

RESEARCH

Low Frequency of Microsatellites in the Avian Genome

Craig R. Primmer,¹ Terje Raudsepp,¹ Bhanu P. Chowdhary,¹
Anders Pape Møller,² and Hans Ellegren^{1,3}

¹Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Box 597, S-751 24 Uppsala, Sweden; ²Laboratoire d'Ecologie, Université Pierre et Marie Curie, F-75252 Paris, France

A better insight into the occurrence of microsatellites in a range of taxa may help to understand the evolution of simple repeats. Previous studies have found the relative abundance of several repeat motifs to differ among mammals, invertebrates, and plants. Absolute numbers of microsatellites also tend to correlate positively with genome size. We analyzed the occurrence, frequency, and distribution of microsatellites in birds, a taxon with one of the smallest known genome sizes among vertebrates. Dot-blot hybridization revealed that about half of 22 different di-, tri-, and tetranucleotide repeat motifs were clearly more common in human than in three species of birds: chicken, woodpecker, and swallow. For the remaining motifs no clear difference was found. From searching avian database sequences we estimated there to be 30,000–70,000 microsatellites longer than 20 bp in the avian genome. The number of (CA)_{≥10} would be around 7000–9000 and the number of (CA)_{≥14} about 3000. The calculated density of avian microsatellites (total, one every 20–39 kb; (CA)_{≥10}, one every 136–150 kb) is much lower than that estimated for the human genome (one every 6 and 30 kb, respectively). This may be explained by the fact that the avian genome contains relatively less noncoding DNA than most mammals and that avian SINE/LINE elements do not terminate in poly(A) tails, which are known to provide a resource for the evolution of simple repeats in mammals. We found no association between microsatellites and SINEs in birds. Primed in situ labeling suggested fairly even distribution of (CA)_n repeats over chicken macrochromosomes and intermediate chromosomes, whereas the microchromosomes, a large part of the Z and W chromosomes, and most telomeres and centromeres had very low concentrations of (CA)_n microsatellites. The scarcity of microsatellites on the microchromosomes is compatible to these regions likely being unusually rich in coding sequences. The low microsatellite density in the genome in general and on the microchromosomes in particular imposes an obstacle for the development of marker-rich genetic maps of chicken and other birds, and for the localization of quantitative trait genes.

Less than 10 years after their introduction, microsatellites have developed into the marker of choice in a number of genetic areas, including genome mapping and medical, evolutionary, and ecological genetics. Despite this widespread use, many questions still remain unresolved regarding the evolution of these simple, repeated sequences, not least the mutational processes involved (Amos et al. 1996; Primmer et al. 1996). Such understanding is important for the proper use of microsatellites in evolutionary contexts, for example, for phylogeny reconstruction (Jarne and Lagoda 1996) and also for addressing general questions relating to genome organization. Knowledge of the patterns of microsatellite evolution may also help to reveal whether these sequences are associated with a functional sig-

nificance. One intriguing question in this respect is why certain repeat motifs are more common than others, and why this varies among taxa. In human, (A)_n and (CA)_n are by far the most common motif variants, the latter being the most widely studied marker. Although a similar situation exists in other mammals (e.g., Beckmann and Weber 1992), the same pattern is not true for all taxa. In at least some insect species, (CT)_n is more common than (CA)_n (Estoup et al. 1993) as is the case for an oyster species *Ostrea edulis* (Naciri et al. 1995). In plants, (AT)_n, which is rare in mammals, clearly outnumbers (CA)_n (Lagercrantz et al. 1993).

Related to evolution and a possible functional role, comprehensive studies of the absolute numbers of microsatellites in various genomes are needed to comparatively address whether microsatellite abundance is a direct function of genome size. Microsatellites predominantly occur in noncoding

³Corresponding author.
E-MAIL Hans.Ellegren@bmc.uu.se; FAX +46-18-504461.

DNA, and if the amount of such sequences acts as the sole constraint for the evolution of simple repeats, the absolute numbers of repeats in the genome should correlate closely to DNA content. That this is at least probable has been indicated from hybridization experiments with divergent taxa (Hamada et al. 1982). Furthermore, bats, which have one of the smallest genomes among mammals, appear to harbor considerably fewer microsatellites than other mammalian species (Van Den Bussche et al. 1995).

In this study we have analyzed the occurrence, frequency, and distribution of microsatellites in birds. Avian genomes are generally small; the domestic chicken (*Gallus gallus*) genome, for instance, is estimated to contain close to one-third the number of base pairs of that in the human genome (Bloom et al. 1993; Wachtel and Tiersch 1993). It would thus be expected that avian genomes harbor significantly fewer microsatellites than most mammals, and there are indications that this may be the case for dinucleotide motifs (Hamada et al. 1982; Manor et al. 1988). Our data from dot-blot hybridizations and database searches now show conclusively that microsatellite repeats generally occur less frequently in birds compared to other vertebrates, and in contrast to mammalian microsatellites, bird microsatellites do not appear to be associated with short interspersed repetitive elements (SINEs).

RESULTS

Dot-Blot Hybridization

To compare the abundance of different repeat motifs in avian and mammalian genomes we performed dot-blot hybridizations with equimolar amounts of chicken, white-backed woodpecker *Den-drocopus leucotos*, barn swallow *Hirundo rustica*, pig, and human DNA. The three bird species were chosen so as to represent three divergent avian lineages; human and pig were included as reference mammalian species. A large set of hybridization probes was applied, including 2 dinucleotide repeats [(CA)₁₅ and (CT)₁₅], all 10 possible trinucleotide repeats [(AAT)₁₀, (AAC)₁₀, (AAG)₁₀, (AGT)₁₀, (GAT)₁₀, (ACG)₁₀, (CAG)₁₀, (CCT)₁₀, (GGT)₁₀, (CCG)₁₀], and 9 tetranucleotide repeats [(AAAT)₁₅, (AAAC)₇, (AAAG)₇, (CATT)₇, (GAAT)₇, (GATA)₇, (GACA)₇, (GGAA)₇, and (GGAT)₇].

In general, hybridization signals from mammalian DNA were stronger than signals from avian samples, whereas there were only slight variations between the signal intensities obtained from the

two mammalian and the three avian species (Fig. 1). We scanned the hybridization filters using a PhosphorImager to quantify these differences and derived hybridization intensity ratios for human/bird signals (Table 1). Dinucleotide repeats were particu-

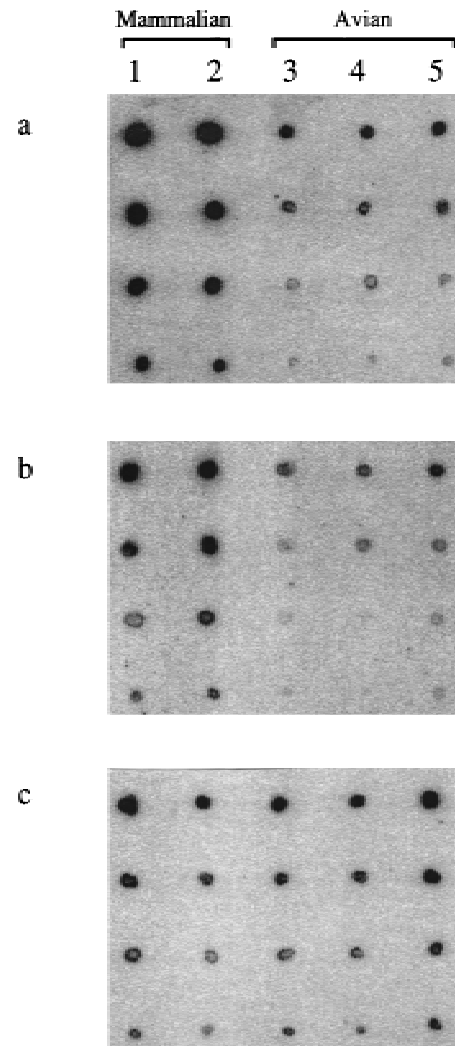


Figure 1 Dot-blot hybridization signals of (CA)₁₅ (a), (AAC)₁₀ (b), and (AAG)₇ (c) to genomic DNA of human (1), pig (2), chicken (3), woodpecker (4), and swallow (5). Intensity ratio calculations from these filters suggested (CA)₁₅ and (AAC)₁₀ to be more common in the mammalian than in the avian genomes, whereas (AAG)₇ appeared to occur equally in all species (refer to Table 1). The uppermost dot on each filter contains 300, 280, 120, 120, and 120 ng of genomic DNA, respectively. The remaining three dots represent 50%, 25%, and 12.5% dilutions of the above concentrations. Different exposure times were used for the filters; therefore signal intensities between filters cannot be compared.

LOW FREQUENCY OF MICROSATELLITES IN BIRDS

Table 1. Intensity Ratios (Human:Bird) of Hybridization Signals to Genomic DNA Probed with Various Microsatellite Motifs

Motif	Chicken	Woodpecker	Swallow	Avian average
AC	14.3	15.0	14.8	14.7
AG	15.9	12.2	10.9	12.7
ACC	7.0	7.6	7.4	7.3
AAAT ^a	≥5.2	≥5.2	≥5.2	≥5.2
AAGT ^a	≥4.5	≥4.5	≥4.5	≥4.5
AAC	4.9	4.6	3.2	4.1
AGAT ^a	≥3.9	≥3.9	≥3.9	≥3.9
ATGG	5.2	3.8	2.0	3.1
AAG	4.5	2.2	2.2	2.6
ACAG	4.5	4.5	1.0	2.0
ATG	3.5	1.8	1.6	2.0
ACG	1.8	5.2	1.1	1.8
AAAC	1.6	2.7	1.2	1.6
AAT	3.2	1.3	1.2	1.6
AGG	1.9	2.0	0.8	1.3
AGC	1.9	1.3	1.0	1.3
AAGG	1.5	1.8	0.8	1.2
CCG	0.8	2.0	0.6	0.8
AAAG	1.6	1.0	0.4	0.7
Mean	4.6	4.2	2.3	3.4

Human and all bird species DNA gave hybridization signals below the background level using the probe AGT.

^aAll three bird species gave hybridization signals below the background level.

larly less frequent in birds, with the CA and CT motifs both being 10–15 times more common in human. Some tri- and tetranucleotide repeat motifs (AAC, ACC, AAAT, AAGT, and AGAT) were similarly biased toward a considerably higher frequency in human. For most of the remaining types of repeats the human/bird intensity ratio fell between 0.5 and 2, which we do not consider to be clear evidence for significantly different frequencies in the two taxa. Only 3 of 22 human/pig ratios fell outside the 0.5–2.0 range, with two motifs being more common in human (AAG ratio = 3.6; AAAT ratio = 2.3) while AAC was more than twice as common in the porcine genome (ratio = 0.4).

For the most commonly used microsatellite marker (CA)_n we also employed dot-blot hybridization to estimate the absolute number of loci present in the chicken genome. Because this estimate is highly dependent on the minimum length of repeats that are detected by the probe, we first determined the lower limit of repeat length uncovered by the hybridization stringency applied. This was accomplished by including a panel of plasmid DNAs with cloned (CA)_n repeats of different length:

(CA)₁₀, (CA)₁₂, (CA)₁₄, (CA)₁₆, and (CA)₁₈. Although little or no signal was obtained with (CA)₁₀ and (CA)₁₂ plasmid DNAs, hybridization intensities for (CA)₁₄, (CA)₁₆, and (CA)₁₈ were about the same (data not shown). From this we concluded that hybridization stringency was adjusted to the detection of repeat lengths of (CA)₁₄ and longer. By relating hybridization intensities of genomic and plasmid DNAs we subsequently estimated there to be ~1500 copies of (CA)_{≥14} in the chicken genome. The corresponding estimates for human and pig DNA were 17,000 and 16,000, respectively.

Survey of Database Sequences

There was ~3.5 Mb of avian non-mitochondrial sequence in Genbank release 93, with the majority constituting chicken sequences. We searched these 2933 sequences for the presence of microsatellite repeats with the criterion of a perfect repetitive

array of at least 20 bp in length, that is, the same search routine as used by Beckmann and Weber (1992) analyzing human repeats. In total, 117 unique bird microsatellites were identified, yielding an average density of 1 every 31 kb (Table 2). This density is clearly lower than one every 6 kb reported for human DNA, but the avian estimate may be flawed, as discussed below. The most common avian repeat type was (A)_n (26 hits) followed, somewhat surprisingly by (AGG)_n (15 hits) and then (CA)_n (10), (CCG)_n (9), (AAAC)_n, and (AAAT)_n (8). Most repeats however, were rather short, and (A)_n and (CA)_n clearly outnumbered all other repeat motifs if only considering loci with at least 10 repeat units (Table 2).

A closer dissection of database sequences indicated that the observed frequencies of different microsatellites may not be representative for the avian (chicken) genome. Avian Genbank entries are significantly biased toward coding sequences, with almost two-thirds of sequences being cDNA clones. When one also considers the proportion of DNA in remaining genomic clones that constitute coding sequences, the total fraction of coding DNA among

Table 2. Number of Avian Microsatellites in the GenBank 93 Database

Motif	Size class (bp)						Total
	20–23	24–27	28–31	32–35	36–39	≥40	
A	13	7	5		1		26
AGG	9	4	2				15
AC	4	2	1	1		2	10
CCG	5	3	1				9
AAAC	6	1	1				8
AAAT	4	1	1			2	8
AGC	4	1		1			6
AT	3					1	4
ACC	3		2				5
AAAG	2					1	3
AAG	1	2					3
AGGG	1	2					3
AAGG	1					1	2
AAT	1	1					2
ACAG	2						2
ACCT	1					1	2
ACGG	1	1					2
AG	1	1					2
C	2						2
AAC	1						1
ACG				1			1
ACGC	1						1
Mononucleotide	15	7	5	0	1	0	28
Dinucleotide	8	3	1	1	0	3	16
Trinucleotide	24	11	5	2	0	0	42
Tetranucleotide	19	5	2	0	0	5	31
Total	66	26	13	3	1	8	117

One hundred eight of the repeats were from domestic chicken, four were from quail, (*Coturnix coturnix*), three were from ducks (*Anas* sp.), one was from turkey (*Meleagris gallopavo*), and one was from pigeon (*Columba livia*).

avian Genbank entries is >60%, a considerably higher proportion than in the genome as a whole. The fact that most microsatellites should be expected to be situated in noncoding DNA suggests therefore that the database screening underestimated the number of bird microsatellites. In support of this, different repeat types showed marked differences with regard to their location within coding or noncoding DNA (Table 3). Importantly, mono-, di-, and tetranucleotide repeats were totally excluded from coding DNA, whereas trinucleotide repeats were found regularly within coding sequences. Therefore, trinucleotide repeats may be over-represented in our sample of database hits.

We used two approaches to derive appropriate

estimates of the absolute numbers of microsatellites in the avian genome. First, similar to Beckman and Weber (1992), we extracted long (>10-kb) genomic clones and tallied the number of microsatellites. Seven repeats found within the 272 kb of DNA contained in 15 such clones give an estimated frequency of one repeat every 39 kb, or a total of ~31,000 loci in the genome. This density, though obviously very approximate, is significantly lower than that calculated by Beckmann and Weber (1992) for the human genome ($\chi^2 = 16.1$, $P = 0.0001$; $df = 1$; 122 microsatellites in 745 kb). Second, we sought to convert the frequencies of avian microsatellites in Genbank to that expected for the whole genome by compensating for the overrepresentation of expressed sequences in the database. This can be done using the simple expression $D = 1/[af_c + (1 - a)f_{nc}]$, where D is the genomic density (one microsatellite every D bp), a is the proportion of coding DNA in the genome and f_c and f_{nc} are the microsatellite frequencies (per bp) in coding and noncoding DNA, respectively, as determined from database occurrence (see Methods). A problem in this context is to derive an estimate of what proportion of the avian genome

is made up of coding DNA. In mammals, a figure of 10% is often assumed (Hochgeschwender and Brennan 1991), but as birds have only about one-third of the mammalian DNA content, avian genes may constitute as much as 30% of the genome. Using gene proportions of both 10% and 30%, the total number of microsatellites ≥ 20 bp was estimated at 60,000–73,000 (one every 16–20 kb; Table 4). The number of $(CA)_{\geq 10}$ copies would be 6600–8500, or one every 140–180 kb. Although data are sparse and must be treated cautiously, the number of $(CA)_{\geq 14}$ would be 2600–3400 (one every 350–450 kb). Interestingly, both database approaches used for estimating total microsatellite numbers in the avian genome gave similar figures, suggesting a markedly

LOW FREQUENCY OF MICROSATELLITES IN BIRDS

Table 3. Distribution of Repeat Classes in Different DNA Regions

Motif	Noncoding	5' UTR	Coding	3' UTR	Unknown
Mono-	12	2	0	6	8
Di-	9	3	0	3	1
Tri-	4	2	28	1	7
Tetra-	15	6	0	3	7

lower density than in the human genome. Furthermore, the estimated number of avian $(CA)_{\geq 14}$ loci was similar to that obtained from the dot-blot hybridization experiments.

No Association with Other Repetitive Elements

Various microsatellites in other taxa are known to be directly associated with SINEs (e.g., Beckmann and Weber 1992) and may even have evolved from sequences contained in such elements (Arcot et al. 1995; Nadir et al. 1996). We therefore investigated whether this type of association is also found in birds. The predominant SINE in chicken is called CR1 (Stumph et al. 1981), and estimates of its copy number range from 7,000 to 20,000 (Hache and Deeley 1988) to >100,000 (Vandergon and Reitman 1994). In 81 genomic sequences harboring CR1 elements (Vandergon and Reitman 1994) we found 14 that also harbored microsatellites. In none of these was the microsatellite directly associated with the CR1 element. In fact only one $(A)_n$ repeat was located 25 bp 3' of a CR1 element whereas all other 13 microsatellite repeats were situated farther than 350 bp, 5' or 3', from the SINE. Moreover, a database search with, on average, 40 bp of DNA flanking each side of all 117 identified microsatellites failed to find significant homology with known repetitive elements, avian or otherwise, or with each other, which also indicates that any association between microsatellites and SINEs in birds would, at best, be rare.

Primed In Situ Labeling

We also investigated whether the apparent lower density of avian microsatellites could be explained by an unusual chromosomal distribution. Primed in situ (PRINS) labeling (Koch et al. 1989) of chicken metaphase chromosome spreads with a $(CA)_{10}$ primer suggested, however, a relatively even distri-

bution of $(CA)_n$ repeats over the macrochromosomes and the intermediate chromosomes, with slight variation in the intensity of fluorescent signals, as exemplified in Figure 2. To get a proper picture of how signal intensities varied over chromosomes, hybridization signals from 40 metaphase spreads were quantified by dividing the larger autosomes and the sex chromosomes into equal sized discrete regions.

This was not possible for the intermediate chromosomes and microchromosomes, as they are extremely difficult to identify by cytological means. Neither were we able to distinguish the two chromosomal arms of the nearly metacentric W chromosome. The signal in each region was classified as weak, medium, or strong (Fig. 3). Several interesting patterns emerged. First, pooled data from all microchromosomes revealed that they generally hybridized less intensely. Second, the telomeric regions of the macrochromosomes also hybridized less brightly than other chromosomal regions. This was also true for most centromeres, although this is not manifested in Figure 3 because of the classified intervals containing centromeres generally being too large to allow such a resolution. Third, although mostly showing only weak signals over large parts of the chromosomes, the telomeric part of the p arm of the Z chromosome and, similarly, the end of one of the chromosomal arms of the W chromosome showed intense signals. Therefore, these data indicate a relatively low frequency of $(CA)_n$ microsatellites on the microchromosomes, on most of the Z and W chromosomes and on telomeres and centromeres.

DISCUSSION

Density of Avian Microsatellites

Dot-blot hybridization, as well as the incidence of repeat sequences in database entries, suggests microsatellites to be generally less frequent in birds than in mammals. For instance, 11 of 22 di-, tri-, and tetranucleotide repeat motifs gave considerably weaker dot-blot hybridization signals to avian than to human or porcine DNA, whereas no probe motif showed the opposite trend. Based on database screening, we estimated the total number of microsatellites ≥ 20 bp in the avian genome to be 60,000–73,000, or 1 every 16–20 kb and if only genomic

Table 4. Estimated Numbers of Microsatellites ≥ 20 bp in Length in the Avian Genome Assuming a Gene Content of 10% or 30% per Genome

Motif	10% coding DNA		30% coding DNA	
	total	spacing (kb)	total	spacing (kb)
A	22,085	54	17,177	70
AC ^a	8,494	141	6,607	182
AAAC	6,795	177	5,285	227
AAAT	6,795	177	5,285	227
AT	3,398	353	2,643	454
AGG	3,191	376	3,912	307
AAAG	2,548	471	1,982	605
AGGG	2,548	471	1,982	605
AAGG	1,699	706	1,321	908
AAT	1,699	706	1,321	908
ACAG	1,699	706	1,321	908
ACGG	1,699	706	1,321	908
ACCT	1,699	706	1,321	908
AG	1,699	706	1,321	908
C	1,699	706	1,321	908
CCG	1,278	939	1,947	616
AGC	1,117	1,074	1,465	819
AAG	956	1,254	993	1,222
ACG	849	1,413	661	1,816
ACGC	849	1,413	661	1,816
ACC	268	4,478	804	1,493
AAC	54	22,388	161	7,463
Mononucleotide	23,784	50	18,498	65
Dinucleotide	13,591	88	10,571	114
Trinucleotide	9,414	127	11,253	107
Tetranucleotide	26,332	46	20,480	59
Total	73,120	16	60,802	20

Avian species as in Table 2.

^aThe proportion of these repeats estimated to be ≥ 14 repeat units in length was 2600–3400 (1 in 350–450 kb).

clones ≥ 10 kb were considered, the estimated number of avian repeats was only 31,000 (1 every 39 kb). The number of $(CA)_{\geq 10}$ copies would be 6600–8500 (1 every 140–180 kb), and the number of $(CA)_{\geq 14}$ loci would be 2600–3400 (1 every 350–450 kb). Dot-blot hybridization gave an estimate of 1500 loci for the latter repeat type. The higher densities are reached if the proportion of genes in the avian genome is 10%, as assumed for many mammals, whereas the lower densities can be estimated if genes constitute 30% of the avian genome. A figure of 30% seems more plausible, as the mean avian genome size is only 1.2×10^9 bp (Tiersch and Wachtel 1991; Bloom et al. 1993). This also depends on whether birds have about the same number of

genes as mammals and if mean gene size is similar in the two taxa. Comparisons of exon lengths in orthologous genes (Hughes and Hughes 1995) lend support to a similar gene size in human and birds. It should be noted that our estimated microsatellite frequencies were obtained mainly from chicken sequence data. The similarity of avian hybridization intensities for most repeat motifs (Table 1) and the remarkable consistency of bird species' DNA content (Tiersch and Wachtel 1991) indicate that these estimated repeat frequencies may also be representative for other birds.

The smaller genome size of birds compared to mammals would logically predict avian genomes to contain lower absolute numbers of microsatellites than, for example, human, as was the case. However, the number of microsatellites in the human compared to the avian genome still exceeds this expectation greatly. Although our dot-blot data suggested that there are 1500 $(CA)_{\geq 14}$ copies in chicken, the number in human was at least 10 times more, $\sim 17,000$. Our human estimate is considerably less than the 50,000–100,000 estimated in earlier studies (e.g., Hamada et al. 1982; Tautz and Renz 1984; Beckmann

and Weber 1992), but these kind of estimates are highly dependent on the length of repeats detected during hybridization and it is evident that at least some of the previous studies of human DNA have included $(CA)_n$ loci with <14 repeat units. Our database screenings gave further support for the non-proportional excess of human microsatellites. We estimated the total microsatellite density to be one every 39 kb in birds, whereas Beckmann and Weber (1992) obtained an estimate of one every 6 kb in humans using similar search procedures. Similarly, our estimate of one $(CA)_{\geq 10}$ copy every 136 kb in birds is clearly lower than one every 30 kb estimated in human (Beckmann and Weber 1992). The general low frequency of avian microsatellites found in

LOW FREQUENCY OF MICROSATELLITES IN BIRDS

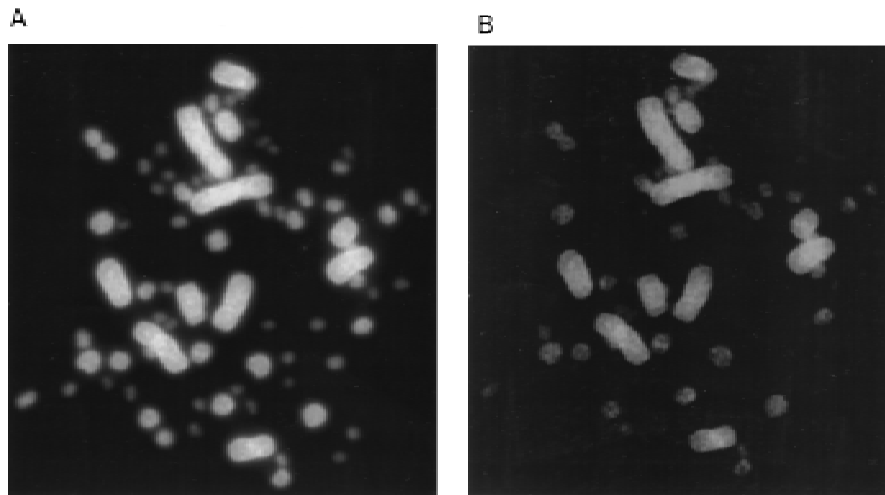


Figure 2 PRINS hybridization of a $(CA)_{10}$ oligonucleotide to a female chicken metaphase spread. (A) Chromosomes were counterstained with DAPI. (B) $(CA)_{10}$ hybridization in the absence of DAPI staining

this study is compatible with circumstantial observations of low numbers of avian dinucleotide repeats found previously (Hamada et al. 1982; Manor et al. 1988; Moran 1993; Cheng et al. 1995).

Is the fact that birds seem to be devoid of microsatellites adaptive, or is this just a matter of coincidence? This question should take into account the low levels of interspersed repetitive DNA observed in birds (~30%; Eppelen et al. 1978; Eden and Hendrick 1978; Venturini et al. 1987) and the shorter average size of avian introns compared with human homologs (Hughes and Hughes 1995). Taken together, this suggests that there is (or has been) a constraint on avian genome size operating. It has been speculated that the physical requirements related to flight may place limits on genome size and therefore on the amount of nonfunctional DNA tolerable (Cavalier-Smith 1978; Hughes and Hughes 1995). In support of this theory, bats have one of the smallest known genome sizes of all mammals (Burton et al. 1989) and have also a reduced frequency of microsatellites (Van Den Bussche et al. 1995).

If one assumes that there is a constraint on avian genome size, it still remains to be explained why the density of microsatellites is low in birds. A higher gene density in birds compared with mammals would imply that the relative amount of noncoding DNA should be less in birds, in turn giving less opportunities for avian microsatellites to evolve. This could be accentuated if unique noncoding DNA is favored over repetitive DNA in birds, a possible scenario if noncoding DNA acts at least

partly as raw material for the evolution of novel genes. If so, it remains to be clarified if the low tempo of microsatellite evolution is governed by factors involved in the generation of simple repeats by replication slippage or solely by direct selection against “superfluous” repetitive DNA.

Another situation that may contribute to low frequencies of microsatellites in birds is the apparent lack of association between avian SINEs/LINEs (long interspersed repetitive elements) and microsatellites as compared to mammals, where the two types of repetitive elements are commonly found

adjacent to each other (Economou et al. 1990; Beckmann and Weber 1992; Buchanan et al. 1993; Ellegren 1993; Varvio and Kaukinen 1993). Mammalian SINEs (and LINEs) regularly terminate in a poly(A) tail derived from retrotransposition, and these stretches provide the basis for the evolution of more complex A-rich repeats like $(A_{1-3}N)_n$ (Arcot et al. 1995; Duffy et al. 1996; Nadir et al. 1996). Because SINEs and LINEs appear to be less abundant in birds than in mammals (Deininger 1989; Hutchison et al. 1989) and, importantly, because those avian interspersed elements characterized so far do not terminate with a poly(A) tail (e.g., Chen et al. 1995), low frequencies of microsatellites in birds may be at least partly attributable to a lack of poly(A) tails of interspersed elements providing a source for transitions into various types of repeats.

Distribution of $(CA)_n$ Loci on Chicken Chromosomes

PRINS labeling of chicken metaphase chromosomes using a $(CA)_{10}$ primer suggested a relatively even distribution of $(CA)_n$ loci across the five largest chromosomes and the intermediate chromosomes. The only exceptions were the centromeres and telomeres, which consistently stained less brightly on all chromosomes. Centromeres and telomeres are known from a wide variety of species to be built up by other specific repeats (e.g., Biessmann and Mason 1992), and the scarcity of $(CA)_n$ microsatellites in these regions is therefore not surprising. A lack of $(CA)_n$ repeats has also been observed on porcine centromeres and telomeres using the same tech-

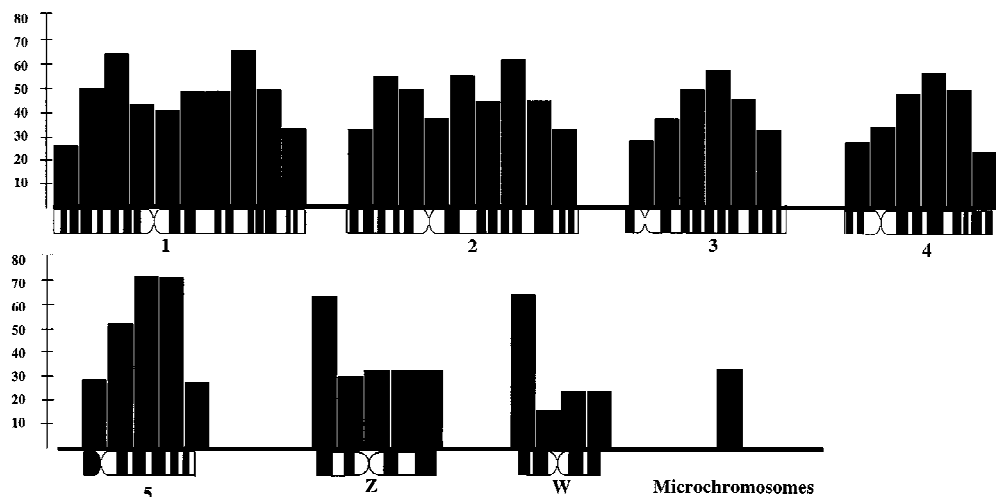


Figure 3 Quantification of PRINS $(CA)_{10}$ hybridization to 40 chicken metaphase spreads. Signal intensities for the five largest chromosomes, the sex chromosome, and the microchromosomes as a whole were classified using a point system as either weak (0), average (+1), or strong (+2). These values were transformed into percentages for defined regions of each chromosome so that, e.g., a classification of average for a particular chromosomal region on all metaphases correlates to a value of 50%. Signals from the W chromosome are shown with the assumption that the bright signal on the end of one of the chromosomal arms was consistently from the end of the p arm (see text).

nique (Winterø et al. 1992). Most of the female-specific W chromosome (birds have a reversed sex chromosome system as compared with mammals, with females being heterogametic ZW and males being homogametic ZZ) also gave rather faint signals with only one of the chromosomal ends staining brightly. We were not able to distinguish the two arms of the W chromosome, but the fact that we never observed a W chromosome with intense signals at both ends suggests that it consistently was the same end that labeled brightly. Some 75% of the chicken W chromosome constitutes huge arrays of satellite DNAs of a few repeat motifs (Saitoh and Mizuno 1992). These fill up the entire q arm and the proximal half of the p arm. Accordingly, the W chromosome harbors very few coding sequences, and the first avian W-linked gene has been identified only recently (Ellegren 1996; Griffiths et al. 1996). The fact that only the distal part of the p arm appears to contain unique DNA (Saitoh and Mizuno 1992) agrees very well with our detection of $(CA)_n$ loci on only one end of the chromosome, which hence should be expected to be the telomeric part of the p arm. The few genes present on the W chromosome should similarly be expected to reside in this region. The observation that the $(CA)_{10}$ probe hybridized more intensely to the telomeric part of the p arm of the Z chromosome than to the remaining part of the chromosome is more surprising. Perhaps the similar distribution of $(CA)_n$ loci on

the Z and W chromosomes reflects a common ancestry.

Generally faint signals from the microchromosomes suggested $(CA)_n$ density to be low on these tiny DNA segments. An alternative explanation could possibly be that the small size of the chromosomes prevented strong signals from emerging to some extent; however, we do not think this was the case because careful examination of 40 metaphases did not reveal any strong signal from any microchromosome. It has become evident from the assignments of a number of genes that the microchromosomes do not represent genetically inert reserves of DNA (e.g., Bloom and Bacon 1985), a situation assumed previously. McQueen et al. (1996) have shown recently that there are a disproportionately high number of CpG islands, indicative of a high gene content, on the chicken microchromosomes. The average density may be as high as one gene per 10 kb, an exceptional value as compared with genome organization in mammals. This high concentration of genes coupled with a generally shorter avian intron size (Hughes and Hughes 1995) would imply reduced possibilities for the evolution of simple repeats from noncoding DNA on the microchromosomes. It should be noted, however, that the lower microsatellite density in the avian genome cannot be attributed entirely to their sparsity on the microchromosomes as they only account for 25% of the DNA in the avian genome (Bloom et al. 1993).

LOW FREQUENCY OF MICROSATELLITES IN BIRDS

Practical Implications for Avian Genome Analysis

Regardless of the reason behind the lower frequency of microsatellites in birds, this factor will limit the ability to develop marker-rich genetic maps to the extent of those achieved in mammalian species. The lack of association with SINES/LINES helps to but does not entirely cancel out this negative effect, as it should be possible to derive locus-specific amplification of a higher percentage of microsatellites than it is when these microsatellites are immediately flanked by other repetitive elements (Economou et al. 1990). As there does not appear to be a single microsatellite motif in birds as common as (CA)_n in mammals, it would be advisable to search with a range of repeat motifs in library screenings. The low density of avian microsatellites also warrants the use of marker-enriched libraries (e.g., Armour et al. 1994) to avoid screening extensive numbers of clones from bulk libraries. Another practical consequence is that cloning the same marker more than once will become increasingly common (Cheng et al. 1995), highlighting the importance of screening sequence databases before developing PCR primers for new clones.

The particularly low microsatellite density on microchromosomes, together with a contrasting high gene density, introduces a special obstacle for genetic mapping in chicken and other birds. Chicken microchromosomes possess an obligate chiasmata, meaning that the genetic length of each chromosome is 50 cM. The 29 chicken microchromosomes contain ~300 Mb (Bloom et al. 1993), and the average ratio of genetic to physical distance would thus be 5 cM per megabase of DNA. This is an extremely skewed figure, paralleling that recorded in telomeric regions of mammalian chromosomes (e.g., Hultén 1974). Frustratingly, the chromosomal regions likely associated with the highest gene contents also coincide with regions of unusually high recombination frequencies but with reduced numbers of potential genetic markers. Coupled with the difficulty in using comparative gene map data from phylogenetically divergent mammals like human and mouse, the localization of trait genes by positional (candidate) cloning will not be easy for avian geneticists.

METHODS

Dot-Blot Hybridization

Genomic DNA was prepared by proteinase K digestion, phenol extraction, and ethanol precipitation from blood of three avian species belonging to different superorders: the domestic

chicken, the white-backed woodpecker, and the swallow and from human and large white domestic pig. The quality of the different DNA preparations was similar, as judged from A₂₆₀/A₂₈₀ ratios. The amount of genomic DNA used on dot-blot filters was calculated taking into account the differing genome sizes so that the quantity of DNA used for each species contained approximately the same number of genome copies (equimolar amounts). Starting from 300 ng of human DNA (assuming a genome size of 3.0×10^9 bp), 280 ng of pig DNA (2.8×10^9 bp; Schmitz et al. 1992) and 120 ng of DNA of the three avian species (1.2×10^9 bp; Bloom et al. 1993), serial dilutions of 50%, 25%, and 12.5% were also analyzed to minimize potential experimental biases (cf. Anchordoguy et al. 1996).

Twelve identical Hybond-N+ filters (Amersham, UK) containing DNA as indicated above were prepared following the manufacturer's instructions by manually dotting DNA solutions using recently calibrated pipettes. Filters were prehybridized for 1 hr in a solution of 7% SDS, 1 mM EDTA, 1% BSA, 0.26 M Na₂HPO₄, and 10% dextran sulfate at 65°C, after which 40 pmoles of single-stranded oligonucleotide, previously end-labeled using 1.6 µCi of [γ -³²P]ATP, was added to the solution and hybridization was allowed to continue overnight. Filters were then washed three times in $0.3 \times$ SSC solution; once for 5 min at room temperature, followed by two 15-min washes at the following temperatures (calculated melting temperature of oligonucleotide in $0.3 \times$ SSC 5°C): (AAT)₁₀ at 32.4°C; (AAAC)₇, (AAAG)₇, (CATT)₇, (GAAT)₇, and (GATA)₇ at 41.0°C; (AAAT)₁₅ at 43.7°C; (AAC)₁₀, (AAG)₁₀, (AGT)₁₀, and (GAT)₁₀ at 46.1°C; (GACA)₇, (GGAA)₇, and (GGAT)₇ at 51.3°C; (CA)₁₅ and (CT)₁₅ at 52.9°C; (ACG)₁₀, (CAG)₁₀, (CCT)₁₀, and (GGT)₁₀ at 59.7°C; and (CCG)₁₀ at 73.4°C. The two remaining dinucleotide motifs, (AT)_n and (GC)_n, were not included in the experiment because of the difficulty of using self-complementary probes. The filters were subsequently exposed to PhosphorImager screens for 4–40 hr, and images were scanned from the screens using a PhosphorImager scanner (Molecular Dynamics). The pixel values above background level were calculated for all dots using ImageQuant software (Molecular Dynamics) with a gray-scale range of 2–100 and local background. The 12 di- and trinucleotide oligonucleotide probes were first hybridized to filters and analyzed, after which filters were stripped with boiling water, exposed overnight as described above to ensure that probe molecules had been removed efficiently, and then screened with the 9 tetranucleotide probes.

To determine the absolute number of (CA)_n microsatellites in the chicken genome along with calibrating hybridization stringency with the minimum length of repeats detected, a filter with human and chicken DNA dilutions together with five plasmid (pUC19) preparations of clones containing different sized (CA)_n repeats ($n = 10, 12, 14, 16, \text{ or } 18$) was also prepared. Plasmid dilutions equivalent to 25,000, 12,500, and 5,000 CA copies were included. Hybridization conditions and method of detection were as described above. The number of (CA)_n repeats in the avian and human genomes was estimated from the calibration of signal intensities against known plasmid copy numbers.

Database Searching

Searches for all possible avian mono-, di-, tri-, and tetranucleotide motifs at least 20 bp in length (no mismatches allowed) were carried out on the (nonmammalian vertebrate)

VRT subsection of GenBank release 93.0, using the FASTA program of the GCG computer package. Sequences annotated as derived from microsatellite library screenings were excluded. Hits corresponding to the poly(A) tails of cDNA clones were also excluded as were mononucleotide stretches apparently originating from terminal transferase treatment. In cases when the same microsatellite locus was present more than once in the database, only a single match was recorded.

One hundred genomic and 100 cDNA GenBank chicken entries were selected randomly using the Entrez World Wide Web site (<http://www3.ncbi.nlm.nih.gov/Entrez/>) to derive an estimate of the relative proportions of noncoding, transcribed but not translated (UTR), and coding DNA in the database. Estimates of the total number of base pairs in the different categories were subsequently obtained by extrapolating figures from this selection. Regions from entries not annotated to a particular category were allocated proportionally to the different groups. By relating the number of microsatellites found in coding and noncoding DNA, respectively, to the estimated number of base pairs contained in GenBank, microsatellite frequencies in the two types of DNA were calculated. Totally, there were 1026 genomic sequences and 1907 cDNA sequences in the database.

Primed In Situ Labeling

Chicken metaphase chromosome spreads were obtained from bone marrow using standard methods (e.g., Christidis 1983). Briefly, bone marrow cells from chicken femur were incubated in RPMI-1640 (Sigma), containing 30% fetal calf serum and 0.1 mg/ml of colcemid for 1 hr, followed by routine hypotonic treatment and fixation. The labeling reaction was based on the study of Winterø et al. (1992) as follows: 50 μ l of reaction mixture that contained 150 pmoles of CA₁₀ primer, 1 mM dNTP, 10 mM spectrum orange dUTP (Vysis Inc.), 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 2.0 mM MgCl₂, and 5 units of AmpliTaq DNA polymerase (Perkin Elmer), was added to a metaphase slide, and a coverslip was added and sealed with rubber cement to limit evaporation. Slides were then incubated at 95°C for 5 min followed by a 1 hr incubation at 60°C or 65°C. The reaction was terminated by washing once in 50 mM NaCl, 50 mM EDTA at 60°C for 5 min followed by two 5 min washes at room temperature in 4 \times SSC, 0.1% Tween 20. Chromosomes were counterstained with 4',6-diamidino-2-phenylindole (DAPI) that produced a G-band like pattern for chromosome identification. The results were analyzed using an Olympus BX60 fluorescence microscope equipped with an IMAC-CCD S30 video camera. Images were captured and analyzed using ISIS 1.65 (Metasystems) software.

ACKNOWLEDGMENTS

We thank Fredrik Söderbom for assistance with the PhosphorImager analysis and Hugh Williams for assistance with database searching.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Amos, W., S.J. Sawcer, R.W. Feakes, and D.C. Rubeinsztein.

1996. Microsatellites show mutational bias and heterozygote instability. *Nature Genet.* 13: 390–391.

Anchordoguy, T.J., D.L. Crawford, I. Hardewig, and S.C. Hand. 1996. Heterogeneity of DNA binding to membranes used in quantitative dot blots. *BioTechniques* 20: 754–756.

Arcot, S.S., Z. Wang, J.L. Weber, P.L. Deininger, and M.A. Batzer. 1995. Alu repeats: A source for the genesis of primate microsatellites. *Genomics* 29: 136–144.

Armour, J.A.L., R. Neumann, S. Gobert, and A.J. Jeffreys. 1994. Isolation of human simple repeat loci by hybridisation selection. *Hum. Mol. Genet.* 3: 599–605.

Beckmann, J.S. and J.L. Weber. 1992. Survey of human and rat microsatellites. *Genomics* 12: 627–631.

Biessmann, H. and J.M. Mason. 1992. Genetics and molecular biology of telomeres. *Adv. Genet.* 30:185–249.

Bloom, S.E. and L.D. Bacon. 1985. Linkage of the major histocompatibility (B) complex and the nucleolar organiser in the chicken. *J. Hered.* 76: 146–154.

Bloom, S.E., M.E. Delany, and D.E. Muscarella. 1993. Constant and variable features of avian chromosomes. In *Manipulation of the avian genome* (ed. R.J. Etches and A.M. Verrinder Gibbins), pp. 39–60. CRC Press, Boca Raton, Florida.

Buchanan, F.C., R.P. Littlejohn, S.M. Galloway, and A.M. Crawford. 1993. Microsatellite and associated repetitive elements in the sheep genome. *Mamm. Genome* 4: 258–264.

Burton, D.W., J.W. Bickham, and H.H. Genoways. 1989. Flow-cytometric analyses of nuclear DNA content in four families of neotropical bats. *Evolution* 43: 756–765.

Cavalier-Smith, T. 1978. Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J. Cell Sci.* 34: 247–278.

Chen, Z.-Q., R.G. Ritzel, C.C. Lin, and R.B. Hodgetts. 1991. Sequence conservation in avian CR1: An interspersed repetitive DNA family evolving under functional constraints. *Proc. Natl. Acad. Sci.* 88: 5814–5818.

Cheng, H.H., I. Levin, R.L. Vallejo, H. Khatib, J.B. Dodgson, L.B. Crittenden, and J. Hillel. 1995. Development of a genetic map of the chicken with markers of high utility. *Poultry Sci.* 74: 1855–1874.

Christidis, L. 1983. Extensive chromosomal repatterning in two congeneric species: *Pytilia melba*, L. and *Pytilia phoenicoptera* Swainson (Estrildidae; Aves). *Cytogenet. Cell Genet.* 36: 641–648.

Deininger, P.L. 1989. SINEs: Short interspersed repeated DNA elements in higher eucaryotes. In *Mobile DNA* (ed. D.E. Berg and M.M. Howe), pp. 619–636. American Society for Microbiology, Washington, D.C.

LOW FREQUENCY OF MICROSATELLITES IN BIRDS

- Duffy, A.J., D.W. Coltman, and J.M. Wright. 1996. Microsatellites at a common site in the second ORF of L1 elements in mammalian genomes. *Mamm. Genome* 7: 386–387.
- Economou, E.P., A.W. Bergen, A.C. Warren, and S.E. Antonarakis. 1990. The polydeoxyadenylate tract of Alu repetitive element is polymorphic in the human genome. *Proc. Natl. Acad. Sci.* 87: 2951–2954.
- Eden, F.C. and J.P. Hendrick. 1978. Unusual organisation of DNA sequences in the chicken. *Biochemistry* 17: 5838–5844.
- Ellegren, H. 1993. Variable SINE 3' poly(A) sequences: An abundant class of genetic markers in the pig genome. *Mamm. Genome* 4: 429–434.
- . 1996. First gene on the avian W chromosome (CHD) provides a tag for universal sexing of non-ratite birds. *Proc. R. Soc. Lond. Biol. Sci.* 263: 1635–1641.
- Epplen, J.T., M. Leipoldt, W. Engel, and J. Schmidtke. 1978. DNA sequence organisation in avian genomes. *Chromosoma* 69: 307–321.
- Estoup, A., M. Solignac, M. Harry, and J.-M. Cornuet. 1993. Characterization of (GT)_n and (CT)_n microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. *Nucleic Acids Res.* 21: 1427–1431.
- Griffiths, R., S. Daan, and C. Dijkstra. 1996. Sex identification in birds using two CHD genes. *Proc. R. Soc. Lond. Biol. Sci.* 263: 1249–1254.
- Hache, R.J.G. and R.G. Deeley. 1988. Organization, sequence and nuclease sensitivity of repetitive elements flanking the chicken apoVLDLII gene: Extended sequence similarity to elements flanking the chicken vitellogenin gene. *Nucleic Acids Res.* 16: 97–113.
- Hamada, H., M.G. Petrino, and T. Kakunaga. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse genomes. *Proc. Natl. Acad. Sci.* 79: 6465–6469.
- Hochgeschwender, U. and M.B. Brennan. 1991. Identifying genes within the genome: New ways for finding the needle in a haystack. *BioEssays* 13: 139–144.
- Hughes, A.L. and M.K. Hughes. 1995. Small genomes for better flyers. *Nature* 377: 391.
- Hultén, M. 1974. Chiasma distribution at diakinesis in the normal human male. *Hereditas* 76: 55–78.
- Hutchison, C.A. III, S.C. Hardies, D.D. Loeb, W.R. Shehee, and M.H. Edgell. 1989. LINEs and related retrotransposons: long interspersed repeated sequences in the eucaryotic genome. In *Mobile DNA* (ed. D.E. Berg and M.M. Howe), pp. 593–617. American Society for Microbiology, Washington, D.C.
- Jarne, P. and P.J.L. Lagoda. 1996. Microsatellites, from molecules to populations and back. *Trends Ecol. Evol.* 11: 424–429.
- Koch, J.E., S. Kolvraa, K.B. Petersen, N. Gregersen, and L. Bolund. 1989. Oligonucleotide-priming methods for the chromosome-specific labelling of alpha satellite DNA in situ. *Chromosoma (Berl)* 98: 259–265.
- Lagercrantz, U., H. Ellegren, and L. Andersson. 1993. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* 21: 1111–1115.
- Manor, H., B. Sridhara Rao, and R.G. Martin. 1988. Abundance and degree of dispersion of genomic d(GA)_n · d(TC)_n sequences. *J. Mol. Evol.* 27: 96–101.
- McQueen, H.A., J. Fantes, S.H. Cross, V.H. Clark, A.L. Archibald, and A.P. Bird. 1996. CpG islands of chicken are concentrated on microchromosomes. *Nature Genet.* 12: 321–324.
- Moran, C. 1993. Microsatellite repeats in pig (*Sus domestica*) and chicken (*Gallus domesticus*) genomes. *J. Hered.* 84: 274–280.
- Naciri, Y., Y. Vigouroux, J. Dallas, E. Desmarais, C. Delsert, and F. Bonhomme. 1995. Identification and inheritance of (GA/TC)_n and (AC/GT)_n repeats in the European flat oyster *Ostrea edulis* (L.). *Mol. Mar. Biol. Biotechnol.* 4: 83–89.
- Nadir, E., H. Margalit, T. Gallily, and S.S. Ben-Sasson. 1996. Microsatellite spreading in the human genome: Evolutionary mechanisms and structural implications. *Proc. Natl. Acad. Sci.* 93: 6470–6475.
- Primmer, C.R., N. Saino, A.P. Møller, and H. Ellegren. 1996. Directional evolution in germline microsatellite mutations. *Nature Genet.* 13: 391–393.
- Saitoh, Y. and S. Mizuno. 1992. Distribution of XhoI and EcoRI family of repetitive DNA sequences into separate domains in the chicken W chromosome. *Chromosoma (Berl.)* 101: 474–477.
- Schmitz, A., B. Chaput, P. Fouchet, M. Noelle Guilly, and G. Frelat. 1992. Swine chromosomal DNA quantification by bivariate flow karyotyping and karyotype interpretation. *Cytometry* 13: 703–710.
- Stumph, W.E., P. Kristo, M.J. Tsai, and B.W. O'Malley. 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. *Nucleic Acids Res.* 9: 5383–5397.
- Tautz, D. and M. Renz. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Res.* 12: 4127–4138.
- Tiersch, T.R. and S.S. Wachtel. 1991. On the evolution of genome size in birds. *J. Hered.* 82: 363–368.
- Van Den Bussche, R.A., J.L. Longmire, and R.J. Baker. 1995.

PRIMMER ET AL.

How bats achieve a small C-value: Frequency of repetitive DNA in *Macrotus*. *Mamm. Genome*. 6: 521–525.

Vandergon, T.L. and M. Reitman. 1994. Evolution of chicken repeat 1 (CR1) elements: Evidence for ancient subfamilies and multiple progenitors. *Mol. Biol. Evol.* 11: 886–898.

Varvio, S.L. and J. Kaukinen. 1993. Bovine microsatellites: racial differences and association with SINE-elements. In *DNA fingerprinting: State of the science* (ed. S.J.D. Pena, R. Chakraborty, J.T. Epplen, and A.J. Jeffreys) pp. 437–443. Birkhäuser Verlag, Basel, Switzerland.

Venturini, G., E. Capanna, and B. Fontana. 1987. Size and structure of the bird genome. II. Repetitive DNA and sequence organization. *Comp. Biochem. Physiol.* 87B: 975–979.

Wachtel, S.S. and T.R. Tiersch. 1993. Variations in genome mass. *Comp. Biochem. Physiol.* 104B: 207–213.

Winterø, A.K., M. Fredholm, and P.D. Thomsen. 1992. Variable (dG-dT)_n · (dC-dA)_n sequences in the porcine genome. *Genomics* 12: 281–288.

Received December 18, 1996; accepted in revised form March 3, 1997.



Low Frequency of Microsatellites in the Avian Genome

Craig R. Primmer, Terje Raudsepp, Bhanu P. Chowdhary, et al.

Genome Res. 1997 7: 471-482

Access the most recent version at doi:[10.1101/gr.7.5.471](https://doi.org/10.1101/gr.7.5.471)

References

This article cites 43 articles, 5 of which can be accessed free at:
<http://genome.cshlp.org/content/7/5/471.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Affordable, Accurate
Sequencing.



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>
