Lucas Jones and Nick Van Gilder

EEB 5300

Final Project Proposal

*Introduction to the species and project*

Northern sand lance (*Ammodytes dubius*) are an important, yet understudied, forage fish. They play a crucial role in the food web as they are a main food source for whales, tunas and sea birds in the Northwest Atlantic. They are a member of the Ammodytidae family, which contains 33 different species. Northern Sandlance are found on well aerated, discontinuous sandbanks. To appropriately protect this species, population structure along the Northwest Atlantic shelf needs to be investigated, as connectivity and inter-reliance of this species is unknown.

Sandlance are found on sandy banks which are areas that tend to have strong currents. Sandlance have benthic eggs that spend roughly two months (21-75 days) on sandy substrate while developing (Staudinger et al). During this time, eggs can be moved by ocean circulation and current patterns. Eggs and larvae were found to drift offshore and south (Richards, Kendall). Population connectivity has yet to be investigated for this species. However, understanding this connectivity is of critical importance with regards to fishery management decisions. Sandy offshore banks, such as Stellwagen Bank, are under threat from beach erosion projects, energy development and fishing gear (Staudinger et al). If these are isolated groups, management of their individual habitats will look much different than a homogenous, mixing population. There have also been observations of large cohort years (Staudinger et al). Where these large cohort populations appear and what contributes to them will be important to understand.

In part, the lack of genomic studies on Sandlance is due to the fact that genetics projects are costly. In 2015, Horne et al. published mitochondrial sequences of both *Ammodytes americanus* and *Ammodytes dubius* to GenBank, stating to understand the population structure we need development of molecular tools. They reported thirteen protein-encoding, two rRNA, and 22 tRNA genes were detected for each fish. Further, distinguishing the two species was difficult (Horne).

In 2014, Hemmer-Hansen et al. reported on advent of Next Gen Sequencing (NGS) allowing scientists to collect genome-wide molecular data and its implications for marine fish research. In this way, we can move from population genetics to population genomics through population scale genome sequencing. There are many methods to take advantage of new NGS technologies, but Therkildsen and Palumbi laid out a workflow for low coverage, whole genome sequencing. With this method you can sequence hundreds of individually barcoded samples. This dramatically brings down the price of sequencing per individual. With coverage ranging from 1 - 4x, analysis of low coverage data allows us to look at population allele frequencies, genotype likelihoods, and polymorphisms based off of SNP’s (Therkildsen, Palumbi).

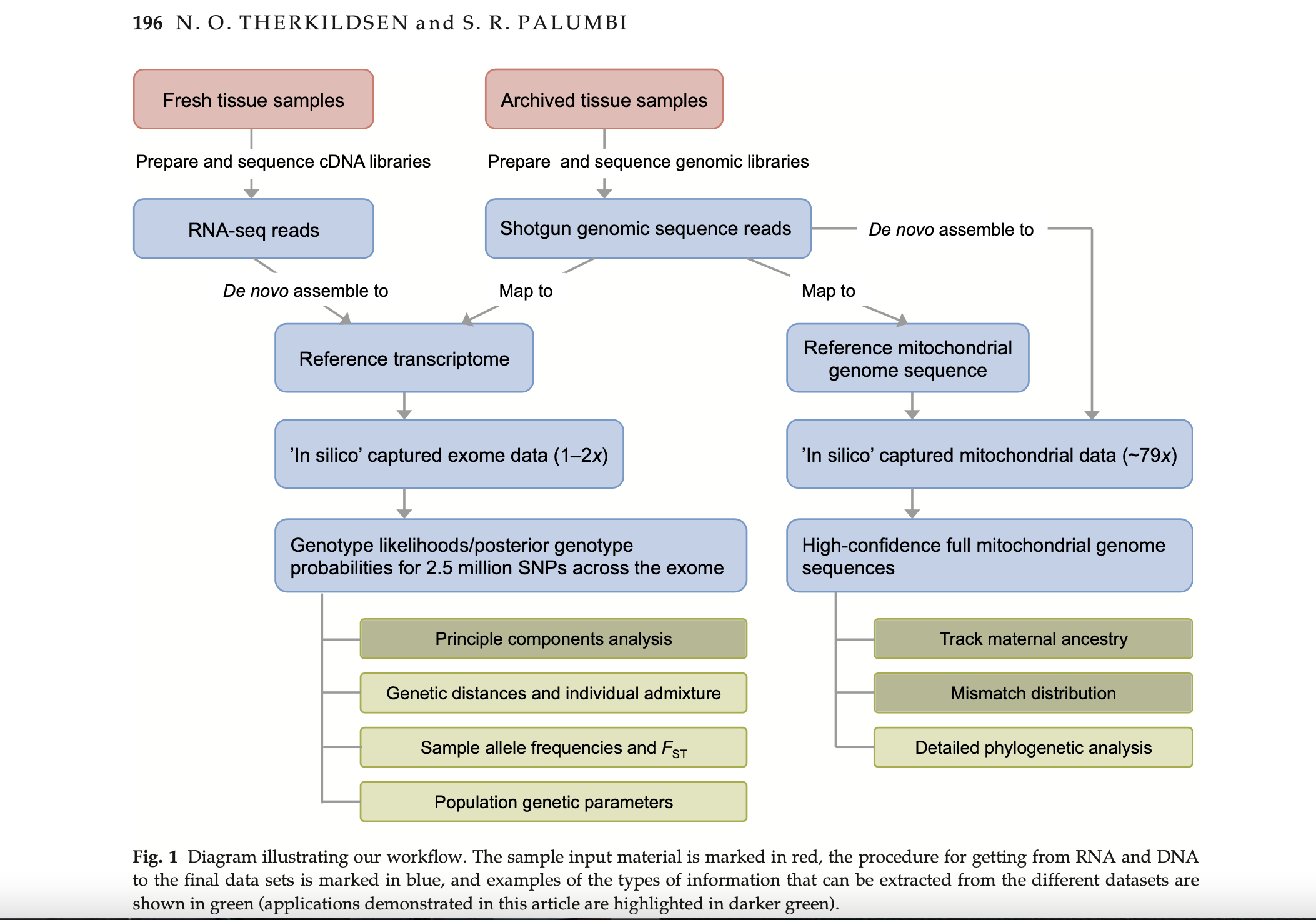
NGS techniques allow us to sequence higher numbers of specimens, allowing us to look at genetic structure, while not compromising on cost or accuracy. Applying this new information to Sandlance, we can now explore their population structures on these sandy offshore banks. NGS prospects allow us to collect samples along the Northwest Atlantic shelf to investigate gene flow; which we have already done. Samples were collected from various populations across the range of the Sandlance, which has culminated in a data set of ~300 individuals sequenced (30/site) across their whole genome with relatively low coverage. Utilizing a deep-sequenced reference genome, we will utilize the statistical power of our replicates to have a means by which to better understand these populations. Understanding more about the population structure and conditions that brought on big cohort years will aid in more informed policy and protection measures with regards to this important forage fish.

*Goals / deliverables*

For this project we seek to answer the following question: are these populations related?

To do this we will follow the workflow laid out in Therkildsen, N. O., & Palumbi, S. R. (2017), allowing us to use SNP lists to compare populations.

This will mean getting individual samples organized, mapping to a reference genome to align individual genomes, generating SNP lists, running PCAngsd, and plotting results.



*Diagram taken from Therkildsen, N. O., & Palumbi, S. R. (2017). This is the workflow that we will follow to answer our population connectivity question.*

*References*

Hemmer-Hansen J, Therkildsen NO, Pujolar JM. 2014. Population genomics of marine fishes: next-generation prospects and challenges. Biol. Bull. 227(2):117–32

John B. Horne, Meghan C. Mcbride, Jackie Lighten, Ian R. Bradbury & Paul Bentzen (2016) Annotated mitochondrial genome assemblies for two sand lances (genus: *Ammodytes*) from the northwest Atlantic,Mitochondrial DNA Part A, 27:6, 4607-4608, DOI: [10.3109/19401736.2015.1101579](https://doi-org.ezproxy.lib.uconn.edu/10.3109/19401736.2015.1101579)

Richards S.W., Kendall A.W.. Distribution of sand lance, Ammodytes sp., larvae on the continental shelf from Cape Cod to Cape Hatteras from R/V Dolphin surveys in 1966, Fishery Bulletin, 1973, vol. 71 (pg. 371-386)

Staudinger MD, Goyert H, Suca JJ, et al. The role of sand lances (*Ammodytes* sp.) in the Northwest Atlantic Ecosystem: A synthesis of current knowledge with implications for conservation and management. *Fish Fish*. 2020;00:1–34. <https://doi.org/10.1111/faf.12445>

Therkildsen, N. O., & Palumbi, S. R. (2017). Practical low‐coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. Molecular Ecology Resources, 17, 194–208.