BMC Genomics

RESEARCH ARTICLE   **Open Access**

Whole genome SNP discovery and analysis of genetic diversity in Turkey (*Meleagris gallopavo*)

Muhammad L Aslam[1*], John WM Bastiaansen[1], Martin G Elferink[1], Hendrik-Jan Megens[1], Richard PMA Crooijmans[1], Le Ann Blomberg[2], Robert C Fleischer[3], Curtis P Van Tassell[4], Tad S Sonstegard[4], Steven G Schroeder[4], Martien AM Groenen[1] and Julie A Long[2]

Abstract

Background: The turkey (*Meleagris gallopavo*) is an important agricultural species and the second largest contributor to the world's poultry meat production. Genetic improvement is attributed largely to selective breeding programs that rely on highly heritable phenotypic traits, such as body size and breast muscle development. Commercial breeding with small effective population sizes and epistasis can result in loss of genetic diversity, which in turn can lead to reduced individual fitness and reduced response to selection. The presence of genomic diversity in domestic livestock species therefore, is of great importance and a prerequisite for rapid and accurate genetic improvement of selected breeds in various environments, as well as to facilitate rapid adaptation to potential changes in breeding goals. Genomic selection requires a large number of genetic markers such as e.g. single nucleotide polymorphisms (SNPs) the most abundant source of genetic variation within the genome.

Results: Alignment of next generation sequencing data of 32 individual turkeys from different populations was used for the discovery of 5.49 million SNPs, which subsequently were used for the analysis of genetic diversity among the different populations. All of the commercial lines branched from a single node relative to the heritage varieties and the South Mexican turkey population. Heterozygosity of all individuals from the different turkey populations ranged from 0.17-2.73 SNPs/Kb, while heterozygosity of populations ranged from 0.73-1.64 SNPs/Kb. The average frequency of heterozygous SNPs in individual turkeys was 1.07 SNPs/Kb. Five genomic regions with very low nucleotide variation were identified in domestic turkeys that showed state of fixation towards alleles different than wild alleles.

Conclusion: The turkey genome is much less diverse with a relatively low frequency of heterozygous SNPs as compared to other livestock species like chicken and pig. The whole genome SNP discovery study in turkey resulted in the detection of 5.49 million putative SNPs compared to the reference genome. All commercial lines appear to share a common origin. Presence of different alleles/haplotypes in the SM population highlights that specific haplotypes have been selected in the modern domesticated turkey.

Background

All commercial turkey lines descend from the South Mexican turkey (*Meleagris gallopavo gallopavo*) indigenous to Mexico, first domesticated in 800 BC [1]. In the US, the turkey is registered as a single breed with eight different varieties as defined primarily by plumage colour. Five of these eight varieties (Bronze, Narragansett, White Holland, Black and Slate) were registered in 1874, while the remaining three (Beltsville Small White, Bourbon Red, and Royal Palm) were registered in 1951, 1909, and 1971 respectively. There are a total of five wild turkey subspecies in North America but none of them contributed to the development of modern commercial lines [1].

Turkey is the second largest contributor of poultry meat consumed worldwide [2]. The production per bird doubled between 1970 and 2008 [2], largely due to selection pressure by the primary breeders for specific economically important traits, such as body weight, meat quality, and egg production [3-6]. Historically, quantitative genetics-based selection has been the primary

* Correspondence: muhammadl.aslam@wur.nl
[1]Animal Breeding and Genomics Centre, Wageningen University, De Elst 1, 6708WD, Wageningen, The Netherlands
Full list of author information is available at the end of the article

strategy of genetic improvement of livestock [7]. This genetic improvement was largely applied to highly heritable traits, such as body size and breast muscle development. Genetic improvement of farm animals through selection may have increased production but has also resulted in a loss of genetic diversity [8]. The efficiency of these classical methods used for genetic improvement decreases when applied to traits that are difficult to measure or have lower heritability [7]. The availability of genome-based selection, based on a large number of SNPs at a density equivalent to the resolution of linkage disequilibrium (LD), has the potential to transform breeding and incorporate previously unavailable genetic information into commercial lines [9] which can be expected to change the impact of commercial breeding on diversity. A tremendous loss of SNP genetic diversity has been observed in chicken with significant absence of rare alleles (50% or more) in commercial breeds compared to ancestral breeds [10].

SNPs are a good marker type to study diversity. SNPs represent the most abundant source of genetic variation within the genome and are linked to heritable differences between individuals [11]. In addition, SNPs have a low mutation rate and are thought to be good genetic markers of potential disease phenotypes as well as for other complex traits [12]. Moreover, SNP markers are amenable to high throughput genotyping platforms and are valuable for a variety of genetic and genomic applications such as the construction of genetic and physical maps and the analysis of genetic diversity [13]. Next generation sequencing (NGS) has proven to be very effective for the large scale, genome-wide discovery of this type of genetic variation [14,15]. When a high quality reference genome sequence is available, genomic sequences of individuals can be aligned more easily to this reference genome to detect nucleotide variation[15,16]. Different studies have applied NGS platforms to achieve highly redundant coverage of the genome, a prerequisite for high quality genome-wide SNP discovery in the complex genomes of plants and animals [17-20].

Turkey genome assembly is based on commercial turkey, containing 39 autosomes and 2 sex chromosomes [21]. The most recent build, UMD 2.01, covers 90% of the genome [22] The size of the turkey genome assembly is 1.1 billion bases and, to date, about 600,000 SNPs [15,21] have been identified within the reference genome assembly. Increasing the number of SNPs identified in the turkey is an essential step for future improvement of economically important traits through genetic association studies [23-25].

Domestication of livestock species and a long history of migrations, selection and adaptation has created an enormous variety in breeds in livestock [8]. Phenotypic selection has created a wide diversity of breeds that are

adopted to different climatic conditions and purposes [26]. Phenotypic variation observed between and among breeds of domestic animals is overwhelming compared with that in natural populations [26]. Chicken is considered the most closely related domesticated agricultural species to turkey. The observed phenotypic diversity in chicken is much larger than that of turkey, [26,27] most likely reflecting a much larger effective population size of chicken, before specialized commercial populations were established during the twentieth century. This is consistent with the extensive sequence diversity present in domestic chicken (5 SNPs/Kb) [28,29].

The presence of genetic diversity in domestic livestock species is of great importance for sustained genetic improvement of selected breeds in various environments, as well as to facilitate rapid adaptation to potential changes in breeding goals [30,31]. In animal breeding, crosses with non-commercial populations are rarely applied and genetically improved animals are often kept in small, closed populations. Small effective population sizes and epistasis can result in loss of genetic diversity, which can lead to reduced individual fitness and reduced response to selection [32,33]. Several studies have assessed genetic diversity in different livestock species [32,34-40] using different types of markers. A number of genetic diversity studies in chicken have reported loss of genetic diversity in commercial chicken populations because of high selection pressure and low effective population size [35,37,41]. A few studies have been published that explored genetic diversity in turkey genetic resources. However, these studies used a limited number of molecular markers [42,43] and only one study has been published that used 9 SNPs along with other molecular markers [44].

The goal of this project was to investigate turkey genome variation and to provide a resource for subsequent genomic work in the turkey and to cover a wide sampling of population for the development of a high-density SNP chip with minimal ascertainment bias. The SNP information will enable or improve application of genomic selection as well as association studies. We have used the identified SNPs to estimate relatedness among the sequenced turkey populations, which will uncover the genetic diversity available to breeders. Information of genetic diversity can be used in the design of breeding programs including making decisions on crosses between lines or introgression of genes from other commercial lines that may affect economically important traits such as growth, meat quality, fitness, and survival traits.

Methods
Populations
Eleven turkey populations were available for this study. Males from seven commercial lines, three heritage

varieties and 113 years old samples of wild turkeys from South Mexico (SM turkeys) were used for whole genome sequencing. The seven commercial lines, L1 through L7, were obtained from two different primary breeding companies. The three heritage varieties were the Beltsville Small White (BvSW), the Royal Palm (RP) and the Narragansett (Nset) [45-47]. Tissue samples representing the wild population were obtained from the Bird Collection of the Smithsonian Institution's National Museum of Natural History (USNM 165490, USNM 166330, and USNM 166329), and were originally collected in 1899 from Chihuahua, Mexico. These samples represent the progenitor subspecies, the South Mexican (SM) turkey. In total 32 individuals were selected for whole genome re-sequencing, with three males per population except for RP, which was represented by 2 males.

Genomic DNA extraction, library preparation and sequencing

Considering mature erythrocytes in poultry are nucleated, genomic DNA was extracted from whole blood of the commercial and heritage lines with the QIAamp DNA blood Midi Kit (Qiagen, Valencia, CA); the procedure included a proteinase K digestion followed by column purification. Integrity of high molecular weight DNA following the extraction was confirmed by agarose gel analysis. Genomic DNA was sheared using the Covaris S2 to yield an average fragment size of 450 bp, as determined with the Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA). The DNA from the three historic SM samples was extracted from the toe-pads in the ancient DNA laboratory at the Smithsonian Institution's Center for Conservation and Evolutionary Genetics, that is fully equipped to avoid contamination with modern DNA. DNA extraction followed a standard protocol of proteinase k and DTT digestion followed by phenol-chloroform extraction and centrifugal dialysis with Centricon concentrators (following methods provided in [48]). An extraction blank sample was used as a no-sample control in each round of extraction. Extractions involved alternation of turkey samples with samples from other avian or non-avian taxa, in order to detect potential cross-contamination among extracts. Extracts of the samples and extract controls were subjected to PCR with standard avian mtDNA primer sets (Cytochrome *b*, ND2; [49]) followed by sequencing of positive products to confirm the isolation of turkey DNA from the toe pads. The genomic DNA of the SM samples ranged from 40-43 bp (Agilent Bioanalyzer).

Genomic libraries were prepared with the Paired-end Sequencing Sample Preparation Kit (Illumina, San Diego, CA) with 5 μg of genomic DNA for commercial and heritage lines according to the manufacturer's

instructions; for the SM samples 0.54 μg was used to construct the libraries. All genomic DNA libraries were validated with the Agilent Bioanalyzer (model 2100). The automated cBot Cluster Generation System (Illumina) was used to generate clusters on the flow cell. Each individual was sequenced (paired-end; read length 120 bp) in a single lane of a flow cell using the Illumina GAIIx. The DNA extracted from museum samples for the SM turkeys was highly degraded, and thus single-end reads of 40 bp were generated from these samples.

Sequence mapping and SNP identification

Sequence reads of each individual from the domesticated populations (heritage varieties and commercial lines) were filtered on base quality; reads were trimmed if three consecutive bases had an average Phred-like quality score of less than 13. Both sequences in a pair needed to exceed 40 bp in length after trimming to be retained for analyses. Sequence reads from the individuals of the SM population were not quality-trimmed before further analyses since they were sequenced to a length of 40 bp only. Sequence reads were aligned against the turkey reference genome (UMD 2.01) using the MOSAIK aligner [50]. Mapping of reads from each individual to the reference genome sequence was performed with hash size 15 (hs), 100 maximum hash positions (mhp), an alignment candidate threshold (act) of 20, and a maximum mismatch percentage (mmp) of 5. Banded Smith-Waterman algorithm (bw = 41) was used to increase the speed of alignments. The algorithm implemented in MOSAIK calculates a mapping quality for each sequence and measures the probability that a sequence belongs to a specific target. The alignments were sorted using MosaikSort. Finally, the file was converted to BAM format using MosaikText. All BAM files have been uploaded to NCBI's Sequence Read Archive (SRA) database under the study accession number "SRP012021.2".

The mpileup function of SamTools version 0.1.12a [51] was used to call variants, separately for each turkey population. The view option of bcftools [51] was used to call the genotype at each variant for each animal. Genotypes were called for each animal with a minimum genotype quality of 20, and a read depth between 1 and 25. At least one individual in a population needed to have a genotype call that met these criteria at a particular position. A SNP that passed the above mentioned criteria were considered as a putative SNP. Putative SNPs were categorized into fixed differences compared to the reference genome and segregating SNPs. Homozygous non-reference genotypes that were the same in all individuals of a population were considered fixed SNPs, while the SNPs that had variable/heterozygous genotypes in a population were considered segregating SNPs.

To estimate heterozygosity (heterozygous SNPs/kb), mpileup genotyping analysis (described above) was used and the number of heterozygous SNPs was calculated at the reference bases covered from 5 to 10 fold. For each individual in a population, heterozygosity was estimated by dividing the total number of discovered heterozygous SNPs by the total genome sequence covered from 5 to 10 fold. Population heterozygosity was estimated by averaging the heterozygosity of all individuals within a population.

Functional annotation of SNPs

The gene-based analysis of ANNOVAR software [52] was used to functionally annotate the putative SNPs. For each putative SNP, the location (exonic, intronic, intergenic, 5'UTR, 3'UTR, splice acceptor or donor site, downstream or upstream) and the functional annotation (nonsynonymous, synonymous, stop codon gain or loss, and amino acid changes) were determined based on the turkey reference genome (UMD 2.01). Gene annotations used in this analysis were taken from Ensembl [53]. Standard settings for gene based analysis of ANNOVAR were used.

Nucleotide diversity and false discovery rate

Genome wide mapping density, or read depth distribution, and the nucleotide diversity across the whole genome were assessed for each individual of the 11 turkey populations. Read depth distribution was used to calculate average sequence coverage across the whole genome. To get genotypes of each individual without imputation, pileup function of SamTools version 0.1.12a [51] was used for the estimation of nucleotide diversity across the whole genome. Genotypes were called for each individual using minimum genotype quality of 20, and a read depth between 3 and 15. The number of heterozygous and homozygous non-reference SNP calls was estimated compared to the reference genome within a 300 Kb window. In order to estimate SNP false discovery rate (FDR), 30 large genomic regions of variable sizes (ranging from 2.7-10.5 Mb on variable positions at chromosomes 1, 3 and 10) were investigated where one individual from each of the 10 domesticated populations was clearly homozygous for a single haplotype. Homozygous regions were identified by visual inspection of the nucleotide diversity plots for turkey chromosome 1, 3 and 10. Any SNP within these regions were considered to be false positives. The false discovery rate was calculated as the total number of heterozygous SNP positions divided by the total number of bases covered (1–25 fold coverage) in these 30 regions.

Genetic diversity analysis

PHYLIP software, version 3.69 [54] was used to calculate pairwise Nei's genetic distance [55] among all the individuals from the 11 turkey populations. SNPs for which genotypes were called in at least 9 turkey populations (irrespective of whether SNPs were segregating in all these populations) were selected and utilized for the genetic diversity analysis. Threshold of at least 9 turkey population was selected to increase number of selected SNPs for analysis and to make sure presence of selected SNPs in maximum populations to have a reliable genetic comparison. Pairwise genetic distance analyses were based on marker data that the individuals had in common, because PHYLIP is unable to deal with missing data [36]. Mega 5.0 [56] was used for hierarchical clustering using a Neighbour-joining procedure on the genetic distance matrix for all the individuals. The wild population was used to root the phylogenetic tree.

Non-reference allelic state

The genome of each individual was screened, using the nucleotide diversity analysis described above, for the occurrence of non-reference allelic states. Determining the ancestral allelic state of SNPs was not possible because species with appropriate evolutionary distance are not available. Chicken is considered a closely related domesticated agricultural species to turkey but the evolutionary distance to the last common ancestor of these two species is around 30 million years [57]. To quantify regional changes in genomic diversity between SM and the domesticated populations, we used heterozygosity as well as the presence of non-reference allelic homozygosity of the positions sufficiently covered by sequencing.

The difference in non-reference allele homozygosity between domesticated and the SM turkey populations was calculated for each bin. This difference was then divided by the average homozygous non-reference allele SNP density for the bin to yield a relative measure that can be compared between bins with different levels of variation.

The ratio of non-reference homozygosity in wild SM vs. domesticated populations was calculated within bin sizes of 300 Kb. A high ratio points to non-reference alleles being lost, or decreased in frequency during domestication and selection. A high ratio of non-reference homozygosity, in combination with low heterozygosity in the domesticated populations, is interpreted as a reduction of allelic variation from wild to domesticated populations, or "fixation of the reference alleles". A bin was considered "fixed for the reference allelic state" in domesticated populations when two conditions were met. First, bins were considered "fixed" when heterozygosity was equal or lower than 0.0002 on average across all domesticated populations. This threshold was chosen because only 5% of the bins had a heterozysity equal or lower than 0.0002 (1 heterozygous position/5000 bp). Second, bins that were considered "fixed" had to have a

ratio of non-reference allele homozygosity above or equal to 1.73, which means that the non-reference allele homozygosity of the wild population must be at least 73% higher than the domesticated populations. This threshold was chosen because only 5% of all the bins in the genome had a ratio equal or higher than 1.73.

Ethical approval for the use of animals in this study

Although animals were used in this study, no direct experiments were performed on them. Blood sample collection was carried out by highly skilled and experienced personnel from the breeding companies. No approval from the ethics committee was necessary according to local legislation.

Results

Whole-genome resequencing and SNPs discovery

The obtained sequence from the DNA samples of the domestic populations (heritage varieties and the commercial lines) varied from 2.30-13.21 Gbp (Giga basepairs) per individual. After quality trimming and alignment of the short reads, the percentage of bases in the reference genome covered by at least 1 and a maximum of 25 reads varied from 47.48% to 86.13% for the animals analyzed (Table 1). The sequences generated from SM turkeys varied from 0.41-0.82 Gb of sequence per individual. The sequence depth at bases covered by at least one read ranged from 1.38 to 1.81 for the SM samples and 2.07 to 6.72 for the domesticated turkey lines (Table 1).

In total, 5.49 million putative SNPs were identified compared to the reference genome (Table 2). Of these 5.49 million SNPs, 4.76 million SNPs were segregating in at least one population (Table 2). The number of segregating SNPs for the different turkey populations varied from 0.12 to 1.58 million, with the highest number of segregating SNPs observed in L3 and the lowest number observed in SM (Table 3). The lowest number of fixed SNPs was observed in L3 and the highest number of fixed SNPs was observed in BvSW (Table 3). The transition to transversion (Ti/Tv) ratio of the SNPs discovered is 2.45. Of the total 5.49 million SNPs discovered, 75,254 were located in exonic regions, including 23,795 nonsynonymous , 52,506 synonymous, 377 stop gain and 8 stop loss variants. The majority of these exonic SNPs, 66,795 or 89% were segregating within the populations analyzed (Table 4).

Heterozygosity

The number of heterozygous genotypes detected within the individuals from the ten domesticated populations (heritage varieties and the commercial lines) varied from 0.08 to 0.80 million with an average of 0.55 million heterozygous genotypes per individual. Individuals from the

SM population showed relatively low numbers of heterozygous SNPs; between 0.01 and 0.07 million. Heterozygosity (heterozygous SNPs/kb) of all individuals from the different turkey populations ranged from 0.17-2.73 while heterozygosity of populations ranged from 0.73-1.64 (Table 2 & 3). The BvSW population had the lowest heterozygosity, while SM showed the highest heterozygosity within the analyzed populations (Table 3). Observed average nucleotide diversity in the 10 largest chromosomes was 0.0005 segregating SNPs per nucleotide position while average nucleotide diversity in the smaller chromosomes (20–30) was 0.0007. Chromosome Z showed the lowest nucleotide diversity with 0.0002 segregating SNPs per nucleotide position. Based on observed homozygous regions (Figure 1), interpreted to represent two copies of the same Identical By Descent (IBD) haplotype, the estimated average heterozygous genotype FDR was 0.00002 per nucleotide position in the reference genome (ranging from 0.000012-0.000023 in the different individuals).

Genetic diversity

There were 223,264 SNPs segregating in at least 9 turkey populations, and these were used to calculate Nei's pair wise genetic distances. The tree based on Nei's genetic distance for the 32 turkey individuals from the 11 different turkey populations presents their genetic relationships (Figure 2). Individuals from a specific turkey population clustered closely together. Inter-population comparisons demonstrated that commercial lines formed a cluster that was distinct from heritage lines with the exception of the L5 line, which exhibited a closer genetic relation to the heritage varieties. Among the heritage varieties, RP and Nset were more genetically related than either to BvSW. Individuals from the SM population also clustered together and showed relatively closer genetic relation with BvSW population.

Non-reference allelic state

Six regions on five different turkey chromosomes (3, 4, 9, 14, and 22) showed differences between the SM and the domesticated populations with respect to the occurrence of no-reference wild type and the reference allelic states (Figure 3). Domesticated populations predominantly showed the reference allelic state, while the SM populations predominantly showed the no-reference wild type allelic state within these regions. These six regions were then examined with respect to the heterozygous SNP density per nucleotide positions within the same bin size. Within these six regions, nucleotide diversity for all the domesticated populations was found to be close to zero, except for one region on chromosome 4 that showed high segregation of non-reference alleles within the domesticated populations (Figure 4). The

Table 1 Alignment statistics for the individuals from different turkey populations

IDs	Sequence coverage (fold)[1]	Assembly coverage (%)[2]	Assembly coverage 1-25X (%)[3]
L1a	5.12	79.04	78.93
L1b	4.72	83.88	84.04
L1c	5.61	84.10	83.85
L2a	6.54	85.91	85.85
L2b	6.72	86.19	86.13
L2c	5.18	80.16	80.05
L3a	6.32	85.98	85.68
L3b	5.75	85.26	85.21
L3c	6.24	85.91	85.72
L4a	6.19	85.58	85.51
L4b	5.75	84.65	84.58
L4c	5.13	84.14	84.12
L5a	3.52	71.18	71.14
L5b	5.18	71.35	71.27
L5c	5.73	68.35	68.08
L6a	2.88	65.14	65.13
L6b	4.50	77.53	77.49
L6c	4.52	81.45	81.43
L7a	5.46	78.59	78.39
L7b	4.61	57.86	57.70
L7c	4.99	70.88	70.78
BvSW1	4.55	83.21	83.19
BvSW2	5.72	48.33	47.48
BvSW3	5.59	82.24	82.13
Nset1	2.07	53.84	53.82
Nset2	5.39	83.94	83.86
Nset3	5.17	79.42	79.29
RP1	5.31	60.31	60.05
RP2	5.00	63.54	63.43
SM1	1.81	47.10	47.06
SM2	1.38	29.32	29.30
SM3	1.73	45.41	45.40

[1] Average sequence depth of each base in the reference genome that is covered by at least 1 read. The used turkey reference genome (UMD 2.01) has genome size of 1,061,982,190 bp, which is 90% of the total turkey genome size. [2] Percentage of reference genome that is covered by at least one read. [3] Percentage of reference genome that is covered by 1–25 reads.

other five genomic regions, two regions in chromosome 22 and one region in each of the remaining three chromosomes, (3, 9 and 14), met the criteria mentioned in the methodology section (Additional file 1). These genomic regions were considered fixed for the reference allelic state in the domesticated populations.

Discussion

In this study, we performed whole genome sequencing for SNPs discovery and used the identified SNPs to characterize genetic diversity in the turkey genome. To avoid imputation of genotype calls across the different populations, mpileup was applied within each population separately because the applied method (mpileup) relies in part on Hardy-Weinberg Equilibrium (HWE) for imputation of genotypes [51].

By using a NGS (Illumina GAIIx) approach, we discovered millions of high quality SNPs in the turkey. Next generation sequencing approaches are considered highly reliable for genome-wide discovery of sequence variation [15], when used to compare different lines/strains to a reference genome [58]. The adoption of NGS platforms

Table 2 Heterozygosity and the number of SNP observed in each individual of different turkey populations

IDs	Homozygous NR SNP [1]	Heterozygous SNP	Heterozygous SNP 5-10X	Genome covered 5-10X (bp)	Heterozygosity Kb^{-1}
L1a	663,406	659,351	369,849	320,663,179	1.15
L1b	686,583	648,928	385,673	396,624,720	0.97
L1c	626,434	737,472	403,423	375,734,398	1.07
L2a	827,249	755,318	504,787	532,961,711	0.95
L2b	896,728	757,226	514,059	554,379,839	0.93
L2c	869,872	562,653	311,525	329,283,144	0.95
L3a	568,439	762,252	519,228	532,049,588	0.98
L3b	434,157	427,393	567,558	527,841,728	0.99
L3c	608,276	834,241	164,167	166,315,925	1.08
L4a	720,530	616,567	440,086	454,905,713	0.80
L4b	760,762	692,079	385,458	439,002,235	0.97
L4c	807,407	618,335	403,201	503,650,627	0.88
L5a	666,287	340,436	160,698	180,577,454	0.89
L5b	652,149	352,682	165,723	144,150,087	1.15
L5c	736,951	520,850	251,977	223,238,275	1.13
L6a	581,773	294,736	109,405	115,435,304	0.95
L6b	644,421	567,275	330,736	306,448,666	1.08
L6c	638,770	579,232	341,869	348,094,277	0.98
L7a	736,881	550,299	300,174	305,785,110	0.98
L7b	698,647	379,941	185,444	161,035,610	1.15
L7c	730,143	504,513	275,118	252,564,184	1.09
BvSW1	1,053,237	417,544	241,641	372,524,318	0.65
BvSW2	1,071,513	269,338	103,333	144,219,590	0.72
BvSW3	1,086,121	525,262	299,713	369,633,525	0.81
Nset1	643,308	79,232	25,217	144,546,998	0.17
Nset2	667,797	519,815	9,929	4,717,330	2.10
Nset3	773,183	804,627	454,052	320,395,210	1.42
RP1	885,734	510,427	154,899	167,716,001	0.92
RP2	842,442	522,599	276,752	208,702,070	1.33
SM1	551,149	69,199	11,106	9,379,558	1.18
SM2	551,380	17,275	2,030	744,899	2.73
SM3	551,543	44,784	6,921	6,868,381	1.01

[1] Homozygous non reference SNPs observed in each individual.

for the discovery of genomic variation has now become mainstream [15,58-60].

The high quality of the SNPs discovery reported here is reflected by the low FDR of 0.00002 per nucleotide in the genome. This FDR suggests around 2.1×10^4 false discovered heterozygous positions per turkey genome (size of 1.1×10^9 base pairs). The SNPs FDR rate for the same 10 animals from distinct turkey populations was estimated after correcting for the coverage and using estimates of FDR per nucleotide position. The SNPs FDR was found to be 2.6%, a number that is similar in magnitude as found previously in the human 1000 Genome Project. In addition to the low FDR, we found a transition/transversion (Ti/Tv) ratio within the expected range. The expected Ti/Tv ratio of true novel variants can vary with the targeted region (whole genome, exome, specific genes), species and also can vary greatly by the CpG and GC content of the region [59-61]. In the case of exomes, an increased presence of methylated cytosine in CpG dinucleotides in exonic regions leads to an increased Ti/Tv ratio [61] due to an easy deamination and transition of a methylated cytosine to a thymine [61]. It is also observed that GC content is higher in birds and mammals than in invertebrates [62]. Observed Ti/Tv ratio in our study of turkey is in concordance with the findings from Dalloul et al. [21], but slightly higher

Table 3 Discovered segregating, and the fixed number of SNPs along with the observed heterozygosity Kb^{-1} in each turkey population

Population ID	Segregating SNPs[1]	Fixed SNPs[2]	Heterozygosity Kb^{-1}
L1	1,563,553	617,893	1.07
L2	1,504,682	781,352	0.94
L3	1,589,525	502,807	1.01
L4	1,441,173	709,507	0.88
L5	950,425	674,038	1.06
L6	1,139,459	613,069	1.00
L7	1,097,788	673,807	1.07
BvSW	926,733	1,047,010	0.73
Nset	1,194,570	708,773	1.23
RP	883,602	813,164	1.12
SM	120,305	552,032	1.64

[1] The total number of SNPs detected compared to the reference genome in which the non-reference allele is segregating in a population. [2] The total number of SNPs detected compared to the reference genome in which only the non-reference allele is found in a population.

(2.45) than that of human. This higher ratio is most likely explained by the smaller genome size and a higher GC percentage in bird genomes.

We report the number of segregating as well as total number of SNPs with their functional annotation. The 23,795 nonsynonymous variants that were observed can potentially change the structure of proteins, possibly resulting in altered phenotypes [63]. Out of these nonsynonymous SNPs, 9,204 were unique to commercial population which may have been detected due to higher coverage and number of individuals for the commercial

Table 4 Number of SNPs detected

Variants	Reference total[1]	Segregating Total[2]
Nonsynonymous	23,795	20,463
Synonymous	52,506	47,281
Stopgain	377	295
Stoplost	8	7
Exonic splice site	1,437	1,256
Exonic	75,254	66,795
Splice acceptor or donor site (interonic)	734	607
5'UTR/3'UTR	8,933	7,661
Upstream/downstream	142,829	124,005
Intronic	1,749,427	1,518,783
Intergenic	3,514,102	3,044,243
ncRNA	1,044	916
Total	**5,493,760**	**4,764,266**

[1] SNPs detected compared to the reference genome in which the non-reference allele is detected in at least one of the 29 individuals.
[2] Detected segregating SNPs within all turkey individuals.

turkey population. We observed 5,417,069 SNPs that were present in non-protein coding DNA. Furthermore, we discovered 1,749,427 intronic variants, some of which may alter gene expression or result in alternative splicing [64,65]. Variants located in intergenic regions, such as promoter, enhancer and silencer regions can result in altered gene expression. The human genome comprises over 98% non-protein coding DNA [66]. Estimates suggest that at least 5.5% of the human genome, including 3.5% of its noncoding fraction, consists of regions under purifying natural selection against deleterious alleles [67–69]. In addition, most of the variants involved in complex genetic diseases in humans are not located in coding regions [59]. Likewise, variation outside of coding regions may be responsible for economically important traits in domesticated species, e.g. disease resistance, meat quality, efficient growth, or high egg production. The functional information of these variants can help in prediction of phenotypes or genetic merit with higher accuracy and selection of individuals can be done accordingly.

The estimated average frequency of 1.07 heterozygous SNPs Kb^{-1} in the turkey is substantially lower than in chicken, which was previously reported as 4.28 and 2.24 heterozygous SNPs Kb^{-1} in two different studies [28,29]. In our study, heterozygous SNP discovery was found to be affected by the sequence coverage (e.g. sequence coverage in L6a, Nset1 and the SM animals was low and as a result the number of observed heterozygous SNPs was also low). Estimates of heterozygosity were therefore obtained only from genomic regions that were covered 5 to 10X to adjust for the effect of low sequence coverage.

Modern commercial turkey lines are derived from historic turkey populations that displayed low variation as a result of small effective population size [70,71]. Heritage (Nset and RP) and the wild SM turkey populations showed higher heterozygosity compared to the commercial populations, which is concordant with the findings of previous studies on ancient and overexploited species [72-74]. The heritage variety BvSW showed the lowest heterozygosity of all turkey populations, which is consistent with the severe bottleneck that this population went through in 2000 (Alexandra Scupham, Personal communications).

Most birds have a characteristic division in chromosome size, with 5 or 6 large chromosomes, around 5 intermediate size chromosomes, and 25 to 30 very small chromosome pairs. In our study, we observed higher nucleotide diversity on smaller chromosomes compared to the larger turkey chromosomes which is in agreement with the previous study [75]. Since the recombination rate is far higher at the smaller sized turkey chromosomes as compared to large chromosomes [76], which leads to lower linkage disequilibrium and higher haplotype diversity on the smaller
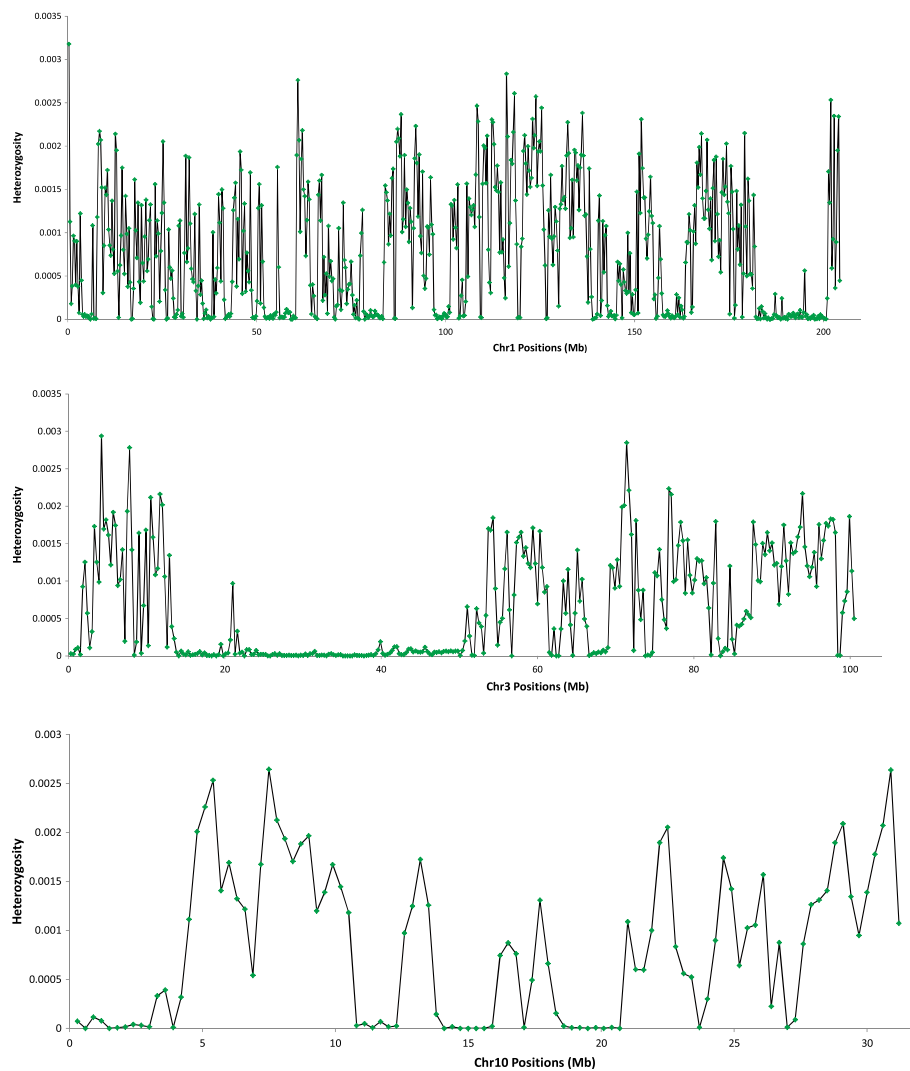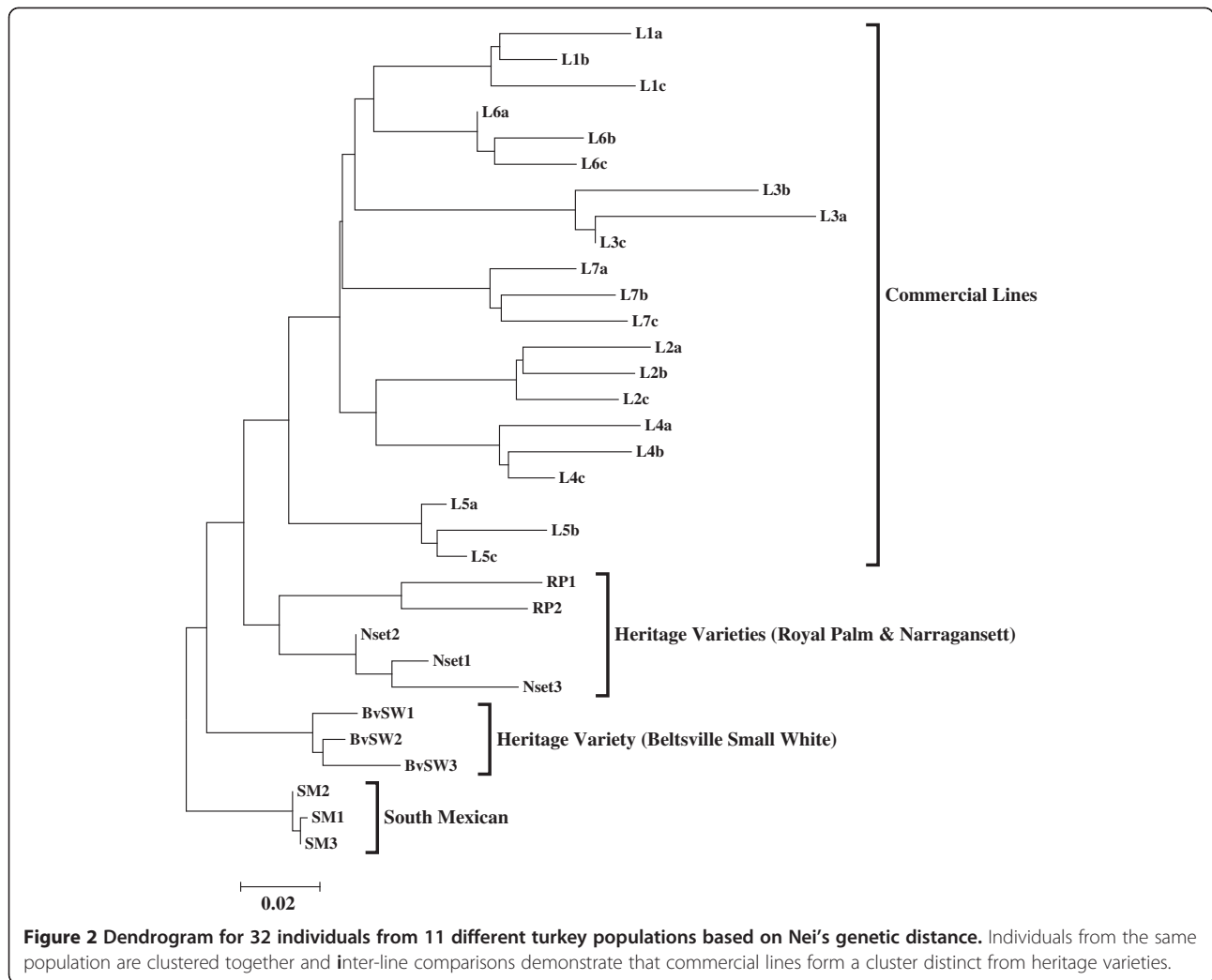
Figure 1 Nucleotide diversity showing heterozygous and homozygous regions within chromosome 1, 3 and 10. Heterozygosity across chromosome 1, 3 and 10 for individual L1c. Chromosome positions (Mb) are presented on the x-axis and on the y-axis heterozygosity is given as the density of heterozygous SNPs corrected for the number of bases covered within a window size of 300 Kb. Note the clear homozygous regions at 188–198 Mb for chromosome 1, 24–38 Mb for chromosome 3 and 18–21 Mb for chromosome 3.

chromosomes [77]. Although the high gene-density of the smaller chromosomes would make them susceptible to hitchhiking effects that could erode genetic variation, hitchhiking effects appear to be offset by the far higher recombination rate of the micro-chromosomes. Chromosome Z showed the lowest nucleotide diversity, which is concordant with the findings of Dalloul et al. [21]. This low nucleotide diversity of chromosome Z is likely the result of a lower effective population size of this chromosome and lower recombination rate [78].

The presence of different allelic states in the wild SM and the domesticated populations is a demonstration of their divergence during the course of domestication event. Domesticated turkey lines were selected (artificially or naturally) for non-wild type alleles. Domestication has
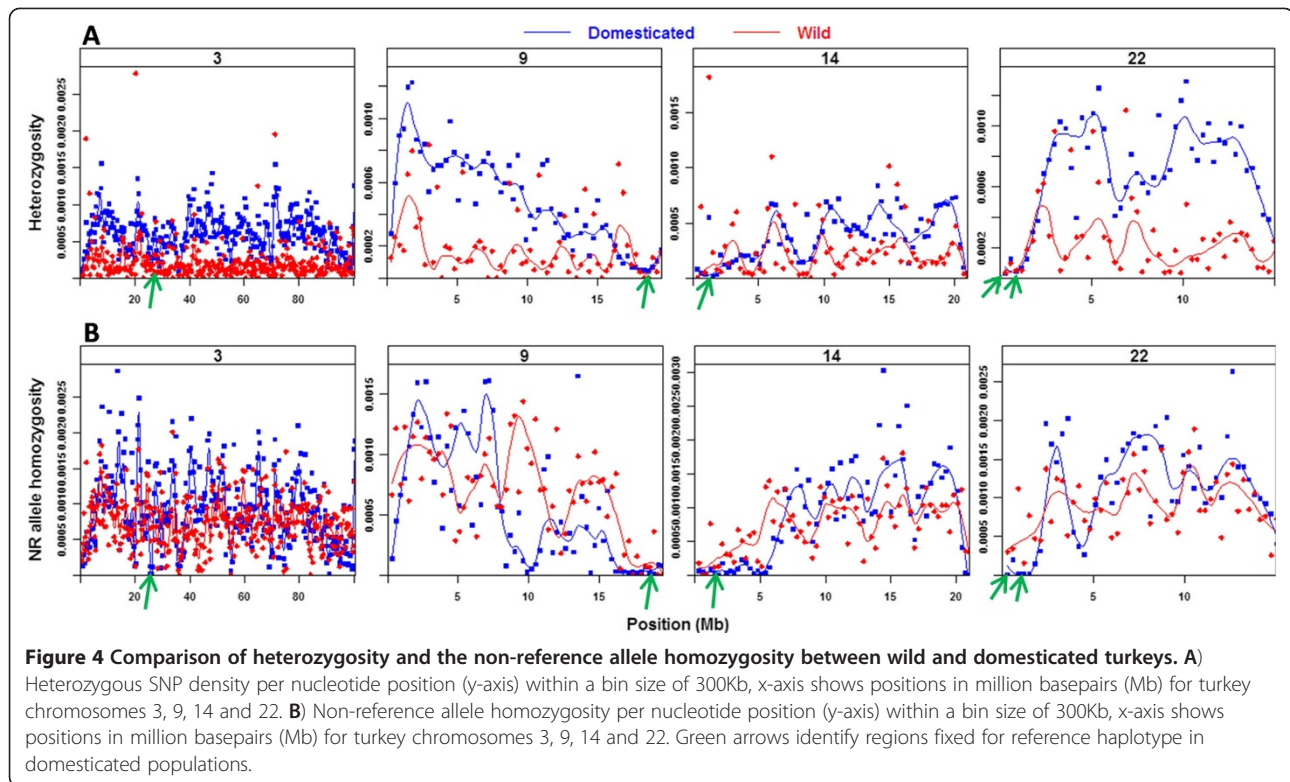
involved the selection on a desired trait(s) [79], and previous studies on domesticated animals have demonstrated selective pressures on genes related to growth [64] and coat colour [80,81]. Such studies have also demonstrated that artificial selection might have contributed to reduced polymorphism levels and increased LD in domesticated species [10,82-84]. On-going directional selection causes footprints of selection identifiable as regions where the derived allele frequency is higher than non-selected regions [29,85,86]. Most of the turkey chromosomes are acrocentric and the five genomic regions that were found to be fixed for the reference alleles within the domesticated populations seem to be located close to the centromere [87]. This may explain the presence of a strong hitchhiking effect due to the

Figure 2 Dendrogram for 32 individuals from 11 different turkey populations based on Nei's genetic distance. Individuals from the same population are clustered together and inter-line comparisons demonstrate that commercial lines form a cluster distinct from heritage varieties.

low recombination rate close to the centromeres. These fixed turkey genomic regions were then investigated for the presence of report QTLs corresponding to these regions. While QTLs were not found within the fixed regions [88], there were QTLs for growth and meat quality on chromosome 3, a QTL for percentage drip loss on chromosome 14 and a growth related QTL on the chromosome 22 [88]. These QTLs for different traits on chromosomes 3, 14 and 22 were located at distinct positions that did not coincide with the observed

Figure 3 Difference in non-reference allele homozygosity per nucleotide position between domesticated and the SM populations. Y-axis denotes difference in non-reference allele density per nucleotide position relative to the mean level of variation discovered between domesticated and the wild SM turkey populations. Five turkey chromosomes 3, 4, 9, 14 and 22 shows visible difference in peaks of these chromosomes.

Figure 4 Comparison of heterozygosity and the non-reference allele homozygosity between wild and domesticated turkeys. A)
Heterozygous SNP density per nucleotide position (y-axis) within a bin size of 300Kb, x-axis shows positions in million basepairs (Mb) for turkey
chromosomes 3, 9, 14 and 22. **B)** Non-reference allele homozygosity per nucleotide position (y-axis) within a bin size of 300Kb, x-axis shows
positions in million basepairs (Mb) for turkey chromosomes 3, 9, 14 and 22. Green arrows identify regions fixed for reference haplotype in
domesticated populations.

regions with high reference allele frequency. Due to the
evidence of the presence of structural and functional
conservation in the turkey and the chicken genomes
[76,88] and also the limited availability of information
on turkey QTLs, these 5 turkey genomic regions that
were found to be fixed for reference alleles within
domesticated populations, were aligned with the chicken
genome sequence (WASHUC2) to determine the pos-
ition of these turkey genomic regions within the chicken
genome (Additional file 1). Regions of the chicken gen-
ome exhibiting synteny with turkey were then examined
for the presence of known chicken QTLs [89]. Several
QTL were identified within these 5 genomic regions
(Additional file 1) and most were related to growth
traits (Additional file 1). Production census of turkeys
from the last few decades [2] show that turkeys are
highly selected for growth and this high selection pres-
sure might have favoured reference alleles in domesti-
cated populations. Since several of the regions identified
in this study are probably close to a centromere, the ef-
fect of selection may have extended over a larger region
due to the likely reduced recombination rate in centro-
meric parts of the genome.

The genetic diversity analysis among the 11 different
turkey lines showed that the heritage varieties and the
commercial populations are derived from the wild South
Mexican population. All of the heritage varieties (BvSW,
RP and Nset) are closely related which is in agreement

with previously published data [43,44]. The relatedness
of these heritage varieties can probably be explained ei-
ther by historic nature, a common origin, selection for
similar traits/phenotype or a relatively low selection
pressure in these varieties. The Nset, RP and BvSW heri-
tage lines were developed in America in 1800, 1920 and
1930, respectively [70,71]. It is assumed that the colour
pattern of RP is derived from crossbreeding with
Narragansett and perhaps another variety, as Nset colour
mutation is a component of the final RP colour (Smith
et al., 2005). The close genetic relatedness observed be-
tween RP and Nset in our study is also concordant with
that assumption and with previous studies [43,44].
According to Figure 2, commercial lines from different
breeding companies did not resolve into two separate
groups. The close relatedness of the L5 commercial line
to the heritage lines is not surprising as it represents a
female line selected for medium weight, conformation
and egg production; selected traits characteristic of the
heritage lines [71]. The other commercial lines that clus-
ter separate from L5 in the dendrogram were selected
for different objectives such as higher body weight and
rapid growth.

Conclusion

The turkey genome is much less diverse with a relatively
low frequency of heterozygous SNPs as compared to other
livestock species like chicken and pig. The whole genome

SNP discovery study in turkey resulted in the detection of 5.49 million putative SNPs compared to the reference genome. All commercial lines appear to share a common origin. Presence of different alleles/haplotypes in the SM population highlights that specific haplotypes have been selected in the modern domesticated turkey.

Additional file

Additional file 1: Positions of turkey genomic regions with their mapping positions and underling QTL in chicken genome. This file contains the start and the end positions of turkey genomic regions that showed fixed haplotype for the reference alleles in domesticated populations. This file also contains information about the start and the end positions of these turkey genomic regions in chicken genome and the chicken QTL reported within these regions.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Animal Breeding and Genomics Centre, Wageningen University, De Elst 1, 6708WD, Wageningen, The Netherlands. [2]Animal Biosciences and Biotechnology Laboratory, Animal and Natural Resources Institute, Beltsville Agricultural Research Center, United States Department of Agriculture, Beltsville, MD 20705, USA. [3]Center for Conservation and Evolutionary Genetics, Smithsonian Conservation Biology Institute, Washington, DC 20013, USA. [4]Bovine Functional Genomics Laboratory, Animal and Natural Resources Institute, Beltsville Agricultural Research Center, United States Department of Agriculture, Beltsville, MD 20705, USA.

References
1. Speller CF, Kemp BM, Wyatt SD, Monroe C, Lipe WD, Arndt UM, Yang DY: Ancient mitochondrial DNA analysis reveals complexity of indigenous North American turkey domestication. *PNAS* 2010, **107**(7):2807–2812.
2. Food and agriculture organization statistical division (FAOSTAT) of the United Nations: *Food and agriculture organization statistical division (FAOSTAT) of the United Nations*. http://faostat.fao.org/.
3. Berri C, Wacrenier N, Millet N, Le Bihan-Duval E: Effect of selection for improved body composition on muscle and meat characteristics of broilers from experimental and commercial lines. *Poult Sci* 2001, **80**:833–838.
4. Li H, Deeb N, Zhou H, Mitchell AD, Ashwell CM, Lamont SJ: Chicken quantitative trait loci for growth and body composition associated with transforming growth factor-β genes1. *Poult Sci* 2003, **82**:347–356.
5. Le Bihan-Duval E, Debut M, Berri CM, Sellier N, Santé-Lhoutellier V, Jégo Y, Beaumont C: Chicken meat quality: genetic variability and relationship with growth and muscle characteristics. *BMC Genetics* 2008, **9**:53.
6. Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Ducro BJ, Vereijken A, Megens H-J, Groenen MAM: Genetic variances, heritabilities and maternal effects on body weight, breast meat yield, meat quality traits and the shape of the growth curve in turkey birds. *BMC Genetics* 2011, **12**(1):14.
7. Montaldo HH, Meza-Herrera CA: Use of molecular markers and major genes in the genetic improvement of livestock. *EJB* 1998, **1**(2).
8. Groeneveld LF, Lenstra JA, Eding H, Toro MA, Scherf B, Pilling D, Negrini R, Finlay EK, Jianlin H, Groeneveld E, et al: Genetic diversity in farm animals - a review. *Animal Genetics* 2010, **41**:6–31.
9. Dekkers JCM, Hospital F: The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews Genetics* 2002, **3**:22–32.
10. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen MAM, Crooijmans RPMA, Megens H-J, Zhang H, Okimoto R, Vereijken A, et al: Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds. *PNAS* 2008, **105**(45):17312–17317.
11. Suh Y, Vijg J: SNP discovery in associating genetic variation with human disease phenotypes. *Mutat Res* 2005, **573**(1–2):41–53.
12. Gray IC, Campbell DA, Spurr NK: Single nucleotide polymorphisms as tools in human genetics. *Human Molecular Genetics* 2000, **9**(16):2403–2408.
13. Duran C, Appleby N, Edwards D, Batley J: Molecular genetic markers: Discovery, applications, data storage and visualisation. *Current Bioinformatics* 2009, **4**:16–27.
14. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES: An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 2000, **407**(6803):513–516.
15. Kerstens HHD, Crooijmans RPMA, Veenendaal A, Dibbits BW, Chin-A-Woeng TFC, Dunnen JT, Groenen MAM: Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey. *BMC Genomics* 2009, **10**:479.
16. Li G, Ma L, Song C, Yang Z, Wang X, Huang H, Li Y, Li R, Zhang X, Yang H, et al: The YH database: the first Asian diploid genome database. *Nucleic Acids Res* 2009, **37**:1025–1028.
17. Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, et al: Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. *PLoS One* 2009, **4**(8):e6524.
18. Stothard P, Choi J-W, Basu U, Sumner-Thomson J, Meng Y, Liao X, Moore S: Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. *BMC Genomics* 2011, **12**(1):559.
19. You F, Huo N, Deal K, Gu Y, Luo M-C, McGuire P, Dvorak J, Anderson O: Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. *BMC Genomics* 2011, **12**(1):59.
20. Kilian B, Graner A: NGS technologies for analyzing germplasm diversity in genebanks. *Briefings in Functional Genomics* 2012, **11**(1):38–50.
21. Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Ann Blomberg L, Bouffard P, Burt DW, Crasta O, Crooijmans RPMA, et al: Multi-Platform Next-Generation Sequencing of the Domestic Turkey (*Meleagris gallopavo*): Genome Assembly and Analysis. *PLoS Biol* 2010, **8**(9):e1000475.
22. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, et al: Ensembl 2012. *Nucleic Acids Research* 2012, **40**(D1):D84–D90.
23. Ma L, Runesha HB, Dvorkin D, Garbe JR, Da Y: Parallel and serial computing tools for testing single-locus and epistatic SNP effects of

quantitative traits in genome-wide association studies. *BMC Bioinformatics* 2008, **9**:315.

24. Boschiero C, Rosario MF, Ledur MC, Campos RLR, Ambo M, Coutinho LL, Moura ASAMT: **Associations between microsatellite markers and traits related to performance, carcass and organs in chickens.** *Int J Poult Sci* 2009, **8**(7):615–620.

25. Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, Li H, Chen K, Ye S, Ge C, *et al*: **Genome-wide association study of body weight in chicken F2 resource population.** *PLoS ONE* 2011, **6**(7):e21872.

26. Andersson L: **Genetic dissection of phenotypic diversity in farm animals.** *Nature Rev Genet* 2001, **2**:130–138.

27. American Poultry Association: **In.** In *APA recognized breeds and varieties*. 2010 http://www.amerpoultryassn.com.

28. International Chicken Polymorphism Map Consortium: **A genetic variation map for chicken with 2.8 million single nucleotide polymorphisms.** *Nature* 2004, **432**:717–722.

29. Rubin C-J, Zody MC, Eriksson J, Meadows JRS, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S, *et al*: **Whole-genome resequencing reveals loci under selection during chicken domestication.** *Nature* 2010, **464**(7288):587–591.

30. Notter DR: **The importance of genetic diversity in livestock populations of the future.** *J Anim Sci* 1999, **77**:61–69.

31. Bijma P, Meuwissen THE, Woolliams JA: **Design of sustainable breeding programs in developed countries.** In *Proceedings 7th World Congress on Genetics Applied to Livestock Production, Montpellier, 19-23 August Vol 30: 2002; [S.l.]:* [s.n.]; 2002: 133.

32. Ryman N, Utter F, Laikre L: **Protection of intraspecific biodiversity of exploited fishes. Rev. Fish Biol. Fish. 5, 417–446.** *Rev Fish Biol Fish* 1995, **5**:417–446.

33. Ficetola GF, Garner TW, De Bernardi F: **Genetic diversity, but not hatching success, is jointly affected by postglacial colonization and isolation in the threatened frog, Rana latastei.** *Mol Ecol* 2007, **16**(15):3285.

34. Kijowski J, Niewiarowicz A: **Emulsifying properties of proteins and meat from broiler breast muscles as affected by their initial pH values.** *J Food Technol* 1978, **13**:451–459.

35. Hillel J, Granevitze Z, Twito T, Ben-Avraham D, Blum S, Lavi U, David L, Feldman MW, Cheng H, Weigend S: **Molecular markers for the assessment of chicken biodiversity.** *World's Poultry Science Journal* 2007, **63**:33–45.

36. Megens HJ, Crooijmans RP, San Cristobal M, Hui X, Li N, Groenen MA: **Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication.** *Genet Sel Evol* 2008, **40**(1):103–128.

37. Fulton JE: **Genomic selection for poultry breeding.** *Animal Frontiers* 2012, **2**(1):30–36.

38. Kijas JM, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, *et al*: **A genome wide survey of SNP variation reveals the genetic structure of sheep breeds.** *PLoS ONE* 2009, **4**(3):e4668.

39. Gautier M, Laloë D, Moazami-Goudarzi K: **Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds.** *PLoS One* 2010, **5**(9). pii: e13038.

40. Lin BZ, SasazaKi S, Mannen H: **Genetic diversity and structure in Bos taurus and Bos indicus populations analyzed by SNP markers.** *Ani Sci J* 2010, **81**(3):281–289.

41. Sawai H, Kim HL, Kuno K, Suzuki S, Gotoh H, Takada M, Takahata N, Satta Y, Akishinonomiya F: **The Origin and Genetic Variation of Domestic Chickens with Special Reference to Junglefowls Gallus g. gallus and G. varius.** *PLoS One* 2010, **5**(5). e10639.

42. Mock KE, Theimer TC, Rhodes OE Jr, Greenberg DL, Keim P: **Genetic variation across the historical range of the wild turkey (Meleagris gallopavo).** *Molecular Ecology* 2002, **11**:643–657.

43. Kamara D, Gyenai KB, Geng T, Hammade H, Smith EJ: **Microsatellite marker-based genetic analysis of relatedness between commercial and heritage turkeys (Meleagris gallopavo).** *Poult Sci* 2007, **86**:46–49.

44. Smith EJ, Geng T, Long E, Pierson FW, Sponenberg DP, Larson C, Gogal R: **Molecular analysis of the relatedness of five domesticated turkey strains.** *Biochemical Genetics* 2005, **43**:35–47.

45. The American Livestock Breeds Conservancy: *Turkeys: Narragansett*. North Carolina, USA: http://www.albc-usa.org/cpl/narragansett.html.

46. Marsden SJ: **The Beltsville small white turkey.** *World's Poult Sci J* 2007:32–41.

47. The American Livestock Breeds Conservancy: *Turkeys: Royal Palm*. North Carolina, USA: http://albc-usa.org/cpl/royalpalm.html.

48. Slikas B, Jones IB, Derrickson SR, Fleischer RC: **Phylogenetic relationships of insular Micronesian white-eyes (Aves: Passeriformes: Zosteropidae), based on mitochondrial sequence data.** *Auk* 2000, **117**:355–365.

49. Fleischer RC, Kirchman JJ, Dumbacher JP, Bevier L, Dove C, Rotzel NC, Edwards SV, Lammertink M, Miglia K, Moore SW: **Mid-Pleistocene divergence of Cuban and North American ivory-billed woodpeckers.** *Biology Letters* 2006, **2**:466–469.

50. Stromberg M: **Mosaik Assembler.** In *110014*. Edited by Lee W-P.: Boston College; 2010.

51. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPDP: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**(16):2078–2079.

52. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.** *Nucleic Acids Res* 2010, **38**(16):e164.

53. Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, *et al*: **Ensembl 2011.** *Nucleic Acids Research* 2011, **39**(suppl 1):D800–D806.

54. Felsenstein J: **PHYLIP (Phylogeny Inference Package).** 369th edition. Washington: Department of genome sciences, university of Washington, Seattle; 2005.

55. Nei M: **Genetic distance between populations.** *American Naturalist* 1972, **106**:283–292.

56. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** *Molecular Biology and Evolution* 2011, **10.1093**.

57. Pereira SL, Baker AJ: **A molecular timescale for galliform birds accounting for uncertainty in time estimates and heterogeneity of rates of DNA substitutions across lineages and sites.** *Molecular Phylogenetics and Evolution* 2006, **38**(2):499–509.

58. Hillier LW, Marth GT, Quinlan AR, Dooling D, Fewell G, Barnett D, Fox P, Glasscock JI, Hickenbotham M, Huang W, *et al*: **Whole-genome sequencing and variant discovery in C. elegans.** *Nature Methods* 2008, **5**(2):183–188.

59. Consortium TGP: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**:1061.

60. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, *et al*: **A framework for variation discovery and genotyping using next-generation DNA sequencing data.** *Nature Genetics* 2011, **43**(5):491–498.

61. Rosenberg MS, Subramanian S, Kumar S: **Patterns of transitional mutation biases within and among mammalian genomes.** *Mol Biol Evol* 2003, **20**(6):988–993.

62. Bernardi G: **Isochores and the evolutionary genomics of vertebrates.** *Gene* 2000, **241**(1):3–17.

63. Strachan T, Read AP: **Human Molecular Genetics.** In *An overview of mutation, polymorphism, and DNA repair*. 2nd edition. New York: Wiley-Liss; 1999.

64. Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, Archibald AL, Haley CS, Buys N, Tally M, *et al*: **A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig.** *Nature* 2003, **425**:832–836.

65. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB: **Alternative isoform regulation in human tissue transcriptomes.** *Nature* 2008, **456**(7221):470–476.

66. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, *et al*: **The Sequence of the Human Genome.** *Science* 2001, **291**(5507):1304–1351.

67. Mouse Genome Sequencing Consortium: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**(6915):520–562.

68. Miller W, Makova KD, Nekrutenko A, Hardison RC: **Comparative genomics.** *Annual Review of Genomics and Human Genetics* 2004, **5**:15–56.

69. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, Clamp M, Chang JL, Kulbokas EJ, Zody MC, *et al*: **Genome sequence, comparative analysis and haplotype structure of the domestic dog.** *Nature* 2005, **438**(7069):803–819.

70. Sponenberg DP, Hawes RO, Johnson P, Christman CJ: **Turkey conservation in the United States.** *Animal Genetic Resources Information Bulletin (AGRI)* 2000, **27**:59–66.

71. Frank R, Reese J, Bender M, Sponenberg DP, Williamson D, Beranger J: *Selecting your best turkeys for breeding.* The American Livestock Breeds Conservancy; http://www.albc-usa.org/documents/MasterBreeder/turkeys_chapter1.pdf.

72. Hauser L, Adcock GJ, Smith PJ, Ramírez JHB, Carvalho GR: **Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (***Pagrus auratus***).** *Proc Natl Acad Sci* 2002, **99**:11742–11747.

73. Larson S, Jameson R, Bodkin J, Staedler M, Bentzen P: **Microsatellite DNA and mitochondrial DNA variation in remnant and translocated sea otter (***Enhydra lutris***) populations.** *J Mammal* 2002, **83**:893–906.

74. Nabata D, Masuda R, Takahashi O, Nagata J: **Bottleneck effects on the sika deer Cervus nippon population in Hokkaido, revealed by ancient DNA analysis.** *Zoolog Sci* 2004, **21**:473–481.

75. Axelsson E, Webster MT, Smith NGC, Burt DW, Ellegren H: **Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes.** *Genome Res* 2005, **15**(1):120–125.

76. Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Vereijken A, Megens H-J, Groenen MAM: **A SNP based linkage map of the turkey genome reveals multiple intrachromosomal rearrangements between the turkey and chicken genomes.** *BMC Genomics* 2010, **11**:647.

77. Megens HJ, Crooijmans RP, Bastiaansen JW, Kerstens HH, Coster A, Jalving R, Vereijken A, Silva P, Muir WM, Cheng HH, *et al*: **Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken.** *BMC Genetics* 2009, **10**(86).

78. Ellegren H: **Molecular evolutionary genomics of birds.** *Cytogenetic and Genome Research* 2007, **117**(1–4):120–130.

79. Price EO: **Behavioral development in animals undergoing domestication.** *Applied Animal Behaviour Science* 1999, **65**:245–271.

80. Moller M, Chaudhary R, Hellmén E, Höyheim B, Chowdhary B, Andersson L: **Pigs with the dominant white coat color phenotype carry a duplication of the *KIT* gene encoding the mast/stem cell growth factor receptor.** *Mammalian Genome* 1996, **7**(11):822–830.

81. Fang M, Larson G, Soares Ribeiro H, Li N, Andersson L: **Contrasting Mode of Evolution at a Coat Color Locus in Wild and Domestic Pigs.** *PLoS Genet* 2009, **5**(1):e1000341.

82. Farnir F, Coppieters W, Arranz J-J, Berzi P, Cambisano N, Grisart B, Karim L, Marcq F, Moreau L, Mni M: **Extensive genome-wide linkage disequilibrium in cattle.** *Genome Res* 2000, **10**:220–227.

83. McRae AF, McEwan JC, Dodds KJ, Wilson T, Crawford AM, Slate J: **Linkage disequilibrium in domestic sheep.** *Genetics* 2002, **160**(3):1113–1122.

84. Amaral AJ, Megens H-J, Crooijmans RPMA, Heuven HCM, Groenen MAM: **Linkage disequilibrium decay and haplotype block structure in the Pig.** *Genetics* 2008, **179**(1):569–579.

85. Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, Simianer H: **A genome-wide scan for signatures of recent selection in Holstein cattle.** *Animal Genetics* 2010, **41**:377–389.

86. Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MAM: **Genome-Wide Footprints of Pig Domestication and Selection Revealed through Massive Parallel Sequencing of Pooled DNA.** *PLoS One* 2011, **6**(4):e14782.

87. Lundberg M, Kesson SA, Bensch S: **Characterization of a divergent chromosome region in the willow warbler Phylloscopus trochilus using avian genomic resources.** *J Evol Biol* 2011, **24**:1241–1253.

88. Aslam ML, Bastiaansen JWM, Crooijmans RPMA, Vereijken A, Groenen MAM: **Whole genome QTL mapping for growth, meat quality and breast meat yield traits in turkey.** *BMC Genetics* 2011, **12**:61.

89. Hu Z-L, Reecy JM: **Animal QTLdb: Beyond a Repository - A Public Platform for QTL Comparisons and Integration with Diverse Types of Structural Genomic Information.** *Mammalian Genome* 2007, **18**:1–4.