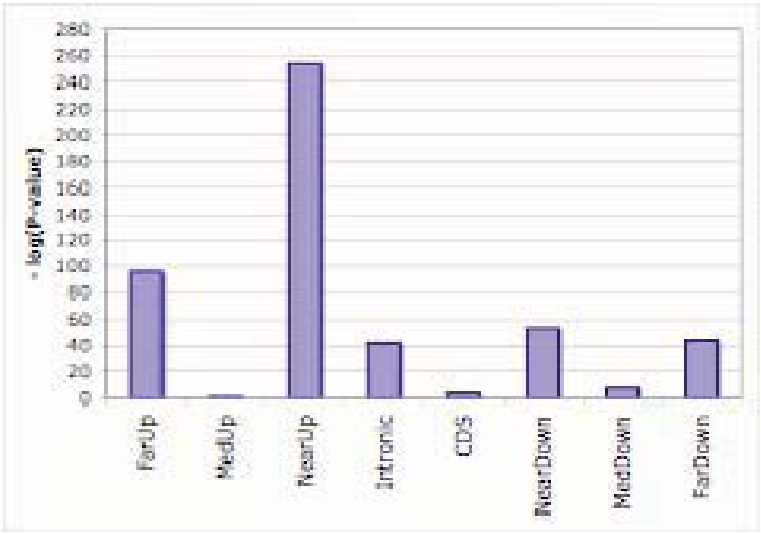
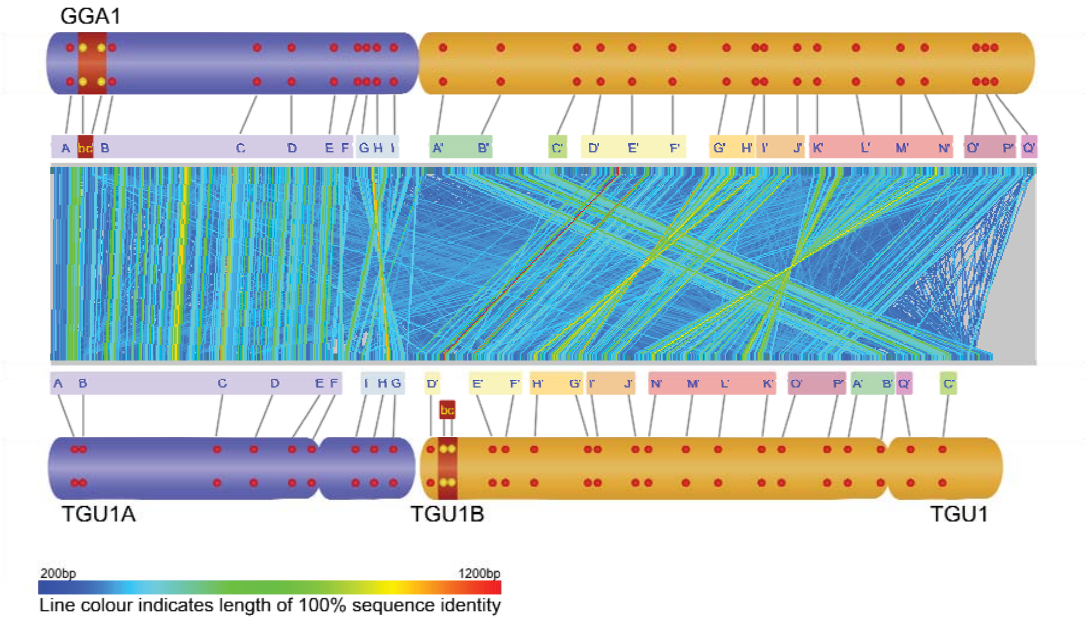
A

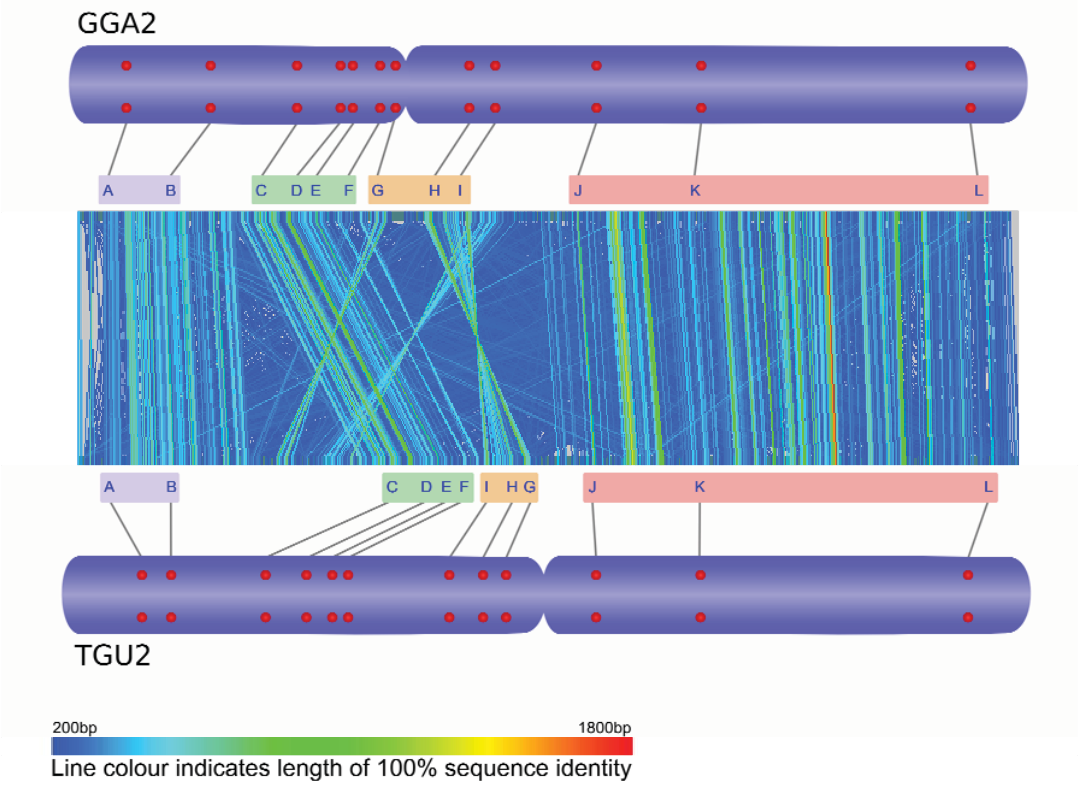
B

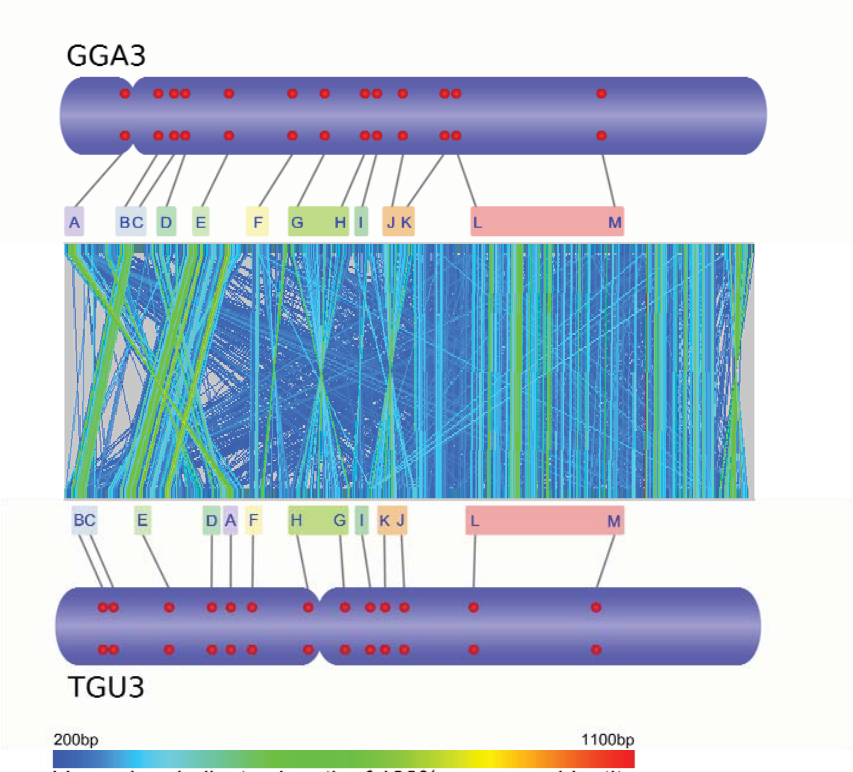


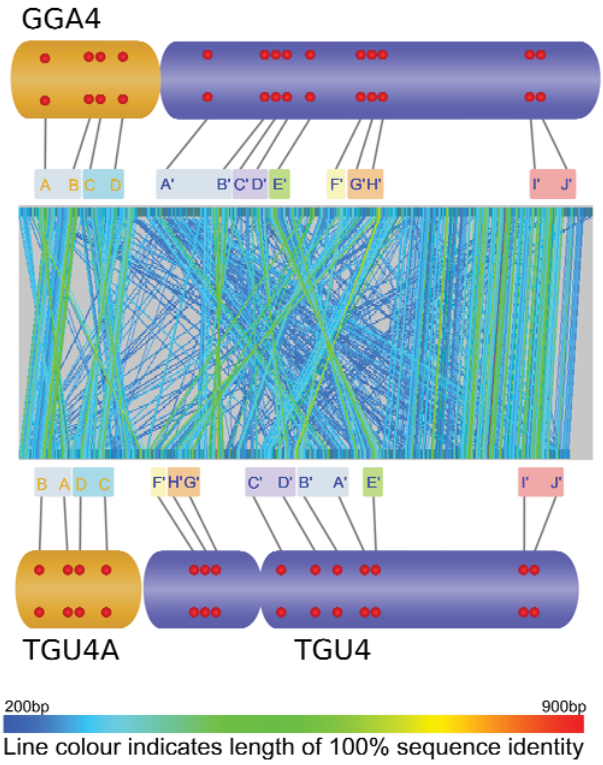

**Accessory Figure 1. Predicted motif genome location distribution.** (A) Density of predicted motifs (y axis) for different categories of regions (x axis) in terms of location with respect to their nearest genes, shown in blue. The horizontal pink line is the genome-wide average. See text for definitions of motif density and various region categories. (B) P-values of enrichment or depletion of motif occurrence in each category of regions, using one-tailed Fishers exact tests. Negative logarithms are shown.

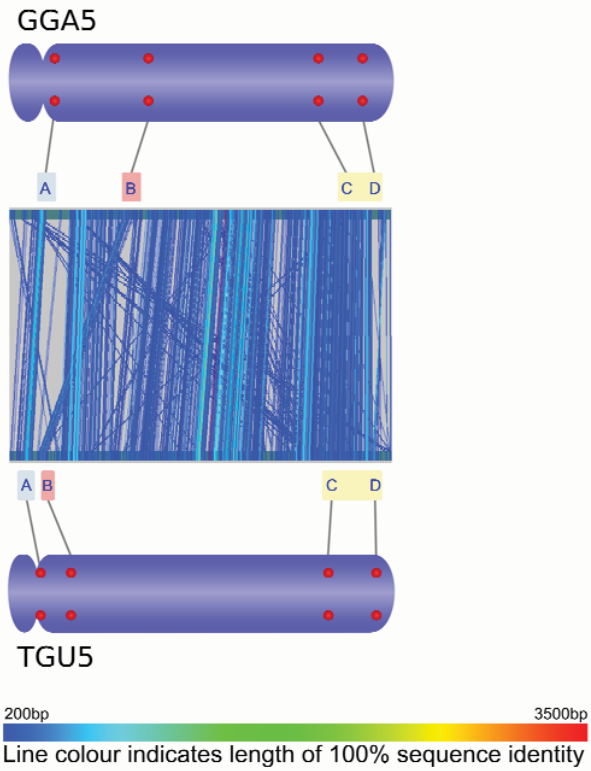
**Accessory Figure 2. Motif target counts by defined location.** Counts of targets windows of individual motifs in different categories of regions are compared to the respective expected values, with colors shown indicating whether the count is greater or less than expectation. Green cells correspond to counts that are higher than the average and red cells correspond to regions with below-average counts.

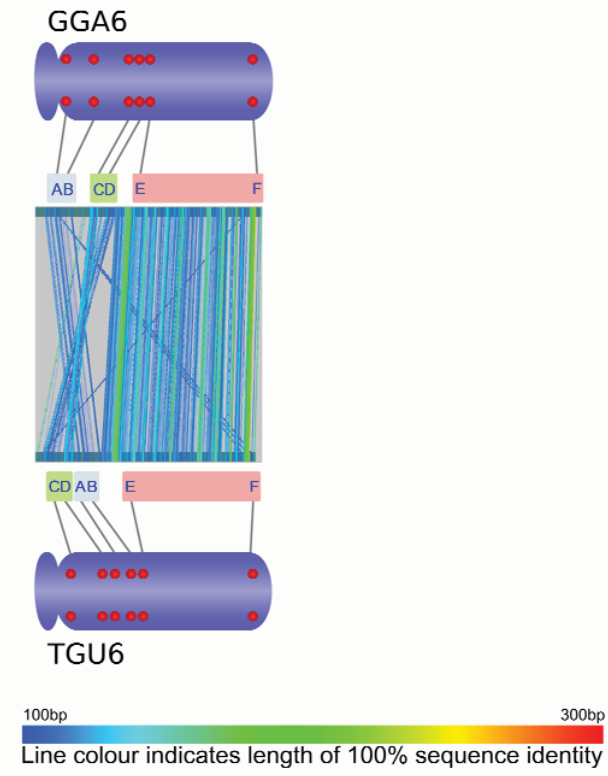

GGA1
TGU1A
TGU1B
TGU1
200bp
1200bp
Line colour indicates length of 100% sequence identity

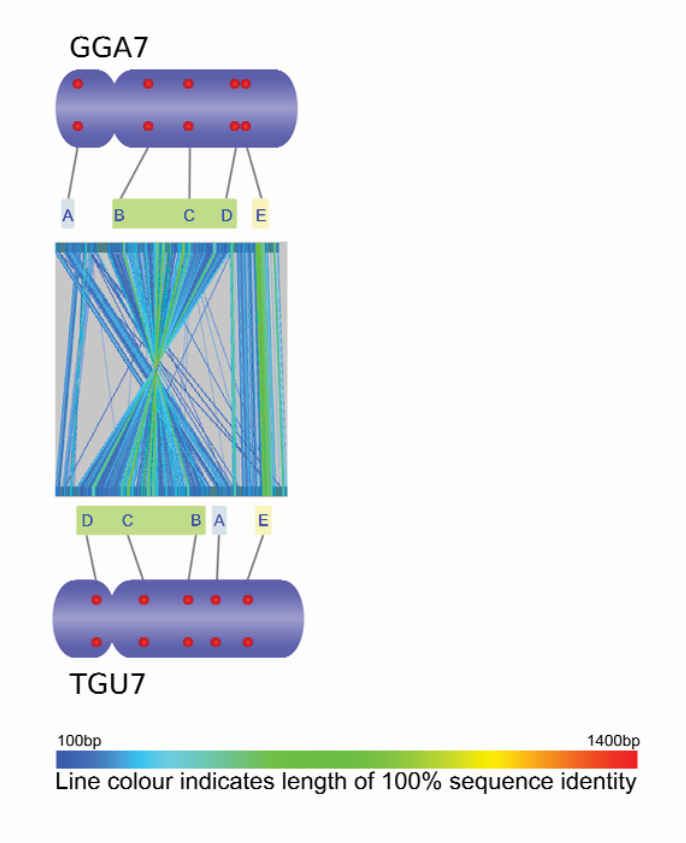

GGA2
TGU2
200bp
1800bp
Line colour indicates length of 100% sequence identity

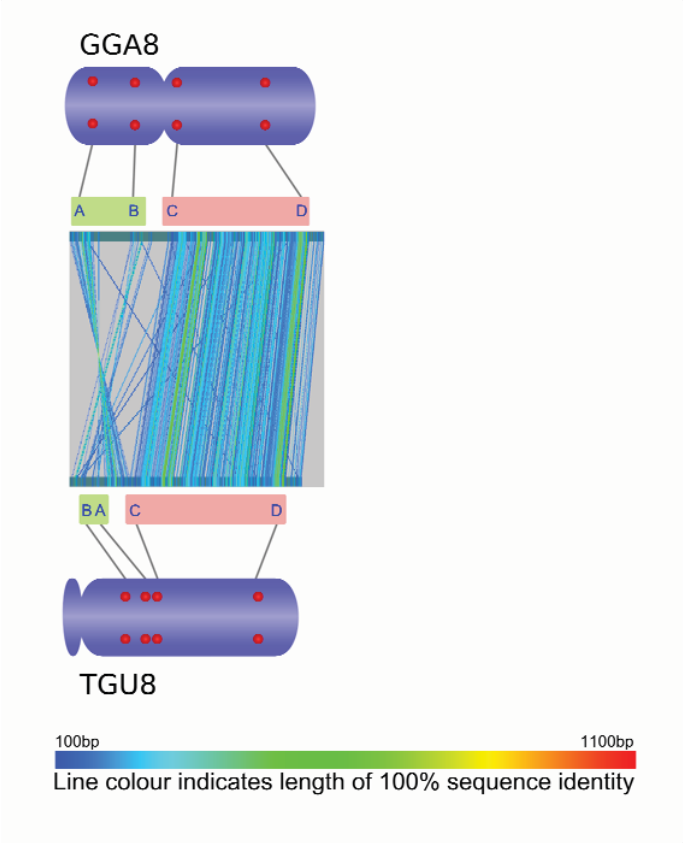

GGA3
A
BC
D
E
F
G
H
I
J K
L
M
BC
E
D A F
H
G
I
K J
L
M
TGU3
200bp
1100bp
Line colour indicates length of 100% sequence identity

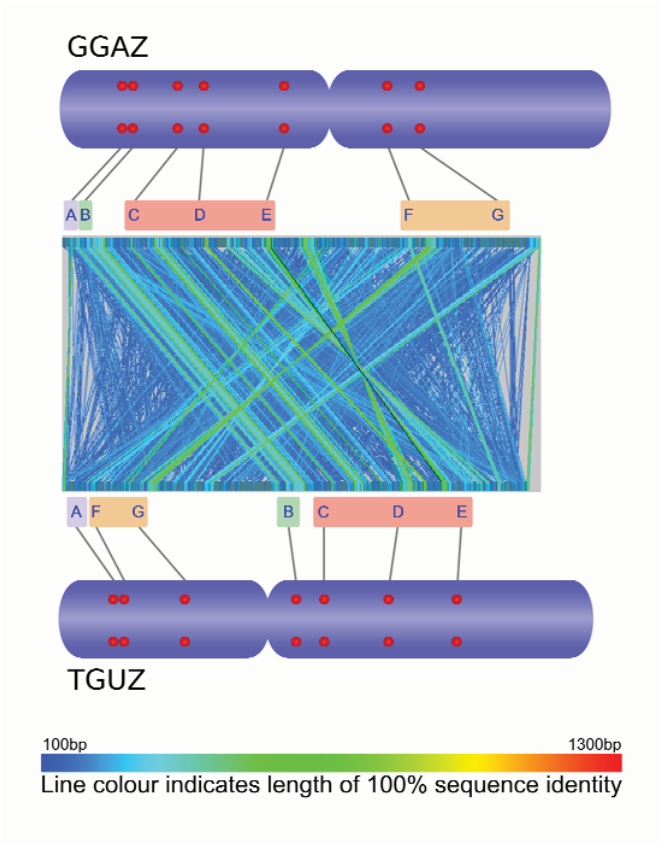

GGA4
A B C D
A'
B' C' D' E'
F' G' H'
I' J'
B A D C
F' H' G'
C' D' B' A'
E'
I' J'
TGU4A
TGU4
200bp
900bp
Line colour indicates length of 100% sequence identity


GGA5
A
B
C
D
A
B
C
D
TGU5
200bp
3500bp
Line colour indicates length of 100% sequence identity

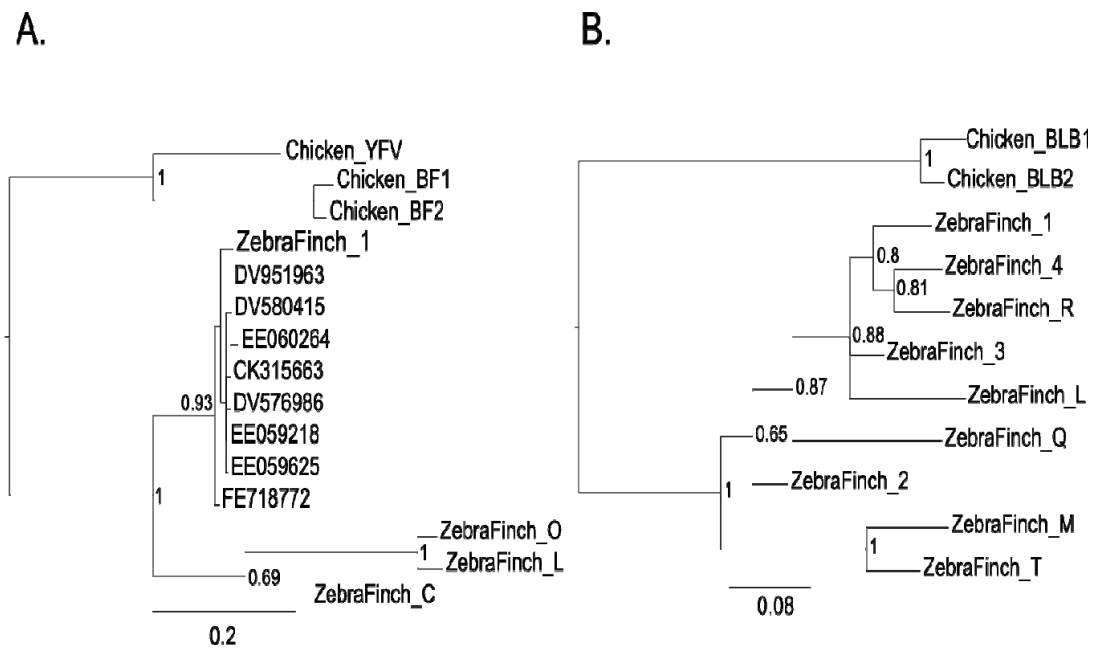

GGA6
AB
CD
E
F
CD
AB
E
F
TGU6
100bp
300bp
Line colour indicates length of 100% sequence identity


GGA7
A
B
C
D
E
D
C
B
A
E
TGU7
100bp
1400bp
Line colour indicates length of 100% sequence identity

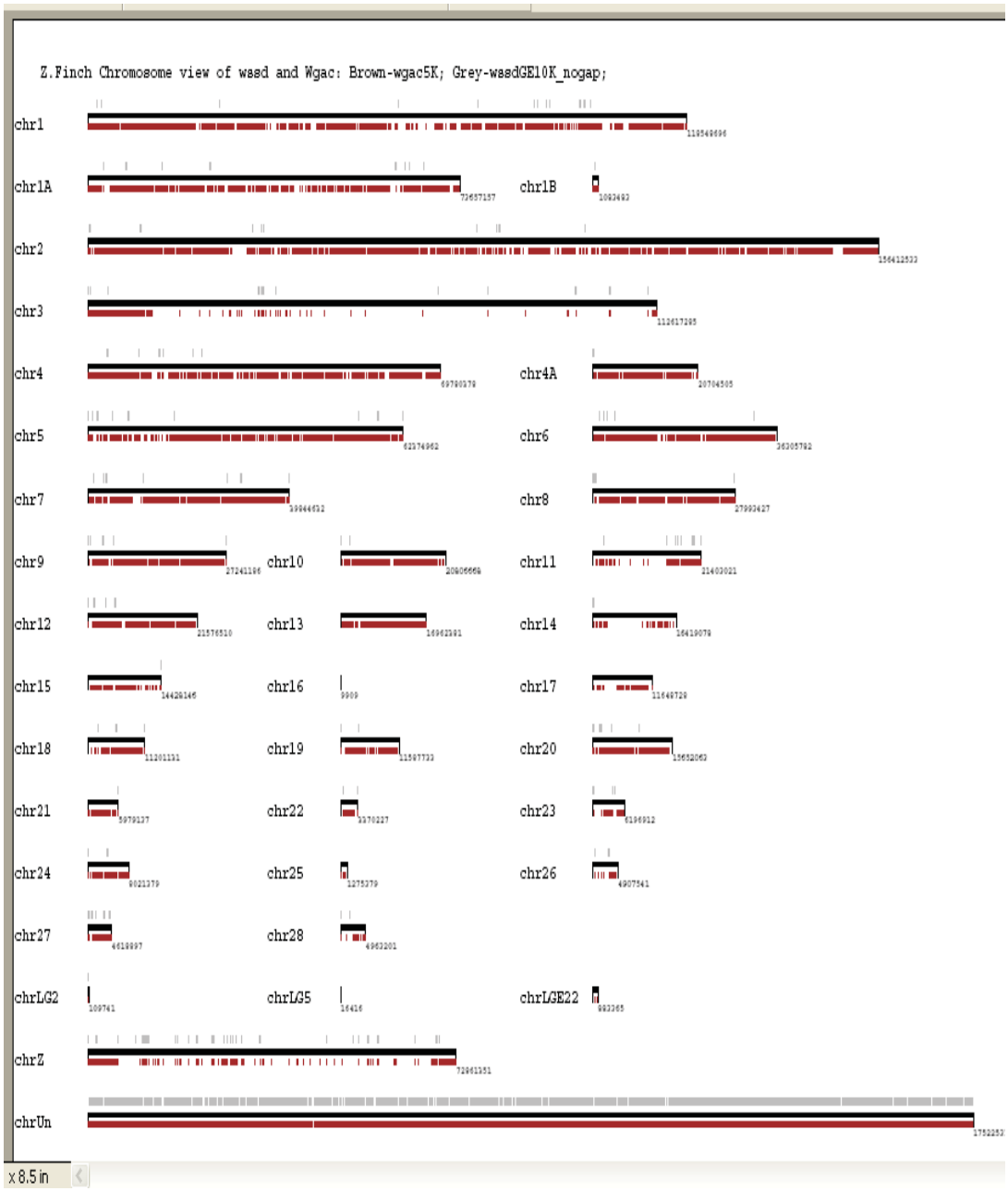

GGA8
A
B
C
D
B A
C
D
TGU8
100bp
1100bp
Line colour indicates length of 100% sequence identity




**Accessory Figure 3:** Comparative analysis of marker order on chicken chromosomes 2-8 and Z (GGA2-8, GGAZ) and their zebra finch orthologues (TGU2-8, TGUZ). The central part of each figure was created by aligning whole chromosomal sequences using the program GenAlyzer. Line colour indicates the length of sequences with 100% sequence identity. The tentative chromosomal rearrangements suggested by this analysis were verified using fluorescent in situ hybridization (FISH). Letters indicate the position of chicken and zebra finch BACs with orthologous sequence content in the genome sequences of both species (see accessory file Physical mapping table 2009-09-16.xls for details on the FISH probes used). Red dots on the ideograms illustrate the physical chromosomal position as determined by FISH.



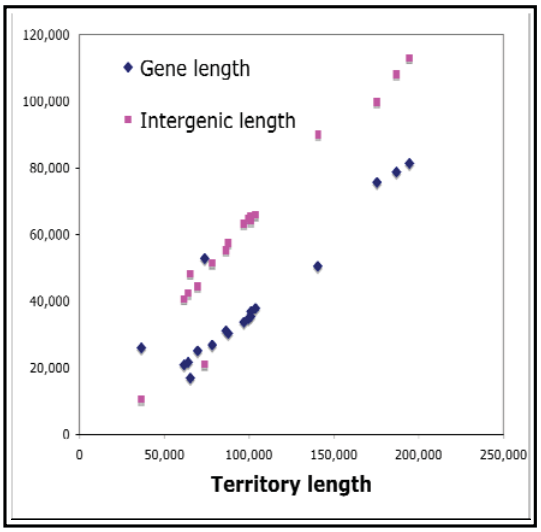

**Accessory Figure 4. Bayesian phylogenies of zebra finch MHC genes.** (A) Class I genes and (B) Class IIB genes were compared to sequences from the chicken MHC-B complex. Putatively functional zebra finch genes with open reading frames are given numerical suffixes and putative pseudogenes are given lettered suffixes. For Class I we also include a chicken sequence from the MHC-Y region (YFV). For zebra finch class I, we show the placement of eight brain ESTs (indicated by their GenBank accession numbers) supporting the expression of MHC Class I genes in the brain. Posterior probabilities are given for well-supported nodes in the tree.




**Accessory Figure 5**. General view showing WGAC (>5kb) and WSSD on all chromosomes. Grey above lines is WSSD and red below lines is WGAC.. ChrUn was treated as a “distinct” chromosome.

| Gene set | size | terr_len | gene_len | intergenic_len | p-value |
|---|---|---|---|---|---|
| All genes | 8438 | 87,768 | 30,355 | 57,413 | - |
| Group 1 | 82 | 61,667 | 20,973 | 40,694 | 6.9E-3 |
| Group 2 | 48 | 65,301 | 16,967 | 48,334 | 4.0E-1 |
| Group 3 | 8 | 73,868 | 52,831 | 21,037 | 9.3E-2 |
| Group 4 | 210 | 194,474 | 81,456 | 113,018 | 1.0E-22 |
| Group 5 | 95 | 175,420 | 75,682 | 99,738 | 1.5E-8 |
| Group 6 | 2 | 36,521 | 26,028 | 10,493 | 7.3E-1 |
| Group 7 | 1100 | 100,885 | 35,416 | 65,470 | 2.0E-10 |
| Group 8 | 1049 | 78,327 | 26,893 | 51,433 | 2.1E-8 |
| Group 9 | 390 | 69,660 | 25,217 | 44,443 | 8.3E-7 |
| Group 10 | 376 | 140,510 | 50,507 | 90,003 | 2.0E-17 |
| down_only (4,5,8) | 1312 | 101,048 | 36,972 | 64,076 | 9.7E-1 |
| fast_down (4,5,6) | 301 | 186,840 | 78,831 | 108,009 | 1.7E-28 |
| fast_up (1,2,3) | 136 | 64,057 | 21,715 | 42,343 | 3.9E-3 |
| slow_down (3,5,8) | 1136 | 86,430 | 31,167 | 55,264 | 1.4E-4 |
| slow_up (2,6,7) | 1138 | 99,655 | 34,799 | 64,856 | 9.3E-10 |
| switch (3,6,9,10) | 768 | 103,752 | 37,866 | 65,886 | 3.3E-2 |
| up_only (1,2,7) | 1212 | 96,910 | 33,772 | 63,138 | 2.4E-7 |

A)




B)

**Accessory Figure 6.** (A) Characteristics of co-expressed gene sets from Dong et al. [58] (See S3 supplementary notes). “Gene set”: name of the gene set, as in the original paper. “All genes” refers to the genes on the array in Dong et al. Numbers in parentheses indicate component subsets of a set. “Size”: number of genes in set. “terr_len”: average gene territory length of a gene set. “gene_len”: average coding sequence length. “intergenic_length”: average of (territory length – gene length). “p-value”: statistical significance of enrichment for short (pink cells) or long (green cells) territories, as measured by 2-tailed Wilcoxon Rank Sum tests. (B) Average gene length and intergenic length of a gene set (y axis) versus average gene territory length. Each point corresponds to a gene set.