

CMSC 22



StruK2Stat

FURUSHO, Kaito O.

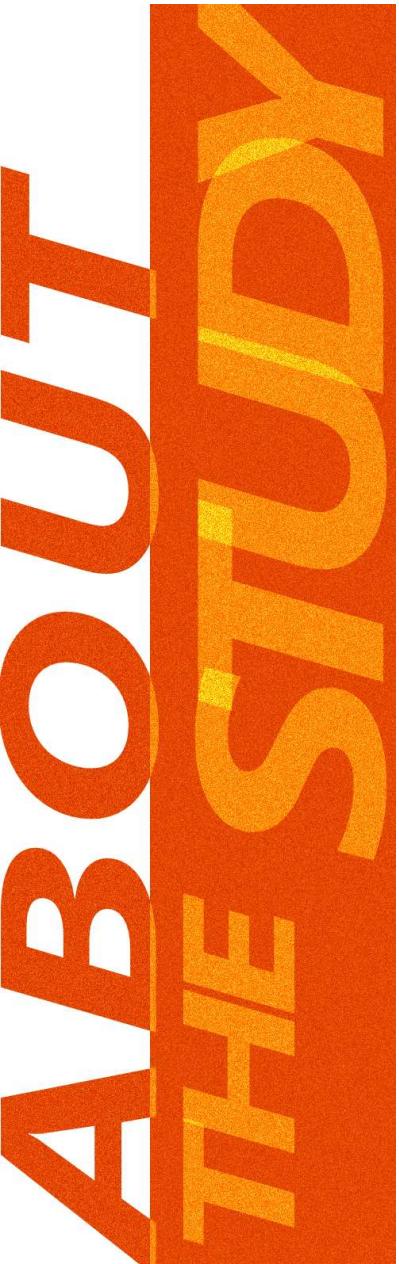
CMSC 22 WX-2L 2nd SEM AY 2024-2025

ABOUT
THE SITE



Genetic Diversity and Population Structure Analysis of Philippine Native Pigs Highlight Five Priority Populations for Conservation

Banayo, J. B. B., Manese, K. L V., Furusho, K. O., Salces, A. J., & Yamagata, Y. (2023).
<https://doi.org/10.1002/ece3.10618>



Sample Published Data

<https://datadryad.org/dataset/doi:10.5061/dryad.b8gtht7gg>



About ▾

Genotype data of Philippine native pigs, Duroc, Landrace, Large White and Berkshire, using 20 ISAG-FAO recommended microsatellite markers

[Banayo, Joy](#) ¹ ; [Manese, Kathlyn Louise](#) ¹ ; [Furusho, Kaito](#) ¹ ; [Salces, Agapita](#) ¹ ;
[Yamagata, Takahiro](#) ²

Author affiliations ▾

Published Oct 04, 2023 on Dryad. <https://doi.org/10.5061/dryad.b8gtht7gg>

Data files

▲ Oct 04, 2023 version files	67.20 MB
DRYAD_DATASET_A.txt	35.34 KB
Microsatellite_raw_data_.fsa_Banayo_2023.zip	67.16 MB
README.md	4.84 KB

[Download full dataset](#)



Genetic Diversity and Population Structure Analysis of Philippine Native Pigs Highlight Five Priority Populations for Conservation

Banayo, J. B. B., Manese, K. L V., Furusho, K. O., Salces, A. J., & Yamagata, Y. (2023).

<https://doi.org/10.1002/ece3.10618>

THE PHILIPPINE NATIVE PIG (PhNP)

The Philippine Native Pig (PhNP), an Asian-type *Sus scrofa* is locally regarded as an indigenous species adapted to local conditions.



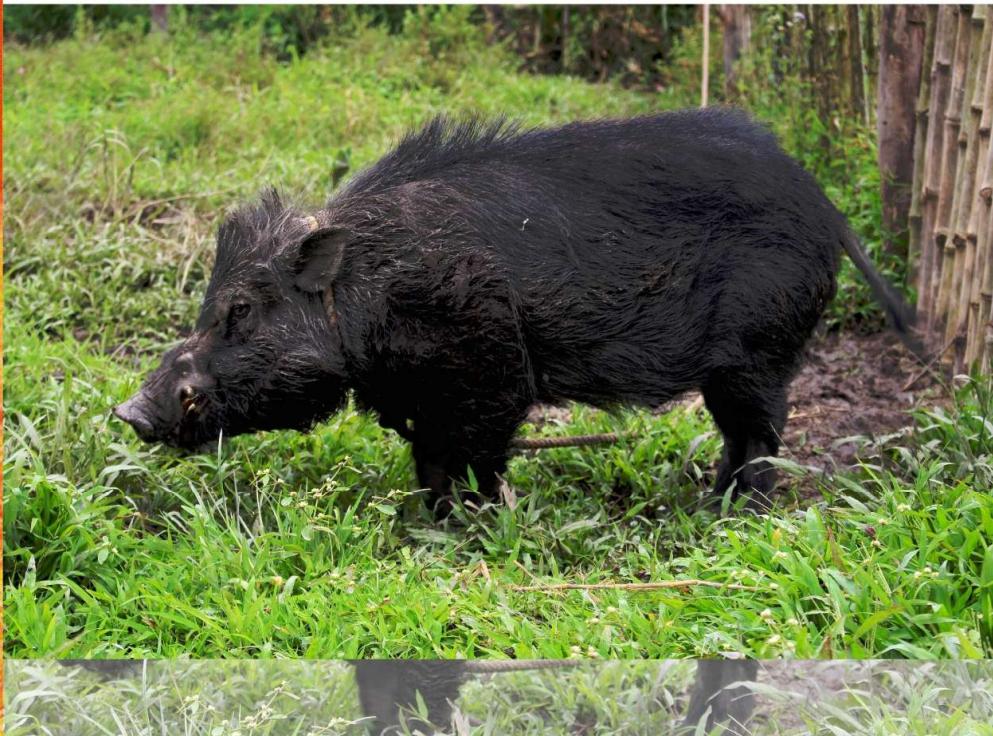
The PhNP is culturally important to many indigenous communities as they are used in many rituals and in preparation of ethnic foods and delicacy (i.e. *lechon*, *etag*, and *kinuday*)

It is also widely known for its superior meat quality, making it an ideal ingredient for many Filipino delicacies.

THE PHILIPPINE NATIVE PIG (PhNP)

The actual proportion of PhNP is vaguely known and is still a minority breed in the Philippines.

Previous phylogenetic studies suggest high genetic variation among PhNPs on different islands. Studies also suggest at least four origins of domestication throughout Asia (i.e. Southeast Asia, East Asia, Cordillera Region/Lanyu Island, and the Pacific)



Possible interbreeding with endemic wild pigs, such as *S. philippinensis*, *S. cebifrons*, *S. ahoenobarbus*, and *S. Olivierii*, may have complicated the PhNP genetics

Nevertheless, population structure based on microsatellite markers showed that the PhNP was distinct from Duroc and Yorkshire breeds (transboundary breeds)

A previous study also showed that subpopulations in Southern Luzon, particularly Quezon and Marinduque, were moderately differentiated with each other despite close proximity suggesting some degree of within breed variation.

Despite being highly adapted in the Philippine condition, the PhNP remains a lowly productive breed.

To improve productivity, raisers would at times breed these PhNPs with transboundary breeds; hence, the genetic integrity of the PhNP may have then been compromised.

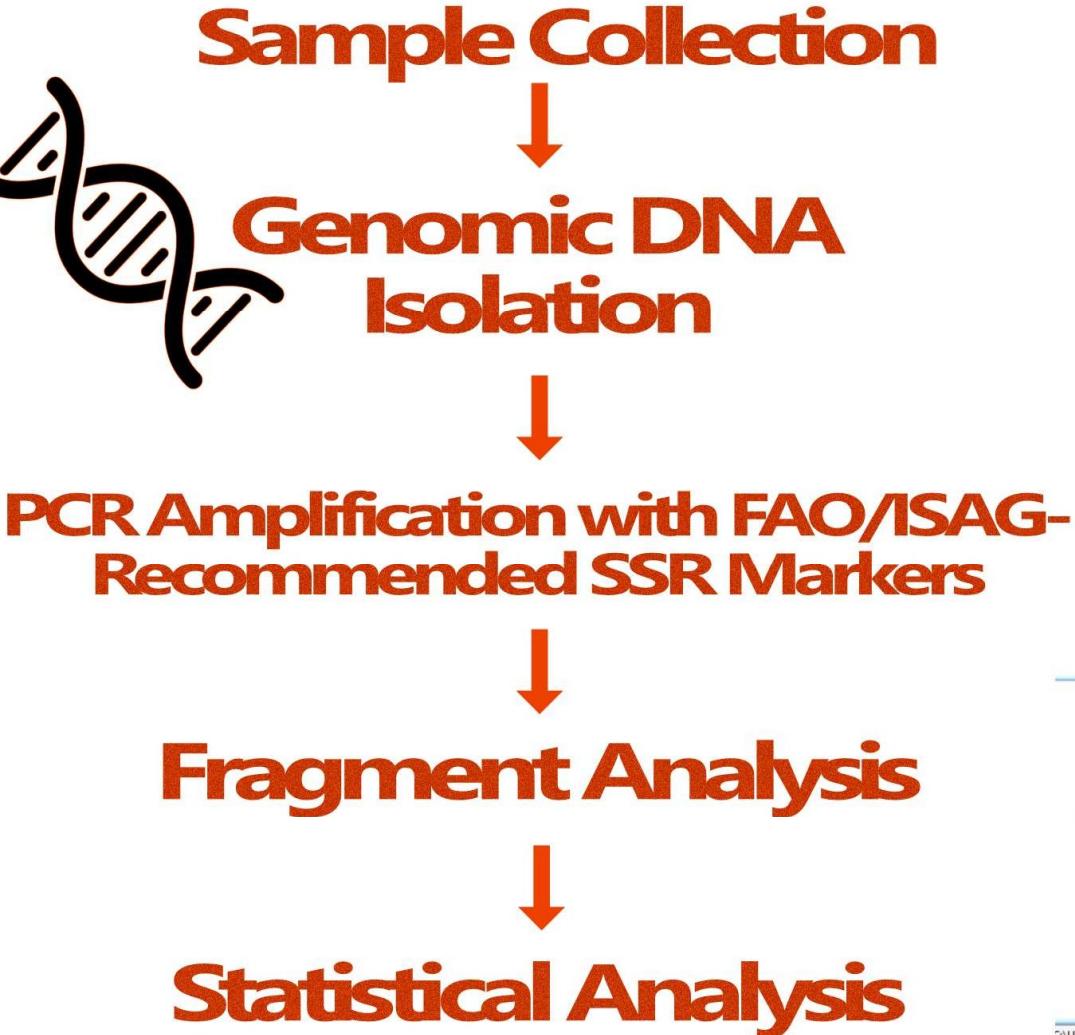
HYPOTHESIS

The proponents hypothesize that there exists variations in the PhNP Population Structure that may provide clues for sustainable conservation practices and maximum utilization of these native pigs.

OBJECTIVE

The study aims to gain deeper understanding of the genetic diversity and population structure of the PhNP by microsatellite analysis and comparison of their molecular variance at the island, administrative region, and population levels, with the goal of informing genetic improvement and conservation programs

Methodology



North Luzon : Benguet, Kalinga, Isabela, Nueva Vizcaya

South Luzon & Visayas : Quezon, Marinduque, Samar

MICROSATELLITES/ SHORT SEQUENCE REPEATS (SSRs)

For Diploids:

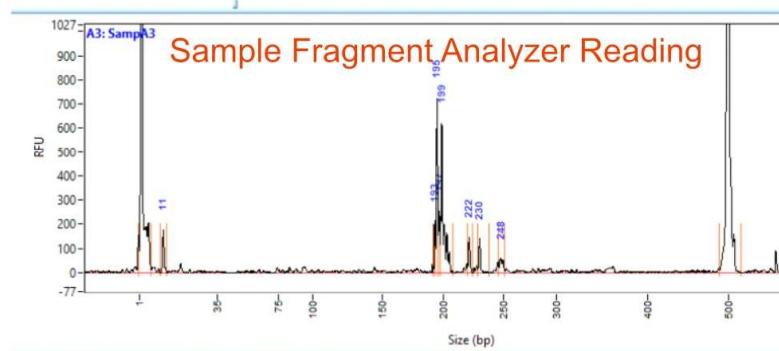
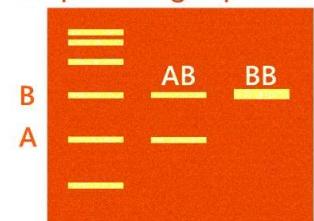
Sample #1 (AB)



Sample #2 (BB)



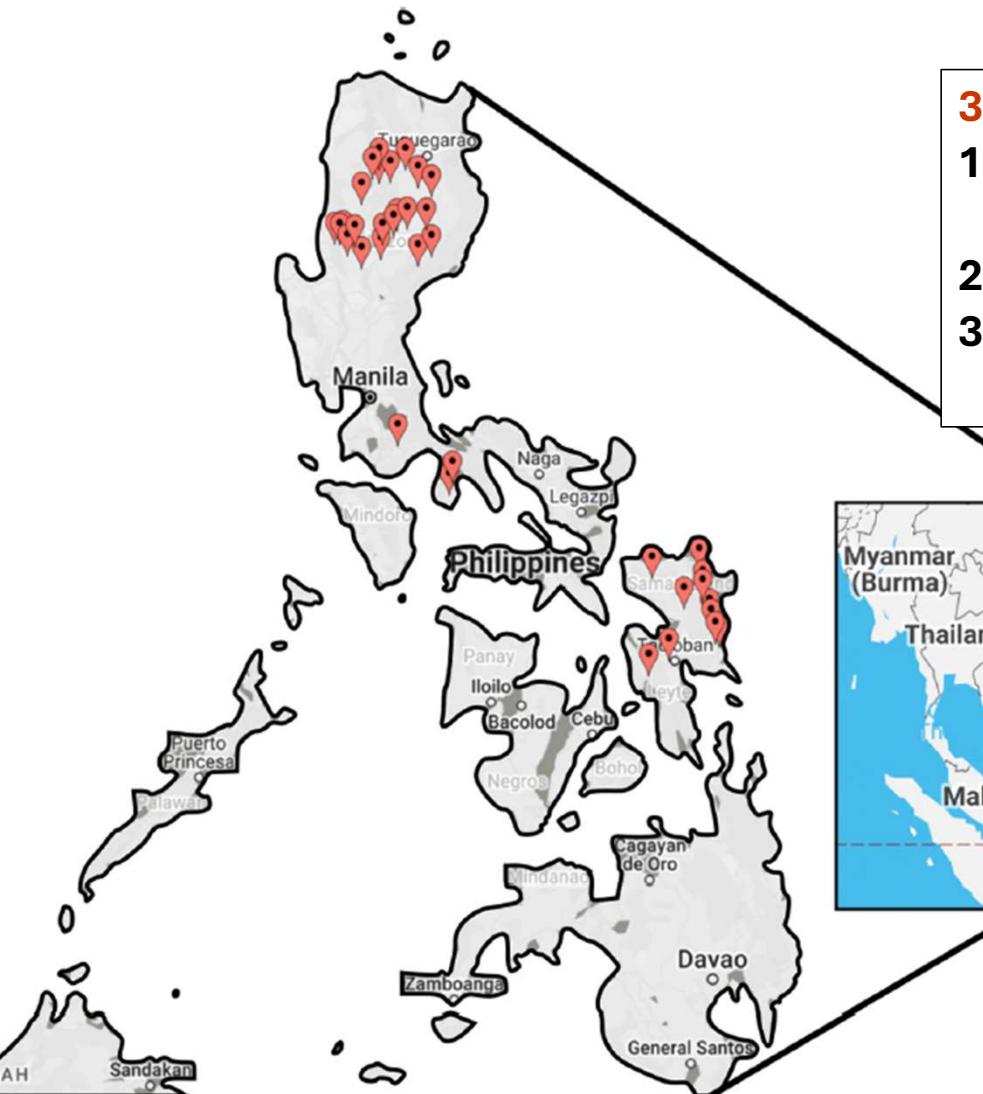
Expected gel product



C:\Users\Kaito Furusho\Desktop\200a\FA\2018 10 30\12-36-12\2018 10 30 12H 36M.raw

Peak Table			
	Size (bp)	ng/uL	% (Conc.) (ng/uL)
1	1 (LM)	0.6824	497.1332
2	11	0.2740	10.3
3	193	0.1654	6.2
4	195	0.6486	24.3
5	197	0.2278	8.5
6	199	0.9564	35.8
7	222	0.1390	5.2
8	230	0.1382	5.2
9	248	0.1208	4.5
10	500 (UM)	0.5000	1.6455
TIC:	2.6702	ng/uL	
TIM:	59.5290	nmole/L	
Total Conc:	4.5342	ng/uL	

Methodology



3 Classes based on ISLAND GROUPING & ORIGIN:

- 1. North Luzon : Benguet, Kalinga, Isabela, Nueva Vizcaya**
- 2. South Luzon - Visayas : Quezon, Marinduque, Samar**
- 3. Transboundary : Duroc, Large White, Landrace, Berkshire**



3 Classes based on ELEVATION & ORIGIN:

- 1. High (Native) : Benguet**
- 2. Low (Native) : Kalinga, Isabela, Nueva Vizcaya, Quezon, Marinduque, Samar**
- 3. Transboundary : Duroc, Large White, Landrace, Berkshire**

Key Statistics and Analysis

F-statistics (Wright's Fixation Indices)

Analysis of Molecular Variance (AMOVA)

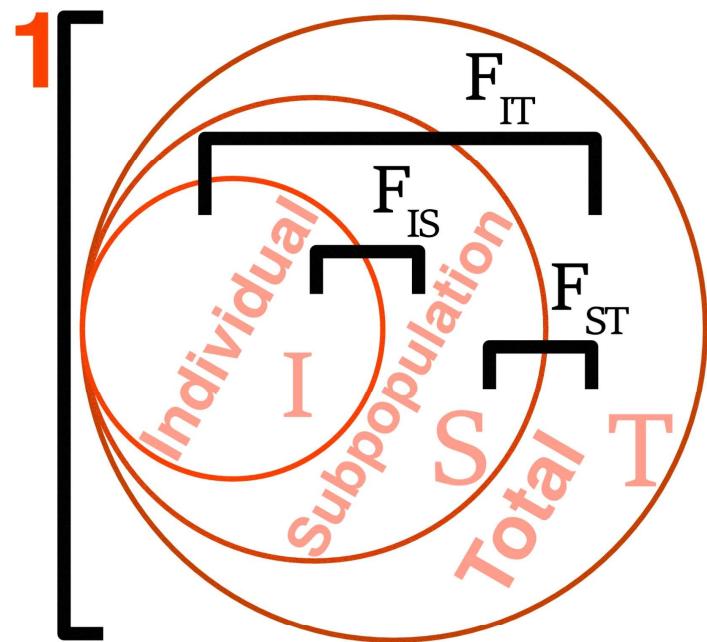
Effective Number of Migrants

Effective Population Number

Genetic Distances (Phylogenetic Analysis)

Bayesian Clustering Analysis

F-statistics (Wright, Fisher, Haldane)



$$F_{IS} = \frac{H_S - H_I}{H_S}$$

Inbreeding
Coefficient
Within
Subpopulations

Deficiency or excess
average heterozygotes in
each sample population

$$F_{IT} = \frac{H_T - H_I}{H_T}$$

Total
Inbreeding
Coefficient

Deficiency or excess
average heterozygotes in
a group of populations

$$F_{ST} = \frac{H_T - H_S}{H_T}$$

Genetic
Differentiation
Among
Subpopulations

Degree of genetic
differentiation among
populations in terms of
allele frequencies

H_T = Average Heterozygosity in the total population

H_S = Average Heterozygosity within subpopulations (Expected Heterozygosity within subpopulations)

H_I = Observed Proportion of Heterozygous individuals in a sample population

Analysis of Molecular Variance (AMOVA)

Different from ANOVA

Partitioning of genetic variation among and within classified groups are described using AMOVA.

2 Genetic Distance Models
(Stepwise Mutation Model,
and Infinite Allele Model)



2 Sets of Populations
(All Native, & With Transboundary)

3 Regions (Classes):

1. High (Native) : Benguet
2. Low (Native) : Kalinga, Isabela, Quezon, Marinduque, Samar
3. Transboundary : Duroc, Large White, Landrace, Berkshire

Analysis of Molecular Variance of PhNP

Genetic distance model	Sample	Region	Number of populations	Variance components (%)		
				Within population	Between populations within regions	Between regions
Allele Size (SMM ^a)	Native	High ^c , Low ^d	7	78	5	18
	All	High, Low, Transboundary ^e	11	88	5	7
Allelic Identity (IAM ^b)	Native	High, Low	7	94	5	1
	All	High, Low, Transboundary	11	88	9	3

^aStepwise mutation model.

^bInfinite allele model.

^cHigh—native pig from a highland elevation (the B population).

^dLow—native pig from lowland elevation (the K, I, N, Q, M, and S populations).

^eDuroc, Large White, Landrace, and Berkshire.

Effective Number of Migrants

$$N_m = \left[\left(\frac{1}{F_{st}} \right) - 1 \right] / 4$$

Fst (degree of population genetic differentiation) is used to estimate Nm

Nm estimates gene flow

Effective Population Number

Linkage Disequilibrium Method

LD:

$$\hat{r}_\Delta = \frac{\hat{\Delta}}{\sqrt{[\hat{p}(1-\hat{p}) + (h_i - \hat{p}^2)][\hat{q}(1-\hat{q}) + (h_j - \hat{q}^2)]}},$$

$$\begin{array}{ll} S \geq 30 & S < 30 \\ \hline \hat{N}_e & \frac{1/3 + \sqrt{1/9 - 2.76\hat{r}^2}}{2\hat{r}^2} \quad \frac{0.308 + \sqrt{0.308^2 - 2.08\hat{r}^2}}{2\hat{r}^2} \end{array}$$

Uses the inverse relationship of linkage disequilibrium (LD) with the effective population number, where high LD reflects low Ne

LD is the non-random association between alleles at different loci

Ne estimates genetic drift and genetic diversity

Bayesian Clustering Analysis

STRUCTURE (Pritchard et al., 2000)

- Examines genetic structure admixture ancestry model and correlated allele frequencies, with location (LOCPRIOR) or without location input

Uses Bayesian Probability:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

A POSTERIORI
Probability of
Event A given
Evidence B

Probability of Evidence
B given Event A

A PRIORI probability
of Event A

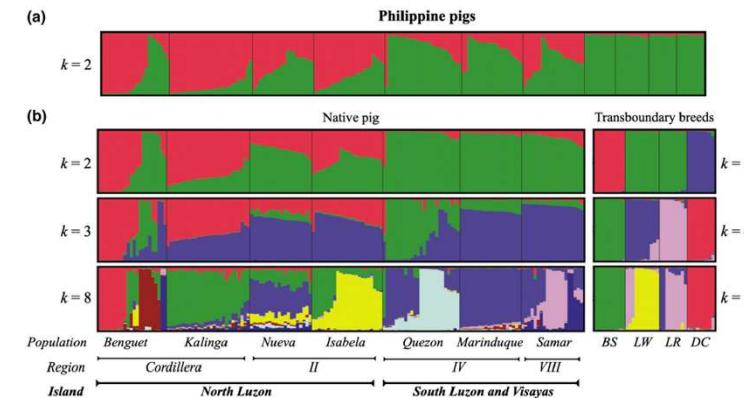
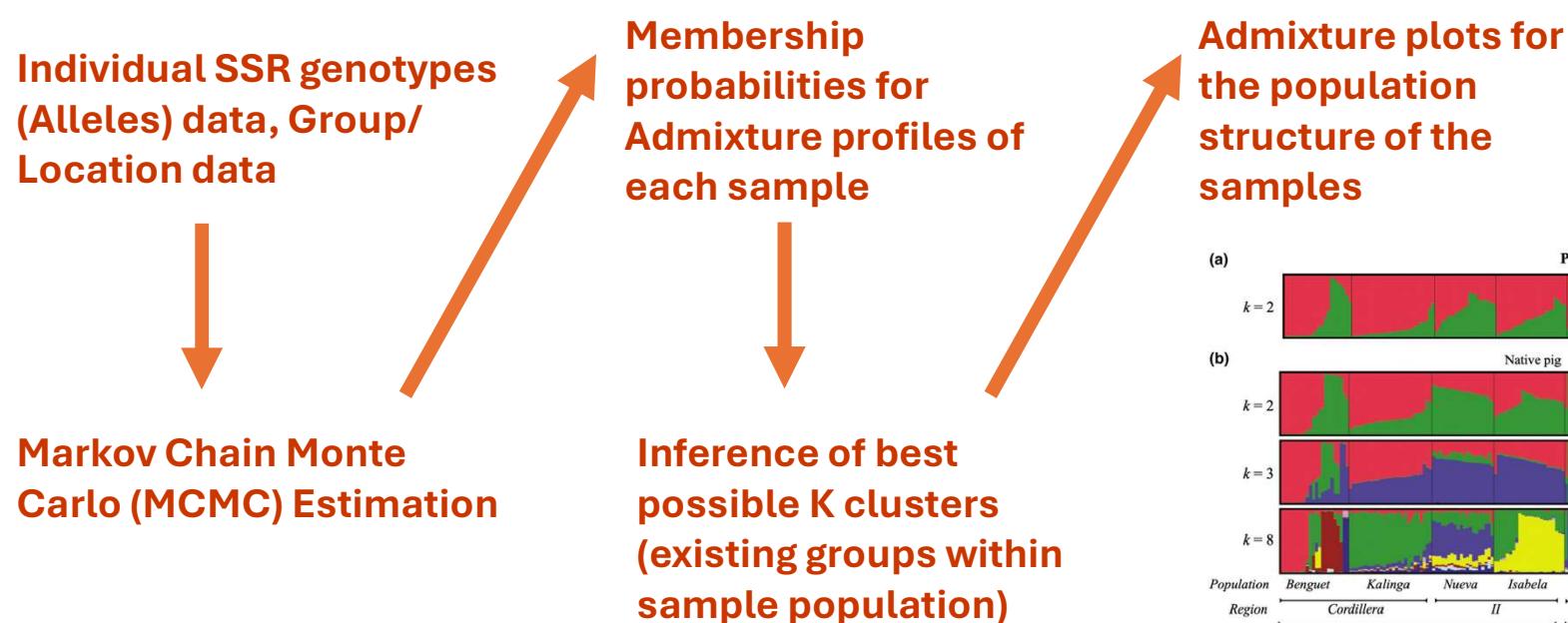
Probability of
Evidence B for all
possible outcomes

The diagram illustrates the components of Bayes' Theorem. At the top, the text 'Probability of Evidence B given Event A' is followed by a downward-pointing orange arrow. Below it, the formula $P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$ is shown. To the right of the formula, an orange arrow points left from the term $P(A)$ to the text 'A PRIORI probability of Event A'. To the left of the formula, an orange arrow points right from the term $P(B|A)$ to the text 'A POSTERIORI Probability of Event A given Evidence B'. At the bottom, an upward-pointing orange arrow is positioned between the terms $P(B|A)$ and $P(B)$, pointing to the text 'Probability of Evidence B for all possible outcomes'.

Bayesian Clustering Analysis

STRUCTURE (Pritchard et al., 2000)

- Examines genetic structure admixture ancestry model and correlated allele frequencies, with location (LOCPRIOR) or without location input





StruK2Stat

**Python-based package for streamlining
population diversity and admixture
analysis using microsatellite data**

Struk2Stat Pipeline

1 STRUCTURE File
*.txt

STRUCTURE
structureFunctions.py
*.exe app

structure2Genalex.py

GenAIEx
genalexFunctions.py

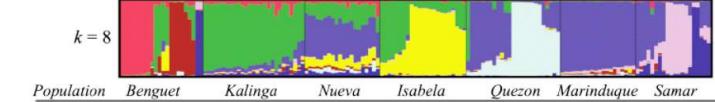
DISPAN
dendrogram.py

STRUCTUREHARVESTER
*.py app

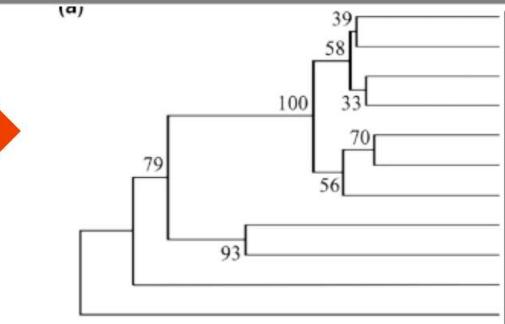
*.py app

CLUMPP
*.exe app

DISTRUCT
*.exe app



Population (R/P)	n	N _A	E _A	I	P _A	H _A	H _E	F	P _{free} ^{exc}	P _{free} ^{def}	Fis	N _E		
NL	B	22	6.850	3.906	1.563	0.050	0.666	0.75	0.090	2.000	.000*	0.113	3.9	
	K	27	8.150	3.840	1.548	0.400	0.662	0.71	0.042	2.219	.011*	0.065	125.9	
	N	20	8.850	5.032	1.810	0.400	0.795	0.81	-0.009	2.874	.459	0.015	69.9	
SLV	I	23	7.300	4.255	1.621	0.300	0.798	0.77	-0.066	495	2.057	-0.040	19.1	
	Q	25	7.500	4.003	1.557	0.100	0.758	0.73	-0.053	2.515	1.061	-0.034	14.2	
	M	20	7.800	4.502	1.69	0.400	0.776	0.78	-0.014	2.529	.460	0.012	44.7	
	S	20	7.950	5.021	1.759	0.300	0.706	0.80	0.099	2.000	.000*	0.125	22.5	
Mean native														
	T	BS	10	2.600	1.930	0.695	0.050	0.530	0.44	-0.254	.008*	1.998	-0.210	19.7
	LW	10	4.600	2.923	1.179	0.000	0.631	0.64	-0.037	2.488	1.256	0.017	16.7	
	LR	9	4.550	2.761	1.117	0.100	0.534	0.60	0.035	2.476	.048	0.114	133.0	
	DC	9	3.450	2.224	0.877	0.150	0.633	0.53	-0.270	.000*	1.818	-0.222	-	
Mean transboundary														
Mean all samples														
	SE		0.442	0.173	0.104	0.032		0.012	0.012					



MEGA
*.exe app

Struk2Stat Scripts

tkMenu.py

tkinter
sys

structureFunctions.py

tkinter
subprocess
os

structure2Genalex.py

pandas
np

genalexFunctions.py

pandas
numpy
defaultdict

dendrogram.py

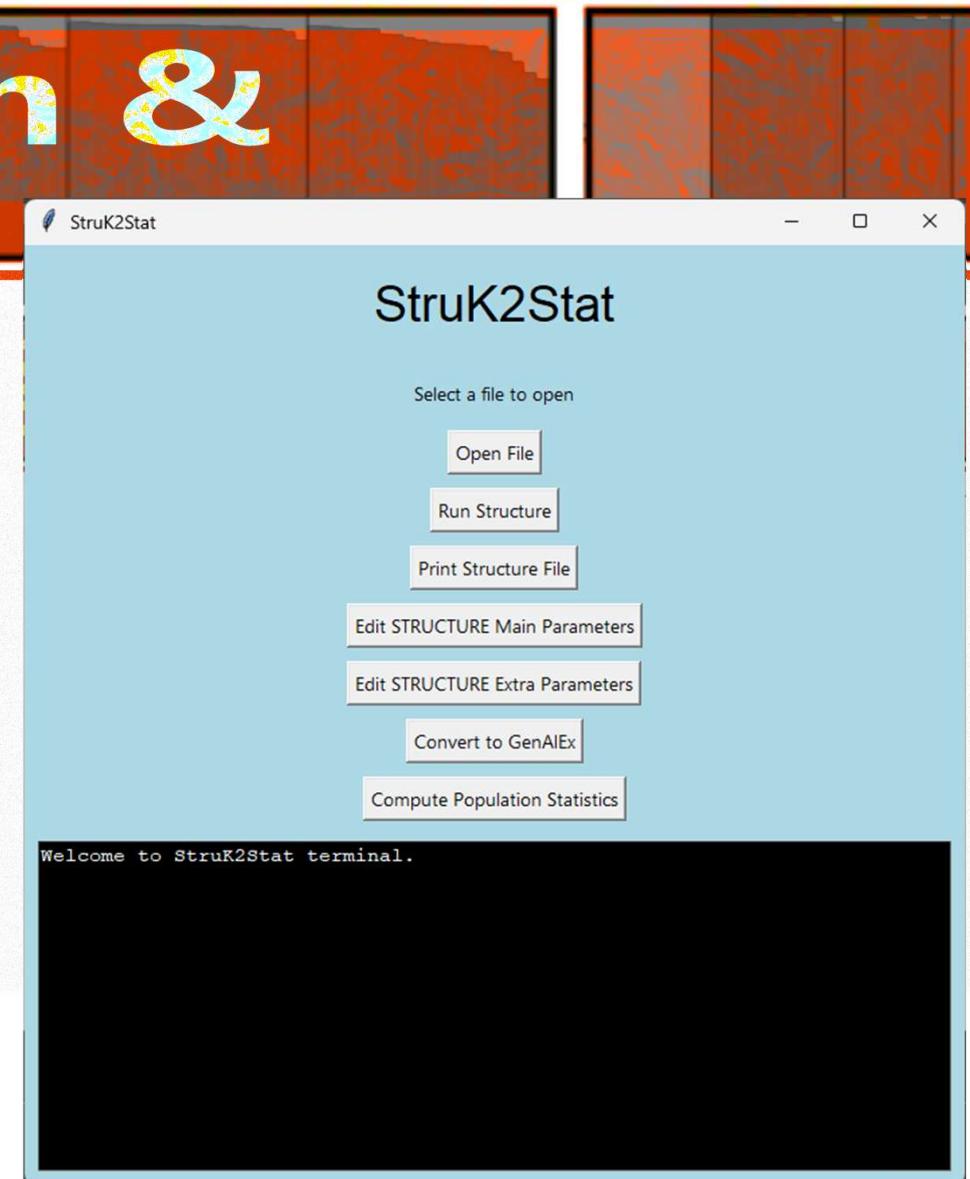
Simulated Run & Results

Current UI:

- tkinter menu

Current Functions:

- Opens and prints STRUCTURE.txt file
- Can run command-line based structure.exe
- An edit function for the Main and Extra Parameter files of STRUCTURE is available
- Can convert STRUCTURE.txt to GenAlex.txt format
- Computes partial population statistics



Simulated Run & Results



Current UI:

- tkinter menu

Current Functions:

- Opens and prints STRUCTURE.txt file
- Can run command-line based structure.exe
- An edit function for the Main and Extra Parameter files of STRUCTURE is available
- Can convert STRUCTURE.txt to GenAlex.txt format
- Computes partial population statistics

DM/	11	200	240	96	248	234	150	194	228	250	122	120	92	226	120	166	114
Names of loci:['IGF1', 'S0005', 'S0026', 'S0090', 'S0097', 'S0155', 'S0226', 'S0228', 'S0355', 'Sw122', 'Sw24', 'Sw240', 'Sw2406', 'Sw2410', 'Sw632', 'Sw72']																	
Number of loci:20																	
Number of individuals:196																	
Unique populations:[1', '10', '11', '2', '3', '4', '5', '6', '7', '8', '9']																	
Number of populations:11																	
Population 1 has 22 unique individuals.																	
Population 2 has 27 unique individuals.																	
Population 3 has 20 unique individuals.																	
Population 4 has 23 unique individuals.																	
Population 5 has 25 unique individuals.																	
Population 6 has 20 unique individuals.																	
Population 7 has 20 unique individuals.																	
Population 8 has 10 unique individuals.																	
Population 9 has 11 unique individuals.																	
Population 10 has 9 unique individuals.																	
Population 11 has 9 unique individuals. IGF1 S0005 S0026 S0090 S0097 S0155 S0226 S0228 S0355 Sw122 Sw24 Sw240 Sw2406 Sw2410 Sw632 Sw72																	
B050	1	192	204	92	246	210	146	198	-1	248	114	108	90	226	118	152	96
B050	1	196	216	94	250	240	158	198	-1	270	122	108	90	226	126	158	96
B051	1	192	204	92	242	240	142	194	-1	244	112	108	90	226	118	152	110
B051	1	192	240	94	246	240	154	210	-1	258	114	108	90	226	118	152	112
B053	1	198	218	92	242	204	154	190	-1	244	116	116	94	224	104	160	106
B053	1	198	246	98	246	234	156	194	-1	244	122	120	94	232	104	166	108
B055	1	202	240	96	244	204	152	182	-1	258	118	116	92	246	104	158	106
B055	1	202	240	104	246	238	158	196	-1	258	118	116	116	256	114	166	114
B056	1	192	218	92	246	204	154	180	-1	252	118	108	90	232	104	160	98
B056	1	196	246	98	252	238	156	182	-1	256	122	116	94	240	104	168	110
B060	1	192	204	90	246	210	146	198	-1	248	114	108	90	226	118	152	96
B060	1	196	216	92	250	240	158	198	-1	270	114	108	90	226	126	158	96
B062	1	192	218	94	244	238	154	182	222	244	118	108	94	226	104	160	108
B062	1	198	244	98	252	240	154	194	222	256	122	120	94	240	130	166	110

Simulated Run & Results

Current UI:

- tkinter menu

Current Functions:

- Opens and prints STRUCTURE.txt file
- Can run command-line based **structure.exe**
- An edit function for the Main and Extra Parameter files of STRUCTURE is available
- Can convert STRUCTURE.txt to GenAlex.txt format
- Computes partial population statistics

```
2 populations assumed
1000 Burn-in period
10000 Reps

-----
Estimated ln Prob of Data      = -4882.3
Mean value of ln likelihood   = -4783.6
Variance of ln likelihood     = 197.5
Mean value of alpha           = 0.5029

Mean value of Fst_1            = 0.1494
Mean value of Fst_2            = 0.0682

-----
Proportion of membership of each pre-defined
population in each of the 2 clusters

Given      Inferred Clusters          Number of
Pop        1       2                   Individuals

1:        0.619  0.381                22
2:        0.785  0.215                27
3:        0.560  0.440                23
4:        0.193  0.807                25
5:        0.180  0.820                20
6:        0.236  0.764                20

-----
Final results printed to file outfile_f

Error:
```

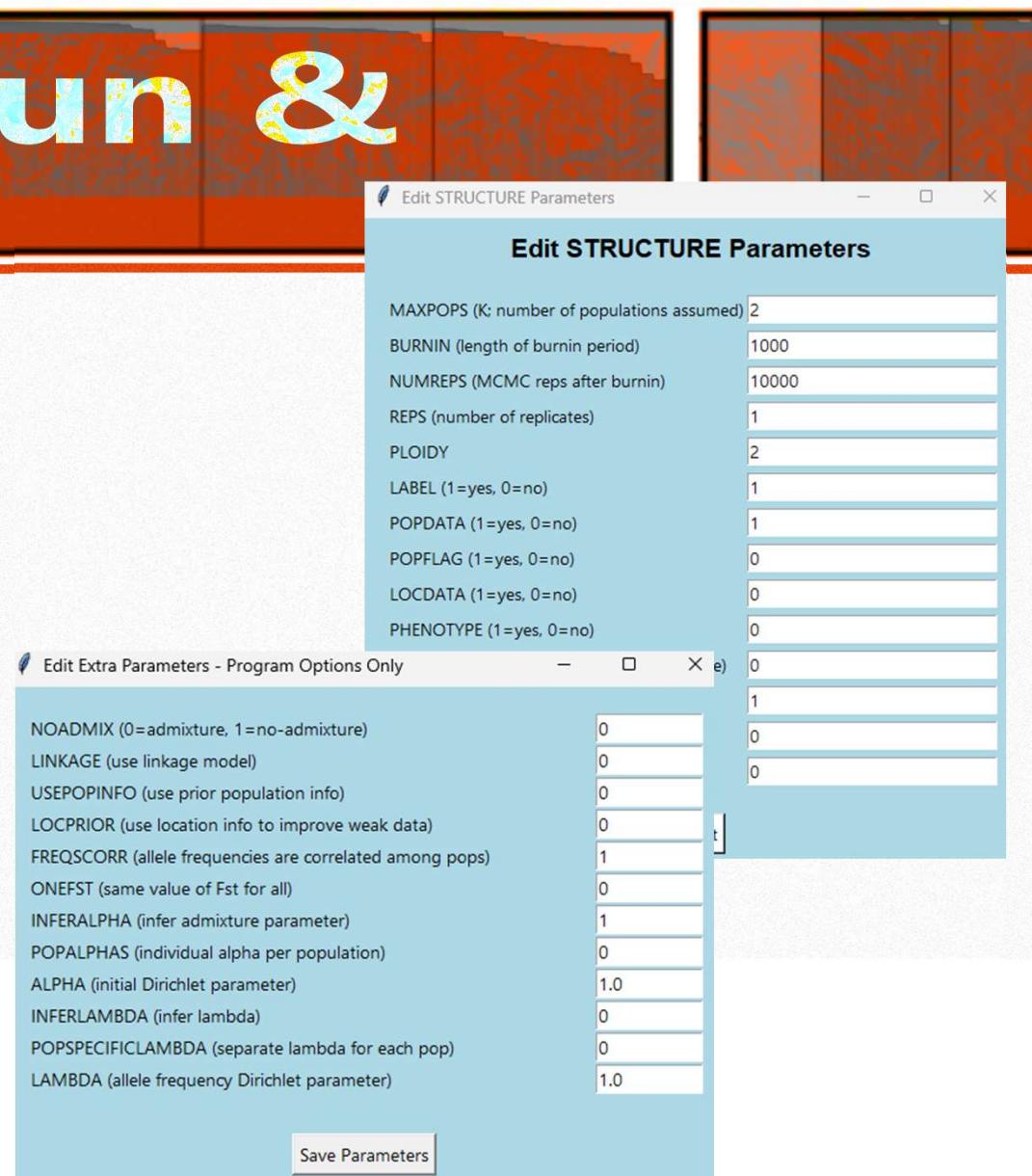
Simulated Run & Results

Current UI:

- tkinter menu

Current Functions:

- Opens and prints STRUCTURE.txt file
- Can run command-line based structure.exe
- An edit function for the Main and Extra Parameter files of STRUCTURE is available
- Can convert STRUCTURE.txt to GenAlex.txt format
- Computes partial population statistics



Simulated Run & Results



Current UI:

- tkinter menu

Current Functions:

- Opens and prints STRUCTURE.txt file
- Can run command-line based structure.exe
- An edit function for the Main and Extra Parameter files of STRUCTURE is available
- Can convert STRUCTURE.txt to GenAlex.txt format
- Computes partial population statistics

```
Names of loci:['s0005', 's0026', 's0090', 's0155', 's0228', 'sw240', 'sw830', 'sw857', 'SWR1941', 'sw2410']
Number of loci:10
Number of individuals:137
Unique populations:['1', '2', '3', '4', '5', '6']
Number of populations:6
Population 1 has 22 unique individuals.
Population 2 has 27 unique individuals.
Population 3 has 23 unique individuals.
Population 4 has 25 unique individuals.
Population 5 has 20 unique individuals.
Population 6 has 20 unique individuals.
Individuals per locus: 22 27 23 25 20 20
Number of rows: 137
Data written to GenAlEx file: output_genalex_file.txt
Creating Genalex DataFrame from data...
GenAlex DataFrame:
   Population Individual    s0005    s0026    s0090    s0155    s0228    sw240    sw830    sw857    SWR1941
0            1        B050  204\t216  92\t94  246\t250  146\t158  0\t0  90\t90  182\t182  152\t152  206\t224
1            1        B051  204\t240  92\t94  242\t246  142\t154  0\t0  90\t90  178\t186  156\t158  206\t216
2            1        B053  218\t246  92\t98  242\t246  154\t156  0\t0  94\t94  172\t184  152\t160  216\t220
3            1        B055  240\t240  96\t104  244\t246  152\t158  0\t0  92\t116  178\t182  150\t150  216\t216
4            1        B056  218\t246  92\t98  246\t252  154\t156  0\t0  90\t94  172\t182  150\t160  218\t220
..          ...
132           6        E531  240\t246  92\t92  244\t248  154\t158  240\t242  90\t94  178\t182  142\t156  206\t216
133           6        E533  204\t240  92\t92  244\t250  158\t158  240\t242  90\t94  178\t182  142\t156  206\t216
134           6        E534  218\t244  92\t92  246\t250  158\t158  240\t242  90\t90  184\t184  154\t158  206\t220
135           6        E536  228\t246  92\t106  242\t250  146\t150  234\t242  94\t94  178\t186  152\t158  218\t220
136           6        E537  222\t244  92\t96  242\t248  158\t160  240\t242  92\t106  178\t178  152\t154  216\t220
[137 rows x 12 columns]
```

Simulated Run & Results



Current UI:

- tkinter menu

Current Functions:

- Opens and prints STRUCTURE.txt file
- Can run command-line based structure.exe
- An edit function for the Main and Extra Parameter files of STRUCTURE is available
- Can convert STRUCTURE.txt to GenAlex.txt format
- Computes partial population statistics

8	8	10	Nan	2.6	Nan	0.695128	0.05	Nan	Nan	Nan	Nan	Nan	Nan	Nan
9	9	11	Nan	4.6	Nan	1.179235	NaN	Nan	Nan	Nan	Nan	Nan	Nan	Nan
10	10	9	Nan	4.55	Nan	1.116596	0.1	Nan	Nan	Nan	Nan	Nan	Nan	Nan
11	11	9	Nan	3.45	Nan	0.87687	0.15	Nan	Nan	Nan	Nan	Nan	Nan	Nan
Result Table after adding expected heterozygosity (He) :														
Population	N	Ne	Na	Ea	I	Pa	Ho	He	Ht	Fst	Fis	Fit		
1	1	22	Nan	6.85	Nan	1.562905	0.05	Nan	0.730032	Nan	Nan	Nan	Nan	Nan
2	2	27	Nan	8.15	Nan	1.54802	0.4	Nan	0.693853	Nan	Nan	Nan	Nan	Nan
3	3	20	Nan	8.85	Nan	1.809537	0.4	Nan	0.786438	Nan	Nan	Nan	Nan	Nan
4	4	23	Nan	7.3	Nan	1.621393	0.3	Nan	0.750978	Nan	Nan	Nan	Nan	Nan
5	5	25	Nan	7.5	Nan	1.556961	0.1	Nan	0.718546	Nan	Nan	Nan	Nan	Nan
6	6	20	Nan	7.8	Nan	1.693396	0.4	Nan	0.763286	Nan	Nan	Nan	Nan	Nan
7	7	20	Nan	7.95	Nan	1.758551	0.3	Nan	0.784097	Nan	Nan	Nan	Nan	Nan
8	8	10	Nan	2.6	Nan	0.695128	0.05	Nan	0.420255	Nan	Nan	Nan	Nan	Nan
9	9	11	Nan	4.6	Nan	1.179235	NaN	Nan	0.611043	Nan	Nan	Nan	Nan	Nan
10	10	9	Nan	4.55	Nan	1.116596	0.1	Nan	0.562664	Nan	Nan	Nan	Nan	Nan
11	11	9	Nan	3.45	Nan	0.87687	0.15	Nan	0.495986	Nan	Nan	Nan	Nan	Nan
Computing heterozygosity and homozygosity ...														
Result Table after adding observed heterozygosity (Ho) :														
Population	N	Ne	Na	Ea	I	Pa	Ho	He	Ht	Fst	Fis	Fit		
1	1	22	Nan	6.85	Nan	1.562905	0.05	0.654545	0.730032	Nan	Nan	Nan	Nan	Nan
2	2	27	Nan	8.15	Nan	1.54802	0.4	0.657407	0.693853	Nan	Nan	Nan	Nan	Nan
3	3	20	Nan	8.85	Nan	1.809537	0.4	0.795	0.786438	Nan	Nan	Nan	Nan	Nan
4	4	23	Nan	7.3	Nan	1.621393	0.3	0.797826	0.750978	Nan	Nan	Nan	Nan	Nan
5	5	25	Nan	7.5	Nan	1.556961	0.1	0.756	0.718546	Nan	Nan	Nan	Nan	Nan
6	6	20	Nan	7.8	Nan	1.693396	0.4	0.76	0.763286	Nan	Nan	Nan	Nan	Nan
7	7	20	Nan	7.95	Nan	1.758551	0.3	0.7025	0.784097	Nan	Nan	Nan	Nan	Nan
8	8	10	Nan	2.6	Nan	0.695128	0.05	0.515	0.420255	Nan	Nan	Nan	Nan	Nan
9	9	11	Nan	4.6	Nan	1.179235	NaN	0.6	0.611043	Nan	Nan	Nan	Nan	Nan
10	10	9	Nan	4.55	Nan	1.116596	0.1	0.516667	0.562664	Nan	Nan	Nan	Nan	Nan
11	11	9	Nan	3.45	Nan	0.87687	0.15	0.633333	0.495986	Nan	Nan	Nan	Nan	Nan

Recommendations



GENETIC DIVERSITY BASED ON THE 20 MICROSATELLITE LOCI

Mean allelic patterns across populations

Region		Type	Province	n	Na	Ea	Pa	Ho	He
North Luzon	Benguet	22	6.85	3.906	0.05	0.666	0.75		
	Kalinga	27	8.15	3.84	0.4	0.662	0.71		
	Nueva								
	Vizcaya	20	8.85	5.032	0.4	0.795	0.81		
	Isabela	23	7.3	4.255	0.3	0.798	0.77		
South Luzon - Visayas	Quezon	25	7.5	4.003	0.1	0.758	0.73		
	Marinduque	20	7.8	4.502	0.4	0.776	0.78		
	Samar	20	7.95	5.021	0.3	0.706	0.8		
22.									
Mean native		3	7.7771	4.3666	0.279	0.737	0.764		
Transboundary	Berkshire	10	2.6	1.93	0.05	0.53	0.44		
	Large White	10	4.6	2.923	0	0.631	0.64		
	Landrace	9	4.55	2.761	0.1	0.534	0.6		
	Duroc	9	3.45	2.224	0.15	0.633	0.53		
Mean transboundary		9.5	3.8	2.46	0.075	0.582	0.552		
17.									
Mean all samples		6	6.327	3.673	0.205	0.681	0.687		

Na – Mean number of alleles per loci across popns

Ea – Mean effective number of alleles (Equivalent number of equally frequent alleles in an ideal population)

$$E_a = \frac{1}{1 - H_e}$$

Pa – Private Alleles (Unique alleles to a single population)

Ho – Observed Heterozygosity
He – Expected Heterozygosity

F-statistics

		F_{IS}
North Luzon	Benguet	0.113
	Kalinga	0.065
	Nueva Vizcaya	0.015
	Isabela	-0.04

		F_{IS}
South Luzon - Visayas	Quezon	-0.034
	Marinduque	0.012
	Samar	0.125

	F_{IS}
Transbound ary	Berkshire
	Large White
	Landrace
	Duroc

Stepwise Mutation Model

Highland (B) vs Lowland (K, N, I, Q, M, S)

$$F_{st} = 0.226$$

Infinite Allele Model

$$F_{st} = 0.059$$

Highland (B) vs Lowland (K, N, I, Q, M, S) vs Transboundary (BS, LW, LR, DC)

$$F_{st} = 0.116$$

$$F_{st} = 0.121$$

F_{IS} values suggested no to moderate inbreeding among PhNPs (-0.04 – 0.125)

Regional variation between highland and lowland PhNP was high (F_{st} : 0.226)

Results show that the SMM attributes larger variance to the regional differences between highland and lowland PhNPs, indicating that it is a better model than IAM when the analysis includes the highland PhNPs (B population)

AMOVA

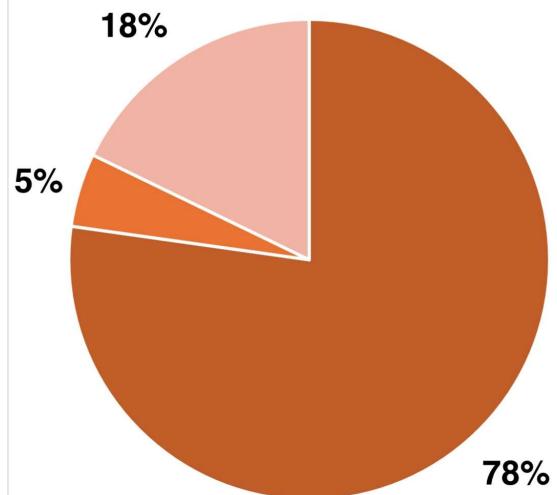
AMOVA

Stepwise Mutation Model

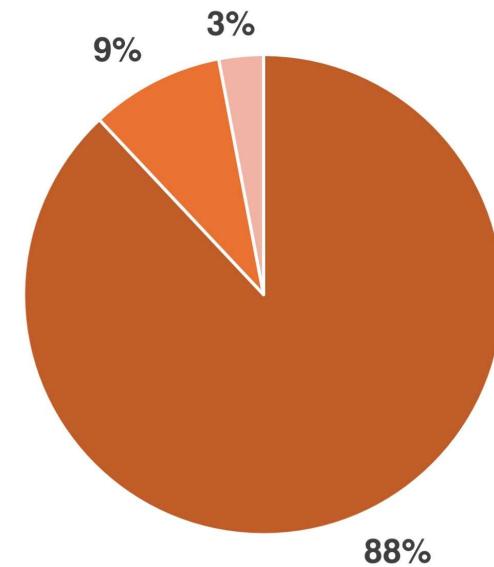
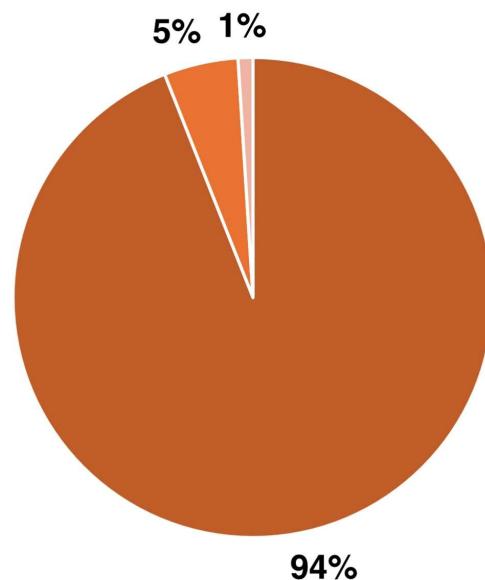
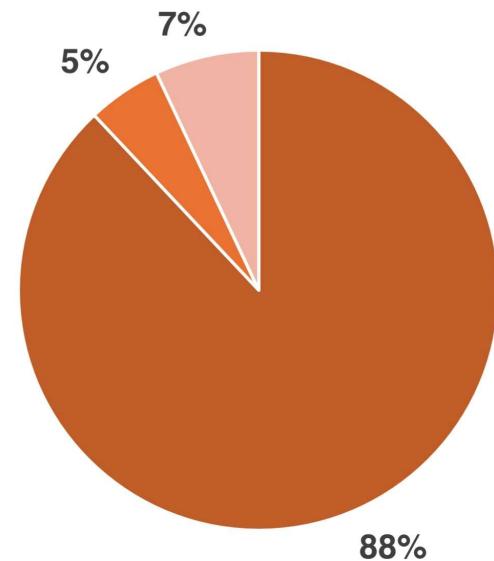
- Within Population
- Between Populations within Regions
- Between Regions

Infinite Allele Model

All Native (# of popns =7)



With Transboundary (# of popns =11)



Effective Population Size (Ne)

		n	Ne
North Luzon	Benguet	22	3.9
	Kalinga	27	125.9
	Nueva Vizcaya	20	69.9
	Isabela	23	19.1

Ne ranged from 3.9 (Benguet) to 125.9 (Kalinga), with the majority (except Kalinga and Nueva Vizcaya) having critical Ne values below 50.

The low and declining effective population sizes in PhNPs highlight the need for immediate conservation actions to prevent further loss of genetic diversity

		n	Ne
South Luzon - Visayas	Quezon	25	14.2
	Marinduque	20	44.7
	Samar	20	22.5

		n	Ne
Transboun dary	Berkshire	10	19.7
	Large White	10	16.7
	Landrace	9	133
	Duroc	9-	-

The Pairwise Effective Number of Migrants and Standardized Fst (F'_st) between populations were computed to further dissect genetic diversity

Pairwise Standardized Fst (F'st; below diagonal) and Nm (above diagonal)

	Benguet	Kalinga	Nueva Vizcaya	Isabela	Quezon	Marinduque	Samar	Berkshire	Large White	Landrace	Duroc
Benguet	-	5.202	8.666	5.661	3.273	3.834	3.728	0.837	1.582	1.338	0.942
Kalinga	0.163	-	4.877	6.085	2.269	2.536	2.687	0.733	1.088	1.114	0.806
Nueva Vizcaya	0.153	0.165	-	9.29	8.042	10.138	8.46	1.12	2.312	2.5	1.524
Isabela	0.175	0.15	0.084	-	3.557	4.324	3.824	0.872	1.524	1.562	1.202
Quezon	0.346	0.397	0.174	0.279	-	9.869	6.625	1.068	2.124	3.152	1.285
Marinduque	0.326	0.379	0.145	0.243	0.137	-	8.174	1.154	2.606	2.788	1.554
Samar	0.312	0.317	0.152	0.248	0.183	0.156	-	1.085	2.288	2.177	1.278
Berkshire	0.7	0.714	0.616	0.667	0.563	0.542	0.614	-	0.777	0.743	0.491
Large White	0.581	0.693	0.485	0.564	0.44	0.372	0.446	0.65	-	2.047	0.933
Landrace	0.626	0.639	0.432	0.538	0.328	0.351	0.455	0.654	0.41	-	0.832
Duroc	0.673	0.7	0.21	0.556	0.535	0.3485	0.579	0.761	0.648	0.611	-

Pairwise Standardized Fst (F'st; below diagonal) and Nm (above diagonal)

	Benguet	Kalinga	Nueva Vizcaya	Isabela	Quezon	Marinduque	Samar	Berkshire	Large White	Landrace	Duroc
Benguet											
Kalinga	0.163										
Nueva Vizcaya	0.153	0.165									
Isabela	0.175	0.15	0.084								
Quezon	0.346	0.397	0.174	0.279							
Marinduque	0.326	0.379	0.145	0.243	0.137						
Samar	0.312	0.317	0.152	0.248	0.183	0.156					
Berkshire	0.7	0.714	0.616	0.667	0.563	0.542	0.614				
Large White	0.581	0.693	0.485	0.564	0.44	0.372	0.446	0.65			
Landrace	0.626	0.639	0.432	0.538	0.328	0.351	0.455	0.654	0.41		
Duroc	0.673	0.7	0.21	0.556	0.535	0.3485	0.579	0.761	0.648	0.611	

Pairwise F_{ST} ranged from 0.145 to 0.397 between NL and SLV , suggesting strong genetic differentiation between populations on different islands.

Within the islands, pairwise F_{ST} values were moderate:

- F_{ST} between the NL populations (B, K, N, I) ranged from 0.084 to 0.175, and
- those between the SLV (Q, M, S) ranged from 0.137 to 0.183.

On the other hand, low F_{ST} values (below 0.150) were observed between the following populations: N and I, N and M, and Q and M.

Pairwise Standardized Fst (F'st; below diagonal) and Nm (above diagonal)

	Benguet	Kalinga	Nueva Vizcaya	Isabela	Quezon	Marinduque	Samar	Berkshire	Large White	Landrace	Duroc
Benguet	-										
Kalinga	0.163-	-									
Nueva Vizcaya	0.153	0.165-	-								
Isabela	0.175	0.15	0.084-	-							
Quezon	0.346	0.397	0.174	0.279-	-						
Marinduque	0.326	0.379	0.145	0.243	0.137-	-					
Samar	0.312	0.317	0.152	0.248	0.183	0.156-	-				
Berkshire	0.7	0.714	0.616	0.667	0.563	0.542	0.614-	-			
Large White	0.581	0.693	0.485	0.564	0.44	0.372	0.446	0.65-	-		
Landrace	0.626	0.639	0.432	0.538	0.328	0.351	0.455	0.654	0.41-	-	
Duroc	0.673	0.7	0.21	0.556	0.535	0.3485	0.579	0.761	0.648	0.611-	-

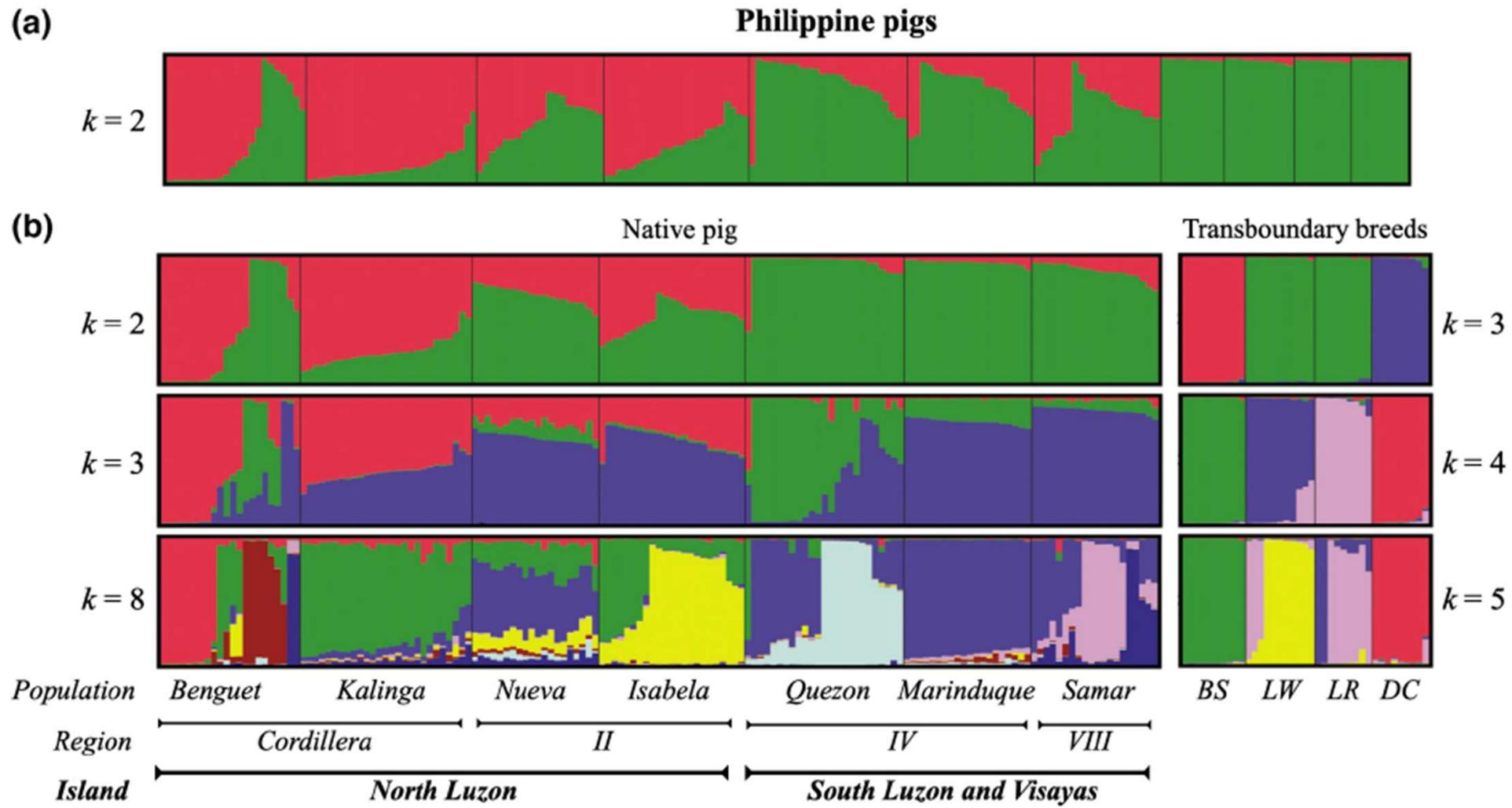
The results indicate that PhNP populations from different islands (NL and SLV) are more likely to have stronger genetic differentiation than those from the same island.

Furthermore, pairwise F_{ST} showed high differentiation between the PhNPs and transboundary breeds, with values ranging from 0.328 (Q vs. LR pigs) to 0.714 (K vs. BS pigs).

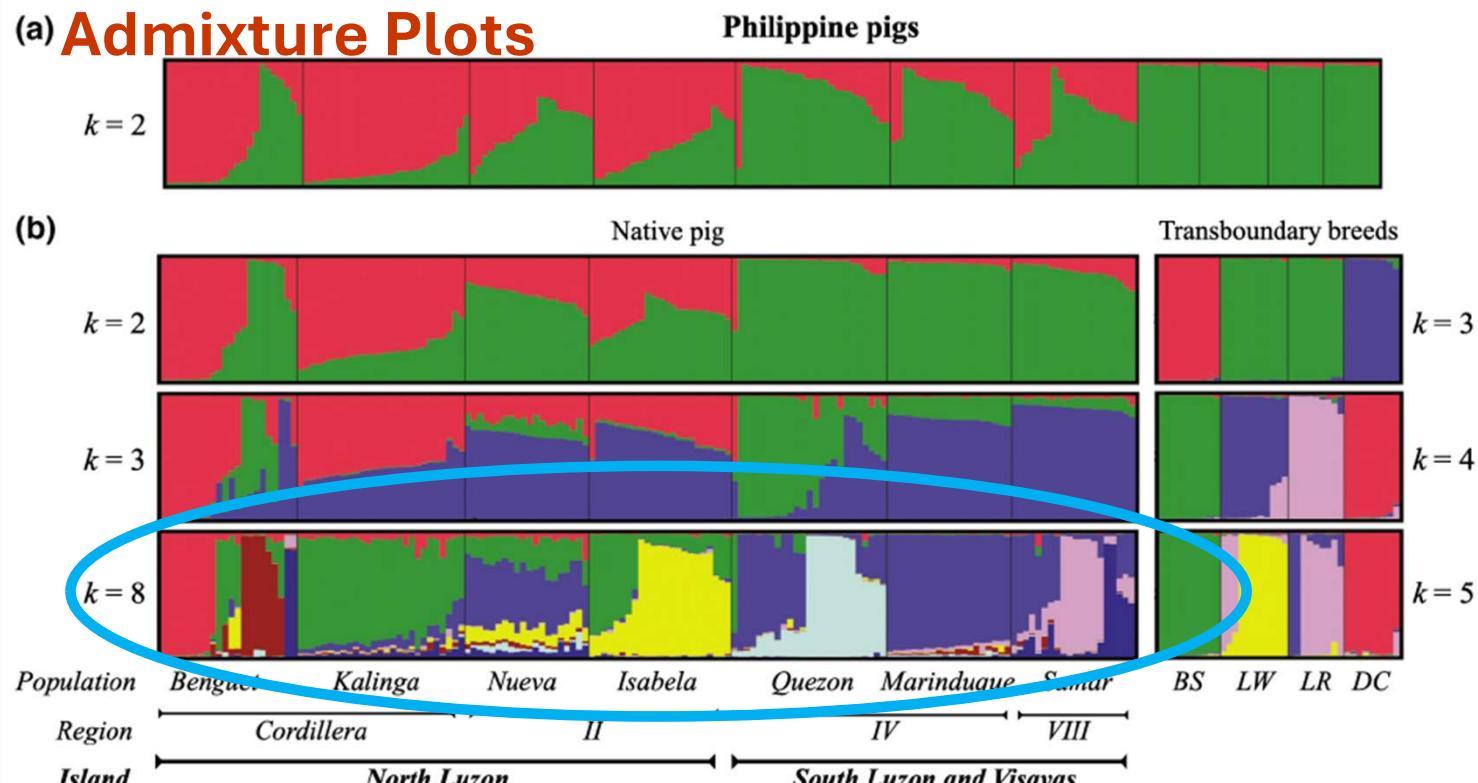
Pairwise Standardized Fst (F'st; below diagonal) and Nm (above diagonal)

BAYESIAN CLUSTERING ANALYSIS

Admixture Plots



BAYESIAN CLUSTERING ANALYSIS

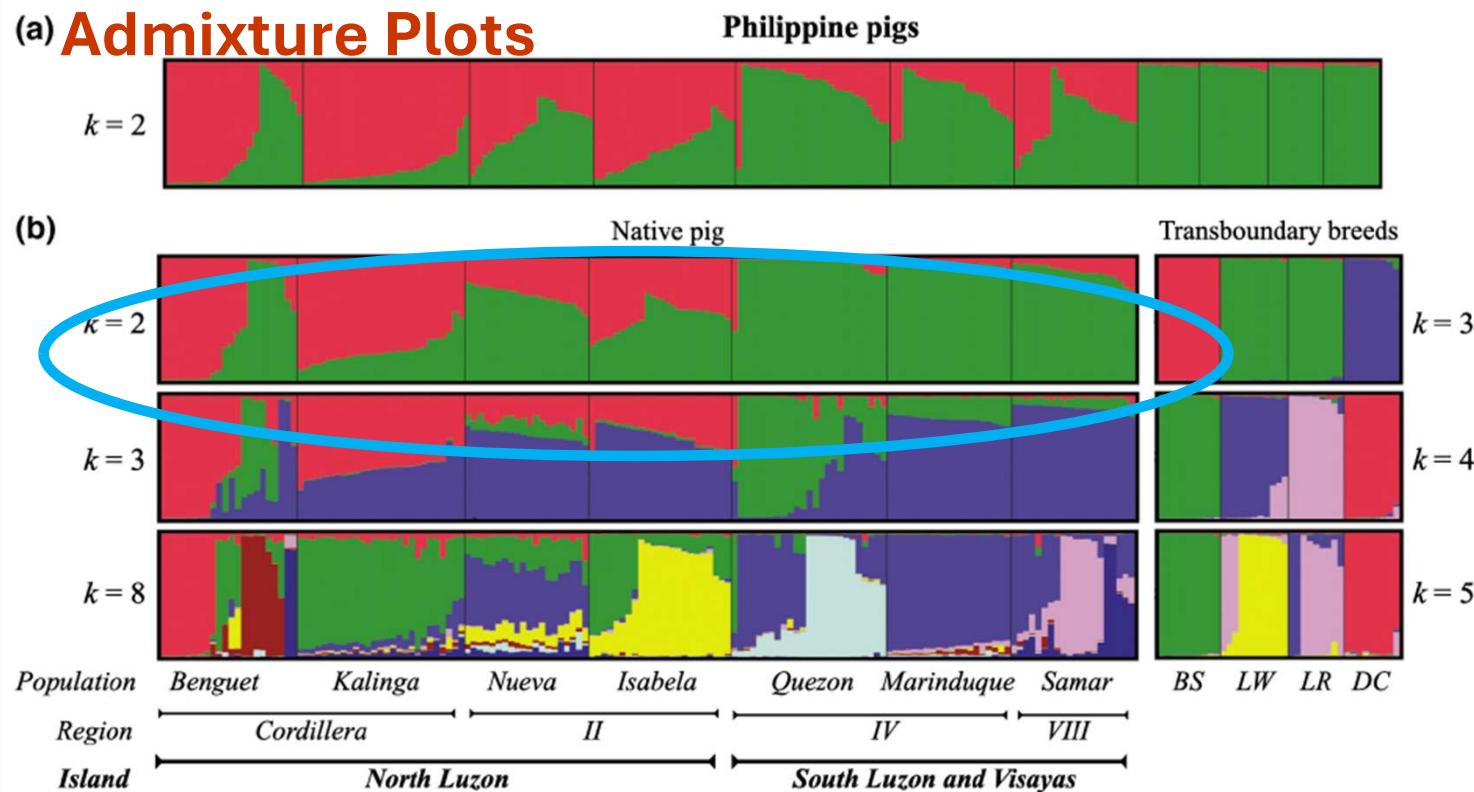


At least six populations were distinct (B, K, I, Q, M, and S) and showed further subdivision below provincial boundaries.

Genetic substructure could be possible at the municipality level especially for B, I, Q, and S.

The homogenous pattern of K and M populations could be due to cultural preference for K native pigs within K, and the local ordinance in M that prohibits outside pigs from entering the island.

BAYESIAN CLUSTERING ANALYSIS



Previous analysis of the mitochondrial DNA (mtDNA) D-loop region revealed three ancestral origins of the PhNPs used in this study, that is East Asia (haplotype D2), Southeast Asia (haplotype D7), and the Cordillera clade (Banayo et al., 2023).

The initial ancestral polymorphisms between the D2 and D7 clades could explain the two initial clusters ($k = 2$) differentiating the NL and SLV regions, respectively

It was previously hypothesized that the pigs of the D7 clade originated from the D2 clade via a genetic bottleneck event upon dispersal to Southeast Asia (Banayo, 2023). Thereafter, upon dispersal to various provinces in the Philippines, members of each clade formed further population structure via genetic drift (i.e., Q, M, and S of the D7 clade) due to the varied agroclimatic conditions and management practices of each province.

Conclusions

- **PhNP exhibited high allelic richness**
 - PhNP is an important breed as it is a repository of genetic variation (possibly novel alleles).
 - This allelic richness may provide clues about how its fitness has been improved despite production constraints.
- **Low admixture between native and transboundary breeds**
 - This should be maintained in favor of conservation of the distinct PhNP populations.
 - This also confirms the historical disregard of native pigs by income-driven production systems
- **The PhNP remains sufficiently differentiated both from transboundary breeds and from each other.**
 - This finding implies the possibility of heterosis by crossbreeding the PhNPs with transboundary breeds and even with other PhNP populations, similar to how the extinct *Berkjala* was developed by crossbreeding SL native pigs with Berkshire pigs in the early 1900s.

Conclusions

- **Benguet population was the most differentiated but suffered from low effective population size.**
 - The majority of the PhNPs (Benguet, Isabela, Quezon, Marinduque, and Samar populations) have a critically low Ne (below 50), necessitating the need for immediate conservation actions.
 - Prioritizing the conservation of the B population is thus a national concern not only to maintain ecological balance but also to protect indigenous and local knowledge, a progressive approach known as biocultural conservation.
 - Understanding indigenous ways of life can inform modern society regarding the sustainable utilization of animal genetic resources.

Conclusions

- Despite Kalinga being in the same administrative region as Benguet (Cordillera) and having indigenous communities with their own rituals and traditions, the pig population in Kalinga had a high effective population size (Ne : 125.9), presumably due to introgression of wild pig alleles via interspecific hybridization.
 - The majority of the PhNPs (Benguet, Isabela, Quezon, Marinduque, and Samar populations) have a critically low Ne (below 50), necessitating the need for immediate conservation actions.
 - Having a high Ne is beneficial in that it increases the fitness of organisms.
 - In addition to the benefits of having high Ne , hybridization is known to introduce novel adaptive variations and further increase fitness.
 - The study suggests that a high Ne in native pigs can be achieved by hybridization with wild pigs, thus necessitating the co-management of PhNPs and Philippine wild pigs.

Conclusions

- The study showed that the Benguet, Isabela, Quezon, Marinduqe, and Samar populations (N_e values <50) are immediate conservation priorities.
 - It was suggested that a minimum effective population size of 50–500 is recommended to maintain short-term population viability, with the ultimate goal of achieving an effective population size >5000 to ensure long-term viability.
 - Therefore, all PhNP populations (N_e values <500) must be conserved to facilitate long-term viability.
 - Additionally, PhNP is an important repository of genetic variation, as shown by its high allele count (N_a : 7.771) and heterozygosity (H_o : 0.737), thereby further warranting investments in their conservation.
 - This study emphasized the importance of PhNP as a repository of genetic variation, primarily for adaptation traits.

Recommendations

- Harness the diversity of the PhNP through crossbreeding (among PhNP populations or between the PhNP and transboundary breeds) to address the needs of modern farmers for a faster-growing native pig while systematically avoiding population admixture
- Indigenous communities that raise PhNPs for cultural traditions, rather than income, must be incentivized to increase private benefits for utilizing PhNPs and subsequently sustaining on-farm conservation.
- Continuously monitor PhNP diversity across different time points to assess the impact of interventions.