Title: Commercial chicken breeds exhibit highly divergent patterns of linkage disequilibrium

Reuben J. Pengelly[1], Almas A. Gheyas[2], Richard Kuo[2], Enrico Mossotto[1], Eleanor G. Seaby[1], David W. Burt[2], Sarah Ennis*[1] & Andrew Collins*[1]

1. Human Genetics & Genomic Medicine, Faculty of Medicine, University of Southampton, UK

2. The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, UK

*These authors contributed equally.

Corresponding author:
Prof. Andrew Collins,
Genetic Epidemiology & Genomic Informatics,
Faculty of Medicine,
University of Southampton,
Duthie Building (MP 808),
Tremona Road,
Southampton,
SO16 6YD, UK

a.r.collins@soton.ac.uk
Tel: +44 (0)2381206939

Running title: Divergent LD structure in chickens

Word count for main text: 3950

Abstract

The analysis of linkage disequilibrium (LD) underpins the development of effective genotyping technologies, trait mapping and understanding of biological mechanisms such as those driving recombination and the impact of selection. We apply the Malécot-Morton model of LD to create additive LD maps which describe the high-resolution LD landscape of commercial chickens. We investigated LD in chickens (*Gallus gallus*) at the highest resolution to date for broiler, white egg and brown egg layer commercial lines. There is minimal concordance between breeds of fine scale LD patterns (correlation coefficient < 0.21), and even between discrete broiler lines. Regions of LD breakdown, which may align with recombination hotspots, are enriched near CpG islands and transcription start sites ($p < 2.2x10^{-16}$), consistent with recent evidence described in finches, but concordance in hotspot locations between commercial breeds is only marginally greater than random. As in other birds functional elements in the chicken genome are associated with recombination, but, unlike evidence from other bird species, the LD landscape is not stable in the populations studied. The development of optimal genotyping panels for genome-led selection programmes will depend on careful analysis of the LD structure of each line of interest. Further study is required to fully elucidate the mechanisms underlying highly divergent LD patterns found in commercial chickens.

Key words: Linkage disequilibrium; *Gallus gallus*; recombination hotspots

Introduction

A detailed understanding of linkage disequilibrium (LD) structure is essential for designing single nucleotide polymorphism (SNP) genotyping arrays, successful association mapping of genetic factors underlying traits of interest, establishing mechanisms of genetic recombination and elucidating patterns of selection and population structure. This is particularly true for agricultural species such as commercial chicken (*Gallus gallus*) where LD analysis has the potential to establish the genetic drivers of selection and therefore contribute to further commercial development of lines.

The chicken genome comprises macrochromosomes, intermediate chromosomes and microchromosomes. The macrochromosomes (GGA1-5) span 50-200 Mb, intermediate chromosomes (GGA6-10) range from 20-40 Mb and 28 microchromosomes (GGA11-38) which average ~12 Mb (Hillier et al. 2004; Megens et al. 2009; Schmid et al. 2015). The microchromosomes are characterised as having higher GC content, gene density and much higher recombination rates compared to macrochromosomes (~50-100 kb/cM versus ~300 kb/cM in macrochromosomes). The latter may reflect the requirement for a minimum of at least one chiasma for each chromosome per meiosis and a higher density of cohesin binding sites (Groenen et al. 2009).

Previous studies of LD in the chicken have established that the microchromosomes show reduced LD compared to macrochromosomes, and that these differences are almost completely explained by differences in the recombination rate (Megens et al. 2009). Studies of egg laying chickens indicate higher levels of LD compared to broilers (Heifetz et al. 2005; Andreescu et al.

2007). Despite relatively low levels of LD in broilers, Andreescu *et al.* (2007) determined that there is significant overlap in LD for marker pairs across nine different commercial broiler lines.

Linkage disequilibrium maps constitute the LD analogue of the genetic linkage map and have been extensively utilised for human data (Maniatis et al. 2002; Tapper et al. 2005), and have also been previously applied to agricultural species (Khatkar et al. 2006). We here construct LD maps according to the Malécot-Morton model, using the program LDMAP (Maniatis et al. 2002; Tapper et al. 2005). This model is defined as:

$$\hat{\rho} = (1 - \mathrm{L})\mathrm{M}e^{-\epsilon \mathrm{d}} + \mathrm{L}$$

where $\hat{\rho}$ is the association between SNPs; the asymptote L is the 'background' association between unlinked markers which is increased in small sample sizes and with residual population structure; M reflects association at zero distance with values ~1 consistent with monophyletic origin and <1 with polyphyletic inheritance; ϵ is the rate of LD decline; and d is the physical distance in kilobases between SNPs (Maniatis et al. 2002).

One LD unit ((LDU, equal to ϵd) corresponds to the (highly variable) physical distance over which LD declines to 'background' levels. Compared to the physical map, the LDU scale delimits regions of LD breakdown (which may represent recombination hotspots) as steps, and broader regions of low haplotype diversity (blocks) as plateaus. The overall LDU map length reflects time since an effective population bottleneck (Zhang et al. 2002; Service et al. 2006). Therefore, populations with shorter LDU maps have been founded more recently, experienced a more recent selective sweep, or have a smaller effective population size compared to those with longer maps.

The close correspondence between LD patterns and much lower resolution linkage maps suggest a dominant role for recombination in LD structure. However, unlike linkage maps, which provide a low resolution description of recombination over recent generations, LD maps are constructed from population data and reflect the historical impacts of recombination, mutation, selection and population history. LDU maps of commercial chicken lines may provide new insights into patterns of recombination and selection. Previous studies have begun to describe differences in recombination across *Gallus* genomes based on linkage and LD structure (Megens et al. 2009), and high resolution genome-wide LD maps have the potential to yield further insights.

Birds appear to lack the zinc-finger protein PRDM9 which is required for recombination hotspot localisation in humans and other mammals (Fumasoni et al. 2007; Oliver et al. 2009; Myers et al. 2010). Despite this, recent work by Singhal *et al.,* (2015) has shown that hotspots are highly concordant between wild bird (finch) populations, apparently due to the concentration of recombination events near functional elements of the genome, namely CpG islands and transcription start sites (TSSs).

Here, we construct genome-wide LDU maps for three chicken breeds: broilers (BRO), white egg layers (WEL) and brown egg layers (BEL). We contrast the LD structure across the three breeds, and evaluate the consistency of LD patterns between breeds, potential recombination hotspots and differences between chromosome types and motifs underlying major features of the maps. Additionally, we compare patterns between specific commercial lines (sub-populations) within breeds.

Materials & Methods

Genotypic data used in this work are as reported in the validation populations of Kranis *et al.* (2013), with all genomic coordinates based on the Gallus_gallus-4.0 (galGal4) reference assembly. The samples described by Kranis *et al.* were supplied by Aviagen (broilers), Hy-line International and Lohmann of Synbreed Consortium (the white and brown egg layers), the Pirbright Institute (inbred lines) and the Roslin Institute (RI-J experimental layer line). As described in this paper the variants genotyped were selected to provide an overall uniform distribution of allele frequencies, across the three breed types, best capturing the diversity of each. All of the genotype data used in this study are from individuals that share < 80% identity by similarity genome-wide to minimise population sub-stratification and have > 95% genotyping completeness. Genotype assessment was performed using PLINK v1.07 (Purcell et al. 2007). Multi-dimensional scaling was undertaken using all autosomal markers in order to evaluate the population structure of the samples.

Once the analytical cohorts were defined, SNP marker filtering was undertaken independently for each cohort. Markers with < 95% genotyping completeness, minor allele frequency (MAF) < 0.05 or Hardy-Weinberg equilibrium (HWE) deviation p-value < 0.001 were removed to leave a dataset containing only common, high quality markers. Within each breed the inbreeding coefficient (*F*) (Wright, 1922) was calculated using the PLINK program. LD maps were generated for the assembled autosomes GGA1-28 on filtered data according to the Malécot-Morton model using LDMAP (Morton et al. 2001; Kuo et al. 2007). Where necessary, filtered genotype data were split into ~25,000 marker segments (with 200 marker overlap) to allow for parallelised processing. Overlapping map segments were then trimmed of the terminal 25

markers, and merged to form complete, contiguous whole-chromosome LD maps for the assembled autosomes.

Comparisons of LD maps with existing linkage maps were made. The order of markers in linkage maps from Elferink *et al.* (2010) was revised in line with galGal4 from the native assembly based upon SNP positions within dbSNP (Sherry et al. 2001). Following transition to the galGal4 order, a small number of disordered markers in the linkage map, which distorted the cumulative map, were removed.

To compare map structure between breeds, we focussed on the macrochromosomes GGA1-5. These were chosen to avoid confounding factors such as potentially incomplete reference assemblies, as well as varying recombination rates for the microchromosomes (Hillier et al. 2004; Groenen et al. 2009; Elferink et al. 2010; Schmid et al. 2015). The Spearman's rank correlation of LDU lengths for all 40 kb regions between the three breeds was calculated, after Rubin *et al*. (2010).

For finer-scale interrogation of the LDU length of 5 kb regions, we investigated the concordance seen for these regions with the longest LDU length. We performed these comparisons based upon LDU percentiles, as opposed to absolute LDU lengths, to allow for the differing global LD map lengths for the breeds. This analysis gives an indication of the extent to which narrow regions of intense LD breakdown (which may align with recombination hotspots) are shared between breeds/lines. A high degree of concordance between long LDU segments might suggest a high proportion of shared recombination hotspots between the samples considered. For comparison, we also compared human European and African populations (Pengelly et al. 2015), the two larger BRO2/3 lines, as well as BRO3a/b, which is a random bisection of the largest

population, BRO3. As a final control, a randomised dataset was used for which an equal number of 5 kb regions were randomly selected independently for each dataset, and the concordance calculated; 100 pseudo-replicates were performed for each percentile cutoff.

For comparing LDU decline rates with genome features we focused on the BRO dataset since it has the largest sample size in the dataset. GC content was calculated directly from the reference sequence for 5 kb regions, CpG islands were defined according the UCSC genome browser, and Ensembl annotations were used to define transcripts (Karolchik et al. 2014). BEDTools was used to calculate the distance between elements and regions (Quinlan and Hall 2010).

Results

Input data

The genotype data used in this work are as reported in the validation populations of Kranis *et al.* (2013), with ~1.8 million genotyped SNPs in total. Multidimensional scaling of all samples shows distinct clustering of breeds, though with three distinct population clusters within each breed, consistent with the three genotyped lines (Supplementary Figure S1). For initial LDU map construction, all lines of a breed were pooled. Sample groups used are described in Table 1. For clarity, we use BRO, BEL and WEL to refer to each pooled breed dataset, and BRO2, BRO3 etc. to refer to a distinct line within breed where analysed.

Global map properties

LD maps were generated for all assembled autosomal chromosomes for the three breeds (Supplementary Figure S2). A representative chromosomal LD map for GGA8 is shown in Figure 1. The physical map of the chromosome is represented on the x-axis, while the y-axis shows the LDU maps for each breed and the linkage map in centimorgans (cM) (Elferink et al. 2010). As in human LDU and cM maps, there is a large central region showing little change in LDU or cM, consistent with the location of the submetacentric centromere, where recombination is supressed and there is therefore intense linkage disequilibrium (Tapper et al. 2005; Krasikova et al. 2006). Summary length statistics for all autosomes are shown in Table 2.

LDU map lengths reflect haplotypic diversity within that population (Service et al. 2006; Pengelly et al. 2015) and can be compared with independent measures of population diversity such as *F* inbreeding coefficients. The mean *F* inbreeding coefficients are 0.21, 0.26 and 0.51 for the BRO, BEL and WEL populations respectively, with the greater value for WEL indicating more limited genetic diversity within the population. In comparison, the ratio of LDU/Mb map lengths is also variable between breeds (3.50, 3.18 and 1.74 for BRO, BEL and WEL respectively across the autosomes; Table 2). This ranking of breeds by LDU length is consistent with the trend obtained from the *F* statistic, in line with expectations. There is a general trend towards a lower LDU/cM ratio of map lengths for the smaller chromosomes (Figure 2), with a negative exponential relationship between length of chromosome in megabases and the LDU/cM ratio ($r^2 > 0.75$, $p < 1.7x10^{-9}$ for all breeds).

LD map structure between breeds

To compare map structure between breeds, we focussed on the macrochromosomes GGA1-5. These were chosen because they are known to have more complete reference assemblies

191 (compared to the microchromosomes), and have less variable recombination rates than the
192 microchromosomes (Hillier et al. 2004; Groenen et al. 2009; Elferink et al. 2010; Schmid et al.
193 2015). Following map generation, we interrogated the fine map structure for the breeds to
194 establish the extent to which patterns of LD are conserved between the breeds. There was a
195 weak, though highly significant correlation between breeds of LDU lengths of corresponding 40
196 kb regions (Spearman's $\rho < 0.21$; $p < 2.2x10^{-16}$ for all pairwise comparisons; Figure 3). Narrow
197 regions where there is strong breakdown of LD may align with recombination hotspots, for
198 which there is a high degree of concordance across human populations (Jeffreys et al. 2004;
199 Myers et al. 2005; Tapper et al. 2005; Pengelly et al. 2015). We investigated the extent to which
200 regions with LD breakdown are conserved across the three breeds. In humans, recombination
201 hotspots span just 1-2 kb (Jeffreys et al. 2004). To test concordance of narrow regions of LD
202 breakdown we obtained LDU lengths within 5 kb windows, this window size selected to allow
203 for the incomplete resolution of array based genotyping (Pengelly et al. 2015). We assessed
204 whether regions with the longest LDU lengths within a breed were conserved between breeds.
205 For pairwise comparisons between breeds, there is low concordance in the top LDU length
206 percentile 5 kb windows, with just ~5% concordance between breeds for the top 5th percentile
207 (Figure 4). This is far lower than the concordance seen in humans, as well as that seen for the for
208 BRO3a/3b comparison which is derived from a random split of the BRO3 line.

209 Characteristics of regions of LD breakdown

210 Despite the low concordance of the alignment of narrow regions of LD breakdown for inter-
211 breed comparisons for chickens, the concordance seen is approximately 2-fold greater than that
212 expected by chance. Recombination intensity is known to be dependent on sequence context,
213 therefore hotspots are not expected to be randomly distributed (Groenen et al. 2009; Elferink et

214 al. 2010). One key determinant of recombination rate, GC content (Groenen et al. 2009), was
215 found to be significantly increased in 5 kb regions in the top 1st percentile of LDU length of any
216 breed when compared to regions of zero LDU length in all breeds (42.0% GC *vs.* 39.2%
217 respectively, $p < 2.2x10^{-16}$). We further compared the LDU/kb ratio for 5kb regions with their
218 distance to the nearest CpG islands and TSS and found a highly significant, though weak,
219 negative correlation in our data for both TSS and CpG islands, consistent with previous findings
220 in finches (Figure 5) (Singhal et al. 2015).

221 In order to better characterise the relationship between TSS and CpG islands with narrow regions
222 of LD breakdown we constructed a 2x4 contingency table for 5 kb regions exhibiting limited LD
223 breakdown (<0.003 LDU/kb) or intense LD breakdown (≥ 0.003 LDU/kb) against whether the
224 regions are within 125 kb of a CpG island, TSS, both or neither (Table 3). These values were
225 selected based upon the approximate points of inflection in Figure 5. There was a highly
226 significant deviation from the expected distributions under the null hypothesis ($p = 9.5x10^{-224}$, χ^2
227 test). Regions within 125 kb of both a TSS and CpG island were 1.36 fold more likely to show
228 strong LD breakdown; this is in excess of the additive enrichment for TSS solely and CpG island
229 solely. This suggests that it is an interplay of factors that underlies the localisation of hotspots,
230 with the location of CpG islands having the strongest effect in isolation.

231

232 **Discussion**

233 The analysis of LD maps for the three breeds indicates extensive LD genome-wide. Since one
234 LD unit represents the distance over which LD declines to background levels the genome-wide
235 Mb/LDU ratio gives an indication of the average extent of LD, or 'swept radius'. Average swept

236 radii for the three breeds are 246 kb for BRO, 272 kb for BEL and 488 kb for WEL. In contrast,
237 the corresponding figures for human populations are 55 kb for Europeans and 39 kb for Africans
238 (Lau et al. 2007). Although estimates of swept radii are weakly influenced by SNP marker
239 density (Pengelly et al, 2015) they provide an indication of, as expected, the strong LD found in
240 commercial chicken populations which have been subject to intense selection. However, the
241 profound differences in fine-scale LD structure we have identified are less expected. Although
242 some large-scale genomic features such as centromeric regions, which typically have extensive
243 and intense LD, are shared across breeds for some chromosomes (e.g. Figure 1) there is
244 relatively little concordance in LD structure genome-wide. The contours of the LD maps show
245 many genome regions with widely divergent LD structure (Figure 3) and the overall correlation
246 in LDU lengths of 40 kb windows is low ($\rho = 0.21$). In contrast the fine-scale LD structure
247 between human populations is sufficiently concordant to support a merged 'cosmopolitan' LD
248 map which recovers 91-95% of the information within population-specific maps (Gibson et al.
249 2005).

250 The LDU/cM ratio of chromosome lengths is known to be virtually constant in human
251 populations strongly suggesting that recombination is the primary determinant of LD structure.
252 However, for the three chicken breeds the linear relationship breaks down, with smaller
253 chromosomes having a lower LDU/cM ratio (Figure 2). Similarly, Megens *et al*. (2009) found
254 that recombination rates estimated from LD data were discordant with those obtained from the
255 linkage map. Specifically, they found that the recombination frequency for two
256 microchromosomes (GGA26 and GGA27) estimated from LD was only 2.8 times greater than
257 that of macrochromosomes GGA1 and GGA2) when the expectation from the linkage map was
258 4.5 fold greater recombination on the microchromosomes (Groenen et al. 2009). This

discrepancy was attributed to biases in fitting a model using effective population sizes computed in physical rather than genetic distance windows.

The breakdown in linear correlation between LDU and cM map lengths could have one or more causes. The evidence suggests that the apparent historical recombination intensity (based on LD maps) is lower than current recombination intensity (based on the linkage map) for the smaller chromosomes. One possible explanation for the deviation from a linear relationship between LD and cM lengths is the intense selection that has underpinned the recent population history of these commercial chicken lines. Selection has been previously reported to distort LD based maps of recombination (O'Reilly et al. 2008), due to the expected reduced haplotypic diversity at loci under purifying selection. This leads to an underestimation of recombination in these regions compared to the linkage map. As an extension of this, genes will be, on average, more conserved than intergenic regions and thus genes will have a reduced LDU/cM ratio, on average (Gibson et al. 2013). Since the smaller chicken chromosomes have higher gene density the impact of selection on reducing haplotypic diversity may have been more intense, compared to chromosomes with lower gene density. There is a strong exponential inverse correlation between chromosome size and gene transcripts per megabase ($r^2 = 0.745$, $p = 3.42x10^{-9}$). The differential impact of selection on increased gene density for the smaller chromosomes may therefore at least partly explain the reduced LDU/cM ratio observed.

However, it is important to recognise that the reference genome sequence is incomplete for several chromosomes, particularly the microchromosomes (Schmid et al. 2015). Since the construction of linkage maps requires a lower density markers and measurable linkage extends much further than LD, the construction of a complete linkage map of a chromosome is less sensitive to the omission of small regions with missing or unreliable sequence. In contrast, LD is

much shorter range and LD maps will be truncated in regions where SNP coverage is incomplete due to assembly gaps and poor SNP coverage. This is however unlikely to be the sole explanation for the lower LDU/cM ratios in the smaller chromosomes due to the close negative exponential relationship between the physical chromosome size and LDU/cM ratio, perhaps indicative of an underlying biological mechanism as opposed to solely a technical artefact due to the incomplete assemblies.

Our finding that the LD structure across the three breeds is highly discordant is in marked contrast to comparisons across human populations (Pengelly et al. 2015). Specifically, the pattern seen in humans of narrow regions of LD breakdown, which align with recombination hotspots and are highly concordant across populations, is not observed in chickens which show little concordance of similar regions across breeds. Comparisons between breeds, and even between lines within a breed (BRO2-BRO3, Figure 4) show alignment of such regions that is only slightly greater than 'random'. Concordance within a random split of a subpopulation (BRO3a-BRO3b) is much higher but even then does not show the degree of alignment in the hotspot landscape human European and African populations.

Although the different extent of genome-wide LD between the breeds has been known for some time (Aerts et al. 2007) and the LD pattern between white and brown egg layers has been recognised as clearly different (Qanbari et al. 2010) this is the among first studies to recognise highly divergent fine-scale LD structure between breeds, as well as distinct lines of each breed. This finding has implications for trait mapping since it suggests that to ensure coverage, panels of tagging SNPs would be optimally selected only within the line of interest and that, unlike in human analyses, a 'standard' linkage map may be less useful if it is not representative of the breed-specific recombination landscape.

Analyses of human LD maps have established that the recombination landscape can be recovered from LD structure (Jeffreys et al. 2004; Tapper et al. 2008; Myers et al. 2010). From the derived recombination landscape the chromatin-modifying zinc-finger protein PRDM9 was shown to regulate recombination at 40% of human hotspots by binding to a degenerate 13 base pair motif (Baudat et al. 2010; Myers et al. 2010). Remarkably, despite genomic similarity between humans and chimpanzees, there is virtually no sharing of recombination hotspot locations. Myers *et al.* (2010) found chimpanzee PRDM9 has a dramatically different predicted binding sequence to that of humans. PRDM9 sequences are known to exhibit extremely rapid evolution that explains lack of hotspot conservation in other species that have PRDM9. However, chicken genomes, along with all other avian genomes tested (48 species) appear to lack PRDM9 (Oliver et al. 2009; Singhal et al. 2015). Singhal *et al.* (2015) examined the genomes of the zebra finch (*Taeniopygia guttata*) and long-tailed finch (*Poephia acuticauda*) and found recombination hotspots are enriched near CpG islands and TSSs. In marked contrast to the lack of hotspot sharing between human and chimpanzee they found a high degree of sharing between these birds, which were selected as showing divergence comparable to human and chimpanzee. They argue that whereas the binding specificity of PRDM9 mediated recombination leads to rapid turnover and evolution of hotspots, association with functional genomic features, such as CpG islands and TSSs, leads to stasis in which there is conservation of hotspot locations over, potentially, 10s of millions of years. However, in contrast, the evidence presented here suggests very limited sharing of regions of LD breakdown for the three chicken breeds based on LD structure.

We have established that the relationship between TSS and CpG islands with regions of LD breakdown/ recombination hotspots, as identified in finches by Singhal *et al.* (2015) is also found

328 in chickens. It is possible that, whilst recombination in chickens is associated with these
329 functional elements, this relationship has not resulted in convergence to the point where
330 recombination hotspots are recognisably stable between lines. Perhaps the population history of
331 the commercial lines studied has not been sufficiently long to allow for convergence of LD
332 patterns. Furthermore, it is possible that small effective population sizes of these breeds also
333 contributes to the observed limited convergence in LD structure. Further work, including the
334 studying of outbred, wild *Gallus* populations, as well as whole-genome sequence based LD
335 mapping would allow for the further elucidation of mechanisms underlying the architecture of
336 patterns of LD in chickens (Gheyas et al. 2015). Improvements to the reference assembly
337 currently underway will provide more complete assemblies for the microchromosomes (Schmid
338 et al. 2015), allowing for further investigation into differential properties of the chromosomes. It
339 would be informative to extend these investigations to other agricultural species, in order to
340 determine whether this effect is specific to *Gallus*, or can also be observed in other selected
341 avian, piscine and mammalian species.

342 LD maps have been shown to be of use in refining the order and orientation of contigs during
343 compilation of reference genomes where there is ambiguity, which may be of utility for enabling
344 further improvements to the assembly (Ennis et al. 2001). LD maps, as described herein, are
345 readily produced for any population for which genotype data is available, without requiring
346 pedigree data. Given the high level of discordance in LD patterns between lines, data derived
347 from the line of interest would provide far greater precision when utilised for trait mapping and
348 other studies, as well as the optimisation of genotyping arrays to allow for maximal coverage for
349 the minimal number of SNP markers for each line.

350

Acknowledgements

This work is supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC) 'Sparking Impact' scheme and by the University of Southampton, Faculty of Medicine Doctoral Training Fund. The authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work.

Conflict of Interest

The authors have no conflict of interests to declare.

Data Archiving

Genotype data used herein are available in the Dryad Digital Repository, http://dx.doi.org/10.5061/dryad.48gp0.

References

Aerts J, Megens HJ, Veenendaal T, Ovcharenko I, Crooijmans R, Gordon L, Stubbs L, Groenen M. 2007. Extent of linkage disequilibrium in chicken. Cytogenet. Genome Res. 117:338–345.

Andreescu C, Avendano S, Brown SR, Hassen A, Lamont SJ, Dekkers JCM. 2007. Linkage disequilibrium in related breeding lines of chickens. Genetics 177:2161–2169.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. Science 327:836–840.

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. 2001. High-resolutions haplotype structure in the human genome. Nat Genet 29(2):229-232.

Elferink MG, van As P, Veenendaal T, Crooijmans RPMA, Groenen MAM. 2010. Regional differences in recombination hotspots between two chicken populations. BMC Genet. 11:11.

Ennis S, Collins A, Tapper W, Murray A, MacPherson JN, Morton NE. 2001. Allelic association discriminates draft orders. Ann. Hum. Genet. 65:503–504.

Fumasoni I, Meani N, Rambaldi D, Scafetta G, Alcalay M, Ciccarelli FD. 2007. Family expansion and gene rearrangements contributed to the functional specialization of PRDM genes in vertebrates. BMC Evol. Biol. 7:187.

Gheyas AA, Boschiero C, Eory L, Ralph H, Kuo R, Woolliams JA, Burt DW. 2015. Functional classification of 15 million SNPs detected from diverse chicken populations. DNA Res. 22:205–217.

Gibson J, Tapper W, Ennis S, Collins A. 2013. Exome-based linkage disequilibrium maps of individual genes: functional clustering and relationship to disease. Hum. Genet. 132:233–243.

Gibson J, Tapper W, Zhang W, Morton N, Collins A. 2005. Cosmopolitan linkage disequilibrium maps. Hum. Genomics 2:20–27.

385 Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens HJ, Crooijmans RPMA, Besnier F, Lathrop
386 M, Muir WM, Wong GKS, et al. 2009. A high-density SNP-based linkage map of the chicken
387 genome reveals sequence features correlated with recombination rate. Genome Res. 19:510–519.

388 Heifetz EM, Fulton JE, O'Sullivan N, Zhao H, Dekkers JCM, Soller M. 2005. Extent and
389 consistency across generations of linkage disequilibrium in commercial layer chicken breeding
390 populations. Genetics 171:1173–1181.

391 Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MA,
392 Delany ME, et al. 2004. Sequence and comparative analysis of the chicken genome provide unique
393 perspectives on vertebrate evolution. Nature 432:695–716.

394 Jeffreys AJ, Holloway JK, Kauppi L, May CA, Neumann R, Slingsby MT, Webb AJ. 2004. Meiotic
395 recombination hot spots and human DNA diversity. Philos. Trans. R. Soc. Lond. B. Biol. Sci.
396 359:141–152.

397 Karolchik D, Barber GP, Casper J, Clawson H, Cline MS, Diekhans M, Dreszer TR, Fujita PA,
398 Guruvadoo L, Haeussler M, et al. 2014. The UCSC Genome Browser database: 2014 update.
399 Nucleic Acids Res. 42:D764–D770.

400 Khatkar MS, Thomson PC, Tammen I, Cavanagh J AL, Nicholas FW, Raadsma HW. 2006. Linkage
401 disequilibrium on chromosome 6 in Australian Holstein-Friesian cattle. Genet. Sel. Evol. 38:463.

402 Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, Talbot R, Pirani A, Brew F, Kaiser P, et al.
403 2013. Development of a high density 600K SNP genotyping array for chicken. BMC Genomics
404 14:59.

405 Krasikova A, Deryusheva S, Galkina S, Kurganova A, Evteev A, Gaginskaya E. 2006. On the positions of
406 centromeres in chicken lampbrush chromosomes. Chromosom. Res. 14:777–789.

407 Kuo T-Y, Lau W, Collins AR. 2007. LDMAP: the construction of high-resolution linkage disequilibrium

maps of the human genome. Methods Mol. Biol. 376:47–57.

Lau W, Kuo TY, Tapper W, Cox S, Collins A. 2007. Exploiting large scale computing to construct high resolution linkage disequilibrium maps of the human genome. Bioinformatics 23:517–519.

Maniatis N, Collins A, Xu CF, McCarthy LC, Hewett DR, Tapper W, Ennis S, Ke X, Morton NE. 2002. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. Proc. Natl. Acad. Sci. U. S. A. 99:2228–2233.

Megens H-J, Crooijmans RPMA, Bastiaansen JWM, Kerstens HHD, Coster A, Jalving R, Vereijken A, Silva P, Muir WM, Cheng HH, et al. 2009. Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken. BMC Genet. 10:86.

Morton NE, Zhang W, Taillon-Miller P, Ennis S, Kwok PY, Collins A. 2001. The optimal measure of allelic association. Proc. Natl. Acad. Sci. U. S. A. 98:5217–5221.

Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of recombination rates and hotspots across the human genome. Science 310:321–324.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. Science 327:876–879.

O'Reilly PF, Birney E, Balding DJ. 2008. Confounding between recombination and selection, and the Ped/Pop method for detecting selection. Genome Res. 18:1304–1313.

Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS Genet. 5.

Pengelly RJ, Tapper W, Gibson J, Knut M, Tearle R, Collins A, Ennis S. 2015. Whole genome sequences are required to fully resolve the linkage disequilibrium structure of human populations. BMC

Genomics 16:666.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Qanbari S, Hansen M, Weigend S, Preisinger R, Simianer H. 2010. Linkage disequilibrium reveals different demographic history in egg laying chickens. BMC Genet. 11:103.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

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

Schmid M, Smith J, Burt DW, Aken BL, Antin PB, Archibald AL, Ashwell C, Blackshear PJ, Boschiero C, Brown CT, et al. 2015. Third Report on Chicken Genes and Chromosomes 2015. Cytogenet. Genome Res. 145:78–179.

Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G, Ospina J, Ruiz-Linares A, Macedo A, Palha JA, et al. 2006. Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. Nat. Genet. 38:556–560.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 29:308–311.

Singhal S, Leffler E, Sannareddy K, Turner I, Venn O, Hooper D, Strand A, Li Q, Raney B, Balakrishnan C, et al. 2015. Stable recombination hotspots in birds. Science 350:928–932.

Tapper W, Collins A, Gibson J, Maniatis N, Ennis S, Morton NE. 2005. A map of the human genome in linkage disequilibrium units. Proc. Natl. Acad. Sci. U. S. A. 102:11835–11839.

Tapper W, Gibson J, Morton NE, Collins A. 2008. A comparison of methods to detect recombination hotspots. Hum. Hered. 66:157–169.

Wright S. 1922. Coefficients of Inbreeding and Relationship. Am. Nat. 56:330.

Zhang W, Collins A, Maniatis N, Tapper W, Morton NE. 2002. Properties of linkage disequilibrium (LD) maps. Proc. Natl. Acad. Sci. U. S. A. 99:17004–17007.

Tables

Table 1: Counts of chickens and SNPs remaining post-filtering.

		BRO	BRO2[d]	BRO3	BRO3a	BRO3b	BEL	WEL
Founders	Males	58	-	50	26	24	12	8
	Females	17	-	9	4	5	40	38
	Total	123[d]	48	59	30	29	52	46
SNPs	Raw count	966355	790531	789359	692467	778135	891200	691954
	MAF[a]	833639	658548	638947	631449	645713	796430	627294
	HWE[b]	760893	789450	788284	691625	777771	763931	420130
	Missingness[c]	966346	788594	787732	690038	776114	888903	690298
	Final count	630435	655905	636535	628382	643554	667605	354737

[a]Markers with minor allele frequency < 0.05 within the cohort excluded. [b]Markers with a Hardy-Weinberg equilibrium deviation p-value < 0.001 within the cohort. [c]Markers with > 5% data missing excluded. [d]Sex data unavailable for BRO2 line.

468 **Table 2**: Physical, linkage, and LDU map lengths for the three breeds.

Chr	Length (Mb)	Linkage cM	BRO LDU	BRO LDU/cM	BEL LDU	BEL LDU/cM	WEL LDU	WEL LDU/cM
1	195.2	413.5	713.7	1.7	612.4	1.5	363.8	0.9
2	148.8	281.3	452.1	1.6	462.8	1.6	207.8	0.7
3	110.4	236.9	439.2	1.9	359.4	1.5	186.3	0.8
4	90.2	195.2	309.8	1.6	277.1	1.4	164.1	0.8
5	59.5	154.4	198.6	1.3	207.1	1.3	126.3	0.8
6	34.9	93.8	171.9	1.8	114.7	1.2	81.4	0.9
7	36.2	103.1	150.3	1.5	153.1	1.5	76.3	0.7
8	28.7	96.6	134.5	1.4	112.2	1.2	57.0	0.6
9	23.4	88.1	123.4	1.4	101.4	1.2	61.0	0.7
10	19.9	80.6	97.4	1.2	77.0	1.0	65.3	0.8
11	19.3	64.0	73.9	1.2	85.4	1.3	34.5	0.5
12	19.9	69.1	101.4	1.5	88.4	1.3	51.6	0.7
13	17.7	62.7	76.3	1.2	85.7	1.4	38.5	0.6
14	15.1	67.4	65.9	1.0	68.1	1.0	46.2	0.7
15	12.6	53.6	61.6	1.1	54.6	1.0	24.8	0.5
16	0.5	59.1	2.0	0.0	2.3	0.0	1.1	0.0
17	10.3	50.9	65.7	1.3	52.6	1.0	24.5	0.5
18	11.2	51.7	56.4	1.1	53.9	1.0	31.1	0.6
19	10.0	52.3	67.9	1.3	57.6	1.1	31.3	0.6
20	14.2	55.1	61.5	1.1	58.6	1.1	31.1	0.6
21	6.8	56.9	56.8	1.0	51.1	0.9	29.8	0.5
22	4.1	56.4	21.9	0.4	21.7	0.4	9.5	0.2
23	5.7	52.3	46.6	0.9	39.2	0.7	27.6	0.5
24	6.2	53.2	57.6	1.1	47.6	0.9	33.0	0.6
25	2.1	57.1	17.8	0.3	23.8	0.4	6.8	0.1
26	4.9	52.3	37.4	0.7	39.5	0.8	21.9	0.4
27	5.2	51.0	32.6	0.6	34.3	0.7	28.2	0.6
28	4.7	53.6	30.4	0.6	37.1	0.7	20.9	0.4
Total	**917.7**	**2762.2**	**3724.6**	**1.3**	**3378.7**	**1.2**	**1881.7**	**0.7**

469

470

Table 3: Odds ratios for recombination hotspot regions (LDU/kb ≥ 0.003) being within 125 kb of genomic features

Feature	Odds ratio
Neither CpG nor TSS	1.00
CpG only	1.12
TSS only	1.06
Both CpG and TSS	1.37

Titles and Legends to Figures

Figure 1: Linkage and LD maps of GGA8 for all breeds. The broadly analogous structure of the linkage map and LD maps for the three populations can be seen. All maps contain a large plateau around 10,000 kb, corresponding to the centromere region. Overall length of the LD maps is inversely related to the strength of LD within a breed. Broilers show the lowest LD overall reflecting relatively high haplotype diversity while white egg layers show strongest LD and lowest population haplotype diversity. Note that there is no expectation that cM and LDU maps be similar lengths, as the LDU map length is dependent upon population diversity.

Figure 2: Relationship between physical chromosome length and LDU/cM ratio for all autosomes in the three breeds. There is a clear trend for the physically smaller chromosomes to exhibit lower LDU/cM ratios, with a negative exponential relationship. Lines indicate best fit for $\log_{10}$(length) *vs.* LDU/cM.

Figure 3: Comparison of LD breakdown intensity on GGA2 for the three breeds. The LDU/kb ratio is shown for sliding 40 kb windows. A common region of strong LD is shown at ~50 Mb, corresponding to the centromere. There are minimal other trends apparent in the localisation of LD intensities between breeds ($\rho < 0.21$; $p < 2.2 \times 10^{-16}$ for all pairwise comparisons).

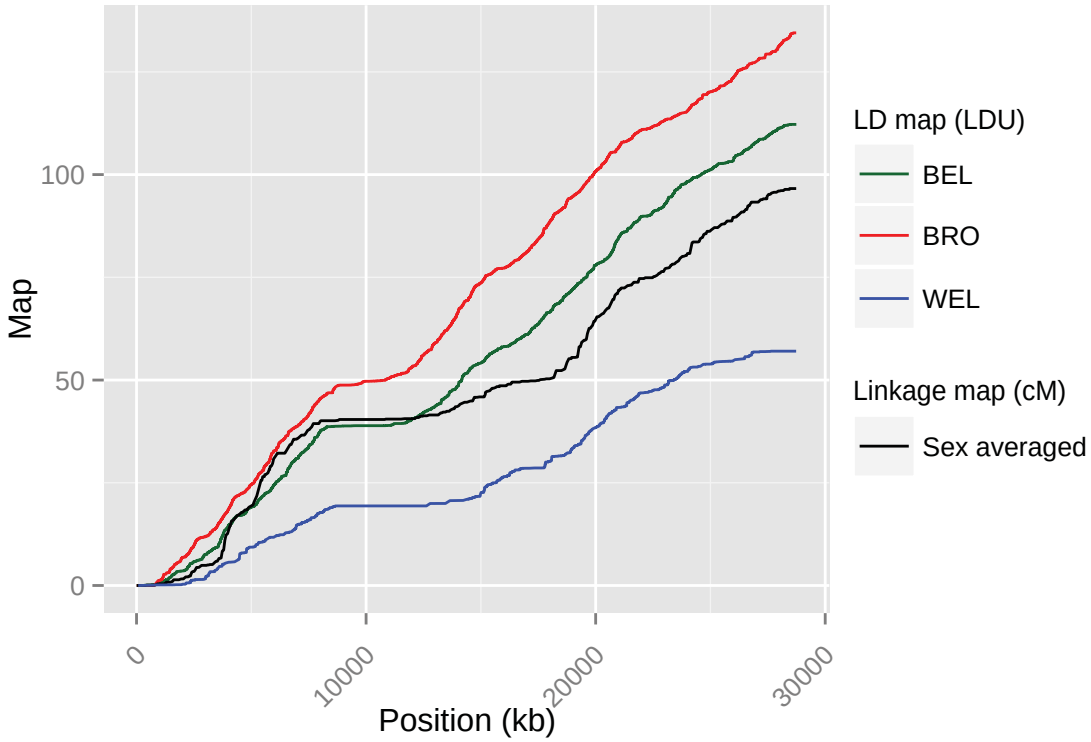
Figure 4: Pairwise concordance of regions of LD breakdown (indicative of recombination hotspots) between populations. Shown is the proportion of regions in the top *n* percentile which intersect between the breeds. For pairwise comparisons between BRO/BEL/WEL ~5% of

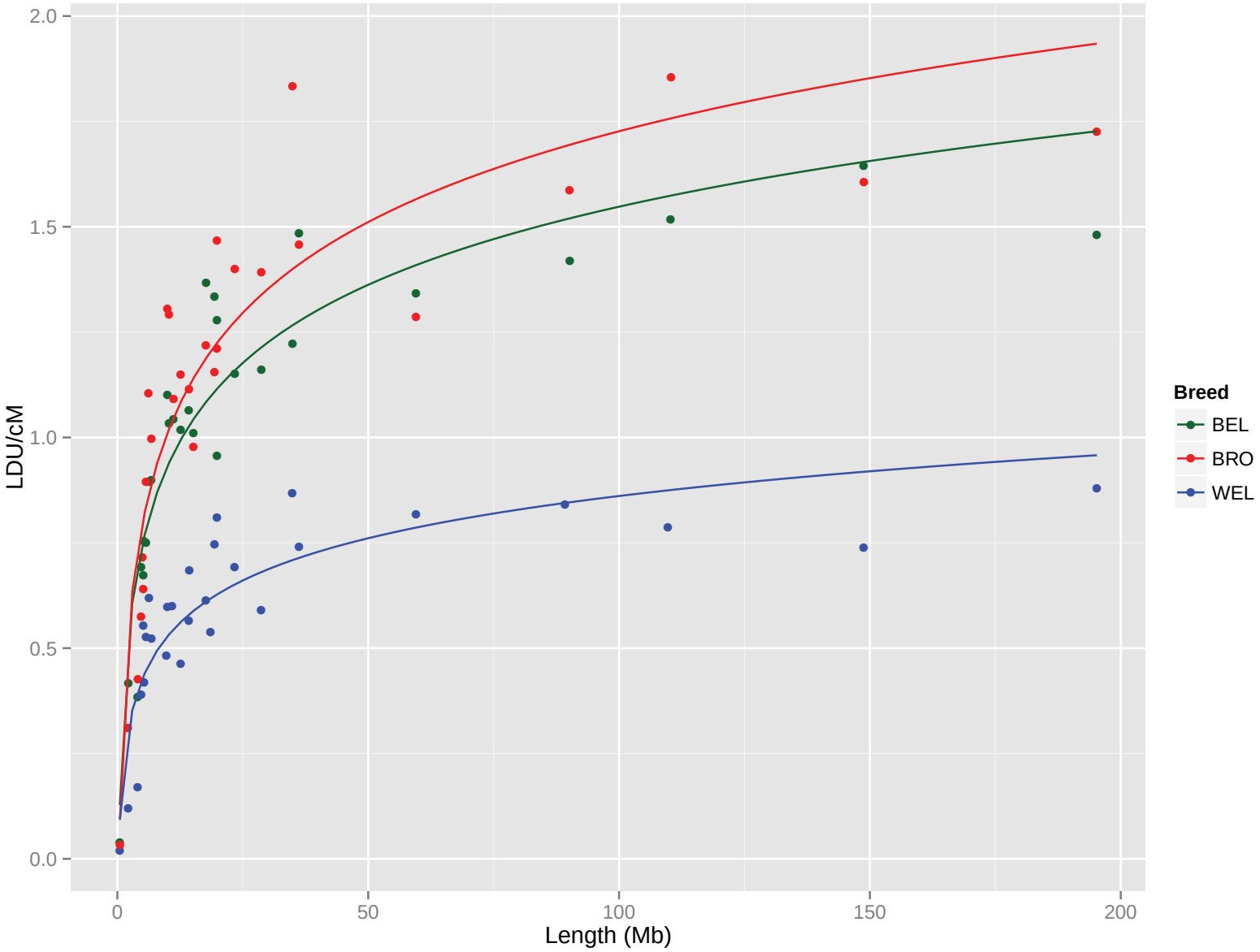
496 regions in the top 5 percentile are concordant. This proportion is also similar where two separate
497 BRO lines (BRO2/BRO3) are compared. The low level of concordance between regions of LD
498 breakdown, even between lines within breed, suggests there may be marked differences in
499 haplotype structure since regions of LD breakdown align with increased haplotype diversity
500 (Daly et al, 2001). When the largest population is bisected (BRO3a/b), this proportion is ~27%,
501 still far lower than the equivalent comparison between European and African human populations
502 (Pengelly et al. 2015). All comparisons show a greater proportion of concordance than the
503 randomised dataset, showing that the patterns are not wholly stochastic. Shaded regions indicate
504 95% confidence intervals of the trendlines, concordance was calculated for 0.1% increments.
505 **Figure 5**: Association of LD breakdown with displacement from nearest functional element,
506 namely CpG islands and TSS. There is a strong negative correlation between the distance from
507 the functional elements and LDU/kb (ρ= -0.12 for CpG islands, ρ= -0.10 for TSS, $p < 2.2x10^{-16}$
508 for each comparison). Shown is the mean LDU/kb ratio for 5 kb bins, shaded area indicated 95%
509 confidence interval.
510

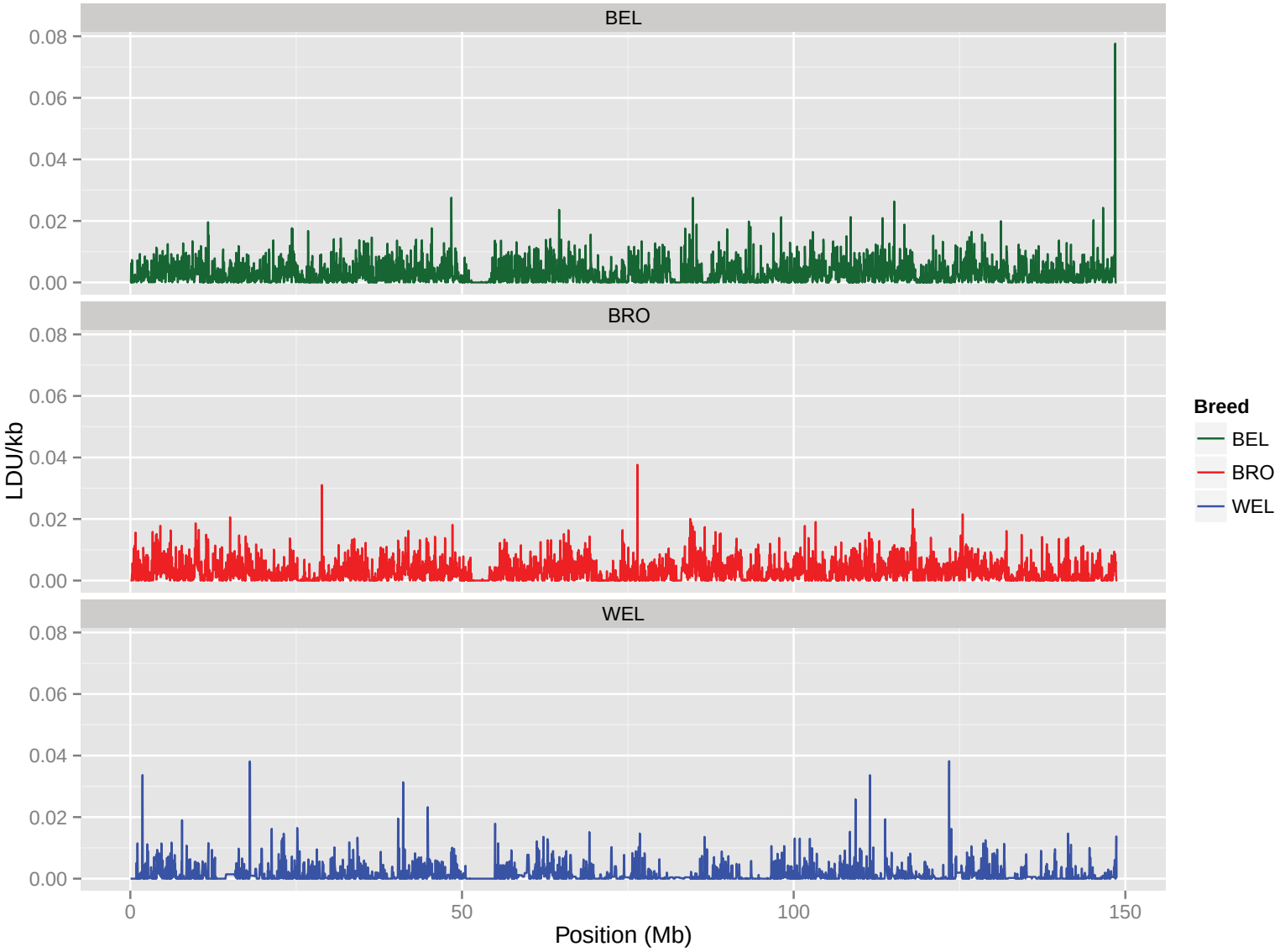
Supplementary files

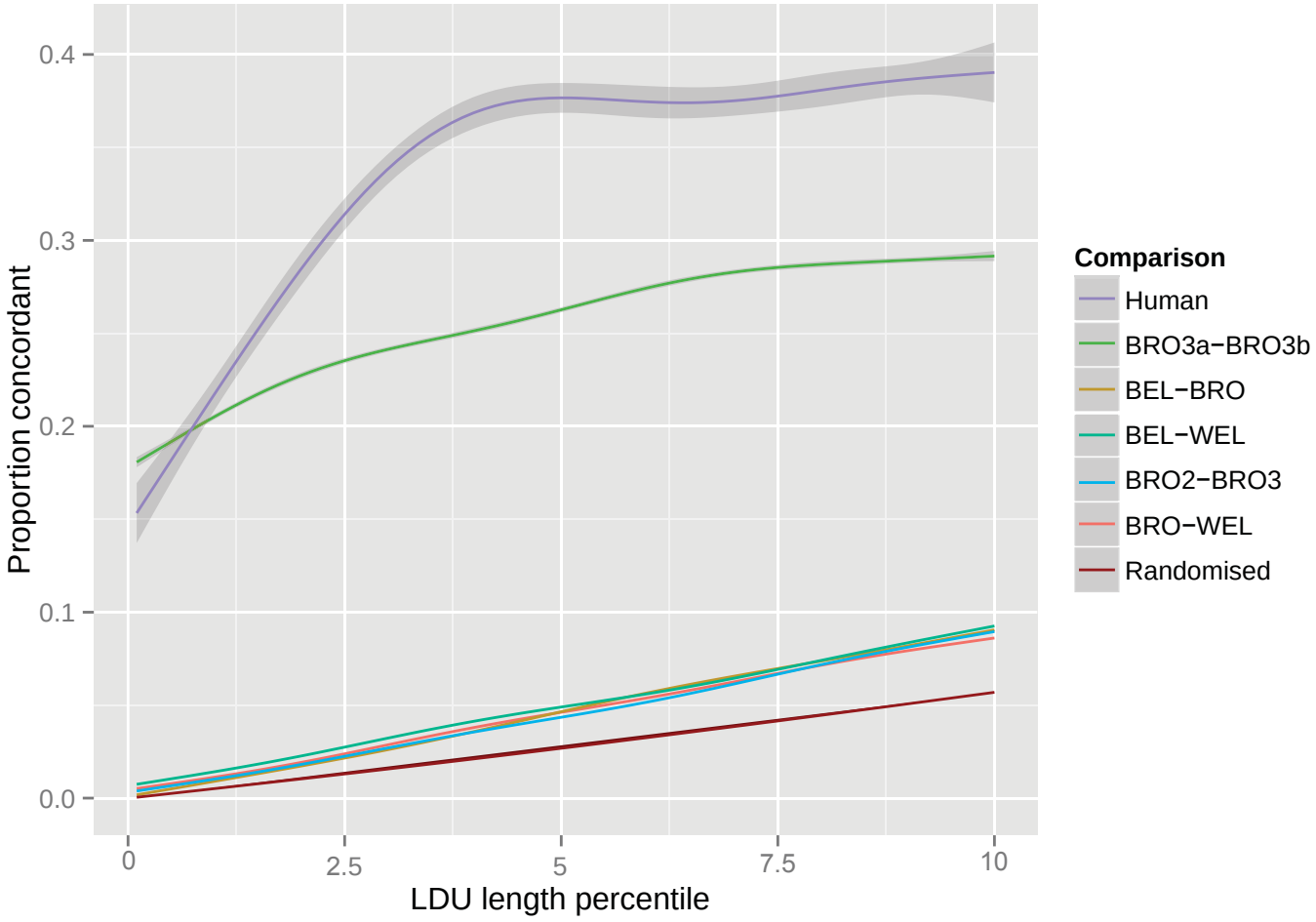
Supplementary Figure S1: Multidimensional scaling analysis of founder chickens. Chicken lines coarsely cluster within breeds, with three clusters for each breed apparent. Population cluster designations are labelled on the plot.

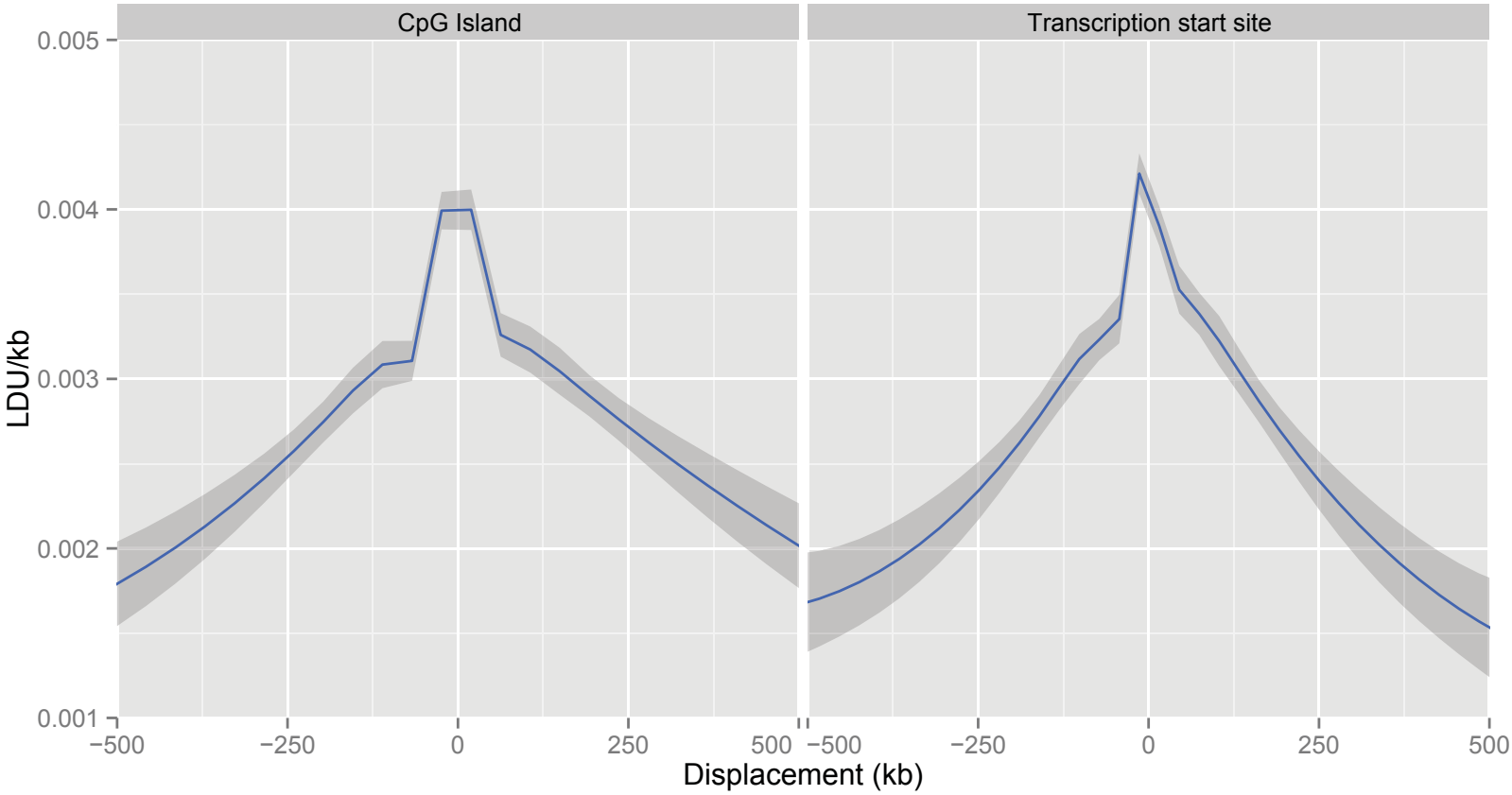
Supplementary Figure S2: Linkage and LD maps of autosomes plotted for all breeds. The broadly analogous structure of the linkage map and LD maps for the three populations can be seen.


Map
100
50
0
0
10000
20000
30000
Position (kb)
LD map (LDU)
BEL
BRO
WEL
Linkage map (cM)
Sex averaged


2.0
1.5
1.0
0.5
0.0
LDU/cM
0
50
100
150
200
Length (Mb)
Breed
BEL
BRO
WEL


BEL
BRO
WEL
0.08
0.06
0.04
0.02
0.00
LDU/kb
0
50
100
150
Position (Mb)
Breed
BEL
BRO
WEL


0.4
0.3
0.2
0.1
0.0
Proportion concordant
0
2.5
5
7.5
10
LDU length percentile
Comparison
Human
BRO3a−BRO3b
BEL−BRO
BEL−WEL
BRO2−BRO3
BRO−WEL
Randomised


CpG Island
Transcription start site
0.005
0.004
0.003
0.002
0.001
LDU/kb
−500
−250
0
250
500
Displacement (kb)