Combining approaches for the analysis of the genetic structure of the Avileña-Negra Ibérica beef cattle breed

Martín Collado, D., Abraham, K.J., Toro, M.A., Carabaño, M.J., Rodriguez, S.T., & Diaz, C.

63rd Annual EAAP Meeting
Session 42. Genetics free communications
Bratislava, Slovakia. 27-31 August, 2012.

Introduction. **Study motivation**

- Avileña-Negra Ibérica (ANI) is under a selection program since 1990.

- Genomic selection could be a potential tool to be used in a context of a multibreed Spanish evaluation.

- ANI breed is involved in GWAS.

- Population stratification has raised concerns in genomic selection (Deckers 2007) and GWAS (Price et al. 2006, 2011).

- Sample of herds for deep phenotyping. **Random?**

Objective

- Determine if there is a hidden structure in the Avileña-Negra Ibérica breed.

Data

- ANI production system is fully extensive.

- Since 2005 ANI association has made a great effort to assess pedigree using a panel of 17 microsatellites.

- As a result we have a database of 13343 individuals from 59 herds.

Microsatellites	n of alleles	H Obs	H Exp
BM1818	9	0.626	0.654
BM1824	6	0.680	0.717
BM2113	11	0.817	0.852
CSRM60	10	0.643	0.689
ETH3	10	0.717	0.747
ETH10	9	0.690	0.721
ETH185	13	0.706	0.727
ETH225	8	0.731	0.762
ILSTS006	12	0.790	0.836
INRA005	5	0.643	0.661
INRA023	11	0.741	0.765
INRA063	6	0.557	0.577
SPS115	8	0.374	0.387
TGLA53	18	0.781	0.811
TGLA122	22	0.831	0.858
TGLA126	8	0.684	0.726
TGLA227	13	0.803	0.830

- None in H-W equilibrium

- Population stratification?

Methods. **Tools to infer population structure**

- We used three tools to infer the population structure: There are two main

  approaches to infer population structure:

1. Model based: STRUCTURE software (Pritchard et al. 2000).

2. Distance based.

   - Based on maximazing Nei's distance (Rodriguez-Ramilo et al. 2009).

   - Based on graphical clustering algorithm (K.J. Abraham and Rohan Fernando

     2012)

STRUCTURE (Pritchard et al. 2000)

■ We used Evanno et al. (2005) criterion to determine the number of clusters (subpopulations).
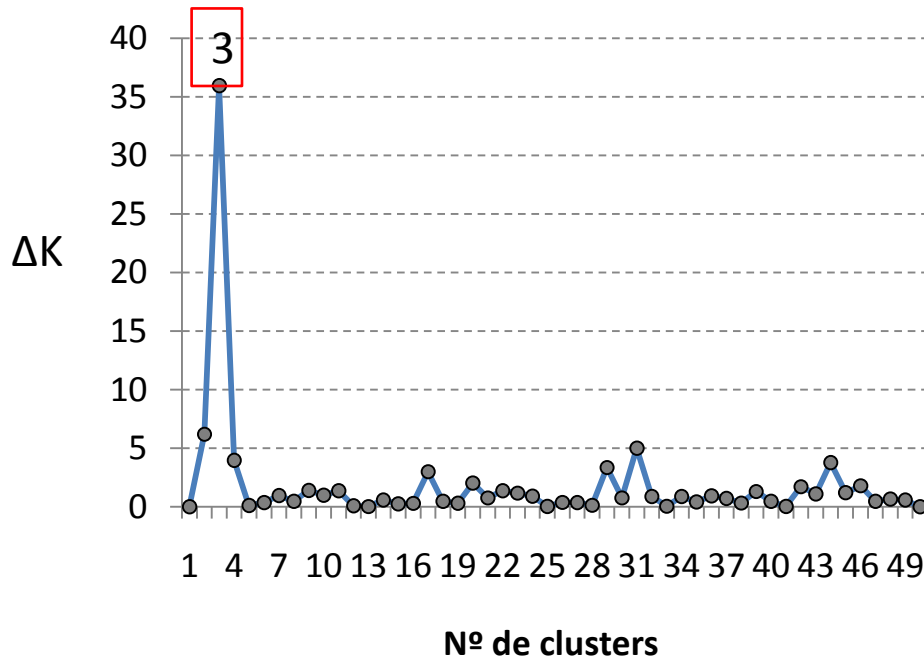
$\Delta K = m(|L(K + 1) - 2\,L(K) + L(K - 1)|)/std[L(K)]$

■ We found H-W equilibrium in:

1. Cluster 1 (n=3077): 12 microsatellites

2. Cluster 2 (n=6055): 4 microsatellites

3. Cluster 3 (n=4211): 7 microsatellites

Nei's distance maximizing method
(Rodriguez-Ramilo et al. 2009)

- We tried from 2 to 5 clusters

- It divided the population in clusters of similar sizes

	2 clusters	
	Cluster 1	**Cluster 2**
N	6670	6673
H-W	1 micro	1 micro

	3 clusters		
	Cluster 1	**Cluster 2**	**Cluster 3**
N	4569	4375	4399
H-W	1 micro	2 micros	2 micros

- This approach does not identify any structure

Graphical clustering: molecular coancestry
(K.J. Abraham and Rohan Fernando 2012)

- It was developed by J.K.Abraham to identify gene networks and it is being adapted to study the population structure.

- It has two steps:

  1. To identify the individuals that will be nodes of the clusters.

  2. Clusters construction: Aggregating individuals to each of the nodes.

- Key points:

  1. Determine the threshold to find the nodes.

  2. Determine the threshold to aggregate individuals.

Graphical clustering. **Identification of cluster nodes**

- Nodes are individuals that have:

    - A minimum coancestry (provided a threshold) among them.

    - A maximum average coancestry with the rest of the population.

- We used 150 sweeps per threshold.

- We used several thresholds.

Nº of nodes	Molecular coancestry treshold for identifying clusters nodes						
	0,147	0,162	0,177	0,191	0,201	0,221	0,235
2	18%	1%	4%				
3	68%	43%	42%				
4	14%	55%	47%	3%			
5		1%	7%	25%			
6			1%	45%	4%		
7				24%	26%		
8				5%	34%	2%	
9					26%	7%	
10					5%	11%	
11					4%	25%	1%
12					1%	28%	2%
13						18%	5%
14						8%	11%
15						1%	21%
16						1%	21%
17							20%
18							9%
19							7%
20							2%

Graphical clustering. Aggregating individuals

- The key point is to define the threshold that determines the minimum coancestry value to aggregate individuals to a cluster.

- An individual will be aggregated iff his coancestry with all individuals previously allocated into the cluster is above the threshold. ☹

- We tried for the case of 3 clusters (similar to STRUCTURE) and we used:

    - an aggregation threshold of 0.184.

    - 150 sweeps.

- Animals were assigned to a cluster in two steps:

    - When they appear at least in 10% of the sweeps associated to a node.

    - The posterior probability of being associated to a node was the highest.

Graphical clustering. Aggregating individuals

- The clusters built were:

1. Cluster 1 (n=2807): H-W in 5 microsatellites

2. Cluster 2 (n=2324): H-W in 5 microsatellites

3. Cluster 3 (n=921): H-W in 11 microsatellites

- Independent individuals according to the aggregating threshold (n=7363): no H-W.

f_{ij}	Cluster 1	Cluster 2	Cluster 3	Group of indep.
Cluster 1	0,3301			
Cluster 2	0,3157	0,3337		
Cluster 3	0,3156	0,3175	0,3366	
Group of indep.	0,2746	0,2741	0,2723	0,246

Conclusions

- Very preliminary results.

- It seems that the proposed method is a good way to infer population structure in a population of related individuals.

- This method determines the population structure in a fast way.

- Up to now results suggest that ANI breed population may be stratified in three subpopulations, leaving a set of individuals that cannot be assigned to any of the subpopulations.

Acknowledgements