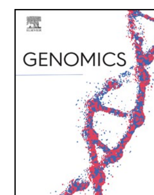
Genomics 112 (2020) 1660–1673

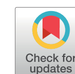
Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno



Variations in genome size between wild and domesticated lineages of fowls belonging to the *Gallus gallus* species



Benoît Piégu[a], Peter Arensburger[b], Linda Beauclair[a], Marie Chabault[c], Emilie Raynaud[c], Vincent Coustham[c], Sophie Brard[d], Sébastien Guizard[a,e], Thierry Burlot[f], Elisabeth Le Bihan-Duval[c], Yves Bigot[a,*]

[a] PRC, UMR INRA0085, CNRS 7247, Centre INRA Val de Loire, 37380 Nouzilly, France
[b] Biological Sciences Department, California State Polytechnic University, Pomona, CA 91768, United States
[c] BOA, INRA, Université de Tours, 37380 Nouzilly, France
[d] SYSAAF, Centre INRA Val de Loire, 37380 Nouzilly, France
[e] Florimond Desprez, 3 Rue Florimond Desprez, 59242 Cappelle en Pélève, France
[f] NOVOGEN, 5 rue des Compagnons, Secteur du Vau Ballier, 22960 Pledran, France

ARTICLE INFO

Keywords:
Repeats
CNV
Domestication
C-value

ABSTRACT

Efforts to elucidate the causes of biological differences between wild fowls and their domesticated relatives, the chicken, have to date mainly focused on the identification of single nucleotide mutations. Other types of genomic variations have however been demonstrated to be important in avian evolution and associated to variations in phenotype. They include several types of sequences duplicated in tandem that can vary in their repetition number.

Here we report on genome size differences between the red jungle fowl and several domestic chicken breeds and selected lines. Sequences duplicated in tandem such as rDNA, telomere repeats, satellite DNA and segmental duplications were found to have been significantly re-shaped during domestication and subsequently by human-mediated selection. We discuss the extent to which changes in genome organization that occurred during domestication agree with the hypothesis that domesticated animal genomes have been shaped by evolutionary forces aiming to adapt them to anthropized environments.

1. Introduction

Domestication of the red jungle fowl (RJF; *Gallus gallus*) was well established 8000 years ago based on archaeological data [1,2], while one study based on molecular data estimated it could have begun as early as 42,000 to 74,000 years ago [3]. It is thought to have emerged simultaneously, or during close periods of time, in several regions of South and Southeast Asia [2,4]. Presumed traces of hybridization with sibling species, the grey (*G. sonneratii*) and the green (*G. varius*) jungle fowls [5–7] were detected both in genes and phenotypic traits of domesticated chicken breeds. Globally, the history of chicken domestication can be summarized in 3 main steps: i) emergence of the first domesticated chicken populations in south Asia, ii) diffusion over the world that led to the emergence of several old breeds depending on their geographic location, and iii) emergence during the twentieth century of modern lines selected by the breeders.

Domestication can be viewed as resulting from an adaptation of animals to an anthropized environment that has led them to lose their ability to survive in the wild [8]. In the case of commercial chicken breeds, these animals are also the result of intense selection processes that occurred since the end of the twentieth century to improve their performance and quality for the production of meat or eggs. Phenotypic

Abbreviations: aCGH, array-based comparative genome hybridization; AR, araucana; BBL1, Brown_Layer_Line_1; BBL2, Brown_Layer_Line_2; bp, base pair; CNV, copy number variation; DAPI, 4′,6-diamidino-2-phenylindole; DNA, deoxyribonucleic acid; DS, DNA size; EDTA, Ethylenediaminetetraacetic acid; GAPDH, glyceraldehyde-3-phosphate deshydrogenase; gDNA, genomic DNA; GO, gene ontology; HTS, high throughput sequencing; KD, korean domestic; lncRNA, long non-coding RNA; PBS, phosphate-buffered saline; PCR, polymerase chain reaction; RD, read density; rDNA, ribosomal DNA; RJF, red jungle fowl; rRNA, ribosomal RNA; SDS, sodium dodesyl sulfate; SNP, single nucleotide polymorphism; WL, white leghorn; WLL, white leghorn line; WLL1, White_Layer_Line_1; WLL2, White_Layer_Line_2; WR, white plymouth rock

* Corresponding author.
E-mail address: yves.bigot@inra.fr (Y. Bigot).

https://doi.org/10.1016/j.ygeno.2019.10.004
Received 11 February 2019; Received in revised form 5 August 2019; Accepted 7 October 2019
Available online 06 November 2019
0888-7543/ © 2019 Elsevier Inc. All rights reserved.

changes in the chicken due to domestication have been characterized at multiple levels including emotional and social behavior, body and organ morphology, different aspects of their physiology, as well as their robustness with respect to the environment (e.g. in [9–15]). At the genetic and genomic levels, selection signatures on allele frequencies of some genes were characterized in domesticated chicken breeds [16]. Similar to several other domesticated animal species and with regard to RJF, chickens were found to display significantly reduced genetic variation, an increased proportion of non-synonymous amino acid changes, and a tendency to harbor higher rates of changes classified as damaging [17]. Interestingly, while the chicken genome was one of the first vertebrate genome to be sequenced, the impact of domestication on global genome features such as size variations, amount of repeated components or telomeric and centromeric sequences between RJF and chicken breeds were overlooked, except by a few studies of centromeric and telomeric repeats [18,19].

We investigated whether there were genome size variations between individuals from different RJF species, chicken breeds and selected chicken lines. Using a cytometry approach we found that the genomes of several domestic breeds and some selected chicken lines had statistically smaller genome sizes than RJF. To understand the origins of these smaller genomes we investigated selected genomic features such as the interspersed repeats, telomeric repeats, and regions that are tandemly duplicated or deleted (regions gathered under the terms “Segmental Duplications” or “Copy Number Variations” (CNVs)). Of these, specific types of tandem repeats (i.e. segmental duplications with variable sized repeated units) were found to vary between breeds and lines and were likely the origin of the observed differences in genome size between RJF and some domesticated chickens.

2. Materials and methods

2.1. Biological samples

In the first sample set blood samples (1 mL, Table S1) were collected from 3 wild fowl species (the red jungle fowl (*Gallus gallus bankiva*), the grey jungle fowl (*G. sonneratii)*, the Sri Lankan junglefowl (*Gallus* lafayettii) and eighteen domestic chicken breeds and lines (Table S1) maintained at the INRA avian experimental unit (PEAT, Centre INRA Val de Loire, France), the “Parc des oiseaux” (Villars les Dombes, France), the “Domaine de la Puisaye” (Grandchamps, France) and the “Centre de Sélection de Béchanne” (Saint-Étienne-du-Bois, France). Blood samples from four commercial lines (White_Layer_Line_01 (WLL1) and White_Layer_Line_02 (WLL2) were selected from a white leghorn line, while Brown_Layer_Line_01 (BLL1) and Brown_Layer_Line_02 (BLL2) originated from a Rhode Island Red line) were supplied by Hendrix Genetics Corporate (Boxmeer, Netherlands). For each breed/line the blood of 1–4 females less than two years old was collected in EDTA tubes and stored at 4 °C. A second sample set of blood samples (1 mL) were collected from four commercial white leghorn lines, WLL_A (48 samples), WLL-_B (45 samples), WLL_C (45 samples) and WLL_D (45 sampes); these were supplied by Novogen SAS (Peldran, France). For each line blood samples from less than two year old females were collected in EDTA tubes and stored at 4 °C.

2.2. Measure of DNA amount in nuclei of red blood cells

The C-value is the amount, in picograms, of DNA contained within a haploid nucleus (e.g. a gamete) or one half the amount in a diploid somatic cell of a eukaryotic organism. In some cases (notably among diploid organisms), the terms C-value and genome size are used interchangeably. For the chicken, the only C-value referenced in the Animal Genome Size Database (http://www.genomesize.com/index.php) is that of the RJF which ranges from 1.25 to 1.28 pg per haploid genome [20]. Since our aim was to measure differences in DNA size (DS) between nuclei of various chickens and RJF, we normalized our calculation by setting the DS of RJF to 1.

The DS in blood cells of each individual was evaluated by cytometry following the staining of nuclear DNA with 4′,6-diamidino-2-phenylindole (DAPI) as described in [21]. Briefly, an aliquot of red blood cells (5 μL) was washed in 5 mL 1× phosphate-buffered saline (PBS). After centrifugation at 200*g* at 4 °C for 5 min, the pellet was decanted and the cells were suspended in 400 μL PBS before being fixed by adding 4.6 mL 70% ethanol and incubated for 2 h at 4 °C. Fixed cells were then washed twice with PBS by centrifuging them at 200*g*, 4 °C for 5 min and each pellet was suspended in 5 mL of PBS. At the end of the second wash each cell pellets was suspended in 2 mL 1× PBS, 0.1% triton ×100, 1 μg/mL DAPI. Cell fluorescence of about 50,000 cells was measured using a flow cytometer with excitation at 340 nm and detection at 380 nm. Five experimental replicates were measured per sample. Because blood samples were harvested at different times and DS measurements of the red blood cells samples with this technique did not show significant variations within 4 days after sampling, the DS of red blood cells from a single AIL strain female (Table S1) was calibrated to RJF as a reference. The blood of this AIL female was thereafter used as a reference to measure DSs of blood samples over time. Cytometery data were analysed using a FlowJo appartus to locate peak fluorescence within each blood sample. To verify that all individual mean DS had a modal distribution, we performed a histogram analysis and tested for a normal distribution using a d'Agostino & Pearson omnibus normality test as recommended in the *Prism 6 package* (*GraphPad Software Inc*). Differences in mean DS between individual chicken species, breeds, or lines and the red jungle fowl or the Houdan lines (first sample set in Table S2) were assayed using *t*-tests as recommended in [22]. To test whether individual DS means in each sample were signficantly smaller than those of RJF individuals we used a one-tailed un*paired* sample *t-test with and without Welch's correction using the Prism 6 package (GraphPad Software Inc) because sample sizes were too small for F-tests on variance in every sample.* In these tests *a significance threshold of 0.05 was used after Bonferroni correction (*i.e. *taking into account the number of tests done in each comparison). Similar tests were* also performed examining whether the mean DS of individual chicken species, breeds, or lines were significantly larger than those of the Houdan individual. *Other comparisons were similarly performed between White Leghorn lines*. For DS differences between the four chicken lines in the second sample set (Table S2) we used a two-tailed un*paired* sample *t-test with a Welch's correction. Most pairwise comparisons (5 of 6 total) displayed significant differences in variance. Two* lines were considered different below *a significance threshold of 0.05 after Bonferroni correction.*

2.3. Genomic DNA purification of red blood cells

The genomic DNA (gDNA) of ten females, 2 RJF, 2 Araucana, 2 Alsacienne, 2 White leghorn, and 2 grey jungle fowl (Table S1) were purified from red blood cells samples in 2 steps as previously described [23]. Briefly, the first step consisted in a Proteinase K treatment in the presence of 100 mM EDTA and 0.5% sodium dodesyl sulfate (SDS) at pH 8 and 65 °C, followed by a phenol/chloroform deproteinisation. The second step was ultracentrifugation on a cesium chloride gradient. DNA concentration was estimated both by UV spectrophotometry and fluorimetry using a Mithras LB 940 (Berthold Technologies, Wildbad, Germany) using Hoechst 33258 stain [24].

2.4. Quantification of genomic components

2.4.1. Interspersed repeats

Five Illumina libraries were made from the gDNA of 1 female of RJF, Araucana, Alsacienne, White leghorn and grey jungle fowl. Average gDNA fragment length of these Illumina libraries was 550–600 bp. Sequencing was performed by the I2BC platform (Gif sur Yvette, France). The quality of reads was analysed with FastQC for each dataset. Reads were thereatfer filtered using TrimomaticSE with

parameters HEADCROP:5-10 (depending on the dataset) SLIDINGWINDOW:4:15 MINLEN:80.

Using these Illumina data, two approaches were used in an attempt to find variations in the amount of interspersed repeats between the RJF, grey jungle fowl, Araucana, Alsacienne, and White leghorn breeds. The first was to use RepeatExplorer [25] and dnaPipeTE [26] software with paired-end reads sequenced using Illumina MiSEq technology (2 × 250 nucleotides) at a low coverage (2–3 ×), except for RJF that was further sequenced ~12 × as input. The second approach was to align the reads of each dataset against the galGal5 (chicken) genome using bowtie2 ([27]; default parameters and the preset –very-sensitive), and to count the number of reads for each species of repeats occurring in the RJF genome using the Bioconductor featureCounts program and a repeat annotation [20].

2.4.2. Genes encoding ribosomal RNA 18S-5.8S-28S

Variations in levels of 18S, 5.8S, and 28S ribosomal RNA transcripts between RJF and the Araucana, Alsacienne and White leghorn breeds (Table S1) were determined by qPCR. The gene coding for glyceraldehyde-3-phosphate deshydrogenase (GAPDH) was used as a reference because it is assumed to be present as a single copy per haploid genome in all chicken genomes. Primer pairs were designed using the sequence of the transcribed ribosomal unit (accession number KT445934) and validated for their efficiency to amplify an inner fragment of 220, 108, or 202 bps within the regions coding for the 18S, 5.8S and the 28S rRNA respectively (Table S3). Chicken GAPDH primers have previously been described [28]. qPCRs were performed in triplicate using 5 ng gDNA, primer pairs and the iQ™ SYBR Green Supermix, as recommended by the manufacturer (Bio-Rad Laboratories, Hercules, USA).

2.4.3. Telomeric repeats

Variation in the amount of $(TTAGGG)_n$ telomeric repeats between RJF and the Araucana, Alsacienne and White leghorn breeds were determined by dot blot hybridization. A dilution range from 1 to 128 ng of oligonucleotides corresponding to the plus or minus strands of a $(TTAGGG)_7$ tract (Table S3) and 200 ng of gDNA of each sample were spotted in triplicate on a Zeta-probe blotting membrane (Bio-Rad Laboratories, Hercules, USA). Membranes were dried after spotting and the gDNA was denaturated by incubation on Watmann paper soaked with 0.5 N NaOH, 1.5 M NaCl for 5 min. Samples were neutralized with 1.5 M NaCl, Tris-HCl 0.5 M pH 7.2 followed by drying. DNA was fixed by UV irradiation (280 nm) for 2 min. For hybridization, membranes were incubated overnight under rolling agitation at 53 °C in 10 mL of 0.5 M sodium phosphate buffer pH 7.2, 7% SDS containing 100 pmole of the labeled oligonucleotide Telom-Rep-ATTO700 (Table S3). Membranes were then washed twice in 0.5 × saline-sodium citrate buffer at 55 °C and the hybridization was quantified by near-infrared fluorometry, using an Odyssey Scanner (LiCor). Experiments were done three times with spots for each sample in triplicates.

2.4.4. Tandem repeats located in subtelomeric and centromeric regions

Most subtelomeric regions and centromeric regions, including tandem repeats corresponding to satellite DNA or sequences assembled from ChIP-seq data [29–31], were not integrated into the galGal5 genome assembly even though there were at least 341 sequences from such regions available in public databases (Supplementary data 1). We created a sequences library of these regions. Furthermore, each sequence was duplicated in tandem in order to catch reads overlapping at a junction. This inventory of subtelomeric and centromeric sequences is probably not exhaustive. Nevertheless, we aligned our Illumina reads after quality filtering to this library and quantified mapping occurrences using the Bioconductor featureCounts program [32].

2.5. CNV analyses

Accuracy of CNV detection using high throughput sequences and programs such as CNVnator [33] largely depends on equimolar coverage of chromosomal regions in the sequenced libraries [34]. Accessibility of some bird genome regions to high throughput DNA sequencing technologies is hampered by their high GC content [35,36]. This results in 6 chromosomes that are currently missing in the galGal5 assembly and may partially explain why approximately 1500 genes present in other vertebrates have yet to be found in chickens [37–39]. Keeping these issues in mind, we filtered public datasets of chicken genome resequencing projects as described in 2.4.1 and using a number of quality criteria on the global genome coverage (described below). Most of these datasets were obtained from research programs aimed at studying single nucleotide polymorphisms. The first of our criteria was that the percentage of the filtered reads from a dataset that aligned to galGal5 had to be at least 90% (in order to avoid datasets containing large amounts of contaminant reads (see Table S4c). The second was that the percentage of unique reads, after removal of PCR duplicates in bam files using the program samtools markdup (option -r), should be at least 90%. The third criterion was that the average coverage depth (using the program samtools flastat) should be at least 8 × after read filtering and removal of duplicated reads. The fourth was the genome model coverage by quality filtered (Table S4) and deduplicated reads that was calculated using bedtools and custom perl scripts. Fourty eight public datasets [40–43] and four private datasets of sequences produced by Illumina and SOLID sequencing technologies were screened (Table S4). Fifteen Illumina datasets were found to match with the requested quality criteria and were further investigated.

Prior to further analyses, filtered reads were aligned against the galGal5 assembly (downloaded at the UCSC website at http://hgdownload.cse.ucsc.edu/goldenPath/galGal5/bigZips/) using bowtie2 with the —end-to-end –very-sensitive options. Duplicated reads were then removed using samtools in the bam files. Filtered bam files were used as input with the program CNVnator which is based on the analysis of a normalized read density (RD) per chromosome for CNV identification and takes into account local GC content variations [33]. All calculations were performed using a bin size of 100 bp. Possible CNV locations were filtered using the following criteria: p-value < .01, size > 0.5 kb, and locations with q0 < 0.5 (zero mapping quality, [33]) within the CNV regions removed. Possible CNV sequences overlapping with gaps (tracts of N's inside the chromosome assembly (Table S5) in galGal5 were excluded. Unlocalized and unplaced chromosomes (chrN_random and chrun_random in UCSC, chrUn) were removed from further analysis due to the short length of the chrUn contigs and mapping ambiguity of chrUn sequence reads. Similarly, the W and Z chromosomes were ignored for this analysis because numerous datasets originated from males which lack W chromosomes, and because males and females have a different level of ploidy for the Z chromosome. Taking into account these constraints and the fact that most telomeric and subtelomeric regions are missing in the assembled galGal5 autosomes, we observed that our analyses were done on 0.9 of the ~1.28 Gbp that composes the haploid genome of RJF.

2.6. Gene ontology (GO) analyses

Our analyses were focused mostly on protein coding genes and those encoding long non-coding RNA (lncRNA). We use galGal5 gene annotations from UCSC. Gene ontology was first investigated using DAVID (https://david.ncifcrf.gov/) and AmiGO2 (http://amigo.geneontology.org/amigo) to assess term enrichment. This was followed up by the Cytoscape plugin ClueGO [44,45].

2.7. Statistics

Statistics and graphic representation were done using R (https://

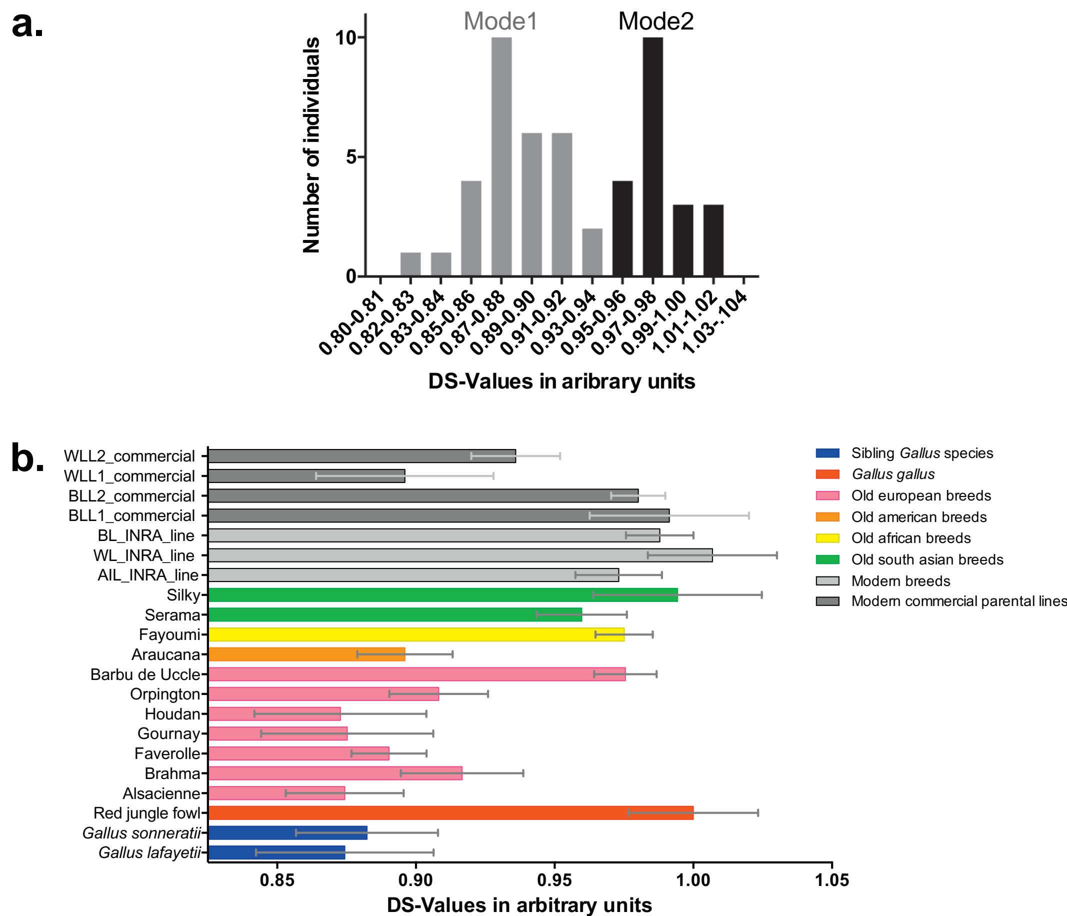


Fig. 1. Variations of DNA content, i.e. DNA size (DS), in red blood cells of females belonging to 3 species of wild fowls (blue and red bars on panel b), eleven old chicken breeds of european, american, african and asian origin (pink, orange, yellow and green bars on panel b, respectively), 3 modern breeds (light grey bars on panel b) and 4 modern commercial lines (dark grey bars). In (a), Distribution of individual DS means among all assayed samples. In (b) Variations of individual DS means in each species, breeds or lines. Bars represent averages with their standard deviation and were calculated from 5, 10, 15 or 20 DS measurements, depending on the number of individuals assayed per species, breed or line (5 measurements per individual). The number of females assayed and the number of replicates per female are supplied in Table S1. Black asterisks indicate statistically significant differences after Bonferroni correction between RJF and all other chicken species, breeds or lines. Asterisks in red indicated statistically significant differences after Bonferroni correction between Houdan all other chicken species, breeds or lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

www.r-project.org/) and Prism 6 (https://www.graphpad.com/scientific-software/prism/).

2.8. Data availability

All raw and processed data are available through the European Nucleotide Archive under Bioproject accessions numbers PRJEB22479 and PRJEB25675.

3. Results

3.1. DNA size in wild fowls and domestic chicken breeds

The genome DNA size (DS) of 3 wild fowl species and 18 chicken breeds were estimated relative to RJF. We first tested whether individual DS means were homogenous. A histogram analysis of DS (Fig. 1a) indicated that their distribution was not modal but was likely at least bimodal. To confirm this observation, we made a d'Agostino & Pearson omnibus normality test and confirmed that the DS distribution was not normal (i.e. not modal). Interestingly, we observed that all individuals from the two sibling species to RJF, all old European and American breeds (excepted the Barbue Uncle) and two commercial lines of layers (WLL1 and WLL2) were gathered in the first mode while all the others, including RJF, were located in the second mode. To further investigate these species, breeds and lines, segregation *t*-tests on small samples were performed as recommended (de Winter 2013). Our statistical analyses (Table S2) confirmed that it was possible to distinguish two groups using individual DS means (Fig. 1, bars with and without black asterisks). In the first group individuals displayed genomes with an average DS that was 5 to 15% smaller than those of RJF individuals. These included individuals from the two sibling species to RJF (Fig. 1, blue bars), four old European and American breeds (Fig. 1, pink and orange bars with a black asterisk at their top), and two commercial lines of layers (Fig. 1, dark grey bars with a black asterisk at their top). Individuals in the other breeds and lines did not show statistically significant differences ($p > .05$). A similar set of comparisons was performed with the smallest genome, the Houdan breed, which confirmed that our RJF individuals had, on average, genomes with larger DS than Houdan (Fig. 1, red bar with a red asterisk). Interestingly, the White Leghorn chicken lines maintained at INRA (WL_INRA_line; Fig. 1b, light grey bar) showed significant differences in DS means from the two commercial lines WLL1 and WLL2 (Table S2). This was further verified using a larger data set of 45–48 individuals from 4 other commercial chicken lines. Our analysis revealed that individuals from one of these four lines (WLL_A) displayed an average DS approximately 3% larger than the DS means of individuals from the three other lines (Table S2).

The analysis of this second sample set also revealed that 5 of 6 pairwise comparisons of the four lines displayed differences in variances (Table S2, F-test results).

Taken together, these results show that there are differences in genome size between RJF and some old breeds, and that genome sizes within some lines, such as those related to the white leghorn breeds, are heterogeneous.

Variations in genome size between sibling animal species are common. For example, the C-value within the *Drosophila* genus varies from 0.14 to 0.38 pg per haploid genome, while in hummingbirds it varies from 0.91 to 1.21 pg, that is 0.8899 to 1183 Gbp (http://www.genomesize.com/index.php, [46]). Examples of intraspecific variation in genome size have been reported for many invertebrate groups including molluscs, crustaceans, insects, as well as among vertebrate groups of fish, reptiles, amphibians and mammals. Fish and amphibian species found in radioactively contaminated or other polluted areas were found to display apparent fluctuations in DNA content (for review, see [47]). However, reports of intraspecific genome size variation should be treated cautiously because this might partially be explained by differences in DAPI-stain efficiency, as described in plants [47]. Using k-mer and similar fluocytometry analyses, it was recently demonstrated that such intraspecific variations occurred in the human and *Arabidopsis* genomes [48]. To further investigate such issues we examined several genome components known to be associated with genome size changes in eukaryotes, namely the amount of interspersed genomic repeats including transposable elements (TEs), the genes coding for ribosomal RNAs 18S-5.8S-28S (rDNA), telomeric (TTAGGG) n repeats, known centromeric and subtelomeric sequences, and segmental copy number variations [47,49,50]. Our hypothesis was that one or several of these components might be the origin of differences observed between RJF and some old breeds and lines.

3.2. Variations in the amount of interspersed or tandem repeats

3.2.1. Interspersed repeats

RepeatExplorer [25,51] and dnaPipeTE [26] are *in-silico* analysis methods that detect and quantify clusters of repetitive DNA sequences in raw genomic reads. These methods are appropriate for investigating variations in the amount of repeated sequences no matter their structure or origin (e.g. satellite DNA, microsatellites, minisatellites, TEs, see [20] for review). We used both of these tools on five low coverage sequence datasets (Table S4b) derived from sequences obtained using either *MiSeq* or Illumina technology. These libraries consisted of one RJF female (MiSeq), grey jungle fowl (*Gallus sonneratii*, Illumina), Araucana (Illumina), Alsacienne (Illumina), and White leghorn (WL_INRA_line, Illumina) breeds. Analyses of these 5 datasets did not reveal any significant variations in repeat content. Similarly, when these 5 datasets were mapped to the galGal5 assembly, no significant differences in the number of reads mapping to interspersed repeat "species" (the 33 TE species found in the RJF genome in [20] were detected between breeds (Table 1). The percent coverage of some repeat species (mainly those on sex chromosomes) were only half as high in our libraries as they were in galGal5 (Table 1, column 3) as would be expected in females (Ancestral_LTR_group_1 and 4). However, other species were equally high (Ancestral_LTR_group_2 and 3) and some cases even higher than in galGal5 (*Z*-rep). This was likely due to the varying quality of chromosome assemblies where the overall presence of repeats was more or less underestimated, depending on the region and the repeat species.

3.2.2. Genes encoding ribosomal RNA 18S-5.8S-28S

In eukaryotic genomes genes encoding the ribosomal RNA 18S-5.8S-28S (rDNA) are organized in one or several blocks located at different chromosomal loci and consist of tandemly repeated units. Each unit is composed of a non-transcribed spacer region that contains at its 3′ end the RNA polymerase I promoter followed immediately downstream by a transcribed region encoding the 18S, 5.8S and 28S rRNAs. In the RJF genome rDNAs are gathered in a single block located on chromosome 16 [52]. However, as in most eukaryotic species, these rDNA units are not included in the assembled galGal5 genome, even as a single copy. Furthermore, the rDNA unit sequence of RJF (~30–40 kbp) has not yet fully been determined, but its transcribed sequence was recently published [53]. The number of tandemly repeated rDNA units per haploid genome have been reported and these vary from 260 to 450 copies depending on the chicken breeds and lines [54]; RJF has 325 copies [55]. These results suggest that at least 0.5 to 2% of the RJF genome consists of rDNA units and might be susceptible to significant variations between RJF and domestic breeds and lines.

Using real-time PCR targeted to the 18S, 5.8S and 28S rRNA (the GAPDH gene was used as a normalization control between samples) we estimated the number of rDNA units in the genomic DNA of two RJF females, as well as two females of the grey jungle fowl, the Araucana, Alsacienne, and the White leghorn (WL_INRA_line). We found that there was approximately half the number of rDNA units in the Alsacienne and Araucana females compared to RJF. The grey jungle fowl and white leghorn females had approximately 75% the number of rRNA units compared to RJF. To verify these ratios, we aligned our low coverage Illumina reads (see Section 3.2.1) as well as fifteen high coverage datasets selected for the CNV analyses (see Section 3.3) to the published transcribed ribosomal unit (NCBI accession number KT445934). These alignments generally confirmed the qPCR results (Fig. 2a and b), with the exception of individual WL7. This may suggest that this trait is polymorphic among lines.

3.2.3. Telomeric (TTAGGG)n repeats

Chicken chromosomes have been estimated to contain approximately 10 times more telomeric (TTAGGG)n repeats than most other vertebrates, covering by some estimates 3%–4% of the genome [19,55–57]. Subtelomeric regions contain multiple classes of larger tandemly repeated units (satellite DNA) as well as genic regions that are difficult to map and sequence [58]. Even though these regions may represent several percents of the whole RJF genome, they are currently not integrated within the galGal5 model. Similar to the rDNA described above, these subtelomeric regions cannot be studied using high throughput sequencing data. Instead, we used dot-plot hybridization of genomic DNA using a $(TTAGGG)_7$ oligonucleotide labeled with an ATTO 700 fluorescent dye, and a $(CCCTAA)_7$ oligonucleotide as a calibration control. We hybridized the genomic DNA of several female individuals: 2 RJF, 2 grey jungle fowl, 2 Araucana, 2 Alsacienne, and 2 White leghorns. We found lower numbers of telomeric (TTAGGG)n repeats in domestic breeds (1.3% in Alsacienne, 1.2% in Araucana, 1.2% in white leghorn compared to 1.9% in RJF) but a slightly higher number in the grey jungle fowl (2.7%).

3.2.4. Overall coverage of subtelomeric and centromeric repeats

To quantify the overall coverage of subtelomeric and centromeric sequences we aligned our low coverage Illumina reads (Section 3.2.1) and fifteen high coverage datasets selected for the CNV analyses (Section 3.3) to our tandem repeat data (a.k.a. satellite DNA, Supplementary data 1). We found that there were on average nearly twice as many subtelomeric and centromeric sequences in RJF compared to domestic breeds and lines, except for WL4, WL7, and WR. This suggested that this trait could be polymorphic at least in modern lines. In RJF individuals 5.5% to 6.5% of high throughput sequencing reads matched these sequences, suggesting that there were at least 5.5% to 6.5% of subtelomeric and centromeric sequences in this species.

3.2.5. Concluding remarks regarding subtelomeric and centromeric repeats

Our analyses showed that there were about half the number rDNA unit copies, (TTAGGG)n repeats and centromeric and subteloemeric sequences in female genomes of old breeds compared to RJF. The same trends were also observed in white leghorn lines but these traits may be

Table 1
Percent coverage of repeat and TE species in galGal5 and 5 chicken breeds based on REPET annotations [20].

Repeat "species"	Type	galGal5*	RJF**	AL**	AR**	GS**	WL**
CR1	LINE	8.8726	8.4658	8.6735	8.6722	8.4107	8.3843
Ancestral_LTR_group_1	LTR#	0.0052	0.0024	0.0024	0.0023	0.0022	0.0023
Ancestral_LTR_group_2	LTR	0.0011	0.0012	0.0012	0.0013	0.0010	0.0010
Ancestral_LTR_group_3	LTR	0.0011	0.0008	0.0009	0.0007	0.0007	0.0008
Ancestral_LTR_group_4	LTR#	0.0112	0.0063	0.0051	0.0062	0.0057	0.0055
BIRDDAWG	LTR	0.2284	0.1878	0.1905	0.1879	0.1811	0.1821
EAV	LTR	0.0169	0.0137	0.0154	0.0164	0.0118	0.0194
EAV-HP	LTR	0.0326	0.0421	0.0359	0.0388	0.0426	0.0375
ERV2	LTR	0.0155	0.0158	0.0166	0.0158	0.0155	0.0152
ERV7	LTR	0.0650	0.0736	0.0682	0.0691	0.0862	0.0695
ERV11	LTR#	0.0132	0.0059	0.0064	0.0062	0.0066	0.0065
Kronos	LTR	0.5329	0.6081	0.6133	0.6098	0.6042	0.5950
putative_LTR_group4	LTR	0.0138	0.0201	0.0200	0.0207	0.0193	0.0196
putative_LTR_group9	LTR&	0.0017	0.0041	0.0043	0.0042	0.0039	0.0039
putative_LTR_group12	LTR	0.0437	0.0506	0.0552	0.0531	0.0523	0.0481
putative_LTR_group22	LTR	0.0257	0.0407	0.0412	0.0423	0.0408	0.0405
putative_LTR_group28	LTR	0.0081	0.0086	0.0075	0.0080	0.0088	0.0081
putative_LTR_group30	LTR&	0.0999	0.1570	0.1601	0.1601	0.1715	0.1556
retroCalimero	LTR	0.0302	0.0263	0.0257	0.0277	0.0252	0.0257
retroSaturnin	LTR	0.0066	0.0092	0.0110	0.0084	0.0096	0.0097
retroTux	LTR	0.0985	0.1043	0.0971	0.0965	0.1122	0.0950
Soprano	LTR	0.1001	0.1000	0.1052	0.0779	0.0956	0.0873
Charlie	TIR	0.2187	0.2259	0.2325	0.2308	0.2230	0.2230
Charlie-Galluhop	TIR	0.4114	0.4589	0.4710	0.4681	0.4550	0.4556
Galluhop	TIR	0.1055	0.1077	0.1091	0.1108	0.1053	0.1076
Mariner1_GG	TIR	0.1303	0.1371	0.1410	0.1407	0.1352	0.1362
Hitchcock	Undefined	0.2139	0.2620	0.2709	0.2699	0.2679	0.2700
undetermined_group_1	Undefined	0.0069	0.0076	0.0070	0.0077	0.0084	0.0077
undetermined_group_2	Undefined	0.0121	0.0146	0.0148	0.0152	0.0164	0.0150
undetermined_group_3	Undefined	0.0038	0.0028	0.0029	0.0027	0.0027	0.0028
undetermined_group_4	Undefined	0.0222	0.0196	0.0197	0.0199	0.0183	0.0188
undetermined_group_5	Undefined	0.0036	0.0030	0.0035	0.0037	0.0043	0.0035
undetermined_group_6	Undefined&	0.0073	0.0148	0.0123	0.0135	0.0163	0.0116
Z_rep	Undefined&	0.1208	0.1537	0.1401	0.1417	0.1653	0.1648
Total		11.4804	11.3521	11.5817	11.5501	11.3257	11.2293

*, coverage of each repeat species in gaGal5 chromosomes. **, coverages were calculated using Illumina reads aligned with Bowtie2 using the "–very-sensitive" option. Read distributions were then counted with featurecounts using gff annotations available at http://chicken-repeats.inra.fr/index.php?. #, coverages that were half as high in our libraries as they were in galGal5. &, coverages that were twiceas high in our libraries as they were in galGal5.

polymorphic because some individuals displayed profiles that were at least partly identical to the RJF. This corresponds to a reduction in genome size of about 5% compared to RJF (~2% rDNA, 2–4% (TTAGGG)n repeats and 6% of centromeric and subtelomeric sequences). These data only partially explain the observed genome size variations between RJF females and Alsacienne and Araucana breeds. However, they may be sufficient to explain differences between RJF and some white leghorn lines. Differences between the grey jungle fowl and RJF cannot be explained using differences in these genomic sequences.

3.3. Extent of variations related to CNVs

CNVs are defined as gains or losses of DNA fragments of 50 bp or longer and are believed to contribute significantly to genome and phenotypic variability. In the chicken genome, chromosomal regions putatively involved with CNVs were previously shown to cover 9 to 10% of the galGal4 genome assembly model [40]. Chromosomal microarray analysis (CMA) and methods using high throughput sequencing (HTS) are the two main approaches used to detect and quantify loci that display copy number variations (CNVs). We decided to use the HTS tool CNVnator because CMAs using array-based comparative genome hybridization (aCGH) or genome hybridization onto single nucleotide polymorphism (SNP) chip have been reported to suffer from several drawbacks, including hybridization noise and low resolution of genome coverage and copy numbers [33,40,59,60]. Our choice of using an HTS tool was also guided by the availability of dozens of whole genome sequencing (WGS) datasets in public databases that looked suitable for our purposes and were not yet analysed in this manner. To be reliable such datasets should be derived from libraries exhibiting good read-depth and uniformity to minimize the detection of false CNVs [60]. Our first step was to identify 48 datasets with a read-depth of over 10×. These were then filtered using 4 stringent criteria (see Section 2.5) to identify those datasets that had a priori the best read-depth uniformity. This step yielded a final 15 datasets (Table S4a) used for analysis as well as stringent CNVnator parameters that should minimize the detection rate of false CNVs [40].

3.3.1. Inventory of CNVs in the 15 selected individual datasets

The 15 selected datasets were representative of our previous investigations regarding variations in genome size because we had fifteen individual genomes sequenced from 3 RJF, 6 individuals from old breeds (3 Araucana (AR) and 3 Korean domestic (KD)), 6 from modern lines (5 white leghorn (WL) and 1 White Plymouth rock (WR) (Table S4a). These datasets were aligned against the galGal5 assembly, a genome model corresponding to that of the red jungle fowl, the common ancestor of all domestic chicken. CNVnator calculated a normalized read density (RD) using a 100-bp window and performed local GC corrections for each chromosome. Because our CNV analyses were performed on diploid genomes and because chromosomal regions outside CNVs displayed a normalized RD of 1 ± 0.25, deletion CNVs could only be detected for heterozygous alleles with an average RD of 0.5 ± 0.25, or for homozygous alleles with a RD ranging from 0 to 0.249 (we were very stringent and only considered as homozygous loci with RD values results below 0.1). CNVs resulting from homozygous deletions could have non-null RD values because multimapping reads to

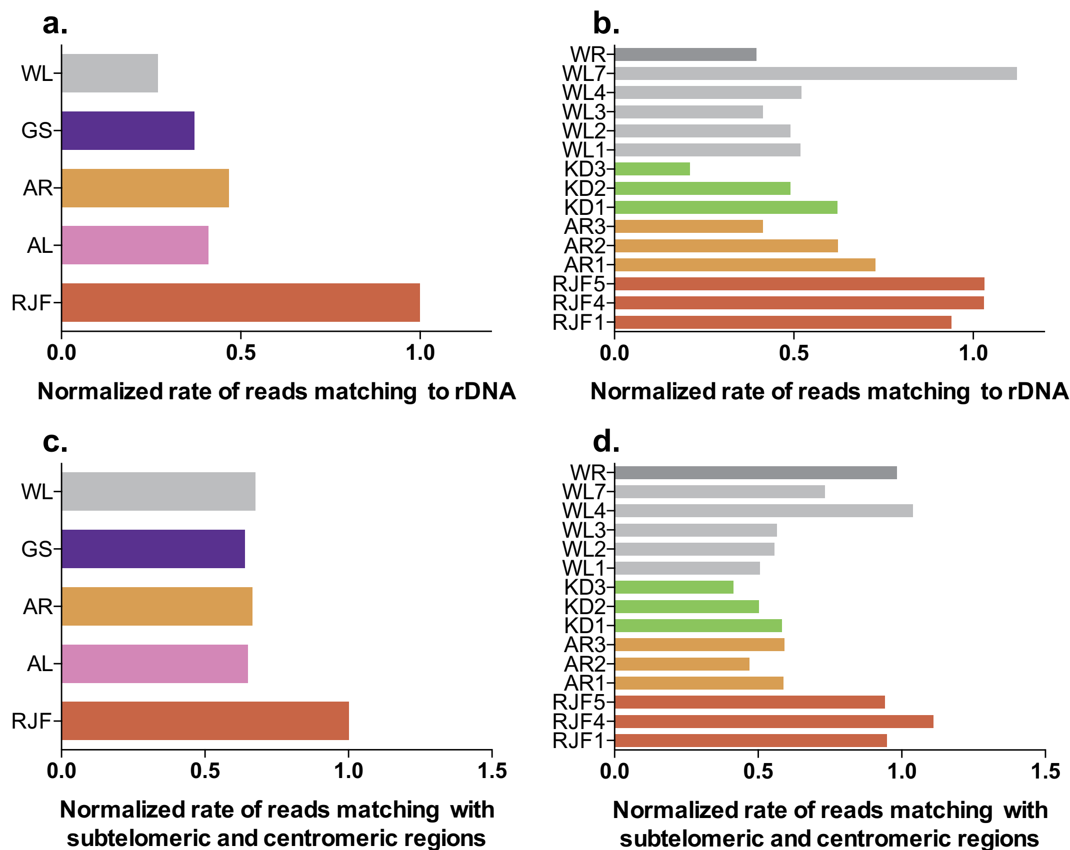


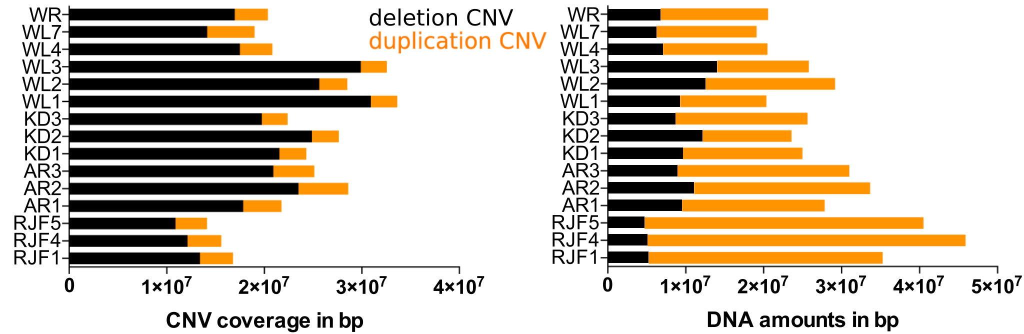
Fig. 2. Relative amounts of rDNA (a and b) and subtelomeric and centromeric sequences (c and d) in individuals of the RJF species and domesticated chicken breeds and lines. Acronyms are described in Table S4. a and c show results obtained with low coverage sequencing. b and d present results obtained with the 15 datasets selected for the CNV study (Section 3.3). The use of colors used to fill bars in histograms was the same as in Fig. 1. Normalized rates were preferred to percentage of reads because the two kinds of sequences studied here are GC-rich and therefore less represented in datasets than sequences with GC content ranging from 35 to 60%. For example, the frequency of reads matching sequences of rDNA units (GC content varying between 70 and 85% along their sequence) displayed an Illumina read frequency ranging from 0.01 to 0.12% depending on the sample while they represented 1 to 2% of the genome.

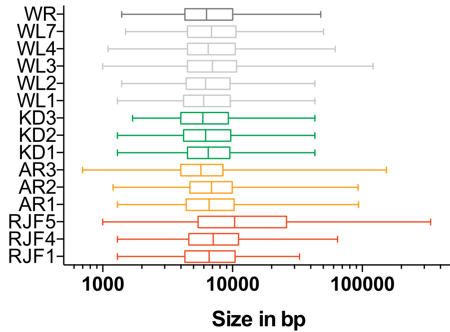
highly conserved sequences could be randomly assigned to one repeat (for further explanations, see bowtie2 and CNVnator manuals). Similarly, CNVs resulting from duplication events can be homozygous (with a RD > 2, depending on the copy number) or heterozygous, their detection being possible from RD values of 1.25 (Table S6).

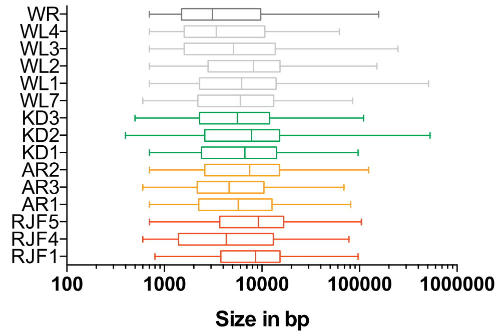
Two CNV sequence profiles could be distinguished for the 15 analysed datasets. The first sequence profile was that of the three RJF individuals. These had between 11 and 14 million bp deletions compared to galGal5 autosomes and between 2 and 3 million bp duplications (Fig. 3a). Concerning the 11–14 million bp deletions between the RJF individuals a significant part of them was shared between the 3 individuals (RJF1, 4 and 7). We interpreted that these deletions might have three origins. The first is related to the fact that the galGal5 genome assembly was built using Illumina and Pacbio reads. This means that regions in the chicken genome that were not accessible by Illumina technology were obtained thanks to the Pacbio technology. For this reason it might not be surprising that at least a part of 11–14 million bp were deleted. A second was that there were copy number variations between individuals, the vast majority of them being heterozygous. Finally, the observed differences might also be due in part to difference between populations (that of the female used for the genomic sequencing project and those of the females used in our study). In terms of DNA amount in their corresponding genome, these CNVs represented between 35 and 45 million bps, ~90% of which were due to deletions and 10% to duplications (Fig. 3b). The second sequence profile was that of domestic chicken breeds, displaying deletions covering between 14 and 30 million bps, duplications spanning 2.5 to 5 million bps in the galGal5 autosomes, but with DNA amounts representing only 20–30 million bps with ~70% coming from deletions and 30% from duplications (Fig. 3a and b). Mann Whitney tests for unpaired data indicated that the numbers of deletion CNVs in RJF and domestic chickens were different ($p < .005$ for both kinds of CNVs) (Fig. 3g). In the first profile, duplication CNVs were twice as important in terms of DNA amount in RJF than in domestic chicken breeds, and coverage of galGal5 autosomes (~3×10^6 bps) was similar. In the second profile, deletion CNVs were on average 2 to 3 times greater in terms of DNA amount in domestic breeds with galGal5 autosome coverages that were 1.2 to 2 fold larger than in RJF. We did not find significant differences in CNV sizes (Fig. 3c and d) or RDs (Fig. 3e and f). However, we did find that there were significantly fewer CNVs corresponding to deletions in RJF than in domestic chicken breeds (Fig. 3g). For the most part, domestic chickens were more often homozygous than RJF (Fig. 3h; $p < .005$, Mann Whitney test for unpaired data, two parameters). Among domestic chickens we also found that individuals from modern lines (WL4, WL7 and WR) displayed significantly more homozygous CNV deletions than other individuals. This agreed with the lower DNA amounts represented by these CNVs in these lines (Fig. 3b).

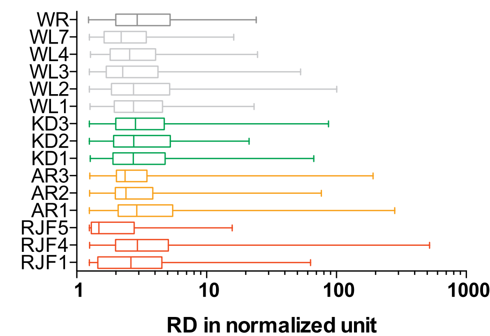
Our results revealed that variations in DNA amounts in regions corresponding to CNVs were significantly different between RJF and domestic chickens. However, these observations were only done on autosomes (~70% of the RJF genome). Considering that CNV coverage in RJF represents 60% of that observed in domestic chickens and DNA amounts in CNVs were on average one third lower in domestic chickens than in RJF, we estimate that at the scale of a complete chicken genome

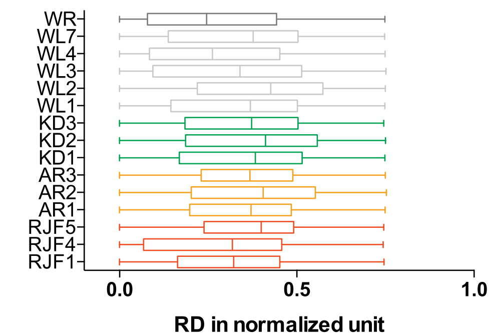
(~1.28 Gbp) differences between RJF and domestic chicken genomes should represent at least 4 to 5% of total DNA content (45 to 55 millions bp). This is similar in size to the amount we previously found for other tandem repeats. In total, duplicated and tandem repeats appear to be responsible for at least 8 to 10% of the differences in genome size between RJF and the domestic chicken breeds and lines. Our results also showed that in modern lines related to WL and WR, significant inter-individual variations of CNVs could be observed in terms of CNV coverage or in terms of the number of homozygous deletions CNVs (WL1-3 versus WL4-7, WR, Fig. 3a and h).

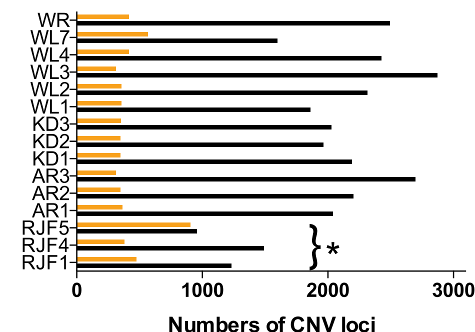


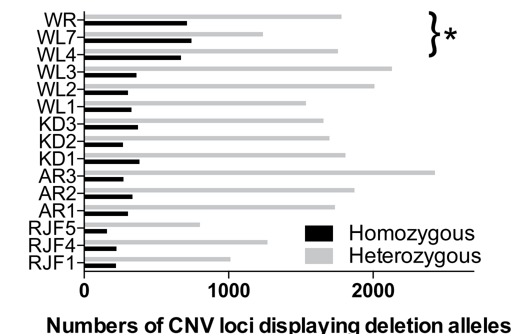












(caption on next page)

Fig. 3. Features of CNVs in the 15 individual chicken datasets analysed. Acronyms are described in Table S4. (a) Histogram of the amount of overlap between CNV sequences and galGal5 chromosomes. (b) Histogram of the amount of CNV sequences in chromosomal sequences for each dataset. In stacked bars, the black sections corresponded to deletions and those in orange to duplications. For each individual, the DNA amounts are calculated from the sum of products between each CNV size and its normalized red density (the RD of each CNV reflecting the number of copy at this locus). In (c), (d), (e) and (f), boxplots with whiskers from minimum to maximum describe the size distributions of CNVs corresponding to duplications and deletions, and of RD in duplications and deletions. Each box represents the quartile 1 and 3, the band inside the box is the median and the whiskers at both ends describe the extent of minimal and maximal values. Boxplots with whiskers are in red for RJF, orange for AR, green for KD, light grey for WL and dark grey for WR. Histogram in (g) describes the numbers of deletions (bars in black) and duplications (bars in orange) in each individual. The parenthesis with the star locates the group of individuals (RJF1, 4 and 5) displaying statistically different numbers of CNV loci corresponding to deletions. Histogram in (h) describes the number of homozygous (bars in black) and heterozygous (bars in grey) deletions in each individual. The parenthesis with the star locates the group of individuals (WL4, WL7 and WR) displaying statistically different numbers of loci homozygous for deletion alleles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Because RJF is the ancestor of domestic breeds and lines, our results supported that domestication led to changes in the genome organization of current domestic chickens by reducing the amount of DNA amount originating from duplication CNVs and increasing the importance of deletion CNVs in domesticated chickens. Below we further examine whether the extent of gene losses and gains were due to CNVs during domestication. We also address the question of whether breeder driven selection might be at the origin of CNVs variations in domesticated chickens. Previous work has shown that CNVs present in certain domestic breed/line have been found to be absent in RJF (e.g. in [61–65]). We therefore investigated the gene content of CNV loci common to RJF individuals (core CNVs) and compared them to those of domesticated chickens. Our goal was to establish whether genes included in CNV loci had been lost or reduced in terms of copy number during domestication, and searched for new CNV loci containing genes that would be common to domesticated chickens and absent in RJF.

3.3.2. Content of CNVs

On the basis of their origin and CNV profiles (Sections 3.2.1 to 3.3.1), samples were divided into five groups: 1) RJF, 2) Araucana (AR), 3) Korean Domestic (KD); the remaining two groups were composed of modern lines 4) WL1, WL2 and WL3 (WL-G1 that gathered individuals from a single line (Table S7)) and 5) WL4, WL7 and WR (WL-G2 that gathered individuals from three lines (Table S7)). Hereafter, we will refer to the last four groups as the "domesticated group" since they are all derived from domesticated sources, as opposed to the RJF. The presence of deletion and duplication CNV loci was investigated separately. Within the groups defined above, between 50 and 75% of both deletion and duplication CNV loci were shared within groups (Tables 2 and 3, values in columns 2 and 3). This suggests that the presence of CNV loci was variable between individuals as well as between breeds and lines. The significance of such variations is difficult to interpret. Indeed, several studies published in the last decade using high throughput sequencing and single cell approaches in mammals have shown that CNVs are dynamic sequences, showing both gains and losses due to recombination in somatic cell genomes and to a lower degree in germ cells (e.g. in [66–73]). The absence of recombination data in chickens, particularly with regard to the intensity and the impact of somatic recombination in different chicken cell types, made it impossible to identify the origin of CNV loci. Indeed, it could not be determined if a CNV locus observed between two individuals originated from the germline, was the consequence of somatic recombination, or both. Therefore we limited our investigations to CNVs that were found in common within each of the five groups.

3.3.2.1. Features of deletion and duplication CNVs. Analysis of core deletion CNV loci (i.e. deletions at loci that did not vary between RJF samples) revealed that 385 deletion loci (including 37 homozygous deletions) were shared between RJF and the 4 domesticated groups (Table 2, columns 3 and 6). Between the 4 domesticated groups there were 614 shared deletion CNVs at the core loci, including 45 that were homozygous loci. When these 614 CNV loci were filtered by intersecting them with deletion CNV loci of each RJF individual, we found that only 29 core deletion CNV loci were not found in our RJF samples, indicating that there were non fixed alleles for most deletion CNVs in RJF. These observations suggested that domestication and selection by breeders might have resulted in the development of only a few new of deletions CNV loci and the fixation, in most cases, of deletions CNV alleles that occurred in RJF populations. We also found that among the 673 RJF core deletion CNV loci, only 2 were not

Table 2
Numbers of CNVs in galGal5 autosomes corresponding to deletions between individuals, or groups of individuals gathered by species, breed or line.

1. Individual	2. N° of deletion CNVloci in individuals*	3. Core deletion CNV loci in each breed or line *,**	4. N° deletion CNVs shared by RJF core CNV loci with each individuals*	5. N° deletion CNV loci shared by RJF core CNV loci with each breed or line*	6. N° of core deletion CNV loci shared by 5 groups, 4 domestic groups or own to RJF*,***
RJF1	1233 (221)	673 (83)	673 (83)	673 (83)	5 groups: 385 (37)
RJF4	1491 (223)		673 (83)		
RJF5	958 (156)		673 (83)		
AR1	2041 (303)	1438 (169)	662 (80)	649 (76)	4 domestic groups: 614 (45)
AR2	2205 (334)		659 (80)		
AR3	2698 (271)		670 (82)		
KD1	2191 (383)	1039 (111)	667 (80)	499 (53)	Own to RJF: 2 (1)
KD2	1964 (267)		499 (49)		
KD3	2303 (372)		659 (75)		
WL1	1863 (326)	1147 (133)	522 (58)	500 (55)	
WL2	2315 (305)		664 (59)		
WL3	2872 (741)		659 (77)		
WL4	2426 (671)	1109 (158)	665 (78)	650 (73)	
WL7	1599 (362)		659 (77)		
WR	2494 (713)		663 (80)		

We have intersected calls that are > 500 bp (these events are detected at maximum sensitivity) from the whole galGal5 models excluding the W and Z chromosomes, and consider two calls concordant if they have > 50% reciprocal overlap. *, Values in parentheses correspond to the number of deletions corresponding to homozygous deletions (i.e. RD < 0.1) compared to the galGal5 autosome models; **, N° of deletion CNVs shared by individuals of each species, breed or line; ***, values were calculated using the N° of core loci in column 3.

Table 3
Numbers of CNVs in galGal5 autosomes corresponding to duplications between individuals, or groups of individuals gathered by species, breed or line.

1. Individual	2. N° of duplication CNV loci in individuals	3. Core duplication CNV loci in each breed or line*,**	4. N° duplication CNVs shared between RJF core CNV loci and individual**	5. N° duplication CNV loci shared by core CNV loci of RJF with each breed or line**	6. N° of core duplication CNV loci by 5 groups, 4 domestic groups or own to RJF*,***
RJF1	475	216	216	216	5 groups: 107
RJF4	380		216		
RJF5	907		216		
AR1	363	200	183	150	4 domestic groups: 120
AR2	347		161		
AR3	314		173		Own to RJF: 22
KD1	348	223	177	155	
KD2	351		184		
KD3	354		181		
WL1	358	221	185	143	
WL2	358		185		
WL3	312		157		
WL4	417	255	191	171	
WL7	566		189		
WR	416		195		

We have intersected calls that are > 500 bp (these events are detected at maximum sensitivity) from the whole galGal5 models excluding the W and Z chromosomes, and consider two calls concordant if they have > 50% reciprocal overlap. *, N° of deletion CNVs shared by individuals of each species, breed or line; ***, values were calculated using the N° of core loci in column 3.

recovered in the 4 groups of domestic chickens, and these two CNV loci did not contain genes.

These analyses revealed that there were a limited numbers of core duplication CNV loci between the five groups, 107 (Table 3, column 6). We found that there were only 120 core duplication CNV loci shared by the domesticated group. Among the 216 core duplication CNV loci in RJF, 22 were found to be specific to this group. It was not possible to determine how important these CNV loci were to RJF physiology because they overlapped 22 genes of unknown function. These results further supported our previous conclusion that there were both losses and gains of duplication CNV loci during chicken domestication.

Analysis of gene content of both deletion and duplication CNV loci in the five groups defined above revealed that 50 to 65% of these loci contained genes (Table S7 and S8), and some of them were shared between groups (Fig. 4). These included protein-coding genes (~75%), ncRNA-coding genes (~22%) and encoded other RNA product genes (~3%; tRNA-coding, pseudogenes and miscellaneous RNA; Table S7 and S8). These observations supported that there were both losses and gains of gene copy number in deletion and duplication CNV loci during chicken domestication since there were respectively more genes in the four groups of domesticated individuals (Table S9a and b, third line: (515 + 112 + 19 + 95 + 31 + 1 = 773) genes) than in RJF genes, and more genes among the CNVs specific to domesticated chickens than in the five groups (Table S9a and b, second line: 391 + 107 + 26 + 168 + 46 + 8 = 746) genes). This also raised the issue of which physiological pathways could have been affected by these changes in copy number during the evolution of domestic chickens. Previous studies have supported that recombination-based processes played a major role in avian genome evolution [74], including during chicken domestication.

3.3.2.2. Definition of GO pathways and processes related to CNVs during domestication. Biological processes and pathways were studied using the ClueGO plug-in in Cytoscape, with the 2018 *Gallus gallus* GO databases from QuickGO (biological process and immune system process), KEGG, and Wikipathways. We used as parameters a default Network specificity (medium) with a $p < .05$ to filter significantly enriched term-clusters that were corrected by the Benjamini-Hochberg method.

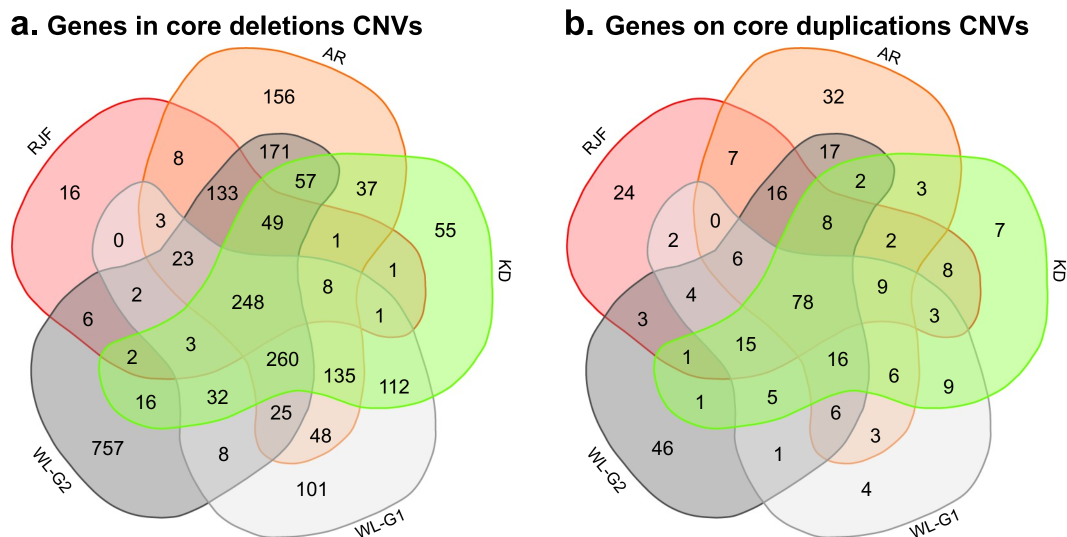


Fig. 4. Venn diagrams representing the gene contents in core CNVs shared or specific to each of the five groups of chicken individuals. For core deletion CNVs (a), the total number of genes was 524 in RJF, 1557 in AR, 1106 in KD, 1135 in WL-G1, and 1979 in WL-G2. For core duplication CNVs (b), the total number of genes was 222 in RJF, 233 in AR, 202 in KD, 199 in WL-G1, and 251 in WL-G2.

Fig. 5. Graph in sections of enriched terms of genes located in core deletion CNVs from RJF (a), AR (b), KD (c), WL-G1 (d) and WL-G2 (e). The size of each sector describes the enrichment and finally depends on the number of involved terms. Sectors describing enriched terms were coloured when they were shared by several groups of individuals. Sectors corresponding to unique enriched terms were filled with different shades of grey. **,* or the absence of * respectively indicated a significance of the cluster of enriched genes < 0.001, 0.01 and 0.05.

The GO analysis of genes contained in core deletion CNVs (Fig. 5) revealed term enrichments that were dramatically different between RJF and the 4 groups of domesticated chickens. This was not unexpected since we had previously observed that the CNV profiles were different between RJF and the 4 domesticated groups. In RJF (Fig. 5a) these conclusions were consistent with 7 pathways involving 22 genes, 5 of them connected (Oocyte meiosis, Apeling signalling pathways, AGE-RAGE signalling pathway in diabetic complications, Glutamate receptor signalling pathway and Lysosomal transport) and the other two independent (embryonic skeletal system development and homologous recombination). Profiles of term-enrichment in the 4 groups of domesticated chicken (Fig. 5b–e) were different from each other but they also displayed some commonalities. Two terms gathered several hundreds of genes under "regulation of the transcription" and "anterior/posterior specification pattern". Others shared terms concerned the development of epithelial cells, the immune system and the functioning of the nervous system through 3 main terms: synaptic vesicle maturation, brain development and the enteric nervous system. Overall, GO analysis of gene content in deletion CNV loci suggested that chicken domestication has favored the fixation or the appearance of deletions in genes affecting these processes and pathways.

GO analyses of core duplication CNVs was less fruitful, likely because these CNVs were less numerous and therefore contained fewer genes. Nevertheless, we found two significant term enrichment ($p < .001$) that were both found in old breed groups Araucana (10 genes) and the Korean domesticated group (4 genes). In the Araucana group, enriched terms included one innate defense mechanism related to a Toll-like signalling pathway directed against microbial and viral cytosolic DNA. One can hypothesize that the tandem amplification of genes involved in this pathway could confer a higher resistance to these kinds of infections. In the Korean domestics, the enriched terms involved cytoskeleton fragment functions.

In all, the analysis of the CNV profiles in the five groups of individuals suggested that the CNV organization in chicken breeds and lines was widely reshaped during their evolution through domestication first, and by breeder selection following that.

4. Conclusions

Our cytometry results showed, for the first time, that differences in genome size could be found between RJF, old breeds, and modern lines. Having established such individual level differences (individual level differences have previously been established in well studied populations such as humans and thale cress [48]) future studies could now examine how these differences are distributed at a population level in *Gallus* species, breeds and lines. Indeed, the sample size for each breed and line used here is limited so larger samples would be required to get more information. Nevertheless, sequence analysis confirmed that these differences were not due to DNA staining problems during sample preparation, but to several types of tandemly repeated sequences. We estimate that these duplicated sequences might explain differences in genome size between 8 and 10%, at the same time our cytometry results revealed differences ranging from 5 to 15%. Given measurement uncertainties of the genome size using cytometry, both of these measurements were similar to one another. Furthermore, our preliminary investigations of CNV amounts in the W chromosome indicated that it was significantly more elevated than in autosomes, suggesting that using only autosomes somewhat underestimated the importance of these sequences in complete chicken genomes. This will need to be further investigated as the accuracy of the Galgal model genome improves, reducing the current uncertainty in repeated regions.

Previous work regarding the profile of single nucleotide polymorphisms in domestic chickens concluded that they displayed significantly reduced genetic variation, an increased proportion of nonsynonymous amino acid changes, and a tendency to harbor higher rates of changes classified as damaging with respect to their RJF ancestor [17]. These data supported the conclusion that domestic animals were genetically impoverished animals. This view is very different from that of neuro-behaviourists working on avian cognition. For these researchers domestic chickens are animals adapted to an anthropomorphised environment and have lost their ability to survive in the wild [14]. Our CNV loci data showing that losses of DNA segments occasionally lead to a decrease in gene copy number, supported that domestication could have led to changes in the number of housekeeping genes in domestic chickens. This was evidenced by fragments deleted heterozygously or homozygously in domesticated chickens, but being present and homozygous in all RJF individuals. These pathways and processes mainly involved transcriptional regulation of biosynthetic processes, anterior/posterior pattern specification, immune system, and brain development (likely in its synaptogenesis; Fig. 5, terms in light green and light blue). This raises questions regarding the interpretation of results obtained from cognition tests aimed at comparing the cognitive capacities of domestic chicken with RJF, the later having been recently shown to have at least their behaviour epigenetically modified after a few generations in captivity [75].

Other kinds of tandem repeat in RJF and domestic chickens brought complementary information. It was initially difficult to evaluate whether the lower number of rDNA units in domestic chickens might have an effect on their phenotype. Indeed, in *Drosophila melanogaster* mutant flies with at least a 2-fold reduction in the number of rDNA copies displayed deleterious pleiotropic phenotypes [76]. However, the difference observed between our limited number of RJF and domestic individuals might not have phenotypic effects since phenotypic changes related to the number of rDNA units only occurred under a certain reduction threshold that remains undefined in chickens. The reduction in length of telomeric regions in domestic chickens was previously described as being an inherited trait in breeds and lines and was related to the longevity in avian species [77]. Long telomeres are thought to provide a buffer against cellular senescence and be generally indicative of genome stability and overall cell health [78]. In chicken somatic cells telomere length was also shown to depend on the environmental and oxidative stressors met by each individual [77,79]. Here, because our investigations were done on red blood cells, it was not possible to determine whether the observations resulted from a lower ability of domestic chicken to maintain the length of their telomeres under husbandry conditions or were an inherited trait resulting from domestication that would be transmitted via the germ cells. Our data also suggested that the content of repeats of subtelomeric and centromeric regions were reduced in domestic chickens. In mammals, size alteration of subtelomeric regions is thought to cause disease due to a position effect that influences the transcription of nearby genes, rather than through the loss of the repeat array itself [80]. These regions were also shown to weaken the innate immune defense [81,82].

Although the analysis of a wider number of individual genomes will be required to further verify the hypotheses and observations presented here, it appears that tandemly duplicated sequences have decreased in domesticated chickens and might have thus reduced the ability of these animals to adapt to varied environments and/or to resist disease. The differences in tandemly duplicated sequences in modern lines might explain why old breeds may have better innate immunity defenses than modern lines. The idea that domestication led to degenerated animals rather than to more adapted ones was used for eugenic purposes by its original author Konrad Lorentz [83]. Nevertheless, this idea cannot be ignored in the context of understanding genomic processes that occurred during domestication, because these processes aim to lower the frequency of undesirable variants in chicken breeds and lines.

Acknowledgements

This work was funded by the Project Région Centre AviGeS, the C.R.B. Anim, the C.N.R.S., the I.N.R.A., the Groupement de Recherche CNRS 2157, and the Ministère de l'Education Nationale, de la

Recherche et de la Technologie. Peter Arensburger holds a senior researcher fellowship from the STUDIUM. The authors thank the staff of the breeding facilities (INRA, Pôle d'Expérimentation Avicole de Tours, Nouzilly, France). We also thank Christophe Bec (the “Parc des oiseaux”, Villars les Dombes, France), Christophe Auzou (Domaine de la Puisaye, Grandchamps, France), the staff of the “Centre de Sélection de Béchanne” (Saint-Étienne-du-Bois, France), and Dr. Pieter Van As (Hendrix Genetics Corporate, Boxmeer, Netherlands) for their friendly welcome and the supply and transport of samples toward our laboratory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2019.10.004.

References

[1] B. West, B.X. Zhoub, Did chickens go North? New evidence for domestication, J. Archaeol. Sci. 15 (1988) 515–533.
[2] M. Tixier-Boichard, B. Bed'hom, X. Rognon, Chicken domestication: from archeology to genomics, Comptes Rendus Acad. Sci. Biol. 334 (2011) 197–204, https://doi.org/10.1016/j.crvi.2010.12.012.
[3] H. Sawai, H.L. Kim, K. Kuno, et al., The origin and genetic variation of domestic chickens with special reference to junglefowls Gallus g. gallus and G. varius, PLoS One 5 (2010) e10639, , https://doi.org/10.1371/journal.pone.0010639.
[4] Y.P. Liu, G.S. Wu, Y.G. Yao, et al., Multiple maternal origins of chickens: out of the Asian jungles, Mol. Phylogenet. Evol. 38 (2006) 12–19, https://doi.org/10.1016/j.ympev.2005.09.014.
[5] M. Nishibori, T. Shimogiri, T. Hayashi, et al., Molecular evidence for hybridization of species in the genus Gallus except for Gallus varius, Anim. Genet. 36 (2005) 367–375, https://doi.org/10.1111/j.1365-2052.2005.01318.x.
[6] J. Eriksson, G. Larson, U. Gunnarsson, B. Bed'hom, M. Tixier-Boichard, L. Strömstedt, D. Wright, A. Jungerius, A. Vereijken, E. Randi, P. Jensen, L. Andersson, Identification of the yellow skin gene reveals a hybrid origin of the domestic chicken, PLoS Genet. 4 (2008) e1000010, , https://doi.org/10.1038/srep35818.
[7] T. Downing, A.T. Lloyd, C. O'Farrelly, et al., The differential evolutionary dynamics of avian cytokine and TLR gene classes, J. Immunol. 184 (2010) 6993–7000.
[8] E.O. Price, Behavioral aspects of animal domestication, Q. Rev. Biol. 59 (1984) 1–32, https://doi.org/10.1086/413673.
[9] B. Agnvall, J. Bélteky, P. Jensen, Brain size is reduced by selection for tameness in red Junglefowl- correlated effects in vital organs, Sci. Rep. 7 (3306) (2017), https://doi.org/10.1038/s41598-017-03236-4.
[10] J. Bélteky, B. Agnvall, M. Johnsson, et al., Domestication and tameness: brain gene expression in red junglefowl selected for less fear of humans suggests effects on reproduction and immunology, R. Soc. Open Sci. 3 (2016) 160033, , https://doi.org/10.1098/rsos.160033.
[11] M. Ericsson, P. Jensen, Domestication and ontogeny effects on the stress response in young chickens (*Gallus gallus*), Sci. Rep. 6 (35818) (2016), https://doi.org/10.1038/srep35818.
[12] M. Johsson, Changes in behavior and emotion under chicken domestication, Animal. Sentience 17 (10) (2017), https://animalstudiesrepository.org/animsent/vol2/iss17/10/.
[13] P. Løtvedt, A. Fallahshahroudi, L. Bektic, et al., Chicken domestication changes expression of stress-related genes in brain, pituitary and adrenals, Neurobiol. Stress. 7 (2017) 113–121, https://doi.org/10.1016/j.ynstr.2017.08.002.
[14] L. Marino, Thinking chickens: a review of cognition, emotion, and behavior in the domestic chicken, Anim. Cogn. 20 (2017) 127–147, https://doi.org/10.1007/s10071-016-1064-4.
[15] A.F. Soleimani, I. Zulkifli, A.R. Omar, et al., Physiological responses of 3 chicken breeds to acute heat stress, Poult. Sci. 90 (2011) 1435–1440, https://doi.org/10.3382/ps.2011-01381.
[16] C.J. Rubin, M.C. Zody, J. Eriksson, et al., Whole-genome resequencing reveals loci under selection during chicken domestication, Nature. 464 (2010) 587–591, https://doi.org/10.1038/nature08832.
[17] T. Makino, C.J. Rubin, M. Carneiro, et al., Elevated proportions of deleterious genetic variation in domestic animals and plants, Gen. Biol. Evol. 10 (2018) 276–290, https://doi.org/10.1093/gbe/evy004.
[18] I. Solovei, A. Ogawa, M. Naito, S. Mizuno, H. Macgregor, Specific chromomeres on the chicken W lampbrush chromosome contain specific repetitive DNA sequence families, Chromosom. Res. 6 (1998) 323–327.
[19] M.E. Delany, L.M. Daniels, S.E. Swanberg, H.A. Taylor, Telomeres in the chicken: genome stability and chromosome ends, Poult. Sci. 82 (2003) 917–926.
[20] S. Guizard, B. Piégu, P. Arensburger, et al., Deep landscape update of dispersed and tandem repeats in the genome model of the red jungle fowl, *Gallus gallus*, using a series of de novo investigating tools, BMC Genomics 17 (659) (2016), https://doi.org/10.1186/s12864-016-3015-5.
[21] Z. Darzynkiewicz, G. Juan, DNA content measurement for DNA ploidy and cell cycle analysis, Curr. Protoc. Cytom. Unit 7 (2001) 5, https://doi.org/10.1002/0471142956.cy0705s00 Chapter 7.
[22] J.C.F. de Winter, Using the Student's *t*-test with extremely small sample sizes, Pract. Assess. Res. Eval. 18 (10) (2013), http://pareonline.net/getvn.asp?v=18&n=10.
[23] Y. Bigot, G. Jegot, S. Casteret, P. Aupinel, J.N. Tasei, DNA modifications and genome rearrangements during the development and sex differentiation of the bumble bee Bombus terrestris, Insect Mol. Biol. 20 (2011) 165–175, https://doi.org/10.1073/pnas.1518008112.
[24] K. Rengarajan, S.M. Cristol, M. Mehta, et al., Quantifying DNA concentrations using fluorometry: a comparison of fluorophores, Mol. Vis. 8 (2002) 416–421.
[25] P. Novák, P. Neumann, J. Macas, Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data, BMC Bioinformatics. 11 (2010) 378, https://doi.org/10.1186/1471-2105-11-378.
[26] C. Goubert, L. Modolo, C. Vieira, et al., De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*), Genome Biol. Evol. 7 (2015) 1192–1205, https://doi.org/10.1093/gbe/evv050.
[27] B. Langmead, S. Salzbergn, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.
[28] D.M. Cairns, R.J. Pignolo, T. Uchimura, et al., Somitic disruption of GNAS in chick embryos mimics progressive osseous heteroplasia, J. Clin. Invest. 123 (2013) 3624–3633, https://doi.org/10.1172/JCI69746.
[29] W.H. Shang, T. Hori, A. Toyoda, et al., Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences, Genome Res. 20 (2010) 1219–1228, https://doi.org/10.1101/gr.106245.110.
[30] W.H. Shang, T. Hori, N.M. Martins, et al., Chromosome engineering allows the efficient isolation of vertebrate neocentromeres, Dev. Cell 24 (2013) 635–648, https://doi.org/10.1016/j.devcel.2013.02.009.
[31] T. Hori, W.H. Shang, A. Toyoda, et al., Histone H4 Lys 20 monomethylation of the CENP-A nucleosome is essential for kinetochore assembly, Dev. Cell 29 (2014) 740–749, https://doi.org/10.1016/j.devcel.2014.05.001.
[32] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics. 30 (2014) 923–930, https://doi.org/10.1093/bioinformatics/btt656.
[33] A. Abyzov, A.E. Urban, M. Snyder, et al., CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing, Genet. Res. 21 (2011) 974–984, https://doi.org/10.1101/gr.114876.110.
[34] S. van Heesch, M. Mokry, V. Boskova, et al., Systematic biases in DNA copy number originate from isolation procedures, Genome Biol. 14 (2013) R33, https://doi.org/10.1186/gb-2013-14-4-r33.
[35] M.G. Ross, C. Russ, M. Costello, et al., Characterizing and measuring bias in sequence data, Genome Biol. 14 (2013) R51, https://doi.org/10.1186/gb-2013-14-5-r51.
[36] M.K. Tilak, F. Botero-Castro, N. Galtier, et al., Illumina library preparation for sequencing the GC-rich fraction of heterogeneous genomic DNA, Gen. Biol. Evol. 10 (2018) 616–622, https://doi.org/10.1093/gbe/evy022.
[37] H. Farkašová, T. Hron, J. Pačes, et al., Identification of a GC-rich leptin gene in chicken, Agri. Gene. 1 (2016) 88–92, https://doi.org/10.1210/en.2015-1634.
[38] T. Hron, P. Pajer, J. Pačes, et al., Hidden genes in birds, Genome Biol. 16 (2015) 164, https://doi.org/10.1186/s13059-015-0724-z.
[39] S. Bornelöv, E. Seroussi, S. Yosefi, et al., Correspondence on Lovell et al.: identification of chicken genes previously assumed to be evolutionarily lost, Genome Biol. 18 (2017) 112, https://doi.org/10.1186/s13059-017-1231-1.
[40] G. Yi, L. Qu, J. Liu, et al., Genome-wide patterns of copy number variation in the diversified chicken genomes using next-generation sequencing, BMC Genomics 15 (2014) 962, https://doi.org/10.1186/1471-2164-15-962.
[41] E. Gering, M. Johnsson, P. Willis, et al., Mixed ancestry and admixture in Kauai's feral chickens: invasion of domestic genes into ancient red Junglefowl reservoirs, Mol. Ecol. 24 (2015) 2112–2124, https://doi.org/10.1111/mec.13096.
[42] D. Oh, B. Son, S. Mun, et al., Whole genome re-sequencing of three domesticated chicken breeds, Zool. Sci. 33 (2016) 73–77, https://doi.org/10.2108/zs150071.
[43] M. Ulfah, R. Kawahara-Miki, A. Farajalllah, et al., Genetic features of red and green junglefowls and relationship with Indonesian native chickens Sumatera and Kedu Hitam, BMC Genomics 17 (320) (2016), https://doi.org/10.1186/s12864-016-2652-z.
[44] G. Bindea, B. Mlecnik, H. Hackl, et al., ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks, Bioinformatics. 25 (2009) 1091–1093, https://doi.org/10.1093/bioinformatics/btp101.
[45] B. Mlecnik, J. Galon, G. Bindea, Comprehensive functional analysis of large lists of genes and proteins, J. Proteome 171 (2018) 2–10, https://doi.org/10.1016/j.jprot.2017.03.016.
[46] T.R. Gregory, C.B. Andrews, J.A. McGuire, et al., The smallest avian genomes are found in hummingbirds, Proc. Biol. Sci. 276 (2009) 3753–3757, https://doi.org/10.1098/rspb.2009.1004.
[47] T.R. Gregory, The C-value enigma in plants and animals: a review of parallels and an appeal for partnership, Ann. Bot. 95 (2005) 133–146, https://doi.org/10.1093/aob/mci009.
[48] H. Sun, J. Ding, M. Piednoël, K. Schneeberger, findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies, Bioinformatics. 34 (2018) 550–557, https://doi.org/10.1093/bioinformatics/btx637.
[49] K.K. Wong, R.J. deLeeuw, N.S. Dosanjh, et al., A comprehensive analysis of common copy-number variations in the human genome, Am. J. Hum. Genet. 80 (2007) 91–104, https://doi.org/10.1086/510560.
[50] T.A. Elliott, T.R. Gregory, What's in a genome? The C-value enigma and the

evolution of eukaryotic genome content, Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 370 (2015) 20140331, , https://doi.org/10.1098/rstb.2014.0331.

[51] P. Novák, P. Neumann, J. Pech, et al., RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads, Bioinformatics. 29 (2013) 792–793, https://doi.org/10.1093/bioinformatics/btt054.

[52] M.M. Miller, C.M. Robinson, J. Abernathy, et al., Mapping genes to chicken microchromosome 16 and discovery of olfactory and scavenger receptor genes near the major histocompatibility complex, J. Hered. 105 (2014) 203–215, https://doi.org/10.1093/jhered/est091.

[53] A.G. Dyomin, E.I. Koshel, A.M. Kiselev, A.F. Saifitdinova, S.A. Galkina, T. Fukagawa, A.A. Kostareva, E.R. Gaginskaya, Chicken rRNA gene cluster structure, PLoS One 11 (2016) e0157464, , https://doi.org/10.4049/jimmunol.0903092.

[54] M.H. Su, M.E. Delany, Ribosomal RNA gene copy number and nucleolar-size polymorphisms within and among chicken lines selected for enhanced growth, Poult. Sci. 77 (1998) 1748–1754, https://doi.org/10.1093/ps/77.12.1748.

[55] M.E. Delany, A.B. Krupkin, Molecular characterization of ribosomal gene variation within and among NORs segregating in specialized populations of chicken, Genome. 42 (1999) 60–71.

[56] M.E. Delany, A.B. Krupkin, M.M. Miller, Organization of telomere sequences in birds: evidence for arrays of extreme length and for in vivo shortening, Cytogenet. Cell Genet. 90 (2000) 139–145.

[57] I. Nanda, D. Schrama, W. Feichtinger, et al., Distribution of telomeric (TTAGGG)(n) sequences in avian chromosomes, Chromosoma. 2002 (111) (2002) 215–227, https://doi.org/10.1007/s00412-002-0206-4.

[58] E. Seroussi, F. Pitel, S. Leroux, et al., M. Friedman-Einat, Mapping of leptin and its syntenic genes to chicken chromosome 1p, BMC Genet. 18 (2017) 77, https://doi.org/10.1186/s12863-017-0543-1.

[59] M. Zhao, Q. Wang, Q. Wang, et al., Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives, BMC Bioinformatics 14 (S11) (2013) S1, https://doi.org/10.1186/1471-2105-14-S11-S1.

[60] B. Trost, S. Walker, Z. Wang, et al., A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data, Am. J. Hum. Genet. 102 (2018) 142–155, https://doi.org/10.1016/j.ajhg.2017.12.007.

[61] D. Wright, H. Boije, J.R. Meadows, et al., Copy number variation in intron 1 of SOX5 causes the Pea-comb phenotype in chickens, PLoS Genet. 5 (2009) e1000512, , https://doi.org/10.1371/journal.pgen.1000512.

[62] X. Wang, S. Byers, Copy number variation in chickens: a review and future prospects, Microarrays. 3 (2014) 24–38, https://doi.org/10.3390/microarrays3010024.

[63] C. Moro, R. Cornette, A. Vieaud, et al., Quantitative effect of a CNV on a morphological trait in chickens, PLoS One 10 (2015) e0118706, , https://doi.org/10.1371/journal.pone.0118706.

[64] M.O. Lee, S. Bornelöv, L. Andersson, S.J. Lamont, J. Chen, J.E. Womack, Duplication of chicken defensin7 gene generated by gene conversion and homologous recombination, Proc. Natl. Acad. Sci. U. S. A. 113 (2016) 13815–13820, https://doi.org/10.1073/pnas.1616948113.

[65] S. Lin, X. Lin, Z. Zhang, et al., Copy number variation in SOX6 contributes to chicken muscle development, Genes (Basel) 9 (2018) E42, https://doi.org/10.3390/genes9010042.

[66] S. Maiti, K.H. Kumar, C.A. Castellani, et al., Ontogenetic de novo copy number variations (CNVs) as a source of genetic individuality: studies on two families with MZD twins for schizophrenia, PLoS One 6 (2011) e17125, , https://doi.org/10.1371/journal.pone.0017125.

[67] P. Scheet, E.A. Ehli, X. Xiao, et al., Twins, tissue, and time: an assessment of SNPs and CNVs, Twin Res. Hum. Genet. 15 (2012) 737–745, https://doi.org/10.1017/thg.2012.61.

[68] M.J. McConnell, M.R. Lindberg, K.J. Brennand, et al., Mosaic copy number variation in human neurons, Science. 342 (2013) 632–637, https://doi.org/10.1126/science.1243472.

[69] X. Cai, G.D. Evrony, H.S. Lehmann, et al., Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain, Cell Rep. 8 (2014) 1280–1289, https://doi.org/10.1016/j.celrep.2015.01.028.

[70] A. Abdellaoui, E.A. Ehli, J.J. Hottenga, et al., CNV concordance in 1,097 MZ twin pairs, Twin. Res. Hum. Genet. 2015 (18) (2015) 1–12, https://doi.org/10.1017/thg.2014.86.

[71] A.P. Morgan, D.M. Gatti, M.L. Najarian, et al., Structural variation shapes the landscape of recombination in mouse, Genetics. 206 (2017) 603–619, https://doi.org/10.1534/genetics.116.197988.

[72] S. Stamouli, B.M. Anderlid, C. Willfors, et al., Copy number variation analysis of 100 twin pairs enriched for neurodevelopmental disorders, Twin. Res. Hum. Genet. 21 (2018) 1–11, https://doi.org/10.1017/thg.2017.69.

[73] Y. Zhang, L. Huang, X. Huang, et al., Chromosomal aberrations and CNVs in twin fetuses with cardiovascular anomalies: comparison between monochorionic diamniotic and dichorionic diamniotic twins, Prenat. Diagn. 32 (2018) 318–327, https://doi.org/10.1002/pd.5238.

[74] M. Völker, N. Backström, B.M. Skinner, et al., Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution, Genome Res. 20 (2010) 503–511, https://doi.org/10.1101/gr.103663.109.

[75] J. Bélteky, B. Agnvall, L. Bektic, et al., Epigenetics and early domestication: differences in hypothalamic DNA methylation between red junglefowl divergently selected for high or low fear of humans, Genet. Sel. Evol. 5 (2018) 13 (0.1186/s12711-018-0384-z).

[76] R. Terracol, N. Prud'homme, Differential elimination of rDNA genes in bobbed mutants of *Drosophila melanogaster*, Mol. Cell. Biol. 6 (1986) 1023–1031.

[77] S.H. Sohn, V.K. Subramani, Dynamic of thelomere length in the chicken, Poult. Sci. 70 (2014) 721–736, https://doi.org/10.1159/000075769.

[78] M.F. Haussmann, R.A. Mauck, Telomeres and longevity: testing an evolutionary hypothesis, Mol. Biol. Evol. 25 (2008) 220–228, https://doi.org/10.1093/molbev/msm244.

[79] P. Monaghan, Organismal stress, telomeres and life histories, J. Exp. Biol. 217 (2014) 57–66, https://doi.org/10.1242/jeb.090043.

[80] H.C. Mefford, B.J. Trask, The complex structure and dynamic evolution of human subtelomeres, Nat. Rev. Genet. 3 (2002) 91–102, https://doi.org/10.1038/nrg727.

[81] F. Bersani, E. Lee, P.V. Kharchenko, et al., Pericentromeric satellite repeat expansions through RNA-derived DNA intermediates in cancer, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 15148–15153, https://doi.org/10.1073/pnas.1518008112.

[82] A. Tanne, L.R. Muniz, A. Puzio-Kuter, et al., Distinguishing the immunostimulatory properties of noncoding RNAs expressed in cancer cells, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 15154–15159, https://doi.org/10.1073/pnas.1517584112.

[83] B. Sac, What is a "Jewish dog"? Konrad Lorentz and the cult of wildness, Soc. Anim. 5 (1997) 3–21 http://www.academia.edu/211878/What_is_a_Jewish_Dog_Konrad_Lorenz_and_the_Cult_of_Wildness.