UPPSALA
UNIVERSITET

CpG islands, but not their methylation level, are key in the regulation of meiotic recombination in chicken (*Gallus gallus*)

Ivar Westerberg

Degree project in biology, Bachelor of science, 2019
Examensarbete i biologi 15 hp till kandidatexamen, 2019
Biology Education Centre and Department of ecology and genetics, Uppsala University
Supervisors: Takeshi Kawakami and Carina Farah Mugal

Contents

Abstract

Meiotic recombination plays a fundamental role in many sexually reproducing species. Recombination shuffles the genetic material during the first meiotic cell division resulting in new combinations of alleles within each chromosome. In many organisms, the rate of recombination is not uniform across the genome but consists of so called hotspots where the recombination rate is remarkably higher than the genome average. In mammals, the regulation and location of recombination hotspots is regulated by a gene called PRDM9. Many non-mammalian animals, like birds, lack this gene and the precise mechanism for recombination rate regulation is still unknown. Previous findings in passerine birds have observed an association between recombination rate and a genomic feature known as CpG islands (CGIs). CGIs are often located in promoter regions of genes and depending on their methylation status constitute accessible chromatin regions. It has therefore been suggested that the proteins involved in the regulation of recombination have better access to less condense chromatin regions. In this study, I tested if the association between recombination rate and CGIs found in passerine birds is also true in chicken. I also tested if methylation levels of CGIs play a role in recombination rate regulation in chicken. To this end, I used previously published data for CGI locations and a methylation map in chicken, and unpublished data of recombination rate estimates. I found that the association between recombination rate and CGIs observed in passerine birds extends to chicken, suggesting that this is an ancestral trait in birds. I did not, however, find a negative association between methylation levels and recombination rate as hypothesised based on a relationship between methylation level and chromatin accessibility. This suggests that DNA methylation level at CGIs is not a strong determinant of recombination in chicken, although there may be some workflow artefacts or unknown factors remaining in my analysis obscuring the relationship between these two variables.

Introduction

Meiotic recombination is a fundamental process in sexually reproducing species. It shuffles genetic materials on homologous chromosomes and creates new combinations of alleles within each chromosome during meiosis. The shuffling of alleles can result in both a creation of new beneficial allelic combinations and separation of already beneficial allelic combinations. Therefore, the impact of recombination on the offspring's fitness is likely dependent on the fitness of the previous generations in a given environment. Consequently, detailed understanding of variation in recombination rate is key toward a better understanding of fitness consequences in sexual reproduction as well as genome divergence between populations.

Meiotic recombination correlates with several features in the genome. One of them is the correlation between the rate of recombination and the amount of C and G nucleotides, referred to as GC content (Marsolier-Kergoat & Yeramian 2009, Berglund *et al.* 2014, Singhal *et al.* 2015, Bolívar *et al.* 2016, Dutta *et al.* 2018). This is primarily mediated by a process called GC-biased gene conversion (gBGC), which happens during meiotic recombination where G and C alleles have a higher fixation probability than A and T alleles (Duret *et al.* 2006). Although the exact mechanistic causes underlying gBGC are not yet known, this leads to a positive correlation between recombination rate and GC content (Han *et al.* 2008, Webster & Hurst 2012). gBGC has been shown to have a large impact on the evolution of base composition in the genomes of several species, including birds and mammals (Bolívar *et al.* 2016, Dutta *et al.* 2018). Therefore, in addition to impacts of recombination on organismal fitness, variation in recombination rate can also affect genome wide patterns of molecular evolution. Another mechanism that affects the evolution of GC content is CpG hypermutability (Bird 1980). Cytosine is sporadically converted to thymine by deamination, particularly if cytosine is methylated and converted into 5-methylcytosine (Cooper et al. 2010)). Methylation of cytosine in CpG dinucleotides is the most common type of DNA methylation (Bird & Taggart 1980). Mutation rate of methylated C is about 10 times higher than the normal transition mutation rate (Cooper & Gerber-Huber 1985, Sved & Bird 1990). This CpG hypermutability decreases GC content in genomic regions with high CpG methylation (Bird 1980, Mugal *et al.* 2015). Because decreased CpG methylation level decreases C-to-T mutation rate, this has been proposed as one explanation to the higher GC content in promoter regions since promoter regions generally are hypomethylated (Berglund *et al.* 2014). Overall genome-wide distribution of GC content is, hence, determined by a

balance between recombination rate (GC-increasing) and CpG methylation rate (GC-decreasing).

Recombination is not uniform across the genome but is instead concentrated in small regions, so called recombination hotspots which show high recombination rate compared to the genome average. Recombination hotspots have been observed in many taxonomic groups (Dooner & Martínez-Férez 1997, Gerton *et al.* 2000, Paigen & Petkov 2010, Kawakami *et al.* 2017). In mammals, location and regulation of the recombination hotspots is determined by a gene called PRDM9 (Myers *et al.* 2010, Parvanov *et al.* 2010, Auton *et al.* 2012). PRDM9 is a DNA binding protein, containing a KRAB domain, a PR/SET domain and a zinc finger domain, which recognises specific DNA sequence motifs. PRDM9 adds trimethylation on lysine 4 of histone H3 (H3K4me3), which is an epigenetic modification required for the initiation of DNA double strand breaks needed for recombination (Hayashi *et al.* 2005). H3K4me3 modifications are placed by the PR/SET domain on adjacent nucleosomes, which attracts other proteins of the recombination machinery (Baudat *et al.* 2013). Location of recombination hotspots in mammals are different to location of hotspots in species that lack a PRDM9 gene. In mammals, recombination hotspots tend to be located away from promoter regions, while in species that lack the PRDM9 gene recombination hotspots are more often found close to promoter regions (Baker *et al.* 2017). For instance, PRDM9 knockout mice show hotspots in promoter more similar to species without PRDM9, which is rarely the case in wildtype mice (Hayashi *et al.* 2005, Brick *et al.* 2012). Another example includes dogs where the PRDM9 gene has lost its function by pseudogenization (or accumulating a loss-of-function mutation) but recombination hotspots is highly localised in promoter regions. Hotspots in dogs are therefore located at similar locations as in species without PRDM9 (Axelsson *et al.* 2012, Berglund *et al.* 2014).

The gene PRDM9 has been identified in several metazoans including species of sea anemone, nematodes and mammals among others (Oliver *et al.* 2009, Ponting 2011), suggesting the ancient origin of this gene. However, there are several animal groups that do not have the PRDM9 gene. One of these groups are birds (Singhal *et al.* 2015). Despite the absence of PRDM9, birds still show a highly variable recombination rate along a genome with distinct recombination hotspots (Singhal *et al.* 2015, Kawakami *et al.* 2017). Studies have shown that recombination hotspots were associated with CpG islands (CGIs) in birds (Kawakami *et al.* 2017). Nevertheless, there are also promoters without CGIs in the vertebrate genome (Davuluri *et al.* 2001, Saxonov *et al.* 2006). Recombination hotspots in birds are enriched

near transcription start sites (TSSs) and transcription termination sites (TTSs). This is thought to be driven by the colocalization of recombination hotspots with promoter-associated CGIs and other functional elements (Singhal *et al.* 2015, Kawakami *et al.* 2017). For example, Kawakami *et al.* (2017) noted that the mean recombination rate in old world flycatchers (genus *Ficedula)* was higher in promoters with CGIs than ones without. However, these studies did not provide any evidence to explain the mechanisms that result in the association between recombination and CGIs.

Previous studies in plants, animals and yeast have suggested that recombination initiating proteins can get access to less condensed chromatin regions, such as less methylated CGIs (Berchowitz *et al.* 2009, Brachet *et al.* 2012, Berglund *et al.* 2014, Shilo *et al.* 2015). The level of DNA methylation has been suggested to be one of the major factors correlated with recombination in humans (Webster & Hurst 2012). A study in dogs, which lack a functional PRDM9 gene, has also shown that recombination hotspots were associated with unmethylated CGIs (Berglund *et al.* 2014). Therefore, the level of CpG methylation is suggested to be a key determinant of the frequency of recombination in species without PRDM9; however, there are only a limited number of studies directly demonstrating the relationship.

The level of CpG methylation is intimately related to chromatin structure. The chromatin is a structurally dynamic molecule that both enables the whole genome to be packed inside of the nucleus and regulates the accessibility of the DNA for several processes including replication, transcription, DNA repair and recombination (Venkatesh & Workman 2015). Methylation on the DNA changes chromatin structure to be more tightly packed, whereas lower methylation levels lead to a more open chromatin structure. There are several studies in plants and yeasts suggesting that meiotic recombination is highly associated with chromatin structure (Berchowitz *et al.* 2009, Shilo *et al.* 2015). Because epigenetic regulation of CGIs has been shown to be associated with chromatin structure in plants (Choi *et al.* 2013, Shilo *et al.* 2015, Yelina *et al.* 2015a), yeasts (Lam & Keeney 2015), and insects (Niehuis *et al.* 2010), it might be hypothesised that the level of DNA methylation impacts the openness of the chromatin structure and as a result affects the recombination landscape.

Given the phylogenetic distribution of PRDM9-driven regulation of recombination, the CpG-associated recombination regulation appears to be an ancestral mode; however, this is a subject of debate (Baker *et al.* 2017). It is therefore important to study and compare the different systems to gain further insight into this question. By studying species that lack

PRDM9, such as birds, one can learn more about the history and evolution of the different mechanisms of recombination regulation. Here, I test whether the level of CpG methylation in promoters with CGIs (hereafter referred to as CGI promoters) is associated with recombination via modification of the chromatin structure and the accessibility of the DNA, using chicken as a model organism.

Chicken (*Gallus gallus*) offers an excellent opportunity to test this hypothesis because of its phylogenetic position, availability of various genomic resources and a fine-scale recombination map. First, chicken represents a basal lineage in the bird phylogeny (Jarvis *et al.* 2014, Prum *et al.* 2015), likely representing an ancestral mode of recombination regulation in birds. Second, chicken has been used as an avian model for studying development, comparative genomics, and functional genomics to explore genome function. Thus, many genomic resources, such as highly contiguous reference genomes, gene and repeat annotation, transcriptome data, epigenetic maps and pedigree-based recombination maps are available (Stern 2004, Burt 2007, Groenen *et al.* 2009). Finally, fine-scale recombination rate estimates (Kawakami unpublished) and methylation data for sperm and testis (Mugal *et al.* 2015) are available. By using these genomic resources, I tested if recombination rate is associated with CpG methylation in promoters by using recombination data and methylation data from the chicken genome. The methylation data in sperm was used because this tissue may sustain DNA methylation levels in meiosis. I also used testis as a somatic tissue that is expected to have different pattern of DNA methylation (He *et al.* 2018) I focused on promoter regions with CGIs to see if recombination rate correlates with methylation levels.

Methods

To get an estimate of the recombination rate in the chicken genome I used raw variant call genome data from a previous study (Li *et al.* 2017). An estimate of the linkage disequilibrium (LD), recombination rate rho (ρ) and CGI positions from the Li *et al* 2017 dataset was used (Kawakami unpublished data). Briefly, following Kawakami *et al*. (2017), $\rho = 4N_e r$ (where N_e denotes the effective population size, and r denotes the recombination rate) was estimated using LDhelmet (Chan *et al*., 2012). CGIs were annotated by using CpGcluster version 1.0 (Hackenberg *et al*., 2006) with default parameter settings and a minimum length of at least 50 bp.

To characterise the distribution of recombination along the chicken genome, the genome was divided into eight different categories: (i) All CGI, (ii) promoter CGI (prom CGI+), (iii) promoters without CGI (prom CGI-), (iv) first introns, (v) other introns, (vi) first exons, (vii) other exons and (viii) intergenic regions. A reference assembly of the chicken genome, galGal4 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000002315.3/) was used to define these regions. Promoter regions were defined as 2 kb upstream of TSSs of genes following Kawakami *et al*. (2017). The intersect function was used from BEDTools v2.27.0 (Quinlan & Hall 2010) to sort out promoter regions with CGIs and to define intergenic and intronic regions. The intersect function from BEDTools v2.27.0 was used to get the LD recombination rate ρ for the different regions. R version 3.4.4 was used to calculate the ρ/kb within the different regions. The recombination rate ρ/kb was then plotted for the different regions.

To test the relationship between recombination rate and CpG methylation, whole genome methylation data was used from both chicken sperm and testis samples sequenced by Mugal *et al* 2015. Because the methylation data were made for the previous galGal3 annotation, liftOver tool from the UCSC Genome Browser (Meyer *et al.* 2013) was used to convert the data into galGal4 format. All methylation data for non-CpG sites were removed. Mean CpG-methylation were calculated for all CGIs in the genome by using a custom perl script. The intersect function in BEDTools v2.27.0 was then used to get the methylation levels of CGIs in: the whole genome, promoter regions, intronic regions and intergenic regions. All statistical tests were carried out by R (version 3.4.4).

Results

Chicken LD recombination rate ρ was significantly higher in CGIs with a mean recombination rate of 14.291 ρ/kb compared to the genome average of 1.208 ρ/kb (Wilcoxon rank sum test, $p < 10^{-16}$). The recombination rate in CGI promoters was also significantly higher than in non-CGI promoters (Figure 1). The mean recombination rate in promoter regions with CGIs was 14.29 ρ/kb compared to 9.42 ρ/kb in promoters without CGIs (Wilcoxon rank sum test $p < 10^{-16}$).

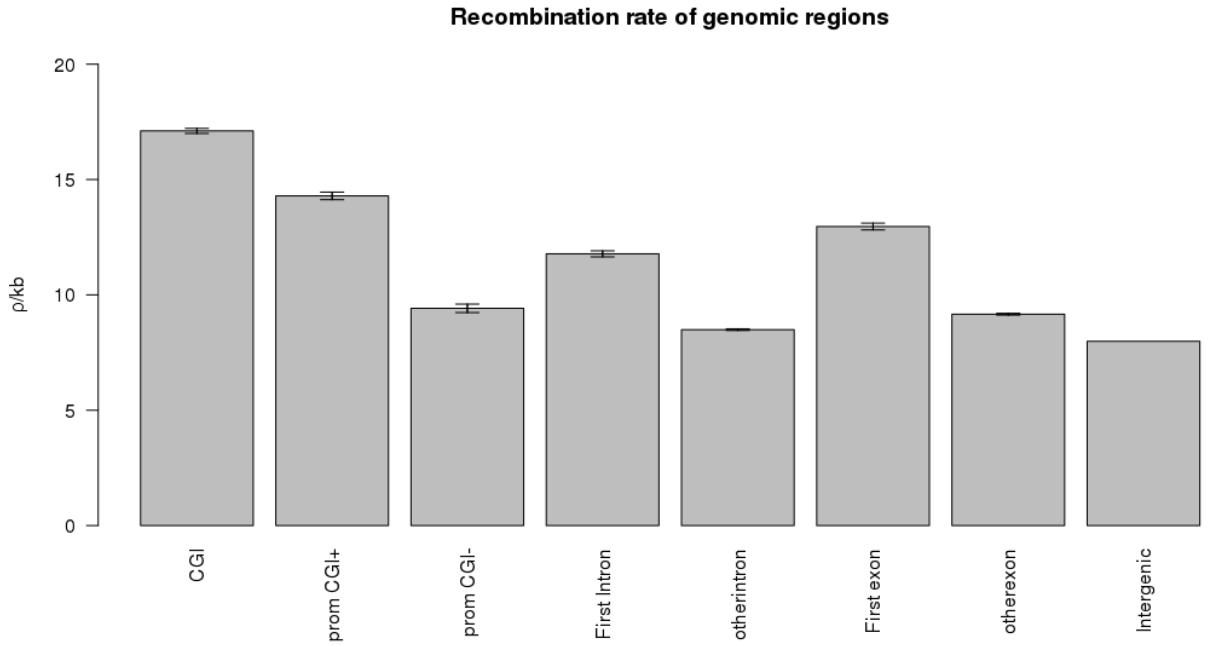

Figure 1: *LD recombination rate ρ/kb in 8 different genomic regions. (i) All CGI, (ii) promoter CGI (prom CGI+), (iii) promoters without CGI (prom CGI-), (iv) first introns, (v) other introns, (vi) first exons, (vii) other exons and (viii) intergenic regions. Error bars show the standard error for all regions except for intergenic where there was only one observation. The genome average recombination rate was 1.208 ρ/kb.*

Next, I tested if CpG methylation was associated with recombination rate in CGIs (Figure 2). I tested this for four different regions: All CGI irrespective of their location, promoter CGI, intronic CGI and intergenic CGI. Methylation levels were divided into size intervals of 10 %. The analysis was limited to the data with a methylation level less than 60 % because of the small number of observations of CGIs with a higher methylation level. All four regions showed the highest recombination rate at CGIs with moderate methylation levels (20-40 %) (Kruskal wallis test, $p < 10^{-16}$).

When testing for a difference in methylation levels between the four regions (All CGIs, promoter CGIs, intronic CGIs and intergenic CGIs) (Figure 3), all regions had a high frequency of locations with low methylation (< 10%). All regions except for promoter regions also had an increase in frequency around methylation levels of 50 %. This pattern was much less pronounced in promoter CGIs. The mean methylation level in promoters was significantly lower (9.9 %) than for the rest of the regions (~25 %) (Wilcoxon rank sum test, $p < 10^{-16}$).

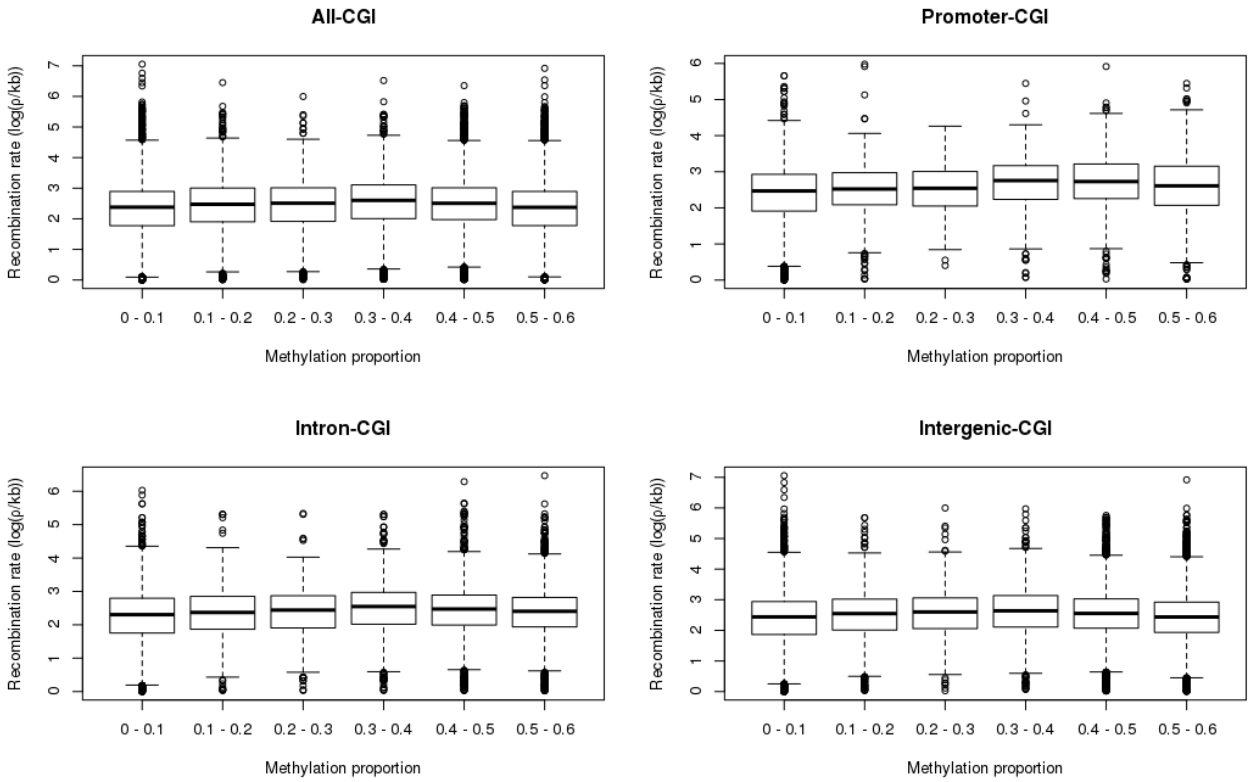

Figure 2: *Relationship between LD recombination rate, log(ρ/kb), on the y-axis and methylation percentages divided into intervals of size 0.1 on the x-axis in four different classes of genomic CGIs. All-CGI refers to every CGI region irrespective of their location, Promoter-CGI featuring only CGIs in promoter regions, Intron-CGI featuring only intronic CGI regions and lastly Intergenic-CGI which features CGI in intergenic regions. Within the boxes are the first and the third quartile and the thick line indicates the median. The analysis was limited to methylation levels less than 0.6.*

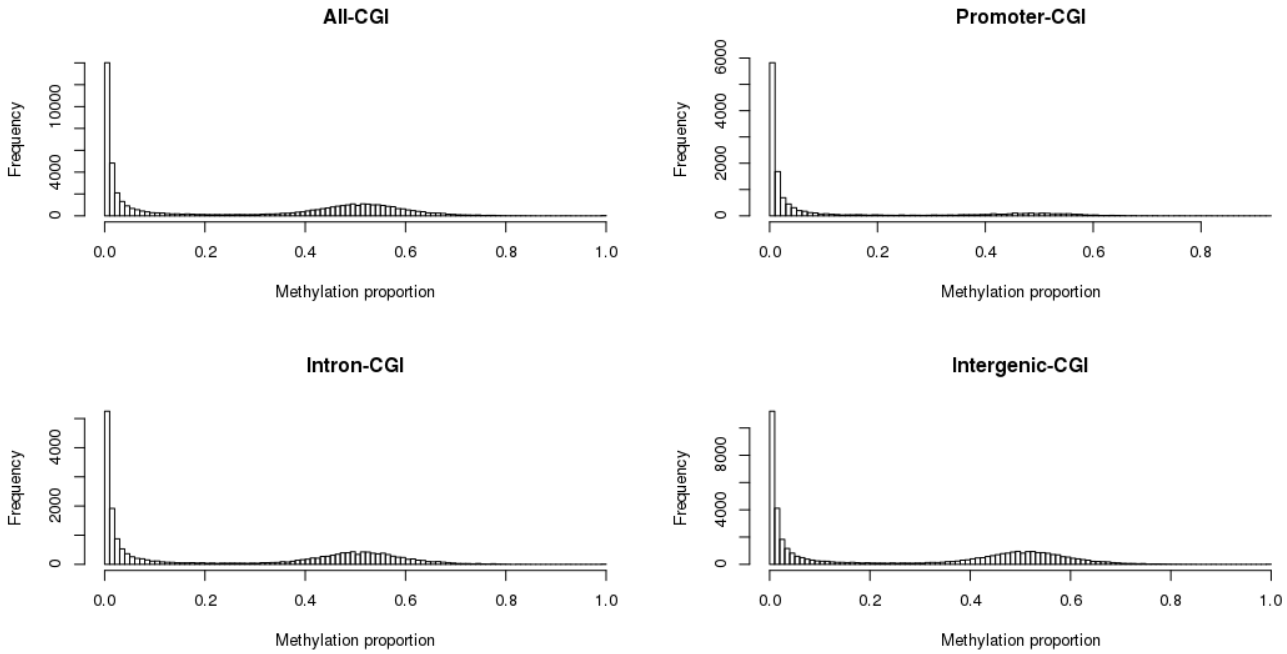

Figure 3: *Distribution of methylation levels at CGIs in the four different genomic regions with frequency of observed data points on the y-axis and methylation proportion on the x-axis. All-CGI refers to every CGI region irrespective of their location, Promoter-CGI featuring only CGIs in promoter regions, Intron-CGI featuring only intronic CGI regions and lastly Intergenic-CGI which features CGI in intergenic regions*

The same pattern was tested for chromosome 1 to disregard possible chromosome size effects of the small microchromosomes. The same pattern as seen for the whole genome was also observed for chromosome 1, all regions show a slight increase in recombination rate around 20-40 % methylation and promoter regions show a slight overall increase (Figure 4). The methylation frequency pattern is also very similar for chromosome 1 compared to the whole genome (Figure 5).

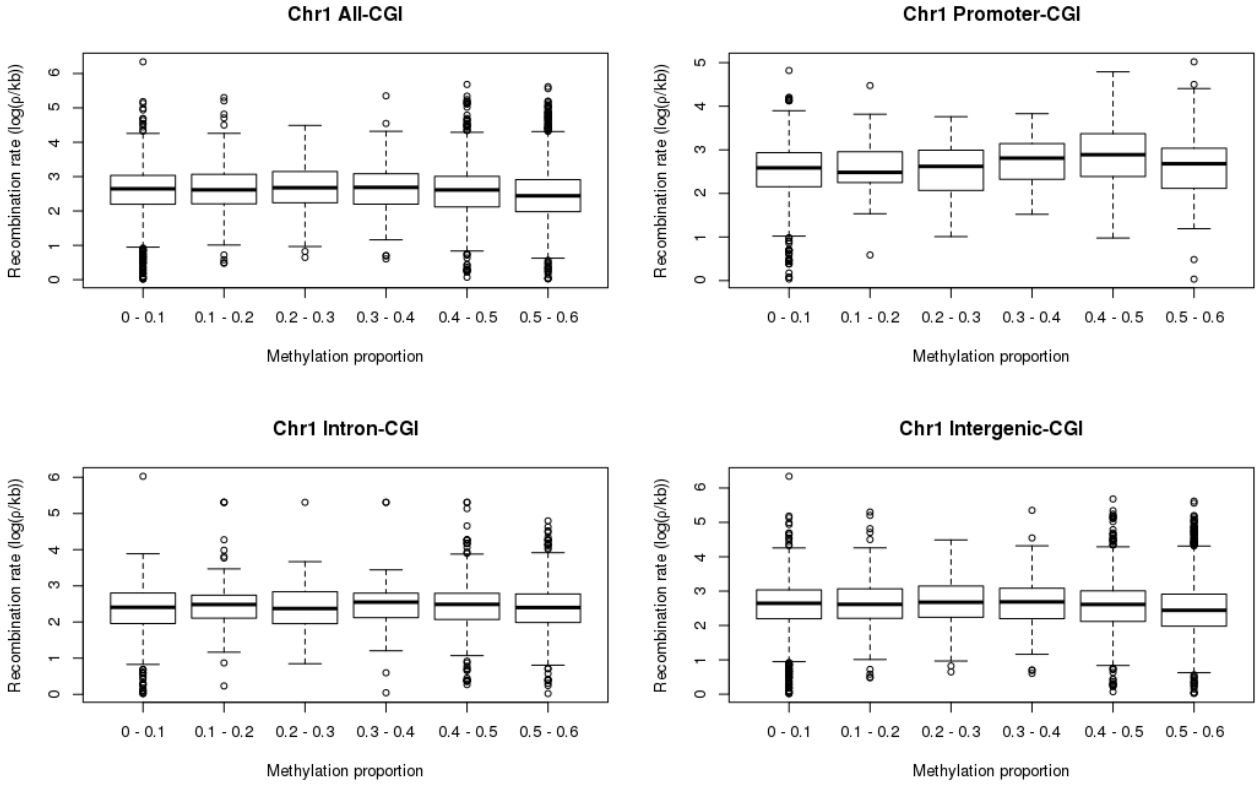

Figure 4: *Relationship between LD recombination rate,* log(ρ/kb)*, on the y-axis and methylation percentages divided into intervals of size 0.1 on the x-axis in four different genomic CGIs on chromosome 1. All-CGI refers to every CGI region irrespective of their location, Promoter-CGI featuring only CGIs in promoter regions, Intron-CGI featuring only intronic CGI regions and lastly Intergenic-CGI which features CGI in intergenic regions Within the boxes are the first and the third quartile and the thick line shows the median. The analysis was limited to methylation levels less than 0.6.*

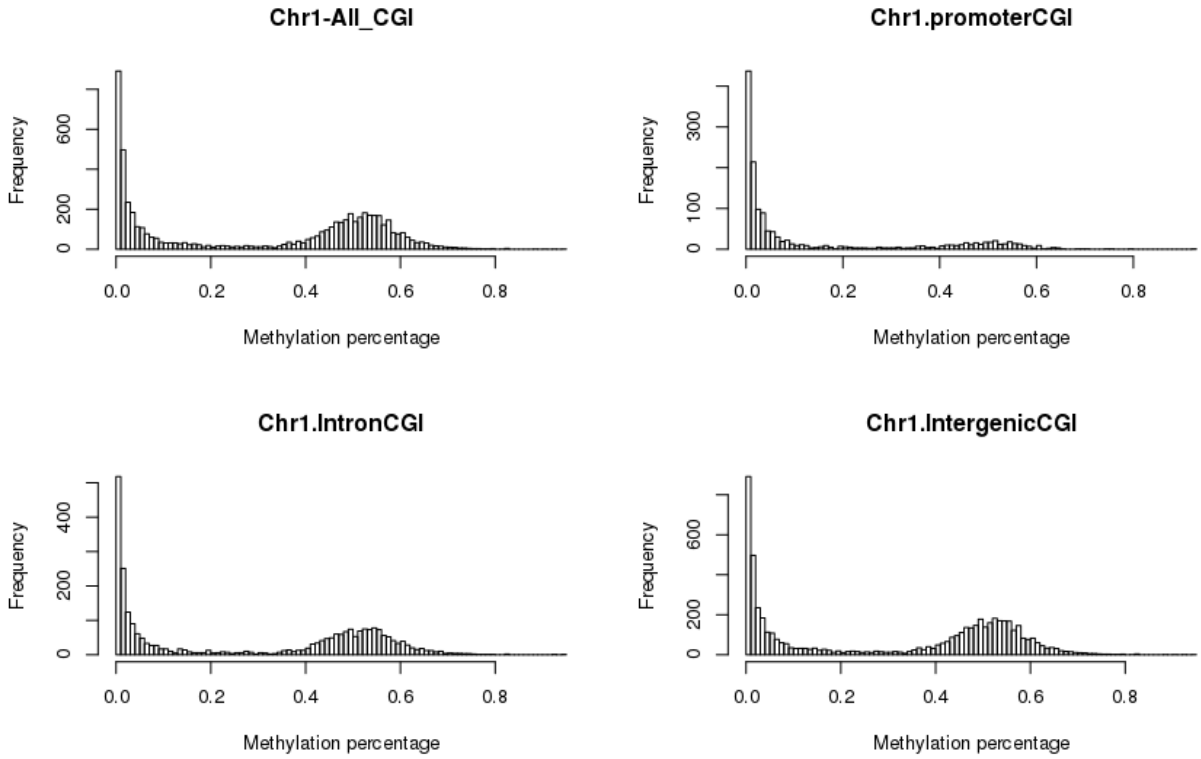

Figure 5: *Distribution of methylation levels for CGIs of four different regions on chromosome 1 with frequency of observed data points on the y-axis and methylation proportion on the x-axis. All-CGI refers to every CGI region irrespective of their location, Promoter-CGI featuring only CGIs in promoter regions, Intron-CGI featuring only intronic CGI regions and lastly Intergenic-CGI which features CGI in intergenic regions.*

This was also tested for methylation data from a testis sample to compare the relationship between recombination rate and the level of DNA methylation at CGIs. The methylation landscape of testis DNA was expected to more resemble somatic cells whereas sperm cells was expected to resemble meiotic cells. Thus, comparing the methylation landscape of these tissues would ascertain these assumptions. Surprisingly, a similar pattern to the sperm data analysis was observed in the testis data. All regions except promoters had a slight increase in recombination rate around methylation levels of 20-40 % (Figure 6). The frequency of methylation levels for the four regions were also like the sperm data. All regions had the highest frequency of locations with low methylation (<10%). All regions except for promoter regions also had an increase in frequency of locations with methylation levels around 50 % (Figure 7).

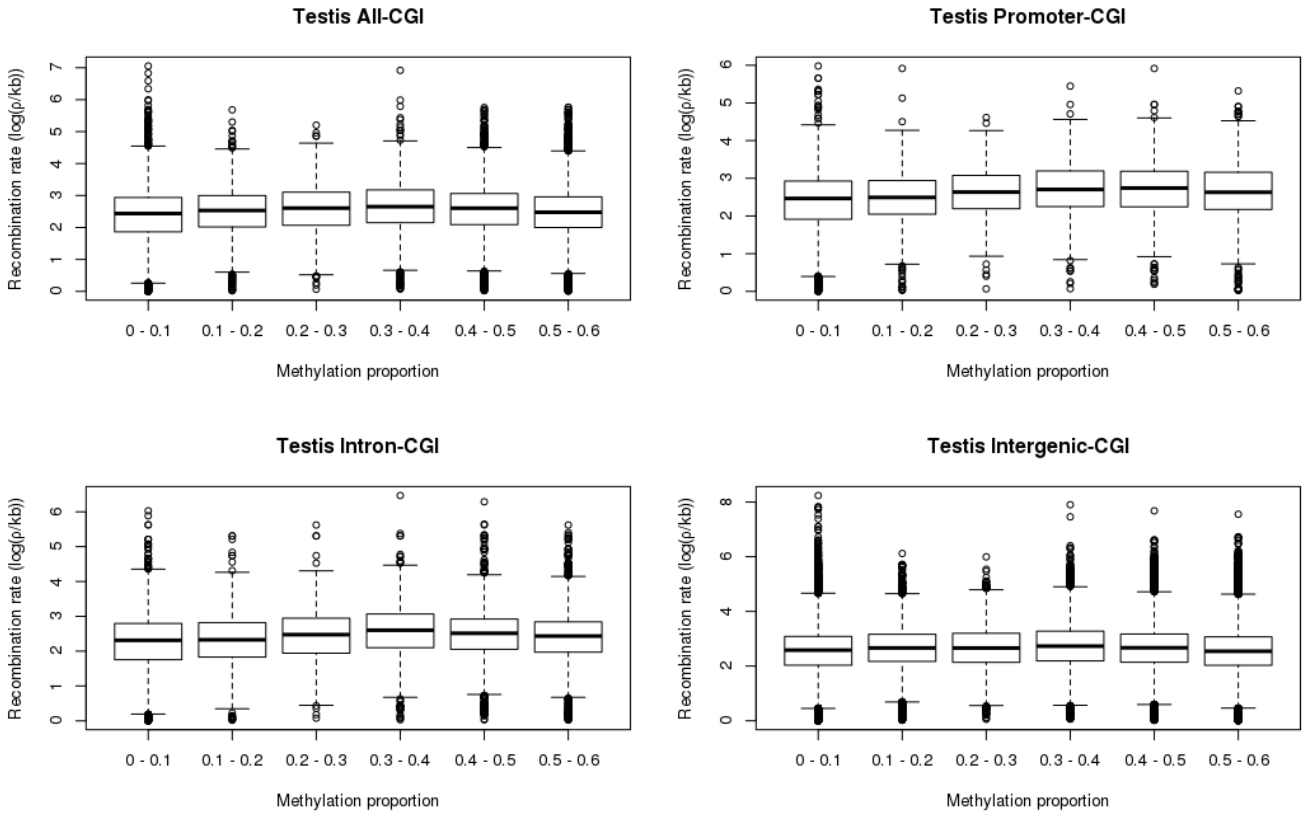

Figure 6: *Relationship between LD recombination rate, log(ρ/kb), on the y-axis and methylation percentages divided into intervals of size 0.1 on the x-axis in different genomic CGIs of DNA from testis. All-CGI refers to every CGI region irrespective of their location, Promoter-CGI featuring only CGIs in promoter regions, Intron-CGI featuring only intronic CGI regions and lastly Intergenic-CGI which features CGI in intergenic regions. Within the boxes are the first and the third quartile and the thick line indicates the median. The analysis was limited to methylation levels less than 0.6.*

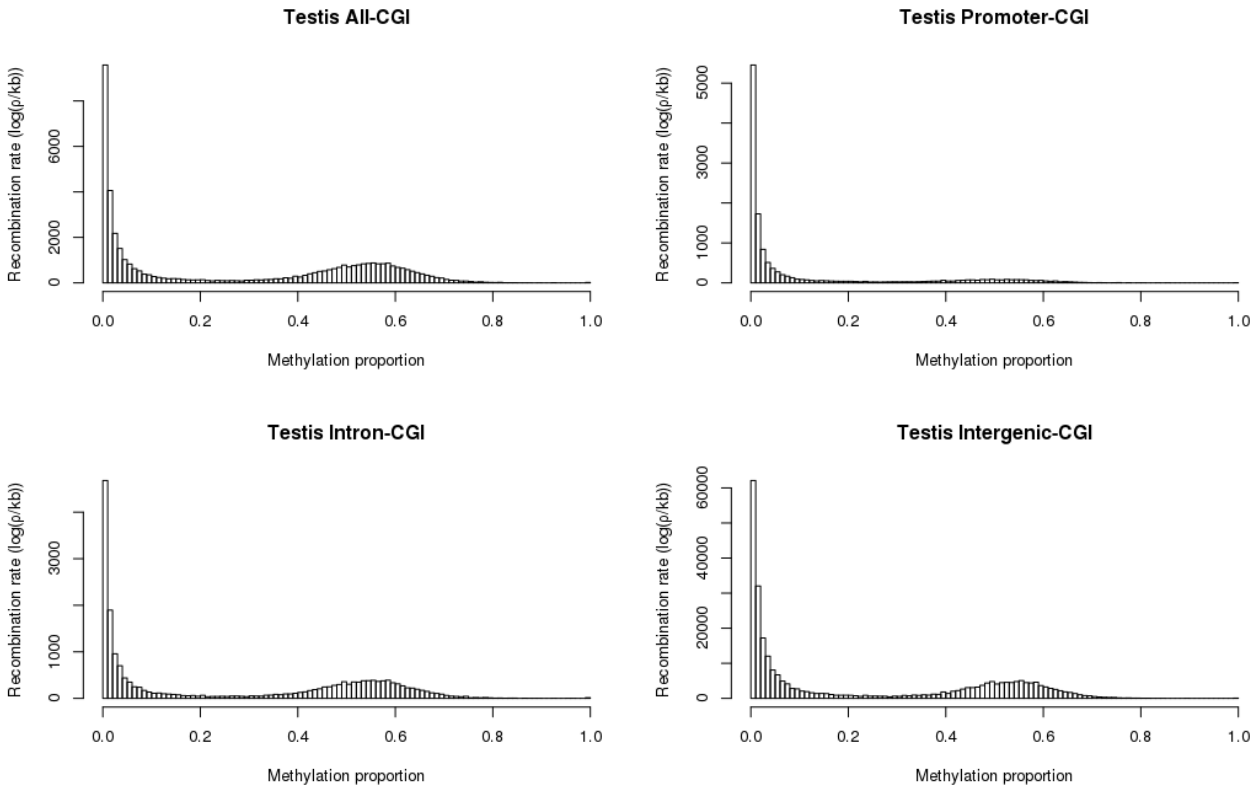

Figure 7: *Distribution of methylation levels for CGIs of different genomic regions of DNA from testis with frequency of observed data points on the y-axis and methylation proportion on the x-axis. All-CGI refers to every CGI region irrespective of their location, Promoter-CGI featuring only CGIs in promoter regions, Intron-CGI featuring only intronic CGI regions and lastly Intergenic-CGI which features CGI in intergenic regions.*

Discussion

By using whole-genome recombination and methylation estimates, I have identified a higher recombination rate in CGI promoters than in non-CGI promoters in chicken. This is consistent with the previous studies in both zebra finch (*Taeniopygia guttata)* & long-tailed finch (*Poephila acuticauda*) (Singhal *et al* 2015), and collared flycatcher (*Ficedula albicollis*) & pied flycatcher (*F. hypoleuca*) (Kawakami *et al.* 2017). In both previous studies, recombination hotspots were found to be associated with CGIs (Singhal *et al.* 2015, Kawakami *et al.* 2017). More precisely, Singhal *et al* 2015 found that recombination hotspots in finches are concentrated in functional elements like TESs, TES and CGI. Kawakami *et al.* 2017 found that the LD recombination rate in flycatchers was much higher in CGIs than for the genome average. They also found that CGI promoters have a higher recombination rate than non-CGI promoters. This suggests that the distribution of recombination is conserved

between chicken (Galliforms), flycatchers and finches (Passeriforms), which show about 90 MYA of divergence (Jarvis *et al.* 2014).

There are at least two explanations for the consistent patterns of the recombination landscape between these two distantly related avian clades. First, this pattern is derived by identity by descent (IBD), indicating that higher recombination rate in CGI promoter regions is the ancestral trait and has evolved before the split between Galliforms and Passeriforms. Second, it could also be due to a convergent evolution between chicken and passerines. This means that the same pattern has independently evolved at least twice in birds. If we look outside the avian phylogeny, many other eukaryotes, such as yeast, plants and insects (Niehuis *et al.* 2010, Yelina *et al.* 2015a, Lam & Keeney 2015) show an association between recombination hotspots and CGIs. These numerous observations make it highly unlikely that the same pattern has independently evolved this many time. Instead, it is much more likely that this has been passed on through evolution. This strongly suggests that the repeated observation of the association between recombination and CGIs especially in promoter regions in chicken and passerines, is due to IBD rather than through convergence. In mammals, however, recombination rate is regulated by the gene PRDM9 and is not in the same way connected to CGIs. In dogs that lack a functional PRDM9 gene the same pattern as in birds can be found (Berglund *et al.* 2014), suggesting that a CGI associated regulation mechanism is likely the ancestral trait.

One would expect that since methylation levels change the chromatin structure, unmethylated DNA should be more open and available for the recombination machinery to induce double strand breaks. In turn this would lead to a negative correlation between methylation levels and recombination rate. However, the data provided no support for a negative relationship between recombination rate and methylation level. Instead there was a slight increase in recombination rate between 20-40 % methylation for all regions. Promoter regions even showed a weak positive relationship between recombination rate and methylation level. A bimodal distribution of CpG methylation at CGIs was observed genome wide as well as in promoter regions, but to a lesser extent in the latter. Although it is not clear what causes this bimodal distribution, regions of moderate methylation level (20-40%) might be associated with some specific genomic features, including centromeres, telomeres and other highly repetitive regions. One such example is the chromosome ends, it has previously been shown that recombination rate gets higher at the ends of chromosomes (Kawakami *et al.* 2014). This shows that there are other genomic features than methylation level that affect the

recombination landscape. Inter-correlations between other genomic features and methylation might blur the association between recombination rate and methylation levels, which makes it harder to detect the association between them. Below I will discuss a few possible genomic features that affect the recombination landscape and could explain the pattern observed in my result.

It is possible that there are other features in the DNA that can cause this increase in recombination rate when methylation levels are around 20-40 %. One other possible reason could the numerous small chromosomes in the chicken genome. It has been previously shown that the recombination rate scales with the size of the chromosome in chicken (Consortium 2004, Groenen *et al.* 2009). This relationship could be masking the expected relationship between recombination rate and methylation levels. However, this explanation is highly unlikely because the same pattern was observed if the analysis was limited only to a single chromosome, i.e. chromosome 1 (Figure 4).

In plants, a similar pattern has been observed where recombination rate was low in regions with low CpG methylation (Yelina *et al.* 2015b). It has been suggested that not only CpG methylation, but also non-CG methylation is associated with recombination rate and the rate of DNA double strand breaks in plants. Two recent studies indicated that CpG methylation can inhibit DNA double strand breaks but that it is only non-CpG and/or H3K9me2 that inhibit the actual crossovers (Choi *et al.* 2018, Underwood *et al.* 2018). In vertebrates, however, non-CpG methylation occurs at very low frequency (Lister *et al.* 2009). Therefore, although I cannot completely rule out this possibility, the impact of non-CpG methylation on recombination appears very small. Another thing to test would be to compare the level of methylation between promoter CGIs associated with recombination hotspots and those not associated with hotspots. This would more directly test if methylation levels regulate recombination hotspots and not only recombination rate in general.

In this analysis, chicken sperm methylation data was used as a proxy for cells where the DNA methylation landscape is similar to meiotic cells. The optimal cell type to test the relationship between recombination and DNA methylation is apparently cells in the meiotic stage where recombination takes place. If the DNA methylation landscape changes between these two cell types, it may result in a weak or non-significant association between the methylation levels and recombination rate. In my study I also tested the same pattern for testis cells to evaluate the use of sperm cells as a proxy. I observed a very similar methylation landscape in both sperm cells and testis cells. If this observation is correct it could suggest that sperm cells may

not be as good of a proxy as earlier thought. However, this observation is not in line with earlier studies, work in the group has noted a decreased methylation in testis tissue compared to germ cell tissue. This pattern is also similar to those observed in human and mouse (He *et al.* 2018). This suggest that there is some other reason why the testis and sperm tissues had very similar patterns of methylation in my result. It could be because of some technical artefacts when quantifying the level of methylation in a given genomic feature. For example, I treated all CGIs equally without taking the number or density of CpG-sites into consideration. The length of CGIs ranged from 8 base pairs to 4900 base pairs with the varying number of CpG sites. The density of CpG sites could affect the methylation pattern observed because CG methylation can only occur at CpG sites. The number of CpG in CGIs sites changes the statistical power of the analysis since a lower number of CpG sites limits where methylation can occur. In future studies one could investigate this further and maybe separate CGIs based on their CpG-density to limit the number of covarying factors.

In conclusion, this study showed that higher recombination rate in CGI promoters compared to non-CGI promoters is an ancestral state and conserved across avian species. Concentration of recombination at CGIs, however, was not strongly associated with the level of methylation at CpG sites. To further elucidate the regulatory mechanisms of recombination hotspots in birds, it is important to use meiotic cells and DNase-seq. DNase-seq is a method used to identify genetic regulatory elements like promoters, silencers and enhancers (Boyle *et al.* 2008). In addition, it would be interesting to further expand this analysis to the most basal avian lineage (Palaeognathae) and non-avian reptiles.

Acknowledgements

I would first like to thank Takeshi Kawakami and Carina Farah Mugal for supervising this project, your help and support through this project has been amazing and I have learned so much from you.

I would like to thank Linnea Smeds at the Program for evolutionary biology at the department of ecology and genetics for helping me with the bioinformatic analyses.

Finally, I would like to thank both the opponent of the final presentation and the coordinators Anna-Kristina Brunberg & Anna Rosling.

References

Auton A, Fledel-Alon A, Pfeifer S, Venn O, Ségurel L, Street T, Leffler EM, Bowden R, Aneas I, Broxholme J, Humburg P, Iqbal Z, Lunter G, Maller J, Hernandez RD, Melton C, Venkat A, Nobrega MA, Bontrop R, Myers S, Donnelly P, Przeworski M, McVean G. 2012. A fine-scale chimpanzee genetic map from population sequencing. Science (New York, NY) 336: 193–198.

Axelsson E, Webster MT, Ratnakumar A, Consortium TL, Ponting CP, Lindblad-Toh K. 2012. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Research 22: 51–63.

Baker Z, Schumer M, Haba Y, Bashkirova L, Holland C, Rosenthal GG, Przeworski M. 2017. Repeated losses of PRDM9-directed recombination despite the conservation of PRDM9 across vertebrates. eLife 6: e24133.

Baudat F, Imai Y, Massy B de. 2013. Meiotic recombination in mammals: localization and regulation. Nature Reviews Genetics 14: 794–806.

Berchowitz LE, Hanlon SE, Lieb JD, Copenhaver GP. 2009. A positive but complex association between meiotic double-strand break hotspots and open chromatin in Saccharomyces cerevisiae. Genome Research 19: 2245–2257.

Berglund J, Quilez J, Arndt PF, Webster MT. 2014. Germline Methylation Patterns Determine the Distribution of Recombination Events in the Dog Genome. Genome Biology and Evolution 7: 522–530.

Bird AP. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Research 8: 1499–1504.

Bird AP, Taggart MH. 1980. Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Research 8: 1485–1497.

Bolívar P, Mugal CF, Nater A, Ellegren H. 2016. Recombination Rate Variation Modulates Gene Sequence Evolution Mainly via GC-Biased Gene Conversion, Not Hill-

Robertson Interference, in an Avian System. Molecular biology and evolution 33: 216–227.

Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. 2008. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell 132: 311–322.

Brachet E, Sommermeyer V, Borde V. 2012. Interplay between modifications of chromatin and meiotic recombination hotspots. Biology of the Cell 104: 51–69.

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. Nature 485: 642–645.

Burt DW. 2007. Emergence of the Chicken as a Model Organism: Implications for Agriculture and Biology. Poultry Science 86: 1460–1471.

Choi K, Zhao X, Kelly KA, Venn O, Higgins JD, Yelina NE, Hardcastle TJ, Ziolkowski PA, Copenhaver GP, Franklin FCH, McVean G, Henderson IR. 2013. *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nature Genetics 45: 1327–1336.

Choi K, Zhao X, Tock AJ, Lambing C, Underwood CJ, Hardcastle TJ, Serra H, Kim J, Cho HS, Kim J, Ziolkowski PA, Yelina NE, Hwang I, Martienssen RA, Henderson IR. 2018. Nucleosomes and DNA methylation shape meiotic DSB frequency in Arabidopsis thaliana transposons and gene regulatory regions. Genome Research 28: 532–546.

Consortium ICGS. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432: 695–716.

Cooper DN, Gerber-Huber S. 1985. DNA methylation and CpG suppression. Cell Differentiation 17: 199–205.

Cooper DN, Mort M, Stenson PD, Ball EV, Chuzhanova NA. 2010. Methylation-mediated deamination of 5-methylcytosine appears to give rise to mutations causing human inherited disease in CpNpG trinucleotides, as well as in CpG dinucleotides. Human Genomics 4: 406–410.

Davuluri RV, Grosse I, Zhang MQ. 2001. Computational identification of promoters and first exons in the human genome. Nature Genetics 29: 412–417.

Dooner HK, Martínez-Férez IM. 1997. Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. The Plant Cell 9: 1633–1646.

Duret L, Eyre-Walker A, Galtier N. 2006. A new perspective on isochore evolution. Gene 385: 71–74.

Dutta R, Saha-Mandal A, Cheng X, Qiu S, Serpen J, Fedorova L, Fedorov A. 2018. 1000 human genomes carry widespread signatures of GC biased gene conversion. BMC Genomics 19: 256.

Gerton JL, DeRisi J, Shroff R, Lichten M, Brown PO, Petes TD. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America 97: 11383–11390.

Groenen MAM, Wahlberg P, Foglio M, Cheng HH, Megens H-J, Crooijmans RPMA, Besnier F, Lathrop M, Muir WM, Wong GK-S, Gut I, Andersson L. 2009. A high-density SNP-based linkage map of the chicken genome reveals sequence features correlated with recombination rate. Genome Research 19: 510–519.

Han L, Su B, Li W-H, Zhao Z. 2008. CpG island density and its correlations with genomic features in mammalian genomes. Genome Biology 9: R79.

Hayashi K, Yoshida K, Matsui Y. 2005. A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. Nature 438: 374–378.

He Y, Zuo Q, Edwards J, Zhao K, Lei J, Cai W, Nie Q, Li B, Song J. 2018. DNA Methylation and Regulatory Elements during Chicken Germline Stem Cell Differentiation. Stem Cell Reports 10: 1793–1806.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, Fonseca RR da, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan X, Dixon A, Li S, Li N, Huang Y, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello CV, Lovell PV, Wirthlin M, Schneider MPC, Prosdocimi F, Samaniego JA, Velazquez AMV, Alfaro-Núñez A, Campos PF, Petersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong Z, Zeng Y, Liu S, Li Z, Liu B, Wu K, Xiao J, Yinqi X, Zheng Q, Zhang Y, Yang H, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jønsson KA, Johnson W, Koepfli K-P, O'Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E, Graves GR, Glenn TC, McCormack J, Burt D, Ellegren H, Alström P, Edwards SV, Stamatakis A, Mindell DP, Cracraft J, Braun EL, Warnow T, Jun W, Gilbert MTP, Zhang G. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346: 1320–1331.

Kawakami T, Mugal CF, Suh A, Nater A, Burri R, Smeds L, Ellegren H. 2017. Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. Molecular Ecology 26: 4158–4172.

Kawakami T, Smeds L, Backström N, Husby A, Qvarnström A, Mugal CF, Olason P, Ellegren H. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Molecular Ecology 23: 4035–4058.

Lam I, Keeney S. 2015. Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. Science 350: 932–937.

Li D, Che T, Chen B, Tian S, Zhou X, Zhang G, Li M, Gaur U, Li Y, Luo M, Zhang L, Xu Z, Zhao X, Yin H, Wang Y, Jin L, Tang Q, Xu H, Yang M, Zhou R, Li R, Zhu Q, Li M. 2017. Genomic data for 78 chickens from 14 populations. GigaScience 6: 1–5.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462: 315–322.

Marsolier-Kergoat M-C, Yeramian E. 2009. GC Content and Recombination: Reassessing the Causal Effects for the Saccharomyces cerevisiae Genome. Genetics 183: 31–38.

Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, Sloan CA, Rosenbloom KR, Roe G, Rhead B, Raney BJ, Pohl A, Malladi VS, Li CH, Lee BT, Learned K, Kirkup V, Hsu F, Heitner S, Harte RA, Haeussler M, Guruvadoo L, Goldman M, Giardine BM, Fujita PA, Dreszer TR, Diekhans M, Cline MS, Clawson H, Barber GP, Haussler D, Kent WJ. 2013. The UCSC Genome Browser database: extensions and updates 2013. Nucleic Acids Research 41: D64–D69.

Mugal CF, Arndt PF, Holm L, Ellegren H. 2015. Evolutionary Consequences of DNA Methylation on the GC Content in Vertebrate Genomes. G3: Genes|Genomes|Genetics 5: 441–447.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. Science 327: 876–879.

Niehuis O, Gibson JD, Rosenberg MS, Pannebakker BA, Koevoets T, Judson AK, Desjardins CA, Kennedy K, Duggan D, Beukeboom LW, Zande L van de, Shuker DM, Werren JH, Gadau J. 2010. Recombination and its impact on the genome of the haplodiploid parasitoid wasp Nasonia. PLoS ONE 5: e8597.

Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS Genetics 5: e1000753.

Paigen K, Petkov P. 2010. Mammalian recombination hot spots: properties, control and evolution. Nature Reviews Genetics 11: 221–233.

Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 Controls Activation of Mammalian Recombination Hotspots. Science 327: 835–835.

Ponting CP. 2011. What are the genomic drivers of the rapid evolution of PRDM9? Trends in genetics: TIG 27: 165–171.

Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526: 569–573.

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

Saxonov S, Berg P, Brutlag DL. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America 103: 1412–1417.

Shilo S, Melamed-Bessudo C, Dorone Y, Barkai N, Levy AA. 2015. DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in Arabidopsis. The Plant Cell 27: 2427–2436.

Singhal S, Leffler EM, Sannareddy K, Turner I, Venn O, Hooper DM, Strand AI, Li Q, Raney B, Balakrishnan CN, Griffith SC, McVean G, Przeworski M. 2015. Stable recombination hotspots in birds. Science 350: 928–932.

Stern CD. 2004. The chick embryo – past, present and future as a model system in developmental biology. Mechanisms of Development 121: 1011–1013.

Sved J, Bird A. 1990. The expected equilibrium of the CpG dinucleotide in vertebrate genomes under a mutation model. Proceedings of the National Academy of Sciences of the United States of America 87: 4692–4696.

Underwood CJ, Choi K, Lambing C, Zhao X, Serra H, Borges F, Simorowski J, Ernst E, Jacob Y, Henderson IR, Martienssen RA. 2018. Epigenetic activation of meiotic recombination near Arabidopsis thaliana centromeres via loss of H3K9me2 and non-CG DNA methylation. Genome Research 28: 519–531.

Venkatesh S, Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. Nature Reviews Molecular Cell Biology 16: 178–189.

Webster MT, Hurst LD. 2012. Direct and indirect consequences of meiotic recombination: implications for genome evolution. Trends in Genetics 28: 101–109.

Yelina N, Diaz P, Lambing C, Henderson IR. 2015. Epigenetic control of meiotic recombination in plants. Science China Life Sciences 58: 223–231.

Yelina NE, Lambing C, Hardcastle TJ, Zhao X, Santos B, Henderson IR. 2015. DNA methylation epigenetically silences crossover hot spots and controls chromosomal domains of meiotic recombination in Arabidopsis. Genes & Development 29: 2183–2202.