
IMAGE

Innovative Management of Animal Genetic Resources

Grant Agreement Number: 677353

Horizon 2020 FRAMEWORK PROGRAMME

TOPIC: MANAGEMENT AND SUSTAINABLE USE OF GENETIC RESOURCES

Topic identifier: SFS-07b-2015

Type of Action: Research and Innovation Action (RIA)

DELIVERABLE D4.5

Deliverable title: A standard multi-species chip for genomic assessment of collections.

Abstract: This deliverable describes the design of multi species SNP arrays based on variation detected within the IMAGE project and from public and partner related information. We made two new arrays: IMAGE001 with 10K SNPs each for cattle, pig, chicken, horse, goat and sheep; IMAGE002 with 10K SNPs each for water buffalo, duck, rabbit, quail, bee and pigeon. Both arrays can capture biodiversity of traditional breeds for each species on the autosomes and sex chromosomes. The arrays also harbour ancestral SNPs, mtDNA SNPs, trait related variation and variation in genes within QTL regions. In addition, for IMAGE001 we included MHC variation for each species. We validated and tested both arrays with 1920 and 1152 DNA samples covering over 300 breeds for IMAGE001 and IMAGE002 respectively. Both tools are recommended for molecular characterisation of stored animal specimens and new entries for genebanks worldwide.

Due date of deliverable: Month 36

Start date of the project: March 1st, 2016

Organisation name of lead contractor: 28-WU,

Contributors: INRA, UCSC, WU, FLI, WR

Dissemination level: PU¹

Initial submission date: Month 48

Duration: 48 months

Revision N°: V1

¹ PU: public

Table of contents

1. Executive Summary	2
2. Array development.....	3
3. Array validation and improvements.....	10
4. Data sharing.....	13
5. Conclusions.....	13
6. Publication bibliography.....	14
7. Table S1	17
8. Table S2	21

1. Executive Summary

Background	<p>Detection of genomic variation by SNP arrays is used in all major livestock species. These arrays are mostly biased to commercial breeds and are subjected to changes in SNP content resulting in little overlap within-species. Some of the arrays are not publicly available. Variation can be best detected by whole genome sequencing but this is still too expensive as a standard procedure in molecular characterization of especially traditional (local) breeds. Standard SNP arrays for each species are needed with more emphasis on traditional breeds and a sufficient overlap with the commercial arrays in order to compare already genotyped populations.</p> <p>Many samples have been stored in gene banks that have not been molecular characterised. It is important to analyse gene bank collections within as well as across countries. A standardized SNP tool can help to open up the national gene banks and help in making decisions what samples to store in the future.</p>
Objectives	<p>To design publicly available, free accessible multi-species single nucleotide polymorphism arrays for the main farm animal species to genotype genetic collections at a low cost (below \$20/sample).</p>
Methods	<p>We use Whole Genome Sequencing (WGS) data described in D4.2 as well as publicly and partner available genomic variation (D4.1). Genotype data provided information for the minor allele frequencies (MAF) of populations. SNP selection for each species was performed based on 1) overlap with existing arrays with a high allele frequency across populations 2) SNPs in genes affecting phenotypic traits; 3) SNPs in the mitochondrial DNA; 4) Ancestral SNPs; 5) SNPs in the MHC region; 6) SNPs in random genes located into QTL regions.</p>
Results & implications	<p>We developed two arrays. IMAGE001 contains 10K SNPs per species for cattle, pig, chicken, horse, goat and sheep. It is publicly available through Affymetrix at a cost of \$19.50 including genotyping at globally active providers (Eurofins or Geneseek). IMAGE002 is currently in production and includes 10K SNPs each from 6 different species: water buffalo, duck, quail, rabbit, bee and pigeon. The IMAGE001 array is validated with 1920 samples and the IMAGE002 array will be validated with 1152 samples, for a total of 300 breeds. Both arrays can be used to measure biodiversity of samples of gene bank collections and <i>in situ</i> populations and to compare the results with existing genotype data sets. They can also be used for sex identification, inheritance checking, mtDNA haplotyping and give insight in the frequency of affected alleles (further described in Task T4.4 and deliverable D4.4). The tools to analyse the genotypes obtained have been developed in task D4.6 and delivered in D5.4.</p>

2. Array development

Introduction

Monitoring genetic variation of animal collections stored in genebanks and comparing them to datasets already available is important to make choices about additional samples that may need to be stored for the future, as well as to assess the value of old collections. This knowledge is not only important within a country but it is also important to compare the variation within the same breed in another country. The challenge is to monitor the genetic variation across all species maintained in a gene bank. For the main farm animal species genotype tools are already available. The problem however is that there are two different main platforms (Illumina and Affymetrix) and that many different arrays are available per species, with variable degrees of overlap between the SNPs depending on the species, as described in D4.1 for cattle, sheep, goat, pig and chicken, and in the literature (Ramos et al, 2009; BovineSNP50 BeadChip, Illumina BovineHD BeadChip, BovineLDv2.0 (Illumina Inc., San Diego, CA); Groenen et al. 2011; Kranis et al. 2013; Schaefer et al. 2017; McCue et al 2012; Colli et al. 2018; Kijas et al. 2012). Moreover, some of the arrays are not freely accessible because they are owned by private companies, like for example in chicken, or are not accessible at all. Most of the SNP arrays contain a large number of SNPs which are used for QTL and GWAS studies. For many other species, no arrays are available and genotyping is performed using genotyping by sequencing (GBS). Thus, it is currently very difficult for gene bank managers, who may not be specialist in genotyping or bioinformatics, to monitor the diversity status of their collections.

To have a complete picture of diversity we need information of the autosomes, the sex chromosomes and the mtDNA, due to different inheritance patterns of these genetic components. In many cases the current SNP assays are biased towards the autosomes and biased to commercial populations (Ramos et al, 2009; BovineSNP50 BeadChip, Illumina BovineHD BeadChip, BovineLDv2.0 (Illumina Inc., San Diego, CA); Groenen et al. 2011; Kranis et al. 2013; Schaefer et al. 2017; McCue et al 2012; Colli et al. 2018; Kijas et al. 2012). Therefore, there is a need for a molecular tool across species covering variation in the three genetic components and more focused on traditional breeds but with sufficient overlap to current arrays in order to enable a comparison with commercial populations.

As described in deliverable D4.2 of task T4.2, we sequenced many genomes of traditional breeds to obtain variation. We combined this with publicly, or IMAGE partner, available data (SNPs as well as population/breed allele frequencies of these SNPs), as described in D4.1 for five of the six species. Scientific referents were identified for each species in order to centralize information on SNPs, which was then used to select 10K SNPs per species to be combined in a new array for six species.

Methods

Species selection:

The species were chosen based on the survey performed in 2017 within WP2 of the IMAGE project which collected information on the species stored in the genebanks. Furthermore, the FAO partner launched a worldwide survey in January 2019 to enquire about the interest of gene banks outside Europe for a standard tool for the genetic characterization of their collections. It was important to identify the potential number of users of a multi-species array in order to convince the provider of the array that it was worthwhile the development. Indeed, cumulating several species on an array is aimed at increasing the number of users which is a necessary condition to get a low price for the array. Since gene banks generally store material from several species, they will particularly benefit from a multi-species array.

Consequently, the first array named IMAGE001 is based on the species for which most samples are collected in the genebanks worldwide: cattle, pig, chicken, horse, goat and sheep. The second array named IMAGE002 is based on species which are less widely present in genebanks but are currently relevant for research or likely to become important for livestock production in the near future.

Platform selection

We contacted Illumina and Affymetrix to get their quotation. For the same budget, Illumina was proposing an array for 10 species, 10K each, and Affymetrix was proposing three arrays for 18 species, 10 k each, six species per array. We finally chose Affymetrix provider.

SNP selection:

The choice of the 10K SNPs for each species was based on the following selection procedure:

- 80% of the variation is derived from existing SNP arrays and SNPs with a MAF >0.3 within traditional breeds
- The remaining 20% of the SNPs are selected based on
 - Sex chromosomes X/Y or Z/W (Felkel et al. 2019; Wong et al. 2004; Colli et al, 2018, Pariset et al. 2011)
 - mtDNA variation covering the different haplotypes within the species (Yang, et al. 2017; Pariset et al. 2011; Colli, et al, 2018; Upadhyay et al. 2016; Jansen et al 2002; Petersen et al. 2013).
 - Ancestral SNPs (depending on the species, they were obtained either from ancient DNA, or from ancestral genomes of live wild ancestor species)
 - Trait markers (derived from OMIA; <https://omia.org>)
 - MHC variation (animal db_SNP; <https://www.ncbi.nlm.nih.gov/snp/>; Ali et al 2017; Matukumalli et al. 2009)
 - Variation detected within genes of QTL regions (animal QTL, Hu et al., 2019, www.animalgenome.org)

Selected SNPs fulfil the criteria recommended by Affymetrix with a convert threshold of 0.6.

All SNPs are aligned to the latest reference genome as indicated in Table 1.

Referent species experts were provided by IMAGE partners and other collaborators:

Multi species array IMAG001:

- Cattle: WU, INRAE, Trinity College , Ireland(D. Bradley for ancestral SNPs)
- Pigs: WU
- Chicken: WU, INRAE, FLI

- Horse: WU, INRAE (a CNRS member, L. Orlando for ancestral SNPs), University of Vienna, Austria (Barbara Wallner for Y chromosome markers) and University of Minnesota, USA (Molly McCue for population frequencies.)
- Goat: INRAE
- Sheep: UCSC

Multi species array IMAGE002:

- Water Buffalo: UCSC
- Rabbit: INRAE
- Quail: INRAE
- Duck: WU , University of Texas El Paso, USA (Philip Lavretsky)
- Bee: INRAE
- Pigeon: WU and University of Utah, USA (M. Shapiro)

Validation:

Genomic DNA samples from European gene banks and from Argentina (INTA as an IMAGE partner) were selected based on the inventory made within IMAGE WP2. For each sample 20 µl of 17 ng/µl DNA was used for genotyping on the Affymetrix-platform. For the IMAGE001(v1) in total 1920 samples were selected covering over 260 breeds (Table 1) with on average 320 samples per species. In Case of IMAGE002(v1) in total 1152 samples will be genotyped with on average of 192 samples per species (Table 1). Based on the results of the validation study of the IMAGE001(v2) and IMAGE002(v1) SNPs which are not working will be replaced by new SNPs for each of the species. The replacement SNPs will be derived from the SNP resources we obtained and from new information publicly available.

Table 1. Species and reference genomes of multi species array IMAGE001 and IMAGE002 and the number of samples and breeds used for validation.

Array					
IMAGE001	species	species	reference genome	# of samples	# of breeds
	Cattle	<i>Bos taurus</i>	UM3.1	281	41
	Pig	<i>Sus scrofa</i>	Susscrofa 11.1	165	31
	Chicken	<i>Gallus gallus</i>	GRCg6a	211	28
	Horse	<i>Equus caballus</i>	EquCab3.0	517	56
	Goat	<i>Capra hircus</i>	ARS1	207	31
	Sheep	<i>Ovis aries</i>	Oar_v3.1	525	66
total				1906	0
IMAGE002	Water Buffalo	<i>Bubalus bubalis</i>	Bubbub1.0	192	tbd
	Duck	<i>Anas platyrhynchos</i>	CAU_duck1.0	192	tbd
	Quail	<i>Coturnix japonica</i>	Coturnix japonica 2.1	192	tbd
	Rabbit	<i>Oryctolagus cuniculus</i>	OryCun2.0	192	tbd
	Bee	<i>Apis mellifera</i>	Amel_HAv3.1	192	tbd
	Pigeon	<i>Columba livia</i>	Cliv_2.1	192	tbd
total				1152	

For every species we collected publicly available genotype data which will be filtered for the SNPs per species which are placed on the IMAGE001 or IMAGE002 arrays. The genotypes of these SNPs are used as a species base population for new entry comparison. We used a principal component analysis in PLINK (Purcell et al. 2007) to show the genetic relationship between samples. This is presented and discussed in detail in D5.4.

For both validation experiments the samples selected are covering as many breeds as possible with the number of samples per breed varying from 1 to 15. Included in the validation are predominantly individual samples but we also included trio's to study inheritance, duplicate samples on different plates to check reproducibility, and some pooled samples derived from evolutionary related species.

Results

Array development: IMAGE001(v1)

This first array design covers the 6 major farm animal species, cattle, pig, chicken, horse, goat and sheep, with on average 9,952 SNP per species. The number of SNPs per species varied from 9,306 in chicken to 10,114 in horse (figure 1). References of the different species on reference genome used and public available genotype data is given in Table 2a. Detailed information on the number of SNPs submitted for assay calling and markers which were finally selected to be placed on the array is given in supplementary Table S2.

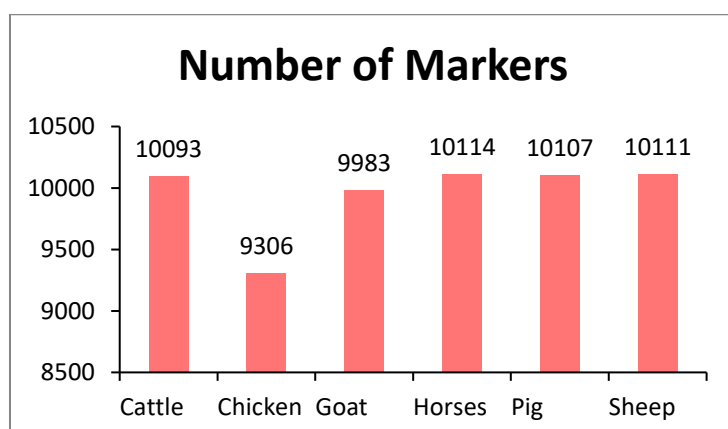


Figure 1. The number of markers per species selected for IMAGE001(v1)

We used SNPs from different categories as described in the M&M section of this deliverable. For this array the majority of the SNP markers (average of 8000 SNPs) are derived from existing commercial arrays in order to be able to compare the results with existing SNP genotyped populations, traditional as well as commercial. The markers covering the other categories are markers derived from the sex chromosomes (to study paternal diversity) and mtDNA (to study maternal diversity).

Table 2a. Overview per species with references of array design and use of the arrays in genotyping population per species used in this study for the IMAGE001 array.

	array	references	references
		design array	use arrays within populations
species			
Cattle	7K; 43K, 777K	BovineSNP50 BeadChip, Illumina BovineHD BeadChip , BovineLDv2.0 (Illumina Inc., San Diego, CA)	Upadhyay et al. 2017
Pig	60K	Ramos et al, 2009	Yang et al. 2017
Chicken	60K; 660K	Groenen et al. 2011; Kranis et al. 2013	Bortoluzzi et al. 2018
Horse	60K; 660K	Schaefer et al. 2017; McCue et al 2012.	Schurink et al. 2018
Goat	52k	Colli et al. 2018	Martin et al. 2018
Sheep	10k,52k	Kijas et al. 2012.	Rochus et al. 2018

Furthermore we selected additional ancestral SNPs and SNPs related to traits, the Major Histocompatibility Complex involved in immune response, and genes in some known QTL and/or selective sweeps regions. Detailed informations of the number of SNP for the different species and different categories are given in Table 3a for the IMAGE001 array

Table 3a. Overview of selected SNPs per species for the IMAGE001(v1) array

species	cattle	pig	chicken	horse	goat	sheep
SNPs						
overlap existing arrays	7817	6173	7366	7748	7979	9583
sex chromosome X/Z	240	539	135	368	12	134
sex chromosome Y/W	5	26	100	80	69	2
mtDNA	13	36	90	0	170	136
ancestral	974	2000	1543	322	1043	256
trait	73	107	20	50	1164	80
MHC	134	9	63	203	0	0
genes in QTL-regions	1723	1289	0	1537	60	800
total	10093	10107	9306	10114	9993	10111

Additional new markers were included for most of the species. For the sex chromosomes especially the Y chromosome, new publicly available markers were added because of improved annotation of the sex chromosomes. The mtDNA markers used on the array are representative for the different haplotypes present in each species. For the ancestral SNPs different approaches were used depending on the species. For cattle and horse, ancient DNA information was used to select variation which was present in the ancestors and no longer seen in the present commercial breeds, In species where the wild relatives are still present (pig,

chicken, goat and sheep) we selected variation in the same way as described above. For the trait markers we looked at the OMIA website and selected SNPs reported to be associated or causal for the traits reported per species. The markers derived from genes in QTL regions did vary per species and are listed together with functional markers in D4.4 as part of task T4.3.

The coverage on the genome for the 6 species represented by the array IMAGE001(v1) is presented in Figure 2. The coverage of the Y chromosome is not always indicated because of insufficient information of the SNPs for this chromosome.

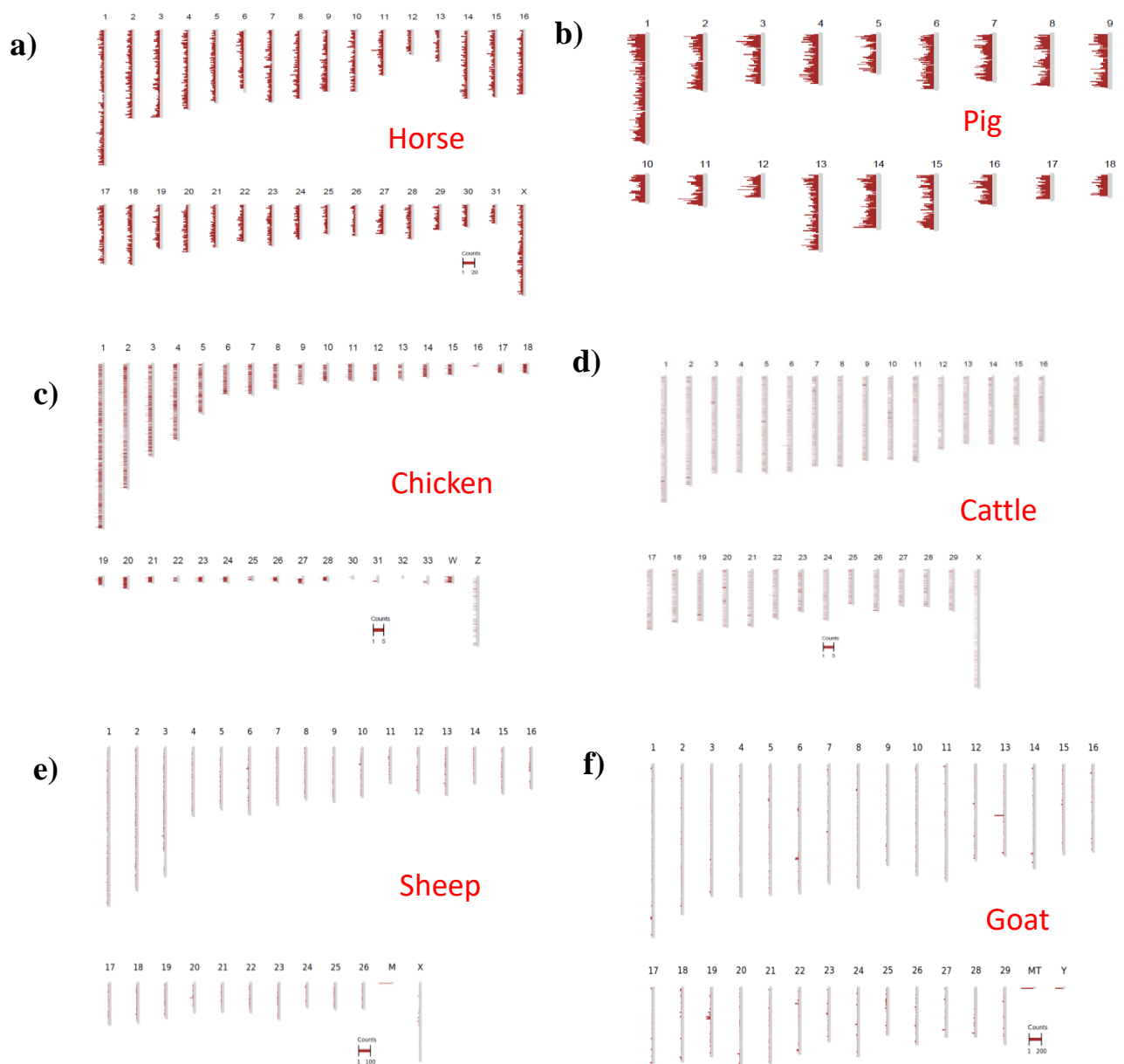


Figure 2. The coverage on the SNPs from IMAGE001(v1) on the genome of the six species: a) horse, b) Pig, c) chicken, d) cattle and e) sheep and f) goat.

Array development: IMAGE002

This second design covers again 6 animal species, water buffalo, duck, quail, rabbit, Bee and pigeon, with an average of 10000 SNPs per species (Table 1). For this array we could only use publicly available SNP data and genotypes for water buffalo and rabbit. For the other 4 species WGS data was used to select high informative SNPs across populations. Selection of the SNPs for this array was performed as much as possible as described in the M&M based on available data. Table 2b shows an overview of the data used to develop the IMAGE002 array. For the species represented on this array the reference genome is less well developed as compared to the species used for the IMAGE001 array. For all species a draft genome is available although some are still arranged in scaffolds instead of chromosomes. Specific trait markers were incorporated if known. MtDNA variation is hardly used in biodiversity studies of the IMAGE002 species. Sex differentiation in bees is different from the other species. Too many variants on the bee genome are involved in sex discrimination. Information on ancestral variation is in many cases not known and therefore this type of variation could not be included in the SNP selection of this array. Details of the SNP selection is given in Table 3b.

Table 2b. Overview per species with references of array design and use of the arrays in genotyping population per species used in this study for the IMAGE002 array.

	Array/genome	references
		design array
species		
Water Buffalo	90k	Iamartino et al. 2017
Duck	600k	Teissier et al. 2019
Quail	Coturnix japonica 2.1	Kawahara-Miki et al 2013
Bee	Amel_HAv3.1	Solignac et al. 2007
Rabbit	200k	Affymetrix Axiom OrcunSNP Array
Pigeon	Cliv_2.1	Shapiro et al. 2013

Table 3b. Overview of selected SNPs per species for the IMAGE002 array.

species	Water buffalo	Duck	Quail	Bee	Rabbit	Pigeon
SNPs						
overlap existing arrays	7991	0	0	0	7897	0
New selection	0	7900	7901	7901	0	7901
sex chromosome X/Z	131		474		296	
sex chromosome Y/W	17	18	2		7	36
mtDNA	198	7	4	7	11	
ancestral	573	1426	335			
trait	201	251	1751			202
MHC	0					
genes in QTL-regions	225		166			1105
total	7901	7900	7901	7901	7897	7901

3. Array validation and improvements

IMAGE001

A set of 1906 individual samples and 14 pooled samples was collected to validate IMAGE001(v1). In total, animals were selected from 246 breeds with an average of 41 breeds per species. The number of breeds varied from 28 breeds for chickens to 66 breeds for sheep. The average number of animals per breed was 8 and this varied from 1 for the Angus breed in Cattle to 37 for the Linca breed in Sheep. All 1920 samples will be used in the analysis and incorporated in the final manuscript which is currently in preparation. The list of breeds per species for the validation is indicated in the appendix table S1.

The number of markers included in the array was 65K of which 61K did perform according to the Affymetrix criteria and 4K markers did not. We obtained genotypes for 1526 samples out of 1536 tested, with a call rate above 95%. Supplementary table S2 (appendix) provides a detailed overview of the submitted SNPs, SNPs placed on the array, successful or failed genotyped markers, per category and per species. Following this analysis, we are going to replace the markers which did fail per species by new markers, including a short list of additional functional markers established for sheep, pig and chicken. This is to increase the number of specific traits that can be genotyped with IMAGE001(v2). The information on the selected SNPs for the improved IMAGE001(v2) is given in Table 4 and is currently in the manufacturing phase to be the final product obtained in IMAGE.

Table 4. Overview of selected SNPs per species for the IMAGE001 (v2) array

species	cattle	pig	chicken	horse	Goat	sheep
SNPs						
overlap existing arrays	7817	6173	7366	7748	7979	9583
sex chromosome X/Z	240	539	635	368	200	134
sex chromosome Y/W	50	26	100	80	69	50
mtDNA	13	36	90	0	170	136
ancestral	974	2000	2361	322	1043	256
trait	73	107	32	50	1164	80
MHC	134	9	63	203	0	0
genes in QTL-regions	1723	1289	0	1537	60	800
total	11024	10179	10647	10308	10685	11039

An example of a PCA plot for the porcine samples genotyped with the IMAGE001(v1) array is given in Figure 3a and 3b for the base population and the new IMAGE001 genotyped porcine samples respectively.

Figure 3a. PCA plot with eigenvectors 1 and 2 for publicly available samples as base. Samples in cluster 1 are European breeds. Cluster 2 are European wild boar samples. 3. Are Asian breeds whereas 4 represent samples from Asian wild boars.

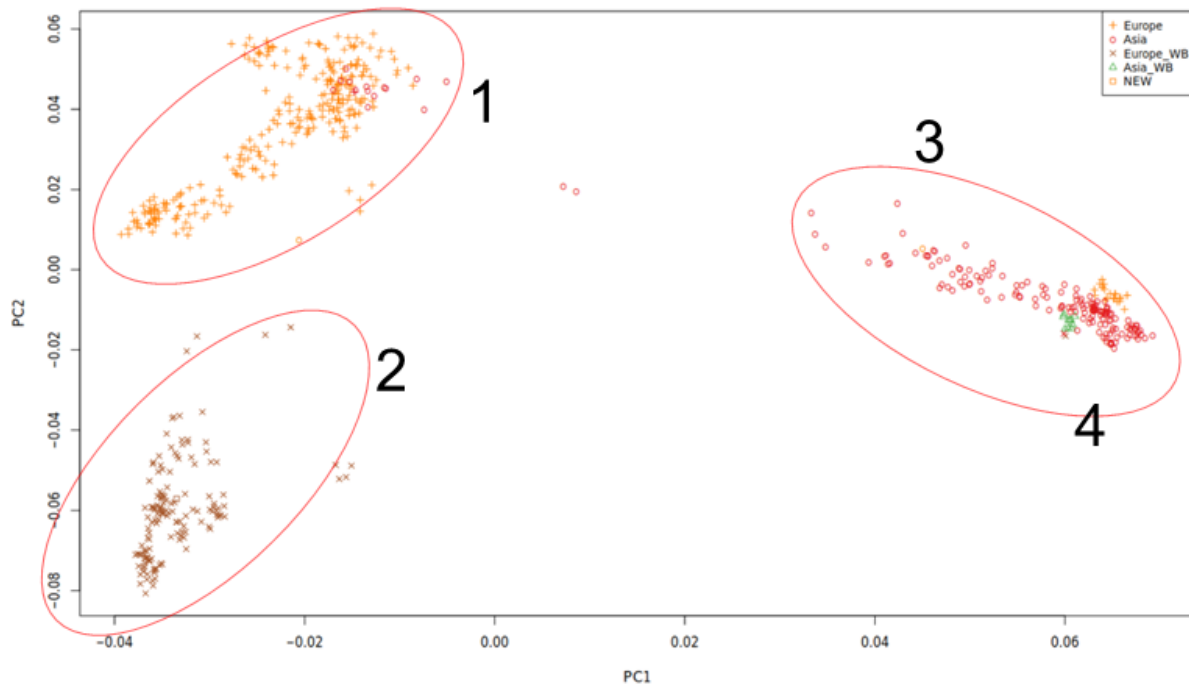
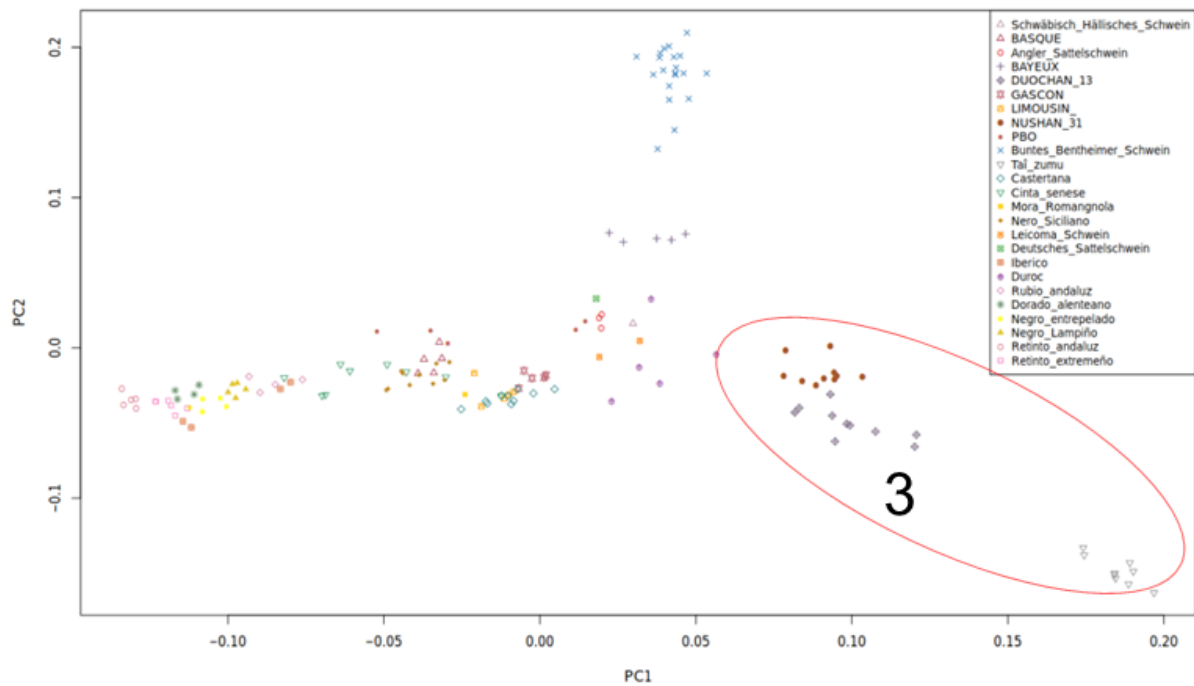


Figure 3b represents the new porcine samples genotyped with the IMAGE001 (v1) array within the samples of cluster 3 having an Asian background, where the other samples of breeds are coming from all across Europe.



The mtDNA markers can be used to reconstruct the haplotypes to be used in phylogenetic studies, as illustrated in Figure 4.

Figure 4. The haplotype tree of the porcine samples genotyped with the IMAGE001(v1) array

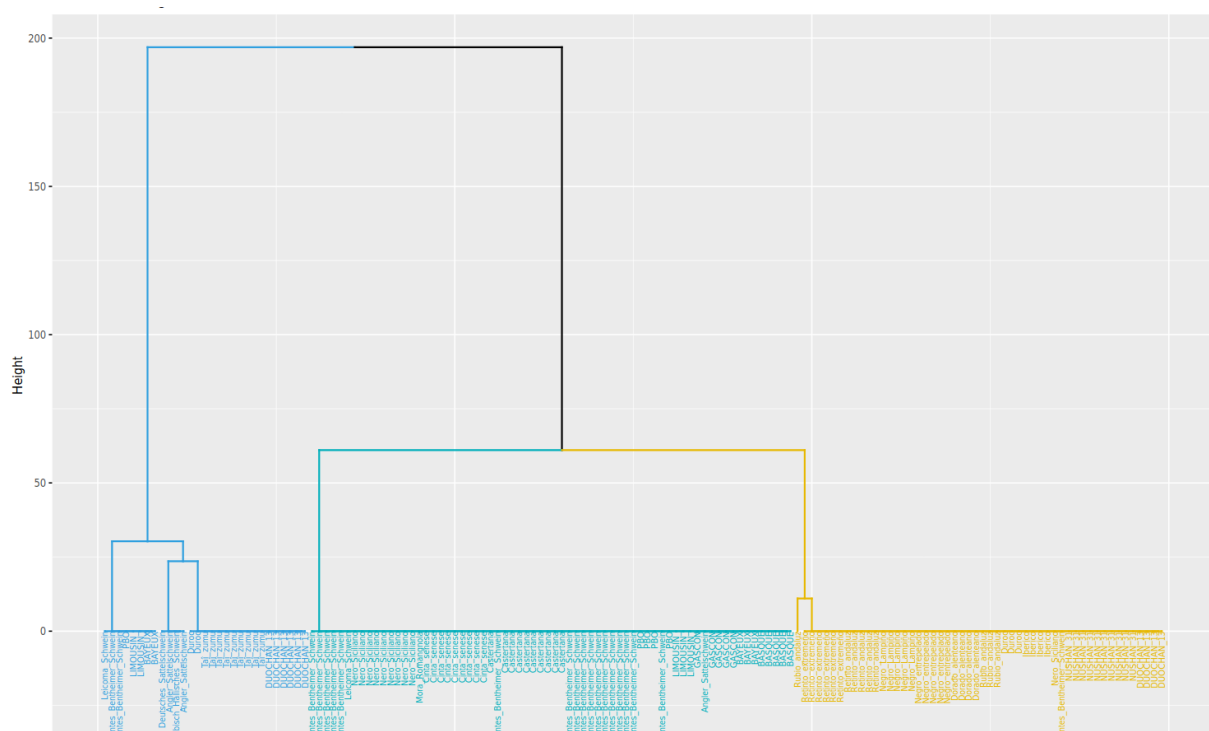


IMAGE002

Validation of the second multi species array will take place by genotyping 1152 samples with an average of 192 samples per species. The array is in manufacturing stage and the samples of breeds and populations are collected in order to represent a wide diversity for each species for genotyping.

4. Data sharing

SNP array and marker information have been provided to Affymetrix (Thermo Fisher Scientific) and will be publicly available. The arrays can be purchased directly through Affymetrix or through globally active genotype providers. The information on the SNPs and selection criteria used will be presented on the IMAGE website at time of publication:

- paper with all data on IMAGE001 is in preparation to be submitted by January 2021
- paper with all data on IMAGE002 will be submitted by April 2021

All genotype data will be uploaded to EBI-EVA following uploading of sample information on BioSamples database.

Discussion will be initiated with the fish scientific community and the AquaExcel infrastructure in order for them to benefit from the agreement with Affymetrix to develop a 3rd IMAGE array.

5. Conclusions

The aim of task T4.5 was to develop a multi species array that is accessible worldwide, without restriction for a low cost. Furthermore, the array should be stable and not subjected to frequent changes. Besides making one array we did have the option to make three arrays. The first array IMAGE001 is available for the community with one modification to come where we will replace SNPs that did not work, by new markers. The second array is in production and will be tested with a broad range of samples. Also, for this array we will have the option to replace markers that do not work. We will design a third multi species array IMAGE003 which will be ready after finishing the IMAGE project which will contain SNPs of only fish species. Specific fish species increasingly represent a very important group of species for animal protein in many countries of the world. A discussion is going on within national genebanks about storing samples of fish species in their genebanks. The developed tools within this deliverable are developed for genebanks but also for animal genetics research institutes in order to genotype their animal collections and to compare genotypes within and over breeds within and over genebanks but also with public available data. For the future it is important that when animal material from a certain species will be stored in a genebank, it is essential to isolate DNA preferably from a blood sample and use this for genotyping to enable a molecular genetic comparison of the new entry with existing data sets.

6. Publication bibliography

1. Affymetrix Axiom OrcunSNP Array
2. Ali AO, Stear A, Fairlie-Clarke K, et al. The genetic architecture of the MHC class II region in British Texel sheep. 2017. *Immunogenetics*. 69:157–163.
3. Bortoluzzi, C., Crooijmans, R.P.M.A., Bosse, M. et al. 2018. The effects of recent changes in breeding preferences on maintaining traditional Dutch chicken genomic diversity. *Heredity* 121: 564–578 .
4. BovineLDv2.0 Microarray (Illumina Inc., San Diego, CA)
5. BovineSNP50 BeadChip (Illumina Inc., San Diego, CA)
6. Bradley DG, MacHugh DE, Cunningham P, Loftus RT. 1996. Mitochondrial diversity and the origins of African and European cattle. *Proc Natl Acad Sci U S A* 93:5131–5135. doi:10.1073/pnas.93.10.5131
7. Bradley Holmes J, Eric Moyer, Lon Phan, Donna Maglott, Brandi Kattman. 2019. SPDI: data model for variants and applications at NCBI, *Bioinformatics*, btz856, <https://doi.org/10.1093/bioinformatics/btz856>
8. Colli L, Milanesi M, Talenti A, Bertolini F, Chen M, Crisà A, et al. 2018. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet Sel Evol*. 50:58
9. Dong Y, Xie M, Jiang Y, Xiao N, Du X, et al. 2013. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol*. 31: 135–41
10. Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, et al. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324: 522–28
11. Felkel, S., Vogl, C., Rigler, D. et al. 2019. The horse Y chromosome as an informative marker for tracing sire lines. *Sci Rep* 9: 6095. <https://doi.org/10.1038/s41598-019-42640-w>
12. Groenen M. A., Megens H.-J., Zare Y., Warren W. C. et al. 2011. The development and characterization of a 60 K SNP chip for chicken. *BMC Genomics* 12:274.
13. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491: 393–98
14. Hu Zhi-Liang, Carissa A. Park, and James M. Reecy. 2019. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Research*, 47:D1.
15. Huang X, Yang Q, Yuan J, et al. Effect of Genetic Diversity in Swine Leukocyte Antigen-DRA Gene on Piglet Diarrhea. 2017. *Genes*. 7:36
16. Iamartino D, Nicolazzi EL, Van Tassell CP, Reecy JM, Fritz-Waters ER, Koltes JE, et al. 2017. Design and validation of a 90K SNP genotyping assay for the water buffalo (*Bubalus bubalis*). *PLoS ONE* 12: e0185220.
17. Illumina BovineHD BeadChip (Illumina Inc., San Diego, CA)
18. Int. Chick. Genome Seq. Consort. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432: 695–716.
19. International Chicken Polymorphism Consortium 2004. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432: 717–722.
20. Jansen T, Forster P, Levine MA, et al. 2002. Mitochondrial DNA and the origins of the domestic horse. *Proc Natl Acad Sci U S A*. 99(16):10905–10910.
21. Jiang Y, Xie M, Chen W, Talbot R, Maddox JF, et al. 2014. The sheep genome illuminates biology of the rumen and lipid metabolism. *Science* 344: 1168–73

22. Kawahara-Miki R, Sano S, Nunome M, Shimmura T, Kuwayama T, Takahashi S. et al 2013. Next-generation sequencing reveals genomic features in the Japanese quail. *Genomics* 101:345–353.
23. Kijas JW, Lenstra JA, Hayes B, Boitard S, Porto Neto LR, San Cristobal M, et al. 2012. Genome-Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PLoS Biol* 10: e1001258.
24. Kranis, A., Gheyas, A.A., Boschiero, C. et al. 2013. Development of a high density 600K SNP genotyping array for chicken. *BMC Genomics* 14, 59
25. Martin P, Palihière I, Maroteau C, et al. 2018. A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. *Sci Rep* 8:4060.
26. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE* 4(4): e5350
27. McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, et al. 2012. A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. Georges M, editor. *PLoS Genet*. Public Library of Science;8:e1002451
28. Miller, M. M., & Taylor, R. L., Jr 2016. Brief review of the chicken Major Histocompatibility Complex: the genes, their distribution on chromosome 16, and their contributions to disease resistance. *Poultry science*, 95:375–392.
29. Online Mendelian Inheritance in Animals, OMIA. Sydney School of Veterinary Science, {date of download}. World Wide Web URL: <https://omia.org/>
30. Pariset L, Mariotti M, Gargani M, et al. 2011. Genetic diversity of sheep breeds from Albania, Greece, and Italy assessed by mitochondrial DNA and nuclear polymorphisms (SNPs). *ScientificWorldJournal*. 11:1641–1659.
31. Petersen JL, Mickelson JR, Cothran EG, Andersson LS, Axelsson J, Bailey E, et al. 2013. Genetic Diversity in the Modern Horse Illustrated from Genome-Wide SNP Data. *PLoS ONE* 8(1): e54997.
32. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D. et al. 2007. PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81.
33. Ramos AM, Crooijmans RP, Affara NA, Amaral AJ, Archibald AL, Beever JE et al. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* 4: e6524.
34. Rochus CM, Tortereau F, Plisson-Petit F, Restoux G, Moreno-Romieux C, et al. 2018. Revealing the selection history of adaptive loci using genome-wide scans for selection: An example from domestic sheep. *BMC Genomics* 19: 1-17
35. Schaefer RJ, Schubert M, Bailey E, et al. 2017. Developing a 670k genotyping array to tag ~2M SNPs across 24 horse breeds. *BMC Genomics* 18:565.
36. Shapiro M.D., Zev Kronenberg, Cai Li, Eric T. et al 2013. Genomic Diversity and Evolution of the Head Crest in the Rock Pigeon. *Science*. 339:1063-1067
37. Schurink, A., da Silva, V.H., Velie, B.D. et al. 2018. Copy number variations in Friesian horses and genetic risk factors for insect bite hypersensitivity. *BMC Genet* 19: 49
38. Solignac M, Zhang L, Mougél F, et al. 2007. The genome of *Apis mellifera*: dialog between linkage mapping and sequence assembly. *Genome Biol*. 8:403.
39. Stella, A., Nicolazzi, E.L., Van Tassell, C.P. et al. 2018. AdaptMap: exploring goat diversity and adaptation. *Genet Sel Evol* 50: 61

40. Teissier M, Noémie Thébault, Juliette Riquet, Christian Diot, Sophie Brard-Fudulea, et al. 2019. Development and validation of high-density SNP array in ducks. XIth European symposium on poultry genetics (ESPG). Prague, Czech Republic. hal-02460148
41. Upadhyay, M., Chen, W., Lenstra, J. et al. 2016. Genetic origin, admixture and population history of aurochs (*Bos primigenius*) and primitive European cattle. *Heredity* 118:169–176 .
42. Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326: 865–67.
43. Wong Ka-Shu, G., Liu, B., Wang, J. et al. A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432, 717–722 (2004).
<https://doi.org/10.1038/nature03156>
44. Yang, B., Cui, L., Perez-Enciso, M. et al. 2017. Genome-wide SNP data unveils the globalization of domesticated pigs. *Genet Sel Evol* 49: 71

7 Table S1

List of breeds per species

Species	Breeds	Samples
Species	Breeds	527
Sheep (67 breeds)	Alpines_Steinschaf	2
	Brianzola	2
	Carinthian	2
	Churra_do_Hinho	2
	Coburg_Foxes	2
	Merino_long_wool_sheep	2
	Merino_Preto	2
	Montafoner_Steinschaf	2
	Racka	2
	Rhoen_sheep	2
	Rouge_Roussillon	2
	Waldschaf	2
	White_headed_mutton_breed	2
	Bovec	3
	Brown_Mountain	3
	Campaniça	3
	Causses_du_Lot	3
	Corriedale	3
	Gentile_di_Puglia	3
	Lacaune	3
	Moutons_Charollais	3
	Rouge_de_louest	3
	Roussin	3
	Tiroler_Steinschaf	3
	Churra_da_Terra_Quente	4
	Comisana	4
	Exmoor_Horn	4
	Kihnu	4
	Massese	4
	Nera_Di_Arbus	4
	Sarda	4
	Scottish_Blackface	4
	Grazalemena	5
	Leine_sheep	5
	Lojena	5
	Merino_negro	5
	North-Holland_sheep	5
	East_frisian_milk_sheep	7
	Veluws_heidenschaap	7
	Blue_Texel	8

	White_horned_heathland_sheep	8
	White_hornless_heathland_scheep	9
	Chamarita	10
	Churra_Tensina	10
	Flevolander	10
	Manchega	10
	Merino_de_los_Montes_Universales	10
	Merino_Negro	10
	Ojalada	10
	Ojinegra	10
	Rasa_Aragonesa	10
	Segurena	10
	Xisqueta	10
	Bentheimer_Sheep	11
	Merino	11
	Friesian/Dutch_Milksheep	19
	Black_blazed_sheep	20
	Drenthe_Heathsheep	20
	Kempen_Heathsheep	20
	Mergelland_sheep	20
	Schoonebeek_Heathsheep	20
	Veluwe_Heathsheep	20
	Castellana	21
	Texel	21
	Churra	22
	Linca	37
Species	Breeds	Samples
Cattle (43 breeds)	Angus	1
	Berrenda_Colorado_BC4	1
	Berrenda_Colorado_BC6	1
	Berrenda_Negro_BN8	1
	Groninger White headed	1
	Deep red	1
	Charolais	1
	Dutch-Friesian	1
	Glanvieh	1
	Holstein-Friesian	1
	Dutch belted	1
	Marchigiana	1
	Marismena	1
	MRY	2
	Rubia_Gallega	1
	Gelbvieh	2
	Angler_Rind	3
	Glanrind	3
	Limousin	3
	Burlina	4

	Berrendo_colorado	5
	Berrendo_en_negro	5
	Cardena	5
	Pajuna	5
	Retinta	5
	Rubia_andaluza	5
	Toro_de_Lidia	5
	Varzese	5
	A_Angus_(traditional_biotype)	7
	Nelore	9
	DSN	11
	A_Angus_(traditional_biotype)	15
	Asturiana_de_la_Montana	15
	Asturiana_de_los_Valles	15
	Guabala	16
	Alistana-Sanabresa	20
	Guaymi	20
	Monchina	20
	Sayaguesa	20
	Creole_(Argentina)	21
	Herford_(traditional_biotype)	21
Species	Breeds	Samples
Goat (32 breeds)	Bunte_Deutsche_Edelziege	1
	Frisa	1
	Mallorquina	1
	Palmera	1
	Weiße_Deutsche_Edelziege	1
	Dutch_Milkgoat	2
	Nicastrese	2
	Orobica	2
	Rove	2
	Poitevine	3
	Saanen_	3
	Tinerfena-Norte	3
	Majorera	4
	Thüringer_Waldziege	4
	Cabra_Blanca_Andaluza	5
	Cabra_Blanca_Celtibérica	5
	Cabra_Negra_Serrana	5
	Payoya	5
	Saanen	5
	Verzaschese	5
	Malaguena	6
	Murciano_granadina	6
	Murciano-Granadina	8
	Florida	9
	Toggenburg	11

	Blanca_Celtib_Ír-ríca	13
	Dutch_Pied_goat	17
	Dutch_White_goat	18
	Dutch_Landrace_goat	19
	Multicerate_Ancestral_flock	20
	Pirenaica	20
Species	Breeds	Samples
Chicken (29 breeds)	Lignee_B13 (MHC)	2
	Lignee_B21 (MHC)	2
	Barbezieux	3
	Gasconne	3
	Lignee_B19 (MHC)	3
	DPF-(low duration of fertility line)	4
	Lignee_B4 (MHC)	4
	DPF+ (high duration of fertility line)	5
	DWNA (dwarf naked neck layer)	5
	Gras (high abdominal fatness)	5
	Lignee_B99 (Bresse breed)	5
	Maigre (Low abdominal fatness)	5
	Blue_Andalusian	10
	Deutsches_Reichshuhn	10
	KrUper	10
	Lachshuhn	10
	Langshan	10
	ostrfriesische_MOwen	10
	RheinlAnder	10
	Sachsenhuhn	10
	Sundheimer	10
	westfAlische_Totleger	10
	Augsburger	11
	Bergische_Schlotterkamm	11
	Deutsche_Sperber	11
	Gauloise_Doree	12
	Gallina_del_Sobrarbe	20

8 Table S2

Table S2 Successful and failed markers per category and per species.

Type	# of SNP Submitted	# of SNPs Removed after QC	# of SNP on array	# of SNPs Failed	# of SNPs working	# of indels genotyped
Horse						
General	10322	1392	8930	172	8758	5
Common	7748	Samples 382	7055	138	6917	1
Genes	1537		1357	0	1357	0
Functional	44		33	6	27	4
X	368		252	1	251	1
Y	100		100	100	0	0
Ancestral	322		46	28	18	0
MCH	203		87	87	0	0
Pig						
General	10179	1987	8192	311	7881	0
Common	6173	Samples 149	6173	170	6003	0
Genes	1289		137	25	112	0
Functional	107		79	11	68	0
MT	36		22	8	14	0
X	539		353	43	310	0
Y	26		4	3	1	0
Ancestral	2000		1422	50	1372	0
MHC	9		2	1	1	0
Chicken						
General	10900	3239	7661	160	7501	2
Common	7366	Samples 199	6284	72	6212	2
Functional	20		3	3	0	0
MT	90		88	0	88	0
W	100		0	0	0	0
Ancestral	2361		948	58	890	0

MHC	63		8	8	0	0
CNV	900		333	20	313	0
Cattle						
General	10975	2371	8604	214	8390	2
Common	7817	Samples 202	6578	137	6441	0
Genes	1723		1700	42	1658	0
functional	73		65	28	37	0
MT	13		8	1	7	0
X	240		184	1	183	0
Y	1		1	1	0	0
Ancestral	974		0	0	0	0
MHC	134		68	4	64	2
Sheep						
General	10111	1476	8635	137	8498	0
Common	8724	Samples 426	7566	56	7510	0
functional	1116		929	22	907	0
MT	136		62	56	6	0
X	134		77	2	75	0
Y	1		1	1	0	0