



This is a repository copy of *Genetic mapping of the major histocompatibility complex in the zebra finch (Taeniopygia guttata)*.

White Rose Research Online URL for this paper:  
<http://eprints.whiterose.ac.uk/152446/>

Version: Accepted Version

---

**Article:**

Ekblom, R., Stapley, J., Ball, A.D. et al. (3 more authors) (2011) Genetic mapping of the major histocompatibility complex in the zebra finch (*Taeniopygia guttata*). *Immunogenetics*, 63 (8). pp. 523-530. ISSN 0093-7711

<https://doi.org/10.1007/s00251-011-0525-9>

---

This is a post-peer-review, pre-copyedit version of an article published in *Immunogenetics*. The final authenticated version is available online at:  
<http://dx.doi.org/10.1007/s00251-011-0525-9>.

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



[eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk)  
<https://eprints.whiterose.ac.uk/>

1 **Genetic mapping of the major histocompatibility complex in the**  
2  
3 **zebra finch (*Taeniopygia guttata*)**  
4  
5

6 3  
7  
8 4 **Robert Ekblom<sup>1,2\*</sup>, Jessica Stapley<sup>2</sup>, Alex D. Ball<sup>2,3</sup>, Tim Birkhead<sup>2</sup>, Terry Burke<sup>2</sup>, Jon**  
9 **Slate<sup>2</sup>**  
10  
11

12 6  
13  
14  
15  
16 7 1) Uppsala University, Department of Population Biology and Conservation Biology,  
17  
18 8 Norbyvägen 18 D, SE-75236, Uppsala, Sweden

19  
20  
21 9 2) University of Sheffield, Department of Animal and Plant Sciences, Sheffield S10 2TN, UK  
22

23 10 3) University of Bath, Department of Biology and Biochemistry, Claverton Down, Bath BA2  
24  
25 11 7AY, UK  
26  
27

28 12  
29  
30 13 \* Corresponding author: [robert.ekblom@ebc.uu.se](mailto:robert.ekblom@ebc.uu.se)  
31  
32

33 14 tel: +46(0)184712756  
34

35 15 fax: +46(0)184716424  
36  
37  
38 16  
39  
40 17  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 **Abstract** Genes of the major histocompatibility complex (MHC) have received much  
2 attention in immunology, genetics and ecology because they are highly polymorphic and play  
3 important roles in parasite resistance and mate choice. Until recently the MHC of passerine  
4 birds was not well described. However, the genome sequencing of the zebra finch  
5 (*Taeniopygia guttata*), has partially redressed this gap in our knowledge of avian MHC genes.  
6 Here we contribute further to the understanding of the zebra finch MHC organization by  
7 mapping SNPs within or close to known MHC genes in the zebra finch genome. MHC class I  
8 and IIB genes were both mapped to zebra finch chromosome 16 and there was no evidence  
9 that MHC class I genes are located on chromosome 22 (as suggested by the genome  
10 assembly). We confirm the location in the MHC region on chromosome 16 for several other  
11 genes (BRD2, FLOT1, TRIM7.2, GNB2L1 and CSNK2B). Two of these (CSNK2B and  
12 FLOT1) have not previously been mapped in any other bird species. In line with previous  
13 results we also find that orthologs to the immune related genes B-NK and CLEC2D, which  
14 are part of the MHC region in chicken, are situated on zebra finch chromosome Z and not  
15 among other MHC genes in the zebra finch.

16  
17 **Key words** MHC, Bird, Chromosome, Linkage map, SNP

# 1 Introduction

2 Major histocompatibility complex (MHC) genes have attracted a considerable amount of  
3 attention in many different fields of biological research due to their important role in  
4 immunity and ecology and their exceptionally high levels of genetic variation (Edwards and  
5 Hedrick 1998; Sommer 2005). Immunoecological studies in different groups of vertebrates  
6 have demonstrated links between MHC variation and several fitness related traits such as  
7 disease susceptibility, mating success and survival (Bernatchez and Landry 2003; Piertney  
8 and Oliver 2006). In birds, however, research in this field has been severely hampered by a  
9 lack of knowledge of the basic features of MHC structure and organization.

10

11 Almost all of the detailed information available on large scale bird MHC organization comes  
12 from investigations in chicken and a few other galliform birds (Chaves et al. 2007; Miller et  
13 al. 2004; Shiina et al. 2006). In those species the classical MHC genes (class I and class IIB)  
14 form a small and closely linked cluster on chromosome 16 (one of the chicken  
15 microchromosomes), characteristics that, together with strong links between specific MHC  
16 haplotypes and disease resistance, have given rise to the concept of a “minimal essential  
17 MHC” (Kaufman et al. 1999b; Kaufman et al. 1995). While MHC sequence variation has  
18 been studied in a large number of other bird species (e.g. Alcaide et al. 2008; Burri et al.  
19 2008; Ekblom et al. 2003; Hughes et al. 2008; Tsuda et al. 2001; Westerdahl 2007), these  
20 studies have only characterized a small part of one or a few loci, while ignoring the large  
21 scale structure and organization of the genes. Preliminary results on MHC organisation from  
22 avian taxa outside the galliformes (primarily the passerine birds) indicate that MHC  
23 organization is more complex than in chicken, and is characterised by a higher degree of gene  
24 duplication, longer introns and intergenic distances resulting in the MHC spanning a larger  
25 region (Hess and Edwards 2002; Hess et al. 2000).

1  
2 Recent technological advances have opened up the field of genomics to researchers studying a  
3 wide variety of non-model organisms (Ekblom and Galindo 2011; Lister et al. 2009; Wheat  
4 2010). Using data from the sequence assembly of the second bird genome (the zebra finch;  
5 *Taeniopygia guttata*), together with targeted sequencing of MHC-containing BACs, FISH  
6 mapping and next generation digital transcriptomics data (RNA-Seq), MHC organization in  
7 this passerine species is now beginning to be described (Balakrishnan et al. 2010; Ekblom et  
8 al. 2010; Warren et al. 2010). Initial findings indicate that the zebra finch MHC is complex  
9 and may have undergone more gene duplication events (especially among the class IIB genes)  
10 than chicken. Intriguingly, this previous work also suggests that the zebra finch homologues  
11 of MHC genes found on chicken chromosome 16 are located on two discrete chromosomes,  
12 with class I genes on a separate chromosome compared to the TAP (antigen peptide  
13 transporter) genes. Furthermore the expressed MHC class I gene is placed on chromosome 22  
14 in the zebra finch genome assembly, while two other genes linked to the chicken MHC region  
15 (B-NK and CLEC2D) are found on zebra finch chromosome Z (Balakrishnan et al. 2010).

16  
17 Here we set out to identify and confirm the location of MHC related genes in the zebra finch,  
18 using linkage mapping of specific single nucleotide polymorphism (SNP) markers. MHC  
19 related genes are here broadly defined as genes known to be situated within the MHC region  
20 in this or other species and genes involved in the function of MHC immune pathways.  
21 Specifically, we tested whether linkage mapping would confirm that zebra finch MHC genes  
22 are situated on several different chromosomes as the current genome assembly suggests,  
23 possibly resulting from a fission (or lack of fusion) of chromosome 16 in this lineage. We also  
24 investigated whether B-NK and CLEC2D genes, linked to the MHC in galliform birds, were  
25 located on chromosome Z in zebra finch, as suggested by the genome sequence assembly.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**2 Methods**

**3 Mapping population**

4 The zebra finch International Mapping Flock (IMF) consists of a 3-generational, 354-bird  
5 pedigree that is part of a larger captive population that has been maintained at the University  
6 of Sheffield since 1985 (Birkhead et al. 2005). Within the mapping population there are 60  
7 G<sub>0</sub>, 43 G<sub>1</sub> and 251 G<sub>2</sub> birds. The mean sibship size among the G<sub>2</sub> progeny is 12.1 (range 9-  
8 27). The IMF was previously used to create a linkage map of 876 SNPs spanning 1068 cM  
9 across 45 linkage groups (Stapley et al. 2008).

**11 Identification of putative MHC SNPs**

12 In order to produce a linkage map of the zebra finch MHC region, we mined several DNA  
13 sequence databases to compile a list of 96 putative SNPs located in different MHC genes or  
14 genetic regions mapping to the MHC in chicken (for details about these see Online Resource  
15 1). It was our intention to genotype up to 48 of these SNPs, but the larger list was initially  
16 compiled to ensure that 48 SNP assays with a high probability of typing success could be  
17 developed. SNPs were detectable because 454 transcriptome sequencing data of expressed  
18 genes had previously been conducted in a pool of six individuals from the mapping  
19 population (Ekblom et al. 2010). To identify those SNPs that were located in MHC genes,  
20 alignments between contigs previously identified as representing MHC genes and all reads  
21 mapping to these contigs were manually searched for polymorphic sites (for details about this  
22 annotation see Balakrishnan et al. 2010; Ekblom et al. 2010). 454 sequencing reads were also  
23 mapped onto the zebra finch genome sequence (chromosome 16 and chromosome  
24 16\_random) and individually sequenced MHC-containing zebra finch BACs (Balakrishnan et  
25 al. 2010). This allowed us to identify SNPs in un-translated regions of MHC genes and in

1 genes placed on chromosome 16 of the genome assembly that had not been previously  
2 annotated. In addition to identifying SNPs from the 454 data (which was generated from birds  
3 in our study population) some putative SNPs were also identified using polymorphism data  
4 from the zebra finch genome project (which used birds from other populations). Genome  
5 project SNPs are available through the ENSEMBL genome browser  
6 ([www.ensembl.org/Taeniopygia\\_guttata/Info/Index](http://www.ensembl.org/Taeniopygia_guttata/Info/Index)) or BioMart  
7 ([www.ensembl.org/biomart/martview](http://www.ensembl.org/biomart/martview)). Note that because the genome project SNPs were not  
8 identified from our mapping population it is possible that many of these SNPs will not be  
9 segregating in our study population.

10

#### 11 SNP typing

12 Of the 96 putative SNPs, we designed a panel of 48 SNPs for typing (for flanking sequences  
13 see Online Resource 2). These were chosen based on Illumina assay design scores and also in  
14 order to get good coverage of the regions of interest. This 48-SNP panel, containing putative  
15 MHC genes and chromosome 16 regions, was genotyped in the IMF as a custom Veracode  
16 GoldenGate kit on the Illumina BeadXpress platform. Genotypes were called using the  
17 Illumina Genome Studio Genotyping Module v1.0, and genotypes were then combined with  
18 the data used to build the linkage map reported in Stapley et al. (2008). SNPs typed in this  
19 study were not selected based on minor allele frequency or sequence depth of the assemblies,  
20 but rather chosen because they were situated in genetic regions of particular interest. This,  
21 combined with the highly variable and extensively duplicated nature of many of the genes of  
22 the MHC in most species, and the fact that some SNPs were identified in a different  
23 population, meant that we expected a rather low genotyping success rate.

24

#### 25 Annotation of SNPs

1 All SNPs used in this study (Table 1) were annotated using a blast approach (Altschul et al.  
2 1997). The SNP flanking sequences (65 – 141 bp, and in a few cases extended genomic  
3 regions surrounding these) were blasted against zebra finch, chicken and human gene  
4 predictions (using blastn) and protein sequences (using blastx). The best blast hits were used  
5 to identify the genes for annotation of each of the SNPs. In addition the SNPs were blasted  
6 against the zebra finch genome assembly to identify the chromosome location. A stand-alone  
7 blast version (2.2.18) was used for the gene annotation while the blat search engine on  
8 ENSEMBL ([http://www.ensembl.org/Taeniopygia\\_guttata/blastview](http://www.ensembl.org/Taeniopygia_guttata/blastview)) was used for the  
9 chromosome location survey. A SNP was classified as either being synonymous, non-  
10 synonymous, intronic, UTR or intergenic, by identifying the position of the SNP within the  
11 alignment to either the zebra finch or chicken genome sequences.

### 12 13 Map construction

14 The SNPs that were successfully typed in this study were combined with the 876 SNPs used  
15 to build the previously published linkage map of the whole zebra finch genome (Stapley et al.  
16 2008). Map construction followed the procedure previously described in Stapley et al. (2008).  
17 A version of CriMap v2.4 (Green et al. 1990) modified by Xuelu Liu (Monsanto) was used to  
18 estimate two-point linkage between all pairs of markers, assign markers into linkage groups  
19 and build a genetic linkage map for the linkage groups containing the new MHC SNPs.  
20 Linkage groups were created between markers that were linked to at least one other marker  
21 with LOD score > 5. For each linkage group a framework map was built; framework maps  
22 contained only those markers whose relative position could be assigned with LOD > 3. Any  
23 remaining markers were then added to the map iteratively using the BUILD command in  
24 CriMap at a lower stringency, such that the final build included all markers at their most  
25 likely positions.



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

## 2 LD analysis

3 The extent of linkage disequilibrium across chromosome 16 was estimated using the program  
4 Haploview v4.2 (Barrett et al. 2005). Unphased genotype data from only the founders was  
5 used for this analysis. All markers and founder individuals passed the selection criteria for  
6 inclusion in analysis (minor allele frequencies,  $MAF > 0.05$ , markers with higher than 65%  
7 call frequency and individuals that were genotyped at  $> 75\%$  of loci). Information from the  
8 LD analysis was also used to infer the order of markers that had zero recombination distance  
9 (cM) between them. In cases where markers share the same map position, changing their  
10 order relative to each other does not change the likelihood of the linkage map. However, LD  
11 analysis can sometimes be informative in this respect and suggest the most parsimonious  
12 order. As a result we used estimates of LD to order three markers at the same genetic map  
13 position (1363, 1365, flot1); marker pairs in greatest LD were placed next to each other.  
14 Ordering tightly linked markers in this way also facilitates the interpretation of the LD  
15 heatmap.

## 17 Haplotype Inference

18 We estimated the number of distinct haplotypes within the founders ( $G_0$  birds) of the mapping  
19 panel. By combining marker distance information with the genotypes at all chromosome 16  
20 SNPs in the founder birds, it is possible to infer the haplotypes (i.e. the phased chromosomes)  
21 in those founders. Haplotype inference was performed using fastPHASE (Scheet and  
22 Stephens 2006) using default parameter settings. We estimated the number of haplotypes  
23 across the length of chromosome 16 and also across the 5 markers spanning the “core MHC”  
24 including the class I and class IIB genes.

## 1 **Results**

### 2 Genetic map

3 Out of 48 genotyped SNPs from zebra finch MHC genes and chromosome 16, eleven were  
4 polymorphic in the mapping population and had sufficiently high GenTrain scores (sufficient  
5 genotype clustering) for genotyping (see Online Resource 1). The mean (SE) call rate and  
6 minor allele frequencies of the 11 SNPs were 0.98 (0.007) and 0.29 (0.029) respectively. Two  
7 of these were found to be identical to SNPs used in the genetic linkage map (Stapley et al.  
8 2008) and were not considered further. None of the 9 remaining SNPs represented amino acid  
9 altering substitutions (Table 1).

10

11 All but one SNP was linked ( $LOD > 5$ ) to other markers on the linkage map, and they were  
12 assigned to two chromosomes. Five SNPs (situated in the genes TRIM7.2, BRD2, FLOT1,  
13 MHC class I, and MHC class IIB) were mapped to chromosome 16, and three SNPs (in genes  
14 RCL1, B-NK and CLEC2D) were mapped to chromosome Z (Fig. 1). Two SNPs located in  
15 presumably functional MHC class I and IIB genes (as inferred from genetic, and  
16 transcriptomic sequences, see Balakrishnan et al. 2010; Ekblom et al. 2010), as well as several  
17 other genes, were thus found to be linked and situated within the zebra finch MHC region on  
18 chromosome 16. The SNP that could not be mapped was from a putative MHC class I  
19 pseudogene ( $\psi C$ ). This marker was weakly linked ( $LOD > 1$ ) to 44 other markers that were  
20 spread across 10 chromosomes. The maximum LOD observed for this marker was 2.21 to a  
21 marker on chromosome 19. However, there was no linkage ( $LOD > 1$ ) to markers on  
22 chromosome 16 or Z. The resulting genetic sex-average map of chromosome 16 contains 9  
23 markers and spans 46 cM (Female map = 46 cM, Male map = 44.5 cM). The genes coding for  
24 two lectin proteins (B-NK and CLEC2D) have been shown to be integrated in the MHC of the  
25 chicken as well as other galliform birds (Hosomichi et al. 2006). Here we verify the finding

1 from Balakrishnan and co-workers (2010) that the zebra finch homologs of these genes are  
2 situated on the Z chromosome, close to the RCL1 (RNA terminal phosphate cyclise-like 1)  
3 gene (Fig. 1). None of the MHC SNPs typed in this study were assigned to zebra finch  
4 chromosome 22 or any other assembled chromosomes other than 16 and Z.

## 5 6 LD analysis

7 Overall LD on zebra finch chromosome 16 was relatively low; average  $r^2$  across all markers  
8 was 0.07 and > 90% of pairwise LD was < 0.3 (Fig. 2). It is not possible to estimate  
9 recombination rate across the chromosome because the physical size is unknown. The genetic  
10 map length of this chromosome is slightly longer than other microchromosomes with a  
11 similar number of markers (4.7 - 39.9 cM) (Stapley et al. 2008), but the physical size of  
12 chromosome 16 is thought to be smaller than these chromosomes (International Chicken  
13 Genome Sequencing Consortium 2004; Warren et al. 2010). Recent estimates of the length of  
14 chicken chromosome 16 suggest that it may be around 10 Mb long (Solinhaç et al. 2010). The  
15 genetic region corresponding to our chromosome 16 linkage group (from MHC class IIB to  
16 TRIM7.2) in chicken is about 100 kb long (Kaufman et al. 1999b) and the size in the turkey is  
17 around 140 kb (Chaves et al. 2007). Although it is likely that the physical size of the zebra  
18 finch MHC is considerably larger than this (Balakrishnan et al. 2010), our results would  
19 suggest unusually high levels of recombination rate in the zebra finch MHC region. The  
20 region containing markers Tgu\_SNP\_01365, Tgu\_SNP\_01363, flot1 and tgu\_class1\_7,  
21 showed slightly higher levels of LD between markers (mean  $r^2 = 0.32$ ) compared to other  
22 parts of chromosome 16. However, without any knowledge of the physical distance between  
23 these markers it is not possible to determine the relevance of this LD, because it may simply  
24 reflect close physical linkage between these markers. The linkage map size of this region was  
25 4.4 cM (Fig. 1).

1

## 2 Haplotype Inference

3 Within the 60 founder birds, we identified 53 unique haplotypes across chromosome 16,

4 suggesting that recombination has helped to maintain diversity on this chromosome.

5 However, the region around the “core MHC”, spanning five markers (1363, 1365, flot1,

6 tgu\_class1\_7, tmptgu393923) contained just three unique haplotypes, indicating that genetic

7 diversity is relatively low in this part of the zebra finch genome. This finding also suggests

8 that recombination is relatively rare in the part of chromosome 16 that contains the core MHC

9 genes.

10

## 11 **Discussion**

12 Balakrishnan and co-authors (2010) reported that zebra finch MHC genes seem to be spread

13 over several chromosomes. Our study reveals that one MHC class I and one MHC class IIB

14 locus are closely linked and placed together on chromosome 16 (Fig. 1). All markers situated

15 on chromosome 16 were assigned to that linkage group with very high confidence. These

16 include two SNPs in expressed MHC genes corresponding to locus 1 for the class I gene and

17 locus 2 for the class IIB gene following the nomenclature of Balakrishnan et al. (2010). Thus,

18 there is no evidence that the homologue of chicken chromosome 16 is represented by two

19 smaller chromosomes in the zebra finch. We found no support for the placement of the MHC

20 class I locus on chromosome 22 as suggested by the recently released zebra finch genome

21 assembly (Warren et al., 2010). A similar conclusion was drawn by Balakrishnan et al. (2010)

22 based on the fact that a class I probe did not co-hybridise with a chromosome 22 probe using

23 FISH mapping. Note, however, that one SNP placed in a putative MHC class I pseudogene

24 (locus  $\psi$ C) was not linked to the rest of the MHC region or any other known linkage group.

25 Furthermore, we were unable to detect any polymorphism within the TAP or TNXB genes,

1 which Balakrishnan et al. (2010) reported were on the BAC that did not co-hybridise with the  
2 rest of the MHC region. Therefore, we could not design SNP assays to map these two genes.  
3 Our study also demonstrated that five other genes (TRIM7.2, GNB2L1, CSNK2B, BRD2, and  
4 FLOT1) are situated in the zebra finch MHC region. These have all previously been reported  
5 as residing in the MHC of chicken and/or humans (The MHC Sequencing Consortium 1999),  
6 but CSNK2B and FLOT1 have not been previously mapped or annotated in the genome  
7 assembly of any bird species apart from zebra finch.

8  
9 The inclusion of several new SNPs to the zebra finch linkage map has improved the original  
10 genetic mapping of chromosome 16. In the first generation linkage map only two markers  
11 mapped to this chromosome and in the second generation map an additional marker (TS1365)  
12 was included (Backström et al. 2010). In the present study two additional SNPs  
13 (Tgu\_SNP\_01363 and Tgu\_SNP\_01547), that were previously mapped by Stapley et al.  
14 (2008) to an unknown linkage group (TguUN4) are shown to be part of chromosome 16.  
15 Together with the new MHC markers described here the linkage group on chromosome 16  
16 now contains nine markers, spanning the region from the MHC class IIB gene to TRIM7.2.

17  
18 In chicken and other galliform birds an independent cluster of MHC genes (MHC-Y) is  
19 located on the same chromosome as the classical MHC (MHC-B) but these are separated by a  
20 region of very high recombination (Briles et al. 1993; Miller et al. 1996). In the chicken  
21 linkage map there are 18 markers on chromosome 16. These are situated in two linkage  
22 groups corresponding to the classical MHC (B) locus and the MHC-Y locus, localized about  
23 60 cM apart, making them effectively unlinked. It has not been possible to provide fine scale  
24 resolution of the 12 markers mapping to the MHC-B region (Groenen et al. 2000). Therefore,  
25 there is a high rate of recombination between the B and Y regions on chicken chromosome 16

1 but low recombination rate within the B-region. It has been argued that the compact MHC  
2 region of the chicken has resulted in tight linkage between functionally interacting immune  
3 genes, enabling co-evolution between them; the minimal essential MHC hypothesis (Kaufman  
4 et al. 1999a). The B-region of chicken has traditionally been divided further into a “B-F/B-L  
5 region” containing the classical MHC class I and class IIB genes and a “B-G region”  
6 containing so called B-G genes (Kaufman et al. 1995). Low but significant rates of  
7 recombination have been observed between these two parts of the chicken MHC by  
8 investigation of crosses between different inbred lines (Skjødt et al. 1985). A more detailed  
9 mapping of the chicken MHC region was recently presented by Solinhac and colleagues  
10 (2010). There the genetic map of chicken chromosome 16 is 130.7 cM long and consists of 33  
11 markers distributed over the MHC-B, MHC-Y and nucleolus organizing region (NOR). The  
12 only additional previous detailed genetic mapping study of the classical MHC region in birds  
13 has recently been performed on the turkey (*Meleagris gallopavo*). Here, 14 SNP and  
14 microsatellite markers were typed in a large number of offspring from two females (Chaves et  
15 al. 2010). The markers were spread over a physical distance of approximately 200 kb across  
16 the complete B-locus (from the TRIM7.2 gene to the C4 gene). After having controlled for  
17 three gene conversion events, the resulting genetic map was 3.6 cM, giving a recombination  
18 rate of 18 cM/Mb (an order of magnitude higher than for the average of the turkey genome).  
19  
20 Our mapping of the zebra finch chromosome 16 found evidence of a considerable amount of  
21 recombination between markers in the MHC region. The low level of LD on this chromosome  
22 is similar to the pattern of LD found on other microchromosomes (Stapley et al. 2010). There  
23 is a strong negative relationship between recombination rate and chromosome length in the  
24 zebra finch (Stapley et al. 2008). Given that chromosome 16 is one of the smallest  
25 chromosomes in the avian genome, it is thus not surprising that we find a high degree of

1 recombination here. However, our finding of a considerable genetic distance (9.1 cM)  
2 between classical MHC class I and class IIB loci is in contrast with the chicken where  
3 recombination within the classical MHC region (B-F/B-L) seems rare (see discussion above  
4 regarding the minimal essential MHC hypothesis). However, our estimate of genetic distance  
5 across chromosome 16 needs to be treated with care, because our estimated genetic map  
6 distance could be inflated by recent gene conversion events, as was observed in the turkey  
7 MHC region (Chaves et al. 2010). Also, since we don't know the physical length of the  
8 mapped zebra finch chromosome 16 region we can not obtain a reliable estimate on local  
9 recombination rate. The fact that we only identified three distinct haplotypes in our 60  
10 founders in the "core MHC" is perhaps consistent with recombination being relatively rare in  
11 this particular region of chromosome 16.

12  
13 Even though there has been two studies of genome-wide linkage disequilibrium of the zebra  
14 finch (Backström et al. 2010; Stapley et al. 2010), neither of these have been able to  
15 successfully map and analyse chromosome 16. Difficulty in mapping this chromosome in the  
16 past is most likely due to a combination of a high recombination rate and very small size.  
17 Recombination is considered an important process creating genetic and haplotype variability  
18 (e.g. Begun and Aquadro 1992; Jaramillo-Correa et al. 2010) and this may be an important  
19 factor contributing to the variability of MHC loci too. Thus, even though different MHC  
20 genes are linked on a physically small chromosome, a high recombination rate in this region  
21 would mean that novel haplotypes between these loci are constantly being generated by  
22 recombination (Schaschl et al. 2006), however our finding of a rather limited number of  
23 haplotypes in the "core MHC" region would argue against this mechanism at least for MHC  
24 class I and IIB loci. Recombination rates across different parts of the MHC region may have

1 profound implications on how the adaptive immune system is evolving in this lineage of  
2 birds.

3

4 **Acknowledgements** We thank Andy Krupa for lab assistance. Christopher Balakrishnan  
5 kindly shared MHC BAC sequences and provided valuable discussions on our results and  
6 three anonymous reviewers provided valuable comments on a previous version of this  
7 manuscript. This work was partially funded by an EC Transfer of Knowledge grant  
8 (MAERO) and a BBSRC grant (BB/E017509/1), both awarded to JS.

9



## 1 References

- 2 Alcaide M, Edwards SV, Negro JJ, Serrano D, Tella JL (2008) Extensive polymorphism and  
3 geographical variation at a positively selected MHC class II B gene of the lesser  
4 kestrel (*Falco naumanni*). *Mol Ecol* 17: 2652-2665
- 5 Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997)  
6 Gapped BLAST and PSI-BLAST: a new generation of protein database search  
7 programs. *Nucl Acids Res* 25: 3389-3402
- 8 Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, Webster MT, Öst  
9 T, Schneider M, Kempnaers B, Ellegren H (2010) The recombination landscape of  
10 the zebra finch *Taeniopygia guttata* genome. *Genome Res* 20: 485-495
- 11 Balakrishnan C, Ekblom R, Volker M, Westerdahl H, Godinez R, Kotkiewicz H, Burt D,  
12 Graves T, Griffin D, Warren W, Edwards S (2010) Gene duplication and  
13 fragmentation in the zebra finch major histocompatibility complex. *BMC Biology* 8:  
14 29
- 15 Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and  
16 haplotype maps. *Bioinformatics* 21: 263-265
- 17 Begun DJ, Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate  
18 with recombination rates in *D. melanogaster*. *Nature* 356: 519-520
- 19 Bernatchez L, Landry C (2003) MHC studies in nonmodel vertebrates: what have we learned  
20 about natural selection in 15 years? *J Evol Biol* 16: 363-377
- 21 Birkhead TR, Pellatt EJ, Brekke P, Yeates R, Castillo-Juarez H (2005) Genetic effects on  
22 sperm design in the zebra finch. *Nature* 434: 383-387
- 23 Briles WE, Goto RM, Auffray C, Miller MM (1993) A polymorphic system related to but  
24 genetically independent of the chicken major histocompatibility complex.  
25 *Immunogenetics* 37: 408-414
- 26 Burri R, Hirzel HN, Salamin N, Roulin A, Fumagalli L (2008) Evolutionary Patterns of MHC  
27 Class II B in Owls and Their Implications for the Understanding of Avian MHC  
28 Evolution. *Mol Biol Evol* 25: 1180-1191
- 29 Chaves L, Faile G, Krueth S, Hendrickson J, Reed K (2010) Haplotype variation,  
30 recombination, and gene conversion within the turkey MHC-B locus. *Immunogenetics*  
31 62: 465-477
- 32 Chaves LD, Krueth SB, Reed KM (2007) Characterization of the turkey MHC chromosome  
33 through genetic and physical mapping. *Cytogenet Gen Res* 117: 213-220
- 34 Edwards SV, Hedrick PW (1998) Evolution and ecology of MHC molecules: from genomics  
35 to sexual selection. *Trends Ecol Evol* 13: 305-311
- 36 Ekblom R, Balakrishnan CN, Burke T, Slate J (2010) Digital gene expression analysis of the  
37 zebra finch genome. *BMC Genomics* 11: 219
- 38 Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology  
39 of non-model organisms. *Heredity Advance online publication*: doi:  
40 10.1038/hdy.2010.152
- 41 Ekblom R, Grahn M, Höglund J (2003) Patterns of polymorphism in the MHC class II of a  
42 non-passerine bird, the great snipe (*Gallinago media*). *Immunogenetics* 54: 734-741
- 43 Green P, Falls K, Crooks S (1990) Documentation for CRIMAP.  
44 <http://compgen.rutgers.edu/Crimap/Default.aspx>.
- 45 Groenen MAM, Cheng HH, Bumstead N, Benkel BF, Briles WE, Burke T, Burt DW,  
46 Crittenden LB, Dodgson J, Hillel J, Lamont S, de Leon AP, Soller M, Takahashi H,  
47 Vignal A (2000) A Consensus Linkage Map of the Chicken Genome. *Genome Res* 10:  
48 137-147

- 1 Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in  
2 birds. *Bioscience* 52: 423-431
- 3 Hess CM, Gasper J, Hoekstra HE, Hill CE, Edwards SV (2000) MHC Class II pseudogene  
4 and genomic signature of a 32-kb cosmid in the House finch (*Carpodacus mexicanus*).  
5 *Genome Res* 10: 613-623
- 6 Hosomichi K, Shiina T, Suzuki S, Tanaka M, Shimizu S, Iwamoto S, Hara H, Yoshida Y,  
7 Kulski J, Inoko H, Hanzawa K (2006) The major histocompatibility complex (Mhc)  
8 class IIB region has greater genomic structural flexibility and diversity in the quail  
9 than the chicken. *BMC Genomics* 7: 322
- 10 Hughes C, Miles S, Walbroehl J (2008) Support for the minimal essential MHC hypothesis: a  
11 parrot with a single, highly polymorphic MHC class II B gene. *Immunogenetics* 60:  
12 219-231
- 13 International Chicken Genome Sequencing Consortium (2004) Sequence and comparative  
14 analysis of the chicken genome provide unique perspectives on vertebrate evolution.  
15 *Nature* 432: 695-716
- 16 Jaramillo-Correa J, Verdu M, Gonzalez-Martinez S (2010) The contribution of recombination  
17 to heterozygosity differs among plant evolutionary lineages and life-forms. *BMC Evol*  
18 *Biol* 10: 22
- 19 Kaufman J, Jacob J, Shaw J, Walker B, Milne S, Beck S, Salomonsen J (1999a) Gene  
20 organisation determines evolution of function in the chicken MHC. *Immunol Rev* 167:  
21 101-117
- 22 Kaufman J, Milne S, Göbel TWF, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S  
23 (1999b) The chicken B locus is a minimal essential major histocompatibility complex.  
24 *Nature* 401: 923-925
- 25 Kaufman J, Völk H, Wallny HJ (1995) A "minimal essential Mhc" and an "unrecognized  
26 MHC": two extremes in selection for polymorphism. *Immunol Rev* 143: 63-88
- 27 Lister R, Gregory BD, Ecker JR (2009) Next is now: new technologies for sequencing of  
28 genomes, transcriptomes, and beyond. *Curr Opin Plant Biol* 12: 107-118
- 29 Miller MM, Bacon LD, Hala K, Hunt HD, Ewald SJ, Kaufman J, Zoorob R, Briles WE  
30 (2004) 2004 Nomenclature for the chicken major histocompatibility ( B and Y )  
31 complex. *Immunogenetics* 56: 261-279
- 32 Miller MM, Goto RM, Taylor RL, Zoorob R, Auffray C, Briles RW, Briles E, Bloom SE  
33 (1996) Assignment of Rfp-Y to the chicken major histocompatibility complex/NOR  
34 microchromosome and evidence for high-frequency recombination associated with the  
35 nucleolar organizer region. *Proc Natl Acad Sci USA* 93: 3958-3962
- 36 Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility  
37 complex. *Heredity* 96: 7-21
- 38 Schaschl H, Wandeler P, Suchentrunk F, Obexer-Ruff G, Goodman SJ (2006) Selection and  
39 recombination drive the evolution of MHC class II DRB diversity in ungulates.  
40 *Heredity* 97: 427-437
- 41 Scheet P, Stephens M (2006) A Fast and Flexible Statistical Model for Large-Scale  
42 Population Genotype Data: Applications to Inferring Missing Genotypes and  
43 Haplotypic Phase. *Am J Hum Genet* 78: 629-644
- 44 Shiina T, Hosomichi K, Hanzawa K (2006) Comparative genomics of the poultry major  
45 histocompatibility complex. *Anim Sci J* 77: 151-162
- 46 Skjødt K, Koch C, Crone M, Simonsen M (1985) Analysis of chickens for recombination  
47 within the MHC (B-complex). *Tissue Antigens* 25: 278-282
- 48 Solinhac R, Leroux S, Galkina S, Chazara O, Fève K, Vignoles F, Morisson M, Derjusheva S,  
49 Bed'hom B, Vignal A, Fillon V, Pitel F (2010) Integrative mapping analysis of  
50 chicken microchromosome 16 organization. *BMC Genomics* 11: 616

1 Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology  
2 and conservation. *Front Zool* 2: 16

3 Stapley J, Birkhead TR, Burke T, Slate J (2008) A linkage map of the zebra finch  
4 *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics*  
5 179: 651-667

6 Stapley J, Birkhead TR, Burke T, Slate J (2010) Pronounced inter- and intrachromosomal  
7 variation in linkage disequilibrium across the zebra finch genome. *Genome Res* 20:  
8 496-502

9 The MHC Sequencing Consortium (1999) Complete sequence and gene map of a human  
10 major histocompatibility complex. *Nature* 401: 921-923

11 Tsuda TT, Tsuda M, Naruse T, Kawata H, Ando A, Shiina T, Fukuda M, Kurita M, LeMaho  
12 I, Kulski JK, Inoko H (2001) Phylogenetic analysis of penguin (Spheniscidae) species  
13 based on sequence variation in MHC class II genes. *Immunogenetics* 53: 712-716

14 Warren WC, Clayton DF, Ellegren H, et al. (2010) The genome of a songbird. *Nature* 464:  
15 757-762

16 Westerdahl H (2007) Passerine MHC: genetic variation and disease resistance in the wild. *J*  
17 *Ornithol* 148: 469-477

18 Wheat C (2010) Rapidly developing functional genomics in ecological model systems via 454  
19 transcriptome sequencing. *Genetica* 138: 433-451

20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

**Table 1** Information, gene annotation and chromosomal positions of all SNP markers used in this study (for sequences see Online Resource 2).

Markers with names beginning with “Tgu\_SNP” were typed in the original genetic map of Stapley et al. (2008). Names beginning with “tmptgu” were taken from the ENSEMBL polymorphism data and other markers where designed from transcriptome and BAC sequencing data. The locus designations of MHC class I and II genes follow the nomenclature of Balakrishnan et al. (2010).

SNP name	Gene name	Gene description	Chromosome	Position	Type
Tgu_SNP_00837	RPL17	60S Ribosomal protein L17	Z	1213283	Synonymous
Tgu_SNP_01249	ACAA2	Acetyl-coenzyme A acyltransferase 2	Z	1464495	3'UTR
Tgu_SNP_01537	ACAA2	Acetyl-coenzyme A acyltransferase 2	Z	1464313	3'UTR
Tgu_SNP_01362	LOC100223824	Similar to tetraspanin-3	Z	1609192	Downstream
Tgu_SNP_01309	FBP1	Fructose-1,6-bisphosphatase 1	Z	10110645	Synonymous
Tgu_SNP_01100	CLTA	Clathrin light chain A	Z	2053857	Synonymous
Tgu_SNP_01413	CLTA	Clathrin light chain A	Z	2057898	3'UTR
Tgu_SNP_00108	BHMT	Betaine--homocysteine S-methyltransferase	Z	56027090	Synonymous
Tgu_SNP_00366	PSIP1	PC4 and SFRS1-interacting protein (Lens epithelium growth factor)	Z	59919835	3'UTR
tmptgu2430803	RCL1	RNA terminal phosphate cyclase-like 1	Z	64096600	Intronic
tmptgu2430825	B-NK	NK receptor-like; Blec2	Z	64154902	Intronic
tmptgu2430826	CLEC2D	C-type lectin domain family 2, member D; Blec1	Z	64164380	Synonymous
Tgu_SNP_00708	GHR	Growth hormone receptor	Z	44207137	Upstream
Tgu_SNP_00517	SMU1	smu-1 suppressor of mec-8 and unc-52 homolog	Z	31446385	Downstream
Tgu_SNP_00109	BHMT	Betaine--homocysteine S-methyltransferase	Z	56026996	3'UTR
Tgu_SNP_00582	LMNB1	Lamin B1	Z	68946310	Downstream
Tgu_SNP_01007	TBCA	Tubulin-specific chaperone A	Z	56676486	Synonymous
Tgu_SNP_00183	VCAN	Versican core protein Precursor	Z	70951190	Synonymous
trim7.2-2	TRIM7.2	Tripartite motif protein 7	16_random	157419	Synonymous
Tgu_SNP_00725	GNB2L1	Guanine nucleotide binding protein (G protein), beta polypeptide 2-like 1; C12-3	16 (Chicken)	Unknown	Synonymous
Tgu_SNP_01547	CSNK2B	Casein kinase II subunit beta	16	270	Synonymous
brd2	BRD2	Bromodomain containing 2; RING3	UN	5958501	Synonymous
Tgu_SNP_01365	-	Unknown Zink finger protein	MHC BAC Contig 93	22810	Unknown
Tgu_SNP_01363	-	Unknown Zink finger protein	MHC BAC Contig 93	23065	Unknown
flot1	FLOT1	Flotillin 1	MHC BAC Contig 93	21298	Intronic
tgu_classI-7	-	MHC class I alpha chain (locus 1)	MHC BAC Contig 93	10081	3'UTR
tmptgu393923	-	MHC class II beta chain (locus 2)	UN	115905700	Synonymous
tmptgu347102	-	MHC class I alpha chain ( $\psi$ -locus C)	UN	70834777	Downstream

## Figure captions

**Fig. 1** Sex averaged genetic maps of zebra finch chromosomes 16 and Z. Linkage map position (cM) to the left and marker names to the right (with corresponding gene symbols within brackets). For brevity TGU\_SNPs are represented by the last 4 numbers only (e.g. TGU\_SNP\_00725 is represented by 0725). Markers “1363” and “1365” are situated in an unknown zink finger protein gene, “tgu\_classI-7” corresponds to MHC class I locus 1 and the marker “tmptgu393923” to MHC class IIB locus 2

**Fig. 2** Linkage disequilibrium heat map for zebra finch chromosome 16. In each square are the pairwise LD values ( $r^2 \cdot 100$ ) between markers and shading corresponds to the amount of LD, ranging from black for high LD ( $r^2 = 1$ ) to white for low LD ( $r^2 = 0$ ). On the top is the genetic map for the linkage group. For clarity TGU\_SNPs are represented by the last 4 numbers only (e.g. TGU\_SNP\_00725 is represented by 0725) and corresponding gene symbols are given within brackets. Markers “1363” and “1365” are situated in an unknown zink finger protein gene, “tgu\_classI-7” corresponds to MHC class I locus 1 and the marker “tmptgu393923” to MHC class IIB locus 2

**Electronic supplementary material**

Online Resource 1: Table with information on all the MHC related SNPs identified in this study (96SNPinfo.xls)

Online Resource 2: Sequences of all SNPs included in this study (SNPs.fas)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Fig. 1

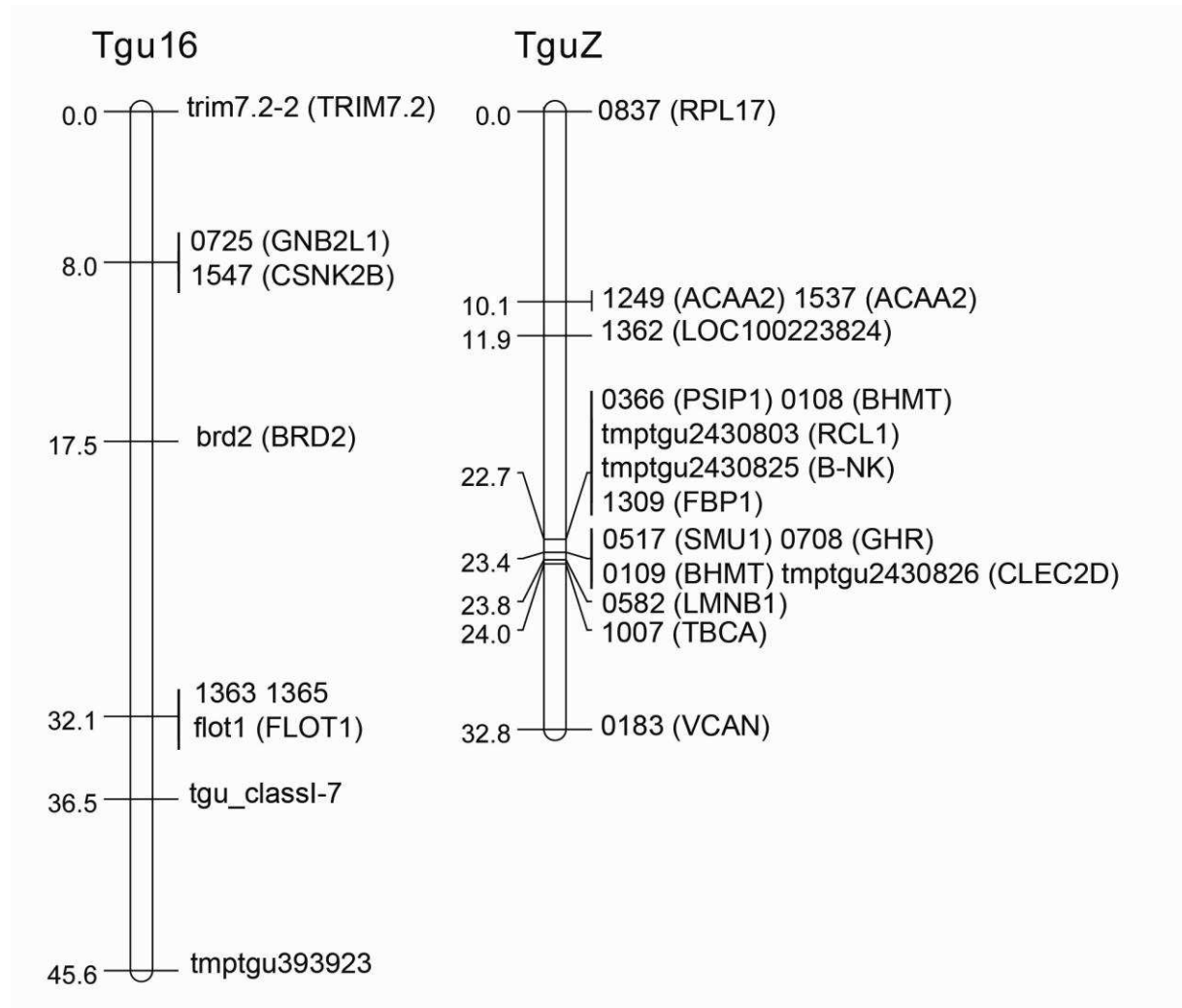
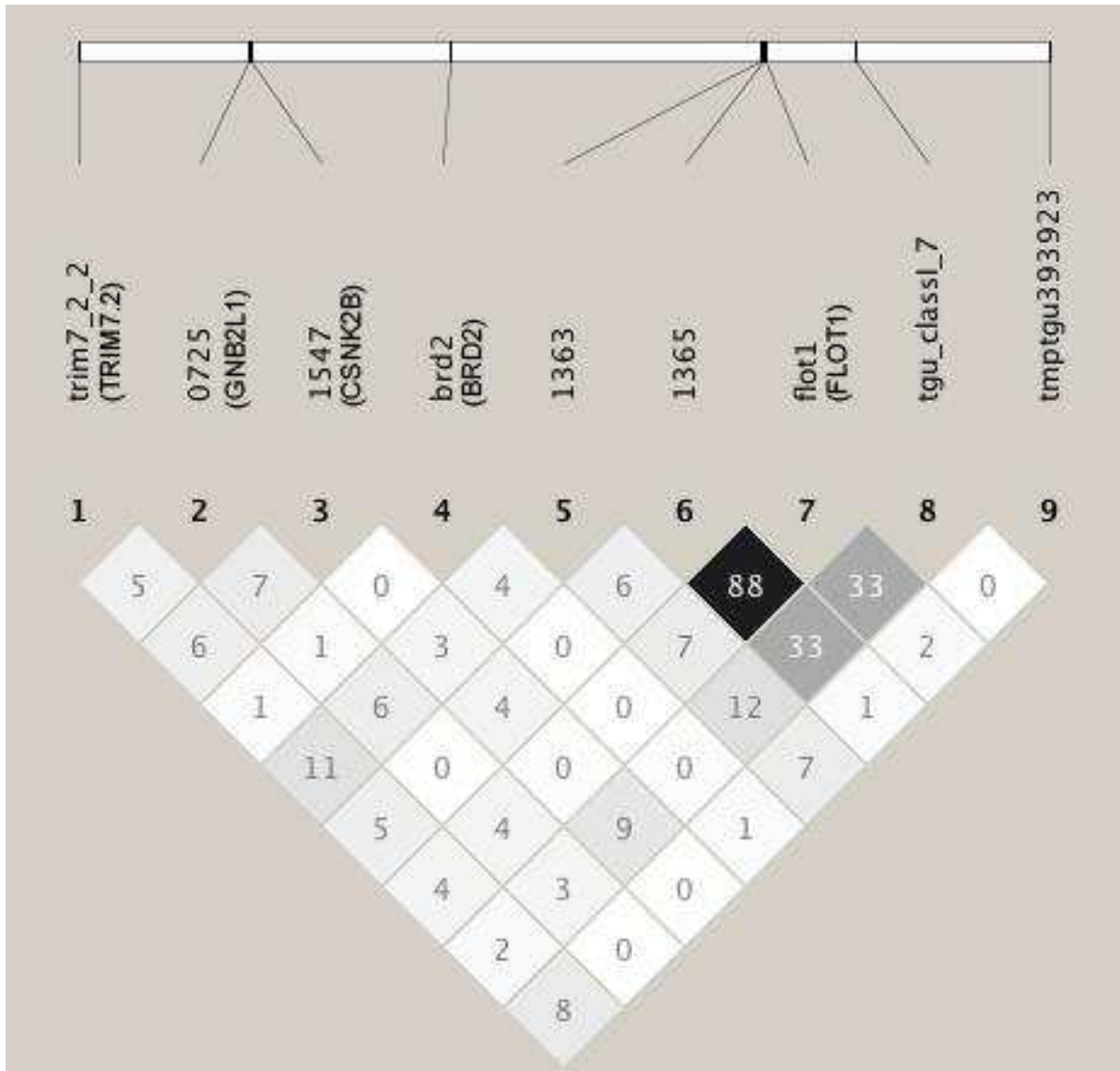
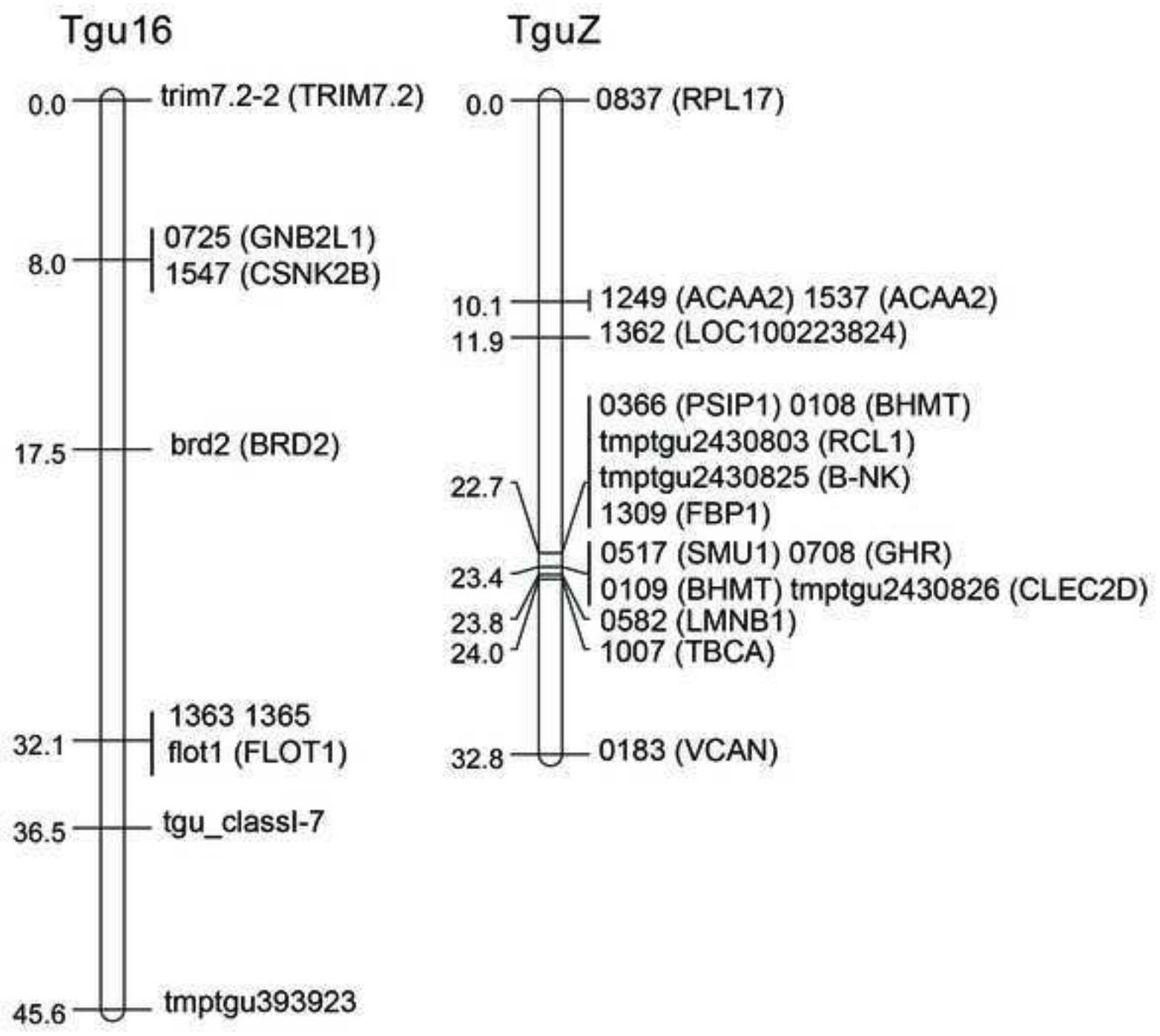


Fig. 2





**Figure 1**  
[Click here to download high resolution image](#)





**Online resource 1**

[Click here to download Supplementary Material: 96SNPinfo.xls](#)

**Online Resource 2**

[Click here to download Supplementary Material: SNPs.fas](#)