# Copy Number Variants & adaptation
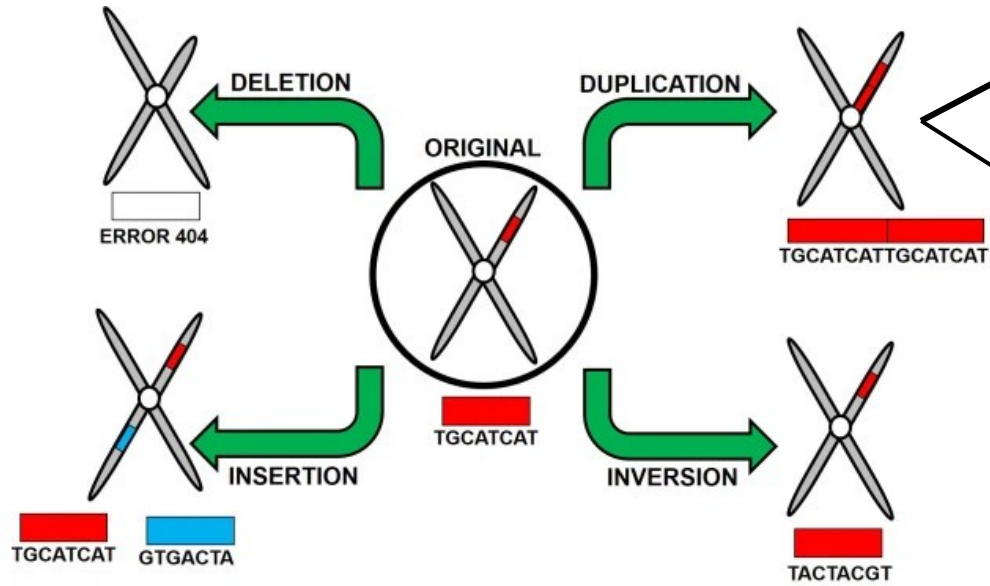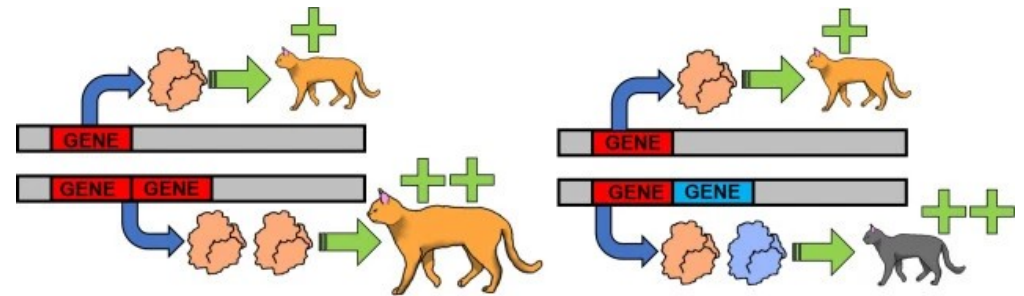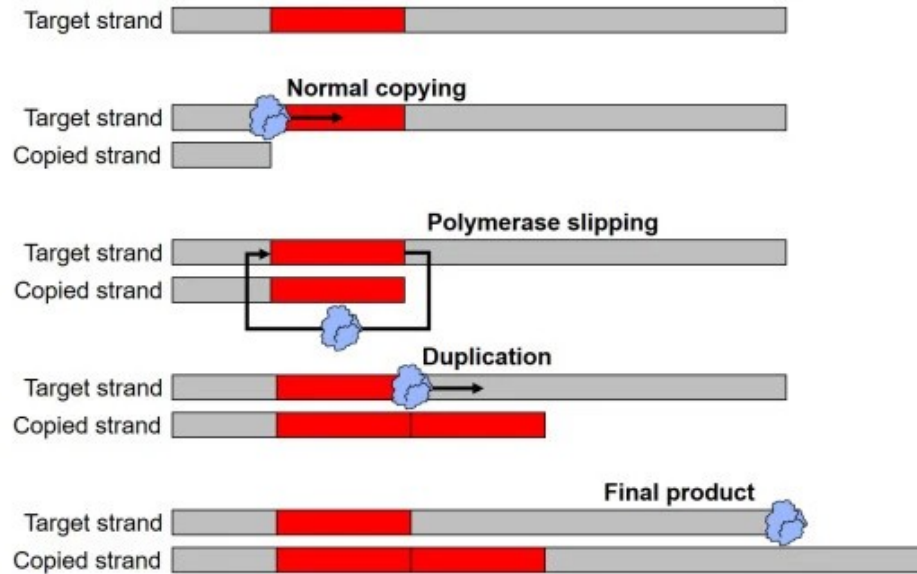
Prepared by Yann Dorant

Physalia courses – Adaptive genomics - 2023

Copy Number Variants (CNVs)

©TheGeneticsCat – Sean Buckley

©TheGeneticsCat – Sean Buckley

**PNAS**
**Polar bear evolution is marked by rapid changes in gene copy number in response to dietary shift**
Rinker et al. 2019

**PNAS**
**Aluminum tolerance in maize is associated with higher MATE1 gene copy number**
Maron et al. 2013

**Molecular Ecology**
**The fire ant social supergene is characterized by extensive gene and transposable element copy number variation**
Fontana et al. 2019

**Molecular Ecology**
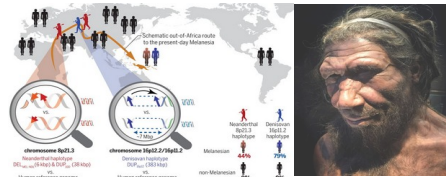**Antifreeze protein gene amplification facilitated niche exploitation and speciation in wolffish**
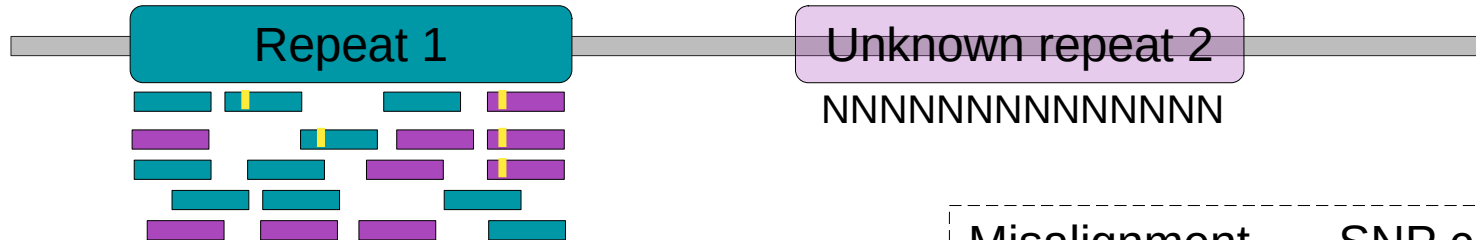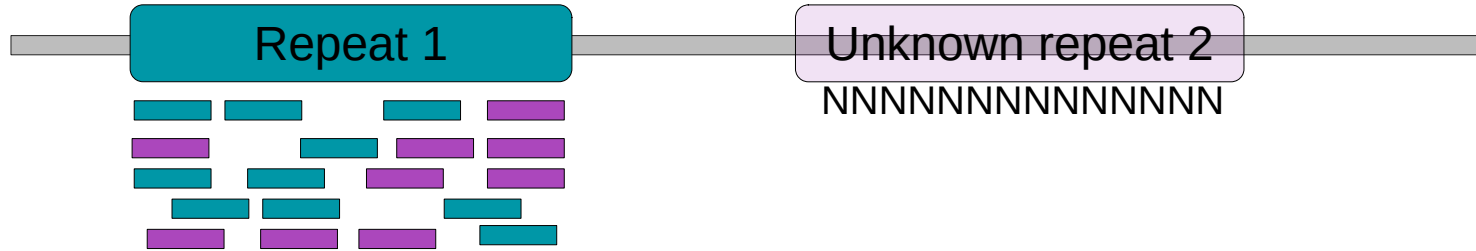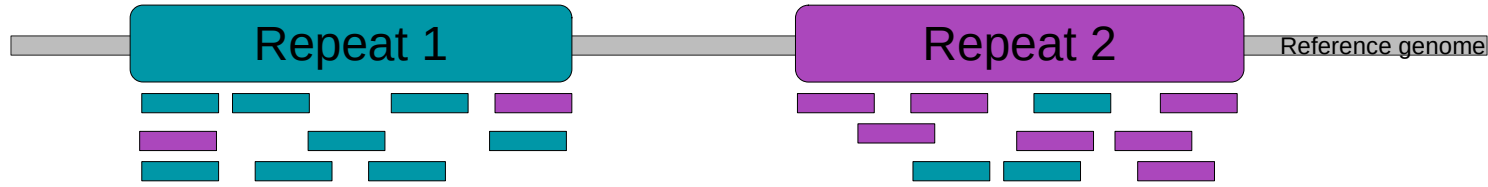Desjardins et al. 2012

E. Bogason

**SCIENCE**
**Adaptive archaic introgression of copy number variants and the discovery of previously unknown human genes**
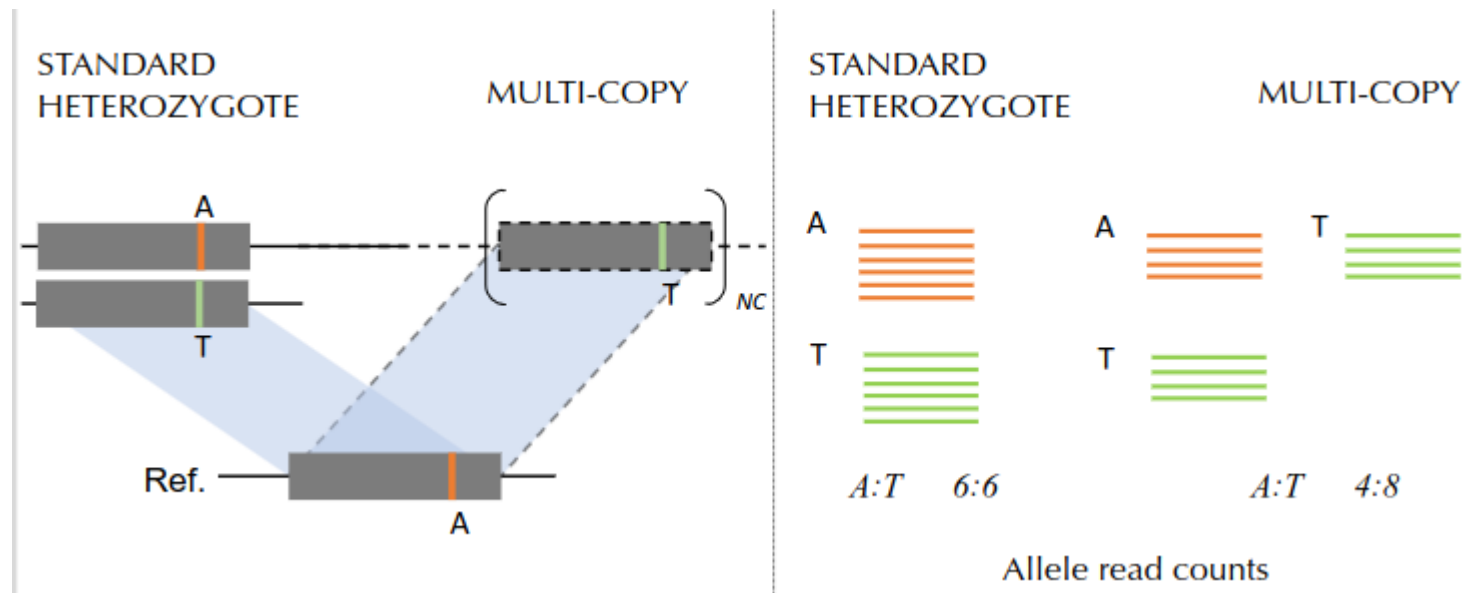Hseih et al. 2019

Adaptation
Resistance
Diseases
Evolutionary
history

RADseq-like

Reference genome

Repeat 1

Repeat 2

Repeat 1

Unknown repeat 2

NNNNNNNNNNNNNN

Repeat 1

Unknown repeat 2

NNNNNNNNNNNNNN

Misalignment → SNP calling artifacts

Repeat 1

Repeat 2

True distribution of SNPs

Karunarathne et al., 2023 (BioRxiv).

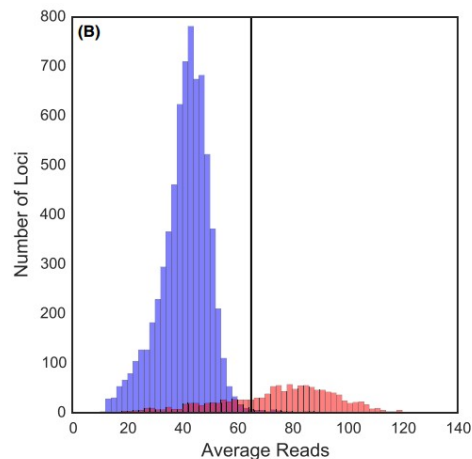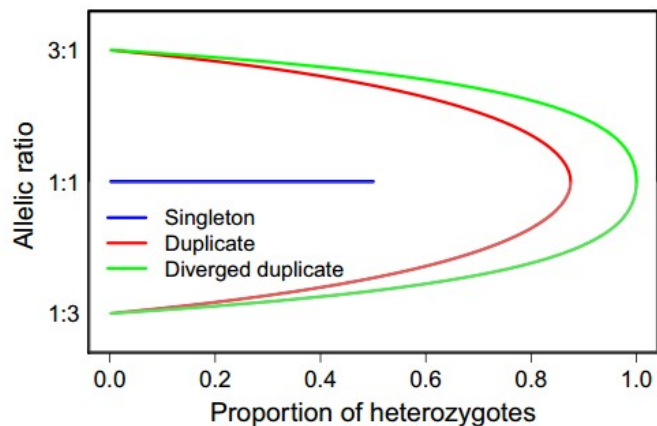Unbalance of allele read ratio mainly observed in heterozygous samples.

# Paralogs are revealed by proportion of heterozygotes and deviations in read ratios in genotyping-by-sequencing data from natural populations

GARRETT J. McKINNEY, RYAN K. WAPLES,[1] LISA W. SEEB and JAMES E. SEEB

*School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle WA 98195-5020, USA*

Saumon Chinook
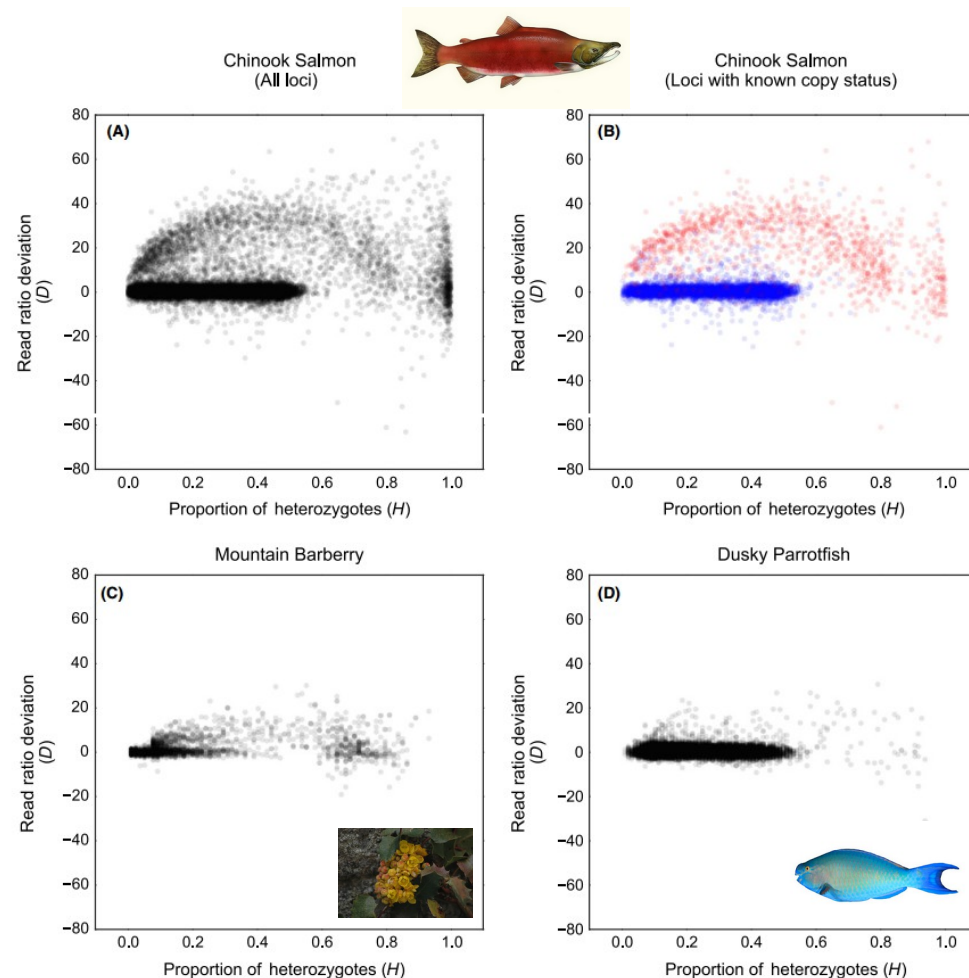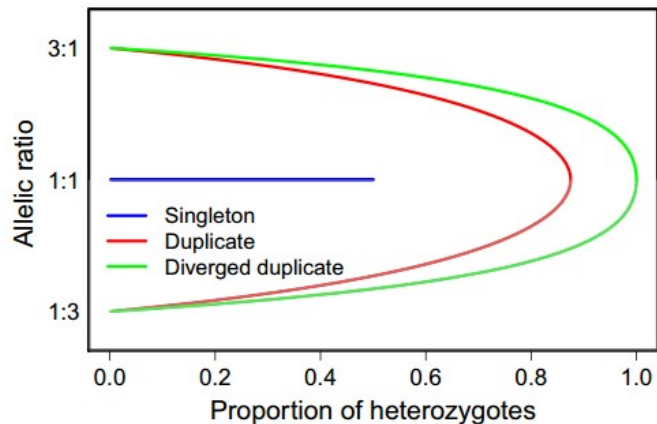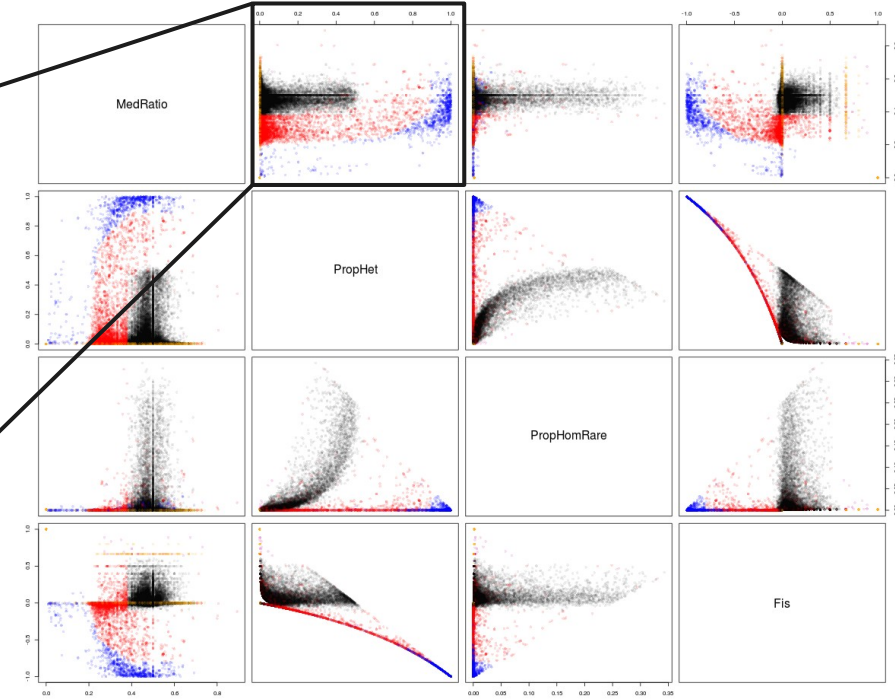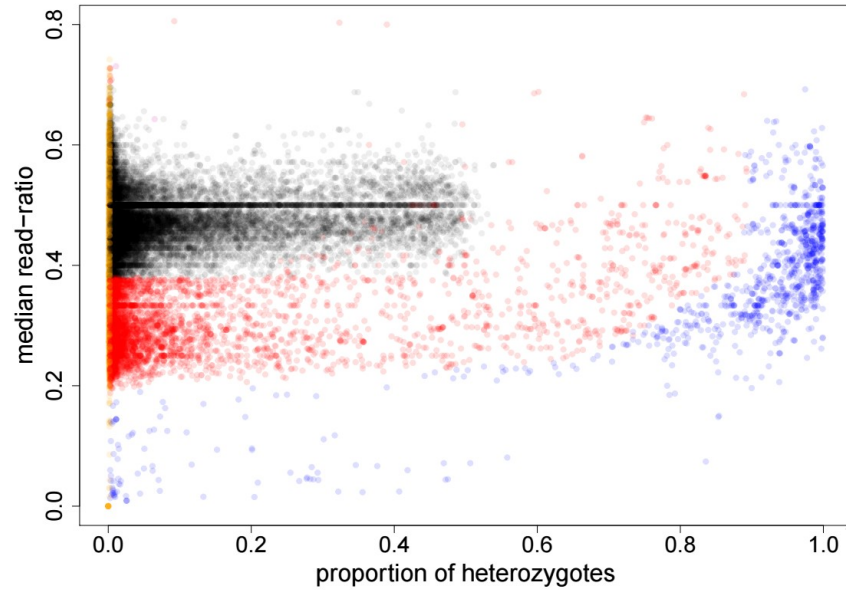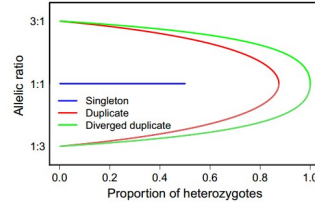
Poisson-Perroquet

*Berberis alpina*

# Paralogs are revealed by proportion of heterozygotes a
deviations in read ratios in genotyping-by-sequencing
from natural populations

GARRETT J. MCKINNEY, RYAN K. WAPLES,[1] LISA W. SEEB and JAMES E. SEEB

*School of Aquatic and Fishery Sciences, University of Washington, 1122 NE Boat Street, Box 355020, Seattle WA 98*
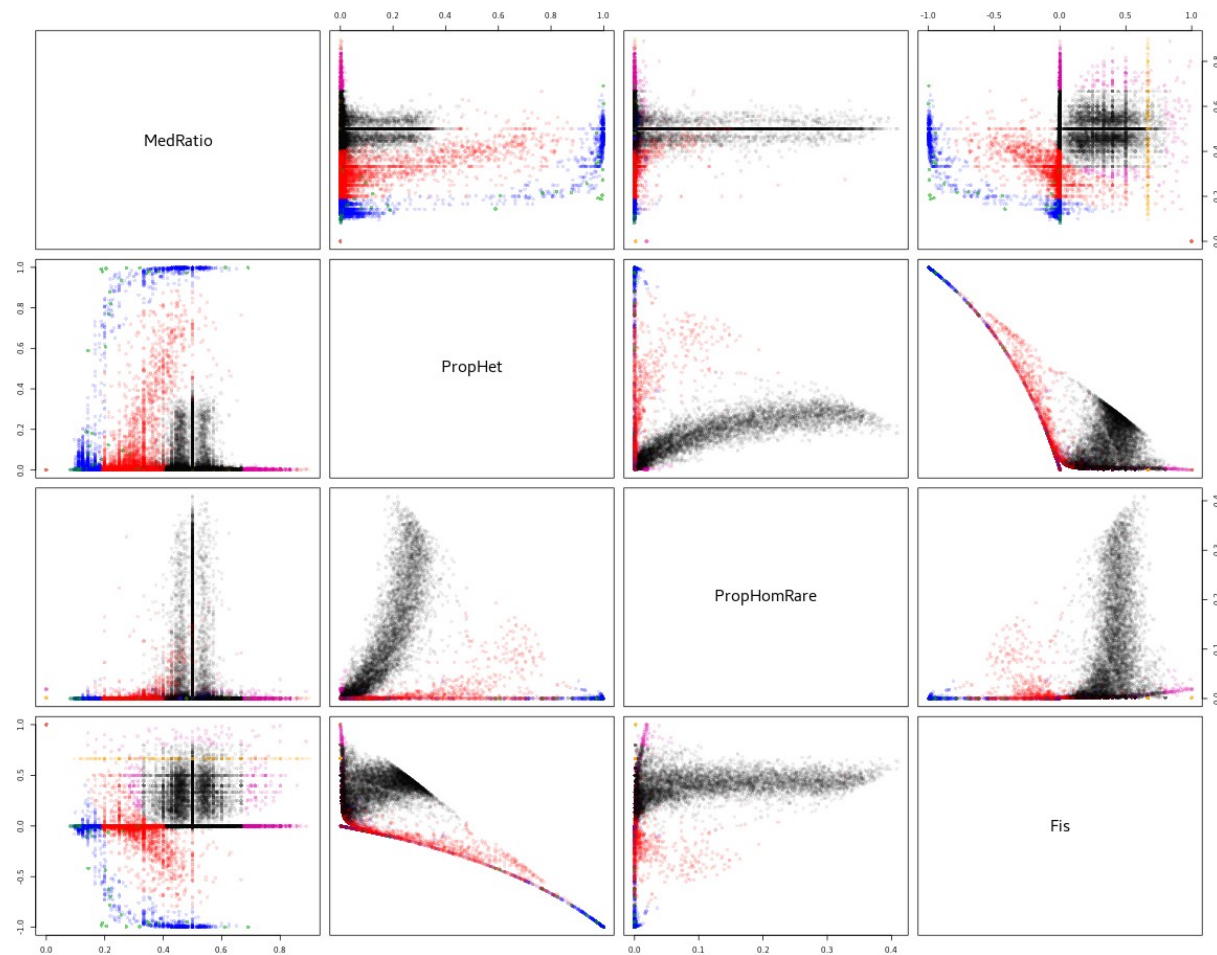*USA*

Legend:
- Singletons (black)
- Duplicated (red)
- Duplicated & diverged (blue)
- Low confidence (yellow)

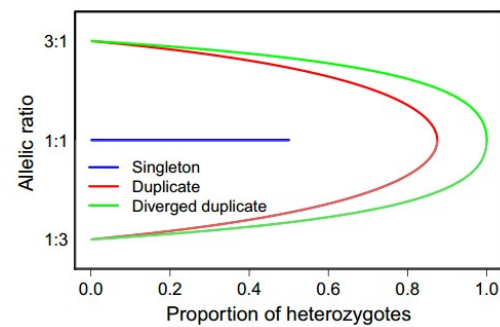→ 13,854 SNPs singletons

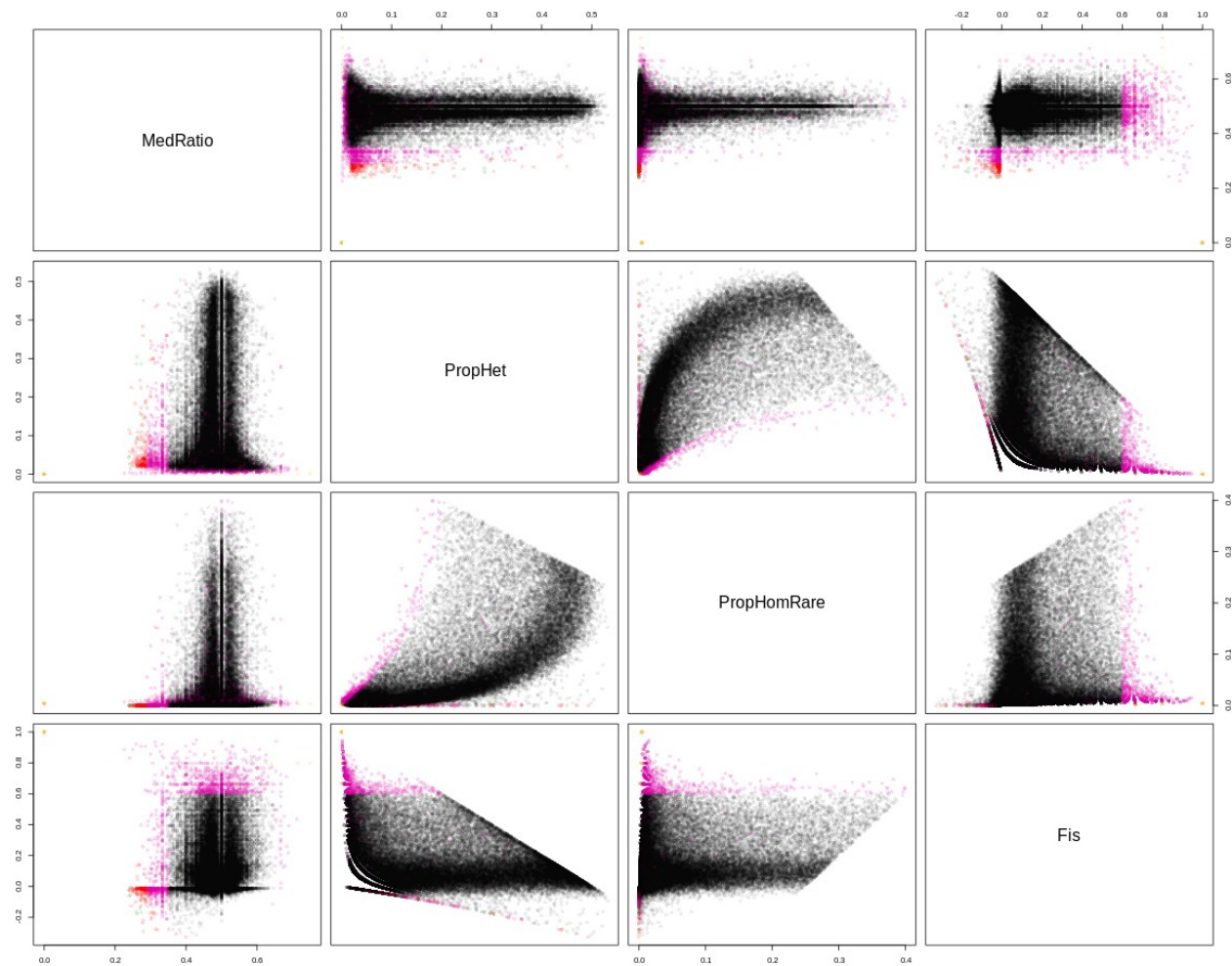→ 1,521 loci duplicated → CNVs
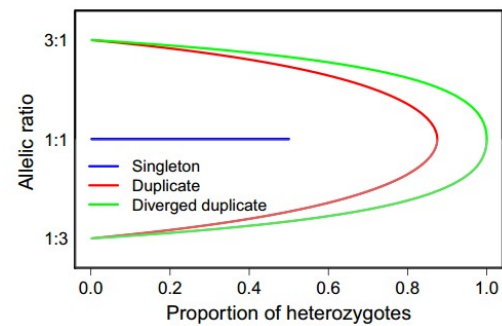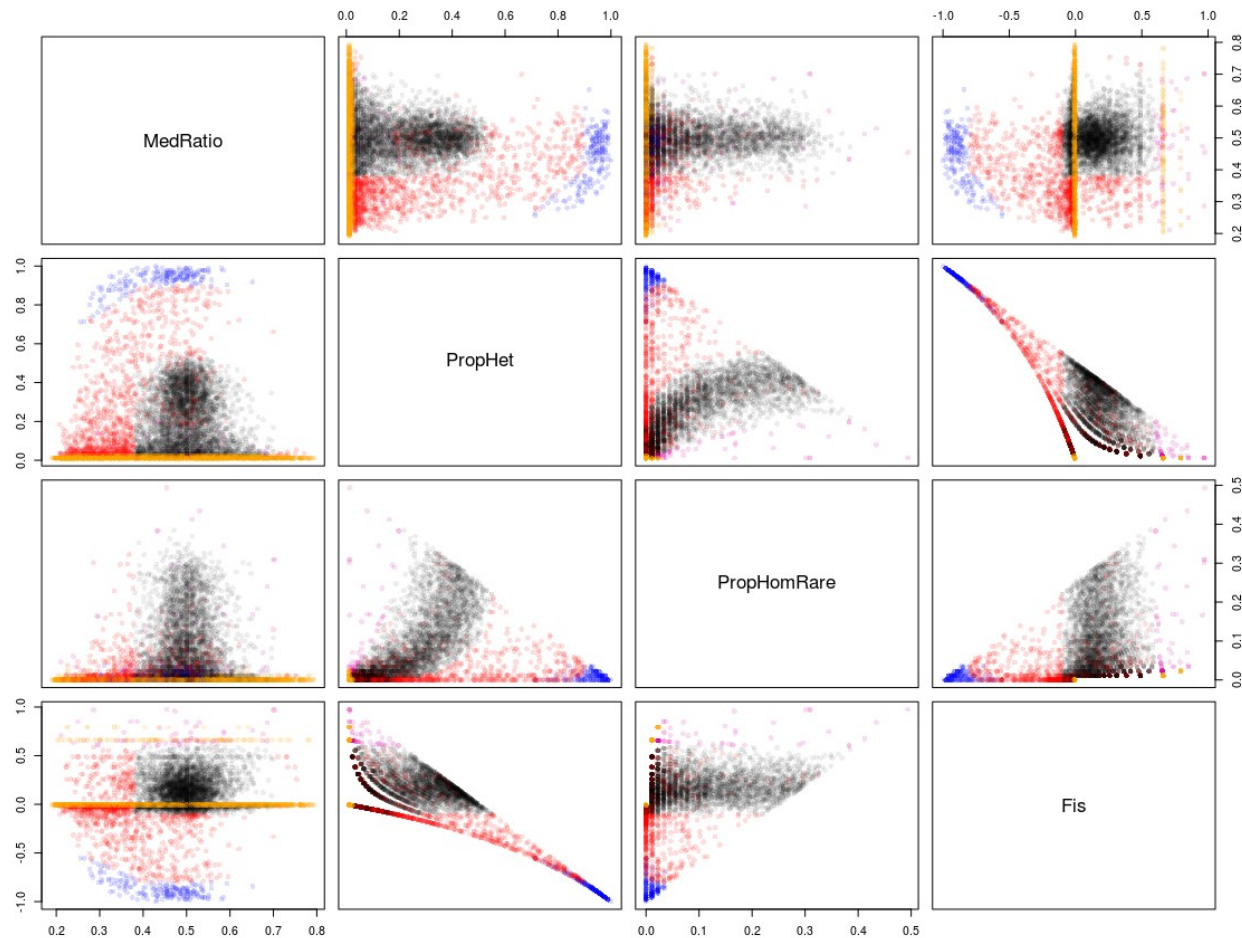
*Salvelinus fontinalis*

Gsize ~ 3.5 Gb

*Reinhardtius hippoglossoides*

Gsize ~ 0.6 Gb

● Singletons
● Duplicated
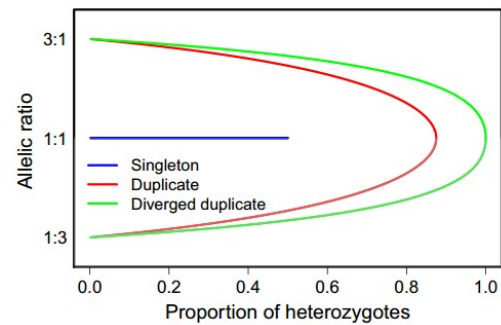● Duplicated & diverged
● Low confidence

*Lithobates sylvaticus*

Gsize ~ 6 Gb

*Filtered VCF from someone who published last oct.2019*

**A new framework for detecting copy number variants from single nucleotide polymorphism data: 'rCNV', a versatile R package for paralogs and CNVs detection**
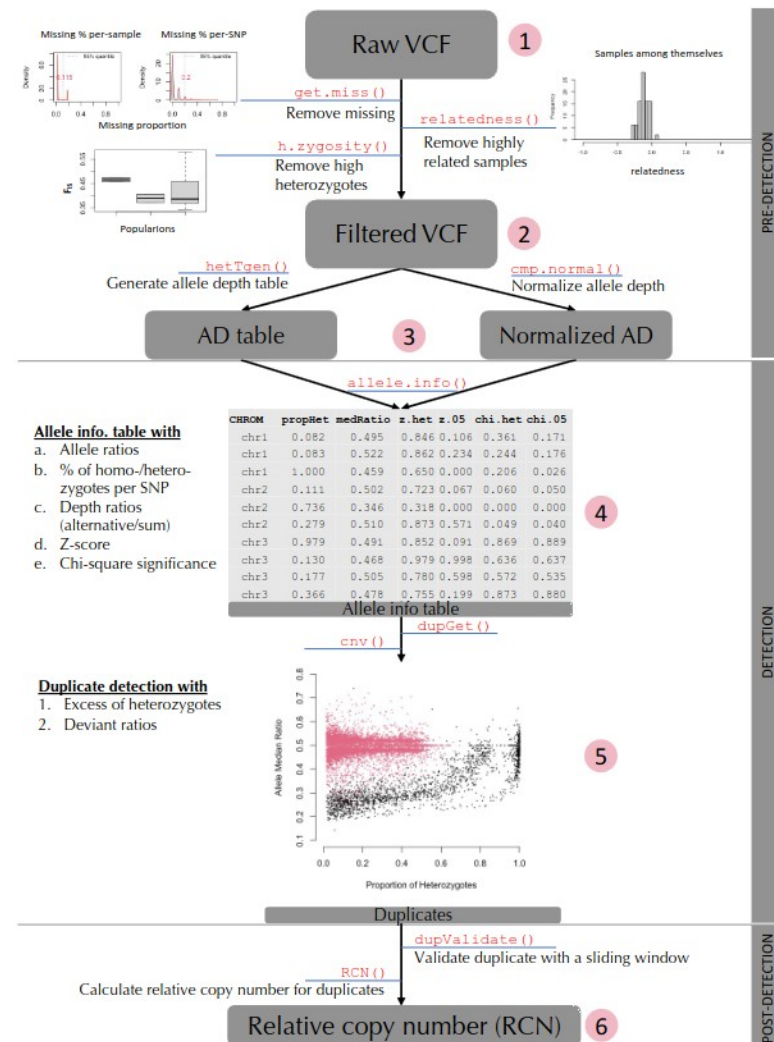
2023

**Running title:** Detecting CNVs from SNPs

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

Piyal Karunarathne[1,2,3]*, Qiujie Zhou[1,2], Klaus Schliep[4], and Pascal Milesi[1,2†]

# A new statistical framework

→ **R package : 'rCNV'**

Sampling map lobsters

**a**
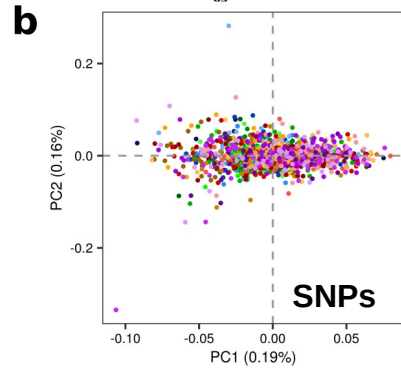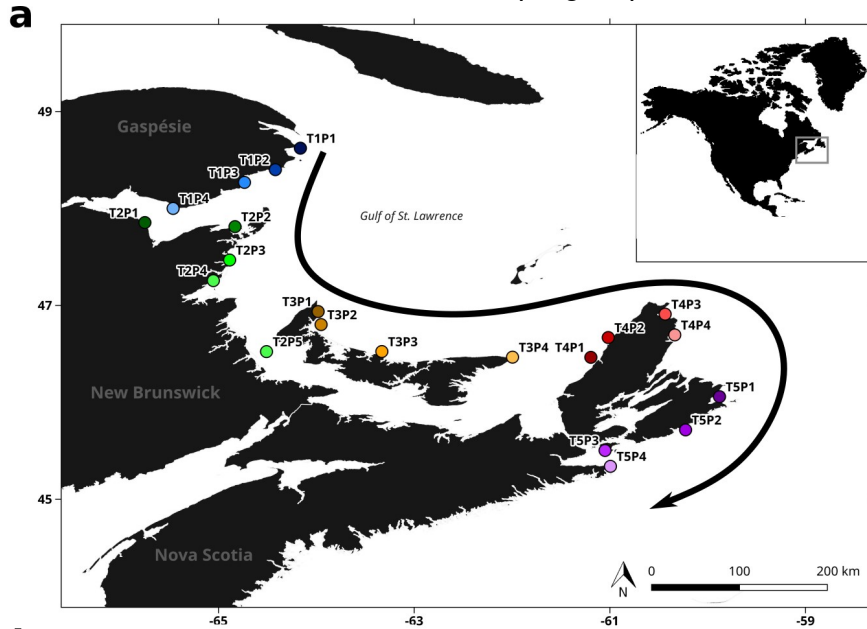
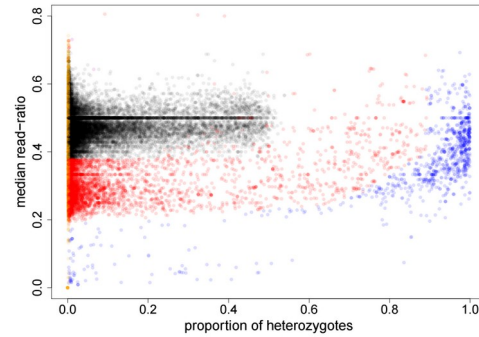**b** SNPs — Genotype info
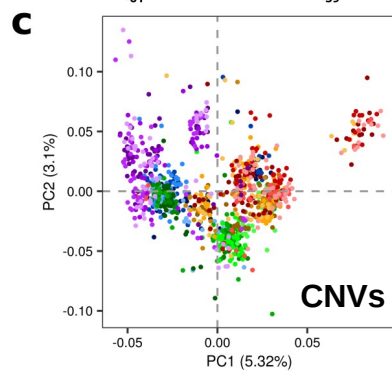
**c** CNVs — Normalized read depth info

**Can we observe patterns of population structure between CNVs and SNPs matrices ?**
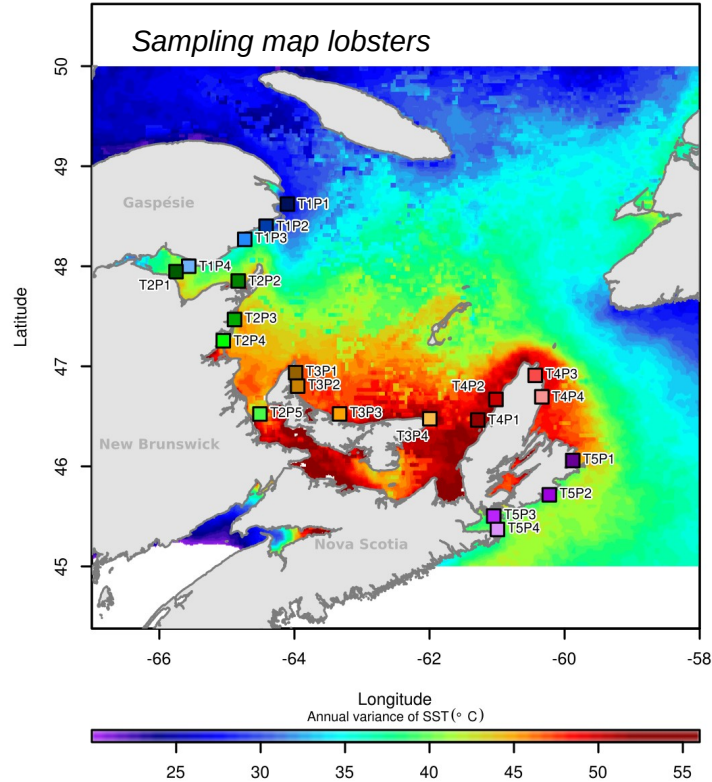
➔ 13,854 SNPs
➔ 1,521 loci CNVs

**→ signal of local adaptation associated with thermal variance of sea surface temperature ?**

**b**



Sampling map lobsters

Data MARSPEC (series 2002-2010)
Source : Sbrocco & Barber 2013

Genome scans GEA (RDA & linear mixed models)

SNPs

CNVs

**➔ signal of local adaptation associated with thermal variance of sea surface temperature ?**

**b**



*Sampling map lobsters*

Data MARSPEC (series 2002-2010)
Source : Sbrocco & Barber 2013

Genome scans GEA (RDA & linear mixed models)

SNPs

CNVs

**Nothing**
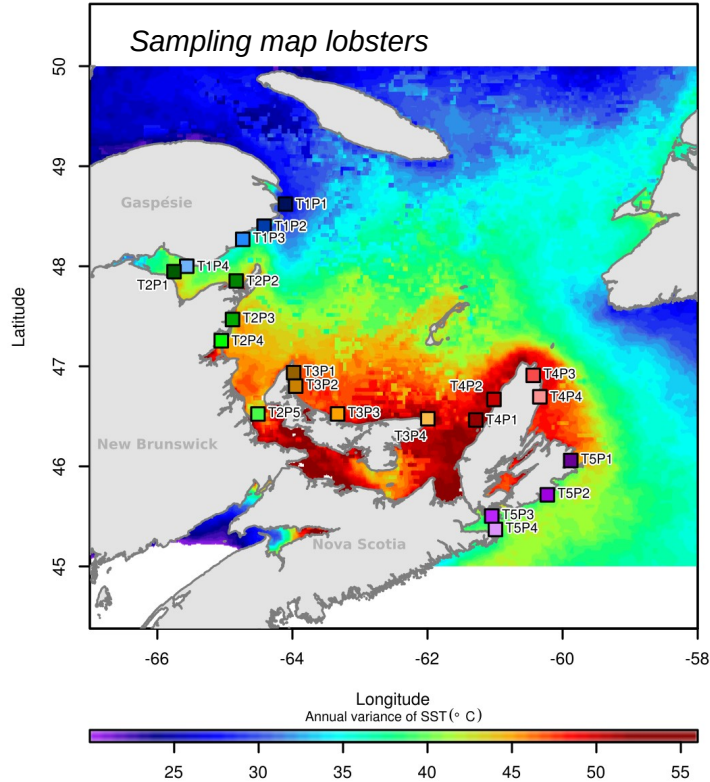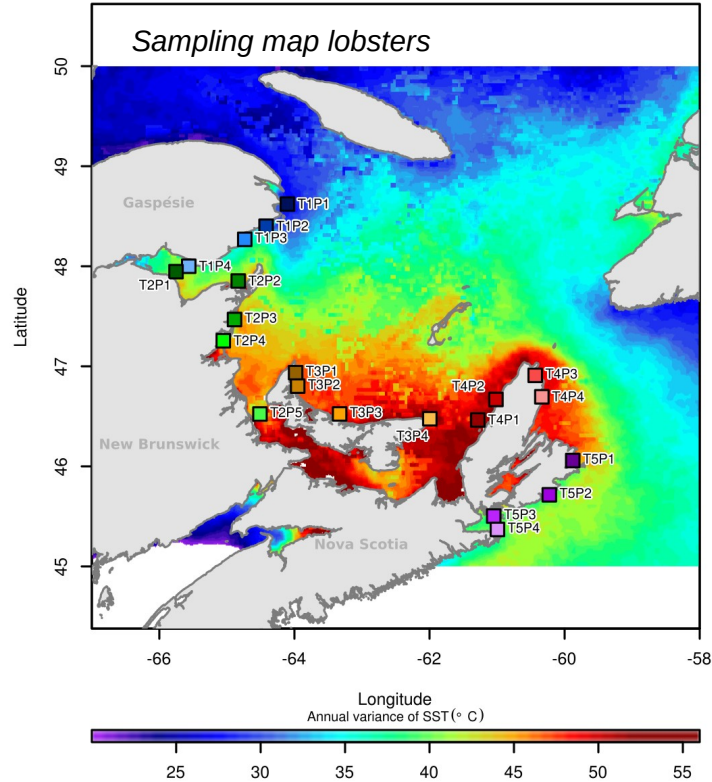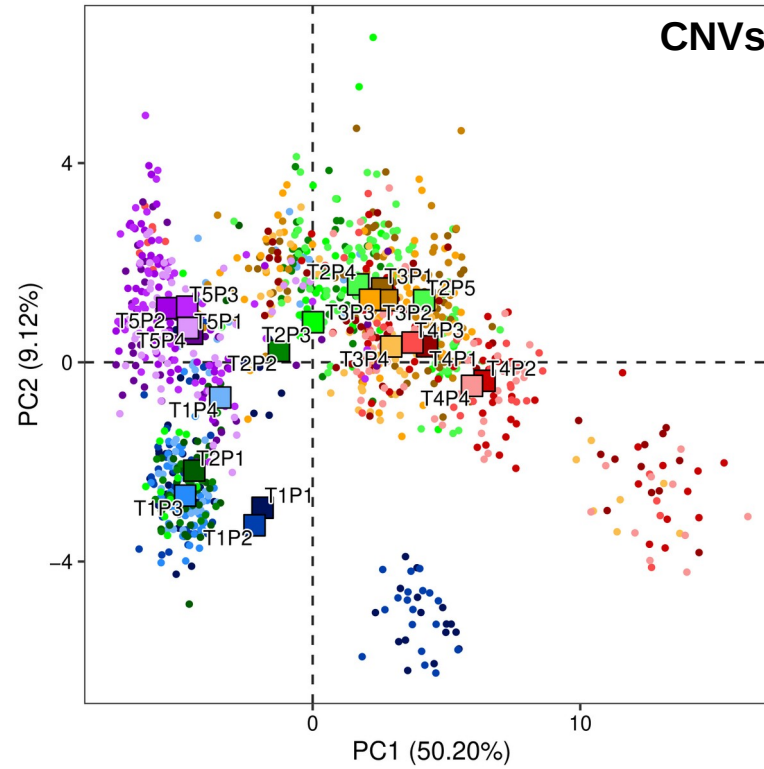
**48 significant loci**

# ➜ Structure pattern for these 48 CNV loci associated with thermal variance.

**b**



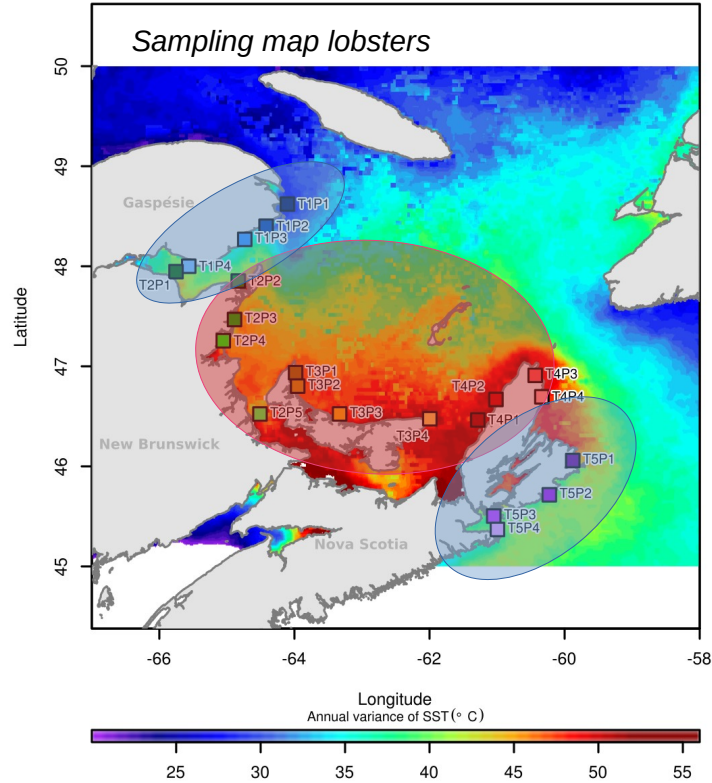Sampling map lobsters

Longitude
Annual variance of SST (° C)

**Data MARSPEC** (series 2002-2010)
Source : Sbrocco & Barber 2013



CNVs

# Structure pattern for these 48 CNV loci associated with thermal variance.

**b**



Sampling map lobsters

Longitude
Annual variance of SST(° C)
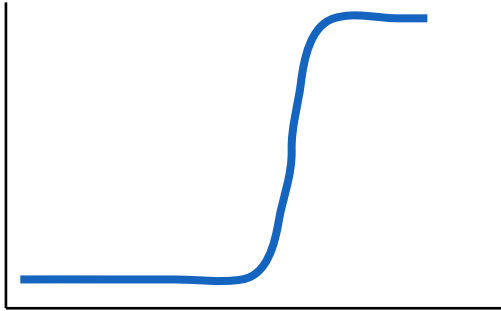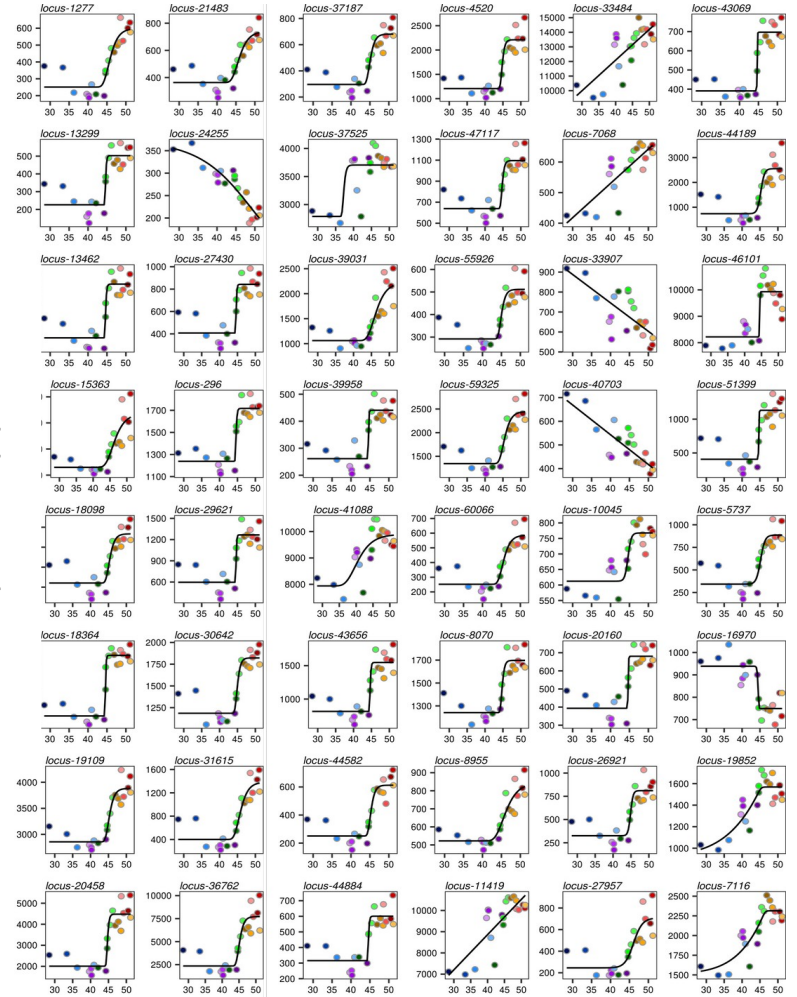
**Data MARSPEC** (series 2002-2010)
Source : Sbrocco & Barber 2013



CNVs
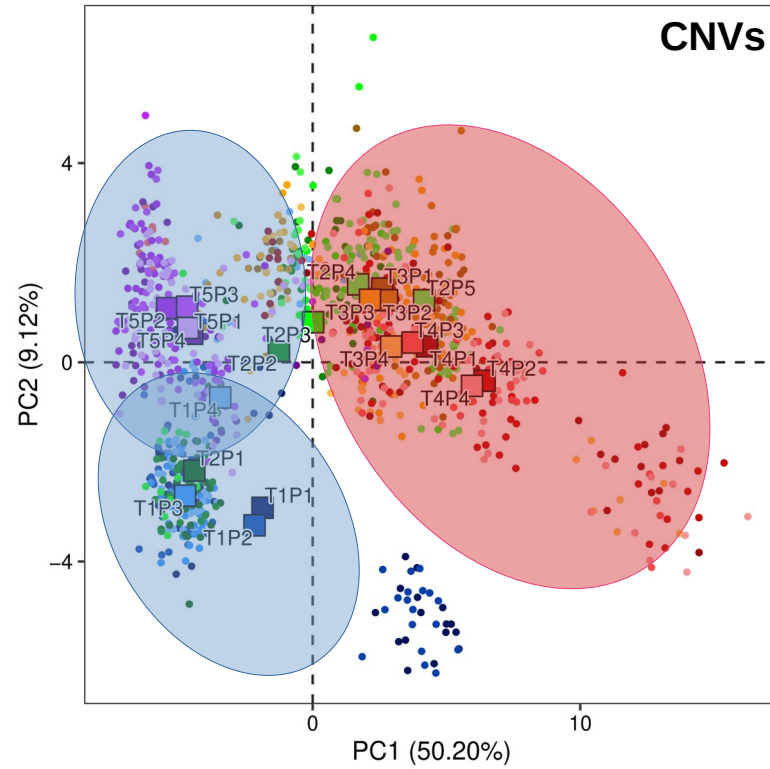
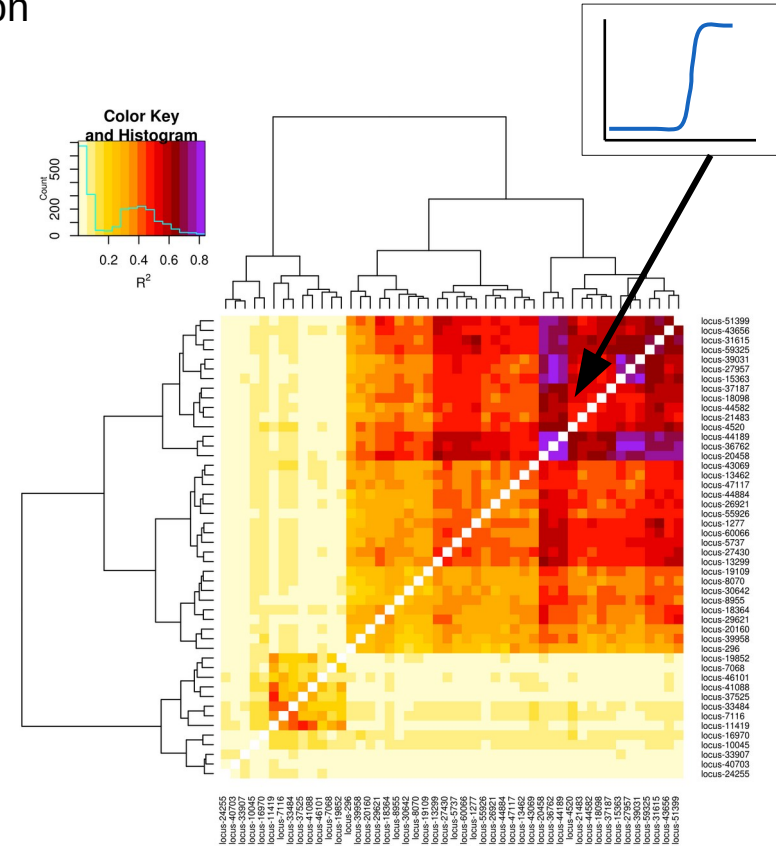**A non-linear relationship that raises questions.**

# Relationship between temperature-associated CNV loci

**Hypothèses :**

1. Physical linkage on the genome.
2. Covariation in copy number among loci driven by selection

**Perspectives :**

➔ Long-read sequencing
➔ Genome assembly



Heatmap of statistical relationship among the 48 CNV loci.

## Summary

- Duplicated loci can introduce bias in SNP genotype calling.

- New way to filter RADseq-like data (duplicated loci)

- Duplicated sequences can be analyses along with SNPs

- Assessing for GEA is possible in both cases

- Specific population structure can be captured by SVs.