Look Over the Horizon - Chicken Linkage Disequilibrium is Much More Complex Over Much Longer Distance than Previously Appreciated

Ehud Lipkin (ehud.lipkin@mail.huji.ac.il)
Hebrew University of Jerusalem
Janet E. Fulton
Hy-Line (United States)
Jacqueline Smith
University of Edinburgh
David W. Burt
University of Edinburgh
Morris Soller
Hebrew University of Jerusalem

Research Article

Keywords: Chicken, long-range linkage disequilibrium, QTL, F6, LD blocks

Posted Date: June 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-598396/v1

Abstract

Background

Appreciable Linkage Disequilibrium (LD) is commonly found between pairs of loci close to one another, decreasing rapidly with distance between the loci. This provides the basis studies to map Quantitative Trait Loci Regions (QTLRs), where it is custom to assume that the closest sites to a significant markers are the prime candidate to be the causative mutation. Nevertheless, Long-Range LD (LRLD) can also be found among well-separated sites. LD blocks are runs of genomic sites all having appreciable LD with one another. High LD and LRLD are often separated by genomic sites with which they have practically no LD. Thus, not only can LD be found among distant loci, but also its pattern may be complex, comprised of fragmented blocks. Here, chicken LRLD and LD blocks, and their relationship with previously described Marek's Disease (MD) QTLRs, were studied in an F_6 population from a full-sib advanced intercross line, and in eight commercial pure layer lines. Genome wide LRLD was studied in the F_6 population by random samples of non-syntenic and syntenic marker pairs. To illustrate the relationship with QTLRs, LRLD and LD blocks in and between the MD QTLRs were studied by all possible marker pairs.

Results

LRLD was defined as $r^2 \geq 0.7$ over a distance ≥ 1 Mb, and 1.5% of all syntenic marker pairs were classified as LRLD. Complex fragmented and interdigitated LD blocks were found, ranging over distances from a few hundred to a few millions bases. Vast high, long-range, and complex LD was found between two of the MD QTLRs. Cross QTLRs STRING networks and gene interactions suggested possible origins of the exceptional LD between these two QTLRs.

Conclusions

All sites with high LD with a significant marker should be considered as candidate for the causative mutation, but, unlike the custom assumption, the causative mutation is not necessarily the one closest to the significant marker. Rather, the present results show that it can be located at a much larger distance from a significant marker than previously appreciated, beyond closer mutations. Thus, LRLD range and LD block complexity must be accounted for while interpreting genetic mapping studies.

Background

Linkage disequilibrium (LD) refers to correlations among alleles of different genomic sites. It quantifies the informativity between different sites [1, 2, 3]. Useful LD indicate non-random association of alleles at different loci [3]. Appreciable LD is commonly found between pairs of loci close to one another, and LD decreases rapidly as distance between the loci increases [e.g., 3, 4]. This provides the basis for Genome Wide Association Studies (GWAS) to map quantitative trait loci (QTL), where it is custom to assume that the closest sites to a significant markers are the prime candidate to be the causative mutation.

Nevertheless, during any one-time snapshot of a population, long-range LD (LRLD) can also be found among loci that are well separated from one another, over millions of bp [3, 4, 5, 6, 7, 8, 9]. Rarely, LD above background level can be found between non-syntenic markers on different chromosomes (chr) [4].

LRLD could just be a matter of sampling variation, especially in small populations [7]. Alternatively, LRLD could be a result of genome assembly errors where SNP locations are misidentified, and thus LRLD may help identify such assembly errors [10]. This phenomenon may, however, also have genuine biological origins, such as co-evolution of genomic sites (coding and non-coding genes, long- and short-range regulatory sites), gene conversion, copy number variation, or demographic factors such as selection, population bottlenecks, nonrandom mating, and epistasis.

LD blocks are runs of genomic sites all having appreciable LD with one another. However, high LD in general and LRLD in particular are not always continuous. Rather, they are often separated by genomic sites with which they have practically no LD

[11, 3, 12, 13]. Thus, not only can LD be found among distant loci, but also its pattern may be complex, comprised of fragmented blocks.

LRLD and LD complexity present concerns for GWAS mapping, as a significant association nay be found between a causative locus and markers far removed from it, thus falsely placing the putative causative locus at a site far away from its actual location [7]. On the other hand, LRLD may point to interactions between unlinked regions in the genome (e.g., a receptor and its ligand or a gene and its regulator). Furthermore, LRLD can identify co-evolution of different genomic regions affected by the same selection, natural or artificial.

The objectives of this study were to characterize LRLD and LD blocks in multiple Hy-Line chicken lines previously used to map quantitative trait loci regions (QTLRs) for MD resistance [14]. This will give a genomic view of the LD complexity, and illustrate its importance to mapping results by assessing the relationship between LRLD and LD blocks and MD QTLRs.

Results

Remapping QTLRs from Galgal4 to Galgal6. The new coordinates of the markers on Galgal6 and association results obtained by Smith et al. (2020) [14] were used to remap the MD QTLRs identified with Galgal4 by Smith et al. (2020) [14]. The same 38 QTLRs were found on each genome build (Table 1). Most changes were negligible, except one movement of a fragment on Chr 1 over about 70 Mb from QTLR 1 to QTLR 4, including a QTLR lncRNA tested by Smith et al. (2020) [14]. This change is detailed in the Appendix. The new QTLR coordinates on Galgal6 were used for the LD analyses in this study.

Linkage disequilibrium in the F_6 population based on 600K genotyping SNP array

Non-syntenic random LD. A total of 923,183 random non-syntenic pairs of markers from different autosomes were used to assess the background level of LD, with an average of 184,636.6 pairs in a family (Table 2). LD averaged 0.011 ± 0.016, comparable to previous reports in chicken [12, 15,], but about ten times higher than values reported in mammals; horse [5), sheep [4], and cattle [16, 17] populations. These differences may represent experimental design, or population sample, size, history and structure or biological differences between birds and mammals.

Mean family LDs and Standard Deviations (SDs) had significant high negative correlation with the size of the population sample (i.e., Ind/pair), r = -0.895 (P = 0.040) and r = -0.884 (P = 0.046). These correlations are also in accord with our previous report in chicken [12], and the expectation of Sved (1971) [18].

With a mean LD of 0.011 ± 0.016, any $r^2 > 0.043$ is above the background LD. Indeed, combining all five F_6 families together, only 3.5% of the r^2 values were above 0.05 (Additional Table 1). A single high LD of $r^2 = 0.991$ was found in Family 2. Without any replication, this was treated as a sampling effect. Based on these results, a conservative critical LD value of $r^2 \geq 0.15$ was chosen for defining significant LD.

Syntenic random LD. A total of 1,008,823 random syntenic marker pairs were used to assess the level of random LDs on the same chromosome (Table 3). Distance between markers in the random pairs varied from 11 to 197,038,449 bp, with an average of 28,976,195.6 bp. Means of r^2 were all in close range around 0.11, averaging 0.114, ten times the means obtained for the non-syntenic LD. Though obtained by random marker pairs, some of which are at a long distance from one another, these means suggest the presence of large number of LDs above the background LD of 0.011 (Table 2). The expected negative correlation between distance and LD was again obtained in all five families (Table 3).

In all families, about two-thirds of the LD values were up to 0.05, dropping rapidly thereafter (Table 4). Interesting, for all families, there was an increase in the range of $r^2 > 0.85$, suggesting existence of large high-LD blocks. Pooled over all families, the proportion of $r^2 \geq 0.15$, set conservatively as a threshold of significance by the non-syntenic LD, was almost 0.2 (Table 4), while less than 5% of the LD values were above 0.7. Hence, the range of $0.15 \leq r^2 < 0.7$ was set as low to moderate LD and used to define moderate LD blocks, and $r^2 \geq 0.7$ was set as high LD and used to define LRLD and high LD blocks.

Long-range LD (LRLD)

Estimating LRLD by random samples of syntenic marker pairs. Pooled over all families, 418,075 pairs had a distance above 20 Mb; as expected, no high LD of $r^2 \geq 0.7$ was found beyond 20 Mb (Additional Table 2).

Detailed inspection of the distances up to 20 Mb showed that all high LDs were in fact within 10 Mb (Additional Table 3).

Pooled over all families together, a total of 50,100 random marker pairs qualified within the LRLD definition, namely $r^2 \geq 0.7$ over a distance ≥ 1 Mb. These LRLDs constitute 30.9% of all pairs within 20 Mb, and 1.5% of the total number of syntenic pairs tested (Additional Table 2).

Among the syntenic pairs, 0.016 had $r^2 > 0.95$, almost 15-times the proportion of the single LD value in this range (0.000001) found among the non-syntenic pairs (Additional Table 1). Thus, the proportion of syntenic high LD was not negligible. LRLDs were distributed over all autosomes in all five F_6 families (Figure 1; Additional Table 4; LD matrices in figshare portal). No LRLD was found on Chromosomes 22 in any of the families.

Though these LRLDs were obtained by random sampling of marker pairs, repeated similar locations of marker pairs suggest the existence of many LRLD blocks. This was indeed found by the LD analysis of the MD QTLRs (see below).

F_6 MD QTLRs and random LRLD. To check for a possible relationship between the LRLDs found here and the F_6 MD QTLRs mapped in the same population (Table 1), LRLDs and QTLRs were aligned together (e.g., Figure 1 and LD matrices in figshare portal), and overlaps were counted (Additional Table 5).

As noted above, with all markers in an interval less than 1 Mb, no LRLD could be found on Chr 16 (Additional Table 4); hence, QTLR 32 was not included in any further analyses. Of the remaining 37 QTLRs, overlaps between 28 QTLRs and LRLDs were found in all families (the non-zeros under 'Families' in Additional Table 5). It seems remarkable that, even though only 1.5% of the random LD values were LDLR, no less than 75.7% of the mapped MD QTLRs overlapped LRLDs. Then again, in Galgal6, QTLRs averaged 1.4 Mb (Table 1), and random LRLDs averaged 2.2 Mb, from 1 to above 12 Mb (Additional Table 6). Thus, such overlap may not be so surprising, but a result of the abundance and size of the QTLRs and LRLDs.

Zooming in on QTLRs clearly showed the overlap between the LRLDs and QTLRs (Figure 2). Not only was LRLD found within QTLRs, but LRLD was found between QTLRs 4 and 5 in all 5 families. The similar locations seen in Figure 2 suggest the presence of LD blocks shared by both QTLRs.

LD in the QTLRs in the F_6 families

The overlaps found in the F_6 families between random LRLDs and the MD QTLRs, led us to examine in more detail the LRLD and LD blocks in these QTLRs, with all informative markers of the five F_6 families (note that this part used *all* pairs of informative markers in the QTLRs, and not only a *sample* of random pairs as in the first LD analysis).

Chromosomes 1, 2, 4, 5, 6 and 14, harbored more than one QTLR (Table 1), thus enabling examination of LD in and between QTLRs. In each F_6 family, Affymetrix SNP array genotypes were used to calculate LD between all possible pairs of all markers in the 21 QTLRs on those chromosomes.

Hundreds of thousands of LRLDs were found in and between the tested QTLRs (Additional Table 7). Total number of marker pairs ranged from below 8 to above 10 million in a family, to a total of more than 43 million pairs. Of these, pooled over all families, 830,182 were LRLDs (62,103 - 227,015 LRLDs in a family). These constitute 0.7 - 2.6% of all pairs in a family, a total of 1.9%, higher than the 1.5% found among the random pairs over all autosomes (Additional Table 3).

A total of 161,832 LRLDs were found between QTLRs (Additional Table 7), 19.5% of all LRLDs found (0.6 - 24.9 % among the families).

Family 5 is an outlier in Additional Table 7, with a much lower number and proportion of total LRLDs and LRLDs across QTLRs compared to the other four families. Further inspection did not identify any source of this difference. Hence, we have no explanation other than sampling variation.

Pooling all families together, LRLDs were found in all 6 chromosomes examined (Table 5). No LRLD could be found *in* the QTLRs on Chromosomes 5 or 6 (Table 5), as no QTLR there was larger than 1 Mb (Table 1). However, LRLDs *between* QTLRs were also found in those two chromosomes.

In all families, LRLDs were found between most pairs of QTLRs (Additional Table 8 a-f). Exceptional among all pairs of QTLRs, an extremely large number of LRLDs (159,413) was found between QTLRs 4 and 5 on Chr 1 in all families, confirming the results of the random samples (Figure 2). The tight LD between these two QTLRs was further confirmed by the LD blocks (below).

Thus, repeating in all F_6 families, LRLDs were found to be frequent, distributing within and between QTLRs in all chromosomes tested.

QTLR LD blocks

LD Blocks in the F_6 QTLR

As shown by the data a complicated LD pattern was found in the F_6 QTLR. Large, fragmented, and interdigitated LD blocks were found in all five families over all six chromosomes examined (LD matrices in figshare portal). The range and complexity would have been even larger if moderate LD blocks were included, with $0.15 \leq r^2 < 0.7$.

An example of fragmented interdigitated blocks is presented in Figure 3a. Close examination of the LD found in Family 1 in this region shows the presence of 3 high LD blocks, all fragmented and all interdigitated with one another: Block 1 includes markers with ID numbers 134-141, 143, 145-149, and 151; Block 2 includes markers 142, 144 and 152; Block 3 includes markers 150 and 337. The fact that, despite their apparent fragmentation, these are indeed genuine blocks is shown in Figures 3 b-d. If the markers in Blocks 2 and 3 were not included in the analysis, (e.g., because they were not on the SNP array or were filtered out by the quality control or were not polymorphic in this family), then three clear unambiguous blocks would have been identified.

Note the distance between the markers in block 3, is above 0.5 Mb. Should the criterion of 0.25 Mb [6] been used, this block would be defined as LRLD.

Blocks shared by QTLRs 4 and 5 in the F_6 families. In accordance with the random sampling of marker pairs and LRLDs in and between QTLRs in the F_6, large and long-range LD blocks were shared by QTLRs 4 and 5 in all five families, as exemplified in Additional Figure 1 and detailed in Additional Tables 8 a-f. In Additional Figure 1, the high LD block distributed from the first marker of QTLR 4 to close to the end of QTLR 5, over 5.7 Mb, with 412 markers included. Considering moderate LD of $0.15 \leq r^2 < 0.7$, would stretch the block all the way to the end of QTLR 5, over more than 7.1 Mb. Thus, the exceptional LD between QTLRs 4 and 5 indicated by the random sample of pairs was confirmed in all F_6 families by both LRLDs and LD blocks between QTLRs.

LD among QTLR elements in the eight pure lines

LD of elements within and between the F_6 QTLRs was further examined within eight Hy-Line elite pure lines. Complex LD blocks between elements within and across QTLRs were found, similar to that found in the F_6 families, over distances from a few bp to a few Mb (Figure 4-6; all LD matrices are in figshare portal).

LD within one QTLR gene. Figure 4 present an example of LD blocks within the QTLR gene *TRANK1* in Line WL1. Despite the short distances (390 bp to 14.5 Kb), a complex pattern was found, with 2 LD blocks, one of which is fragmented around the

other. There was high to complete LD between markers 5, 8 and 13-36. These markers had practically no LD with markers 11 and 12, which were in complete LD with one another. Thus, in the gene *TRANK1* in Line WL1, Block 1 starts before, but ends after Block 2. The association test P values [14] completely matched the LD blocks, with the same or close P values in each block. This match was found in all other combinations of QTLR - line (Figures 5 - 7).

LD between QTLR elements. An example of a more complex LD pattern with interdigitated blocks is shown in Figure 5, this time across QTLR elements (3 lncRNAs).

Careful inspection of Figure 5 shows 2 interdigitated blocks: Block 1 includes Markers 6-8, 15, and 19-20; Block 2 comprise of Markers 11-14, 16-17 and 30-33. Thus, the high LD Block extend over the 3 QTLR lncRNAs. The middle lncRNA05 is split among the 2 blocks. Some of the markers are in LD with upstream lncRNA02, while other markers of the same lncRNA05 form a block with the downstream lncRNA02. The 2 groups of lncRNA05 are interdigitated. That is, Markers 6-8 of lncRNA02 are in the same block with 2 separate regions in the next lncRNA05 - Markers 15 and then 19-20 but not with the other markers in the same lncRNA; Markers 12-14 and 19-20 of lncRNA05 are in LD with all 4 markers of lncRNA04. It would be interesting to find out what are the sources of such complex LD patterns.

LD was found between other types of QTLR elements as well. Figure 6 present such LD between the QTLR genes *TLR4* and *BRINP1* in QTLR 33 on Chr 17. The first marker of *TLR4* has high LD to the first 2 markers of *BRINP1*, and the 3 markers are not linked to other markers of their own gene. The other 6 markers of *TLR4* form a tight LD block. Complexing it even further, the last marker of *BRINP1* (Marker 15) had low to moderate LD with all markers in QTLR 33, both genes included.

LD between QTLRs 4 and 5. Markers on both QTLRs 4 and 5 were informative only in Lines WL3, WPR1, WPR2, and RIR1, up to only 4 markers in a line in QTLR 4 (Lines' LD matrices in figshare portal). Thus, information on the LD between the QTLRs was limited in this dataset. Nevertheless, in accord with the random LDs in the F_6 families (Figures 1 and 2) and cross QTLRs LRLD in these families (Additional Tables 8 b-f), moderate LD blocks among elements in these QTLRs crossed their boundaries in Lines WPR1, WPR2 and RIR1. In Line WRP1, 2 clear high LD blocks were found, one in each QTLR (Figure 7). However, the 2 QTLR blocks had moderate LDs of $r^2 = 0.478$ among them, thus forming one moderate LD block. Note that the distances between the cross QTLR pairs, varied from 5.135 to 5.138 Mb.

Looking for a source of such vast, high, and complex long-range LD between QTLRs 4 and 5, a bioinformatics search found 10 and 68 genes in QTLRs 4 and 5, respectively (Figure 8 and Table 6). STRING network analysis revealed five networks of 2 to 28 genes. Two of the networks ('Net' 2 and 3 in Table 6), are comprised of genes from both QTLRs (Figure 8 and Table 6). Of 'Net' 2, the 2 genes in QTLR 4 and 17 of the 26 genes in QTLR 5 are located in the LD blocks extending over the two QTLRs found in F_6 ('+' in the column 'B4-5' in Table 6). Both genes in 'Net' 3 are in those blocks. Finally, the two networks with genes from both QTLRs included 6 genes interacting with a gene from another QTLR (Figure 8), all of which located in the cross QTLRs LD blocks. The gene networks and interactions shared by both QTLRs could be the origin of the LD between QTLRs 4 and 5. In fact, the phenomenon of genes whose products work together tending to be on the same chromosomal region is quite common. For example, the Major Histocompatibility Complex (MHC) on chicken chromosome 16 and the Regulators of Complement Activation cluster (RCA) on chromosome 26 [19, 20, 11]. In fact, the networks presented in Figure 8 is a good examples for this colocation of genes working together.

Discussion

Chicken LD over a range of distances, and patterns of LD blocks, were studied in five F_6 families from a Full Sib Advanced Intercross Line (FSAIL), and eight commercial pure layer lines, thus allowing to study the repetition of the results. LRLD was studied in the F_6 population by random non-syntenic and syntenic samples of marker pairs genotyped by the Affymetrix HD SNP array. In face of the LRLD results, and to illustrate the importance of LRLD to QTL mapping results, LRLD and LD blocks were studied with all possible marker pairs of all markers in previously described MD QTLRs [14].

This study started with SNP location information from the previous chicken genome build Galgal4, and was subsequently updated to the Galgal6 assembly. This change necessitated remapping of the QTLRs described in Smith et al. (2020) [14], resulting in negligible changes of most QTLR coordinates. Nevertheless, the change of genome versions moved a segment of Galgal4 QTLR 1 (Chr 1) to Galgal6 QTLR 4 (Chr 1) (Appendix). The moved segment included one QTLR lncRNA described in Smith et al. (2020) [14], thus emphasizing the importance of basing genomic analyses on the most updated genome version. These results also present the power of LD to identify mapping errors, as already noted by Utsunomiya et al. (2016) [10].

Long-range LD (LRLD) was defined as $r^2 \geq 0.7$ over a distance ≥ 1 Mb. These criteria are more restricted than previously used [6, 22]. Nevertheless, repeated appreciable numbers of LRLDs within chromosomes were found repeated in all five F_6 families by the random sampling of syntenic marker pairs, far above the numbers of high LD found between non-syntenic markers from different chromosomes. The LRLDs were further found in all five F_6 families by all QTLR array markers on chromosomes with more than one QTLR. These results could be an underestimate, as the F_6 population was designed to fragment the genome for high-resolution QTL mapping [23, 24].

High LD blocks were defined as a group of markers located on the same chromosome, having $r^2 \geq 0.7$ with each other, even if markers with low LD appeared between them. This definition allowed "a look over the horizon" and identification of complex blocks. The phenomenon of fragmented and interdigitated LD blocks were repeatedly found in all five F_6 families and in all eight pure lines over a vast range of distances, from hundreds of bp to mega bases. The FSAIL population was composed of five families, and they showed similar results. Strength of this analysis was that it repeated five times. Then the same phenomenon was seen in the eight elite lines, further adding strength to the validity of these results, which also agree with previous studies from us and others [e.g., 3, 11, 12, 13].

A strong linkage was found between QTLRs 4 and 5. LRLD between them was found while analyzing the F_6 random samples of SNPs within all autosomes. High LD blocks were found in all five F_6 families, comprised of markers from both QTLRs. Moderate LD blocks between QTLRs 4 and 5 were also found in 3 of the 8 pure lines by QTLR elements' markers. These results raise the question of what elements on both QTLRs are in high LD over such large distances. Of course, being MD QTLRs, the LD between the QTLRs could be a result of a co-selection for MD resistance. But what make these two QTLRs so different from all other pairs of QTLRs? Why is their LD so exceptional?

To answer this, we looked at the gene content of QTLRs 4 and 5. Ten and 68 genes were found in QTLRs 4 and 5, respectively. STRING protein network analysis revealed five gene networks. Two of the networks include genes from both QTLRs, most of which are located within the LD blocks extending over the two QTLRs found in the F_6. All 6 genes interacting with a gene from another QTLR are in the cross QTLRs LD blocks. Obviously, the shared gene networks and interactions could be the origin of the LD between QTLRs 4 and 5. However, assessing the uniqueness of the LD between the QTLRs necessitates further study on the distribution of cross QTLR networks and interactions among other pairs of QTLRs with less LD among them. Furthermore, assessing the real effect of the gene networks and interactions needs more molecular, quantitative and population studies. All of these are beyond the scope of the present study.

In general, the complex LD found in this study could stem from technical reasons such as sampling variation. It can also be a result of mapping errors, as was indeed found in this study for regions on chromosome 1 with build 4 (Appendix). However, it could have genuine biological meaning, through processes such as co-evolution of genomic sites (as a result natural or artificial selection), gene conversion, copy number variation, population bottlenecks, non-random mating, and epistasis. The repeatability in different analyses, different datasets and different populations, and the agreement with previous reports strengthen the case for the present results as being a genuine biological phenomenon.

Conclusions

The observed LRLD and fragmented interdigitated LD blocks imply that the causative element is not necessarily the closest, or even close at all to the significant marker, and maybe not even to the significant block of markers. Thus, mapping results and

searches for causative elements must consider the complexity of the LD.

Methods

Populations

Nine populations described by Smith et al. (2020) [14] were again used in the present [14] study. These comprised five families of F_6 birds from an FSAIL used by Smith et al. (2020) [14] to map QTLRs affecting MD resistance, and eight pure lines used in the same study to test these QTLRs.

A priory, it is expected that LRLD in F_6 will be at higher frequency than in pure lines, because the families start with only 4 chromosomes each, and there are only a few generation of intercrossing to break up the haplotype blocks. However, this study did not compare populations, families or lines. Rather, it aim to present the phenomena of LD long range and complexity.

Trait

The trait for which these QTLR are associated is resistance to the avian oncogenic alpha herpes virus, Marek's Disease (MD virus [14]. This trait association data set was used, as the phenotype and genotype information was available. It is used as an illustration for the QTLR and LD associations that were identified.

Genotypes

Only the autosomal genotypes as used by Smith et al. (2020) [14] to map and test QTLRs were used in the present study. Genotypes on the Z chromosome will be analyzed in detail in a different manuscript. Genotypes were obtained from the HD 600K Affymetrix SNP chicken array [25] in the F_6 population, and marker genotypes used to test the QTLRs by the eight pure lines [14]. The only difference was that instead of a minimum MAF ≥ 0.01 used for the association tests by Smith et al. (2020) [14], a threshold of 0.10 was used here for the LD analysis, to avoid spurious high LD due to rare alleles [7]. However, to test LD with exactly the same markers used for association tests in the eight pure lines, the threshold of MAF ≥ 0.01 was also used for LD for these lines.

Genome assemblies and remapping QTLRs

As described in Smith et al. (2020) [14], the initial analysis in this study was based on the Galgal4 genome build. The Lift Genome Annotations tool [26] within the UCSC Browser was used to remap markers from Galgal4 to Galgal6 (GRCg6a; Acc. No.: GCA_000002315.5). The new coordinates were then used to remap the F_6 QTLRs as was done by Smith et al. (2020) [14]. The results were very similar on both assemblies, and hence only the results from Galgal6 will be presented here. Nevertheless, there was one change worth noting, detailed in the Appendix.

Linkage disequilibrium (LD)

LD Measure. LD r^2 within each F_6 family and each pure line were obtained using JMP Genomics software (JMP Genomics, Version 9, SAS Institute Inc., Cary, NC, USA, 1989–2019).

Non-syntenic LD. Background LD over all autosomes was estimated utilizing Affymetrix 600K genotypes of two random combined samples with return of non-syntenic marker pairs from each of the five F_6 families.

Long-Range LD (LRLD). Koch et al. (2013) [6] used all pairs of SNPs on each human chromosome and defined LRLD between haplotype-blocks rather than between SNP pairs. However, due to computer limitations, we used samples of random SNP pairs on the Affymetrix 600K SNP array to assess LRLD in the autosomes of the five F_6 families. Koch et al. (2013) [6] defined LRLD in human as high LD (their low p_D) over a distance ≥ 0.25 cM (≈ 0.25 Mb). Vallejo et al. (2018) [22] used a very relaxed LD threshold of $r^2 > 0.25$ in rainbow trout studies, but a larger minimum distance of 1.0 Mb. Conservatively, and based on the present results, in this study we defined LRLD as a marker pair with $r^2 \geq 0.7$ over a distance ≥ 1.0 Mb. As noted in the

Introduction, LRLD between elements in a QTLR and across QTLRs can indicate a relationship between the element and between the QTLRs. Hence, to illustrate its importance, LRLD in and between MD QTLRs [14] was studied by all F_6 Affymetrix genotypes in the QTLRs.

F_6 random LRLD and QTLRs. F_6 LRLD of random marker pairs were aligned by chromosomal location with the F_6 MD QTLRs [14], and all overlaps between the LRLDs and the QTLRs were counted.

LD blocks. High LD blocks were defined as a group of markers located on the same chromosome having $r^2 \geq 0.7$ with each other. The definition was applied even if markers with low LD appeared between the LD markers. This definition allowed a "look over the horizon" and identification of fragmented and interdigitated blocks.

Gene network analysis

To investigate the genes underlying QTLRs 4 and 5, the BioMart tool within the Ensembl database (https://www.ensembl.org/info/data/biomart/index.html) was used to identify genes in these regions. These identified genes were then subject to network analysis using the STRING database (v11) [27], which provides an overview of known protein interactions.

Abbreviations

chr: chromosome

FSAIL: Full Sib Advanced Intercross Line

GWAS: Genome Wide Association Studies

LD: Linkage Disequilibrium

LRLD: Long-Range LD

MD: Marek's Disease

QTL: Quantitative Trait Locus

QTLR: QTL Region

SD: standard deviation

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of Data: Data have been submitted to the European Nucleotide Archive (ENA) at EMBL-EBI under study accession numbers PRJEB39142 (WGS) and PRJEB39361 (RNAseq). LD matrices are available at figshare.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Biotechnology and Biological Sciences Research Council (grant number BB/K006916/1).

Authors' contributions

EL conceived of and designed the study, performed the LD analyses, generated the figures and tables, and wrote the manuscript. JS performed the STRING analysis and generated Figure 8. All authors wrote, read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Reich D, Cargill M, Bolk S. Ireland J, Sabeti PC, et al. Linkage disequilibrium in the human genome. Nature;411:199–204. https://doi.org/10.1038/35075590.
2. Ardlie K, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet*. 2002;3:299–309 https://doi.org/10.1038/nrg777.
3. Aerts J, Megens HJ, Veenendaal T, Ovcharenko I, Crooijmans R, Gordon L, et al. Extent of linkage disequilibrium in chicken. Cytogenet Genome Res. 2007;117(1-4):338-45. doi: 10.1159/000103196. PMID: 17675876.
4. García-Gámez E, Sahana, G, Gutiérrez-Gil, B, Arranz, JJ. (2012). Linkage disequilibrium and inbreeding estimation in Spanish Churra sheep. BMC genetics;13: 43. https://doi.org/10.1186/1471-2156-13-43.
5. Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. Anim Genet. 2010;41 Suppl 2:8-15. doi: 10.1111/j.1365-2052.2010.02092.x. PMID: 21070270.
6. Koch E, Ristroph M, Kirkpatrick M. Long Range Linkage Disequilibrium across the Human Genome. PLOS ONE 2013;8(12): e80754. https://doi.org/10.1371/journal.pone.0080754.
7. Skelly DA., Magwene PM., Stone EA. Sporadic, global linkage disequilibrium between unlinked segregating sites. Genetics, 2016;202(2):427–437. https://doi.org/10.1534/genetics.115.177816.
8. Park L. Population-specific long-range linkage disequilibrium in the human genome and its influence on identifying common disease variants. Sci Rep. 2019;9:11380. https://doi.org/10.1038/s41598-019-47832-y.
9. Peters SO, Kızılkaya K, Ibeagha-Awemu EM, Sinecen M, Zhao X. Comparative accuracies of genetic values predicted for economically important milk traits, genome-wide association, and linkage disequilibrium patterns of Canadian Holstein cows. J Dairy Sci. 2021;104(2):1900-1916. DOI: 10.3168/jds.2020-18489.
10. Utsunomiya ATH, Santos DJA, Boison SA. Utsunomiya YT, Milanesi M, Bickhart DM, et al. Revealing misassembled segments in the bovine reference genome by high resolution linkage disequilibrium scan. BMC Genomics. 2016;17:705. https://doi.org/10.1186/s12864-016-3049-8.
11. Allabi AC, Horsmans Y, Issaoui B, Gala JL. Single nucleotide polymorphisms of ABCB1 (MDR1) gene and distinct haplotype profile in a West Black African population. Eur J Clin Pharmacol. 2005;61(2):97-102.
12. Lipkin E, Dolezal M, Bagnato A, Fulton J, Settar P, Arango J, et al. Linkage Disequilibrium and Haplotype Blocks in Cattle and Chicken Populations. Proc. Plant & Animal Genomes XXI Conf. 2013: P0680.
13. Pérez O'Brien AM, Utsunomiya YT, Mészáros G. Bickhart DM, Liu GE, et al. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. Genet. Sel. Evol. 2014;46:19. https://doi.org/10.1186/1297-9686-46-19.
14. Smith J, Lipkin E, Soller M, Fulton JE, Burt DW. Mapping QTL Associated with Resistance to Avian Oncogenic Marek's Disease Virus (MDV) Reveals Major Candidate Genes and Variants. Genes 2020;

11:1019. https://doi.org/10.3390/genes11091019.

15. Seo D, Lee DH, Choi N, Sudrajad P, Lee SH, Lee JH. Estimation of linkage disequilibrium and analysis of genetic diversity in Korean chicken lines. PLOS ONE. 2018;13(2):e0192063. doi: 10.1371/journal.pone.0192063. PMID: 29425208; PMCID: PMC5806858.
16. Khatkar MS., Nicholas FW., Collins AR. Zenger KR, Cavanagh JAL, Barris W, et al. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics. 2008;9:187. https://doi.org/10.1186/1471-2164-9-187.
17. Sargolzaei M, Schenkel FS, Jansen GB, Schaeffer LR. Extent of Linkage Disequilibrium in Holstein Cattle in North America, J Dairy Sci. 2008;91:2106-17. https://doi.org/10.3168/jds.2007-0553.
18. Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. Theor. Pop. Bio. 1971;2: 125–41.
19. Oshiumi H, Shida K, Goitsuka R, Kimura Y, Katoh J, Ohba S, et al. Regulator of complement activation (RCA) locus in chicken: identification of chicken RCA gene cluster and functional RCA proteins. J Immunol. 2005;175(3):1724-34. doi: 10.4049/jimmunol.175.3.1724. PMID: 16034113.
20. Hosomichi K, Miller MM, Goto RM, Wang Y, Suzuki S, Kulski JK, et al. Contribution of mutation, recombination, and gene conversion to chicken MHC-B haplotype diversity. J Immunol. 2008;181(5):3393-9. doi: 10.4049/jimmunol.181.5.3393. Erratum in: J Immunol. 2010;184(9):5415. PMID: 18714011; PMCID: PMC2657362.
21. Michalak P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics. 2008;91(3):243-8. doi: 10.1016/j.ygeno.2007.11.002. Epub 2007 Dec 21. PMID: 18082363.
22. Vallejo RL, Silva RMO, Evenhuis JP, Gao G, Liu S, Parsons JE, et al. Accurate genomic predictions for BCWD resistance in rainbow trout are achieved using low-density SNP panels: Evidence that long-range LD is a major contributing factor. J Anim Breed Genet. 2018;135(4):263–74. https://doi.org/10.1111/jbg.12335.
23. Heifetz EM, Fulton JE, O'Sullivan NP, Arthur JA, Wang J, Dekkers JC, et al. Mapping quantitative trait loci affecting susceptibility to Marek's disease virus in a backcross population of layer chickens. Genetics. 2007;177: 2417–31.
24. Heifetz EM, Fulton JE, O'Sullivan NP, Arthur J, Cheng H, Wang J, et al. Mapping QTL affecting resistance to Marek's disease in an F6 advanced intercross population of commercial layer chickens. *BMC* Genom. 2009;10:20.
25. Kranis A, Gheyas AA, Boschiero C, Turner F, Yu L, Smith S, et al. Development of a high density 600K SNP genotyping array for chicken. BMC Genom. 2013;14:59. [CrossRef]
26. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019;8;47(D1):D853-D858. PMID: 30407534; PMC: PMC6323953
27. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8–a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009;37(Database issue):D412-6. doi: 10.1093/nar/gkn760. Epub 2008 Oct 21. PMID: 18940858; PMCID: PMC2686466.

Tables

Table 1. Remapping Galgal4 QTLRs on Galgal6.

QTLR	Chr	Gagal4			Galgal6		
		Start	End	Length	Start	End	Length
1	1	8,854,589	9,261,317	406,729	9,510,148	9,902,036	391,889
2	1	13,268,810	14,294,877	1,026,068	13,994,599	14,950,768	956,170
3	1	52,012,381	52,517,953	505,573	52,166,588	52,643,244	476,657
4	1	71,738,977	73,085,956	1,346,980	71,892,917	73,277,481	1,384,565
5	1	75,288,912	77,765,499	2,476,588	75,513,671	79,029,197	3,515,527
6	1	91,482,501	91,803,475	320,975	93,533,567	93,853,587	320,021
7	1	101,761,945	104,477,713	2,715,769	103,738,415	106,416,920	2,678,506
8	1	109,395,871	110,913,451	1,517,581	111,372,866	112,400,685	1,027,820
9	1	169,656,958	172,228,091	2,571,134	171,680,812	174,306,953	2,626,142
10	1	174,394,951	175,634,386	1,239,436	176,474,702	177,748,402	1,273,701
11	1	194,193,118	194,788,548	595,431	196,152,404	196,750,875	598,472
12	2	15,960	883,257	867,298	48,621	959,053	910,433
13	2	46,423,161	46,868,789	445,629	45,786,534	46,247,754	461,221
14	2	105,493,833	108,988,920	3,495,088	105,791,822	109,334,178	3,542,357
15	2	125,168,753	127,089,304	1,920,552	125,532,963	127,219,187	1,686,225
16	2	138,767,830	139,701,774	933,945	139,198,404	140,160,087	961,684
17	3	108,220,206	109,260,649	1,040,444	108,593,746	109,643,999	1,050,254
18	4	8,308,498	11,268,107	2,959,610	8,328,709	11,309,259	2,980,551
19	4	84,393,173	88,579,400	4,186,228	84,829,085	89,057,374	4,228,290
20	5	7,568,851	8,147,837	578,987	8,388,371	8,967,466	579,096
21	5	18,806,924	19,673,354	866,431	19,753,005	20,610,009	857,005
22	6	2,077,640	2,709,412	631,773	3,323,132	3,946,659	623,528
23	6	29,536,109	29,817,337	281,229	30,954,349	31,233,344	278,996
24	6	31,006,769	31,448,342	441,574	32,440,880	32,888,648	447,769
25	7	13,062,871	16,436,053	3,373,183	13,563,779	16,986,311	3,422,533
26	10	22,643	1,713,384	1,690,742	1,025,523	2,668,959	1,643,437
27	11	7,397,790	8,440,259	1,042,470	7,912,510	8,959,749	1,047,240
28	12	8,996,686	9,432,693	436,008	9,414,714	9,845,036	430,323
29	13	10,363,430	12,176,727	1,813,298	11,756,937	13,566,822	1,809,886
30	14	8,085,563	9,335,685	1,250,123	8,499,374	9,745,708	1,246,335
31	14	13,138,194	15,087,518	1,949,325	13,542,085	15,384,231	1,842,147
32	16	1,630	490,907	489,278	1,852,095	2,669,032	816,938
33	17	3,442,598	5,634,042	2,191,445	3,808,082	5,932,858	2,124,777

QTLR	Chr	Gagal4			Galgal6		
		Start	End	Length	Start	End	Length
34	18	3,196,488	4,093,129	896,642	3,221,049	4,118,252	897,204
35	24	4,489,675	5,514,833	1,025,159	4,160,414	5,498,172	1,337,759
36	26	4,378,168	5,036,699	658,532	4,438,584	5,002,302	563,719
37	27	1,540,112	2,270,461	730,350	3,930,559	4,689,821	759,263
38	28	1,282,726	1,571,011	288,286	1,447,725	1,687,264	239,540

QTLR, ordinal number of the QTLR [14]; Chr, chromosome; Start/End, QTLR coordinates of the first and last SNP in the QTLR; Length, size of the QTLR (bp).

Table 2. Summary statistics of non-syntenic random LD between 600K markers in the F_6 families.

Family	Pairs	Ind/ pair	r^2			
			Avg	SD	Min	Max
1	190,232	176.6	0.014	0.019	0.000	0.368
2	182,366	231.7	0.012	0.016	0.000	0.991
3	184,399	354.0	0.008	0.011	0.000	0.372
4	187,406	219.5	0.010	0.014	0.000	0.246
5	178,780	200.7	0.012	0.016	0.000	0.243
All	184,636.6	236.3	0.011	0.016	0.000	0.443

Pairs, number of pairs r^2 values obtained; Ind/pairs, average number of individuals used to calculate a markers pair LD in a family; r^2: Avg, average; SD, standard deviation; Min, minimum; Max, maximum; All: Pairs, average number of pairs in all families combined; Ind/pairs r^2, means weighted by the number of Pairs.

Table 3. Summary statistics of syntenic random LD between 600K markers in the F_6 families.

Family:		1	2	3	4	5	All
Pairs	No.	207,770	200,411	200,866	204,453	195,323	1,008,823
	Ind	176.6	231.7	354.0	219.5	200.7	236.2
bp	Avg	28,745,980.9	30,164,000.4	28,092,640.5	29,039,387.5	28,844,817.9	28,976,195.6
	Min	11	11	40	64	165	57.3
	Max	196,356,451	197,038,449	196,167,202	196,407,037	196,676,774	196,526,525.6
r^2	Avg	0.119	0.112	0.109	0.110	0.119	0.114
	SD	0.221	0.218	0.216	0.218	0.224	0.220
	Min	0.000	0.000	0.000	0.000	0.000	0.000
	Max	1.000	1.000	1.000	1.000	1.000	1.000
$r(bp\text{-}r^2)$		-0.346	-0.346	-0.347	-0.336	-0.351	-0.345

Pairs: No., number of pairs r^2 values obtained; Ind, average number of individuals used to calculate a markers pair LD in a family; bp: average, minimum and maximum bp between markers in a pair; r^2: Avg, average; SD, standard deviation; Min, minimum; Max, maximum; $r(bp\text{-}r^2)$, correlation between the distance and r^2 of a marker pair; All: Pairs, total number of pairs in all families combines; Ind, bp and r^2: means of weighted by the number of pairs.

Table 4. Distribution of syntenic random LD values among the F_6 families.

r^2	Fam1	Fam2	Fam3	Fam4	Fam5	All
≤ 0.05	0.6429	0.6778	0.6920	0.6920	0.6920	0.6707
>0.05 - ≤0.10	0.1036	0.0898	0.0757	0.0757	0.0757	0.0879
>0.10 - ≤0.15	0.0495	0.0451	0.0423	0.0423	0.0423	0.0457
>0.15 - ≤0.20	0.0351	0.0277	0.0292	0.0292	0.0292	0.0315
>0.20 - ≤0.25	0.0242	0.0192	0.0221	0.0221	0.0221	0.0219
>0.25 - ≤0.30	0.0180	0.0165	0.0171	0.0171	0.0171	0.0174
>0.30 - ≤0.35	0.0165	0.0138	0.0152	0.0152	0.0152	0.0152
>0.35 - ≤0.40	0.0120	0.0110	0.0120	0.0120	0.0120	0.0122
>0.40 - ≤0.45	0.0101	0.0101	0.0103	0.0103	0.0103	0.0102
>0.45 - ≤0.50	0.0094	0.0092	0.0081	0.0081	0.0081	0.0091
>0.50 - ≤0.55	0.0076	0.0087	0.0079	0.0079	0.0079	0.0080
>0.55 - ≤0.60	0.0064	0.0079	0.0064	0.0064	0.0064	0.0070
>0.60 - ≤0.65	0.0066	0.0076	0.0072	0.0072	0.0072	0.0069
>0.65 - ≤0.70	0.0067	0.0069	0.0062	0.0062	0.0062	0.0066
>0.70 - ≤0.75	0.0056	0.0070	0.0066	0.0066	0.0066	0.0062
>0.75 - ≤0.80	0.0061	0.0062	0.0063	0.0063	0.0063	0.0060
>0.80 - ≤0.85	0.0060	0.0063	0.0067	0.0067	0.0067	0.0063
>0.85 - ≤0.90	0.0074	0.0068	0.0065	0.0065	0.0065	0.0068
>0.90 - ≤0.95	0.0080	0.0073	0.0082	0.0082	0.0082	0.0081
>0.95 - ≤1.00	0.0180	0.0151	0.0139	0.0139	0.0139	0.0162
Sum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$r^2 > 0.15$	0.2040	0.1872	0.1900	0.1886	0.2089	0.1957
$r^2 > 0.70$	0.0512	0.0486	0.0482	0.0486	0.0518	0.0497

All, all families combine; $r^2 \geq$ (last 2 rows), frequencies of r^2 above the indicated value.

Table 5. Chromosomes and LRLDs in and between QTLRs.

LRLDs	Chr	Fam1	Fam2	Fam3	Fam4	Fam5	All
In	1	83,455	52,985	55,025	39,612	0	231,077
QTLRs	2	15,854	5,195	0	0	0	21,049
	4	87,859	106,654	70,245	88,091	61,605	414,454
	5	0	0	0	0	0	0
	6	0	0	0	0	0	0
	14	503	511	489	116	151	1,770
	Sum	187,671	165,345	125,759	127,819	61,756	668,350
Between	1	39,244	54,674	33,243	33,937	321	161,419
QTLRs	2	52	122	1	1	0	176
	4	26	35	24	33	22	140
	5	1	1	0	0	0	2
	6	2	2	2	2	0	8
	14	19	32	8	24	4	87
	Sum	39,344	54,866	33,278	33,997	347	161,832
Total	1	122,699	107,659	88,268	73,549	321	392,496
	2	15,906	5,317	1	1	0	21,225
	4	87,885	106,689	70,269	88,124	61,627	414,594
	5	1	1	0	0	0	2
	6	2	2	2	2	0	8
	14	522	543	497	140	155	1,857
	Sum	227,015	220,211	159,037	161,816	62,103	830,182

Number of F_6 array LRLDs in and between QTLRs on the six chromosomes with more than one QTLR. Chr, chromosome; Fam, family; All, All families combined; In QTLRs, LRLDs within the MD QTLRs; Across QTLRs, LRLDs between QTLRs; Total, total number of LRLDs.

Table 6. Genes and protein networks in QTLRs 4 and 5, ordered by location.

Q	Gene	Start	End	Net	B4-5	Q	Gene	Start	End	Net	B4-5
4	DUSP16	71,872,763	71,904,322	1	+	5	ING4	77,734,990	77,749,769		+
4	CREBL2	71,960,834	71,970,851	1	+	5	ZNF384	77,755,770	77,781,779		+
4	GPR19	71,980,272	71,988,795		+	5	PIANP	77,803,621	77,808,412		+
4	CDKN1B	72,102,062	72,105,400	1	+	5	COPS7A	77,817,350	77,820,482		+
4	MRPS35	72,622,696	72,644,857	2	+	5	MLF2	77,831,588	77,841,435	5	+
4	MANSC4	72,649,269	72,664,039		+	5	PTMS	77,844,882	77,848,395		+
4	KLHL42	72,671,873	72,682,871	3	+	5	LAG3	77,862,795	77,868,711	2	+
4	PTHLH	72,752,038	72,764,874	2	+	5	CD4	77,873,930	77,885,897	2	+
4	CCDC91	72,867,675	73,073,035		+	5	GPR162	77,894,615	77,899,113		+
4	FAR2	73,191,857	73,320,608			5	P3H3	77,900,316	77,910,271		+
5	SLC2A14	75,548,453	75,558,943			5	GNB3	77,917,296	77,922,102		+
5	NANOG	75,593,243	75,596,024	2		5	CDCA3	77,922,769	77,924,912	5	+
5	AICDA	75,632,084	75,637,754	2		5	USP5	77,924,975	77,939,984	5	+
5	MFAP5	75,647,640	75,660,701	2		5	TPI1	77,940,005	77,943,711	2	+
5	RIMKLB	75,676,381	75,723,185	2		5	LRRC23	77,945,286	77,949,479		+
5	PHC1	75,865,514	75,884,654			5	ENO2	77,952,924	77,962,832	2	+
5	M6PR	75,883,096	75,891,432	2		5	C1H12ORF57	77,991,895	77,993,033		+
5	OVST	76,362,509	76,397,638			5	PTPN6	77,994,636	78,014,698	2	+
5	MAN1A2	77,147,561	77,281,468		+	5	PHB2	78,015,335	78,020,461	2	+
5	CD86	77,307,980	77,318,936	2	+	5	EMG1	78,020,566	78,022,842	2	+
5	CASR	77,388,128	77,429,662	2	+	5	LPCAT3	78,023,023	78,038,708	5	+
5	CSTB	77,433,979	77,438,548	2	+	5	C1S	78,045,961	78,055,022	4	
5	CSTA	77,439,994	77,445,026		+	5	C1R	78,058,485	78,065,761	4	
5	CCDC58	77,454,326	77,464,115	2	+	5	RBP5	78,070,071	78,071,702		
5	FAM162A	77,463,747	77,472,850		+	5	CLSTN3	78,071,723	78,086,627	5	
5	KPNA1	77,477,877	77,514,773	2	+	5	PEX5	78,099,886	78,110,916	5	
5	FBXO40	77,526,063	77,541,005	3	+	5	EPHA1	78,162,527	78,195,976		
5	TAPBPL	77,546,387	77,554,991		+	5	ZYX	78,201,045	78,212,460	5	
5	gga-mir-6553	77,567,295	77,567,395		+	5	FAM131B	78,257,533	78,259,472		
5	SCNN1A	77,568,659	77,576,825		+	5	CLCN1	78,276,084	78,333,916		
5	VAMP1	77,579,970	77,584,568		+	5	CASP2	78,336,971	78,362,958		
5	MRPL51	77,585,845	77,587,396	2	+	5	TMEM139	78,364,767	78,369,487		
5	NCAPD2	77,587,505	77,611,310	5	+	5	RAP1GAP1	78,380,088	78,418,997		

Q	Gene	Start	End	Net	B4-5	Q	Gene	Start	End	Net	B4-5
5	SCARNA10	77,588,152	77,588,472		+	5	GSTK1	78,425,057	78,436,557	5	
5	CNP1	77,615,840	77,617,557		+	5	TAS2R40	78,481,425	78,482,360		
5	GAPDH	77,619,214	77,623,350	**2**	+	5	TRPV6	78,604,630	78,633,785	**2**	
5	IFFO1	77,633,594	77,642,717		+	5	EPHB6	78,678,054	78,732,359		
5	NOP2	77,649,727	77,654,873	**2**	+	5	PRSS2	78,802,240	78,805,418	**2**	
5	LPAR5	77,694,435	77,702,350	**2**	+	5	PRSS3	78,879,681	78,923,402	**2**	

Q, QTLR; Start, End, genes' coordinated on Galgal6; Net, arbitrary number of a network seen in Figure 8, given by order of location of the first gene (not by order of appearance in Figure 8); Net bolded, net comprise of genes from both QTLRs; B4-5, location in high LD blocks extending over the two QTLRs in the five F_6 families.

Figures

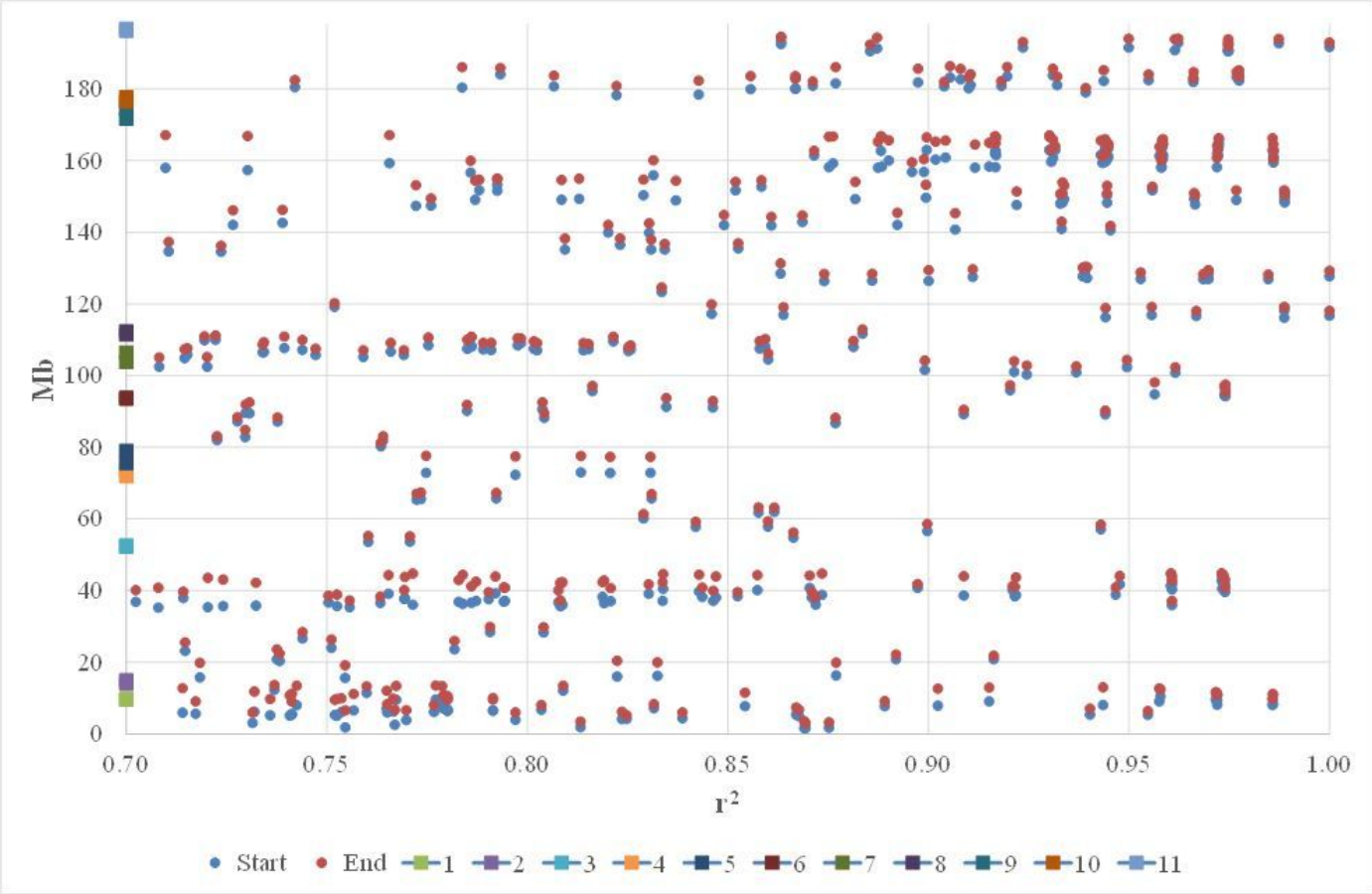


Figure 1

Distribution of random LRLDs over Chr 1 and overlaps with QTLRs (Table 1) in F6 Family 1. Location of marker pairs plotted against r2. Start, End, locations of the markers in a pair (for each Start dot there is a matched End dot; see Figure 2 for clarity);

numbers, QTLR numbers (Table 1).

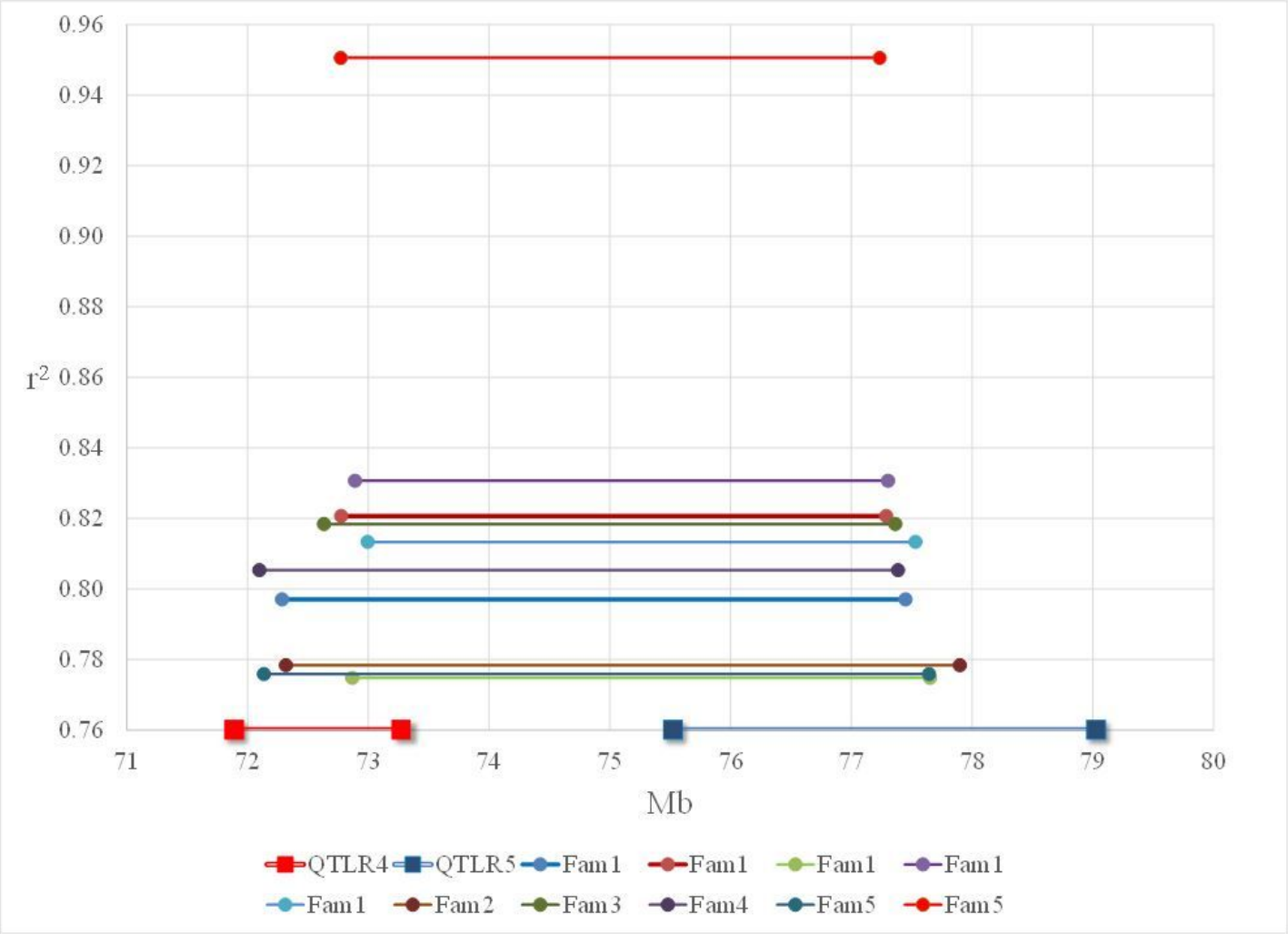
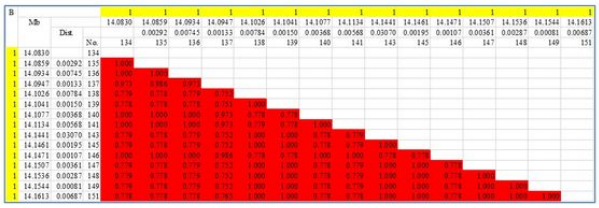


Figure 2

Overlaps between QTLRs 4-5 and random LRLD in all F6 families. Fam, family. There were 5 LRLDs in Family 1, 1 LRLD in Families 2 - 4, and 2 LRLDs in Family 5. Each series of an LRLD is composed of two circles connected by a single line; the circles represent the locations on the x-axis of 2 random markers constitute a pair, and the LD r2 of this pair is presented on the y-axis. There could be more than one LRLD in a family. Similarly, the series of QTLRs 4 and 5 are presents by squares connected by a double line; the squares represent the locations of the QTLR boundaries. QTLRs do not have an r2 of course; they are simply presented on the x-axis, aligned with the LRLD.

a. all blocks.

b. Block 1. Blocks 2 and 3 omitted.



c. Block 2. Blocks 1 and 3 omitted.

B				2	2	2
	Mb			14.1403	14.1460	14.1739
		Dist.			0.00570	0.02783
			No.	142	144	152
2	14.1403		142			
2	14.1460	0.00570	144	0.947		
2	14.1739	0.02783	152	1.000	0.947	

d. Block 3. Blocks 1 and 2 omitted.

B				3	3
	Mb			14.1563	14.7193
		Dist.			0.56293
			No.	150	337
3	14.1563		150		
3	14.7193	0.56293	337	1.000	

Figure 3

Fragmented interdigitated blocks in QTLR 2 on Chr 1 found in F6 Family 1. B, block serial number ordered by the location of the first marker, colored by block; Mb, location on Galgal6 in Mb; Dist., distance in Mb; …, unpresented intermediate markers; No., serial number of the marker; red, LD r2 ≥ 0.7.

B				1	1	2	2	1	1	1	1	1	1
	Mb			45.9001	45.9006	45.9013	45.9021	45.9025	45.9151	45.9157	45.9188	45.9214	45.9359
		Dist.			0.0005	0.0007	0.0008	0.0004	0.0126	0.0007	0.0031	0.0026	0.0145
			No.	5	8	11	12	13	18	20	28	29	36
1	45.9001		5										
1	45.9006	0.0005	8	0.895									
2	45.9013	0.0007	11	0.016	0.041								
2	45.9021	0.0008	12	0.016	0.041	1.000							
1	45.9025	0.0004	13	1.000	0.895	0.016	0.016						
1	45.9151	0.0126	18	1.000	0.895	0.016	0.016	1.000					
1	45.9157	0.0007	20	1.000	0.895	0.016	0.016	1.000	1.000				
1	45.9188	0.0031	28	1.000	0.895	0.016	0.016	1.000	1.000	1.000			
1	45.9214	0.0026	29	1.000	0.895	0.016	0.016	1.000	1.000	1.000	1.000		
1	45.9359	0.0145	36	0.895	1.000	0.041	0.041	0.895	0.895	0.895	0.895	0.895	
	Association test P:			3E-02	5E-02	5E-01	5E-01	3E-02	3E-02	3E-02	3E-02	3E-02	5E-02

Figure 4

LD within one QTLR gene. Line WL1, the gene TRANK1 in QTLR 13 on chromosome 2. B, block serial number ordered by the location of the first marker (same LD block have the same color); Mb, location on Galgal6 in Mb; Dist., distance in Mb; Red, r2 ≥ 0.7; similar P values has similar colors.

B					1	1	1	2	2	2	1	2	2	1	1	2	2	2	2
	Mb				52.2768	52.2769	52.2769	52.3375	52.3532	52.3826	52.4001	52.4082	52.4281	52.4684	52.4745	52.6418	52.6419	52.6422	52.6424
		Dist.				0.0001	0.0000	0.0605	0.0157	0.0294	0.0175	0.0081	0.0199	0.0403	0.0060	0.1674	0.0001	0.0003	0.0002
			E		lncRNA02			lncRNA05								lncRNA04			
				No.	6	7	8	11	12	14	15	16	17	19	20	30	31	32	33
1	52.2768		lncRNA02	6															
1	52.2769	0.0001		7	1.000														
1	52.2769	0.0000		8	1.000	1.000													
2	52.3375	0.0605	lncRNA05	11	0.006	0.006	0.006												
2	52.3532	0.0157		12	0.006	0.006	0.006	1.000											
2	52.3826	0.0294		14	0.016	0.016	0.016	0.959	0.959										
1	52.4001	0.0175		15	1.000	1.000	1.000	0.006	0.006	0.016									
2	52.4082	0.0081		16	0.016	0.016	0.016	0.958	0.958	0.998	0.016								
2	52.4281	0.0199		17	0.016	0.016	0.016	0.945	0.945	0.985	0.016	0.986							
1	52.4684	0.0403		19	1.000	1.000	1.000	0.006	0.006	0.016	1.000	0.016	0.016						
1	52.4745	0.0060		20	1.000	1.000	1.000	0.006	0.006	0.016	1.000	0.016	0.016	1.000					
2	52.6418	0.1674	lncRNA04	30	0.015	0.015	0.015	0.921	0.921	0.960	0.015	0.961	0.978	0.015	0.015				
2	52.6419	0.0001		31	0.015	0.015	0.015	0.921	0.921	0.960	0.015	0.961	0.978	0.015	0.015	1.000			
2	52.6422	0.0003		32	0.015	0.015	0.015	0.921	0.921	0.960	0.015	0.961	0.978	0.015	0.015	1.000	1.000		
2	52.6424	0.0002		33	0.015	0.015	0.015	0.921	0.921	0.960	0.015	0.961	0.978	0.015	0.015	1.000	1.000	1.000	
	Association test P:				7E-01	7E-01	7E-01	1E-01	1E-01	2E-01	7E-01	2E-01	2E-01	7E-01	7E-01	3E-01	3E-01	3E-01	3E-01

Figure 5

LD blocks across QTLR elements. Line WL3, QTLR 3 on chromosome 1. B, block serial number ordered by the location of the first marker (same LD block have the same color); Mb, location on Galgal6 in Mb; Dist., distance in Mb; E, QTLR element [14]; Red, r2 ≥ 0.7; similar P values has similar colors.

B					1	2	2	2	2	2	2	2	1	1	
	Mb				3.9390	3.9418	3.9420	3.9421	3.9428	3.9435	3.9443	3.9447	4.3588	4.3642	4.3689
		Dist.				0.0029	0.0001	0.0001	0.0007	0.0008	0.0007	0.0005	0.4141	0.0054	0.0047
			E		TLR4								BRINP1		
				No.	2	4	5	6	7	8	9	10	11	14	15
1	3.9390		TLR4	2											
2	3.9418	0.0029		4	0.135										
2	3.9420	0.0001		5	0.135	1.000									
2	3.9421	0.0001		6	0.135	1.000	1.000								
2	3.9428	0.0007		7	0.135	1.000	1.000	1.000							
2	3.9435	0.0008		8	0.135	1.000	1.000	1.000	1.000						
2	3.9443	0.0007		9	0.135	1.000	1.000	1.000	1.000	1.000					
2	3.9447	0.0005		10	0.135	1.000	1.000	1.000	1.000	1.000	1.000				
1	4.3588	0.4141	BRINP1	11	0.830	0.102	0.102	0.102	0.102	0.102	0.102	0.102			
1	4.3642	0.0054		14	0.830	0.102	0.102	0.102	0.102	0.102	0.102	0.102	1.000		
	4.3689	0.0047		15	0.157	0.661	0.661	0.661	0.661	0.661	0.661	0.661	0.171	0.171	
Association test P:					3E-01	7E-01	7E-01	7E-01	7E-01	7E-01	7E-01	7E-01	2E-01	2E-01	8E-01

Figure 6

LD between QTLR genes. Line WL2, QTLR 33 on chromosome 17. B, block serial number ordered by the location of the first marker (same LD block have the same color); Mb, location on Galgal6 in Mb; Dist., distance in Mb; E, QTLR element [14]; Red, r2 ≥ 0.7; purple, 0.15 ≤ r2 < 0.7; same LD block have the same color; similar P values has similar colors.

Q						4	4	4	4	5	5	5
	B					1	1	1	1	2	2	2
		Mb				72.3073	72.3074	72.3076	72.3076	77.4421	77.4449	77.4449
			Dist.				0.0001	0.0002	0.0000	5.1345	0.0028	0.0000
				E		lncRNA01.03				CSTA		
					No.	34	36	37	38	42	44	45
4	1	72.30728		lncRNA 01.03	34							
4	1	72.30737	0.00009		36	1.000						
4	1	72.30762	0.00025		37	1.000	1.000					
4	1	72.30763	0.00001		38	1.000	1.000	1.000				
5	2	77.44215	5.13451	CSTA	42	0.478	0.478	0.478	0.478			
5	2	77.44490	0.00276		44	0.478	0.478	0.478	0.478	1.000		
5	2	77.44493	0.00003		45	0.478	0.478	0.478	0.478	1.000	1.000	
Association test P:						8E-02	8E-02	8E-02	8E-02	4E-01	4E-01	4E-01

Figure 7

LD between QTLRs 4 and 5. Line WPR1, the lncRNA01 in QTLR 4 and the gene CSTA in QTLR 5. Q, QTLR serial number (Table 1); B, high LD block serial number ordered by the location of the first marker (same LD block have the same color); Mb, location on Galgal6 in Mb; Dist., distance in Mb; E, QTLR element [14]; Red, r2 ≥ 0.7; purple, 0.15 ≤ r2 < 0.7; same LD block have the same color; similar P values has similar colors.

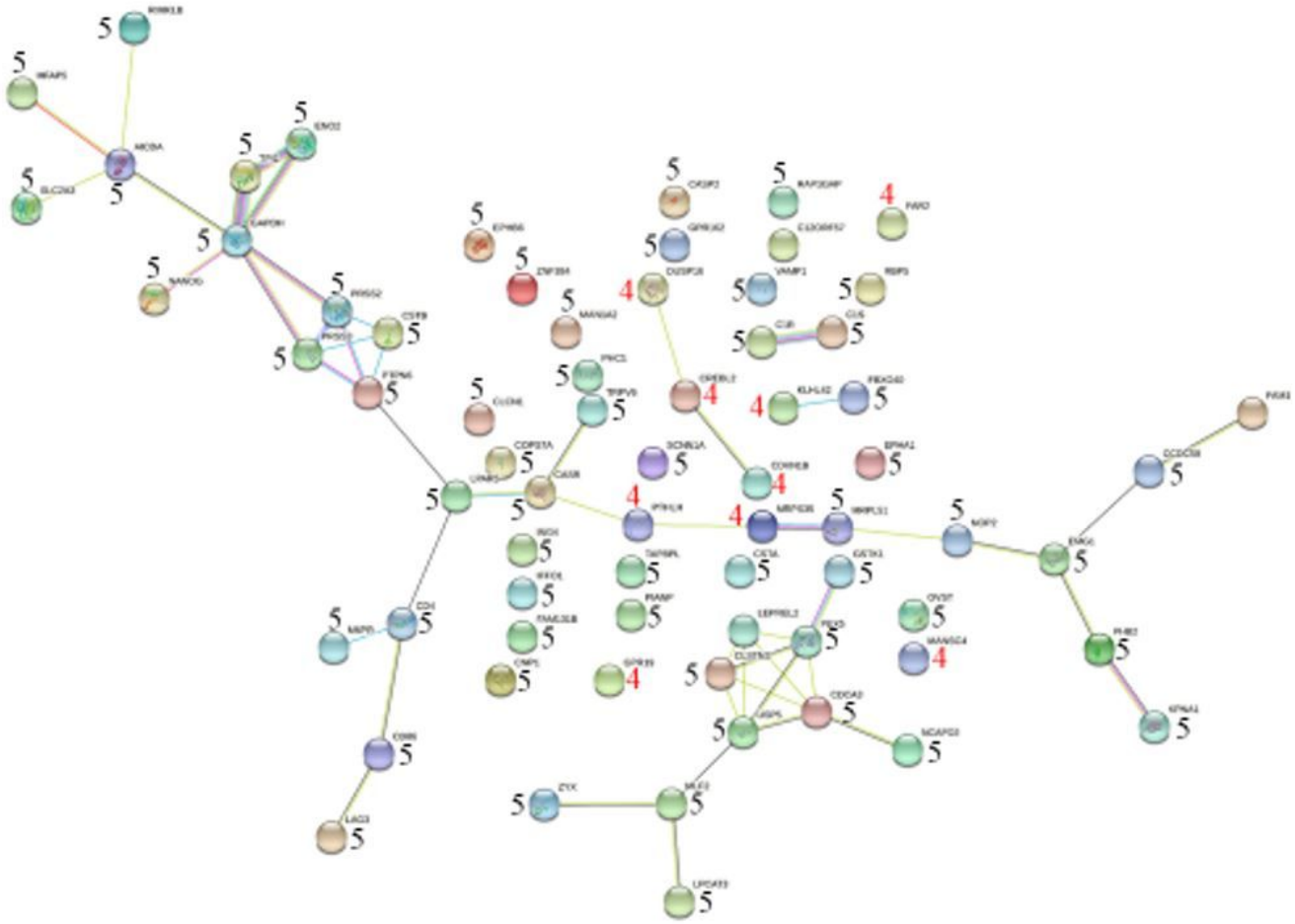


Figure 8

please see the manuscript file for the full caption

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ChickenLDadditionaltablesandfigureBMCGenomics.docx