Cost-efficient assignment panel for ducks. Setup of a cost-efficient assignment panel for duck populations.

Hervé Chapuis, Sophie Brard-Fudulea, Azélie Hazard, Alain Vignal, Julie Demars, Romuald Rouger, Marc Teissier, Hélène Gilbert

HAL Id: hal-04542880
https://hal.inrae.fr/hal-04542880

Preprint submitted on 11 Apr 2024

1 # Cost-efficient assignment panel for ducks

2 Setup of a cost-efficient assignment panel for duck populations. An illustration with
3 experimental data.

4

5 Hervé Chapuis*[1], Sophie Brard-Fudulea[2], Azélie Hazard[3], Alain
6 Vignal[1], Julie Demars[1], Romuald Rouger[2], Marc Teissier [12], Hélène
7 Gilbert[1]

8 [1] UMR GenPhySE, INRAE Toulouse, BP52626, 31326 Castanet Tolosan, France

9 [2] SYSAAF, INRAE Centre Val de Loire, 37480 Nouzilly, France

10 [3] AVIPOLE, INRAE, 40280 Benquet, France https://doi.org/10.17180/R1VP-T666

11

12 *Corresponding author

13 Correspondence: herve.chapuis@inrae.fr

14

15

16 **ABSTRACT**
17 The setup of a flexible and cost-effective 96-SNP assignment panel to be used in Pekin duck
18 (*Anas platyrhynchos*), Muscovy duck *(Cairina moschata)* and their mule duck hybrid, is
19 presented. SNP were selected on the available 600K array in ducks. This SNP array is made of
20 two libraries (one for the Muscovy duck, the other for the common duck which encompasses
21 the Pekin duck), the intersection of which, after a preliminary elimination on the primer
22 length, contained only 399 SNP that were considered a starting point to obtain a final list. A
23 first step was to obtain a list of 192 SNP, based on technical properties, using a reference set
24 of 600K genotypes from commercial lines. In a second step, to obtain the final 96 markers, a
25 subset of the previous reference set was combined with genotypes from 134 Pekin and 128
26 Muscovy, which were the parents of the experimental populations to assign. Assignment
27 rates were 99%, 96% and 88% in the mule, Pekin and Muscovy populations respectively. The
28 lower-than-expected assignment rate in the Muscovy population was due to the absence of
29 16 parental samples. Availability of an effective and affordable assignment panel was
30 deemed necessary after switching from a system where breeders are housed in individual
31 cages to a system where females are housed and inseminated in groups. In the latter case, a
32 factorial mating design replaced the hierarchical design, common in poultry. This new design
33 impacted the population structure, creating more sire x dam combinations, offering
34 possibilities for a better estimation of non-additive genetic effects, which could prove
35 relevant in the foie gras sector. Finally, a list of 135 markers resulted from this study that
36 could be used to build an efficient 96 SNP panel for any local or commercial population.

37

38 *Keywords:* duck, parentage assignment, SNP

39

Introduction

41   In most poultry species, selection is carried out using individual cages in order to easily trace the
42   pedigree of hatched chicks. Equipped with sloped floor allowing eggs to roll to the front of the cage where
43   they are out of the hen's reach and can be collected by the farmer, these cages gained popularity since
44   their introduction in the early twentieth century (Arndt, 1931). Compared with a system where hens lay in
45   a pen equipped with trap-nesting devices, broodiness and floor eggs are eliminated and eggs are cleaner.
46   In addition, more birds can be housed in a given floor space.

47   Yet, in 2021, the European Citizens' Initiative (ECI) "End the Cage Age" called on the European
48   Commission to propose legislation to prohibit the use of cages for a wide range of farm animals. The
49   Commission now assesses the feasibility of working towards the proposed legislation expected in 2027.
50   The poultry breeding companies will then need alternative solution to safely establish pedigree of their
51   stocks. Electronic nests relying on RFID can be used to establish a link between the egg and the layer
52   (Marx et al., 2002) but they remain to be perfectioned in each concerned species to deliver reliable data.
53   In addition, they can only help to build the maternal pedigree. By contrast, the use of molecular markers
54   is susceptible to bring a complete solution to the issue. However, developing such tools for duck
55   populations, that rely on mixtures of purebred and interspecies crossbreds, presents specific challenges,
56   as markers should exist in the two species, and show variability. Indeed, A microsatellite panel had been
57   developed for duck populations in France (Chapuis et al., 2010), and was deemed usable in various
58   purebred and crossbred populations. However, this panel exhibited assignment rates to a unique parental
59   pair too low to be routinely used at a large scale, mainly because markers revealed to be poorly
60   polymorphic within the Pekin and Muscovy populations (Chapuis et al., 2010).

61   Here we present the setting of an efficient and affordable assignment panel that can be used to assign
62   pedigree in populations of Muscovy and Pekin ducks, and also their hybrids traditionally raised for fatty
63   liver production. To build a posteriori the pedigree in these populations, the KASPar technology was
64   retained, as providing access to affordable small SNP arrays. We will present and discuss its performances
65   to assign pedigree in a genetic experimental design. The possible use of the developed molecular tools in
66   other populations, such as local breeds, will also be discussed.

67

Material and methods

Designing the Assignation Panel

Development Strategy of a Cost-Efficient panel

71   Our objective was to assign pedigree in an experimental population of hybrid mule ducks and their
72   purebred half-sibs, namely Muscovy duck (*Cairina moschata*) for the sire line and Pekin duck (*Anas
73   platyrhynchos*) for the dam line. Therefore, we aimed at organizing mating plans and building an
74   affordable 96 SNP panel to retrieve the pedigree using molecular information. The two parental lines
75   pertained to populations sampled to previously develop the ThermoFisher Axiom HD SNP duck array,
76   hereinafter referred to as 600K array (Teissier et al., 2019). This collection of genotypes, already available
77   (hereinafter labelled as "reference dataset"), was used as a starting point to build the desired panel. The
78   600K genotypes from *Anas platyrhynchos* (n=139), *Cairina moschata* (n=79) and some mule ducks (n=45)
79   were used to assess allele frequencies. However, as among these genotypes only 15% originated from the
80   same populations as our parental lines, a two-step strategy was adopted. In a first instance, a set of 192
81   SNPs eligible for the chosen technology was developed, based on both their frequencies in the three
82   populations and their technical properties. This first set was used to obtain first genotypes in our parental
83   lines and in some triplets of mule progeny and their parents, i.e. with known pedigree. In a second step,
84   the 96 SNPs with best technical outcomes and frequencies within and across parental lines were selected
85   among these 192 to obtain an efficient panel. They were later used to establish pedigree of our offspring
86   batches. Note that the mule duck is the hybrid obtained by crossing Muscovy drakes and Pekin females,

87 while the common duck populations (Anas platyrhynchos) represented on the 600K reference dataset
88 encompassed many other breeds than Pekin.

Selection of 192 SNP eligible for KASPar technology

90 The 600K chip contained 334,950 SNPs segregating in the Muscovy duck library and 331,241 SNPs
91 segregating in the common duck library. A preliminary step was to select only markers without
92 polymorphism in the 50 bp before and after the SNP, as primer length is longer (50 bp) with the KASPar
93 technology than with the Axiom technology (35bp). For that purpose, pool-sequenced DNA from 50
94 males, sampled from several French populations (wild mallard and commercial Pekin and Muscovy) were
95 used (Teissier et al., 2019). Primers for markers found on the 600K chip were aligned on the reference
96 genome (*Anas platyrhynchos* genome from Huang et al. 2013, and *Cairina moschata* genome from
97 Thébault et al. 2019). Only SNPs exhibiting an identical primer sequence in the Muscovy and common
98 duck populations were kept. After this step, 229,138 SNP remained in the Muscovy library while the
99 common duck library contained 198,091 markers. The intersection of both led to a list of 399 candidate
100 SNPs, susceptible to be amplified in the mule duck population. Only 396 were awarded the recommended
101 PolyHighResolution status from the Axiom Analysis Suite software distributed by ThermoFisher, meaning
102 they were found high quality and polymorphic. The final list of 192 SNPs was to be built among these 396,
103 applying filters to individuals and triplet genotypes available in the reference dataset. PLINK V2.0 (Purcell
104 et al., 2007) was used to perform filtering operations on missingness, both for genotypes and SNPs, minor
105 allele frequency (MAF), and Mendelian mismatches. The retained criteria were values of 0.95 for call rate
106 (CR) and call frequency, and 0.10 for MAF within Pekin and Muscovy populations. About 100 trios
107 representing various genetic types were available in the reference dataset and could, therefore, be used
108 to track markers leading to Mendelian incompatibilities. Such incompatibilities disqualified the concerned
109 markers. An ultimate filter was applied based on linkage disequilibrium (LD), aiming to choose
110 independent markers.

Setup of the final cost-efficient 96 SNP panel

112 A mixture of two groups of animals was used to evaluate the properties of the 192 selected SNP. The
113 first group was a subset of the reference dataset composed of 72 individuals: 44 Pekin, 15 Muscovy and
114 13 mule ducks, in order to ensure consistency between KASPar and Axiom results. The second group
115 encompassed most of the parents (134 Pekin and 128 Muscovy ducks) of the experimental batches to
116 assign. To select the final 96 markers with desired properties, similar criteria as for the previous step were
117 used: markers were kept when they had maximum call-rate of 5% missingness, a within line MAF of 0.15
118 and absence of Mendelian incompatibilities, the latter being assessed using samples with known kinship
119 (nine offspring-sire-dam triplets in Pekin, four offspring-sire pairs and two offspring-dam pairs for mule
120 ducks).

Assessment of the assignment power of the 96 SNP panel

122 An evaluation of the assignment power of the marker set was carried out by computing the exclusion
123 probability (Vandeputte, 2012), which is the probability of a randomly chosen parent-pair being
124 genetically excluded as parents of a randomly chosen offspring, when that parent pair did not produce
125 that offspring (Dodds et al., 1996). It depends on the number of parents and the allele frequencies in the
126 parental population. It provides a good quality criterion for the set of markers once the parental
127 population is genotyped.

Sample Collection and Genotyping

129 Blood samples from offspring and their parents were collected after slaughtering and sent to the
130 INRAE genotyping platform Gentyane (Clermont-Ferrand, France) for DNA extraction and
131 genotyping. Genomic DNA extraction was performed using GenFind V2™ (Beckman Coulter) commercial
132 kit. The offspring were genotyped for parentage assignment using 96 SNP in KASPar . Dynamic Array™ IFC
133 96 * 96 chips were used with Biomark™ HD Reader to perform the competitive PCR and chip reading. The
134 Fluidigm® SNP Genotyping Analysis software was used to analyze the genotyping results.

135

Parentage Assignment Validation in an Experimental Design

Ethical statement

The present study was conducted in agreement with the 2010/63/EU regulation for use of animals for research purposes. Animals were bred at the INRAE Duck farm (UEPFG, Benquet, France) which has been approved for animal experimentation (C40-037-1). Experiments were carried out following a protocol approved by the French Ministry of Higher Education, Research and Innovation, abiding by European regulations for animal care (APAFIS# 2018013116519672).

Mating design

The mating plan was designed with the double purpose of achieving pedigree assignment in a limited size population (our testing capacity did not exceed 280 ducklings in Pekin, 220 animals in Muscovy and mule ducks) with related breeders, while preserving enough genetic diversity in the offspring population to estimate genetic parameters. The retained strategy was i) to split related breeders in separate factorial designs and ii) to ensure that the largest possible number of maternal origins was represented among ducklings. Each female stock (N= 96 for Muscovy ducks and N=99 for Pekin ducks) was split in three 35 m² cells with slatted floor. These cells were equipped with nests lined with wood shavings to limit the number of floor eggs. To respond to the species specificities, 15 partially closed nests were available in each cell for the Muscovy ducks, whereas for the Pekin ducks, cells were equipped with two large collective nests without roof. Drakes (N=48 for Muscovy ducks and N=40 for Pekin ducks) were kept in individual cages, to avoid aggressive behaviors. A factorial design was organized, where groups of females in a given cell were inseminated with different pre-designed pools of semen from 4 drakes. Females from each cell were split into four groups of eight individuals in the Muscovy population and three groups of eleven females in the Pekin population, each group being identified using a colored leg ring. Thus, in the Muscovy population, the number of possible parental pairs of an egg reduced from 48 males*96 females = 4608 to 3 cells*4 groups*8 females*4 males = 384. In the Pekin population, this number reduced from 3564 to 396. In both populations, based on preliminary genotypes results, this maximum number of parental pairs was considered adequate to properly estimate genetic parameters. Then, the dams and sires were allocated to each cell and grouped depending on their relatedness. In order to facilitate parentage assignment, this setup aimed at avoiding situations where one parent could not be decided between two sibs. First, among breeders, the number of full-sibs never exceeded three males or three females. They were, therefore, allocated to different cells. Second, a maximum of two half-sib dams was allocated to a given cell and they belonged to two distinct groups. In addition, in order to limit inbreeding, no female was allocated with one male sib in the same group. During the two-week reproduction period, each group of females was repeatedly inseminated with pooled semen from the same group of drakes. Following common practices, insemination doses were calibrated to provide 100 million spermatozoids for Muscovy females and 150 million spermatozoids for Pekin females. Contribution of each male was monitored prior to mixing based on optical density of ejaculates, to provide an equal number of spermatozoids from each drake within an insemination dose.

Egg collection and hatching

Eggs were harvested daily during the egg collection period. Day of lay and cell number were written on the shell. After candling prior to the hatcher transfer, eggs were put into hatching baskets (one hatching basket per day of lay and cell number) and then were ordered in the hatcher based on decreasing number of viable eggs. At hatch, ducklings were identified with a wing band until the desired number of ducks was reached, *i.e.* not all hatching baskets were collected. Given the above-mentioned limited testing capacity and assuming a female lays only one egg each day, the ranking of the baskets based on egg numbers was retained to maximize the number of dams contributing to the final retained population. The correspondence between the wing band and the cell number was recorded.

182

A posteriori pedigree assignment

The experimental population to assign was composed of three batches, each related to a genetic type: 157 male Muscovy ducks, 207 male mule ducks and 273 Pekin ducks of both sexes, all issued from the parents first genotyped with the 192 SNP panel. The APIS software (Griot et al., 2020) was used for pedigree assignment. The two available methods were compared. One is based on the maximization of the average Mendelian transmission probability of the markers for a given offspring and all the possible parental pairs. The other one is based on the exclusion principle, where any Mendelian incompatibility eliminates a parental pair until only the true one remains. In order to account for genotyping errors, a user-tuned number of mismatches can be allowed and was set to two. Offspring exhibiting more than 5% missingness in genotypes were excluded from the assignment process, leading to the removal of 7 individuals (*i.e.* 2.6 % of the initial 273 offspring to be assigned) in the Pekin population only. Knowing the effective factorial design, we were able to produce a positive list of possible parental pairs and challenge the putative pedigree produced by the software with factual elements.

Results and discussion

First List of 192 Markers

Among the birds with 600K genotypes available in the reference dataset, only those exhibiting a call rate over 0.95 (*i.e.* with less than 5% of missing information) were retained, leading to a subset of 139 Pekin, 79 Muscovy and 39 mule ducks with genotypes, and a final number of 94 offspring sire dam triplets. Call-rate filtering for markers (maximum 5% missingness) led to a list of 348 SNPs, among which twelve were discarded because of Mendelian mismatch occurrences. SNPs were kept when minor allele frequency exceeded 0.10 in each of the *Anas platyrhynchos* and *Cairina moschata* populations, which led to a list of 232 SNPs. Only SNPs showing some polymorphism in the 39 mule duck samples were kept, reducing the number to 210. This criterion was applied to make sure the retained markers were not monomorphic among mules, as assignment of mule ducks was of prime interest. Finally, the list of 192 primers was obtained after eliminating SNPs exhibiting a LD above 0.25 with other markers.

Design of an Operational 96 SNP Panel

Table 1 - Call-Rate and Minor Allele Frequency (MAF) observed for the 192 SNPs in the parental populations

	Anas platyrynchos N=134		*Cairina moschata* N=128	
	Call-rate	MAF	Call-rate	MAF
minimum	0.940	0.026	0.258	0.047
1st quartile	0.993	0.222	0.984	0.236
median	0.993	0.338	0.992	0.323
3rd quartile	0.993	0.412	1.000	0.418
maximum	1.00	0.500	1.000	0.500

Elementary statistics about CR and MAF of the 192 SNP for our parental populations are displayed in table 1. These results were obtained for the parents of our experimental populations (134 Pekin and 128 Muscovy ducks), which explains why MAF were lower than 0.1 for some markers, as initial thresholds were set on a different population (our reference dataset). In our experimental Muscovy population call-rates were lower than expected. Fifty-seven SNPs exhibited missingness rate ranging from 0.42 to 0.75, while they were below 5% in the Muscovy samples previously genotyped with the 600K chip. Our hypothesis is that undetected polymorphisms in the primer sequences can be incriminated for these poor results. Such polymorphisms remained undetected in the few individuals sampled from the same line as our experimental populations. These 57 SNPs were discarded from the list. This endorses the strategy of

starting with 192 SNPs to retain a final list of 96. Six additional markers exhibiting at least one Mendelian mismatch were deleted, reducing the list to 133. The minimal MAF criterion was set to 0.10 in each parental population, resulting in a list of 111 SNPs. Finally, to ensure desirable properties in the mule duck population, 7 SNPs with a call rate below 0.95 in the 39 mule samples were discarded. Eight additional markers were thrown away based on the clustering quality of their genotypes in the Fluidigm® SNP Genotyping Analysis software, resulting in the final list of 96 SNPs.

Table 2 - Name and position of the 96 SNP retained in the final list

KB745320.1 is a scaffold.

Chromosome	Position (bp)	Marker name	Chromosome	Position (bp)	Marker name
1	109061561	AX-247363485	7	639397	AX-247355830
1	198136954	AX-247363213	7	6642882	AX-247355836
2	9314971	AX-247354978	7	6784807	AX-247364551
2	22038866	AX-247363748	7	7458603	AX-247364557
2	25524298	AX-247355025	7	7903291	AX-247355848
2	48224427	AX-247355091	7	17149047	AX-247364577
2	57105300	AX-247363838	7	37659499	AX-223686578
2	72878000	AX-247363840	8	5024747	AX-247355910
2	95527796	AX-247355149	8	9828535	AX-247364640
2	106227402	AX-247363883	8	18077068	AX-247355936
2	125817433	AX-247355201	8	20064891	AX-247364660
2	130944301	AX-247363942	8	23941232	AX-247364672
2	133449691	AX-247363956	8	25365172	AX-247364675
2	142558953	AX-247355235	8	26073249	AX-247364679
2	148407413	AX-247355249	9	6446865	AX-247364711
2	152370825	AX-247355261	9	10829712	AX-247356029
2	152906965	AX-247355267	9	11668820	AX-247364749
3	178108	AX-247364000	9	13906991	AX-247364763
3	22898020	AX-247355316	9	14469818	AX-247364765
3	34203102	AX-247364053	10	11096372	AX-247356148
3	41332352	AX-247364072	11	15392465	AX-247364917
3	49962556	AX-247364080	12	4812384	AX-247356238
3	53539930	AX-247355356	14	6336270	AX-247356370
3	66856580	AX-247364116	14	14544447	AX-247365129
3	68837303	AX-247364118	14	14827130	AX-247365133
3	74410901	AX-247364122	16	2984766	AX-247356455
3	110150507	AX-247355450	16	3718731	AX-247356463
3	110627101	AX-247355452	16	9044628	AX-247356481
4	6220620	AX-247364191	16	9063242	AX-247356483
4	14309946	AX-247355482	16	13744076	AX-247365233
4	25865050	AX-247355506	16	14448873	AX-247356512
4	60721998	AX-247364276	18	5084874	AX-247356525
5	2505593	AX-247364303	19	10473301	AX-247365299
5	6739459	AX-247364317	19	10494308	AX-247365301
5	7253477	AX-247364320	20	2186582	AX-247365309
5	26939905	AX-247364353	20	6341095	AX-247365329
5	27529421	AX-247355645	20	8185628	AX-247356614
5	36094216	AX-247364375	20	9156425	AX-247356617
5	42690814	AX-247355671	20	11133474	AX-247365344
5	45865637	AX-247355673	21	12160575	AX-247365370
5	54586500	AX-247355693	22	2485025	AX-247365384
5	54717023	AX-247364410	22	2576730	AX-247365386

5	58368151	AX-247364419	24	2923576	AX-247365450
5	58440894	AX-247364421	24	4729515	AX-247365466
5	59563000	AX-247355708	24	5386920	AX-247356749
5	62080514	AX-247355726	25	1070592	AX-247356762
6	28171519	AX-247364508	25	1255481	AX-247356766
6	31606612	AX-247364526	KB745320.1	252400	AX-247364465

230

231

232 The final list of 96 SNP is displayed on table 2, while the list of 192 markers is given as supplementary
233 material.
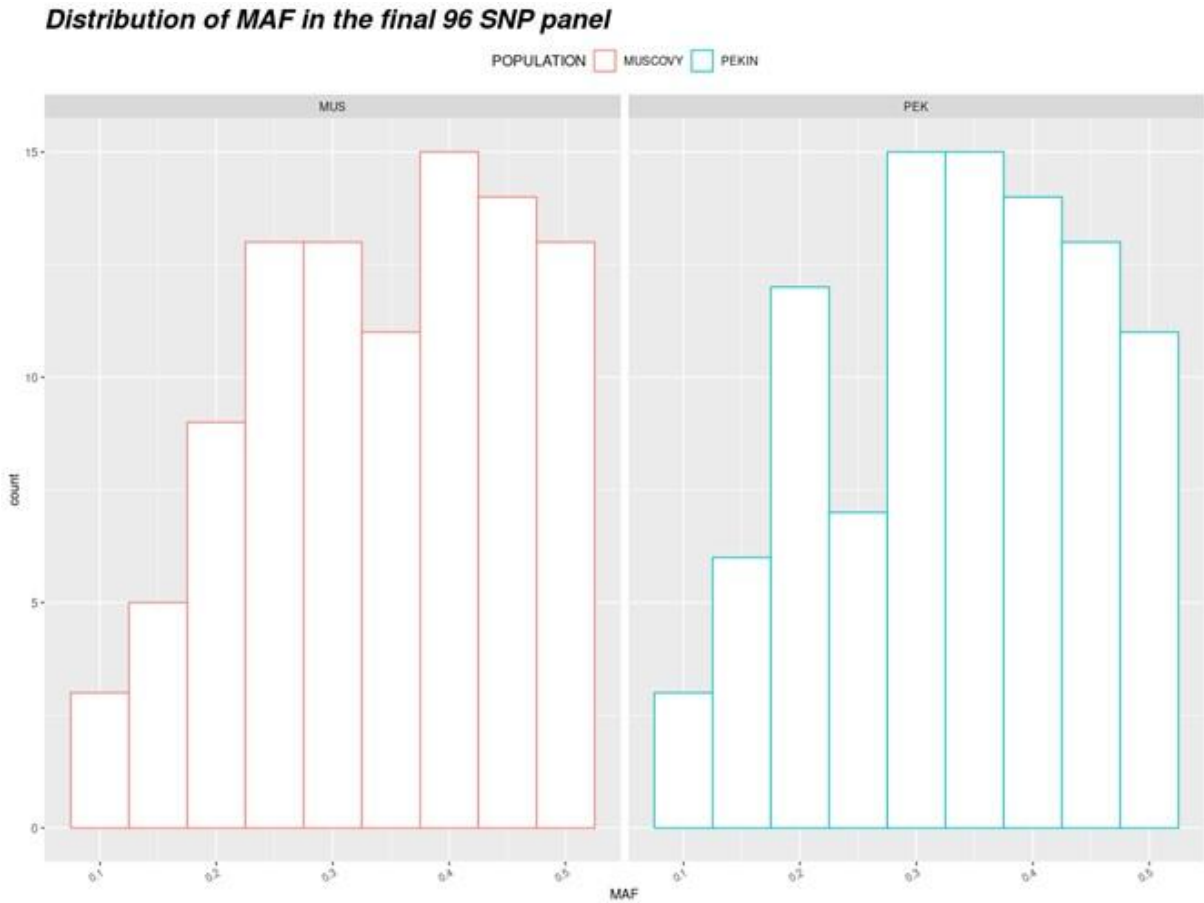
235 **Figure 1 -** Minor Allele Frequency distribution of the final 96 assignment markers in the
236 experimental population

237 The MAF distributions of these 96 SNP in our experimental populations are displayed on figure 1.
238 Figure 2 shows the location of the SNPs on the different chromosomes. The localization of the 192 SNPs
239 (upper panel) was somehow consistent with the size of chromosomes, with a larger number of SNPs on
240 macro-chromosomes compared with micro-chromosomes. No SNP was located on chromosome 17 and
241 23. For the final set (lower panel), the priority was given to technical proprieties of the markers, leading to
242 some gaps (no SNP on chromosomes 13, 15, 17 and beyond 25) and only two on chromosome 1.
243 Nonetheless, the vast majority of SNPs in the final set were located on macro-chromosomes (numbered
244 from 1 to 8, following Skinner et al.(2009)).

245 A consistency (i.e. percentage of identical genotypes) of 0.997 was observed between the genotypes
246 of the 72 individuals in the reference panel, which were obtained with both technologies (KASPar and

247 Axiom). Three individuals were genotyped twice with the KASPar technology with complete (100%)
248 consistency. As previously stated, the set of animals used to obtain the final list of 96 markers contained
249 nine individuals of complete known pedigree. Using these 96 SNPs, all offspring in the 9 trios of known
250 pedigree were correctly assigned to their true parental pair using the APIS R package.

251
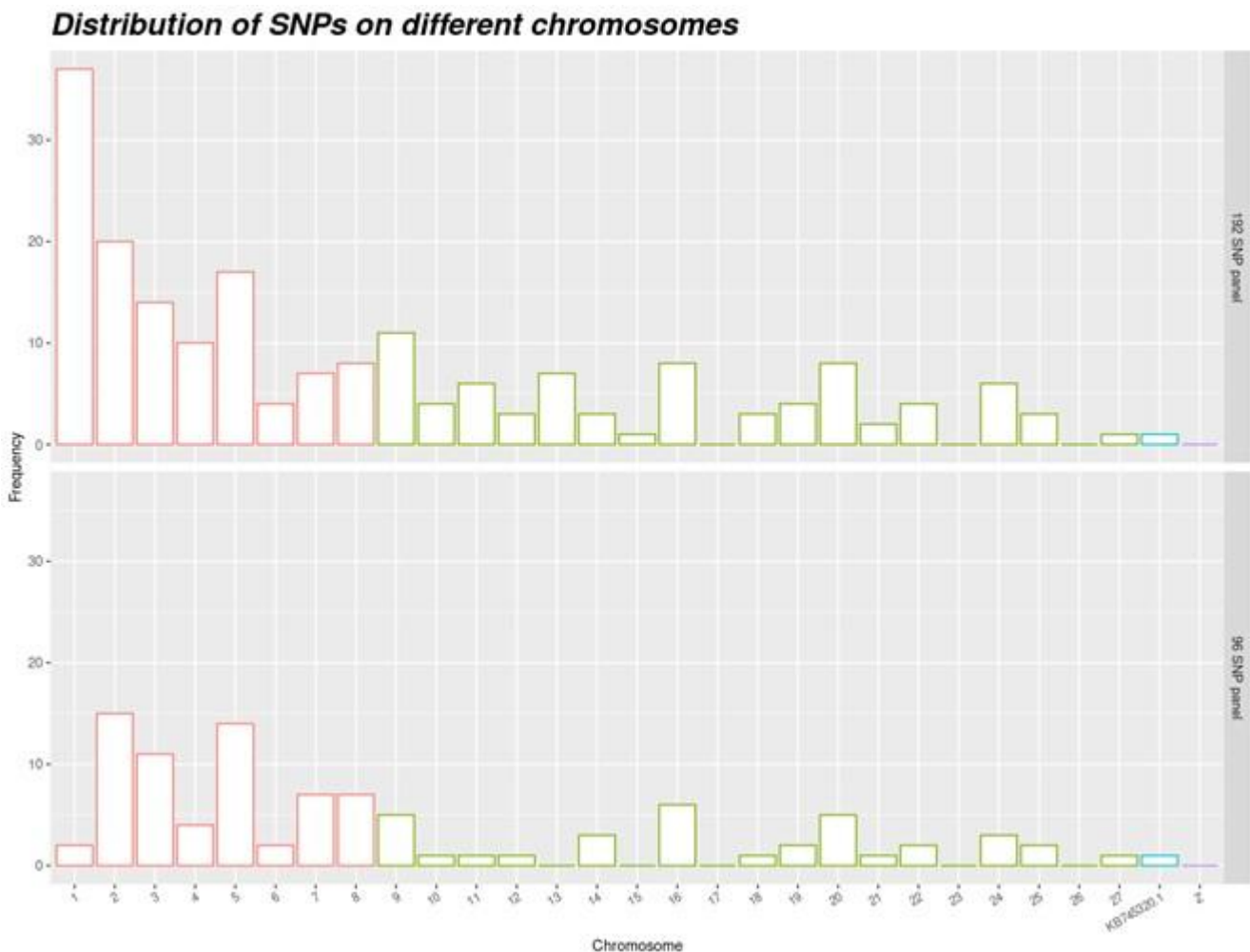
252 **Figure 2 -** Location of the SNPs on the chromosomes (upper part: 192 SNP panel and lower part:96
253 SNP panel) *Chromosomes 1 to 8 are macro-chromosomes, chromosomes 9 to 27 are micro*
254 *chromosomes, Z is a sexual chromosome and KB745320.1 is a scaffold.*

255 Following Vandeputte (2012), the exclusion power of the 96 SNP panel, based on the allele
256 frequencies in the parental population and assuming random mating, was computed and found above
257 0.99999 in all the populations These values were an encouraging result before attempting to build the
258 pedigree of our experimental batches.

259 **Obtention of a DNA-based Pedigree of our three Experimental Populations**

260 *Assignation rate*

261 The assignment rate to a unique parental pair was 99% for the mule ducks (204 over 207), 96% in the
262 Pekin population (261 over 273), and 88% in the Muscovy population (138 over 157). A posteriori, this
263 poor performance in the Muscovy population could be attributed to the absence of 16 parental samples
264 in the genotyped populations (fourteen females and two males). Yet, with 88% of success this set of
265 markers performed at least as well as the previous microsatellite panel (Chapuis et al., 2010). In this
266 study, assignment failures occurred when the most probable putative parent pairs identified had a
267 relatively high number of Mendelian incompatibilities (above eight, when the threshold was set to two

268 mismatches). In addition, this was confirmed by the two-peaked distribution of the difference in
269 Mendelian transmission probability between best and second-best putative parents (figure 3) for the
270 Muscovy offspring, unlike the two other populations. As stated by Griot et al. (2020) assuming a sufficient
271 power of the panel (exceeding 0.99999 here), this situation clearly signaled missing parents. This
272 demonstrates that the main obstacle for a posteriori building of pedigree is the absence of one or both
273 parents. To confirm this hypothesis, the absence of the same number of parents (two sires and fourteen
274 dams randomly discarded) was simulated in the Pekin population and, over 50 replicates, the average
275 assignment rate dropped to 0.80±0.01, i.e. a loss of 16 percentage points. In these replicates, the
276 maximum number of observed mismatches in the assigned individuals was 2, while, in the non-assigned
277 Muscovy individuals, it ranged between 5 and 11, indicating a clear cut-off when one parent is missing.
278 Another cause of APIS assignment failures may be the wrong estimation of the empirical threshold to be
279 set in Mendelian transmission probability. According to (Griot et al., 2020), a minimal number of 200
280 offspring is required to properly estimate this threshold, while we had only 157 Muscovy.

281 *Benefits of Mating Plan Knowledge*

282 The outcomes of an APIS run can be split into three situations: i) direct successful assignment to the
283 rightful parental pair, ii) wrong assignment to an erroneous parental pair, or iii) failure to return a unique
284 parental pair. In our case, thanks to the availability of the mating plan, the two latter situations could be
285 sorted out in most cases. As an illustration, the vast majority of assigned parental pairs was fully
286 compatible with both the list of possible mating and the cell number where the egg was collected,
287 associated with the wing band. They were, therefore, considered as correct, and corresponded to case i.
288 In addition, these pieces of information allowed to detect and fix one single wrong assignment returned
289 by the software. In this case, the parental pair ranking first on Mendelian transmission probability could
290 materially not be the true one, unlike the second ranking pair, exhibiting a Mendelian transmission
291 probability only slightly lower than the first one (case ii). Assignment failure (case iii) occurred in very few
292 situations (less than 5% of cases in Pekin and mule populations), for instance, when the two most
293 probable parental pairs featured the same sire while the different dams could not be separated based on
294 Mendelian transmission probability only. In these cases also, supplementary information brought by the
295 wing band, which identified which cell number the egg originated from, and thus which mating was
296 possible, helped to designate the true pair among the putative pairs proposed by APIS.

297 **Consequences on the Population Structure**

298 Avian pedigreed populations are usually bred using individual cages for females, applying a
299 hierarchical mating design (a single male used to inseminate p females, a dam having offspring from one
300 sire only). In factorial designs allowed by group housing, a female can give birth to ducklings with multiple
301 sires, up to four different drakes in our case. Table 3 displays, for each of three experimental batches, the
302 proportion of dams which had progeny identified from k males, k varying from 1 to 4. The population
303 structure here is different from a hierarchical mating design, as less than half of the dams had offspring
304 from only one sire. This remarkable change in the mating design is displayed on figure 4, which shows the
305 last batch of Pekin and its two generations of closest ancestors (parents and grand-parents). When the
306 hierarchical mating plans operated, much less combinations of sires and dams were recruited than when
307 the mating scheme was factorial. Population structure varied among the three genetic types displayed in
308 table 3. Without any replicate, however, it is not possible to infer the differential consequences to be
309 expected in the three populations once the hierarchical mating plan is replaced by a factorial one.

310
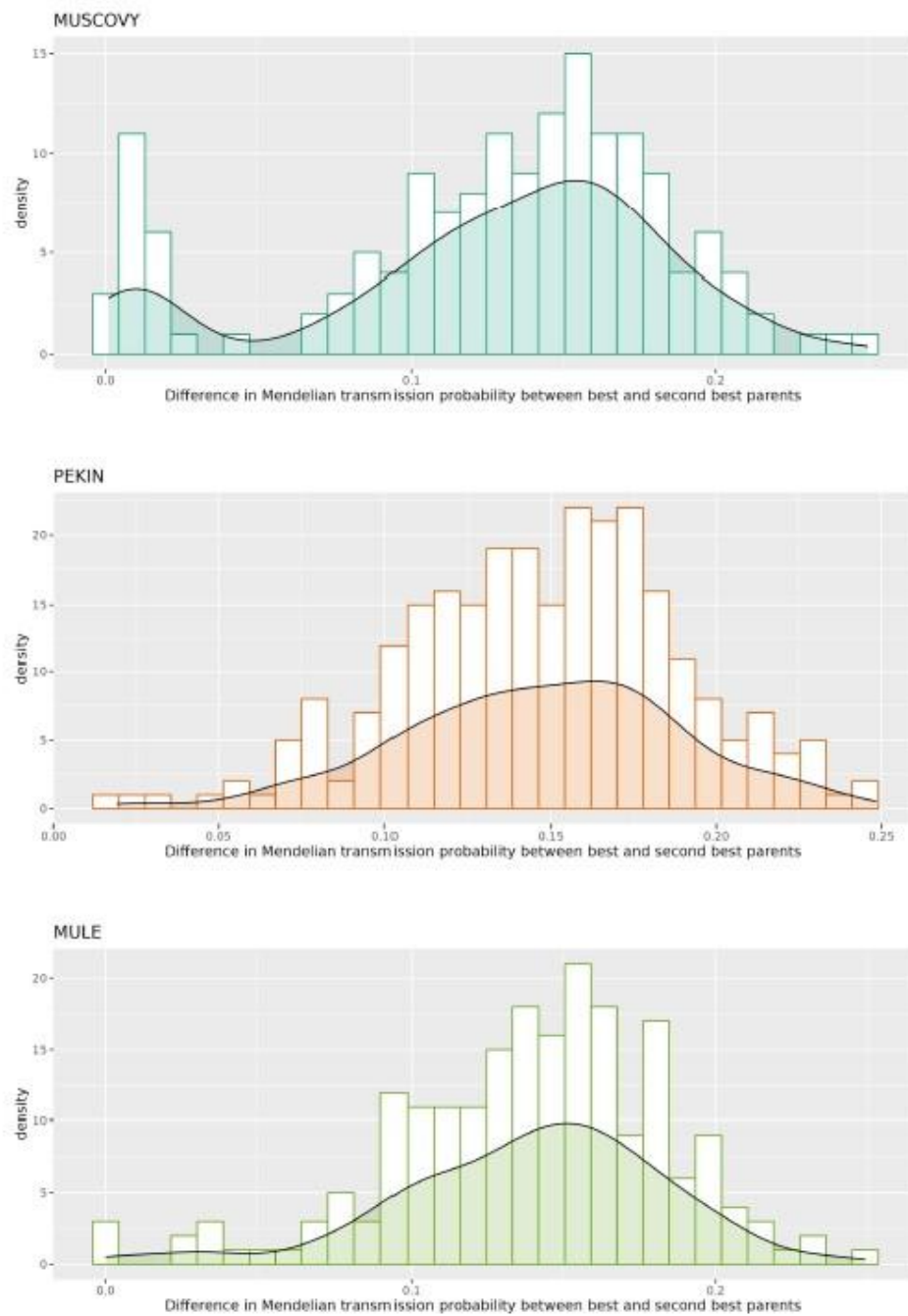
Figure 3 - Distribution of differences in Mendelian transmission probability between best and second-best putative parents in the three populations

Table 3 - Proportion of dams giving birth to ducklings with k different sires

	Population		
k	*Cairina moschata*	Mule ducks	*Anas platyrynchos*
4	_	33%	12%
3	16%	41%	30%
2	36%	20%	27%
1	48%	6%	30%

318

319   It is useful here to remind that, given the characteristics of poultry reproduction, in particular the
320   presence of sperm storage tubules in the oviduct of females, the hierarchical mating plan carried out for a
321   long enough period was, regardless of the housing system, the only way to ascertain the pedigree of
322   newborn chicks before the availability of molecular tools allowing for parentage assignment. Thus,
323   females could be housed in cages or in pens, but they were mated to a single male during a given egg
324   collection period. Switching from hierarchical to factorial design is recommended first for practical
325   convenience when individual cages are banned: it is indeed easier to pick a female based on its colored
326   leg ring and inseminate it with a prepared semen pool than randomly pick a female, read its wing band
327   and inseminate it with sperm from the single relevant male. Besides, not only the SNP panel allows for
328   parentage assignment but it also provides context for the correct estimation of maternal effects, which
329   are no longer confounded with a sire-dam interaction in a given laying series, as can be seen on figure 4.
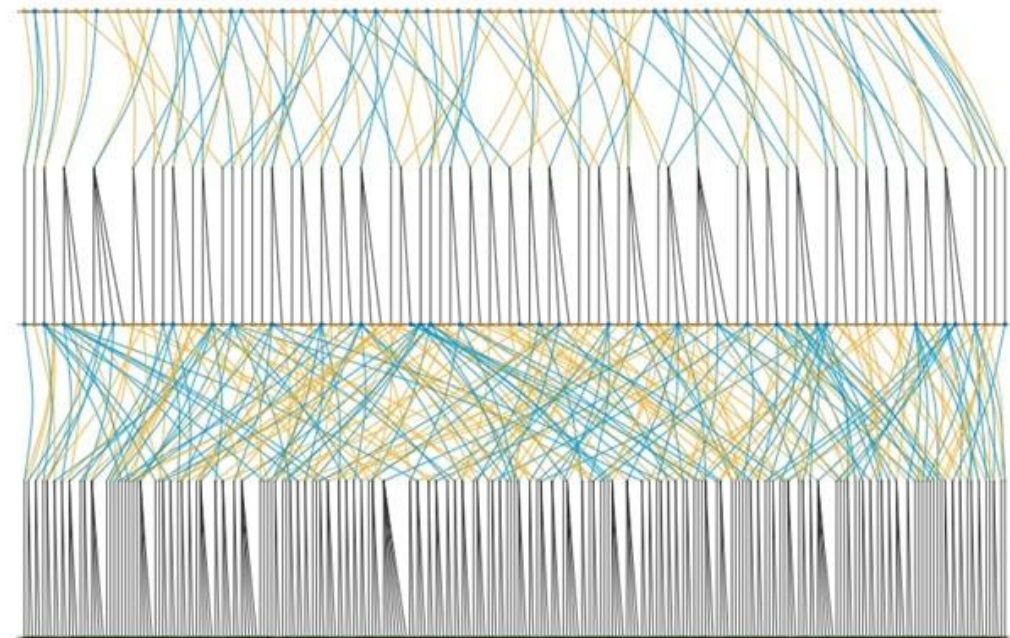

330

331   **Figure 4 -** Pedigree representation of the last two generations in the Pekin line

In the context of duck breeding for fat liver production, such a change in breeding schemes is prone to dramatically impact the way Pekin lines (i.e. the dam pathway of the mule duck) are selected. Indeed, their breeding values used to be computed based on purebred performances (body weight and laying performances) and crossbred performance measured on mule offspring. When these offspring are obtained through a hierarchical mating design, the dam estimated breeding value is confounded with the Muscovy drake potential, which may lead to bias, if the sire breeding value is not properly estimated, a common situation when evaluations for both Pekin and Muscovy lines are not carried out simultaneously. If mule offspring are obtained with multiple drakes for each dam, the bias partly wipes out. Besides, in the case of low male fertility, a Pekin female will potentially have lesser progeny with a hierarchical mating design than with a factorial mating design, due to the male side. Switching from a hierarchical to factorial mating design should, therefore, improve the selection process on the dam pathway.

Table 4 - Dam family structure in three successive batches of mule ducks

M1 and M2 were obtained using individual cages and a hierarchical mating design. M3 was *obtained using a factorial design and pedigreed through genotyping.*

batch	# anim	# dam	Dam family size									MEAN	VAR
			1	2	3	4	5	6	7	8	9		
M1	247	**87**	4	17	55	11						2,84	0,49
M2	282	**84**	10	15	22	18	14	2	2	1		3,36	2,33
M3	204	**69**	18	14	15	11	4	2	4		1	2,96	3,40

On the other hand, management of breeding resources raises new issues in the case of floor reproduction and late pedigree knowledge. When the parents of the egg are known at egg collection (i.e. with a hierarchical design applied to individually caged females), it is easy to monitor family size at hatch and obtain a balanced family representation for a given batch size. This can be assessed looking at table 4, which displays dam family structure in three successive batches of mule ducks. In the latest mule batch (obtained under factorial design), dams had from 1 to 9 offspring, with an average of 2.96 ±1.88. Only 70% of the dams had male offspring in this latest batch. This proportion was above 85% in the previous batches- with hierarchical designs. This drop can originate from the sampling of ducklings at hatch (males kept until the desired number was reached), when the dam is not yet known, and may also be due to zootechnical issues, if some females did not lay hatchable eggs, or only floor eggs. Such an unbalanced contribution of breeders to the progeny due to free mating system has been described by Brard-Fudulea et al. (2023) in red partridge. Therefore, pen size (cell size in our situation) and animal sorting should be carefully organized, lest origins may be lost. In addition, there is room for optimization of the mating design. Usually mating plans are designed in order to monitor the increase of inbreeding rate, for instance by avoiding common ancestors between associated groups of males and females. Here another constraint should be imposed on the common ancestors within a group of breeders, lest difficulties arise to find the true parental pair. One solution could be to use, in the optimization process, a kinship matrix based on genotypes instead of the numerator relationship matrix derived from pedigree. One could also imagine minimize the expected inbreeding of future progeny, as do most mating plan setup software, while setting a constraint on a molecular kinship of breeders computed using marker genotypes. A similar algorithm (simulated annealing mixed with Lagrangian multiplier) was used by Chapuis et al. (2016) to optimize breeder selection under a constraint on kinship.

373   Last, but not least, here females were inseminated and doses were calibrated to equilibrate male
374   contributions. Ultimately, in breeding companies with large populations, one could be tempted to rely on
375   natural mating, using pens with p males and q females, like at the multiplication stage. Such condition
376   would add another heterogeneity factor with mating behavior likely to dramatically impact family
377   composition. A thorough modeling of selection schemes is, therefore, necessary, to face the replacement
378   of hierarchical mating design with factorial ones.

379   **Assignation power in other duck and poultry populations**

380   This 96 SNP panel was explicitly designed to perform in our experimental population. Yet, eight 95 x
381   96 chips were used to obtain 192 SNP genotypes, leaving some spots available that were used to collect
382   genotypes for local breed samples. Thirty-four Duclair and 10 Rouen individuals (two local breeds of Anas
383   platyrhynchos) were thus genotyped. Minor allele frequencies averaged 0.26 and 0.29, respectively, in
384   these two populations. These values were lower than those reported for our experimental lines in table 1.
385   They nonetheless led to exclusion probabilities above 0.99 in these two populations, giving way to a
386   potential use for improved management of genetic resources. Practically, a side outcome of this study is a
387   list of 135 SNPs (i.e. the initial list of 192 SNPs, deprived of the 57 markers that did not work in our
388   Muscovy population) with reliable properties being now available in *Anas platyrhynchos*, *Cairina*
389   *moschata* and their hybrid offspring, to setup SNP sets for any commercial or local population.
390   Commercial populations undergoing genomic selection are not concerned with the need of an efficient
391   assignment marker set, as the thousands of SNP on a chip can also be used to build pedigree. Yet, the
392   question remains for mule offspring, as usually, for cost reasons, only selection candidates are genotyped
393   using medium or low density (MD) chips featuring 10 to about 50K SNPs. Should individual cages be
394   banned in European breeding companies, mule ducks would also require genotyping and then the cost
395   benefit ratio of using a 96 SNP set vs. a MD chip should be carefully reevaluated.

396   As previously stated, the setup of an operative assignment panel is not an issue in widely distributed
397   poultry species, where genomic material has already been developed (chicken, turkey, ducks). This can be
398   more complicated with minor species such as guinea fowl or game (partridge or pheasant). Yet, in Europe,
399   breeders operating in these species could be also concerned with the ban of individual cages. Recently, in
400   red partridge, assignment rate reached 90% using a 96 SNP panel (Brard-Fudulea et al., 2023). In their
401   review, Flanagan and Jones (2019) noted that as few as 31 SNPs could be used to assign all offspring with
402   >99% confidence in a population of wild birds.. They also reported many examples (mostly in fish, some in
403   mammals) where 96 SNP panels would be sufficient to provide a unique parental pair for each offspring.
404   In our situation, we benefited from previous work carried out in ducks and the availability of a 600K
405   microarray. Assignment panels could also be obtained de novo using Next Generation Sequencing (NGS)
406   methods. As stated by Guichoux et al. (2011) these technologies enable the identification of large
407   numbers of microsatellite loci at reduced cost in non-model species. Consequently, more stringent
408   selection of loci is possible, thus further enhancing multiplex quality and efficiency. This potentially could
409   allow for a microsatellite panel avoiding the pitfall encountered by Chapuis et al., (2010) where the
410   available microsatellites were not sufficiently polymorphic in both parental populations simultaneously.
411   NGS methods also provide different ways to obtain sets of SNPs that could be used for parental
412   assignment. For instance, in Atlantic salmon, Holman et al.(2017) used RAD markers (Miller et al., 2007) to
413   identify SNPs to be developed into a marker set. Knowledge of the mating plan allowed for a 100%
414   accuracy in parentage resolution with no more than 94 SNPs, even when putative parents were related.
415   These results, in accordance with our own, leaves to hope that a set of 96 SNP and some practical rules
416   for bird management could be enough to provide an affordable tool for effective parentage assignment in
417   most commercial poultry populations.

418

419   **Conclusion**

420   In this study, starting from a 600K Axiom chip, a 96 SNP panel was developed and proved effective to
421   correctly assign parentage in an experimental population of three connected genetic types. Technical

422 steps, including an intermediate selection of 192 SNPs first evaluated in the populations of interest,
423 revealed that the selection of markers to transfer from one technology (Axiom) to another (KASPar)
424 deserves thorough attention. Besides, as poultry populations have limited effective sizes, an optimization
425 of the factorial design was needed to avoid genetically similar types of progenies in the same pen (issued
426 from sibling breeders), which resolved most of the dubious assignations, and the pending ones actually
427 pointed out missing samples in the parents.

428 The ban of individual cages is likely to dramatically impact selection schemes in poultry species. Here,
429 we suggest to switch from a hierarchical to a factorial mating design, which leads to clear changes in the
430 population structure. Their consequences in the long term for selection schemes still remain to be
431 investigated, and the management of the mating plans (*i.e.* pen size) will have to be optimized
432 accordingly. In addition to an impact on the pedigree, banning individual cages will also affect the
433 individual recording of laying traits, and the development of connected nesting devices to record laying
434 performances of female ducks will also be a concern.

435

440

Conflict of interest disclosure

442 The authors declare that they comply with the PCI rule of having no financial conflicts of interest in
443 relation to the content of the article.

444

447

References

449 Arndt, M. (1931). *Battery Brooding* (2nd ed.). Orange Judd Publishing.

450 Brard-Fudulea, S., Rouger, R., & Tricoire, S. (2023). Main challenges for red partridge mating in cage-free
451 system: reproduction efficiency, pedigree recording, and management of genetic diversity. In WPSA
452 (Ed.), *12th European Symposium on Poultry Genetics, Hannover, Germany*.

453 Chapuis, H., Faugeras, R., Rossignol, M. N., Feve, K., Genestout, L., Chantry-Darmon, C., & Guémené, D.
454 (2010). Vers l'utilisation des marqueurs moléculaires pour l'assignation des parentés en canard
455 grâce au programme Apache. *9ièmes Journ. Rech. Palmipèdes à Foie Gras*, 1–4.

456 Chapuis, H., Pincent, C., & Colleau, J. J. (2016). Optimizing selection with several constraints in poultry
457 breeding. *Journal of Animal Breeding and Genetics*, *133*(1), 3–12. https://doi.org/10.1111/jbg.12178

458 Dodds, K. G., Tate, M. L., McEwan, J. C., & Crawford, A. M. (1996). Exclusion probabilities for pedigree
459 testing farm animals. *TAG. Theoretical and Applied Genetics. Theoretische Und Angewandte Genetik*,
460 *92*(8), 966–975. https://doi.org/10.1007/BF00224036

461 Flanagan, S. P., & Jones, A. G. (2019). The future of parentage analysis: From microsatellites to SNPs and

462   beyond. *Molecular Ecology*, *28*, 544–567. https://doi.org/DOI: 10.1111/mec.14988

463   Griot, R., Allal, F., Brard-Fudulea, S., Morvezen, R., Haffray, P., Phocas, F., & Vandeputte, M. (2020). APIS:
464   An auto-adaptive parentage inference software that tolerates missing parents. *Molecular Ecology*
465   *Resources*, *20*(2), 579–590. https://doi.org/10.1111/1755-0998.13103

466   Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Lé Ger, P., Lepais, O., Lepoittevin, C., Malausa, T.,
467   Revardel, E., Salin, F., & Ti, E. (2011). Current trends in microsatellite genotyping. *Molecular Ecology*
468   *Resources*, *11*(4), 591–611. https://doi.org/10.1111/j.1755-0998.2011.03014.x

469   Holman, L., Garcia de la serrana, D., Onoufriou, A., Hillestad, B., & Johnston, I. (2017). A workflow used to
470   design low density SNP panels for parentage assignment and traceability in aquaculture species and
471   its validation in Atlantic salmon. *Aquaculture*. https://doi.org/10.1016/j.aquaculture.2017.04.001

472   Huang, Y., Li, Y., Burt, D. W., Chen, H., Zhang, Y., Qian, W., Kim, H., Gan, S., Zhao, Y., Li, J., Yi, K., Feng, H.,
473   Zhu, P., Li, B., Liu, Q., Fairley, S., Magor, K. E., Du, Z., Hu, X., … Li, N. (2013). The duck genome and
474   transcriptome provide insight into an avian influenza virus reservoir species. *Nature Genetics*, *45*(7),
475   776–783. https://doi.org/10.1038/NG.2657

476   Marx, G., Klein, S., & Weigend, S. (2002). An automated nest box system for individual performance
477   testing and parentage control in laying hens maintained in groups. *Archiv Für Geflügelkunde*, *66*(3),
478   141–144.

479   Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnson, E. A. (2007). Rapid and cost-effective
480   polymorphism identification and genotyping using restriction site associated DNA (RAD) markers.
481   *Genome Research*, *17*(2), 240–248. https://doi.org/10.1101/gr.5681207

482   Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., De
483   Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association
484   and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575.
485   https://doi.org/10.1086/519795

486   Skinner, B. M., Robertson, L. B. W., Tempest, H. G., Langley, E. J., Ioannou, D., Fowler, K. E., Crooijmans, R.
487   P. M. A., Hall, A. D., Griffin, D. K., & Völker, M. (2009). Comparative genomics in chicken and Pekin
488   duck using FISH mapping and microarray analysis. *BMC Genomics*, *10*(1), 1–11.
489   https://doi.org/10.1186/1471-2164-10-357/TABLES/1

490   Teissier, M., Thébault, N., Riquet, J., Diot, C., Brard-Fudulea, S., Guemene, D., Alletru, B., Cornil, M.,
491   Bouleau, P., Blanchet, M., Guémené, D., Le Mignon, G., Demeure, O., & Vignal, A. (2019).
492   Development and validation of high-density SNP array in ducks. *XIth European Symposium on*
493   *Poultry Genetics (ESPG)*, 62–64. http://broadinstitute.github.

494   Thébault, N., Riquet, J., Diot, C., Brard-Fudulea, S., Guémené, D., Blanchet, M., Le Mignon, G., Demeure,
495   O., Alletru, B., Cornil, M., Bouleau, P., & Vignal, A. (2019). Développement d'une puce de
496   génotypage haute densité 600K pour le canard commun et le canard de Barbarie. *13ièmes Journées*
497   *de La Recherche Avicoles et Palmipèdes à Foie Gras*.

498   Vandeputte, M. (2012). An accurate formula to calculate exclusion power of marker sets in parentage
499   assignment. *Genetics Selection Evolution*, *44*(1), 2–5. https://doi.org/10.1186/1297-9686-44-36

500

501

502