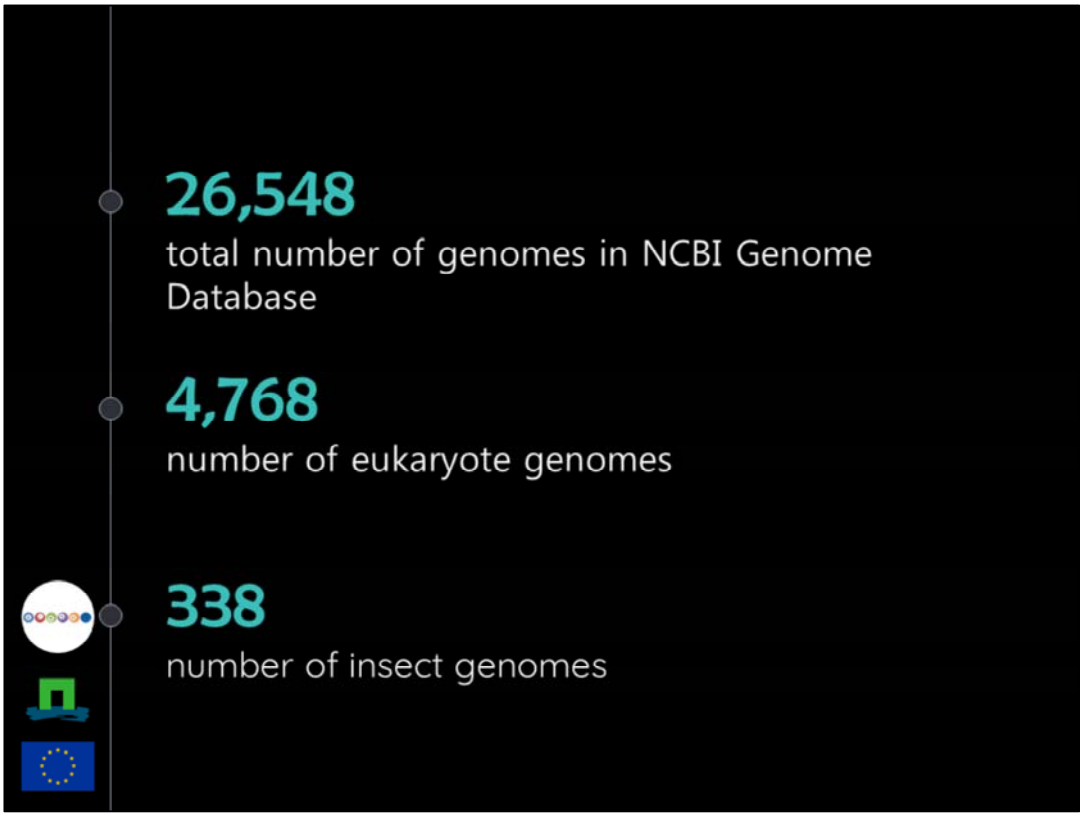


<<Ferguson, K.B., S. Chattington, M. Chinchilla-Ramírez, A. Paspati, E.C. Verhulst, B.J. Zwaan, and B.A. Pannebakker. 2017. "**Effective applications of genome projects: a tale of three biocontrol species.**" Presentation at IEIC5, October 18, 2017. [DOI available online]>>

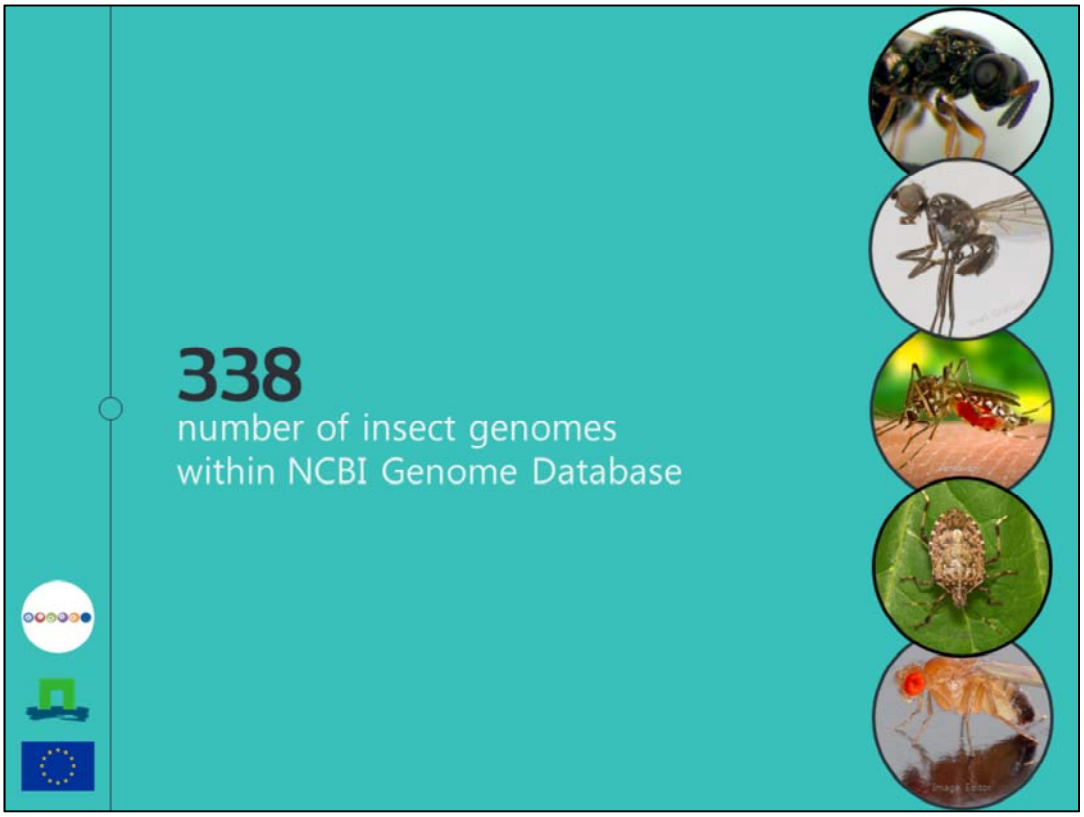
Hey everyone, my name is Kim Ferguson, and I'm a PhD Candidate at Wageningen University. Thank you for having me today. Two important notes before I start: First, this presentation will be made available online by the end of the week. Second, when you see the upper left symbol of a graph, that indicates that I have a previous talk that deals with this topic, and will have those links listed later. With that, let's get started, starting with some numbers.



26, 548 – the total number of genomes in the NCBI Genome Database (as of some time in late September)

Of these, 4, 768 are eukaryotic genomes. Already, we're narrowing the field here.

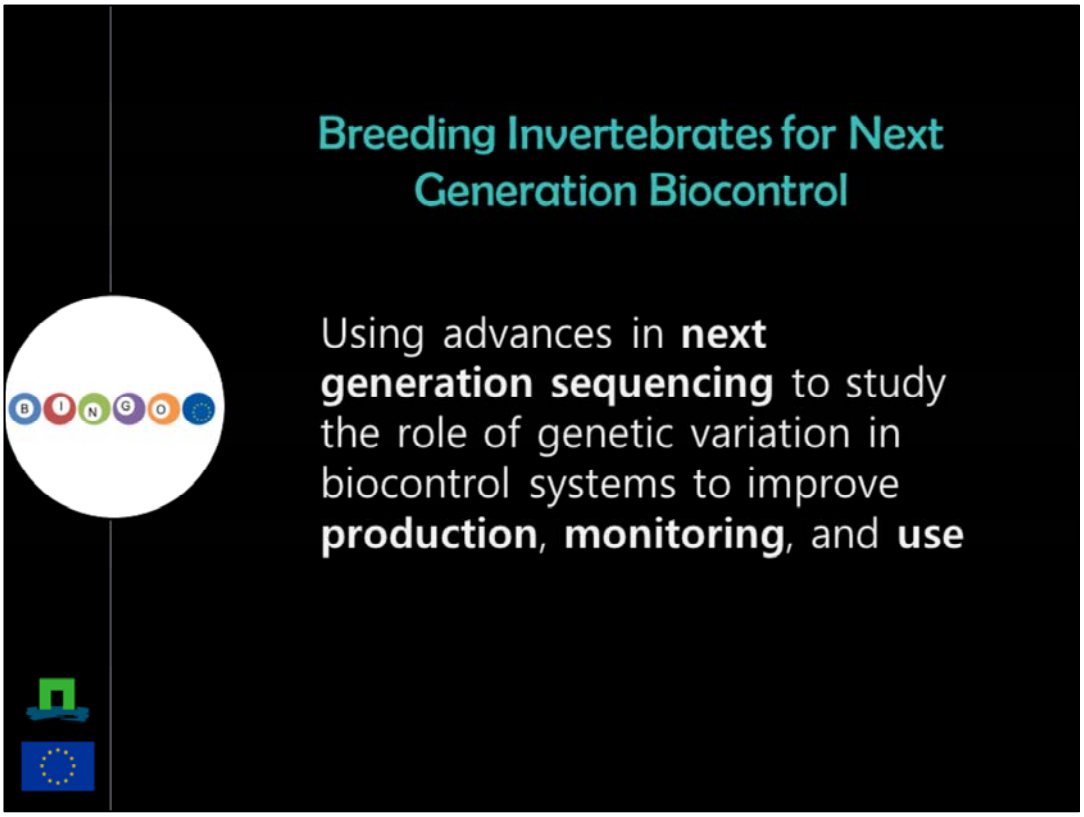
So how about insects? 338!



Well, 338, its quite a bit I guess on it's own. And here are some of the species that have their genome sequenced and available for public access. The expected heavy-hitter genuses – Nasonia, Aedes, and Drosophila, as well as some more recent genomes – *Themia minor* and *Halyomorpha halys*.
These are all fairly different insects – different lifestyles and roles in ecosystems or in agriculture, as well as different study uses. Some we would call model systems. The availability of these genomes can open up tremendous possibilities

Is there anything holding us back from adding to this number?

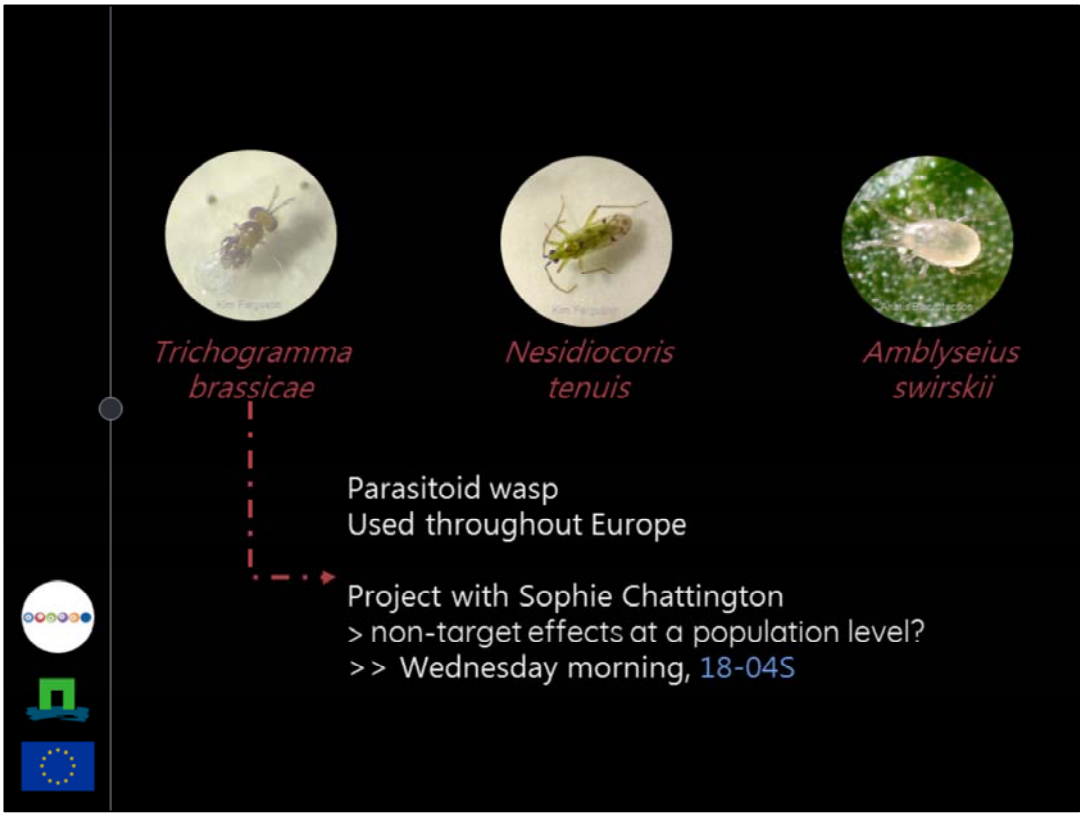
I think that we've all heard some variation of this phrase within the last few years – essentially, the approaches to sequencing genomes and other –omes are growing in approach and popularity, while the costs are decreasing. Decreasing costs are one thing though, what remains is the accessibility of genome projects – both in access to the market as well as really narrowing the focus of a project to a particular goal – do we as researchers have enough skills, time, and money to make a genome project worth it or is it all a lot of work with a slow payoff. And I hope to give you a few examples today as to why genomes aren't so scary and can be super useful



My project is within a larger project, called BINGO – Breeding Invertebrates for Next Generation Biocontrol. It is an integrated training network funded by the EU under the Marie Skłodowska-Curie Actions program, and is all about training researchers.

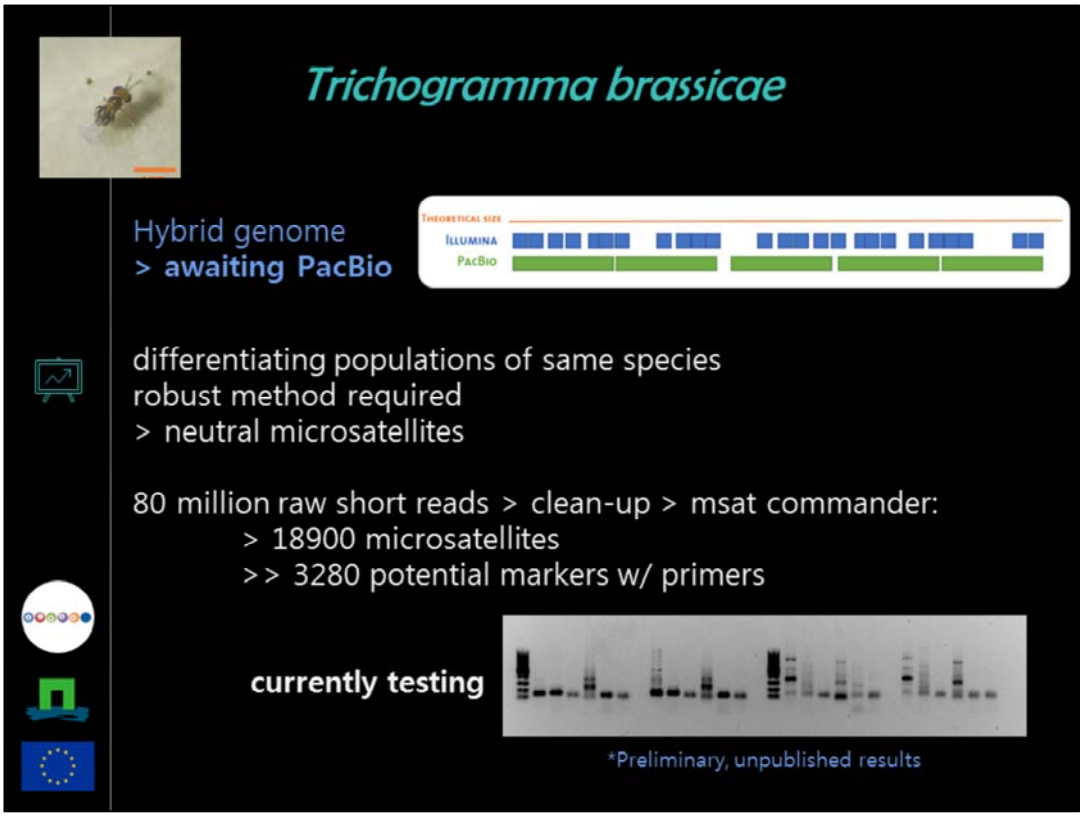
What our network hopes to do is combine the advances in next generation sequencing with biocontrol systems to improve the production, monitoring, or use of biocontrol agents.

There are a lot of people involved in the BINGO project, and some are here this week, so you can look forward to their presentations and posters



So my project is about working with three different genomes. My collaborators on these projects will all be giving their own presentations this week, so I won't tell you too much about their perspective. While these are very different species, the goals of the projects tend to go into the same territory.

The first genome I will talk about will be *Trichogramma brassicae*, a parasitoid wasp used throughout Europe in crop biocontrol. This project is with Sophie Chattington, who will speak tomorrow morning.



The approach we chose for the Trichogramma was a hybrid de novo genome, where there will be two different sequencing techniques used. To simplify it, there are short reads and long reads. This method was chosen because of the likely large amounts of tandem repeats often found in hymenoptera genomes, and would have meant a confusing assembly if it was just composed of short reads. So we bought some extra insurance. At the moment, we're just waiting for the long reads, the PacBio reads, but there are still things we can do in the meantime

So, what are we doing with this genome? A desired outcome of this project is to be able to provide markers that can differentiate between different field populations of same species. We also want a way that can be repeated reliably across labs, so while we have the time, we will start with microsatellites, repeating segments of DNA, and get a robust amount of data To do this, we will use neutral microsatellites across the genome, however many as is needed.

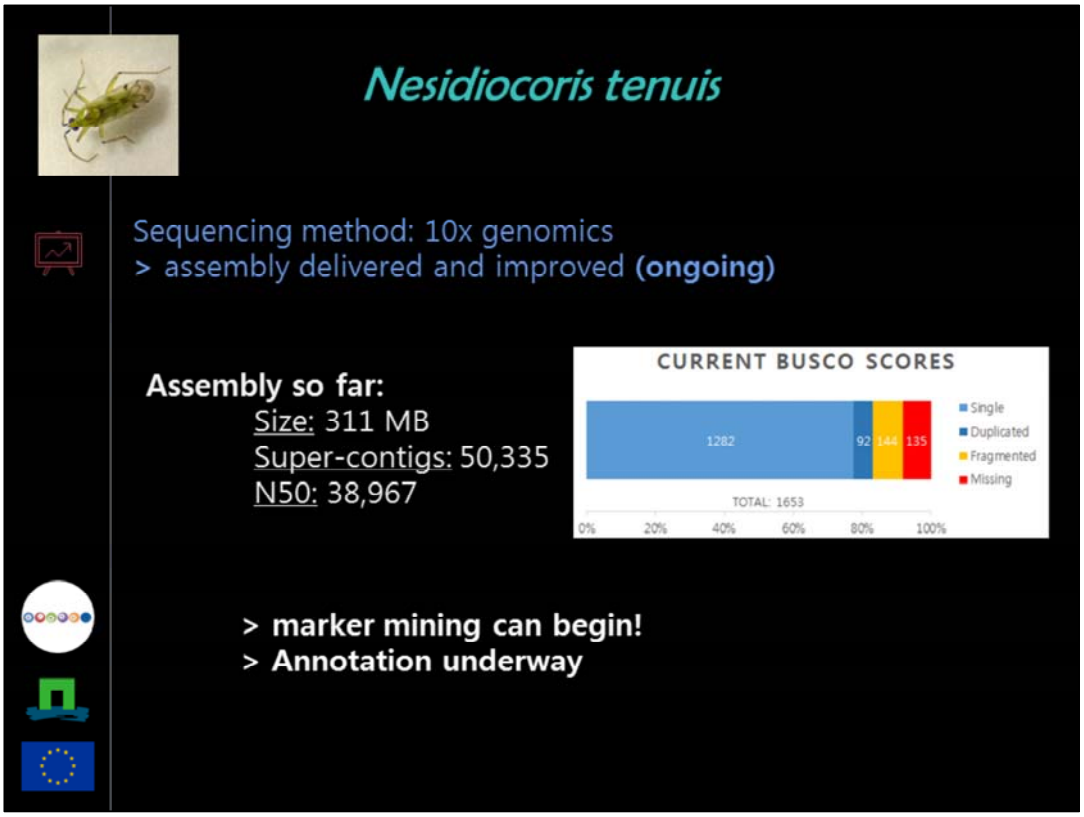
Now, we can already start looking at microsatellites with that short read data that we have.

80 million raw short reads > clean-up > microsat mining with msat commander, a python program which finds different types of microsatellites based on input parameters, and also uses the PRIMER3 program to create primers based on the flanking regions

> 18900 microsatellites
> 3280 potential markers w/ primers

> currently testing a selection on different populations and species – I can explain more of this after, but suffice it to say that things look promising, but we need to test more markers, this is still in the initial stages

Moving on to *Nesidiocoris tenuis*, we've already assembled the genome of this mirid bug. This zoophytophagous insect is popular in Spanish tomato greenhouses, but it is far from perfect – and Milena Chinchilla-Ramirez will tell you why tomorrow afternoon.



From single female from Koppert Spain population, we were able to use the 10x Genomics platform to go forward with the sequencing and assembly – for more on this sequencing and assembly method, a previous presentation that I'll link to in references will tell you all about it, or you can ask me later.

> assembly delivered and improved – this is an ongoing process

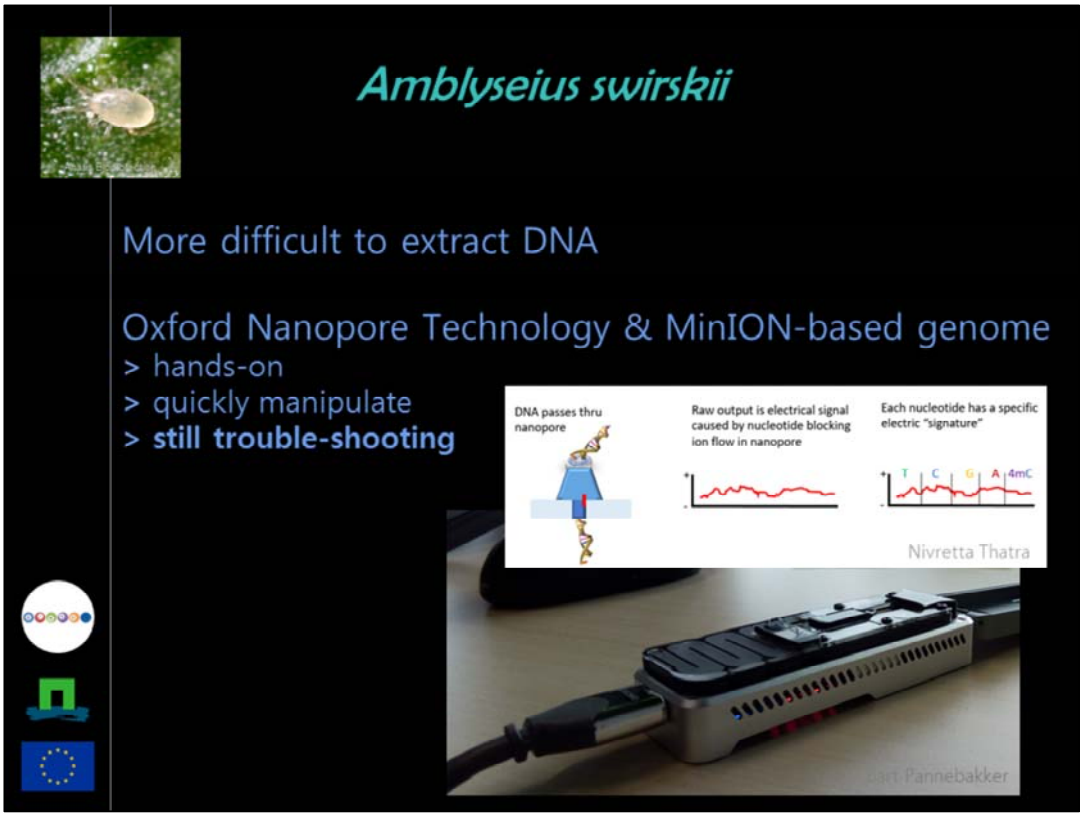
> Assembly so far: So now, I will put a bunch of numbers up and a graph, and if you are familiar with genome assembly, these will be informative, but I'm just going to tell you some descriptive information that will tell a similar story.

The assembly so far is looking good. After some initial clean-up, we were able to get an assembly that is in just over 50,000 chunks of various sizes. The chart on the right is a BUSCO analysis – short for Benchmarking Universal Single-Copy Orthologs, and looks at key genes that should be present in a genome. If you take the puzzle analogy a bit further, the blue is fully assembled pieces, some we have extra copies. The yellow bit are partially assembled pieces, and the red are missing pieces from the puzzle. What you would want is all blue, and this is fairly close and will get better with annotation.

We're happy with these results and want to move right on to annotation alongside some marker mining that will be useful for Milena's work to test for existing genetic variation within the commercial population. As I said before, microsatellite mining can happen over a weekend, so we will start testing those while working

on annotation in the meantime. So I encourage anyone working with these species to get in touch if you want to get involved or benefit from the work we've been able to complete so far.

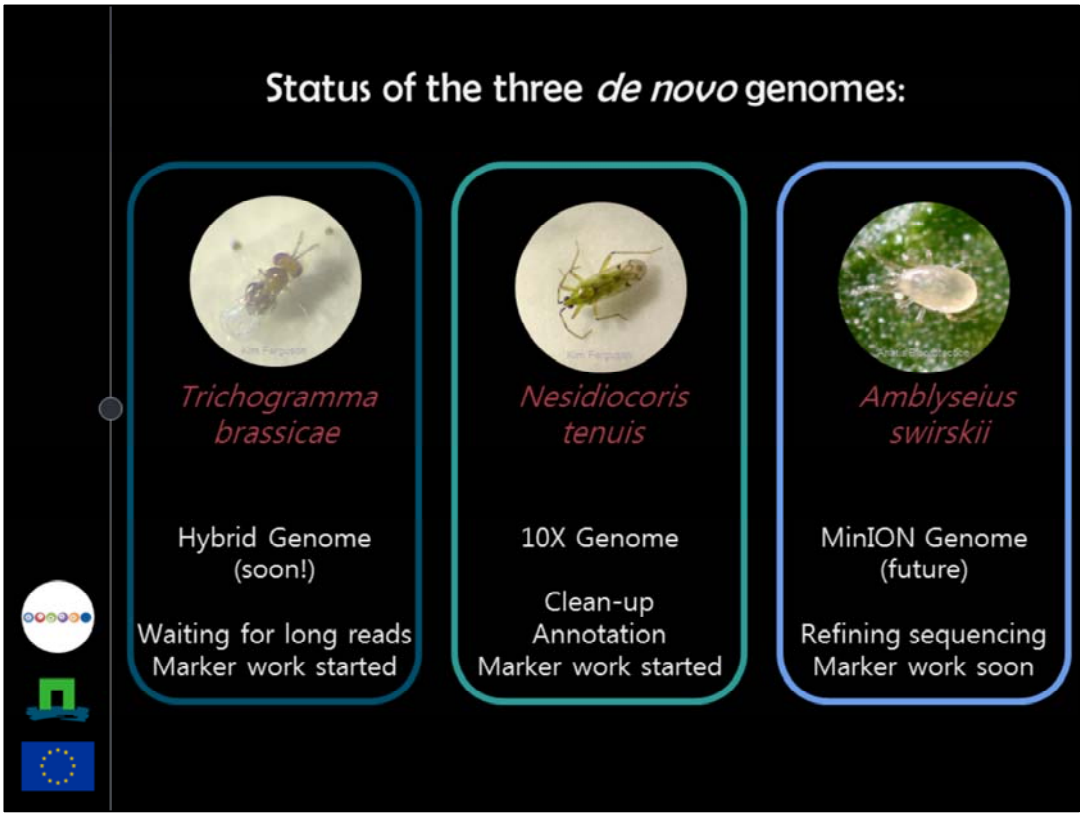
And last but not least, the mighty *Amblyseius swirskii* – one of the most widely used predatory mites in biocontrol. The key goal to look at the population-level genetic variation to see if selection is possible. Angeliki Paspati will be speaking about her work with swirskii on Thursday afternoon.



We're working here with commercial populations from Koppert (a dutch biocontrol company). It's quite a tough nut to crack, and with DNA extractions being a bit difficult (^pollen, detritus, small amount of genetic material per individual),

we decided to go with Oxford Nanopore Technology and the MinION – a desktop sequencer that can (in theory) run for two days and deliver results. It uses the electric potential of nucleotides as they go through nanopores, enabling a massive parallel sequencing output in a device that is smaller than a banana. Not only is this speedy, it means that we can tweak things in the moment, rather than shipping a sample and waiting two weeks to find out that it's poor quality and we need to do everything again. What's required then is the time and effort put in to the troubleshooting.

The future will also include marker mining, to be used to screen for variation between populations in addition to other tools

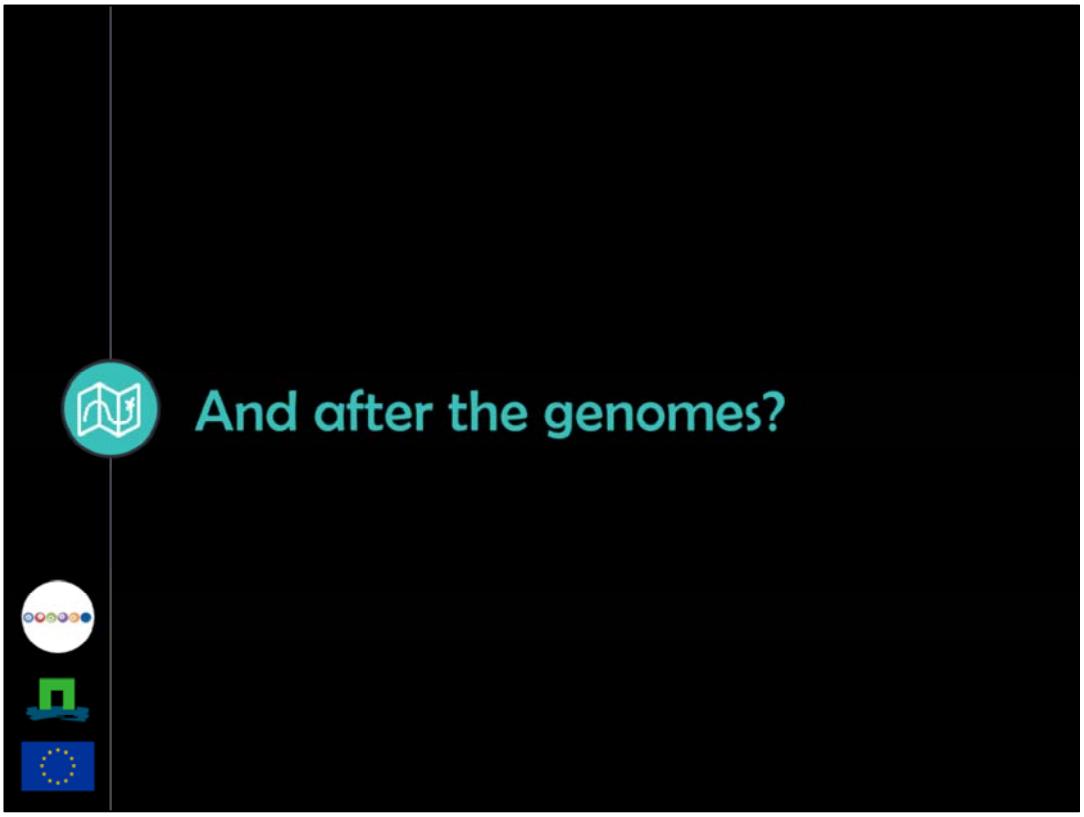


So to sum it up, here is our progress so far on these genomes

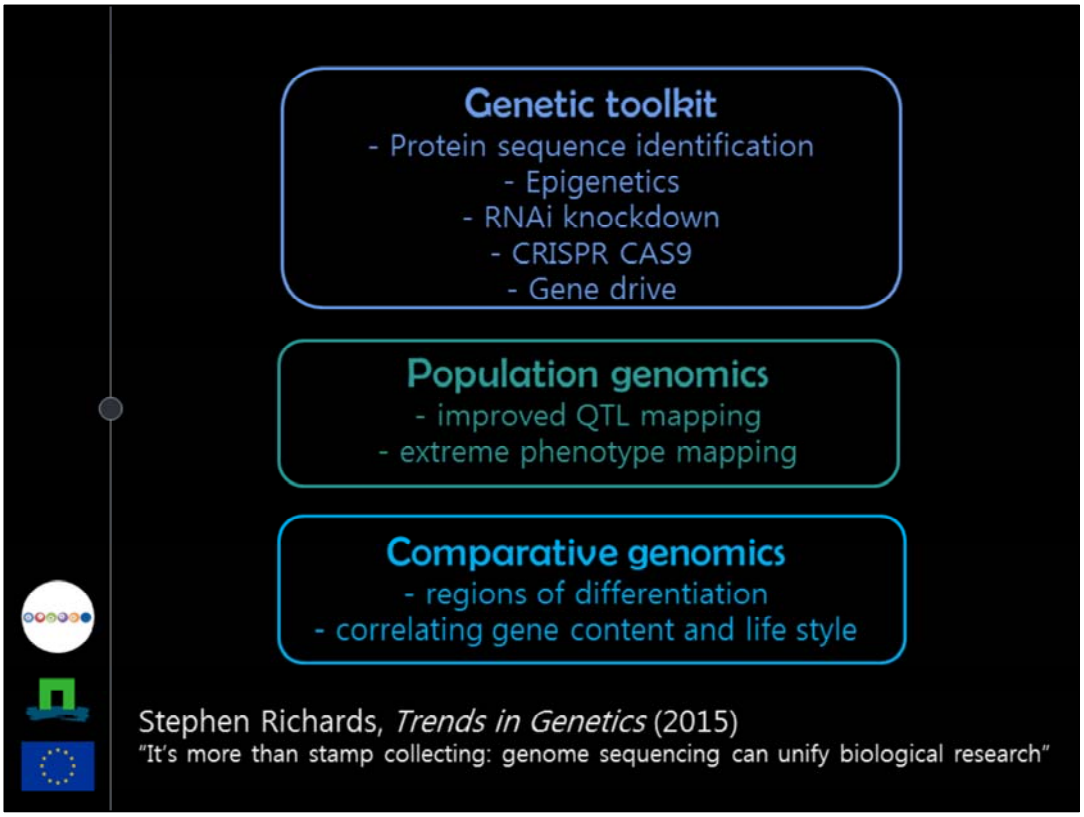
For *Trichogramma brassicae* we will soon have the hybrid genome and then would move on to annotation, but in the meantime, we have already found the microsatellites and are starting to test them.

For *Nesidiocoris tenuis*, we have the genome, and it's already being cleaned up and annotation work is in process. We are concurrently generating markers to be tested, so that will be coming down the pipeline soon

Finally, with *Amblyseius swirskii*, we are still working on it. BUT, it looks promising!



So, what about after all of the sequencing, assembly, and annotation is finished?



So once you have a genome, what are the possibilities? While I point out a few, there are several scales that genomes can be used for studies.

Talking about genetic/molecular/genomic toolkits, there are a variety of ways that having access to full genome can assist

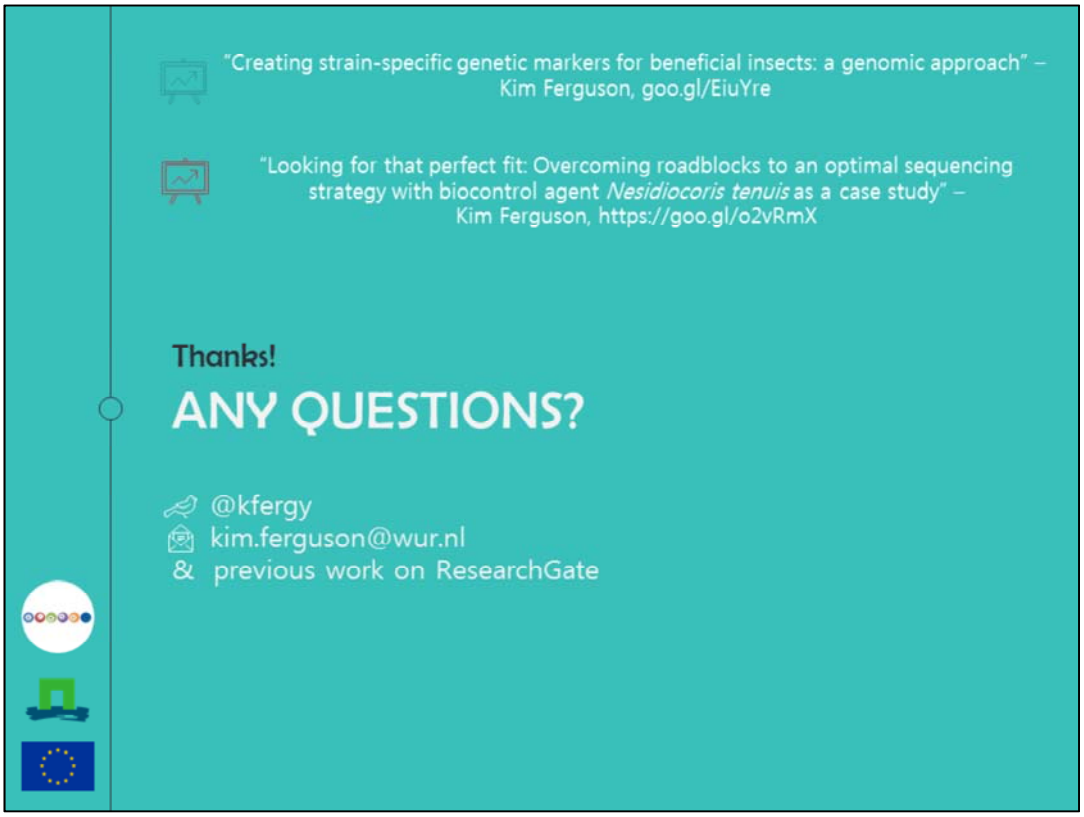
For population scale studies, genomes can improve QTL mapping as well as looking into extreme phenotypes.

Then there are larger scale applications, such as comparing species or genus, and looking at regions of differentiation or looking at connections between genes and lifestyles. This also works for phytogenetic placements.

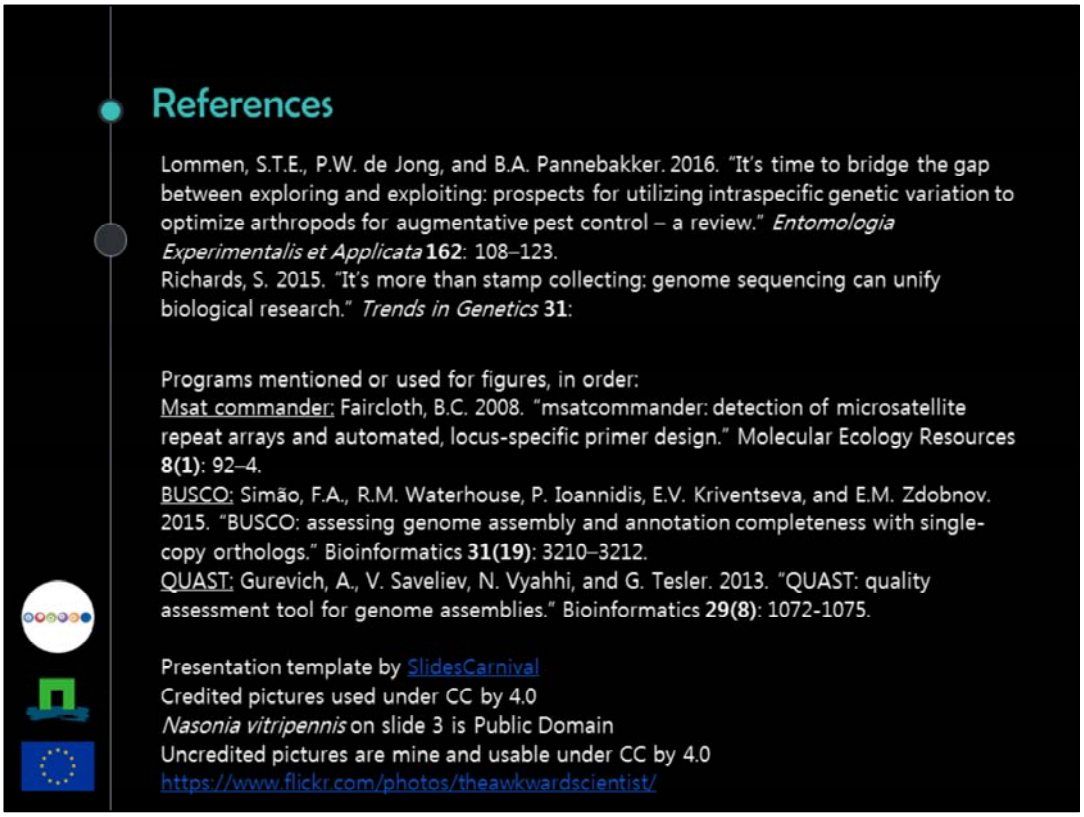
As for my work, after the genomes are complete I'll be looking at population genomics, because the questions that I want to answer my project relate to the genetic variation between native/wild and commercial populations, and look for evidence of genetic drift or lab adaptation. By having the full genome, I can better isolate areas of interest for further investigation. So while the microsatellites I find will be useful for other projects, it will also be part of a larger study into variation that I will also supplement with population genomics to see a wider picture of the variation.

And that's really what genomes can deliver – a wider or deeper view that can improve our understanding. So I hope that if you were a bit skeptical or felt like it was a bit too much of a challenge, this gives you a bit more insight into how these projects can work for you, as well as add to the overall state of the art of a species or study system.

So with that, I'm going to thank my department back home as well as my BINGO colleagues. On a project like this, there are a lot of collaborators and my co-authors are some of them, so I want to thank them for their help and feedback. The two presentations I mentioned previously are linked here, and as I said before, this presentation will be made available online by the end of the week. You can find it at my ResearchGate page, or send me an email! Our funding comes from the European Union – if you want more information on BINGO, you can go to our website.



Any questions?



<< References for papers, programs, slides, and photo credits >>

Please make sure to cite this presentation correctly

Ferguson, K.B., S. Chattington, M. Chinchilla-Ramírez, A. Paspati, E.C. Verhulst, B.J. Zwaan, and B.A. Pannebakker. 2017. "**Effective applications of genome projects: a tale of three biocontrol species.**" Presentation at IEIC5, October 18, 2017. [DOI available online]