Plant & Food Research
Rangahau Ahumāra Kai

Investigating Hoki Stock Structure Using Genomics

Maren Wellenreuther, Emily Koot, Elena Hilario, Igor Ruza, Roy Storey, Chen Wu and David Chagné, Plant and Food Research

Talk outline

1. **The question**
2. **Data generation**
   - Sites sampled
   - Sample extraction, genotyping success, and genome assembly
3. **Results & Discussion**
   - Genetic diversity
   - Stock structure
4. **Next steps**

1. The question

The question we are trying to answer

How many sub-populations of hoki are there?

- How different are they?
- Do juveniles mix but adults segregate as per model assumptions?

We are applying a genomics approach to answer these questions

The two stock model

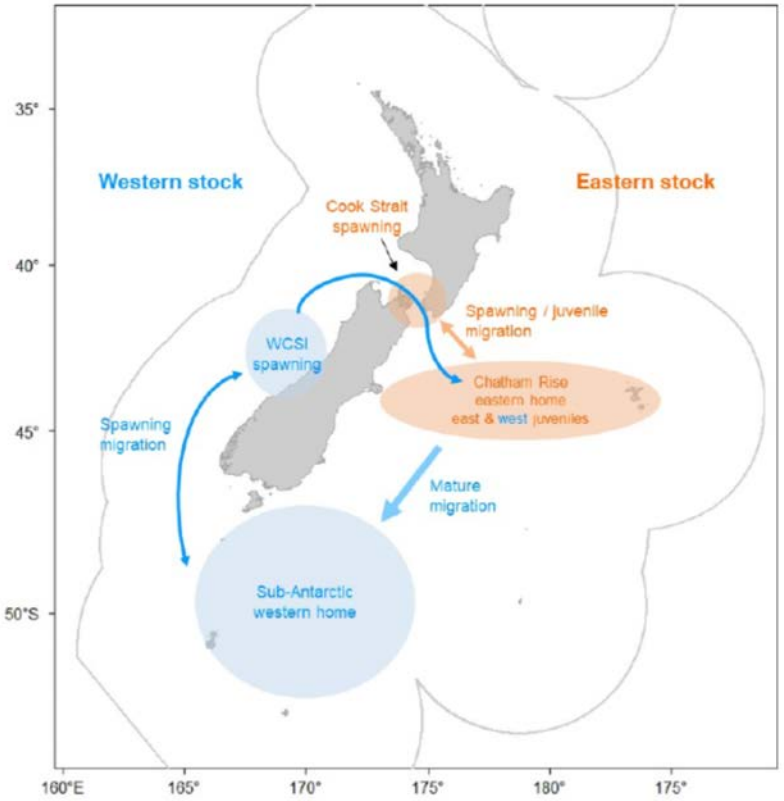
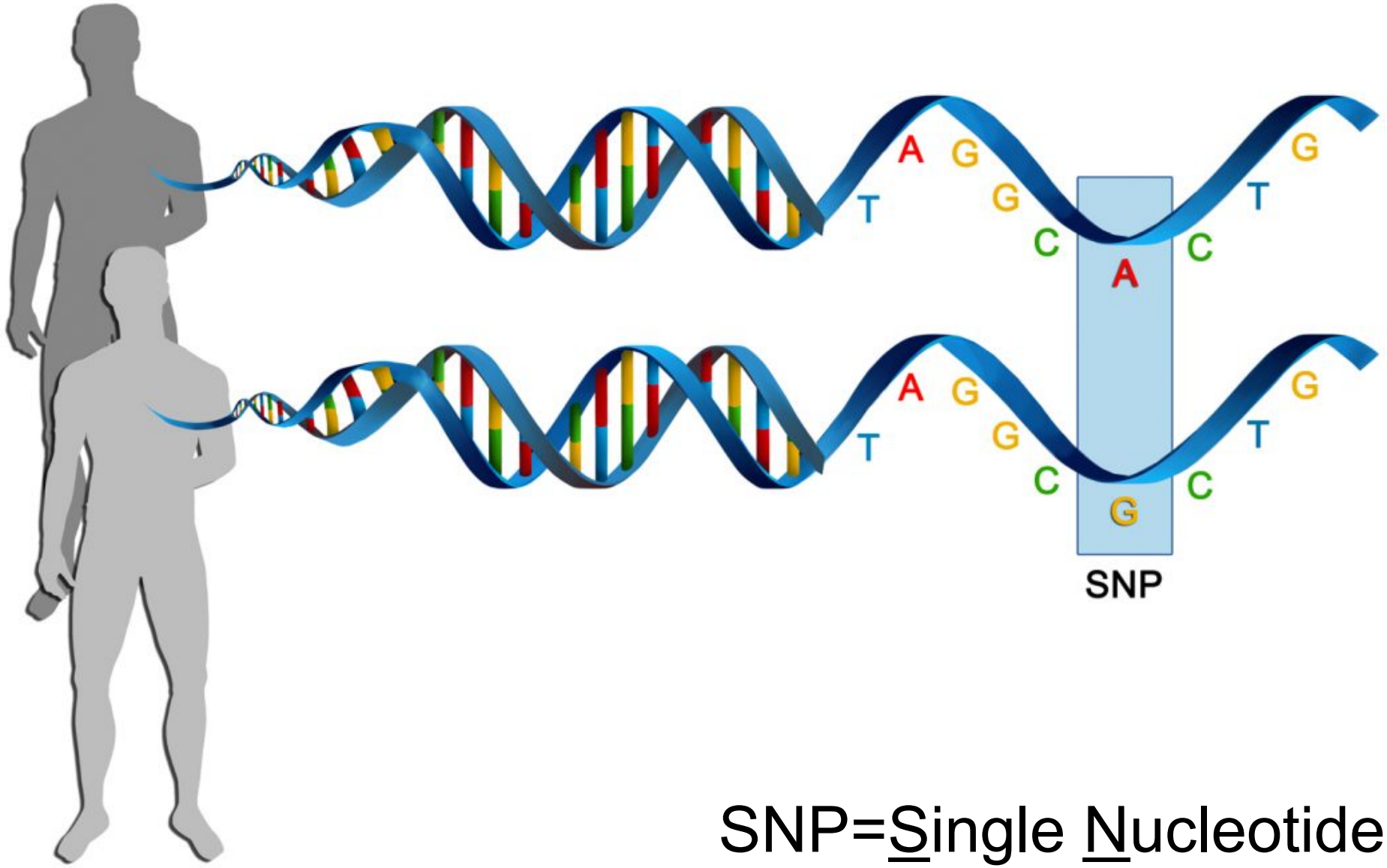


Figure 1: Hoki juvenile nurseries, spawning grounds and migration routes for the eastern and western stocks.

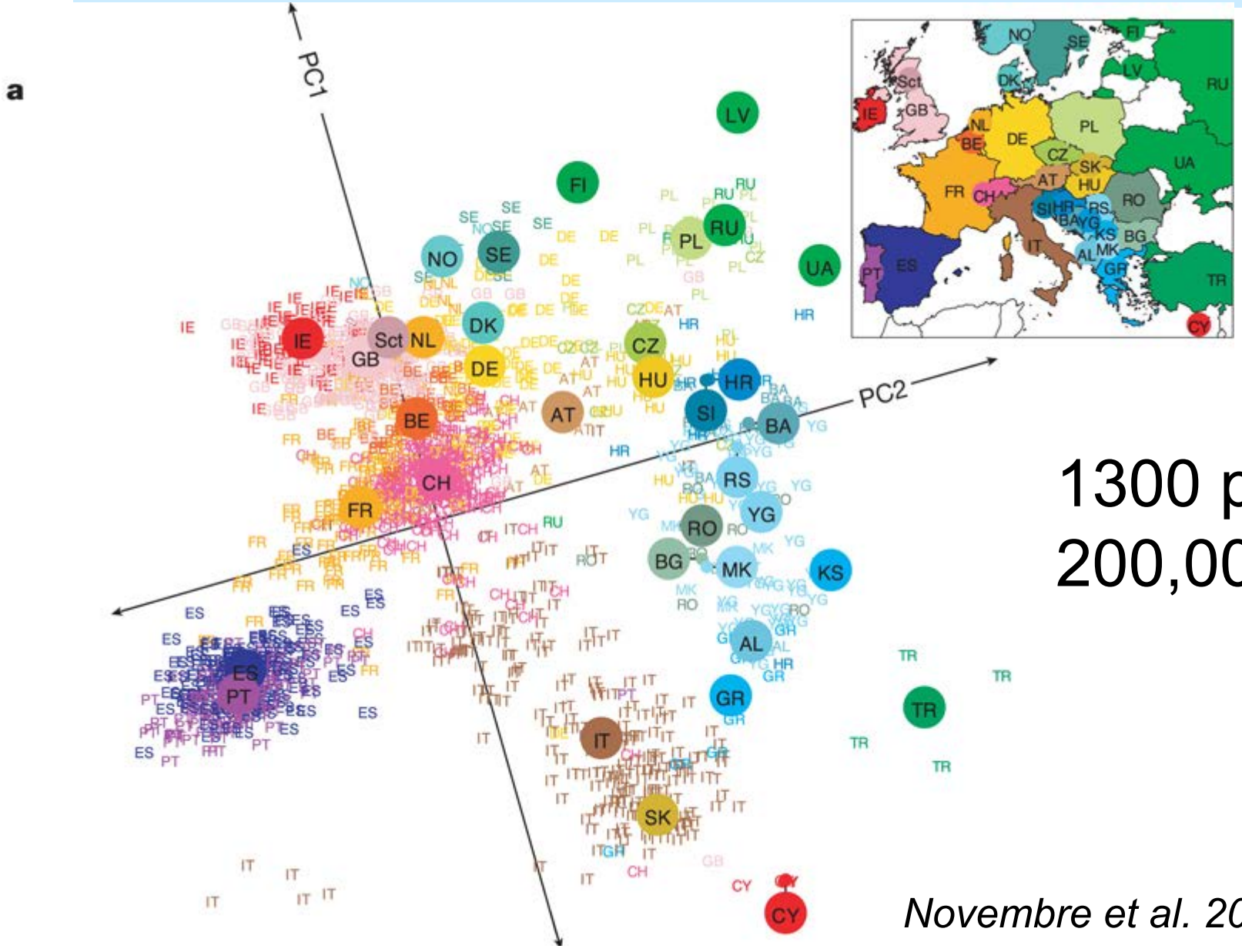
Fisheries New Zealand Plenary document

The data to answer this question: SNPs



SNP=Single Nucleotide Polymorphisms

SNPs mirror geography in Europe

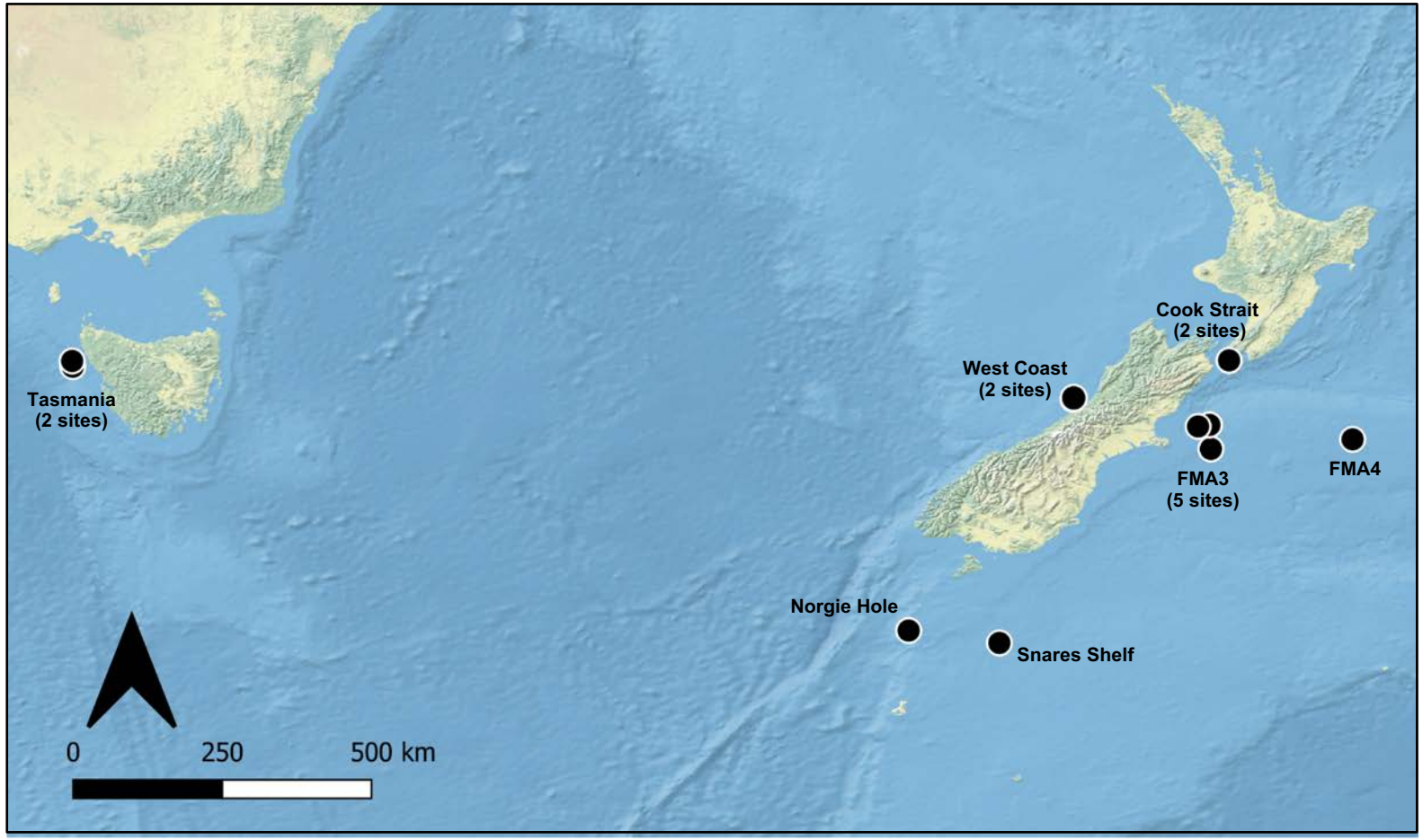


1300 people across Europe
200,000 SNPs

Novembre et al. 2008, Nature

3. Hoki data

Sampling sites analysed



- Samples were collected by fishing boats
- 14 sampling sites
  - Totalling 510 individuals
  - 2x sampling sites for juveniles (FMA3)
  - Including 60 from Tasmania

Sample extraction and genotyping success

- We had a 100% success rate for the DNA extractions
- Frozen samples provide high quality DNA
- We had 100% success rate for the sequencing
  - Whole Genome Sequencing (WGS with Novaseq) (genome size is 600-700 MB)
- Obtained data for **all** 510 hoki individuals
- **Workflows are very reliable and produce high-quality data**

Hoki reference genome

- One sample of a female fish chosen as a reference (Cook Strait)
- Genome assembled using a combination of short and long read sequencing technology
- High quality assembly produced
- This genome will provide a valuable resource for all future work

Hoki reference genome

- 22.2Gb data that needed to be assembled into a spatial organisation=the genome
- The assembly is in 589 fragments
- Assembling into 21 chromosomes in progress: will allow us to see if our markers are evenly distributed
- Comparison of scaffold lengths: N50 scaffold length is 11 MB, for comparison
  - Chinook salmon its is 2MB
  - Atlantic cod 28MB
- A high quality genome allows us to find variants across entire genome as well as with higher accuracy

	Data & assembly statistics
Data depth	44X
Total number	589
Total size	501,547,618
Longest	25,972,667
Shortest	1,004
No. > 10kb	168
No. > 100kb	114
No. > 1Mb	69
No. > 10Mb	18
N50	11,052,189

Christensen et al. (2018) Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. PLOS ONE 13(4)

Assembly completeness

- BUSCO using the conserved vertebrate gene set 95.5%
- BUSCO using the conserved Actinopterygii gene set 90.8%
- This is comparable to the BUSCO scores for
  - Atlantic Cod – 93.2%
  - Chinook Salmon – 90.3%

DNA sequence alignment

Data from 510 individuals aligned to the genome to call genetic variants (SNPs)
Totalling **179,181** SNPs

4. Results & Discussion

Summary Stats

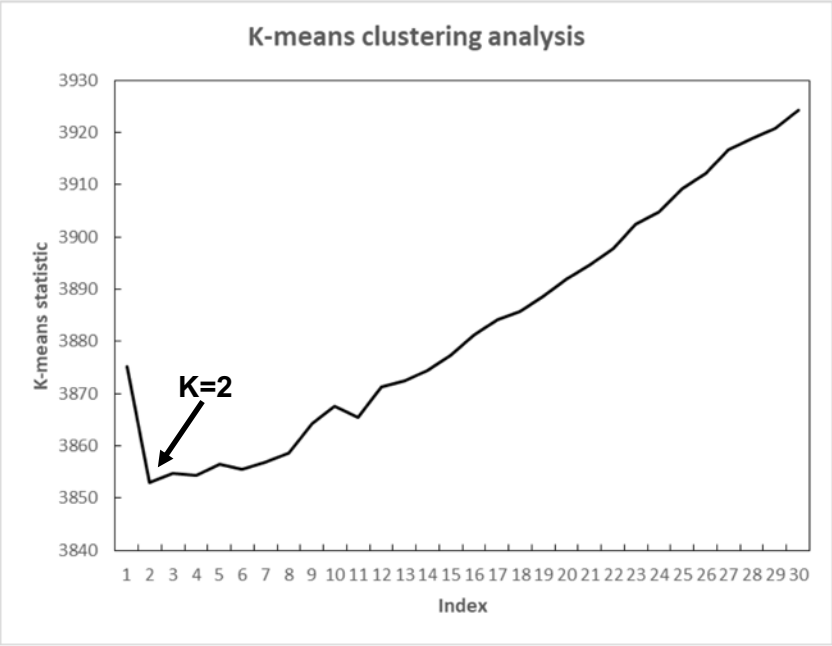
1. Moderate genetic diversity (π~12-16%) for NZ and Tasmanian sites

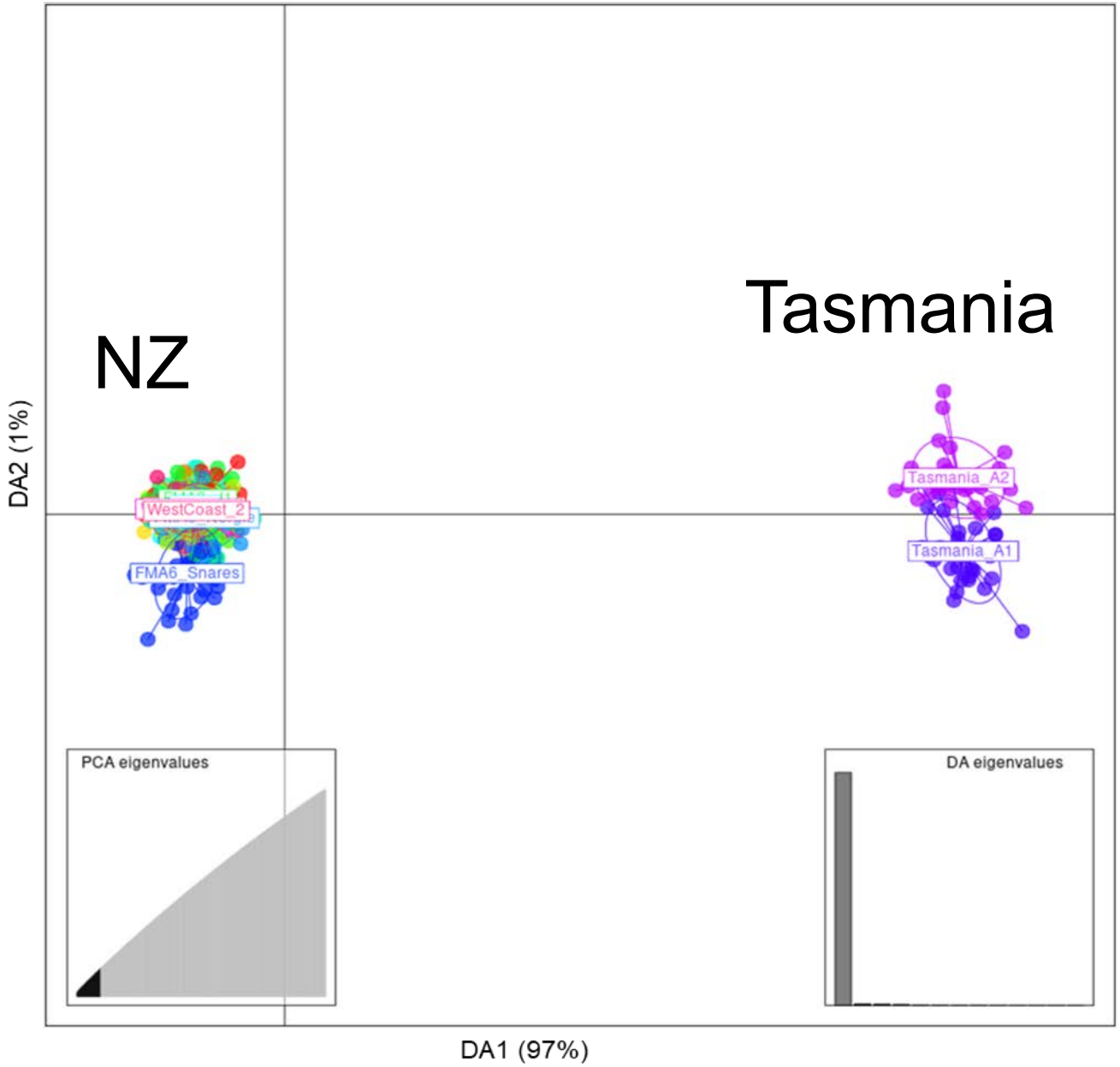
2. Moderate-high Het values (>0.1) indicating large stable sampling sites with high migration rates

3. Negative Tajima's D: sites have few rare variants – may be a result of recovery from contraction/bottleneck

Sampling sites	π (mean)	Het obs (exp)	Tajima's D
CookStrait_spawning_trawl1	0.132	0.137 (0.130)	-0.707
CookStrait_spawning_trawl2	0.128	0.127 (0.126)	-0.753
FMA3_adult_trawl1	0.130	0.132 (0.129)	-0.725
FMA3_adult_trawl2	0.128	0.129 (0.126)	-0.748
FMA3_juvenile_trawl1	0.132	0.138 (0.131)	-0.699
FMA3_juvenile_trawl2	0.128	0.127 (0.126)	-0.749
FMA4_adult	0.127	0.125 (0.125)	-0.760
WestCoast_spawning_trawl1	0.129	0.130 (0.127)	-0.743
WestCoast_spawning_trawl2	0.129	0.130 (0.127)	-0.744
Tasmania_adults1	0.168	0.203 (0.165)	-0.354
Tasmania_adults2	0.153	0.178 (0.151)	-0.506
FMA3_adult_trawl3	0.129	0.129 (0.126)	-0.801
FMA6_adults_Norgie	0.128	0.127 (0.126)	-0.805
FMA6_adults_Snares	0.132	0.135 (0.130)	-0.761
NZ	0.131	0.132 (0.130)	-0.0741
Tasmania	0.161	0.191 (0.160)	-0.2529

Do we find clusters?

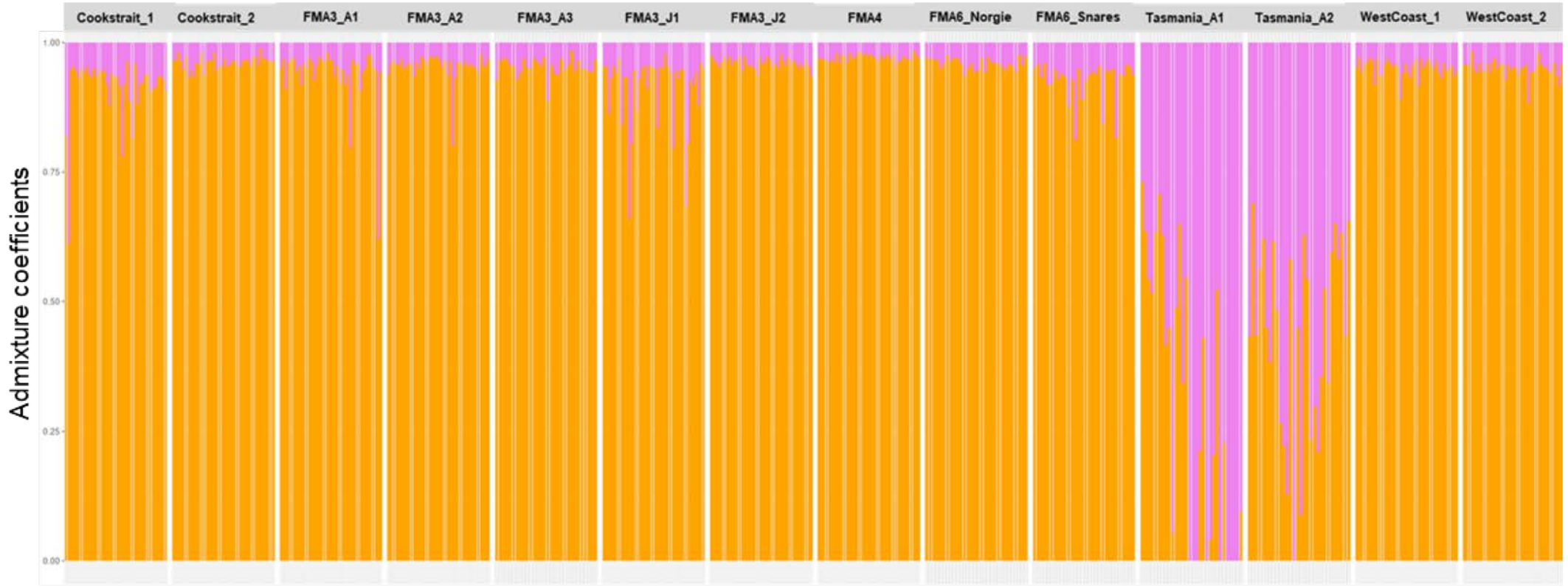




Clustering analysis and Discriminant Analysis of Principal Components (DAPC)

Do hoki from NZ and Tasmania originate from the same source?

- Identification of number of source ancestral populations from which hoki are derived
- Shared colours indicate shared ancestry between individuals/populations and subsequent genetic structure

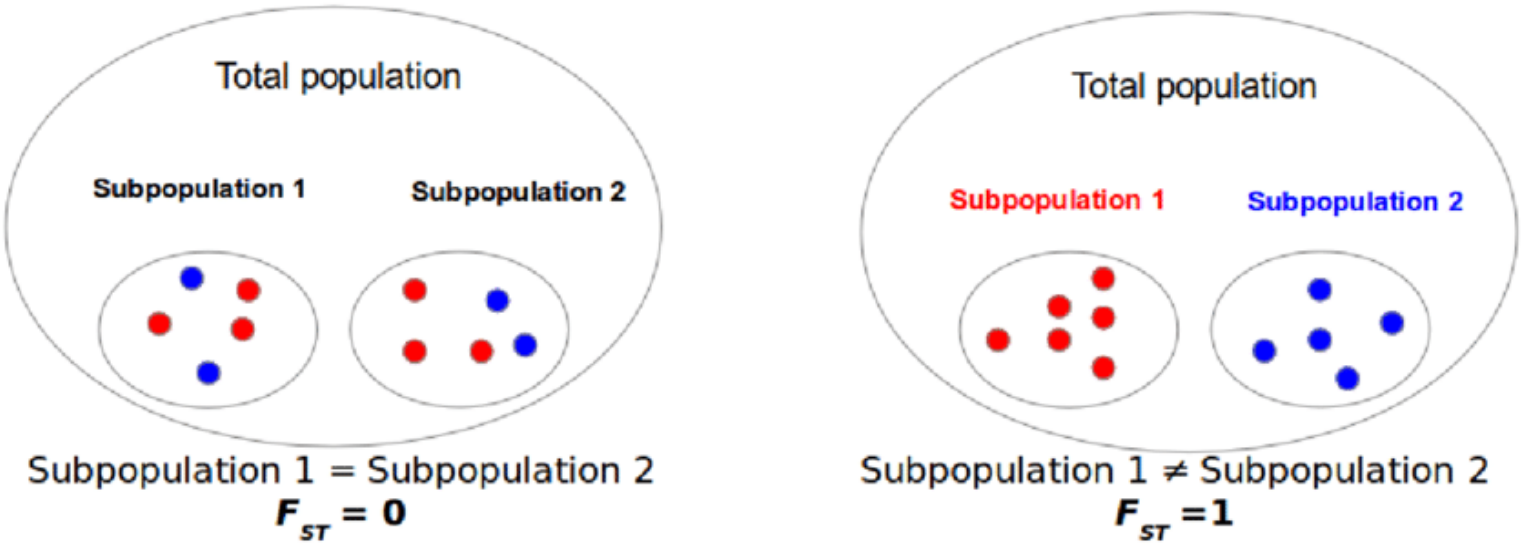


LEA (sNMF) Ancestry Admixture estimation

Genetic differentiation between populations F_{ST}

Fixation Index (F_{ST}): a measure of population differentiation

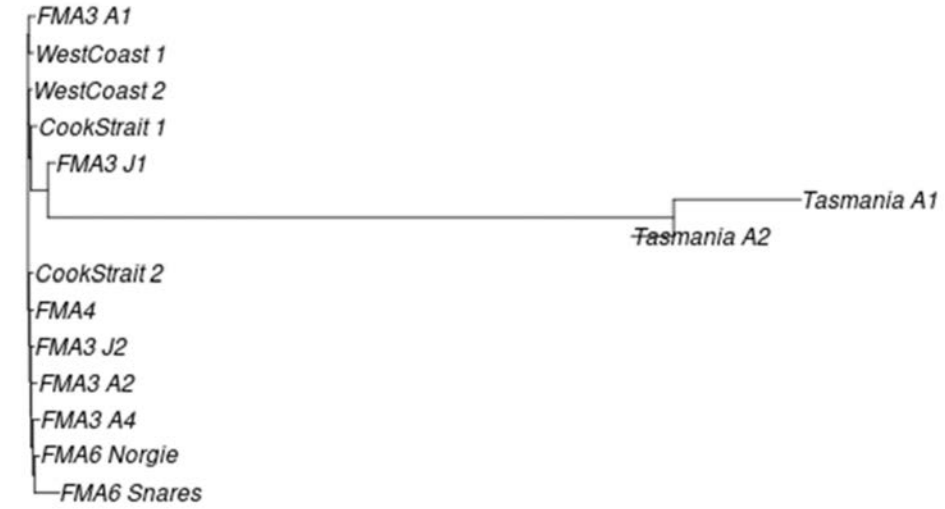
>10% migration rate to achieve F_{ST} 0.0001 – 0.0003

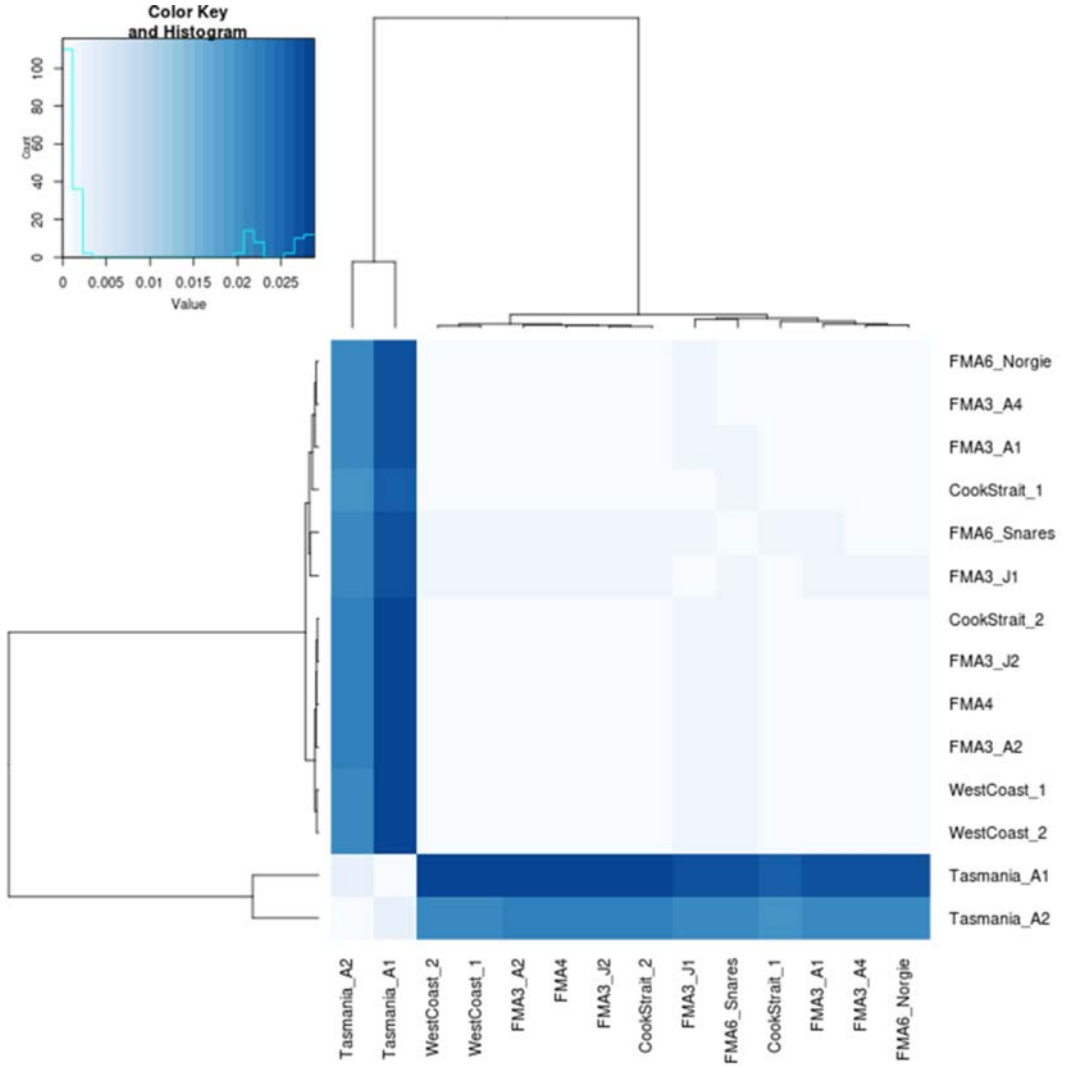


Genetic differentiation – F_{ST}

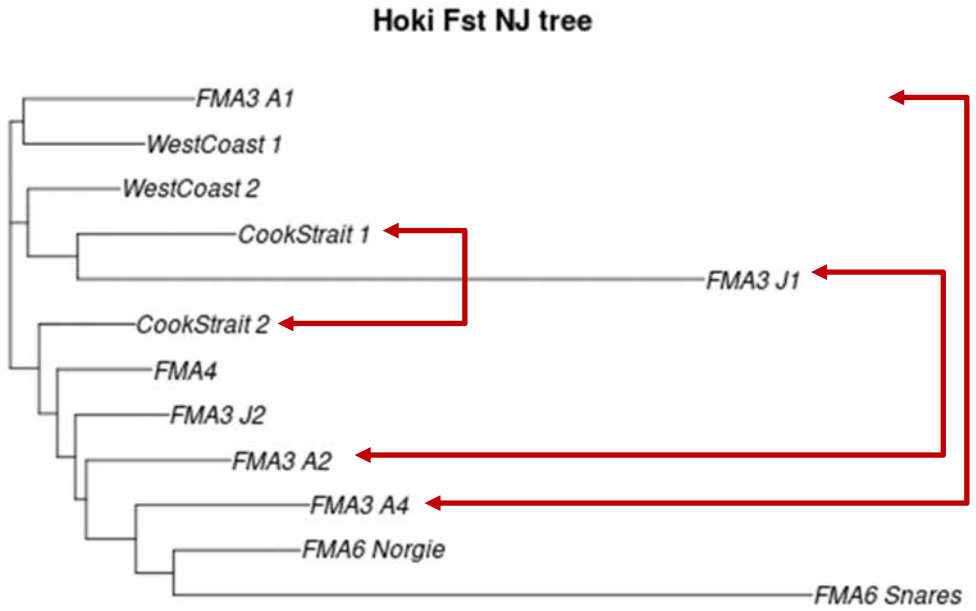
	NZ												
Tasmania	0.0268	Cook Strait 1	Cook Strait 2	FMA3 A1	FMA3 A2	FMA3 A3	FMA3 J1	FMA3 J2	FMA4	FMA6_Norgie	FMA6_Snares	Tasmania A1	Tasmania A2	West Coast 1
Cook Strait 2	0.0005												
FMA3 A1	0.0005	0.0004											
FMA3 A2	0.0005	0.0004	0.0005										
FMA3 A3	0.0006	0.0005	0.0007	0.0005									
FMA3 J1	0.0011	0.0012	0.0012	0.0012	0.0012								
FMA J2	0.0006	0.0003	0.0005	0.0004	0.0005	0.0012							
FMA4	0.0006	0.0003	0.0005	0.0004	0.0004	0.0012	0.0003						
FMA6_Norgie	0.0007	0.0005	0.0006	0.0005	0.0005	0.0012	0.0004	0.0005					
FMA6_Snares	0.0013	0.0013	0.0012	0.0013	0.0011	0.0018	0.0013	0.0014	0.0011				
Tasmania A1	0.0262	0.0284	0.0272	0.0283	0.0273	0.0266	0.0284	0.0288	0.0276	0.0266			
Tasmania A2	0.0205	0.0221	0.0213	0.0221	0.0214	0.0210	0.0222	0.0223	0.0216	0.0212	0.0030		
West Coast 1	0.0005	0.0003	0.0004	0.0005	0.0007	0.0012	0.0005	0.0004	0.0007	0.0014	0.0281	0.0219	
West Coast 2	0.0004	0.0004	0.0004	0.0005	0.0006	0.0012	0.0004	0.0003	0.0006	0.0012	0.0280	0.0218	0.0003

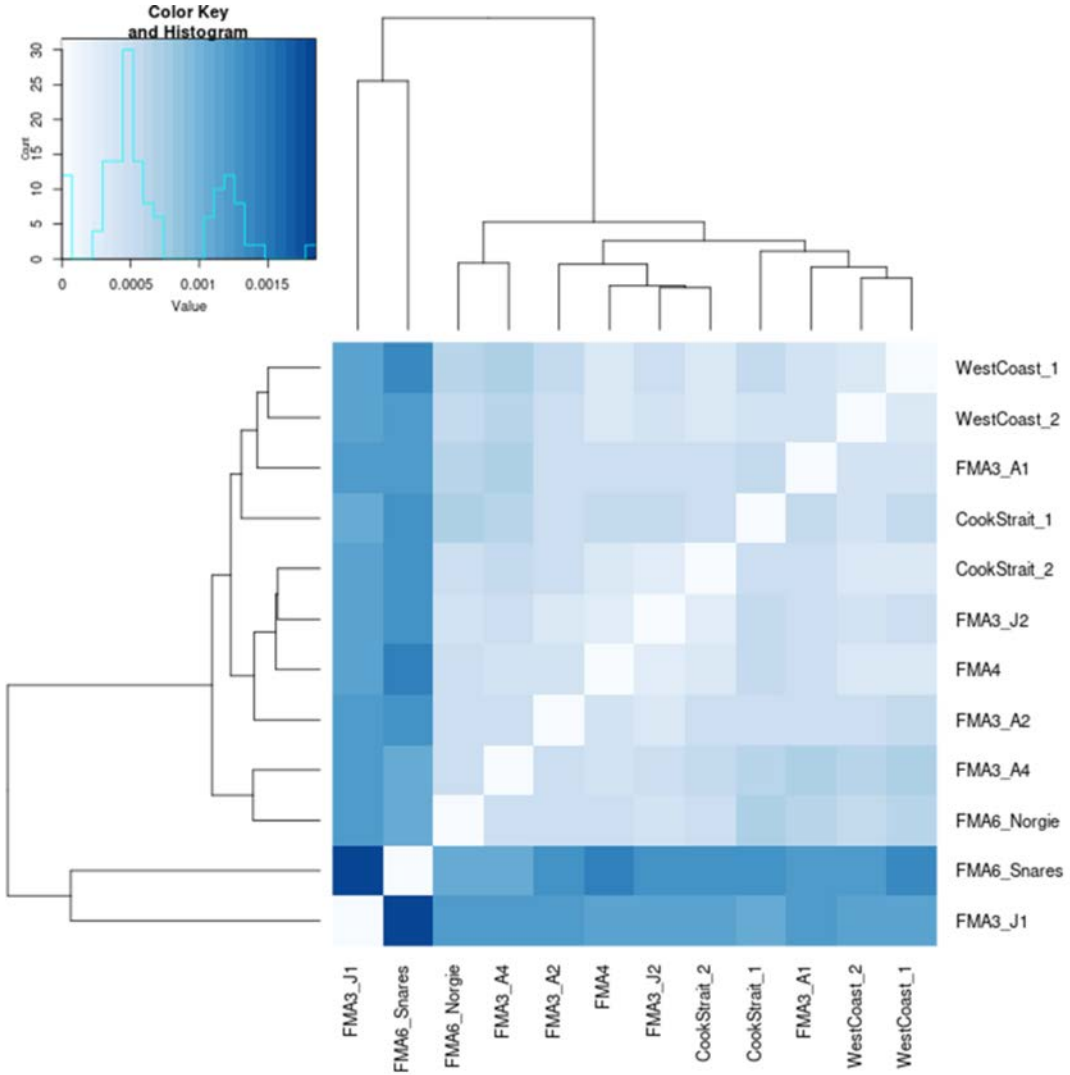
Heat map and NJ tree based on F_{ST}





Heat map and NJ tree based on F_{ST} – NZ only





Questions

1. Is there a genetic difference between adult hoki spawning contemporaneously in Cook Strait and on WCSI?

2. Do the genetic data provide evidence that juvenile hoki spawned on WCSI occur on the western Chatham Rise?

3. Do the genetic data provide evidence that adult hoki in the Sub Antarctic are consistent with spawning hoki on WCSI and not consistent with spawning hoki in Cook Strait and Pegasus?

4. Do the genetic data provide evidence that adult hoki on the Chatham Rise are consistent with spawning hoki in Cook Strait/Pegasus and not consistent with spawning hoki in WCSI?

Questions

1. Is there a genetic difference between adult hoki spawning contemporaneously in Cook Strait and on WCSI?
    - » Clustering analyses indicate one population, however there are signs there is minor genetic differentiation
    - » Adaptive SNPs may help us tease out this signal

Questions

2. Do the genetic data provide evidence that juvenile hoki spawned on WCSI occur on the western Chatham Rise?
    - » Data indicate juveniles from Chatham Rise are very genetically similar with WCSI, with DAPC unable to differentiate two sites

Questions

3. Do the genetic data provide evidence that adult hoki in the Sub Antarctic are consistent with spawning hoki on WCSI and not consistent with spawning hoki in Cook Strait and Pegasus?
   - » Clustering indicates one admixed population, with similar genetic relatedness between Cook Strait & West Coast with Snares shelf sites.
   - » Possible mixing with external population
   - » We need to analyse the Tangaroa samples (Sub-Antarctic)

Questions

4. Do the genetic data provide evidence that adult hoki on the Chatham Rise are consistent with spawning hoki in Cook Strait/Pegasus and not consistent with spawning hoki in WCSI?
    - » Hoki on Chatham Rise are genetically very similar to both Cook Strait and WCSI population

5. Next steps

Take home messages from this talk

- Data workflow is very reliable and achieves high quality data
- Genome of high quality-valuable resource for the future
- Tasmanian sites are clearly different from all NZ sites
- Proof that the method works

Take home messages

- NZ hoki sites show generally low F_{ST} and high connectivity, and significant exchange of individuals

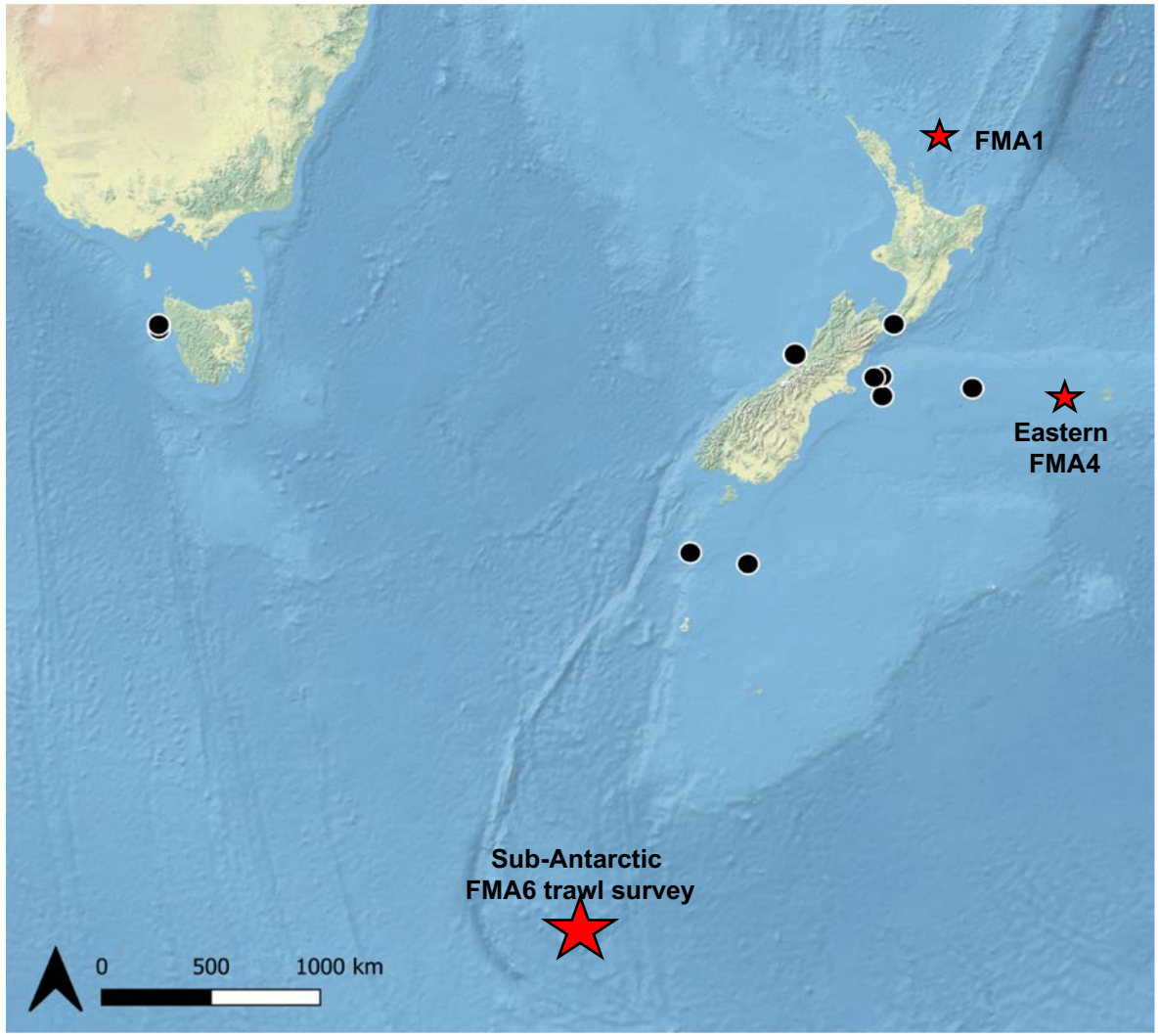
» No clear differences between adults and juveniles, and west and east coast populations

» Temporal and geographical distribution of sampling between WCSI and Cook Strait

Next steps

- Finish reference genome
- Re analyse markers with new genome, check distribution of markers across the genome
- Include adaptive markers in analyses (only neutral now)
- Check if specific SNPs are fixed geographically to assign individuals to geographic origin
- Publication(s)

Additional samples



- Additional samples from all 14 analysed sites are available (around 70 per site)
- Samples indicated by stars are in storage (each site 100 samples)
- Cook Strait samples from 2005 and 2006

Time plan of next steps

- April-Continue engagement with Te Ohu Kaimoana
- April-Start building data repository
- May-Finalise genome and analyses
- June-Final report for DWG
- June-Steering Group meeting
- August-Publication(s)

Plant & Food™ Research
Rangahau Ahumāra Kai

Acknowledgements

All people that collected samples
PFR staff
Observers
Richard Wells
SIL
Adam Langely
DWG (George Clements)
NIWA for collection of hoki samples

plantandfood.co.nz

Maren.Wellenreuther@plantandfood.co.nz

MW