Title

The role of LINEs and CpG islands in dosage compensation on the chicken Z chromosome

Journal

Chromosome Research: Cytogenetics, genomics, chromatin and the nucleus, 17(6)

ISSN

1573-6849

Authors

Melamed, Esther
Arnold, Arthur P.

Publication Date

2009-08-01

DOI

10.1007/s10577-009-9068-4

Peer reviewed

Chromosome Research (2009) 17:727–736
DOI 10.1007/s10577-009-9068-4

The role of LINEs and CpG islands in dosage compensation on the chicken Z chromosome

Esther Melamed · Arthur P. Arnold

Received: 17 June 2009 / Accepted: 2 July 2009 / Published online: 12 August 2009

Abstract Most avian Z genes are expressed more highly in ZZ males than ZW females, suggesting that chromosome-wide mechanisms of dosage compensation have not evolved. Nevertheless, a small percentage of Z genes are expressed at similar levels in males and females, an indication that a yet unidentified mechanism compensates for the sex difference in copy number. Primary DNA sequences are thought to have a role in determining chromosome gene inactivation status on the mammalian X chromosome. However, it is currently unknown whether primary DNA sequences also mediate chicken Z gene compensation status. Using a combination of chicken DNA sequences and Z gene compensation profiles of 310 genes, we explored the relationship between Z gene compensation status and primary DNA sequence features. Statistical analysis of different Z chromosomal features revealed that long interspersed nuclear elements (LINEs) and CpG islands are enriched on the Z chromosome compared with 329 other DNA features. Linear support vector machine (SVM) classifiers, using primary DNA sequences, correctly predict the Z compensation status for >60% of all Z-linked genes. CpG islands appear to be the most accurate classifier and alone can correctly predict compensation of 63% of Z genes. We also show that LINE CR1 elements are enriched 2.7-fold on the chicken Z chromosome compared with autosomes and that chicken chromosomal length is highly correlated with percentage LINE content. However, the position of LINE elements is not significantly associated with dosage compensation status of Z genes. We also find a trend for a higher proportion of CpG islands in the region of the Z chromosome with the fewest dosage-compensated genes compared with the region containing the greatest concentration of compensated genes. Comparison between chicken and platypus genomes shows that LINE elements are not enriched on sex chromosomes in platypus, indicating that LINE accumulation is not a feature of all sex chromosomes. Our results suggest that CpG islands are not randomly distributed on the Z chromosome and may influence Z gene dosage compensation status.

Keywords dosage compensation · Z chromosome · DNA sequence · LINEs · CpG · chicken · sex chromosome · X chromosome

Responsible Editor: Mary Erin Delany.

Electronic supplementary material The online version of this article (doi:10.1007/s10577-009-9068-4) contains supplementary material, which is available to authorized users.

E. Melamed · A. P. Arnold (✉)
Department of Physiological Science and Laboratory of Neuroendocrinology of the Brain Research Institute, University of California,
621 Charles Young Drive South,
Los Angeles, CA 90095-1606, USA
e-mail: arnold@ucla.edu

E. Melamed
e-mail: emelamed@ucla.edu

Abbreviations

LINE	long interspersed nuclear element
MHM	male hypermethylated
SVM	support vector machine (SVM)
XIC	X inactivation centre
XIST	X-inactive specific transcript

Introduction

On the basis of previous studies of XX female / XY male systems, chromosome-wide gene dosage compensation of the heterogametic sex chromosome was thought to be critical (Meyer et al. 2004). However, recent studies in birds and in moths have definitively shown that most genes on a large chromosome (Z) can remain uncompensated (Itoh et al. 2007; Ellegren et al. 2007; Zha et al. 2009). Further study of ZZ/ZW sex chromosome systems can help to uncover important adaptations that reduce the disadvantages of constitutively higher expression of Z genes in one sex.

Bird sex chromosomes, Z and W, bear many similarities to the mammalian X and Y chromosomes. Like the mammalian X, the Z chromosome is large and gene rich, whereas the W chromosome, like the mammalian Y chromosome, is small and gene poor. Unlike the sex chromosome constitution in mammals, where the homogametic female is XX and the heterogametic male is XY, in birds homogamety is found in ZZ males whereas females are heterogametic with ZW sex chromosomes. Monotreme X and Y chromosomes are thought to have evolved from the common ancestor of birds and mammals via addition of genes to the ancestral Z and W (Grutzner et al. 2004; Rens et al. 2007). However, unlike birds and eutherian mammals, the monotreme platypus has 5X and 5Y sex chromosomes and the X and Y appear to be intermediate between bird Z/W and eutherian mammalian X/Y in gene content and in level of dosage compensation (Deakin et al. 2008).

Mammalian X chromosome inactivation results in dosage compensation between males and females via silencing of one X chromosome in each female cell (Lyon 1961; Gupta et al. 2006; Nguyen and Disteche 2006). X inactivation occurs in the epiblast early in embryonic development (Nguyen and Disteche 2006), but the inactive X chromosome is re-activated in the female germline during meiosis to allow for recombination between the two X chromosomes (Ross et al. 2005). X inactivation is initiated with the transcription of the XIST non-coding RNA from the X inactivation center. XIST RNA coats the X chromosome from which it was transcribed and recruits repressive Polycomb complexes PRC1 and PRC2 to the inactive X that aid in heterochromatin formation and maintenance of gene silencing (Chadwick and Willard 2003; Plath et al. 2003; Silva et al. 2003). LINE elements are thought to help in the spread of X inactivation along the entire chromosome by acting as 'way stations' or 'booster elements' (Lyon 1998). Long-term maintenance of gene silencing also requires histone modifications and CpG island methylation on the inactive X (Mohandas et al. 1981; Pfeifer et al. 1990).

In contrast to mammals, dosage compensation of Z genes is ineffective in birds, and average Z gene mRNA is at least 30–40% more highly expressed in males than in females (Itoh et al. 2007; Ellegren et al. 2007; Arnold et al. 2008). Although there is no Xist in birds, dosage compensation may be mediated by an Xist-like non-coding RNA transcribed from the MHM (male hypermethylated) locus on the Z chromosome. The MHM RNA is expressed only in females, associates with the Z chromosome near the site of transcription, and is correlated with female-specific hyperacetylation of lysine 16 on histone 4 near the MHM locus (Teranishi et al. 2001; Bisoni et al. 2005). Importantly, the MHM region shows greater concentration of dosage-compensated genes on the Z chromosome (Melamed and Arnold 2007; Melamed et al. 2009). The greater compensation of dosage could be explained by female-specific upregulation of genes and/or by downregulation of genes in males (Teranishi et al. 2001; Bisoni et al. 2005; Ellegren et al. 2007; Melamed and Arnold 2007).

Dosage compensation in monotremes has not been assessed for all X chromosome genes. However, among 19 X genes on the platypus' X chromosomes, some genes appear not to be compensated, whereas others are dosage compensated by stochastic transcriptional inhibition with variable expression patterns between genes (Deakin et al. 2008). Paternal inactivation or imprinted partial expression do not appear to be likely mechanisms of dosage compensation in platypus (Deakin et al. 2008). Xist has not been found in marsupial or monotreme genomes and the region homologous to the XIC in mammals is disrupted (Hore et al. 2007; Shevchenko et al. 2007).

Although most genes on the X chromosomes are well compensated, about 15% of human X genes escape inactivation (Carrel and Willard 2005). Previous studies in humans explored how to predict X gene inactivation status. Wang et al, showed that primary DNA features together with support vector machines (SVM) can predict gene compensation status with high accuracy (Wang et al. 2006). Here we sought to evaluate whether primary DNA sequences on the chicken Z chromosome could also be used to predict gene compensation status. To answer this question, we evaluated all primary sequences on the Z chromosome using linear SVM and statistical analysis. We found that CpG islands and CR1-LINE elements are significant primary DNA features on the chicken Z chromosome. CR1-LINE elements are enriched on the Z chromosome compared to autosomes in chicken, but do not predict Z gene compensation status, are not differentially associated with compensated or uncompensated genes, and are not enriched in specific regions of the Z chromosome. In contrast, CpG islands are significantly enriched around uncompensated genes and have the highest predictive power of genes' compensation status compared with other repetitive features. We further show that LINEs are not enriched on the platypus X chromosome compared with the LINE enrichment on the human X and chicken Z chromosomes, suggesting that LINE accumulation is not a general feature of all sex chromosomes.

Materials and methods

Sample collection and microarray data analysis

Sample collection and microarray data analysis have been described previously (Itoh et al. 2007; Melamed and Arnold 2007).

Statistical analysis

Statistical analysis of the feature sequence difference between compensated and uncompensated genes as well as the SVM analyses were performed using R (R Development Core Team 2006). The libSVM package was used for data classification and gene prediction using linear SVM (Chang and Lin 2001). R's base statistics package was used to perform Fisher's exact tests on the CpG island and LINE-CR1 features.

For the SVM analysis in chicken, a total of 310 compensated (lowest 1/3 quantile with M:F ratios <1.21) and uncompensated (upper 1/3 quantile with M:F ratios >1.54) Z brain-expressed genes were considered. Thus, 'dosage compensation' is used here to refer to any process that reduces the high M:F ratios assumed to result from the double genomic dose of Z genes in males relative to females. Z gene compensation status was determined by calculating male to female expression ratios (M:F ratios) for all Z genes in brain. The SVM was trained to associate the set of compensated and uncompensated genes with feature scores. The feature list consisted of 329 unique repetitive elements on the Z chromosome obtained from the UCSC genome browser RepeatMasker track (see Supplementary Table 1). The training set consisted of 280 genes. The leave-one-out cross validation sample was performed either with a set of 30 genes chosen randomly on the Z chromosome or with a segment of 30 contiguous Z genes in separate analyses. SVM was trained on 280 genes, and then tested to determine whether mathematical association of features with compensation status could accurately predict the compensation status on the 30 genes that were left out of the training sample. The procedure was repeated 100 times for different sets of 30 genes in the leave-out sample, each time scoring the number of correct and incorrect predictions of the genes' compensation status.

For both Fisher's exact test and SVM analyses, we considered a total of 23 types of CR1 elements present on the Z chromosome based on information obtained from UCSC genome browser (Supplementary Table 1) and CpG islands. Gene transcription is known to be influenced by both the promoter region sequence and the larger genomic sequences upstream of genes (Bailey et al. 2000; Ke and Collins 2003; Wang et al. 2006). We therefore chose distances upstream of genes to interrogate genomic windows within gene-promoter regions near gene start points as well as genomic windows with other regulatory units in between genes. Average intergene distance on the Z chromosome was calculated to be 134.9 kb. Therefore, a repeat was considered to be associated with a gene if any part of it was within the search width of 2 kb, 10 kb, 50 kb, or 100 kb upstream of the gene's start.

For Fisher's exact test, Z genes were classified into compensated and uncompensated as described above. We counted the number of CpG or LINE elements

within 2 kb, 10 kb, 50 kb, or 100 kb genomic windows upstream of genes and established mean numbers of CpG or LINE elements for each genomic window. Genes with less than the mean number of CpG or LINE elements were called low-CpG or low-LINE, and genes with greater than the mean number of elements were called high-CpG or high-LINE.

Genes and repetitive elements sampling

LINE and CpG island information for the chicken (May 2007 freeze, build May 2006), human (May 2007 freeze, build March 2006), and platypus genomes (February 2009 freeze, build March 2007) were extracted from the UCSC genome annotation database using the Repeats and Variation track and the CpG island track (http://genome.ucsc.edu/) (Supplementary Table 1). Repeat information is based on the RepeatMasker annotation (repeat name, repeat Class, and repeat Family, and start and end positions on the chromosome), which uses a modified Smith–Waterman algorithm to align sequences against the Repbase library (http://www.girinst.org). The CpG island track is based on the definition of DNA sequence >200 bp in length, GC content >0.5, CpG_{obs}/CpG_{exp} ratio (observed to expected ratio based on GC content) ≥0.6 (Gardiner-Garden and Frommer 1987).

Information about LINE element positions and length of chromosomes was obtained from the UCSC genome browser RepeatMasker track. We calculated what percentage of each chromosome consisted of LINE-CR1 elements relative to the length of the chromosome. For these analyses, chromosomes of

Overall Correct Prediction of Compensated and Uncompensated Genes

	% Correct Prediction
329 features	62.4 `
CpG Island	63
LINE-CR1	52.6
LINE-CR1 and CpG	61.6

All features

		Predicted	
		Compensated	Uncompensated
Observed	Compensated	0.656	0.344
	Uncompensated	0.408	0.592

CpG islands

		Predicted	
		Compensated	Uncompensated
Observed	Compensated	0.716	0.284
	Uncompensated	0.455	0.545

LINE-CR1

		Predicted	
		Compensated	Uncompensated
Observed	Compensated	0.855	0.145
	Uncompensated	0.803	0.197

LINE-CR1 and CpG islands

		Predicted	
		Compensated	Uncompensated
Observed	Compensated	0.672	0.328
	Uncompensated	0.440	0.560

Fig. 1 SVM results. SVM was trained to associated genes' compensation status in a training set of genes and then tested to predict the compensation status of a leave-out sample of genes. The tables show the percentages of correctly and incorrectly predicted compensation status of genes based on the actual observed gene compensation status for different features. An overall percentage of correctly predicted compensated and uncompensated genes is included, calculated by averaging the correctly predicted compensated and uncompensated genes for the feature. The search width for these results was 2 kb

length less than 100 kB (i.e., in chicken, chromosomes 16, 32, and W) were excluded.

Results

Machine learning classifiers such as support vector machine algorithms (SVM) have previously been successfully applied to predict X inactivation status of mammalian genes using different repetitive element features around X chromosome genes (Wang et al. 2006). We asked whether the 329 repetitive features on the Z chromosome are predictive of the compensation status of Z genes using SVM. As described in Methods, the SVM was trained to associate repetitive features and compensation status of 310 compensated and uncompensated Z genes. When all 329 repetitive features were used together, SVM correctly predicted 62.4% of genes. CpG island feature was the most predictive feature of compensation status and alone correctly predicted 63% of genes. LINEs and CpG islands together gave a similar predictive power (61.6%) as CpG islands alone (Fig. 1). Using LINE elements alone decreased SVM's predictive power to 52.6%.

It has previously been shown that the mammalian X chromosome is enriched 2-fold in LINE elements compared with chromosomes 6, 7, 20, 21, and 22 (Bailey et al. 2000). Similarly, our analysis of LINE element composition on the human X compared with all 22 autosomes confirmed an X-specific enrichment of 1.7-fold (32.3% for X vs. 18.2% for 22 autosomes) (Fig. 2). Chromosomal size and LINE element content were significantly correlated in the human genome (r=0.645, p<0.0007). The chicken Z chromosome contained 2.76-fold more LINE elements (9.77%) than the average autosome (3.55%) (Table 1). In addition, chromosome size was highly correlated with percent LINE element composition (r=0.815, p<0.0000001) (Fig. 2). The higher percentage of LINE elements on the Z chromosome, however, is not just a function of size, so that the Z vs. autosome difference holds even when comparing the Z chromosome to autosomes of comparable size (Fig. 2). In both the chicken and the human genomes, the Z and X chromosomes contained the highest LINE content of any chromosome (Fig. 2, Table 1).

We found that in platypus, the four X chromosomes contain the average percentage of LINE elements (18.73%) compared with autosomes (18.55%) (Table 1). Interestingly, the percentage of LINE elements on the platypus X chromosomes was between the LINE percentage compositions of chicken and human.

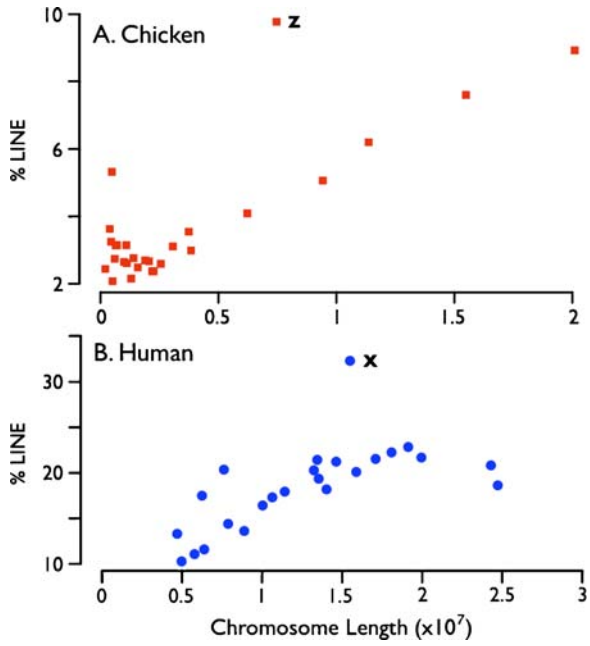

Fig. 2 Percentage LINE element composition on chicken and human chromosomes. In both chicken **(a)** and human **(b)**, the percentage LINE elements increases monotonically with chromosome length, but in each case the larger sex chromosome (X in human, Z in chicken) has a disproportionately high percentage of LINE elements

Given that CpG islands and LINE elements were the most predictive features for dosage compensation, we next evaluated how the density of CpG islands or LINE elements around Z genes related to their compensation status. Using Fisher's exact test, we observed that in all four genomic windows the low-CpG-score group contained more compensated genes, and the high-CpG-score group contained more uncompensated genes. These differences were significant for the 2 kb (p<0.00003), 10 kb (p<0.0006), and 50 kb (p<0.012) windows, but not in the 100 kb window (p=0.12) (Fig. 3, Table 2). LINE elements considered as a combined single feature or as separate classes of elements (23 different classes) did not yield a significant association with gene compensation status in any of the considered windows after Bonferroni correction.

We have previously shown that Zp is enriched in compensated genes in the MHM region (MHM

Table 1 Percentage of LINE elements per length of chromosome in the chicken, human, and platypus genomes. Compared with the chicken Z and human X chromosomes, both of which have a higher percentage LINE composition, the platypus X chromosome has a similar percentage of LINES compared with autosomes

Human			Chicken			Platypus		
Name	Length	Percent	Name	Length	Percent	Name	Length	Percent
chr22	49691432	10.3	chr25	2031799	2.44	chr17	1399469	10.56
chrY	57772954	11.1	chr22	3936574	3.63	chr20	1816412	21.44
chr19	63811651	11.6	chr28	4512026	3.25	chr14	2696122	18.75
chr21	46944323	13.3	chr27	4841970	5.32	chr15	3786880	18.47
chr16	88827254	13.6	chr26	5102438	2.07	**chrX2**	**5652501**	**17.87**
chr17	78774742	14.4	chr23	6042217	2.74	**chrX3**	**5951358**	**14.71**
chr15	100338915	16.4	chr24	6400109	3.14	chr18	6611290	16.19
chr14	106368585	17.3	chr21	6959642	3.14	chr11	6809224	17.41
chr20	62435964	17.5	chr19	9939723	2.64	chr10	11243762	15.72
chr13	114142980	17.9	chr18	10925261	3.14	chr12	15872666	18.19
chr9	140273252	18.2	chr17	11182526	2.61	chr6	16302927	17.87
chr1	247249719	18.6	chr15	12968165	2.15	chr5	24609220	20.16
chr10	135374737	19.4	chr20	13986235	2.76	**chrX5**	**27786739**	**21.99**
chr7	158821424	20.1	chr14	15819469	2.48	chr7	40039088	18.40
chr12	132349534	20.3	chr13	18911934	2.69	**chrX1**	**45541551**	**20.35**
chr18	76117153	20.4	chr12	20536687	2.67	chr1	47594283	22.17
chr2	242951149	20.8	chr11	21928095	2.37	chr2	54797317	21.78
chr8	146274826	21.2	chr10	22556432	2.37	chr4	58987262	20.07
chr11	134452384	21.4	chr9	25554352	2.59	chr3	59581953	21.15
chr6	170899992	21.5	chr8	30671729	3.11			
chr3	199501827	21.7	chr6	37400442	3.54			
chr5	180857866	22.3	chr7	38384769	2.99			
chr4	191273063	22.8	chr5	62238931	4.09			
chrX	**154913754**	**32.3**	**chrZ**	**74602320**	**9.77**			
			chr4	94230402	5.06			
			chr3	113657789	6.2			
			chr2	154873767	7.6			
			chr1	200994015	8.92			

valley, 2.5–3.5×10^7 bp) whereas the distal part of Zq (Zq peak, 5.5–7.5×10^7 bp) is enriched in non-compensated genes (Melamed and Arnold 2007, 2009). We next asked whether there were any regional differences in CpG islands or LINE concentration in the MHM valley or in the Zq peak. We computed the number of CpG islands relative to the size of the region in basepairs (CpG proportion) inside the region (Zq peak or MHM valley) and outside of the region. A CpG island was considered to be inside or outside of the region if its midpoint was inside or outside of the region. We found that the CpG proportion inside the Zq peak was 1.13 times higher than on the rest of the Z chromosome. The CpG proportion on the entire Z chromosome was 1.22 times higher than inside of the MHM valley (Table 3). Using a similar method for LINE elements, we did not find a difference in proportion of bases inside or outside of the MHM valley and the Zq peak compared to the rest of the Z chromosome (Table 3).

Discussion

Here, we evaluated the contribution of primary DNA sequences to chicken Z gene dosage compensation. We find that CpG islands and LINE elements are the most predictive features of 329 repetitive Z chromosome features. CpG islands are significantly

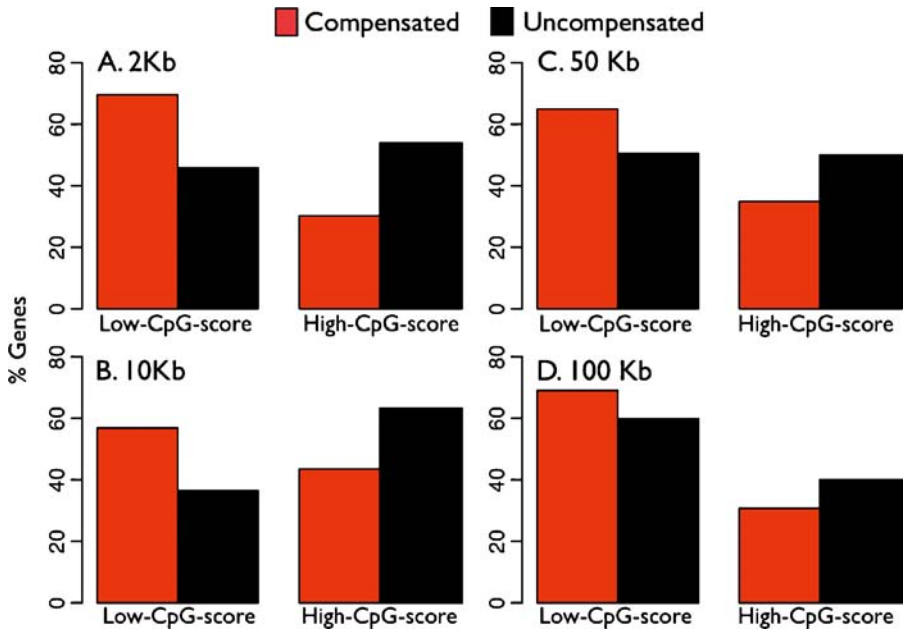

Fig. 3 CpG islands are differentially associated with compensated and uncompensated genes. **a** 2 kb ($p=3.19\times10^{-5}$), **b** 10 kb ($p=6\times10^{-4}$), **c** 50 kb ($p=0.012$), and **d** 100 kb ($p>0.05$) genomic windows upstream of gene start. Fisher's exact test)

enriched upstream of uncompensated Z genes and within the Zq uncompensated region. LINE elements are enriched on both the chicken Z and human X compared with autosomes but not on the platypus X chromosomes. The position of LINE elements does not appear to be significantly associated with dosage compensation status of Z genes. Thus, these analyses implicate LINE elements as important features of the Z chromosome and suggest that CpG islands are not randomly distributed on the Z chromosome with regard to dosage compensation. We also find that LINE elements are not enriched on the platypus X chromosomes, suggesting that LINE accumulation is not a common feature of all sex chromosomes.

In mammals, LINE elements have been estimated to occupy as much as 40% of the genome (Waters et al. 2007). Unlike the pattern in mammals, repetitive elements make up just 10% of the bird genome, with most elements belonging to the CR1 family of LINEs (Hughes and Piontkivska 2005). In addition, unlike mammalian LINEs, which belong to a single lineage (Smit et al. 1995; Furano 2000), chicken CR1 elements come from several distinct divergent lineages (International Chicken Genome Sequencing Consortium

Table 2 CpG islands are differentially associated with compensated and uncompensated genes. Fisher's exact test results for 2 kb, 10 kb, 50 kb, and 100 kb genomic windows upstream of gene start. In all four genomic windows, the low-CpG-score group contains a disproportionate number of compensated genes, and the high-CpG-score group contains a disproportionate number of uncompensated genes, reaching significance in the 2 kb, 10 kb, and 50 kb

	Low CpG score	High CpG score	*p* value
	2 kb window		
Number of compensated genes	108	47	3.19E-05
Number of uncompensated genes	71	84	
	10 kb window		
Number of compensated genes	88	67	6.0E-4
Number of uncompensated genes	57	98	
	50 kb window		
Number of compensated genes	101	54	0.012
Number of uncompensated genes	78	77	
	100 kb window		
Number of compensated genes	107	48	0.1223
Number of uncompensated genes	93	62	

Table 3 CpG island and LINE-CR1 element proportions on different regions of the Z chromosome. Number indicates the proportion of DNA sequence occupied by CpG islands and LINE-CR1 elements calculated for the entire Z chromosome or inside the Zq peak or MHM valley

	CpG	LINE
Entire Z chromosome	1.07×10^{-5}	2.34×10^{-4}
Zq peak	1.21×10^{-5}	2.31×10^{-4}
MHM valley	8.80×10^{-6}	2.32×10^{-4}

2004). Mammalian LINEs are mostly found in AT-rich genomic regions (Soriano et al. 1983; Lander et al. 2001; Pavlicek et al. 2001; Hackenberg et al. 2005) while chicken CR1s are located in both AT- and GC-rich regions of the genome (Abrusan et al. 2008). It is unclear at present whether any of the chicken CR1 elements are active, since most are truncated at the 5′ end (Silva and Burch 1989; Stumph et al. 1981; International Chicken Genome Sequencing Consortium 2004; Wicker et al. 2005) although at least one CR1 appears to have an open reading frame on chromosome 6 (International Chicken Genome Sequencing Consortium 2004).

There are several explanations for the overall lower percentage of repetitive elements in the bird genome compared with mammals. One reason may be that chicken CR1 is more specific than the mammalian CR1, resulting in fewer pseudogenes and a smaller genome in chickens (International Chicken Genome Sequencing Consortium 2004; Shedlock et al. 2007). Another explanation may be that LINEs are preferentially removed from the chicken genome via ectopic exchange between repeats with subsequent accumulation in AT- and GC-rich regions (Wichman et al. 1992; Abrusan and Krambeck 2006; Abrusan et al. 2008). Still another speculation is that the lower number of repetitive elements in birds' genomes is related to the need for smaller cell size for the requirements of flight and efficient gas exchange (Szarski 1983; Wachtel and Tiersch 1994).

Other factors, however, are probably relevant to the observation that the Z chromosome is enriched in LINE elements compared with autosomes in a manner similar to the enrichment on X chromosomes of mammals. One question is whether LINE elements are important in dosage compensation on sex chromosomes. In mammals, LINEs are considered to be 'way stations' for the propagation of the dosage compensation signal (Lyon 1998, 2000). However, the association of LINEs with inactivated vs. escapee genes is not strong. Only some mammalian studies have reported an association of LINE-L1 elements with inactivated X genes in mammals, whereas other studies did not find an association (Bailey et al. 2000; Wang et al. 2006). In our study, we did not observe a strong relationship between CR1 elements and gene compensation status, suggesting that in birds LINE elements on the Z chromosome may not be active players in a Z-chromosome-specific process of dosage compensation.

The chicken Z chromosome accumulation of LINEs may have more to do with the low level of recombination of the sex chromosomes. The low meiotic recombination rate results in higher accumulation of repetitive elements on the mammalian X and Y chromosomes (Wichman et al. 1992; Jensen-Seaman et al. 2004). When the full chicken W chromosome sequence becomes available, it will be interesting to investigate in future studies whether the W also has a high LINE content given its low recombination. However, the platypus X chromosomes do not have an accumulation of LINEs compared to autosomes in our study. The opossum sex chromosomes are also not enriched in LINEs (Mikkelsen et al. 2007). Therefore, LINE accumulation may not be strictly limited to sex chromosomes. In mammals, chromosomal size is thought to have a positive correlation with repetitive element enrichment (Wichman et al. 1992; Jensen-Seaman et al. 2004), likely explaining higher LINE content on chicken macrochromosomes in our study.

Interestingly, our results show that the distribution of CpG islands on the chicken Z chromosome is not random, and CpG islands tend to be associated with uncompensated genes. This result is similar to the enrichment of CpG islands around highly expressed genes in mammals. For example, in mammals CpG islands are commonly associated with housekeeping genes, which have generally higher expression than other genes in various tissues (Cross et al. 2000; Hurst et al. 2004). In addition to being highly expressed in males, uncompensated Z genes are also enriched in catalytic activity and other housekeeping functions in birds (Melamed and Arnold 2007). In mammals, methylation of CpG islands contributes to the maintenance of gene silencing on the inactive X chromosome (Cross et al. 2000). X escapees have been found to have fewer CpG islands than inac-

tivated X genes in some studies in mammals, suggesting that lack of DNA methylation is related to the escape from inactivation (Bailey et al. 2000). In future studies in birds, it will be important to assess Z-chromosome-wide methylation and association with dosage compensation.

Our analysis using support vector machine algorithms showed that the CpG island feature is the most predictive of dosage compensation compared to all other types of repetitive elements on the Z chromosome. CpG islands alone could be used to correctly predict compensation status of 63% of genes compared to the random chance of 50%. The percentage prediction rate is lower than in mammalian studies, where 12 features can be used to correctly predict compensation status of about 80% of genes (Wang et al. 2006). It is possible that other types of sequences that were not considered in this study may help improve the predictive power over our current result. Some candidates may be present within the MHM region, which contains concentration of compensated genes (Melamed and Arnold 2007). Alternatively, the lack of an effective chromosome-wide mechanism of dosage compensation suggests that the compensated Z genes have each evolved gene-specific adaptations leading to lower M:F ratios of expression. Such gene-specific adaptations may be heterogeneous, and not involve common DNA elements near compensated genes.

In conclusion, primary DNA sequence information can be used to assess dosage compensation status of genes on the Z chromosome. CpG islands are not randomly distributed on the Z chromosome and may influence Z genes' lack of dosage compensation. CpG islands are also the most predictive feature of Z gene compensation status of all repetitive elements on the Z chromosome, but this feature leads to only a modest level of prediction. LINE elements are important features of the chicken Z and mammalian X but not of the platypus X chromosome suggesting that LINE element accumulation is not a feature of all sex chromosomes. LINEs do not appear to play a role in dosage compensation in chicken compared with their significant role in dosage compensation signal propagation in mammals.

Acknowledgements Thanks to Yuichiro Itoh, Kathy Kampf, Jake Lusis, Daniel Geschwind, and Barney Schlinger for assistance and discussion. Supported by NIH grants DC000217, HD 007228, GM08042, and the NIH Neuroscience Microarray Consortium.

References

Abrusan G, Krambeck HJ (2006) The distribution of L1 and Alu retroelements in relation to GC content on human sex chromosomes is consistent with the ectopic recombination model. J Mol Evol 63:484–492

Abrusan G, Krambeck HJ, Junier T, Giordano J, Warburton PE (2008) Biased distributions and decay of long interspersed nuclear elements in the chicken genome. Genetics 178:573–581

Arnold AP, Itoh Y, Melamed E (2008) A bird's eye view of sex chromosome dosage compensation. Ann Rev Genomics Hum Genet 9:109–127

Bailey JA, Carrel L, Chakravarti A, Eichler EE (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc Natl Acad Sci U S A 97:6634–6639

Bisoni L, Batlle-Morera L, Bird AP, Suzuki M, Mcqueen HA (2005) Female-specific hyperacetylation of histone H4 in the chicken Z chromosome. Chromosome Res 13:205–214

Carrel L, Willard HF (2005) X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature 434:400–404

Chadwick BP, Willard HF (2003) Chromatin of the Barr body: histone and non-histone proteins associated with or excluded from the inactive X chromosome. Hum Mol Genet 12:2167–2178

Chang CC, Lin CJ (2001) LIBSVM: a library for support vector machines. http://www.csie.ntu.edu.tw/~cjlin/libsvm. Accessed January 15, 2009

Cross SH, Clark VH, Simmen MW et al (2000) CpG island libraries from human chromosomes 18 and 22: landmarks for novel genes. Mamm Genome 11:373–383

Deakin JE, Hore TA, Koina E, Marshall Graves JA (2008) The status of dosage compensation in the multiple X chromosomes of the platypus. PLoS Genet 4:e1000140

Ellegren H, Hultin-Rosenberg L, Brunstrom B, Denker L, Kultimaa K, Scholz B (2007) Faced with inequality: chicken does not have a general dosage compensation of sex-linked genes. BMC Biol 5:40

Furano AV (2000) The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. Prog Nucleic Acid Res Mol Biol 64:255–294

Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. J Mol Biol 196:261–282

Grutzner F, Rens W, Tsend-Ayush E et al (2004) In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. Nature 432:913–917

Gupta V, Parisi M, Sturgill D et al (2006) Global analysis of X-chromosome dosage compensation. J Biol 5:3

Hackenberg M, Bernaola-Galvan P, Carpena P, Oliver JL (2005) The biased distribution of Alus in human isochores might be driven by recombination. J Mol Evol 60:365–377

Hore TA, Koina E, Wakefield MJ, Marshall Graves JA (2007) The region homologous to the X-chromosome inactivation centre has been disrupted in marsupial and monotreme mammals. Chromosome Res 15:147–161

Hughes AL, Piontkivska H (2005) DNA repeat arrays in chicken and human genomes and the adaptive evolution of avian genome size. BMC Evol Biol 5:12

Hurst LD, Pal C, Lercher MJ (2004) The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet 5:299–310

International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 432:695–716

Itoh Y, Melamed E, Yang X, Kampf K et al (2007) Dosage compensation is less effective in birds than in mammals. J Biol 6:2

Jensen-Seaman MI, Furey TS, Payseur BA et al (2004) Comparative recombination rates in the rat, mouse, and human genomes. Genome Res 14:528–538

Ke X, Collins A (2003) CpG islands in human X-inactivation. Ann Hum Genet 67:242–249

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190:372–373

Lyon MF (1998) X-chromosome inactivation: a repeat hypothesis. Cytogenet Cell Genet 80:133–137

Lyon MF (2000) LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? Proc Natl Acad Sci U S A 97:6248–6249

Melamed E, Arnold AP (2007) Regional differences in dosage compensation on the chicken Z chromosome. Genome Biol 8:R202

Melamed E, Elashoff D, Arnold AP (2009) Evaluating dosage compensation on the chicken Z chromosome: should effective dosage compensation eliminate sexual bias? Heredity (in press)

Meyer BJ, Mcdonel P, Csankovszki G, Ralston E (2004) Sex and X-chromosome-wide repression in *Caenorhabditis elegans*. Cold Spring Harb Symp Quant Biol 69:71–79

Mikkelsen TS, Wakefield MJ, Aken B et al (2007) Genome of the marsupial *Monodelphis domestica* reveals innovation in non-coding sequences. Nature 447:167–177

Mohandas T, Sparkes RS, Shapiro LJ (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science 211:393–396

Nguyen DK, Disteche CM (2006) Dosage compensation of the active X chromosome in mammals. Nat Genet 38:47–53

Pavlicek A, Jabbari K, Paces J, Paces V, Hejnar JV, Bernardi G (2001) Similar integration but different stability of Alus and LINEs in the human genome. Gene 276:39–45

Pfeifer GP, Steigerwald SD, Hansen RS, Gartler SM, Riggs AD (1990) Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. Proc Natl Acad Sci U S A 87:8252–8256

Plath K, Fang J, Mlynarczyk-Evans SK et al (2003) Role of histone H3 lysine 27 methylation in X inactivation. Science 300:131–135

R Development Core Team (2006) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Rens W, O'Brien PC, Grutzner F et al (2007) The multiple sex chromosomes of platypus and echidna are not completely identical and several share homology with the avian Z. Genome Biol 8:R243

Ross MT, Grafham DV, Coffey AJ et al (2005) The DNA sequence of the human X chromosome. Nature 434:325–337

Shedlock AM, Botka CW, Zhao S et al (2007) Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. Proc Natl Acad Sci U S A 104:2767–2772

Shevchenko AI, Zakharova IS, Elisaphenko EA et al (2007) Genes flanking Xist in mouse and human are separated on the X chromosome in American marsupials. Chromosome Res 15:127–136

Silva R, Burch JB (1989) Evidence that chicken CR1 elements represent a novel family of retroposons. Mol Cell Biol 9:3563–3566

Silva J, Mak W, Zvetkova I et al (2003) Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev Cell 4:481–495

Smit AF, Toth G, Riggs AD, Jurka J (1995) Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. J Mol Biol 246:401–417

Soriano P, Meunier-Rotival M, Bernardi G (1983) The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. Proc Natl Acad Sci U S A 80:1816–1820

Stumph WE, Kristo P, Tsai MJ, O'Malley BW (1981) A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. Nucleic Acids Res 9:5383–5397

Szarski H (1983) Cell size and the concept of wasteful and frugal evolutionary strategies. J Theor Biol 105:201–209

Teranishi M, Shimada Y, Hori T et al (2001) Transcripts of the MHM region on the chicken Z chromosome accumulate as non-coding RNA in the nucleus of female cells adjacent to the DMRT1 locus. Chromosome Res 9:147–165

Wachtel SS, Tiersch TR (1994) The search for the male-determining gene. In: Wachtel SS (ed) Molecular genetics of sex determination. Academic Press, New York, pp 1–22

Wang Z, Willard HF, Mukherjee S, Furey TS (2006) Evidence of influence of genomic DNA sequence on human X chromosome inactivation. PLoS Comput Biol 2:e113

Waters PD, Dobigny G, Waddell PJ, Robinson TJ (2007) Evolutionary history of LINE-1 in the major clades of placental mammals. PLoS ONE 2:e158

Wichman HA, Van Den Bussche RA, Hamilton MJ, Baker RJ (1992) Transposable elements and the evolution of genome organization in mammals. Genetica 86:287–293

Wicker T, Robertson JS, Schulze SR et al (2005) The repetitive landscape of the chicken genome. Genome Res 15:126–136

Zha X, Xia Q, Duan J, Wang C, He N, Xiang Z (2009) Dosage analysis of Z chromosome genes using microarray in silkworm, *Bombyx mori*. Insect Biochem Mol Biol 39:315–321