

Openscapes Pathway: Therkildsen Lab

Conservation Genomics and Molecular Ecology

Lab background

The Therkildsen lab was established in the Department of Natural Resources at Cornell University in January 2016. Our research aims to improve our understanding of how species adapt to their environment, and how quickly they can respond to altered conditions caused by selective harvest, climate change, or other anthropogenic pressures. With a primary focus on marine fish, we study how spatial and temporal variation in selection pressures interact to shape patterns of genetic variation within species and explore the roles of ongoing genetic adaptation and geographic distribution shifts in promoting species persistence in our rapidly changing world.

We currently have five lab members: the PI, our lab manager, a postdoc, and two PhD students. We interact formally and informally on a daily basis, get together for weekly lab meetings and now primarily communicate online via Slack, and use Zoom for telecommuting while traveling.

A bit about us:

Nina Overgaard Therkildsen: I'm an Assistant Professor in the Department of Natural Resources at Cornell University. My undergraduate training was in marine biology and environmental policy, and it was only during my Master's work that I discovered how population genomics can be an extremely powerful tool for addressing current problems in conservation and natural resource management and I decided I wanted to pursue that line of research. So I had a lot to learn! Since then, the tremendous advances in DNA sequencing technology have revolutionized the field, and dealing with the tsunamis of data we now are faced with requires a whole new skill set I was never formally trained in. Through a few key courses and a whole lot of google-assisted learning-by-doing, I am now able to deal with those challenges. Yet, I have felt under-qualified to adopt more efficient ways to share research pipelines with others (and my future self!), so I was very excited about the opportunities offered through Openscapes to start making a transition to more open science within a supportive framework.

Nicolas: I am a graduate student in the lab. I am broadly interested in the fields of population genetics and conservation genetics. My current research focuses on the population structure, local adaptation, and fishery-induced evolution in the Atlantic cod.

Arne: I am a postdoc with Nina Therkildsen in the Department of Natural Resources at Cornell. I am broadly interested in understanding the processes and mechanisms underlying rapid (in the span of a

10s to a few thousand generation) adaptation and diversification in species, mostly fish. To address these questions I tend to combine a range of approaches, combining phenotypic, population genomics and functional genomics in systems of parallel and rapid evolution. In the Therkildsen lab I am focusing on understanding the population genomic patterns of local adaptation with gene flow in Atlantic silversides and the functional genomic mechanisms underlying adaptation and phenotypic divergence. My undergraduate training was mostly in developmental biology and functional genetics, and for my PhD I decided to move into population genomics to widen my skill set. This meant that I had a lot to learn, which is also the fun part of science, and heavily relied on the resources and code openly shared by others. Thus, I was excited to be part of the Openscapes community and work toward a more open scientific community that will hopefully lead to better science by allowing scientists to learn from each other.

Maria: I am a PhD candidate in the Department of Ecology and Evolutionary Biology at Cornell, co-advised by Drs. Nina Therkildsen and Kelly Zamudio. I am broadly interested in ecological and evolutionary genomics, with a focus on understanding the mechanisms underlying adaptation and divergence in natural populations of vertebrate species. Currently, I am working on characterizing the genomic architecture of local adaptation with gene flow in Atlantic silversides. Specifically, I am describing structural variation between locally adapted populations and identifying genomic regions that underlie adaptive traits. I am very excited about increasing access to genomic data, which provides a powerful tool for studying ecological and evolutionary processes, but demands unique skills to analyze and interpret. Having to learn the computational skills required for my research is one of the most challenging, empowering, and rewarding aspects of my training. One of my greatest struggles is overcoming my tendency to work too independently and suffer in silence when issues arise, so I am especially thrilled to now be part of the amazing Openscapes community.

Our Openscapes story

Although we all work on different projects, many of our workflows are shared or similar. In the past, we had tried to synergize by sharing data files and analysis scripts by emailing to each other, or making lots of parallel randomly named copies on our shared server. It worked to some extent, but we often would get confused about what version of a script a certain analysis had been run with and each person would frequently have to reinvent a small part of the wheel. Neither efficient, nor always effective! We realized there was a better way with formalized version control, but none of us had the skills or experience to set it up, and - more importantly - nobody felt they could invest the time in acquiring those skills and transitioning our workflows.

Openscapes provided a really useful framework for us that boosted our motivation to transition towards more open and reproducible workflows, provided technical guidance and support to make the barriers to entry seem less daunting, included bi-weekly accountability to a supportive and driven group that kept us focused on the process even when the semester had us very busy with other tasks,

and both the direct program material and the community of champion labs gave us lots of inspiration for how we want to organize our workflows and data management better in the future.

Our [Lab Pathway](#) summarizes how we used to do things and the goals we have set for moving forward. Some of the goals we have achieved, for others we are not quite there yet. But there is no doubt that Openscapes has given us a really important push to initiate a transition process that will make our work much more open, collaborative, and easily reproducible moving forward.

Key accomplishments

- We have backed up all of our raw data on a secure server.
- We have started to use Google Drive to keep track of our sample and sequence metadata, lab notes, as well as experiment protocols (instead of having various randomly named files scattered across everybody's personal computers).
- We have transitioned to use Slack as our primary online communication tool (instead of email). We like it a lot because it lowers the barrier to initiate conversations and helps to facilitate collaboration among lab members. We also found the various features in Slack very helpful in keeping our communication organized by having channels for specific purposes and keeping track of prior discussions and lab results by having all prior conversation and image sharing on a topic listed in one place with the ability to reference previous comments, star and pin items, and etc.
- We have incorporated GitHub as a central part of our workflow. We have made our central data processing pipeline and part of the data analysis pipeline available on GitHub, have made the pipeline flexible to the need of other users, and have provided clear documentation, so that our collaborators (and our future selves) can easily access it without the need to reinvent the wheel. In addition, we have started to keep track of our entire analysis pipeline and result visualization using GitHub as well, enabling more efficient communication among lab members on various projects. We have also recommended (and taught) GitHub to some collaborators and colleagues.
- We have come up with a lab code of conduct.

Key understandings

Nicolas: One key understanding for me is that I learned spending some extra time to make my workflow readable and tractable can really help me save more time in the long run. One other thing: I found that some tools that were initially developed for the industry (e.g. GitHub and Slack) can be quite helpful to the academia as well.

Nina: With a desire to always find optimal solutions, I had initially felt very overwhelmed by the jungle of tools available out there to help promote open science, and the resulting insecurity about where to start prevented me from starting at all. A key thing Openscapes has illustrated to me, is that there

obvious is not one single solution that will work best for all aspects of our work. But also, that we don't have to completely change everything we do all at once. There are lots of little modular ways in which we can start to improve our workflows, and over time these will build on each other to really make a big difference in how efficiently and transparently we operate.

Arne: The two key points I have taken away from this experience is to spend more time making my analyses and workflow more easily accessible for others (and my future self) as this will increase the impact of my work and will save time in the long run. I have also learned that it is okay to share my workflow with the lab and the wider scientific community even if it is not perfect and that this will ultimately improve my work, a point I have struggled with and still struggle with. It ultimately taught me that we all seem to struggle with similar things.

Maria: Dedicating time to discuss data openness, workflow reproducibility, and a lab code of conduct was extremely valuable. I think it strengthened the sense of community and accountability in our lab, and has helped set a foundation upon which we can continue to slowly build our better data practices.

Next steps

Our immediate focus moving forward is to continue the process of getting all our analysis scripts on Github. We have decided that moving forward, for all new papers coming out of the lab, the full data analysis pipeline, including the code to reproduce all results and figures, must be made available in a well organized way in a Github repository. This requires a big change of habit for most of us, but the benefits are so obvious to us that we are committed to the challenge.

We are also in the process of getting all our old sample and sequence metadata uploaded and reorganized to a standardized format in our shared Google Drive folder.

To facilitate smooth transitions for new lab members coming into the lab and ensuring that all relevant data and analysis-related documentation stays in the lab after people leave, we are also working on developing onboarding and offboarding processes/guidelines.

Staying connected

We will definitely stay connected with Openscapes by following blog posts, links to tutorials etc. Nina will also be designing a new course for incoming graduate students (offered in Spring 2020 for the first time) that will introduce them to key open science tools, building to a large extent on the lessons from our own process. This course will connect closely with the mission of Openscapes and the program content!

Acknowledgements

We are very grateful to Julie and the entire Openscapes community for all the support, inspiration, and empowerment throughout this program. Thank you so much for helping push us towards better science in less time!

Last updated June 7 2019