

Génétique des populations des souris des Orcades (marqueurs RADs)



ANR
BIGTOOTH

Pascale Chevret



Souris des Orcades

Les Orcades



Mus musculus domesticus

Colonisation récente

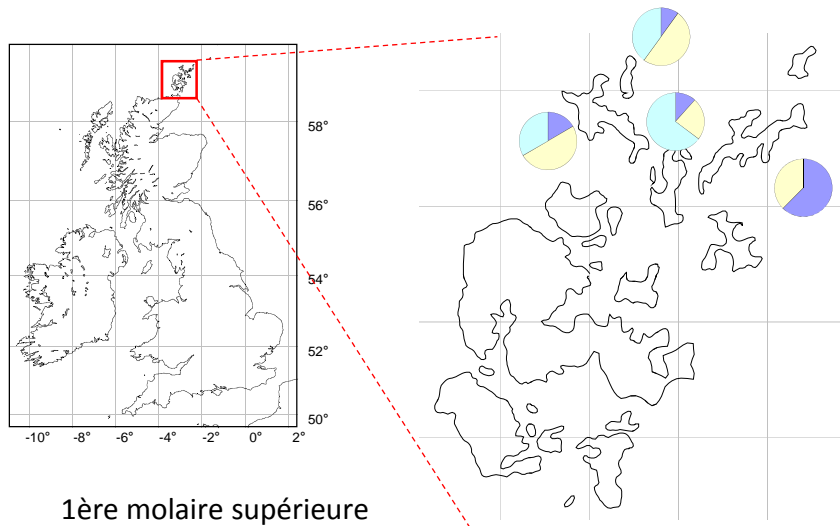


Signature mitochondriale viking
~1200 ans

« Orkney mtDNA lineage »

Searle et al. 2009
Jones et al. 2011

Mission 1992 (Guila Ganem)
Grande diversité morphologique



1^{ère} molaire supérieure

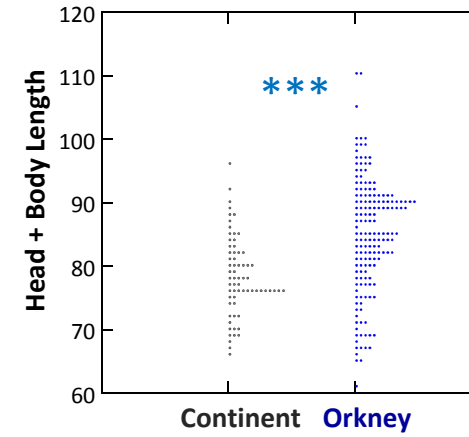


Evolution morphologique

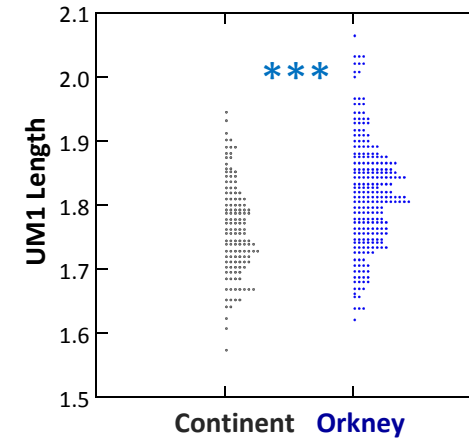
Structure génétique aux Orcades (différents marqueurs)

Relation entre évolution morphologique et moléculaire

Evolution insulaire ?

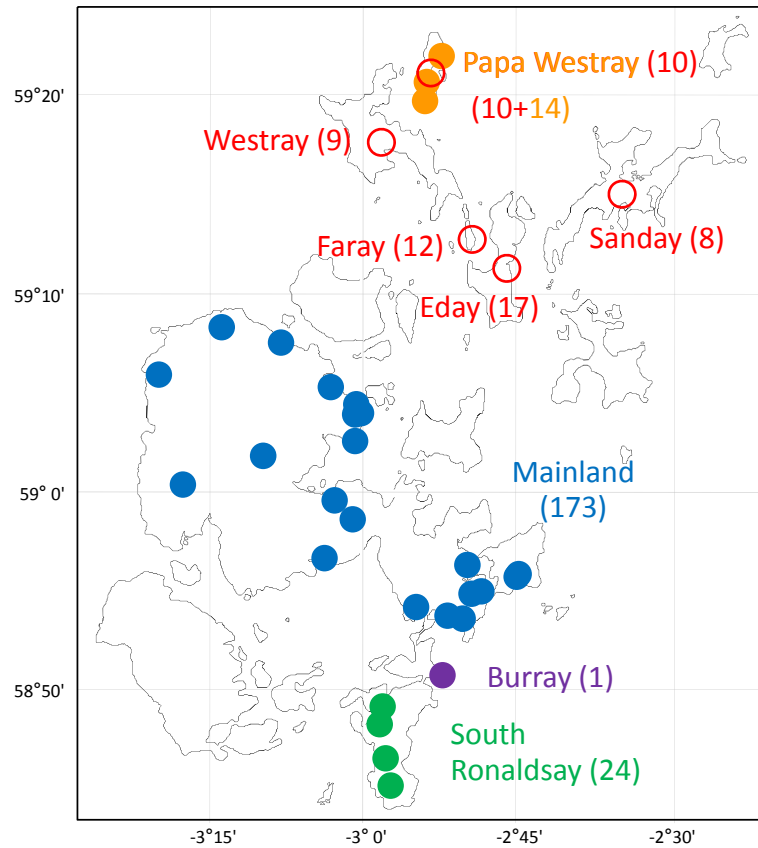


1^{ère} molaire supérieure



Echantillonnage aux Orcades

Iles Orcades



2 missions

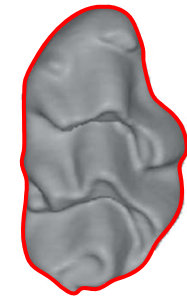
○ 1992 ● 2012



- **Souris mises en élevage**

- **Analyses morphométriques**

Contour de la M¹
Analyses multivariées

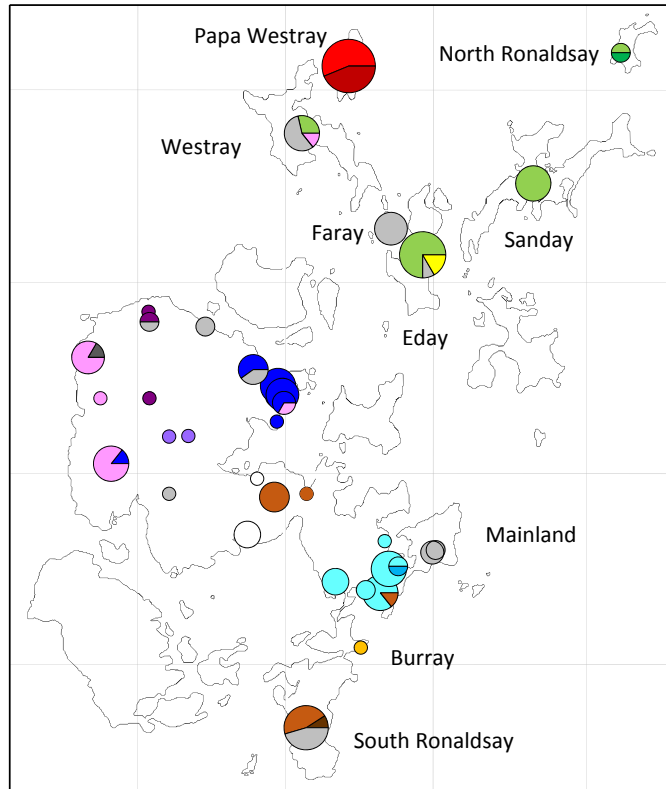


- **Analyses moléculaires**

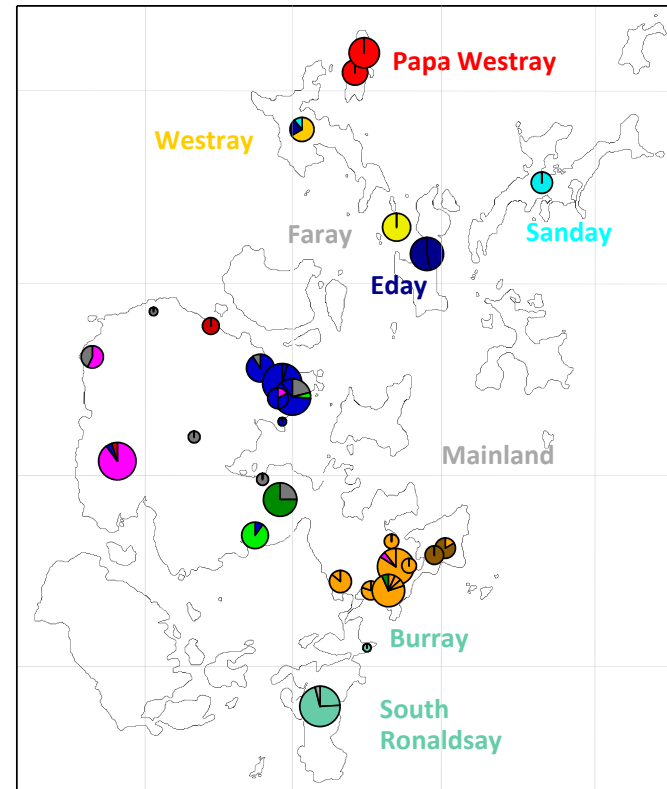
- Dloop
- microsatellites
- RADseq

Structure génétique aux Orcades

Dloop (17 haplotypes)
n= 146



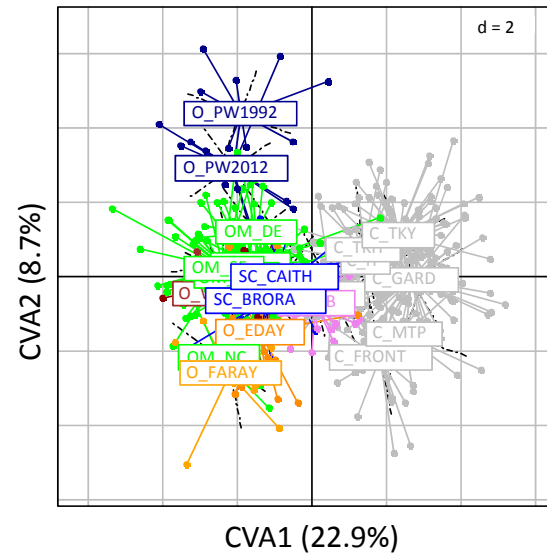
19 microsatellites (21 clusters)
n= 279



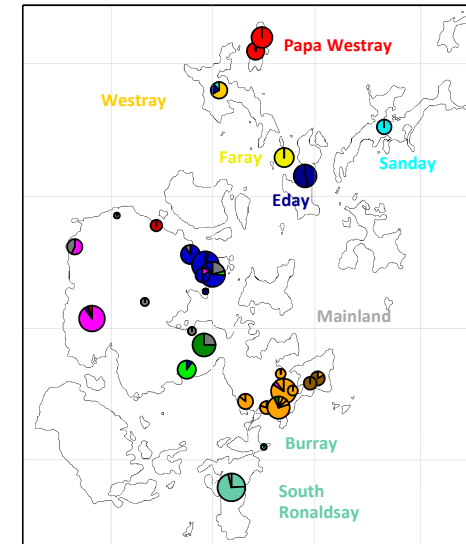
- Congruence des données mitochondriales et microsatellites
- Différenciation entre Mainland et les autres îles
- Structure intra-mainland



Morphométrie



Microsatellites



- Diversification morphologique très importante, accélération de l'évolution morphologique sur les Orcades
- Corrélation entre évolution moléculaire (microsatellites) et morphologique à l'échelle des Orcades
- Congruence entre les données mitochondriales (Dloop) and microsatellites

Marqueurs RADseq ?

RADseq des souris des Orcades

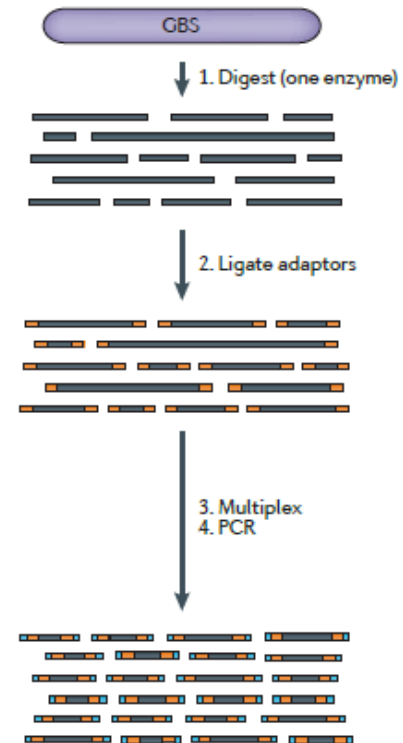
2 approches différentes :

- GBS (Genotyping By Sequencing)
- ddRAD (double digest RADseq)

GBS (Genotyping by sequencing)

Genomic Diversity Facility (Cornell University)

Enzyme PstI



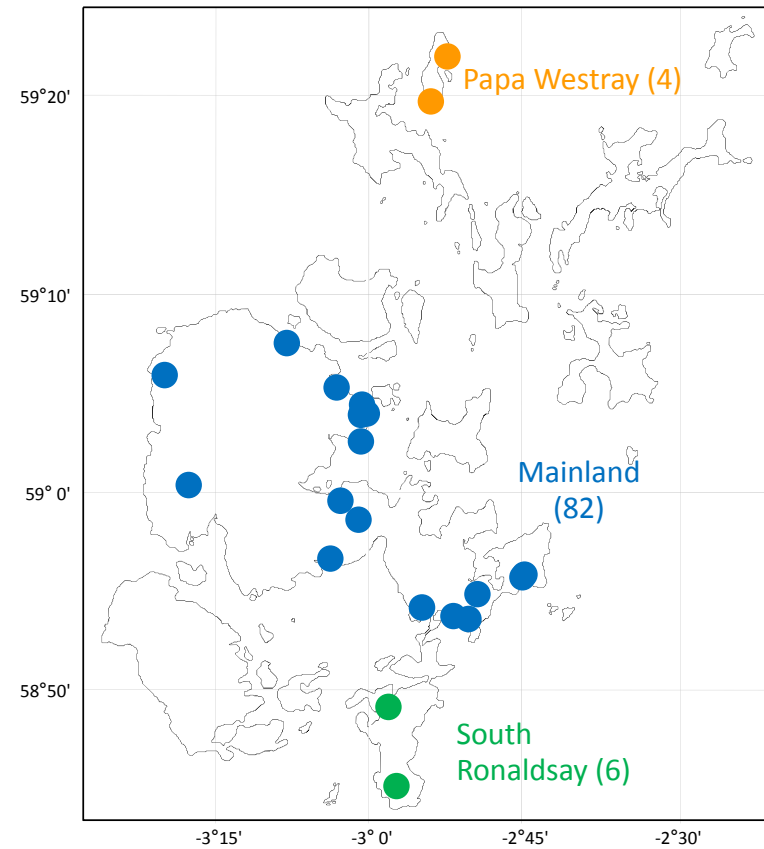
GBS Orcades

Echantillonnage

566 souris de l'élevage
fondateurs et descendants
6 plaques, 1 plaque/ligne



92 souris sauvages
21 localités (31 échantillonnées)
3 îles



GBS Orcades

Genomic Diversity Facility (Cornell University)

- Librairies
- Séquençage
- **SNPs calling : TASSEL (Bradbury et *al.*, 2007), génome de référence**

Résultats

129 698 SNPs, couverture moyenne : 8.1x, données manquantes : 35 %
(filtre : minor allele frequency >0.01, missing data per site < 90%).

Données manquantes/site < 15 %

-> 92 genotypes, 48 434 SNPs, 5.9 % données manquantes

LBBE

SNPs calling : stacks (Catchen et *al.*, 2013), génome de référence

$r = 0.1$ (=filtre Cornell)

150 686 SNPs, couverture moyenne : 7.8x, données manquantes : 36 %

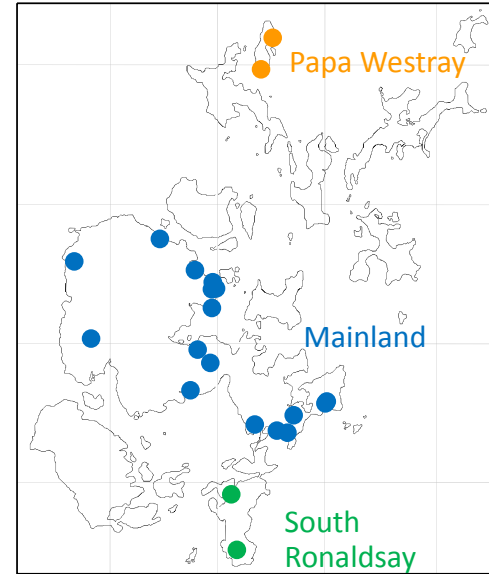
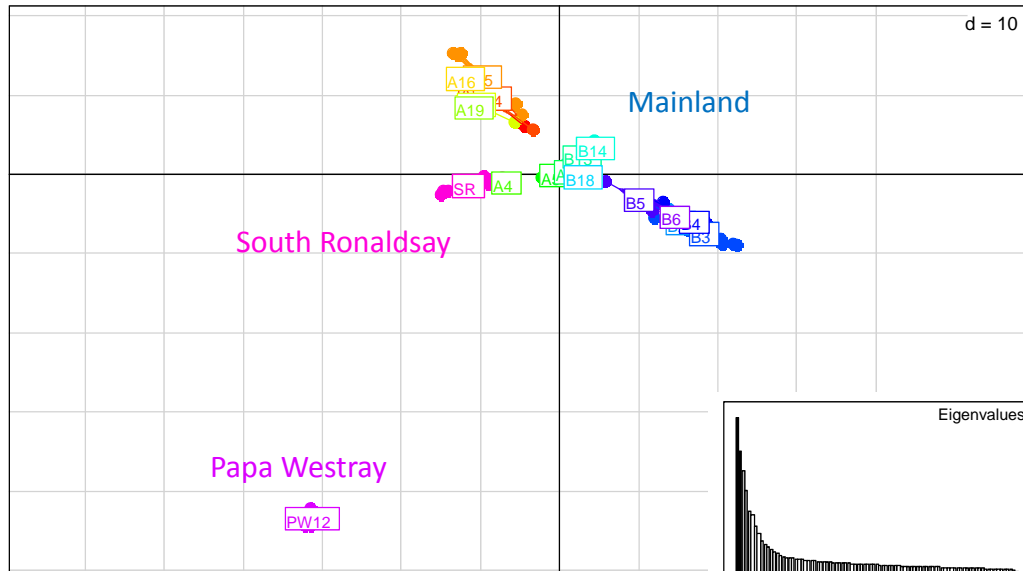
$r = 0.85$ (données manquantes/site < 15 %)

-> 92 genotypes, 44 567 SNPs, 7.6 % missing data

GBS Orcades

Analyse : R package **adegenet**, *Jombart, 2008*

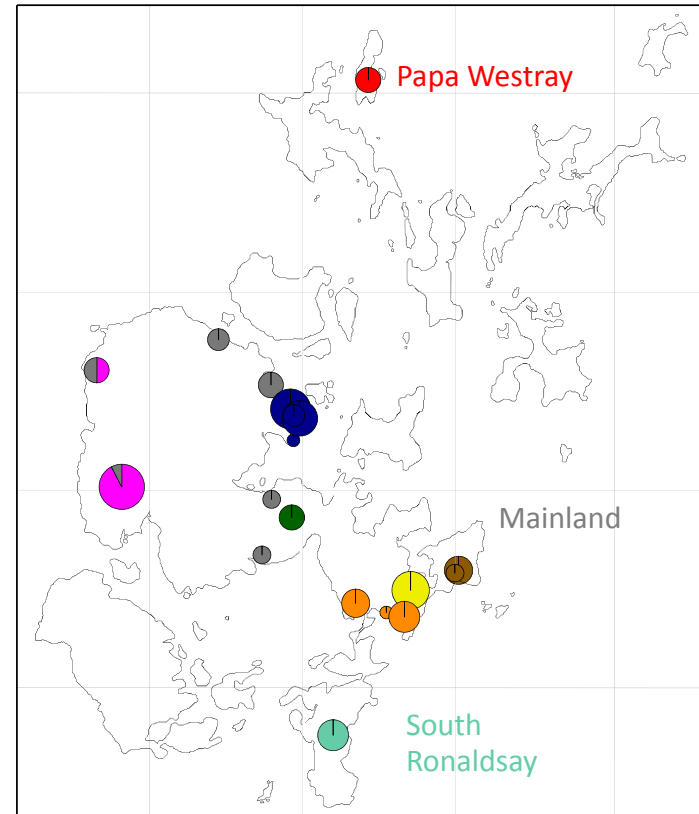
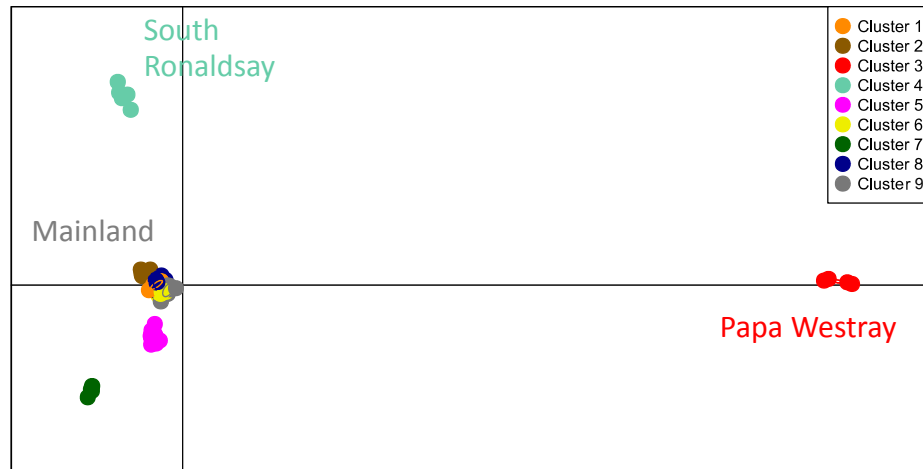
ACP, axes 1 & 2



GBS Orcades

Analyse : R package **adegenet**, *Jombart, 2008*

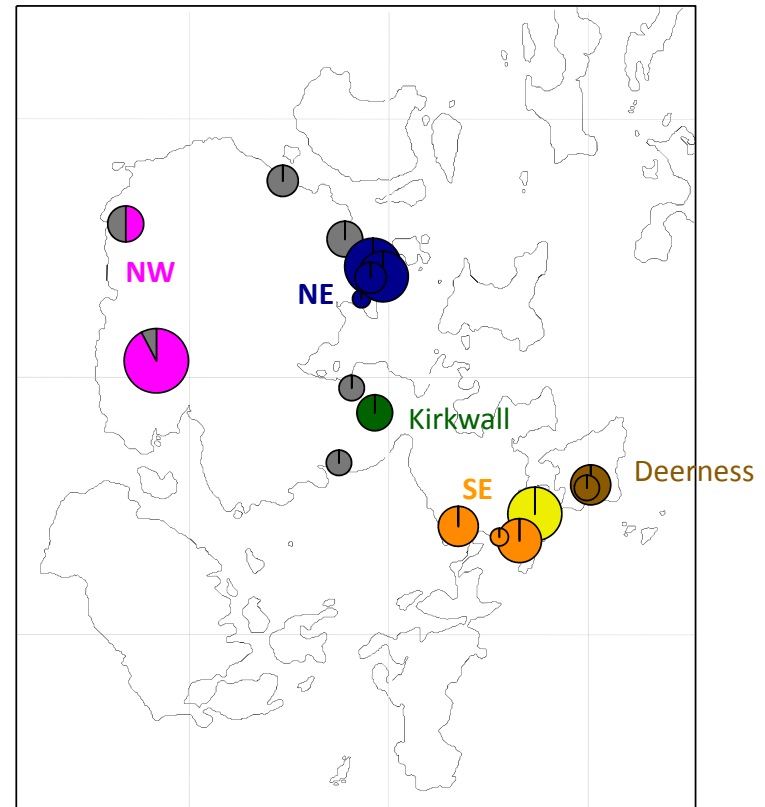
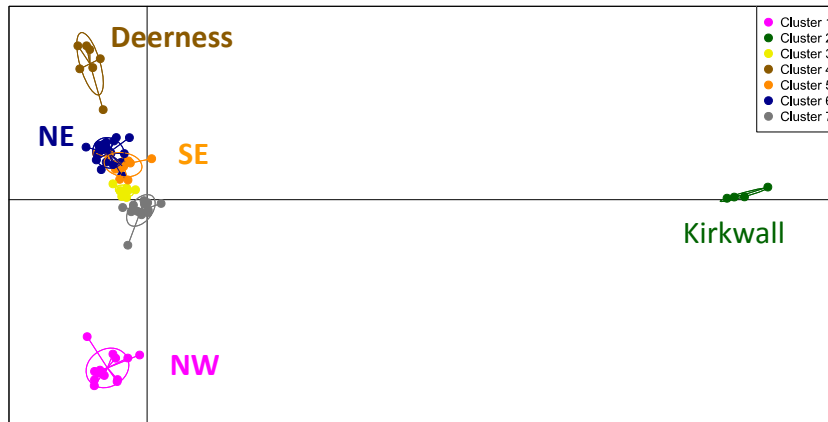
Discriminant Analysis of Principal Components
« find cluster » K = 9



GBS Mainland

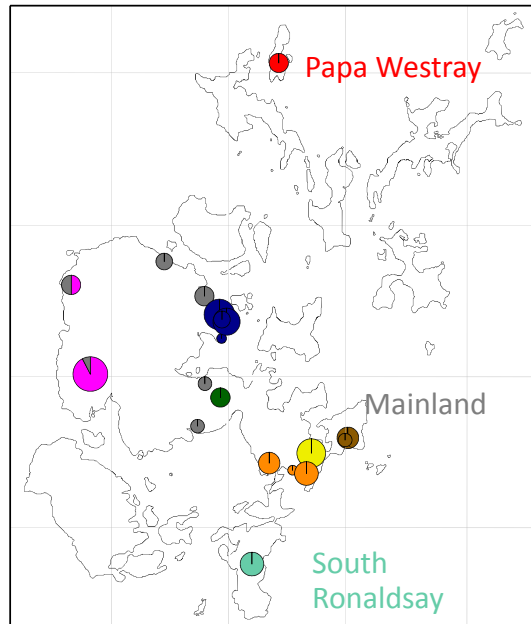
82 souris, 40 496 binary SNPs, 8.08 % données manquantes

Discriminant Analysis of Principal Components
« find cluster » K = 7



Structure génétique aux Orcades

GBS, 44 567 SNPs
n = 92



- Différenciation entre Mainland, Papa Westray et South Ronaldsay

- Structure intra-Mainland

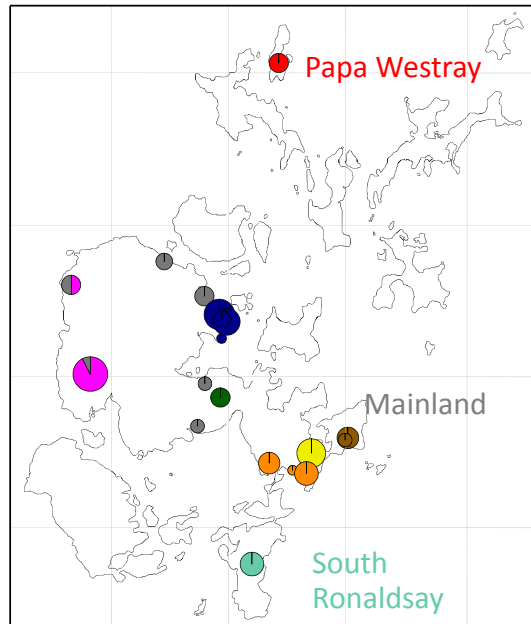
- Isolation par la distance

Orcades : ✓

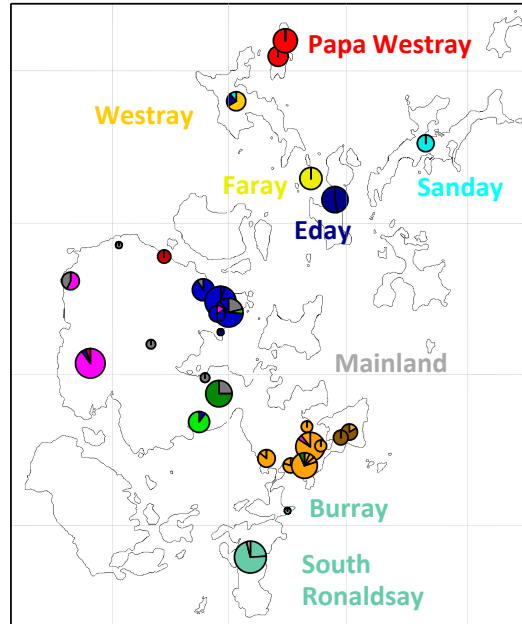
Mainland : ✗

Structure génétique aux Orcades

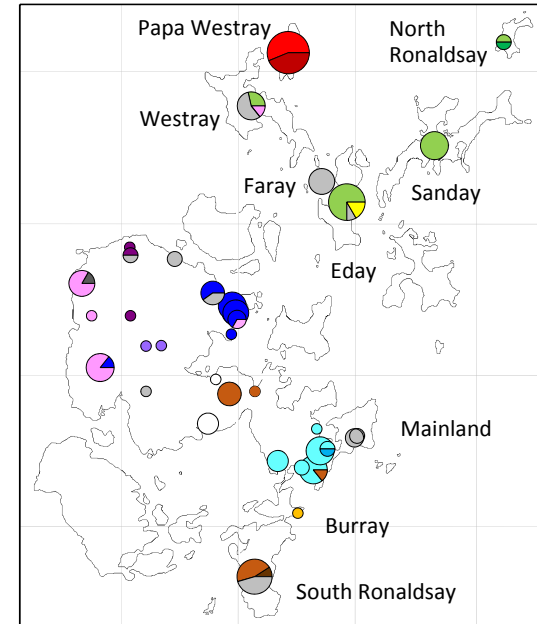
GBS, 44 567 SNPs
n = 92



19 microsatellites
n = 279



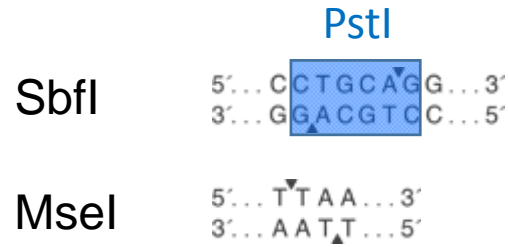
Dloop
n = 146



- Différenciation entre Mainland et les autres îles
- Structure intra-mainland

ddRAD Orcades

Enzymes

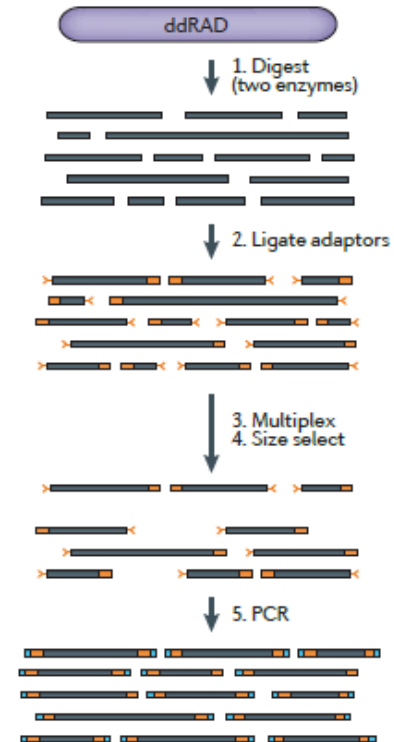


Réalisation des bibliothèques

1 par espèce (souris/chat/sanglier)

Pool des 3 bibliothèques pour séquençage

80 échantillons/ligne



ddRAD Orcades

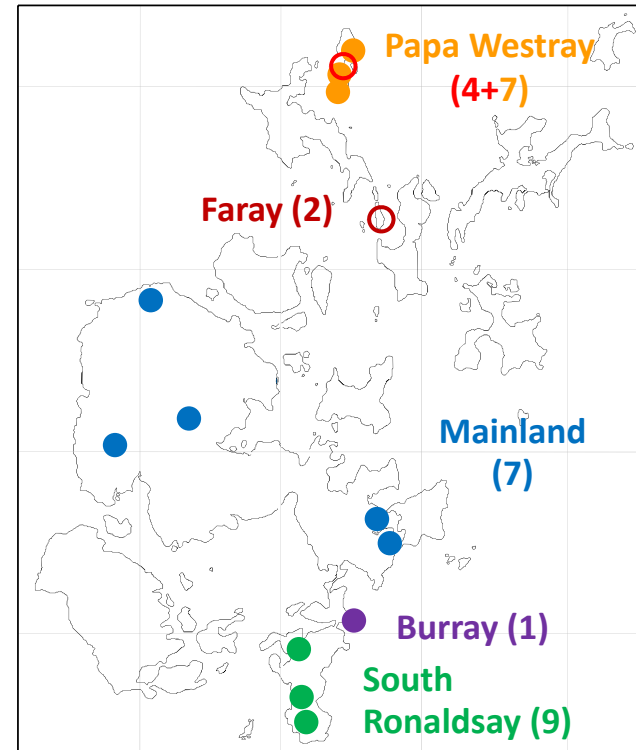
Echantillonnage Orcades

30 souris

10 localités

5 îles

Echantillons 1992 et 2012



○ 1992

● 2012

ddRAD Orcades

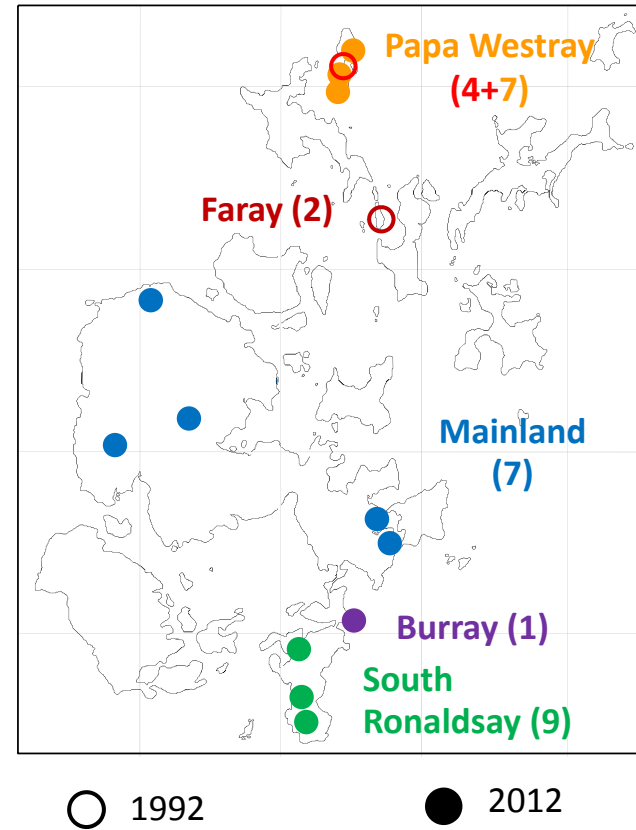
Echantillonnage Orcades

30 souris

10 localités

5 îles

Echantillons 1992 et 2012

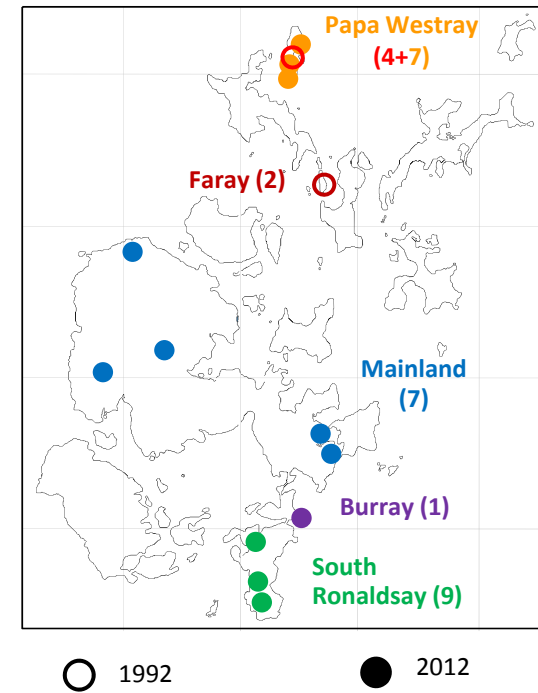
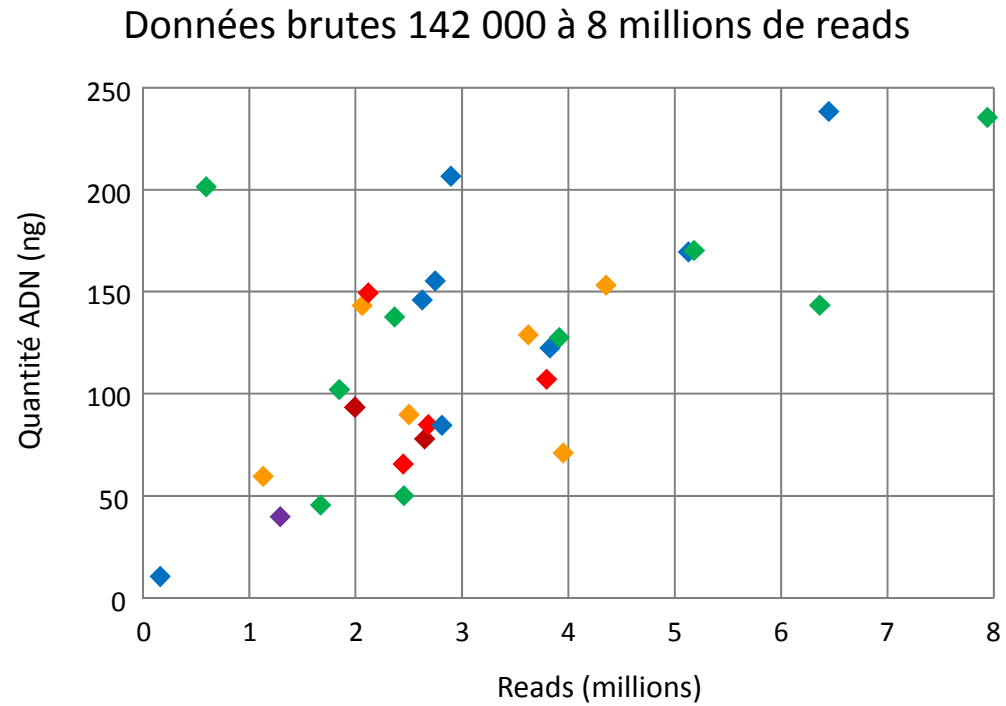


SNPs calling : stacks (Catchen et *al.*, 2013), génome de référence

Analyse : R package **adegenet**, Jombart, 2008

ddRAD Orcades

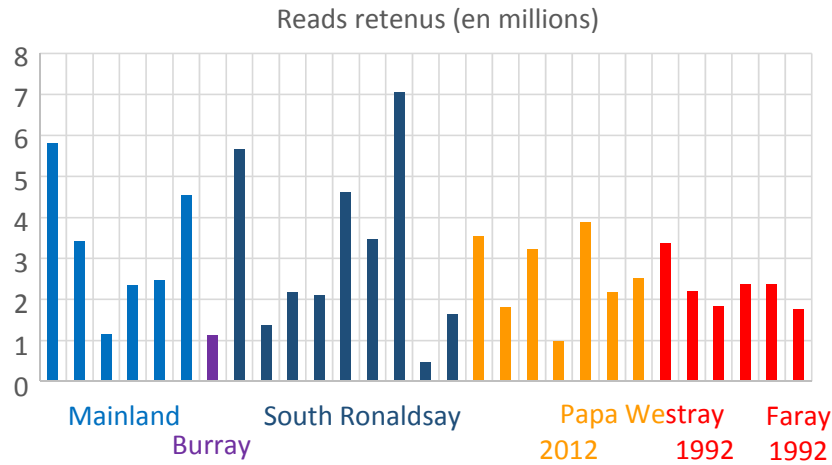
Différences 1992 – 2012 ?



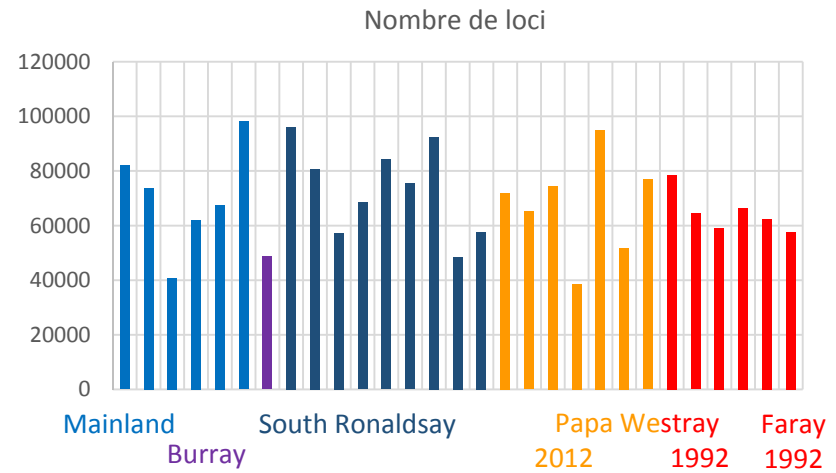
Suppression de l'échantillon le plus faible → 29 souris

ddRAD Orcades

Comparaison 1992 - 2012



86 - 92 % des reads sont retenus



Couverture moyenne : 10x - 76x

stacks

$r = 0.1$

87 637 SNPs, couverture moyenne : 58.5x, 38 % données manquantes

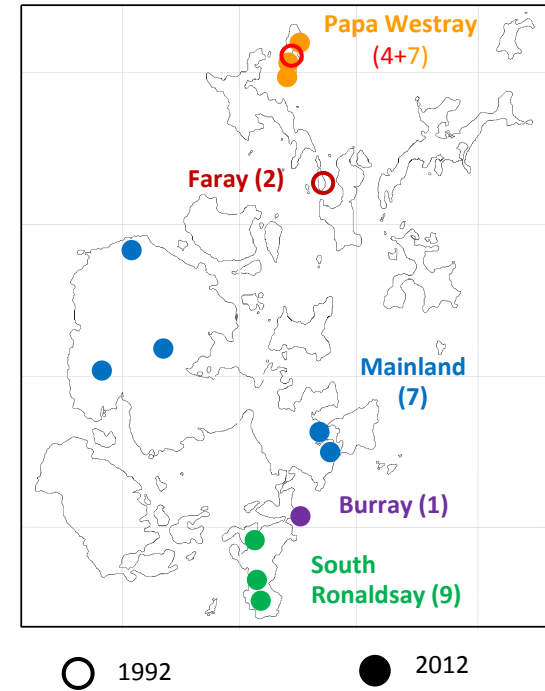
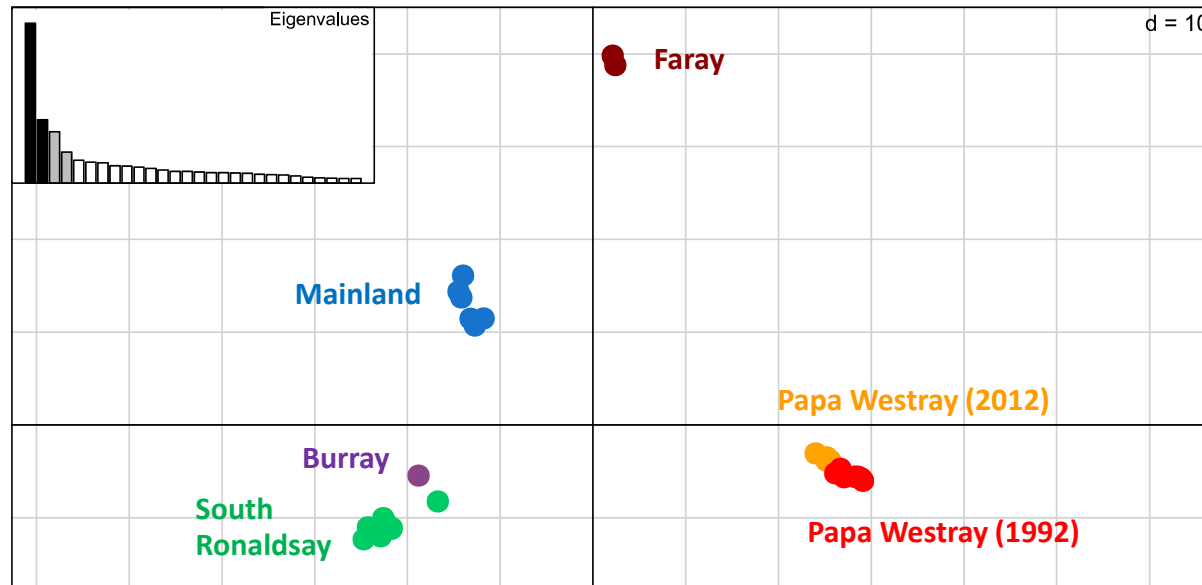
$r = 0.85$

26 896 SNPs, 5.1 % données manquantes

ddRAD Orcades

Analyse : R package **adegenet**, *Jombart, 2008*

ACP, axes 1 & 2



GBS - ddRAD



Comparaison / combinaison fichiers SNPs (sorties stacks, format vcf) 

Analyses stacks combinées 

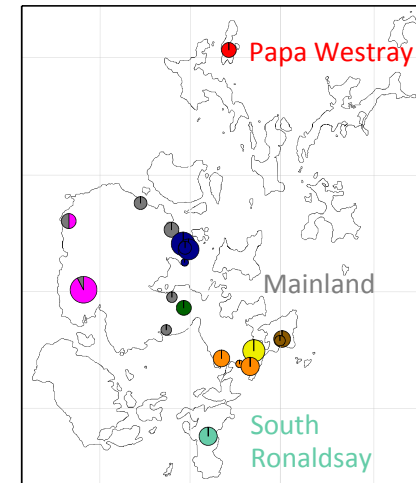
TagDigger: user-friendly extraction of read counts from GBS and RAD-seq data

Lindsay V. Clark* and Erik J. Sacks

 
(en cours)

| Paramètres stacks | GBS | | ddRAD |
|---|---|------|--|
| % reads retenu | 32 - 48 | < 50 | 86 - 92 |
| Couverture moyenne | 7.8x | < 10 | 58.5x |
| SNPs, r=0.85 | 44 567 | | 26 896 |
| stacks denovo (sans génome de référence) |  | |  12599 SNPs |

Conclusions



Qualité données ddRAD > GBS

RADseq : Structuration génétique

Congruence avec microsatellites et Dloop

Origine ?

Combinaison des jeux de données RADseq

Comparaison avec autres données NGS

Remerciements

Orcades 2012, ANR Bigtooth



Sylvie Agret

Jean-Christophe Auffray

Josette Catalan

Lionel Hautier

Annie Orth

RADseq



Julien Claude

Claire Dufour

Marie Pagès



Hélène Henri

Corine Régis

Adil El Filali

Orcades 1992

Guila Ganem



Sabrina Renaud

