Mapping genetic diversity in European gene banks: preliminary results on chickens for the validation of IMAGE001 array

Michèle Tixier-Boichard, Nathalie N. Peynot, Delphine Duclos, Steffen Weigend, L. Monteagudo, A. Martínez, Juan V. Delgado Bermejo, Rayner González-Prendes, Richard P. M. A. Crooijmans, Gwendal Restoux

Mapping genetic diversity in European gene banks: preliminary results on chickens for the validation of IMAGE001 array

M. Tixier-Boichard[1*], N. Peynot [1], D. Duclos[2], S. Weigend[3], L. Monteagudo[4], A. Martinez Martinez[5], J.V. Delgado[5], R. Gonzalez Prendes[6], R. Crooijmans[6], and G. Restoux[1]

[1] Université Paris-Saclay, INRAE, AgroParisTech, GABI, 78352 Jouy-en-Josas Cedex, France; [2]Institut de l'Elevage, 75595 Paris Cedex 12, France ; [3]Friedrich-Loeffler-Institut, Institute of Farm Animal Genetics, 31535 Neustadt, Germany ; [4] University of Zaragoza, Facultad de Veterinaria, 50013 Zaragoza, Spain ; [5]University of Cordoba, Animal Breeding Consulting, S.L, 14071 Cordoba, Spain ; [6] Wageningen University and Research, Animal Breeding and Genomics, 6708 PB Wageningen, The Netherlands. [*]michele.tixier-boichard@inrae.fr <

Abstract

Gene banks are a component of a national strategy for the preservation of genetic diversity. Gene bank managers need to have a global and comparable picture of the diversity in their collections in order to rationalize them. Facing a diversity of molecular tools is a difficulty. The IMAGE H2020 project aimed at developing a low cost 60k SNP array to facilitate the mapping of diversity in gene banks. The first test of this array was performed for chicken with 204 samples from 18 local breeds and nine experimental lines provided by Germany, France and Spain. The MAF across population was 0.34, showing that this tool is useful over a range of populations. The principal component analysis and the Neighbor-joining tree showed that local breeds did not cluster according to country and were generally homogenous. Comparison with on-farm populations remains to be done to assess the value of the gene bank collections.

Introduction

Genetic diversity is essential for the sustainability of animal breeding. Countries are investing in long term conservation of genetic diversity with the creation of gene banks. FAO promotes this strategy in the Global Plan of Action for Animal Genetic Resources. In Europe, the European Regional Focal Point is gathering national coordinators of animal genetic resources from all countries and includes a permanent working group on ex situ cryoconservation.

Until now, sampling strategies remain national and may vary according to countries, resulting in a variety of situations in terms of species and breeds represented. The ERFP is currently promoting the set up of a European network of Animal Gene Banks (EUGENA) to increase collaboration between national gene banks in Europe. An economic analysis of gene banking was recently performed in the frame of the IMAGE H2020 project (Grant # 677353). Results showed that the cost of preservation of a given number of breeds could be decreased by 20% by a better coordination across European countries (de Oliveira Silva et al., 2018).

Assessing the value of gene bank collections for breeding programmes require data on within-breed diversity as well as between-breeds, since both levels are important for the management of farmed animal species. Whereas pedigree data have been used since a long time to characterize within-breed diversity, for *in situ* populations as well as *ex situ* collections, only molecular data make it possible to map the diversity at all levels within a species.

Mapping diversity has made significant progress thanks to the development of genotyping arrays in all farmed species, whereas whole genome sequencing can provide a deeper analysis of the status of collections. However, sequencing costs are still relatively high for gene bank

managers, who generally lack the capacity to analyse such data. Thus, genotyping arrays are still a relevant tool to provide a global description of the diversity present in gene bank collections and to make comparisons between gene banks as well as between gene bank collections and on farm populations. However, several SNP arrays for chickens coexist, either private (around 60K) or public (around 580K), at a rather high cost (from 90 to 150 € per animal), with a low % of markers shared between the arrays. This is complicated for gene bank managers who need a global and comparable picture of the diversity in their collections. To facilitate a common approach across gene banks, the IMAGE H2020 project has developed a multi-species 60k array, with the aim to provide 10k markers for each of six species, i.e. cattle, sheep, goat, horse, pig and chicken, at a price of 20 US$ per sample maximum for both the array and the genotyping, thus excluding the cost of DNA extraction.
A subset of gene bank collections from different countries was genotyped to validate the IMAGE001 array. In this paper, we show how the first version of IMAGE001 can map diversity of chicken collections from gene banks in Germany, France and Spain.

Table 1- List of chicken populations included into the genotyping test for IMAGE001 array

Type of Population	Breeds	Samples	Country
Experimental lines	Congenic line_LB13 (MHC)	2	France
	Congenic line LB21 (MHC)	2	France
	Congenic line LB4 (MHC)	4	France
	Congenic line_LB19 (MHC)	3	France
	DPF-(low duration of fertility line)	4	France
	DPF+ (high duration of fertility line)	5	France
	DWNA (dwarf naked neck layer)	5	France
	Fat line(high abdominal fatness)	5	France
	Lean line Low abdominal fatness)	5	France
Local breeds	High quality Bresse chicken (LB99)	5	France
	Barbezieux	3	France
	Gasconne	3	France
	Gauloise_Doree	12	France
	Blue_Andalusian	10	Spain
	Deutsches Reichshuhn	10	Germany
	Krüper	10	Germany
	Deutsches Lachshuhn	10	Germany
	Langshan	10	Germany
	Ostfriesische Moewen	10	Germany
	Rheinlaender	10	Germany
	Sachsenhuhn	10	Germany
	Sundheimer	10	Germany
	Westfaelische Totleger	10	Germany
	Augsburger	11	Germany
	Bergische Schlotterkamm	11	Germany
	Deutsche Sperber	11	Germany
	Gallina_del_Sobrarbe	20	Spain

Materials & Methods

Gene bank samples.
A total of 211 chicken DNA samples were provided for 27 breeds from the gene banks of three countries (Table 1). Depending on the breed, the DNA had been extracted from frozen semen or from blood. Quality and concentration were determined with spectrophotometry and at least 1.5 µg of DNA was provided to the IMAGE partner at Wageningen University who was coordinating the test. Sampling represented only local breeds for Germany and Spain, with 10 to 20 animals per breed, whereas experimental lines were also provided by France, with a lower number of animals (2-5) per population (Table 1).

SNP selection.
SNP selection was performed based on 1) overlap with existing arrays with a high allele frequency across populations; 2) location in genes affecting phenotypic traits; 3) mtDNA; 4) Ancestral SNPs as compared to wild species of jungle fowl; 5) MHC region; 6) random genes located into QTL regions. A total of 9,306 SNPs were initially selected to be tested for the chicken, covering 32 autosomes (9071 markers) and the sex chromosomes (235 markers).

Genotyping.
Genotyping was performed on an Affymetrix platform by Eurofins® biotech company.

Analysis.
Analysis was performed with *plink* software. The Minor Allele Frequency (MAF) was calculated per chromosome and per breed. Autosomal markers were used to determine the genetic structure with a Principal Component Analysis and a neighbour-joining tree.

Results
Genotyping
Altogether 7743 autosomal markers passed the calling rate and were included in the analysis. Seven samples did not provide satisfactory results, leading to a total of 204 individual genotypes. The mean MAF per chromosome was above 0.30 for most of the autosomes, exceptions being chr 16 (0.03), chr 25 (0.27), chr 27 (0.20) chr 31 (0.02) and chr 33 (0.04). The overall MAF across all populations was 0.34. At population level, the median MAF was above 0.25 except for three congenic lines (LB4, LB13, LB19), which were less variable.

Principal component analysis
The first four principal components (PC) explained only 16% of the total variance. PC1 explained 7%, populations were scattered across a gradient from Langshan breed to congenic lines, Westfaelische Totleger and Ostfriesische_Moewen breeds. PC2 isolated the Gallina_del_Sobrarbe from other breeds, PC3 isolated the Deutsche Lachshuhn and PC4 isolated the Andalusian.

Neighbor-joining tree
The NJ tree (Figure 1) featured 4 main branches: branch A included only one Spanish breed (Blue Andalusian) and one German breed (Augsburger); branch B included three German breeds (Rheinlander, Westfaelische Totleger and Ostfriesische Moewen) and the four congenic lines which are all of White Leghorn origin; branch C included three German breeds (Krüper, Deutsche Sperber and Bergische Schlotterkamm); branch D included the remaining German breeds, most animals of the Gallina del Sobrarbe, the French local breeds and the experimental lines of a non White Leghorn origin. The divergent lines such as Fat/Lean lines and DPF+/DPF- clustered together.

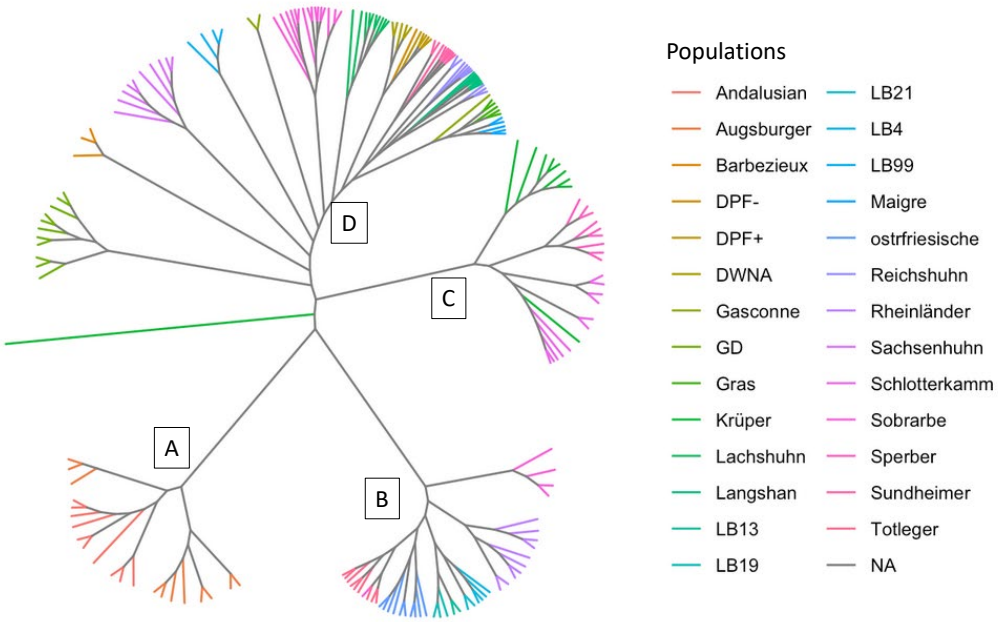

Figure 1. Neighbor-Joining tree of the populations. (A, B, C, D are discussed in the text)

Discussion

First of all, the high MAF value observed across all populations demonstrates the ability of the array to be used over a large range of populations. The total number of markers was quite reduced and another set of markers will be added to reach the target of 10k markers. Yet, a few chromosomes, (16, 31, 33) are likely to remain underrepresented in the array.

The populations did not particularly cluster according to country, which shows the interest of sharing information across gene banks to assess their complementarity. This finding, however, may result from the unbalanced distribution of breeds according to country in the data set. There was only one branch made up of breeds from a single country (Germany) which is consistent with the fact that these breeds are typical Northwestern European breeds and quite different from Mediterranean breeds such as the White Leghorn. Enlarging the sampling to Northern Europe would make possible to check this hypothesis. More information is needed on the history of the two Spanish breeds to understand their separate clustering pattern in the NJ tree. Besides, the history of breeds from different countries which are located within the same branch of the NJ tree also needs to be further explored.

Each population appeared in a single branch of the NJ tree, with a few exceptions, such as the Gallina del Sobrarbe, for which most animals appeared mainly in branch D but a few were in branch B. Local chicken breeds managed by a breeders'club generally appear to be homogenous, unless the breed is distributed across farmers who do not exchange their animals, thus leading to genetic fragmentation.

In conclusion, using a single SNP array for mapping diversity in gene bank collections is feasible and the next step is to include data from on-farm populations in order to map the diversity within and outside gene banks.

References

De Oliveira Silva R., Ahmadi B.V, Hiemstra S.J., and Moran D. (2018) J. Anim. Breed. 136(1) : 63 -73. https://doi.org/10.1111/jbg.12368

Acknowledgments : this work was funded by H2020 IMAGE grant #677353.