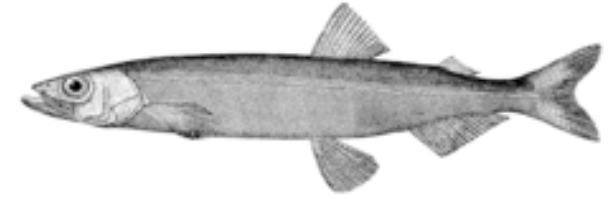


# Capelin dataset








DOI: 10.1111/mec.15499

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY



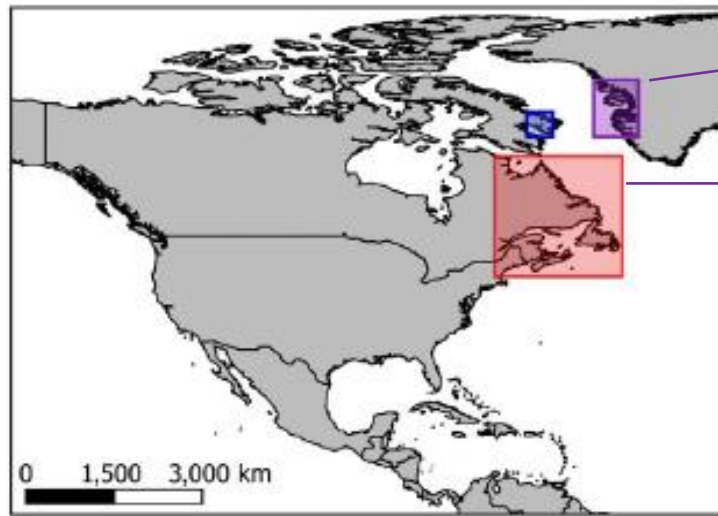
## Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish

Hugo Cayuela<sup>1\*</sup>  | Quentin Rougemont<sup>1\*</sup>  | Martin Laporte<sup>1</sup>  | Claire Mérot<sup>1</sup>  |  
Eric Normandeau<sup>1</sup> | Yann Dorant<sup>1</sup>  | Ole K. Tørresen<sup>2</sup> | Siv Nam Khang Hoff<sup>2</sup>  |  
Sissel Jentoft<sup>2</sup>  | Pascal Sirois<sup>3</sup> | Martin Castonguay<sup>4</sup> | Teunis Jansen<sup>5,6</sup> |  
Kim Praebel<sup>7</sup> | Marie Clément<sup>8,9</sup> | Louis Bernatchez<sup>1</sup>

- *Mallotus villosus*
- Small fish
- Spawn on beaches
- Cold waters of the North Atlantic Ocean



# Capelin dataset



Greenland species

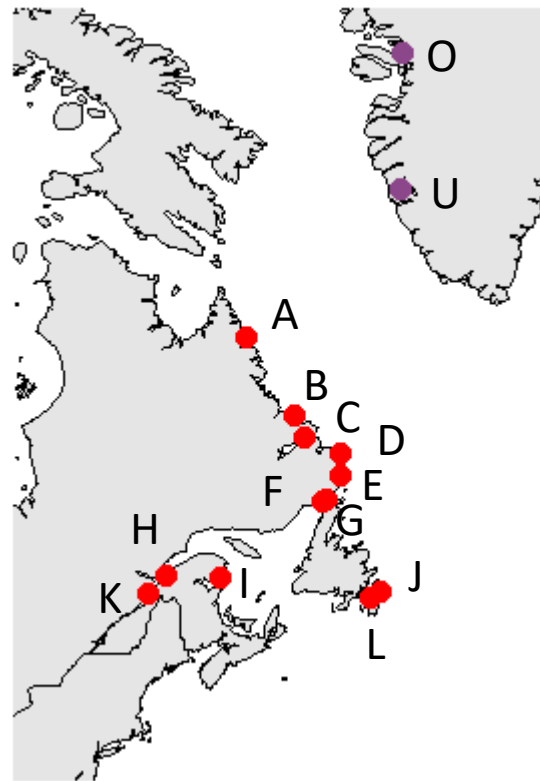
North American species

2 population

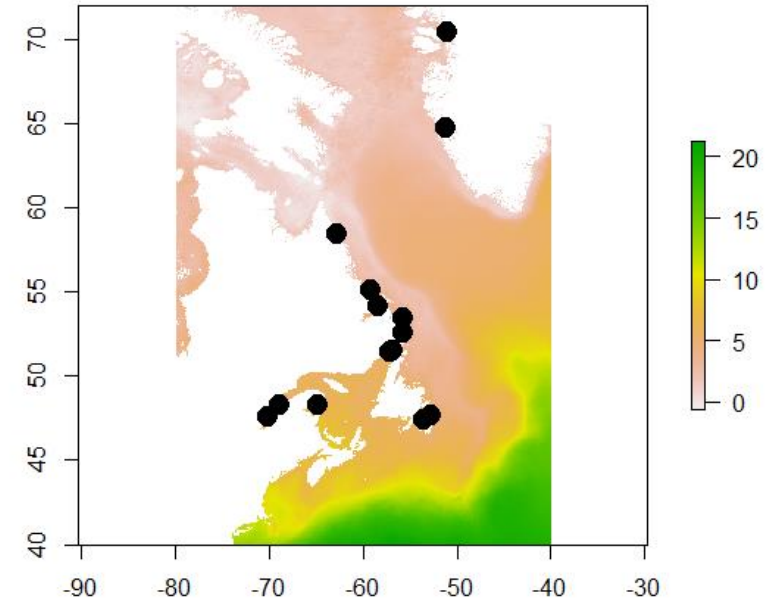
N= 40

12 populations

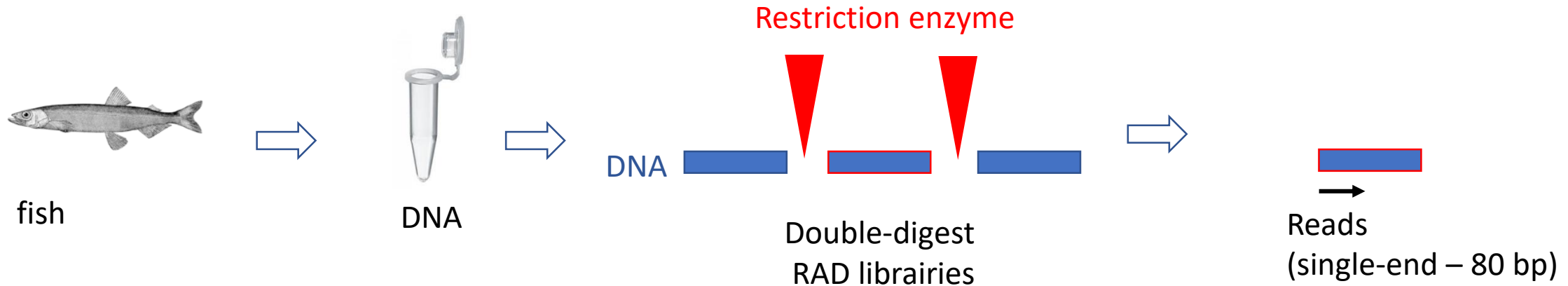
N= 240 ( 20/pop)



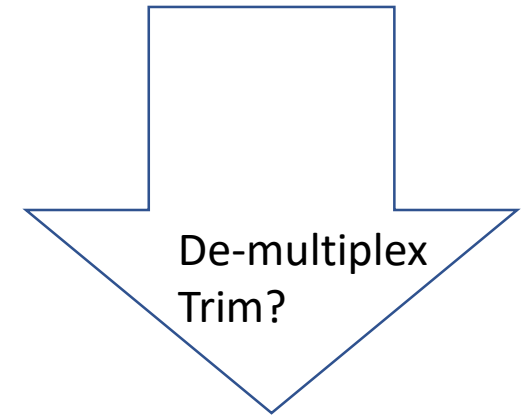
Sea temperature  
(from MARSPEC)



# Capelin dataset



```
@70ZFD:01332:11598/1
TGCATCAACTTTAAGATACGCTATTGGAGCTGGAATTACCGCGGCTGCTGGCACCCAGACTTGCCCTCCAATGGATCCTC
+
7<<=<;<4676*115345::=<;<=6;5<;<;7<1918<199<6<<::9:5:556+38469166=3;<6<655-477-4/
@70ZFD:01334:11636/1
TGCATCCTGTGGAAGTAGCTGCACACCTGCTCATGCTGTGCCAGGAAGGGAGGGTGGGATCAGCCAATCGGGGAACAGAG
+
5;?;;;5855;4:4<A<;<<;<<<B9B=<<<<<<;<;<<:69:58-55)533)/893<;<;9:496888<:1;599;;B
@70ZFD:01335:11615/1
TGCATGGCAGAGTGGAGAGGAGCGCCCTCTACTGGAACCTTCTGGAACAGGTCCTCCGAATGTCCAAGGTACAACGGTTC
```



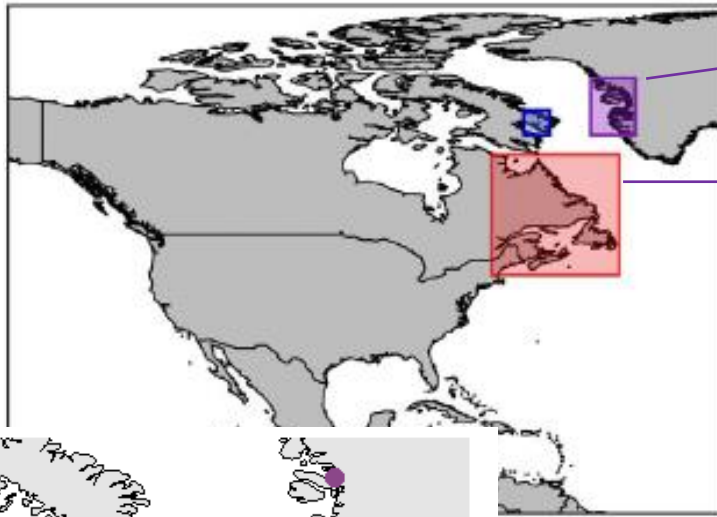
N fastQ files

# Dummy genome

Smaller genome :  
5 chromosomes

We aligned fastq files = the raw reads on that dummy reference genome

⇒ BAM files that you will play with in STACKS.

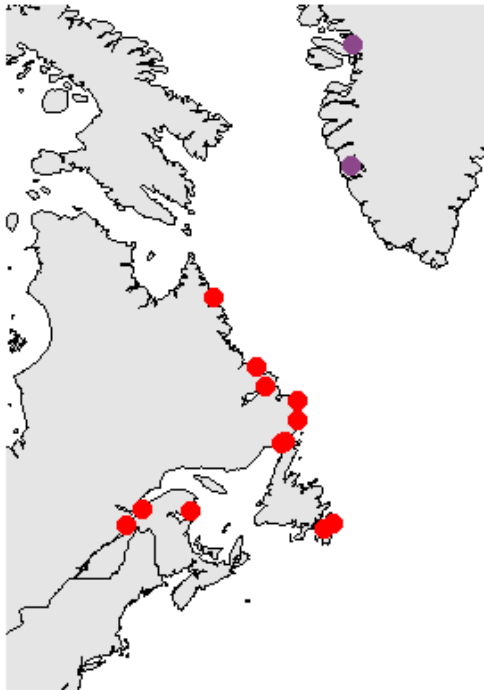


Greenland species

North American species

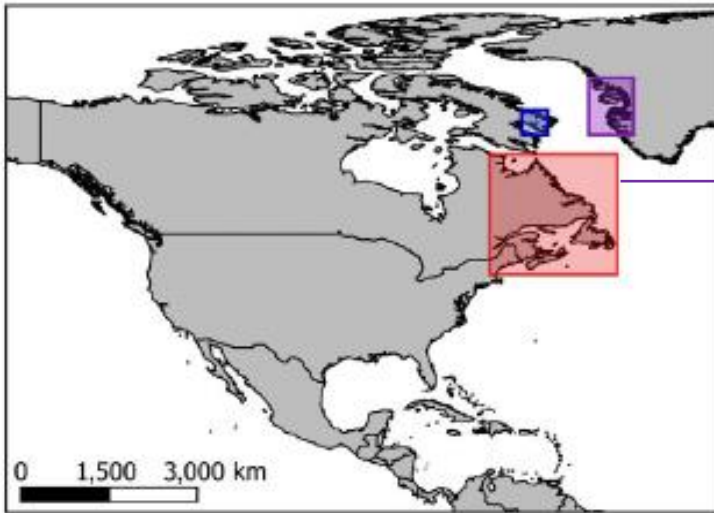
2 population  
N= 40

12 populations  
N= 240 ( 20/pop)



Dataset 1	Dataset 2	Dataset 3
« 2_lin »	« all »	« canada »
4 populations (2 greenland /2 canadian) => 80 samples	14 populations (2 greenland /12 canadian) => 280 samples	12 populations (12 canadian) => 240 samples
Fst (vcftools) PCA	Faststructure DAPC	PCA DAPC
Optional (Fst with Stacks)		Optional (Pairwise Fst)
		-> ALL analyses of day 3-day4-day5

# For thursday: whole-genome sequencing



North American species

12 samples from  
different canadian  
populations

Whole-genome sequencing = much bigger files  
BUT useful for SV detection or for a higher density of SNPs

Here we pick a very reduced dataset to make things run fast!!