NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

UNIVERSITY OF ALBERTA

STUDIES ON TWO REPETITIVE DNA FAMILIES IN THE AVIAN GENOME

BY

Zhi-Qing Chen

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

DEPARTMENT OF GENETICS

EDMONTON, ALBERTA

FALL, 1990.

ISBN 0-315-65006-0

University of Alberta
Edmonton
Canada T6G 2E9

Department of Genetics
Faculty of Science
G216 Biological Sciences Centre, Telephone (403) 432-3290

To Zhi-Qing Chen

Department of Genetics
University of Alberta
Edmonton, Alberta
T6G 2E9

October 4, 1990

Dr. P. Moens
Department of Biology, York University
4700 Keele Street
North York, Ont.
M3J 1P3

Dear Dr. Moens:

I am writing to request written permission from you to use a paper entitled Chen et al, 1989, Cloning and characterization of a tandemly repeated DNA sequence in the crane family (Gruidae). Genome 32 (4): 646-654 as a part of my Ph. D. thesis. I would appreciate very much if you could FAX your permission before October 10th. I apologize for such short notice.

Sincerely yours,

Zhi-Qing Chen

You have my permission
Peter Moens

FAX 403 492 7033

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Zhi-Qing Chen

TITLE OF THESIS: Studies on two repetitive DNA families in the avian genome.

DEGREE: Doctor of Philosophy

YEAR THIS DEGREE GRANTED: 1990.

(SIGNED) .....

(Address) Department of Genetics
University of Alberta
Edmonton, Alberta, T6G 2E9

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled STUDIES ON TWO REPETITIVE DNA FAMILIES IN THE AVIAN GENOME, submitted by ZHI-QING CHEN in partial fulfillment of the requirements for the degree of DOCTOR OF PHILOSOPHY.

.....R.B.Hodgetts.....

Dr. R. B. Hodgetts (Supervisor)

.....C.C. Lin.....

Dr. C. C. Lin (Supervisor)

.....

Dr. J. B. Bell

.....John Locke.....

Dr. J. Locke

.....

Dr. A. Maresca

Date: October 5, 1990

ABSTRACT

This thesis consists of studies on two avian DNA families. The first study (Chapter II) reports the characterization of a tandemly repeated DNA sequence in several species of the crane family. This "*Pst*I family" is comprised of at least 8800 monomer units 187 bp in length and constitutes 0.14% of the genome of the sarus crane (*Grus antigone*). The Pst array is located mainly in the centromeric heterochromatin of chromosome 2 in the two species where *in situ* hybridizations of a cloned monomer to metaphase chromosome spreads were carried out. DNA sequence comparisons between five monomer units from *G. antigone* revealed a high degree of homology between four of the individual repeats, while the fifth was somewhat divergent. The G+C content deduced from the DNA sequence makes it likely that the Pst family constitutes part of a density satellite seen in profiles of crane DNA centrifuged to equilibrium in CsCl. The common occurrence of tandem arrays such as the *Pst*I family, with repeat lengths close to 200 bp, leads us to an hypothesis implicating nucleosomes in the evolution of such families.

The second study (Chapter III) reports the identification of a short interspersed repetitive DNA element (SINE), called CR1, in representatives of nine avian orders which comprise a wide spectrum of the class Aves. CR1 was first identified in chickens; however, unlike virtually all

other SINEs whose distribution is species limited, CR1 is probably a ubiquitous component of the avian genome. This was ascertained using the polymerase chain reaction (PCR), which revealed interspecific homologies not detected by conventional Southern analysis. DNA sequence comparisons between a CR1 element isolated from a sarus crane and those isolated from an emu showed that two short highly conserved regions were present. These coincide with two previously characterized regions of the chicken CR1. One of these behaves as a transcriptional silencer and the other is a binding site for a nuclear protein. Our observations suggest clearly that CR1 has evolved under functional constraints and that SINEs as a whole may constitute a more significant component of the eukaryotic genome than is generally acknowledged.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Dr. R. B. Hodgetts and Dr. C. C. Lin for their guidance, enthusiastic support and encouragement throughout the course of these studies. I thank other committee members Drs. J. Bell, J. Locke and A. Maresca for their helpful suggestions; Dr. C. Strobeck, Dr. D. Boag, Dr. M. Russell and Dr. B. Addison for their interest in these studies and for their expert advice. I would like to extend my appreciation to Dr. J. Kuspira for his advice on preparation for the candidacy exam, Dr. F Nargang for his tolerance to my playing on his computer, Gary Ritzel for his helpful advice on double stranded DNA sequencing and the PCR. A boatload of thanks to Bill Clark for what words alone can not do it justice. Thanks the following folks for their friendship and help and in addition: Rahul for the tape of "Adam's song", Dianne for horse-riding, Tim for his guitar music, Millan for arranging a Lab golf tournament, Jim and Audrey for organizing camping trips, and numerous others in the Department for their help. I also would like to thank Tomomi for her excellent cytogenetic techniques, Dr. Sasi for his DNA samples, and the ladies in the general office for their daily greeting.

Very special thanks to my wife Xiuhong for her persistent love and support, and my two sons, Tao and Chen, for bringing me hope and joy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I. INTRODUCTION

Early Studies

Despite the fact that most of the protein-coding DNA sequences are single copy, that is, present only once in a haploid genome, a significant portion of the higher eukaryotic genome consists of various redundant DNA sequences. These include repetitive gene families and repetitive DNA families (DFs). Repetitive gene families have highly conserved DNA sequences which encode well defined products such as the histones, the actins, and rRNA. By contrast, the DFs are those repetitive DNA sequences which do not encode specific functional products. In this thesis, only DFs are considered.

Discovery of the existence of repetitive DNA sequences in the higher eukaryotic genome stemmed from an unexpected observation by Kit (1961). In his cesium chloride gradient centrifugation study, Kit noticed that high molecular weight DNA prepared from various mouse tissues formed small discrete (satellite) bands in addition to the expected principle one. The highly repetitious nature of these satellites was not revealed until a DNA reassociation experiment was performed, which showed that the DNA contained in these bands reannealed much more rapidly than the DNA contained in the main band and in fact, more rapidly than DNA prepared from bacterial cells (Waring and Britten, 1966). The first systematic study on

repetitive DNA should be credited to Britten and Kohne (1968) who, from a series of reassociation kinetic experiments on DNA prepared from various species, deduced the following: repetitive DNA is probably an ubiquitous portion of the higher eukaryotic genome; it consists of various DFs with heterogeneous members in each family; and it is dispersed throughout the genome (Britten and Kohn, 1968). Development of the chromosomal *in situ* hybridization technique made it possible to survey the deposition patterns of DFs in the genome. For example, the mouse satellite DNA was found to be clustered mostly in the centromeric regions of chromosomes (Pardue and Gall, 1969). All of the initial studies on DFs were performed on mixtures of repetitive DNA fragments. With the advent of recombinant DNA technologies, numerous repetitive DNA families in various species were cloned and characterized.

Definitions

As these data accumulated, it was realized that DFs as a class are so heterogeneous that their only common features are their repetition and their lack of a coding function. Therefore, various terminologies have evolved to describe the different classes of DF.

The analysis of reassociation profiles of higher eukaryotic genomic DNA revealed two distinguishable repetitive fractions: a highly repetitive class with a C_0t

value of 10^{-2} to 10^{-4}, and a middle repetitive class with a $C_o t$ value of 0.2 to 100. The terms "highly repetitive DF" and "moderately repetitive DF" were introduced by Britten and Davidson (1971), to describe these two classes. In the literature, the term highly repeated DF is often used interchangeably with satellite DNA, while the term moderate, or middle, repetitive DF is also referred to as interspersed DNA. Sometimes these terms can be misleading, since there is no real correlation between family size of a DF and its deposition pattern. One obvious example is the Alu family in humans (Houch et. al., 1979) and other primates' (Grimaldi et. al., 1981) genomes. It has over 10^6 members in a haploid complement and is the most redundant DF known, but its family members generally are dispersed throughout the genome.

Lengths of repeat units have also been used for classification purposes (Singer, 1982). The Alu family, with a unit length of less than 500 bp, is representative of SINE (short interspersed DNA element), whereas the L1 family of primates, with a size of 6-7 kb for the intact members, has been referred to as typical LINE (long interspersed DNA element).

Tandemly arrayed and interspersed, or dispersed, DFs are another pair of commonly used terms, which describe different deposition patterns of DFs. Most tandemly arrayed DFs tend to cluster in heterochromatic regions. By contrast, there are apparently no such constraints on the location of interspersed DFs. However, a recent study combining high

resolution *in situ* hybridization with quantitative solid state imaging revealed that the two major interspersed DFs in primates alternatively punctuate chromosomes: the Alu family preferentially resides in G-banding negative regions whereas the L1 family tends to be located in G-banding positive regions (Korenberg and Rykowski, 1988). These results actually refined a previous study by Manuelidis and Ward (1984). It seems that the assumption of random deposition of the interspersed DFs needs some revision, at least in reference to a genomic scale. However, it would be premature to draw any general implications from these findings.

Clearly, none of these terms by itself is sufficient to describe a DF and probably each DNA family should be treated individually, just as single copy genes, each with their own identity, have been dealt with. However, in the discussion to follow, I shall group studies of DFs under the following two headings: tandemly arrayed repeats and interspersed repeats.

Tandemly arrayed repetitive DNA families

Tandemly arrayed DFs may be cloned from satellite bands obtained from isopycnic gradient centrifugation, or from gel electrophoresis of digested genomic DNA. In addition, they may also be cloned from reannealed DNA fragments collected on HAP columns following re-association to a low $C_o t$ value. The physical features of these clones can be ascertained either

by restriction or DNA sequence analysis. Their copy number in the genome can be quantified by various DNA-DNA hybridization methods. Such DFs have been found in various species. Table I-1 presents the DFs whose characteristics were revealed through analysis of cloned elements from each DF rather than by the analysis of a mixture of DNA fragments in satellite bands.

In the eukaryotic genome there also exist "simple repeats" (Tautz and Renz, 1984) and "minisatellites" (Jeffreys et al., 1985). Simple repeats are stretches of DNA consisting of one or a few tandemly repeated nucleotides, such as poly (dC-dA), common in eukaryotic genomes. Minisatellites are short (about 10 bp) regions forming part of the tandemly arrayed repeating unit in hypervariable DNA clusters.

Why and how these DFs were generated and maintained are the essential theoretical concerns in this field. Tandemly arrayed DFs are found mostly within centromeric and telomeric heterochromatin of chromosomes (Hilliker et. al., 1980; Singer, 1982). One explanation is that the recombination rate is low in centromeric heterochromatic regions so that DFs there might persist longer than those located elsewhere. This would increase their probability of being discovered there by researchers (Charlesworth et. al., 1986; Walsh, 1987). Alternatively, others view it as evidence that these DFs have evolved under certain selective pressures.

Table I-1. Representitive studies of tandemly arrayed DNA families

Organism	DNA family designation	Approximate Length	Reference
Mouse		234	Horz and Altenberger, 1981
Rat	Satellite I	370	Pech et al., 1979a
Root rat	I' satellite	185	Witney and Furrano, 1983
Calf	Bovine satellite	23, 46, 1400	Pech et al., 1979b
Monkey[1]	alpha satellite	172	Zhang et al., 1983
Marmoset	Cgo A, Cgo B	338, 916	Fanning et al., 1989
Human	alphoid satellite	172	Mitchell et al., 1985
Muntjac		870	Yu et al., 1986
Kangaroo rat	HS-a satellite	6	Fry and Salser, 1977
		3.3 kb	Keim and Lark, 1987
Sheep	II satellite	176, 235	Maio et al., 1977
Guinea pig	III satellite	215	Altenberger et al., 1977

Table I-1. Continue

Whale			Arnason and Widegren, 1989.
Falcon		174	Longmire et al., 1988
Snake	Satellite III & IV	varies	Singh et al., 1980; Epplen et al., 1982
Newt	Sat G	310, 330	Barsacchi-Pilone et al., 1986
	Satellite I	222	Diaz et al., 1981
Frog		741	Lam and Carroll, 1983
Carp		245	Datta et al., 1988
Arabidopsis		180	Martinez-Zapater et al., 1986
Maize	Knob satellite	185	Peacock et al., 1981
Mustard		172	Capesius, 1983
Radish		177	Grellet et al., 1986
Rye	350-family	350	Xin and Appels, 1988

1. Africa green monkey

By comparison to the first question, the explanations of how the numerous copies of a DF may have been generated in the genome seem less controversial. A detailed analysis of a mouse satellite DNA family (Southern, 1975) revealed that its 234 bp monomer unit consists of repeats of a 9 bp basic sequence and its variants. Later, this observation was explained by a saltatory replication model (Fry and Salser, 1977) which suggested that DNA families were generated by sudden amplification of certain templates randomly chosen from a set of candidate DNA sequences in the genome. These events must have been relatively recent so that the resulting DFs are usually conserved in closely related species.

Smith (1973, 1976) proposed a different model suggesting that random unequal crossing over was a major mechanism for the generation of tandem arrays. Such crossing over could occur between two sister chromatids and/or between two non-sister chromatids of a pair of homologs in the germ line. By computer simulation, Smith's model was proven plausible mathematically. This model was revised by Maio et al. (1977), who postulated that such crossing over events would be restricted by the subunit structure of constitutive heterochromatin, occurring at the spacer regions between nucleosomes.

Recently, Walsh (1987) pointed out that crossing over by itself is not sufficient to expand an array significantly, since intramolecular crossing over, a counteracting event, would occur with the same frequency as the intermolecular

one. Alternatively, he proposed that more likely causes to account for the generation of tandem arrays were gene amplification-like processes such as unproportional DNA replication (Schimke 1984) and rolling circle replication (Hourcade et. al., 1973; Flavell 1982).

The models described above all suggest that tandem repeats can arise without selection ; on the other hand, some tandemly repeated gene families can arise under conditions of high selective pressure, such as in the amplification of dihydrofolate reductase genes (Kaufman et. al., 1979) and multi drug resistant genes (Roninson et al., 1984).

To elucidate the evolution of the tandemly arrayed DFs, investigators have recently started studying the hierarchical structure of satellite DNA (Willard and Waye, 1987). It has been found that a "high-order repeat unit", in addition to monomer units, comprises the basic building block of the arrays. For example, alpha satellite in the human X chromosome is characterized by a higher order repeat unit of 2 kb in size, containing twelve 171 bp monomer units and spanning about 500 to 900 kb (Waye and Willard, 1985), whereas the one in the human Y chromosome has 5.5 kb repeating units, each consisting of 32 monomers (Wolfe et al., 1985; Tyler-Smith and Brown, 1987).

To date, no function has been found for tandemly repeated DFs, and most believe that perpetuation of these DFs in eukaryotic genomes does not necessarily depend on their sequences having a structural or functional role. However, a

function for certain DFs is implied by a number of observations: specific proteins that bind to certain satellite sequences have been reported (Straus and Varshavsky, 1984; Linxweiler and Horz, 1985), and there exists a coincidence between the size of repeat units found in a large number of satellite DNAs and the length of mononucleosome DNA (Martinez-Zapater et. al., 1986). Secondly, despite the fact that most of the tandem repeats are confined to closely related species, conservation of some satellite sequences in distantly related species has been documented (Brutlag, 1980; Grellet et. al., 1986). Cooke et. al., (1982), using a specific cloned DNA fragment of a human satellite as a hybridization probe, showed that these sequences are highly methylated in somatic tissues but selectively unmethylated in the germ line. This may suggest a germ line function.

Interspersed repetitive DNA families

Comprehensive studies on interspersed DFs have focused on the genomes of Drosophila and primates, one for its convenience and the other for its relevance to humans, and very different pictures have been obtained. There are about 50 or more dispersed DFs in the genome of *Drosophila melanogaster*, each consisting of between 10 and 100 members (Spradling and Rubin, 1981). In human and other primate genomes, the best defined dispersed DFs are Alu and L1. Each

of them consists of about 10^6 and 10^4 copies respectively, accounting for one third of the total repetitive component of human genome (Sun et al., 1984; Hwu et al., 1986). Characteristics of DFs in other organisms have also been reported. Table I-2 lists some representative clones.

Hypotheses of why the interspersed DFs were generated are speculative and controversial. On one hand, the DFs are believed to bestow upon the genomes a better fitness; therefore they are the inevitable result of genome evolution. As the minimum C-values of many metazoan groups increase with developmental complexity, generation of DFs was suggested to be a reasonable solution for increasing cellular volumes (Britten and Davidson, 1969; Cavalier-Smith, 1978).

Noticing that the copy number of B1, an interspersed distributed Alu-equivalent DNA family of mice (Krayev et. al., 1980; Pan et. al., 1981), is equivalent to the number of replication origins in the genome, Georgiev et. al. (1981) speculated that this DF may be involved in DNA replication. Furthermore, their experimental results showed that the B1 family member indeed can function as a replication origin *in vivo*.

Britten and Davidson's model (1969; 1971), which suggests regulation of gene expression as a function of interspersed DFs, had the most influence on the attitude of subsequent researchers on the study of DFs.

On the other hand, DFs have been described as selfish entities (Doolittle and Sapienza, 1980; Orgel and Crick,

Table I-2. Representitive studies of interspersed DNA families

Organism	DNA family designation	Approximate Length	Family size	Reference
Rabbit	C family	300	$1.7x10^5$	Cheng et al., 1984
Goat		300		Schon et al., 1981
Cow		120	10^5	Watanable et al., 1982
Mouse	L1md	>5Kb	$2x10^4$	Menunier-Rotival et al., 1982.
	B1	130	10^5	Krayev et al., 1980
	B2	190	10^5	Krayev et al., 1982
Frog	1723 element	6-10 kb	$8.5x10^3$	Kay and Dawid, 1983
		77-79	10^5	Spohr et al., 1981

1980) not fulfilling any function in the genome. Their arguments were based mainly on the available data showing a high degree of sequence variation and poor cross-species conservation to be associated with almost all known DFs.

Regarding the proliferation of the dispersed DFs, most people believe that the DFs are generated by some kind of transpositional mechanism (Bouchard 1982; Rogers, 1985; Doolittle 1985). In fact, the majority of the interspersed DFs in the genomes of Drosophila are transposable elements (Spradling and Rubin, 1981). Chromosomal location and copy number of a given transposable element are under close genetic control specified by the interaction between the DF and the genome (Ananiev et al., 1984), and the transposition mechanism of a few of these DFs such as the Copia-like families (Scherer et. al., 1982) and P elements (Rubin et. al., 1982) has been determined. Special structural features of some other DFs also provided evidence substantiating the hypothesis of a transposon-like origin of the interspersed DFs. During the last years, some reports in literature have shown that the Alu and L1 families in the primates are now actively undergoing transposition (Lin et al., 1988; Di Nocera and Sakaki, 1990; Kazakian et al., 1988). A typical Alu element has a polyadenylated tail at its 3' end and is flanked by a direct repeat. These interesting physical featurs are now accorded a special name, retroposon (Rogers, 1985). L1 elements also have poly $[A]^+$ tails at their 3' ends, but most of them have truncated 5' ends. By

comparison, CR1 family members in the chicken genome have a conserved 3' end instead of a poly $[A]^+$ tail and truncated 5' ends. RNA intermediates have been speculated to be involved in transposition of these elements. However, not every interspersed DF has such a structure, and further studies are required to elucidate the mechanism of their origin.

DNA families in the avian genome

In the avian genome, only some preliminary data on DFs have been accumulated. Rosen et. al., (1973) estimated that the total fraction of repeated sequences represents 30% of the chicken genome. Biederman (1981) studied the buoyant density profiles of twelve species from four avian orders, and found heavy satellite bands in all cases. These results suggested the ubiquity of tandemly arrayed DNA components in the avian genome; however, quantitative studies were not reported. *In situ* hybridization studies of satellite DNA sequences isolated from the Japanese quail genome showed that they are mainly located on microchromosomes (Comings and Mattoccia, 1970; Brown and Jones, 1972). Subsequently, sex chromosome specific satellite DNAs were cloned (Tone et al., 1982; Quinn et. al., personal communication). Very recently, a study of a tandemly arrayed DF characterized in the falcon genome indicated a potential use of the DF as a polymorphic marker (Longmire et. al., 1988).

An electron microscopic study on reannealed duplex repeats suggested that about 40% of the chicken genome contains an interspersed repeat component (Eden and Hendrick 1978), and the total of the interspersed repeats was estimated as 11-13% of the chicken genome (Eden and Hendrick, 1978; Arthur and Straus, 1978). By analysis of C_0t curves, Epplen et. al., (1979) reported that interspersed repeats in duck, chicken and pigeon genomes were of sizes averaging more than 1500 bp, which was consistent with a conclusion drawn from an independent study on the chicken genome (Arthur and Straus, 1978). Conversly, the best characterized (and perhaps the only cloned) dispersed DF from the chicken genome, CR1, has 20000 members of less than 1 kb in length (Stumph et. al., 1981; Hache and Deeley, 1988).

Applications of DFs to broader biological problems

Studies on DFs have enriched not only our understanding of the evolution and the organization of DFs *per se*, but also our knowledge of other biological phenomena.

The study of muntjac chromosome karyotype evolution represents an excellent example. The Indian muntjac (*M. Muntjak. vaginalis*) has 6 or 7 (depending on sex) chromosomes in each somatic cell, whereas the Chinese muntjac (*M. reevesi*) has a diploid chromosome number of 46. Despite this, the two are closely related based not only on their morphological similarities, but also on their ability to

produce viable hybrids (Shi et al., 1980). It was speculated that the giant chromosomes in Indian muntjac cells were formed by tandom fusions of small chromosomes in its ancestral species. However, direct evidence in support of this has only recently been provided. I sequenced a repetitive DNA family member isolated from a cell line derived from a Chinese muntjac (Lin et. al., personal communication), and found a high homology of the clone with a DF previously characterized in Indian muntjacs (Yu et. al., 1986). Using *in situ* hybridization to locate this DF on chromosomes of both muntjac species, we discovered that in Chinese muntjacs the DF was clustered on centromeric heterochromatin of all the chromosomes, but in Indian muntjacs the DF was clustered on specific interstitial regions in addition to the centromeric heterochromatic regions (Lin et. al., personal communication). Therefore, this study provided clear positive evidence substantiating the tandem fusion hypothesis.

Since the inception of chromosomal banding techniques (Casperson et al., 1968), cytogenetic analysis of the complex genomes of vertebrates has accelerated a great deal (Sumner, 1982), yet the biochemical basis of these staining procedures was poorly understood. By cloning and sequencing repetitive DNA fragments that reside in centromeric heterochromatin, it has been possible to look into the issue. For example, certain C-band regions (centromeric heterochromotin) in some mammalian species can be differentiated further by

fluorescent dyes which are base specific with regards to G-C content of these DNA domains (Lin et al., 1980). However, these data are limited and further studies are required to test the binding specificity of the various fluorescent dyes.

Since *in situ* hybridization with single copy DNA sequences is difficult (Harper and Saunders, 1981), a series of probes has been developed from repetitive DNA sequences, and they turned out to be very useful tools for distinguishing individual human chromosomes and chromosome regions (Moyzis et. al., 1987). Since such probes are also species specific, they have been applied to distinguish, for example, human chromosomes from rodent ones in somatic hybrid cell lines (Durnam et. al., 1985).

It has been suggested that DFs can be very useful for the study of speciation (Funderburk et. al., 1987). Traditional approaches to establish phylogenic relationships between species involve morphological comparisons. Because of the difficulty in distinguishing homology from convergency, fossil and anatomic data can be very inconclusive (Cracraft, 1982). Alternative data sources have been utilized, such as comparative studies of protein (egg white) profiles (Sibley, 1960) and interspecific DNA-DNA association kinetics (Sibley et al., 1988). However, no satisfactory conclusions have been made from these attempts. Recently, some studies have used repetitive DNA sequences as a data base to approach phylogenic problems (Maeda et. al., 1988), and the results were promising. However, one of the

factors preventing a large scale taxonomic survey by this approach is the limited number of highly conserved repetitive DNA families.

Scope of this thesis

This study was initiated for several reasons. First, although numerous studies on DFs had been performed on a variety of species, information on avian species was very limited. Theoretical concerns of the evolutionary processes affecting DFs, such as whether DFs evolve under functional constraints, and how interspecific and intraspecific sequence homology is maintained amongst DF members, need to be addressed . Restricted by a poor fossil record, a large number of questions about avian taxomony are to be solved. For population studies and the management of endangered avian species such as the whooping crane, easily detectable and highly variable genetic markers would be an asset. Therefore, we set out to clone and characterize some DFs from the sarus crane (*Gruidae antigone*), and to survey the distribution of these DFs among avian genomes, in an attempt to address questions concerning DF evolution. In addition, we had hoped that some of the DFs could be developed into a useful system for individual bird genotyping. In this thesis two studies were undertaken. In the first, the cloning and characterization of a tandemly repetitive DNA family in the cranes is described and in the second the cloning and

characterization of an interspersed repetitive DNA family in avian species is presented.

REFERENCES

Altenburger, W., Horz, W., and Zachau, H.G. 1977. Comparative analysis of three guinea pig satellite DNAs. Eur. J. Biochem. 73: 393-400.

Ananiev E. V., Barsky V. E., Ilyin Yu V. and Ryzic M. V. 1984. The arrangement of transposable elements in the polytene chromosomes of *Drosophila melanogaster*. Chromosoma 90: 366-377.

Arnason U. and Widegren B. 1989. Composition and chromosomal location of cetacean highly repetitive DNA with special reference to the blue whale, *Balaenoptera musculus*. Chromosoma 98: 323-329.

Arthur R. R. and Straus N. A. 1978. DNA-sequence organization in the genome of the domestic chicken (Gallus domesticus). Can. J. Biochem. 56: 257-263.

Barsacchi-Pilone, G., Batistoni, R., Andronico, F., Vitelli, L., and Nardi, I. 1986. Heterochromatic DNA in *Triturus* (Amphibia, Urodela). Chromosoma 93: 435-446.

Biederman B. M. 1981. A comparison of avian genomes. Ph. D. thesis. The University of Calgary.

Britten R. J. and Davidson E. H. 1969. Gene regulation for higher cells: a theory. Science 165: 349-357.

Britten R. J. and Davidson E. H. 1971. Repetitive and nonrepetitive DNA sequences and a speculation on the origin of evolutionary novelty. Quart. Rev. Biol. 46: 111-137.

Britten R. J. and Kohne D. E. 1968. Repeated sequences in DNA. Science 161: 529-540.

Brutlag D. L., 1980. Molecular arrangement and evolution of heterochromatin DNA. Anne. Rev. Genet. 14: 121-144.

Bouchard R. A. 1982. Moderately repetitive DNA in evolution. pp. 113-194. In Int. Natl. Rev. Cytol. Vol. 76.

Brown, J.E., and Jones, K.W. 1972. Localisation of satellite DNA in the microchromosomes of the Japanese quail by <u>in situ</u> hybridization. Chromosoma 38: 313-318.

Capesius, I. 1983. Sequence of the cryptic satellite DNA from the plant *Sinapis alba*. Biochim. Biophys. Acta. 739: 276-280.

Casperson T., Farber S., Foley G. E., Kudynowski J., Modest E. J., Semonsson E., Wagh V. and Neck L. 1968. Chemical differentiation along metaphase chromosomes. Exp. Cell Res. 49: 219-222.

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.

Charlesworth B., Langely C. H. and Stephan W. 1986. The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics 112: 947-962.

Cheng J.-F., Printz R., Callagham T., Shuey D. and Hardison R. C. 1984. The rabbit C family of short, interspersed repeats nucleotide sequence determination and transcriptional analysis. J. Mol. Biol. 176: 1-20.

Comings, D.E., and Mattoccia, E. 1970. Studies of microchromosomes and G-C rich DNA satellite in the quail. Chromosoma 30: 202-214.

Cooke H. J., Schmidtke J. and Gosden J. R. 1982. Characterisation of a human Y chromosome repeated sequence and related sequences in higher primates. Chromosoma 87: 491-502.

Cracraft J. 1982. Phylogenetic relationships and monophyly of loons, grebes, and hesperornithiform birds, with comments on the early history of birds. Syst. Zool. 31: 35-56.

Di Nocera P. P. and Sakaki Y. 1990. Lines: a superfamily of retrotransposable ubiquitous DNA elements. Trends in Genet. 6: 29-30.

Diaz M. O., Barsacchi-Pilone G., Mahon K. A. and Gall J. G. 1981 Transcript from both strands of a satellite DNA occur on lampbrush chromosome loops of the newt *Notophthalmus*. Cell 24: 649-659.

Datta, U., Dutta, P., and Mandal, R.K. 1988. Cloning and characterization of a highly repetitive nucleotide sequence. Gene 62: 331-336.

Durnam D. M., Gelinas R. E. and Myerson D. 1985. Detection of species specific chromosomes in somatic cell hybrids. Somatic Cell. Mol. Genet. 11: 571-577.

Doolittle W. F. 1985. The evolutionary significance of middle repetitive DNAs. In Cavalier-Smith (Ed), The

evolution of genome size. John Wiley and Sons Ltd. pp 443-487

Doolittle W. F. and Sapienza C. 1980. Selfish genes, the phenotype paradigm and gene evolution. Nature 284: 601-603.

Eden F. C. and Hendrick J. P. 1978. Unusual organization of DNA sequences in the chicken. Biochemistry 17: 5838-5844.

Epplen J. T., Diedrich U., Wagenmann M., Schmidtke J. and Engel W. 1979. Contrasting DNA sequence organization patterns in sauropsidian genomes. Chromosoma 75: 199-214.

Epplen J. T., McCarrey J. R., Sutou S. and Ohno S. 1982. Base sequence of a cloned snake w-chromosome DNA fragment and identification of a male-specific putative mRNA in the mouse. Proc. Natl. Acad. Sci. USA. 79: 3798-3802.

Fanning T. G., Seuanez H. N. and Forman L. 1989. Satellite DNA sequences in the neotropical marmoset *Callimico goeldii* (Primates, Platyrrhini). Chromosoma 98: 396-401.

Flavell R. B., 1982. Sequence amplification, deletion and rearrangement: major sources of variation during species divergence. pp. 301-323. In Genome Evolution, Edited by Dover G. A. and Flavell R. B. Academic Press, New York.

Fry K. and Salser W. 1977. Nucleotide sequences of HS-a satellite DNA from kangaroo rat Dipodomys ordii and characterization of similar sequences in other rodents. Cell 12: 1069-1089.

Funderburk S. J., Klisak I., Law M. L., Ma N., Neiswanger K. and Sparkes R. S. 1987. Primate evolution of a dispersed human repetitive DNA sequence. Chromosoma 95: 148-153.

Georgiev G. P., Ilyin Y. V., Chmeliauskaite V. G., Ryskov A. P., Kramerov D. A., Skryabin K. G., Krayev A. S., Lukanidin E. M. and Grigoryan M. S. 1981. Mobile dispersed genetic elements and other middle repetitive DNA sequences in the genomes of Drosophila and mouse; transcription and biological significance. Cold Spring Harbor Symp. Quant. Biol. 45: 641-654.

Grellet, F., Delcasso, D., Panabieres, F., and Delseny, M. 1986. Organization and evolution of a higher plant alphoid-like satellite DNA sequence. J. Mol. Biol. 187: 495-507.

Grimaldi G., Queen C. and Singer M. F. 1981. Interspersed repeated sequences in the African green monkey genome that are homologous to the human Alu family. Nucleic Acids Res. 9: 5553-5568.

Hache R. J. G. and Deeley R. G. 1988. Organization, sequencing and nuclease hypersensitivity of repetitive elements flanking the chicken apo VLDII gene: extended sequence similarity to elements flanking the chicken vitellogenin gene. Nucleic Acids Res. 16: 97-113.

Harper and Saunders, 1981. Localization of single copy DNA sequences on G banded human chromosomes by in situ hybridization. Chromosoma 83: 431-439.

Hilliker A. J., Appels R. and Schalet A. 1980. The genetic analysis of D. melanogaster heterochromatin. Cell 21: 607-619.

Horz, W., and Altenburger, W. 1981. Nucleotide sequence of mouse satellite DNA. Nucleic Acids Res. 9: 683-696.

Houch C. M., Rinehart F. P. and Schmid C. W. 1979. A ubiquitous family of repeated DNA sequences in the human genome. J. Mol. Biol. 132: 289-306.

Hourcade D., Dressler D. and Wolfson J. 1973. The amplification of ribosomal RNA genes involving a rolling circle intermediate. Proc. Natl. Acad. Sci. USA 70: 2926-2930.

Hwu H. R., Roberts J. W., Davidson E. H. and Britten R. J. 1986. Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. PROC. NATL. ACAD. SCI. USA. 83: 3875-3879.

Jeffreys A. J., Wilson V. and Thein S. L. 1985. Hypervariable 'minisatellite' regions in human DNA. Nature 314: 67-73.

Kaufman R. J., Brown P. C. and Schimke R. T. 1979. Amplified dihydrofolate reductase genes in unstablly methotrexate-resistant cells are associated with double minute chromosomes. PROC. NATL. ACAD. SCI. USA. 76: 5669-5673.

Kay B. K. and Dawid I. B. 1983. The 1723 element: a long, homogeneous, highly repeated DNA unit interspersed in the genome *Xenopus laevis*. J. Mol. Biol. 170: 583-596.

Kazazian Jr H. H., Wong C., Youssoufian H., Scott A. F., Phillips D. G. and Antonarakis S. E. 1988. Haemophilia A resulting from *de novo* insertion of L1 sequences

represents a novel mechanism for mutation in man. Nature 332: 164-166.

Keim P. and Lark K. G. 1987. The sequence of the 3.3-kilobase repetitive element from *Dipodomys ordii* suggests a mechanism for its amplification and interspersion. J. Mol. Evol. 25: 65-73.

Kit S. 1961. Equilibrium centrifugation in density gradients of DNA preparations from animal tissues. J. Mol. Biol. 3: 711-716.

Korenberg J. R. and Rykowski M. C. 1988. Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 53:391-400.

Krayev A. S., Kramerov D. A., Skryabin K. G., Ryskov A. P., Bayev A. A. and Georgiev G. P. 1980. The nucleotide sequence of the ubiquitous repetitive DNA sequence B_1 complementary to the most abundant class of mouse fold-back RNA. Nucleic Acids Res. 8: 1201-1215.

Krayev A. S., Markusheva T. V., Kramenov D. A., Ryskov A. P., Skryabin K. G., Bayev A. A. and Georgiev C P. 1982. Ubiquitous transposon-like repeats B1 and B2 of the mouse genome: B2 sequencing. Nucleic Acids Res. 10: 7461-7475.

Lam B. S. and Carroll D. 1983. Tandemly repeated DNA sequences from Xenopus laevis I. Studies on sequence organization and variation in satellite 1 DNA (741 base-pair repeat). J. Mol. Biol. 165: 567-585.

Lin C. C., Jorgenson K. F. and van de Sande J. H. 1980. Specific fluorescent bands on chromosomes prduced by acridine orange after prestaining with base specific non-fluorescent DNA ligands. Chromosoma 79: 271-286.

Lin C. S., Goldthwait D. A. and Samols D. 1988. Identification of Alu transposition in human lung carcinoma cells. Cell 54: 153-159.

Linxweiler, W., and Horz, W. 1985. Reconstitution experiments show that sequence-specific histone-DNA interactions are the basis for nucleosome phasing on mouse satellite DNA. Cell 42: 281-290.

Longmire J. L., Lewis A. K., Brown N. C., Buckingham J. M., Clark L. M., Jones M. D., Meincke L. J., Meyne J., Ratliff R. L., Ray F. A., Wagner R. P. and Moyzis R. K. 1988. Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family falconidae. Genomics 2: 14-24.

Maeda N., Wu C. I., Bliska J. and Reneke J. 1988. Molecular evolution of intergenic DNA in higher primates: pattern

of DNA changes, molecular clock, and evolution of repetitive sequences. Mol. Biol. Evol. 5: 1-20.

Maio J.J., Brown F. L. and Musich P. R. 1977. Subunit structure of chromatin and the organization of eukaryotic highly repetitive DNA: recurrent periodicities and models for the evolutionary origins of repetitious DNA. J. Mol. Biol. 117: 637-656.

Manuelidis L. and Ward D. C. 1984. Chromosomal and nuclear distribution of the HindIII 1.9 kb human DNA repeat segment. Chromosoma 91: 28-38.

Martinez-Zapater J.M., Estelle M.A., and Somerville C. R. 1986. A highly repeated DNA sequence in Arabidopsis thaliana. Mol. Gen. Genet. 204: 417-423.

Menunier-Rotival M. Soriano P., Cuny G., Strauss F. and Bernardi G. 1982. Sequence organization and genetic distribution of the major family of interspersed repeats of mouse DNA. PROC. NATL. ACAD. SCI. USA. 79: 255-259.

Mitchell, A.R., Gosden, J.R., and Miller, D.A. 1985. A cloned sequence, p82H, of the alphoid repeated DNA family found at the centromeres of all human chromosomes. Chromosoma 92: 369-377.

Moyzis R. K., Albright K. L., Bartholdi M. F., Cram L. S., Deaven L. L., Hildebrand C. E., Joste N. E., Longmire J. L., Meyne J. and Schwarzacher-Robinson T. 1987. Human chromosome-specific repetitive DNA sequences: Novel markers for genetic analysis. Chromosoma 95: 375-386.

Orgel L. E. and Crick F. H. C. (1980) Selfish DNA: the ultimate parasite. Nature 284: 604-607.

Pan J., Elder J. T., Duncan C. H. and Weissman S. M. 1981. Structural analysis of interspersed repetitive polymerase III transcription units in human DNA. Nucleic Acids Res. 9: 1151-1170.

Pardue M. L. and Gall J. G. 1969. Molecular hybridization of radioactive DNA to the DNA of cytological preparations. Proc. Nat. Acad. Sci. USA 64: 600-604.

Peacock, W.J., Dennis, E.S., Rhoades, M.M., and Pryor, A.J. 1981. Highly repeated DNA sequence limited to knob heterochromatin in maize. Proc. Nat. Acad. Sci. USA 78: 4490-4494.

Pech M., Igo-Kemenes T. and Zachau H. G. 1979a. Nucleotide sequence of a highly repetitive component of rat DNA. Nucleic Acids Res. 7: 417-432.

Pech M., Streek R. E. and Zachau H. G. 1979b. Patchwork structure of a bovine satellite DNA. Cell 18: 883-893.

Rogers J. H. 1985. Origin and evolution of retroposons. Int. Rev. Cytol. 93: 231-279.

Roninson I. B., Abelson H. T., Housman D. E., Howell N. and Varshavsky A. 1984. Amplification of specific DNA sequences correlates with multi-drug resistance in CHinese hamster cells. Nature 309: 626-628.

Rosen J. M., Liarakos C. D. and O'Malley B. W. 1973. Effect of estrogen on gene expression in the chicken oviduct. I. renaturation studies. Biochemistry 12: 2803-2809.

Rubin G. M., Kidwell M. G. and Bingham P. M. 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell 29: 987-994.

Scherer G., Tschudi C., Perera J., Delius H. and Pirrotta V. 1982. *B104*, a new dispersed repeated gene family in *Drosophila melanogaster* and its analogies with retroviruses. J. Mol. Biol. 157: 435-451.

Schimke R. T. 1984. Gene amplification in cultured animal cells. Cell 37: 705-713.

Schon E. A., Cleary M. L., Haynes J. R. and Lingrel J. B. 1981. Structure and evolution of goatγ-, β^C- and β^A-globin genes: three developmentally regulated genes contain inserted elements. Cell 27: 359-369.

Shi L., Ye Y. and Duan X. 1980. Comparative cytogenetic studies on the red muntjac, Chinese muntjac, and their F_1 hybrids. Cytogenet. Cell Genet. 26: 22-27.

Sibley C. G. 1960. The electrophoretic patterns of avian egg white pattern as taxonomic characters. Ibis 102: 215-284.

Sibley C. G., Ahlquist J. E. and Monroe B. L. 1988. A classification of the living birds of the world based on DNA-DNA hybridization studies. The Auk 105: 409-423.

Singh.L., Purdom I. F. and Jones K. W. 1980. Conserved sex-chromosome- associated nucteotide sequences in eukaryotes. Cold Spring Harbor Symp. Quant. Biol. 45: 805-814.

Singer, M.F. 1982. Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76: 67-112.

Smith G. P. 1973. Unequal crossover of the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. 38: 507-513.

Smith G. P. 1976. Evolution of repeated DNA sequences by unequal crossovers. Science 191: 528-535.

Southern E. M. 1975. Long range periodicities in mouse satellite DNA. J. Mol. Biol. 94: 51-69.

Spohr G., Reith W. and Sures I. 1981. Organization and sequence analysis of a cluster of repetitive DNA elements from *Xenopus laevis*. J. Mol. Biol. 151: 573-592.

Spradling A. C. and Rubin G. M. 1981. Drosophila genome organization: conserved and dynamic aspects. Annu. Rev. Genetics 15: 219-264.

Strauss F. and Varshavsky A. 1984. A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. Cell 37: 889-901.

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

Sumner A. T. 1982. The nature and mechanisms of chromosome banding. Cancer Genet. Cytogenet. 6: 59-87.

Sun L., Paulson K. E., Schmid C. W., Kadyk L. and Leinwand L. 1984. Non-Alu family interspersed repeats in human DNA and their transcription. NUCLEIC ACIDS RES.. 12: 2669-2690.

Tautz D. and Renz M. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. NUCLEIC ACIDS RES.. 12: 4127-4138.

Tone, M., Nakano, N., Takao, E., Narisawa, S., and Mizuno, S. 1982. Demonstration of W chromosome-specific repetitive DNA sequences from the domestic fowl, Gallus g. domesticus. Chromosoma 86: 551-569.

Tyler-Smith C. and Brown W. R. A. 1987. Structure of the major block of alphoid satellite DNA on the human Y chromosome. J. Mol. Biol. 195: 457-470.

Walsh J. B. 1987. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. Genetics 115: 553-567.

Waring M. and Britten R. J. 1966. Nucleotide sequences repetition: a rapidly reassociating fraction of mouse DNA. Science 154: 791-794.

Watanabe Y., Tsukada T., Notake M., Nakanishi S. and Numa S. 1982. Structural analysis of repetitive DNA sequences in the bovine corticotropin-B-lipotropin precursor. Nucleic Acids Res. 10: 1459-1469.

Waye J. S. and Willard H. F. 1985. Chromosome-specific alpha sarellite DNA: nucleotide sequence analysis of the 2.0 kb repeat from the human X chromosome. Nucleic Acids Res. 13: 2731-2743.

Willard H. F. and Waye J. S. 1987. Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet. 3: 192-198.

Witney, F.R., and Furano, A.V. 1983. The independent evolution of two closely related satellite DNA elements in rats (*Rattus*). Nucl. Acids Res. 11: 291-304.

Wolfe J., Darling S. M., Erickson P. R., Craig I., Buckle V., Rigby P., Willard H. F. and Goodfellow P. N. 1985. Isolation and characterization of a alphoid centromeric repeat family from the humne Y chromosome. J. Mol. Biol. 182: 477-485.

Xin Z.-Y. and Appels R., 1988. Occurrence of rye (Secale cereale) 350-family DNA sequences in Agropyron and other Triticeae. Pl. Syst. Evol. 160: 65-76.

Yu, L.-C., Lowensteiner, D., Wong, E.F.-K., Sawada, I., Mazrimas, J., Schmid, C. 1986. Localization and characterization of recombinant DNA clones derived from the highly repetitive DNA sequences in the Indian muntjac cells: Their presence in the Chinese muntjac. Chromosoma 93: 521-528.

Zhang, X.Y., Fittler, F., and Horz, W. 1983. Eight different highly specific nucleosome phases on a-satellite DNA in the African green monkey. Nucleic Acids Res. 11: 4287-4306.

CHAPTER II

Cloning and characterization of a tandemly repeated DNA sequence in the crane family (*Gruidae*)

A version of this chapter has been published.
Chen Z.-Q., C. C. Lin and R. B. Hodgetts, 1989. Cloning and characterization of a tandemly repeated DNA sequence in the crane family (Gruidae). Genome 32: 646-654.

INTRODUCTION

Repetitive DNA sequences are very common and ubiquitous components of the higher eukaryotic genome. They can be found as single units, dispersed throughout the genome, or in tandemly repeated arrays. The latter are sometimes referred to as satellite DNA since they are often observed as mini bands on density gradients. If these sequences comprise more than 0.5% of the total genome, they may also be noted as distinct bands against the background smear after genomic DNA has undergone restriction digestion, electrophoretic fractionation and subsequent staining of the gel (Singer, 1982).

Because of the ease with which satellite sequences can be cloned and sequenced, they have become very useful in the study of multigene families. As a result, they have been characterized in a number of species of plants, invertebrates, fish, amphibians and mammals (Martinez-Zapater *et al.*, 1986; Miklos, 1982; Datta *et al.*, 1988; Barsacchi-Pilone *et al.*, 1986; Yu *et al.*, 1986; Maio *et al.*, 1981). We are interested in the characterization of satellite sequences in birds where little data have been accumulated and for some time now, our studies have been focused on the genome of the crane family (*Gruidae*). Biederman *et al.* (1982) have described two G+C rich satellites in the genome of the whooping crane, *Grus americana*, following isopycnic centrifugation in CsCl. Knowing that there was a significant

repetitive fraction in the crane genome, we set out to characterize it. In this study we describe the cloning of a repetitive DNA sequence from the sarus crane, *Grus antigone*. It differs from the two other satellites described for avian species to date in being confined to the centromeric heterochromatin of a single, large autosome. By contrast, the quail satellite DNA is found in the microchromosomes (Comings and Mattoccia, 1970; Brown and Jones, 1972) and the highly repeated sequences found in the chicken are W-chromosome specific (Tone *et al.*, 1982). Since many of the repeat units in the array we have studied possess a unique *Pst*I site, we refer to this repetitive fraction as the "*Pst*I family". Our characterization of this family prompts us to offer an hypothesis concerning a constraint imposed by nucleosomes on the evolution of multigene families.

MATERIALS AND METHODS

Specimens

Unless otherwise noted, all blood samples were provided from birds held in the Calgary Zoo by Dr. R.M. Cooper and Dr. M. Mainka, to whom we are indebted.

DNA extraction and molecular cloning

To isolate genomic DNA, 0.1 to 1 ml whole blood samples were adjusted to a final concentration of 0.85% NaCl. Triton X-100 was added to 1% and the nuclei were isolated by sucrose

gradient centrifugation as described by Wu *et al.* (1980). The purified nuclei were lysed at 50^0C for about one hour in a solution containing 100 mM Tris-HCl (pH 8.0), 40 mM EDTA, 0.4% SDS and 50 µg/ml proteinase K. After incubation, the lysis mixture was extracted with phenol, phenol/chloroform and chloroform, and high molecular weight DNA was then recovered by ethanol precipitation.

For molecular cloning of tandemly repeated elements, DNA from *Grus antigone* was digested to completion with the restriction enzyme *Pst*I, then fractionated on a 0.8% agarose gel. The monomer band was sliced out and the DNA was electroeluted into dialysis tubing (McDonnell *et al.*, 1977) and purified through an Elutip column (Schleicher and Schuell). The DNA was then ligated with the pUC19 plasmid which had been linearized by *Pst*I. The ligation mix was used to transform *E. coli* strain JM83. Recombinant clones were transferred to Gene Screen Plus membranes (Dupont), and those containing repetitive sequences were identified by the intensity of their signal after hybridization with ^{32}P labelled sarus crane genomic DNA. Rapidly prepared plasmid DNA (Ish-Horowicz and Burke, 1981) from randomly chosen clones was digested with several restriction enzymes to characterize the insert. Ligation, bacterial transformation, preparation of plasmid DNA and agarose gel electrophoresis were all performed as described by Maniatis *et al.* (1982), or with minor modifications thereof.

Southern and slot blot hybridization

For Southern blots, total genomic DNA was digested with restriction enzymes, fractionated on 0.8% agarose gels and transferred onto Gene Screen Plus membrane as suggested by the manufacturer. For slot blots, genomic DNA was digested with restriction enzymes, denatured in 0.25N NaOH, then diluted to a final concentration of 0.125N NaOH and 0.125X SSC. The DNA samples were drawn through a Gene Screen Plus membrane, under vacuum in a slot blot apparatus (Tyler Research).

The membrane was prehybridized in a sealed bag for at least one hour with a solution containing 50% formamide, 1% SDS, 1 M NaCl, 10% dextran sulphate and heat-denatured, sonicated salmon sperm DNA at 100 µg/ml. Hybridization was carried out at 42^0C for about 20 hours following the addition of probe (10^6 cpm/ml) labelled according to Feinberg and Vogelstein (1984) with ^{32}P-dCTP (3000 Ci/mmole; ICN Radiochemicals). Two washes of 30 minutes were carried out at 65^0C in 1% SDS and 2x SSC. The filters were exposed to X-ray film (Kodak XAR 5) at -70^0C for various times.

DNA sequencing

Double stranded DNA of the cloned plasmids was sequenced using ^{35}S-deoxythioadenosine triphosphate (600 Ci/mmole; Amersham) by the chain termination method of Sanger *et al.* (1977). Oligonucleotide primers were purchased from the Regional DNA Synthesis Laboratory, University of Calgary.

Reaction products were fractionated on a 6% acrylamide buffer gradient gel (Biggin *et al.*, 1983) and visualized (by autoradiography) following drying of the gel.

Chromosome preparation

The method of obtaining metaphase spreads from avian blood samples was based on the technique described previously (Biederman and Lin, 1982) with some modifications. About 0.5-1.0 ml of plasma containing lymphocytes of a female sarus crane and of a juvenile female whooping crane (from Wood Buffalo National Park, Northwest Territories, Canada) were added into 10 ml McCoy's 5a medium (in 25 mM Hepes) supplemented with 25% fetal calf serum, 1% L-glutamine and 0.2 ml of phytohemaglutinin for 65 hrs at 40^0C. Colcemid (0.1 mg/ml) was added to the cell culture 1-15 hrs prior to harvesting. The cells were treated with 0.075M KCl for 15 min and fixed in three changes of 3:1 methanol/glacial acetic acid. Chromosome spreads were achieved by releasing 2-3 drops of cell suspension onto a slide from a height of about 12 cm. The slides were placed on a rack over a 65^0C waterbath, where they were allowed to dry.

Chromosomal *In situ* hybridization

In situ hybridization of cloned DNA fragments to the metaphase chromosome spreads was based on the methods of Harper and Saunders (1981) and Zabel *et al.* (1983) and is described in detail elsewhere (Lin *et al.*, 1985). The DNA

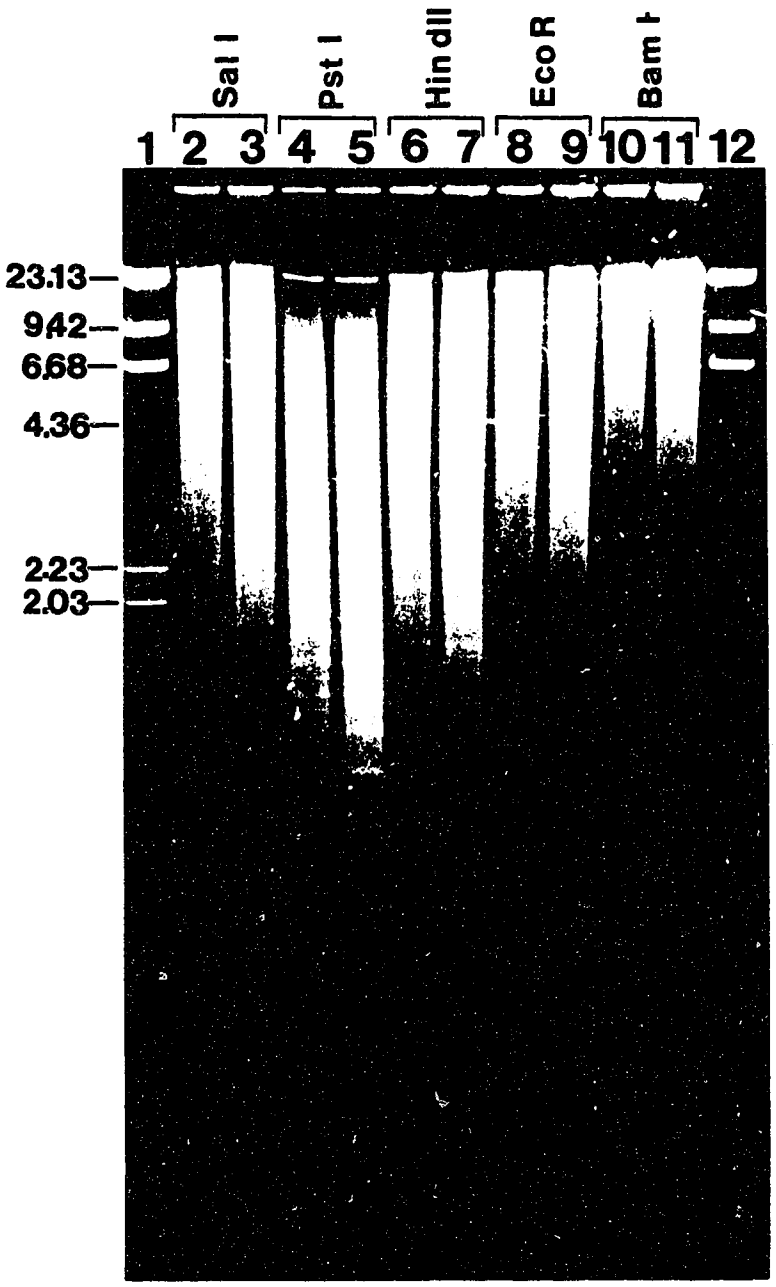
fragment from one of the Pst clones (pSC2) was oligo-labelled with ^{3}H-dCTP (55 Ci/mmole; New England Nuclear) to a specific activity of 6 x 10^{7} cpm/µg DNA. Approximately 2 ng of radio-labelled SC2 DNA (1.2 x 10^{5} cpm) in 100 µl of hybridization mixture containing 10% dextran sulfate were used on each slide and the hybridization was carried out at 37^{0}C for 16 hrs. Following the hybridization and washing procedures, the slide was then coated with Kodak NTB2 nuclear track emulsion and kept at 4^{0}C for 72 hrs prior to developing.

RESULTS

Detection, isolation and identification of the repetitive elements

Genomic DNA of the sarus crane was digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I and *Sal*I, respectively, and the digests were fractioned on a 1% agarose gel which was stained with ethidium bromide. As shown in Figure II-1, the most prominent ladder of bands was visible with the *Pst*I digested sample. The smallest band migrated to about 200 bp and the size difference between any two adjacent bands was also about 200 bp. We assumed the ladder was formed by the clustered repeats of a repetitive DNA family whose units were approximately 200 bp in length. To confirm this, the DNA in the 200 bp band was eluted from the gel, and cloned into *Pst*I digested pUC19 plasmid as described in Materials and Methods. Five clones (pSC1, 2, 6, 8 and 9) of the monomer unit were

Figure II-1. An ethidium bromide stained agarose gel (0.8%) on which 10 µg samples of *G. antigone* DNA digested with various restriction enzymes has been fractionated. Lanes 1 and 12 contain *Hin*dIII digests of λ DNA used as molecular weight markers. Even numbered lanes contain female DNA, odd numbered lanes male contain male DNA. Lanes 2, 3, *Sal*I; lanes 4, 5, *Pst*I; lanes 6, 7, *Hin*dIII; lanes 8, 9 *Eco*RI; lanes 10, 11, *Bam*HI.


Sal I
Pst I
Hin dII
Eco R
Bam I
1 2 3 4 5 6 7 8 9 10 11 12
23.13
9.42
6.68
4.36
2.23
2.03

randomly chosen for further studies. The insert in the clone pSC2 was labelled and hybridized to a Southern blot of sarus crane genomic DNA digested with several restriction enzymes. The autoradiograph in Figure II-2 shows a pronounced ladder of bands in the *Pst*I digest (lane 4) corresponding to the one seen previously on the gel stained with ethidium bromide (Fig. II-1). This indicates that these bands have substantial sequence homology and confirmed that the ladder was formed by a repetitive DNA family. Lower intensity ladders with almost negligible components of the monomer and dimer units were seen with the other enzymes as well. In the case of *Bam*HI, *Hin*dIII and *Sal*I, a single base pair change in the consensus sequence of the repeating unit would generate a unique recognition sequence for these enzymes in each repeat (Fig. II-3). Two base pairs must be altered in the consensus sequence to generate an *Eco*RI site, which accounts for the low abundance of *Eco*RI restriction fragments less than seven repeats long.

Sequence organization of the cloned elements

DNA sequences of the five independent clones, pSC1, 2, 6, 8 and 9, are presented in Figure II-3. As estimated by Biederman *et al.* (1982), in cranes the G+C content is about 39% in the genomic and 51-56% in the two satellite species. The average G+C content in our cloned fragments is 50% which suggests they may constitute part of the crane satellite found at 1.710 gm cm^{-3}. As a group, the nucleotide sequences

Figure II-2. Autoradiograph of a Southern hybridization of ^{32}P-labelled pSC2 probe with *G. antigone* DNA digested with various restriction enzymes. Lane 1, *Bam*HI; lane 2, *Eco*RI; lane 3, *Hin*dIII; lane 4, *Pst*I; lane 5, *Sal*I.

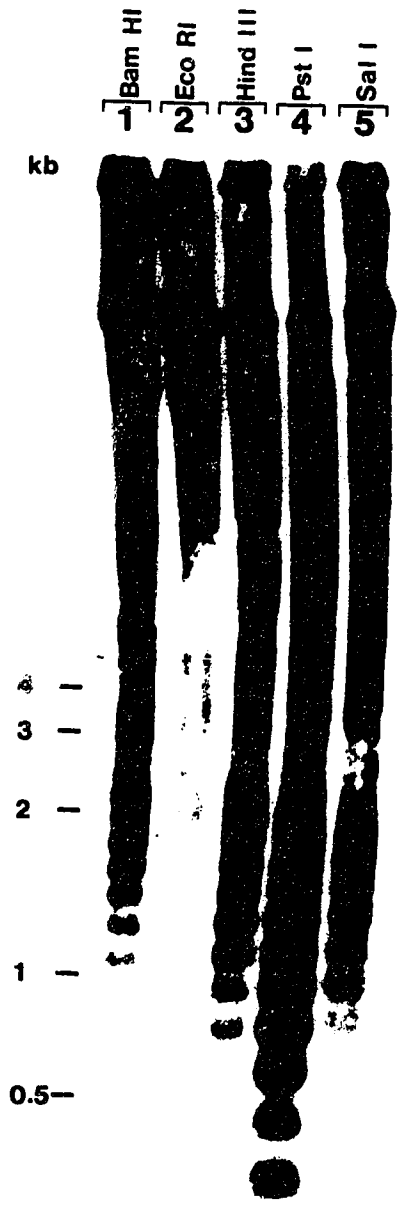
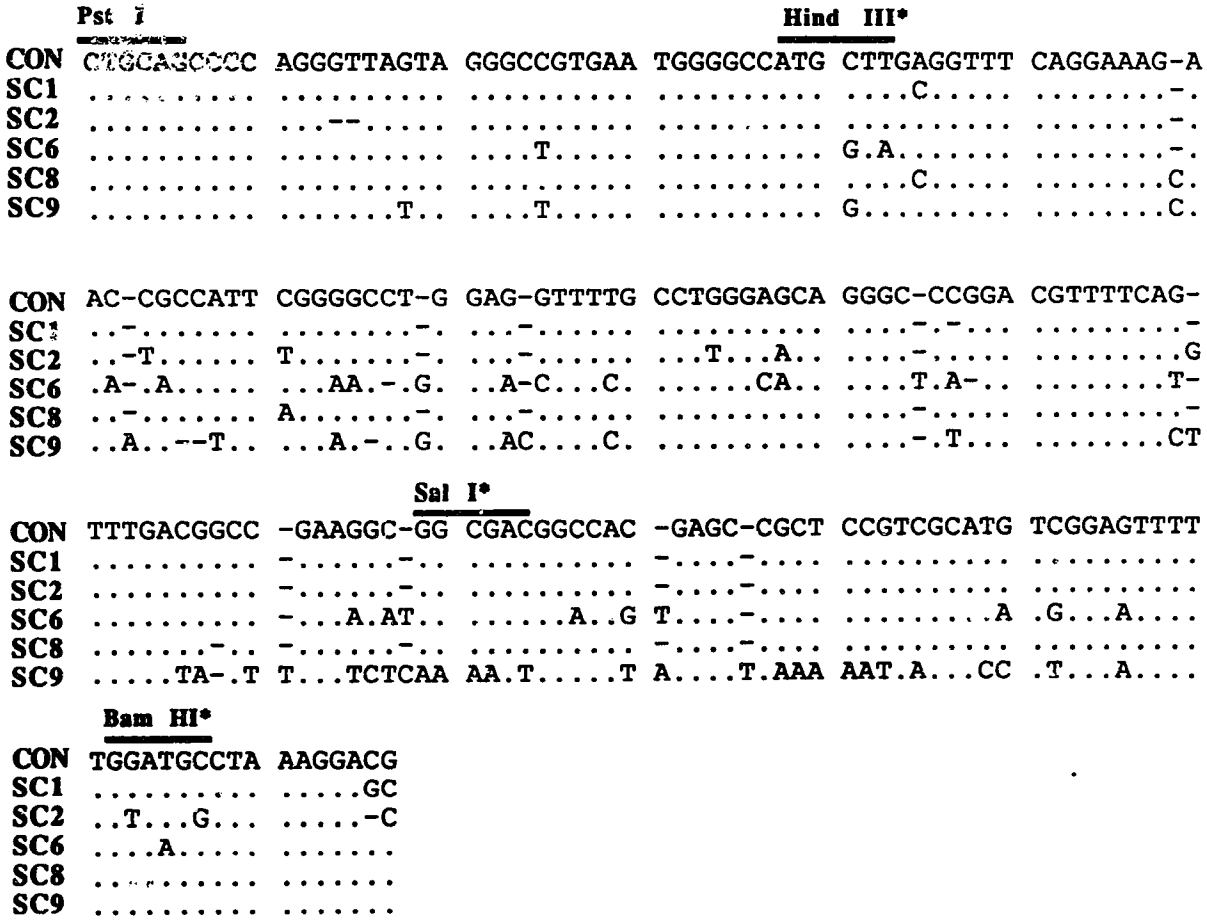

Bam HI
Eco RI
Hind III
Pst I
Sal I
1
2
3
4
5
kb
4
3
2
1
0.5

Figure II-3. Nucleotide sequences of the five cloned *Pst*I elements of *G. antigone*. The consensus DNA sequence of the five monomers is shown in 5' to 3' orientation at the top of each line. For clones SC1, SC2, SC6, SC8 and SC9, only the non-homologous nucleotides are indicated. Potential restriction sites that could be created by single base substitutions are indicated by an asterisk.

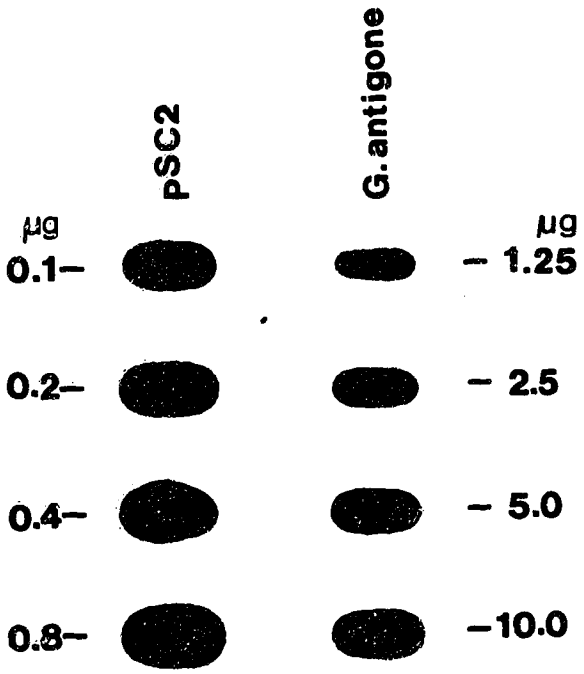

Pst I
Hind III*
CON CTGCAGCCCC AGGGTTAGTA GGGCCGTGAA TGGGGCCATG CTTGAGGTTT CAGGAAAG-A
SC1 ....C..... ....-.
SC2 ....--..... ....-.
SC6 ....T..... .... G.A....... ....-.
SC8 ....C..... ....C.
SC9 ....T.. ....T..... .... G......... ....C.
CON AC-CGCCATT CGGGGCCT-G GAG-GTTTTG CCTGGGAGCA GGGC-CCGGA CGTTTTCAG-
SC1 ..-....... ....-. ...-...... ....-.-... ....-
SC2 ..-T...... T.......-. ...-...... ...T...A.. ....-..... ....G
SC6 .A-.A..... ...AA.-.G. ..A-C...C. ......CA.. ....T.A-.. ....T-
SC8 ..-....... A.......-. ...-...... ....-..... ....-
SC9 ..A..--T.. ...A.-..G. ..AC....C. ....-.T... ....CT
Sal I*
CON TTTGACGGCC -GAAGGC-GG CGACGGCCAC -GAGC-CGCT CCGTCGCATG TCGGAGTTTT
SC1 .... -......-.. .... -....-.... ....
SC2 .... -......-.. .... -....-.... ....
SC6 .... -...A.AT.. ......A..G T....-.... ....A .G...A....
SC8 ....-.. -......-.. .... -....-.... ....
SC9 .....TA-.T T...TCTCAA AA.T.....T A....T.AAA AAT.A...CC .T...A....
Bam HI*
CON TGGATGCCTA AAGGACG
SC1 ....GC
SC2 ..T...G... .....-C
SC6 ....A..... ....
SC8 ....
SC9 ....

of the five clones exhibit greater than 90% homology to the consensus sequence shown in the figure. The mismatches are mainly base substitutions or small insertions and deletions. In this random sampling of monomer units bounded by *Pst*I sites, one unit, SC9, exhibits a significantly higher degree of divergence than the others from the consensus (24%). But since we have shown that SC2 hybridizes strongly to SC9 (data not shown), our estimate of the number of monomer units in the *Pst*I family (see below) encompasses members which have diverged from SC2 by at least 25%. The length of the monomers ranges from 185 to 192 bp and the consensus sequence is 187 bp.

Quantification

The copy number of the 200 bp repeat was determined by a slot blot hybridization experiment. Various amounts of pSC2 plasmid DNA and sarus crane genomic DNA were digested with *Pst*I, denatured and blotted onto a Gene Screen Plus membrane which was then hybridized to the pSC2 insert fragment that had been purified from a low melting point agarose gel. As shown by the autoradiograph in Figure II-4, the amount of signal given by 5 µg of the crane genomic DNA approximates that given by 0.1 µg of the plasmid. The respective signals were quantified by scintillation counting of slots cut from the blot. Assuming that the haploid DNA content in cranes is 1.4 picograms (Biederman *et al.*, 1982), we estimate that the monomer sequence is repeated 8800 times and comprises 0.14%

Figure II-4. Autoradiograph of slot blot hybridization to estimate the relative amount of the SC2 sequence in the *G. antigone* genome. The probe used was oligolabelled ^{32}P-labelled pSC2 insert DNA isolated from low melting point agarose.

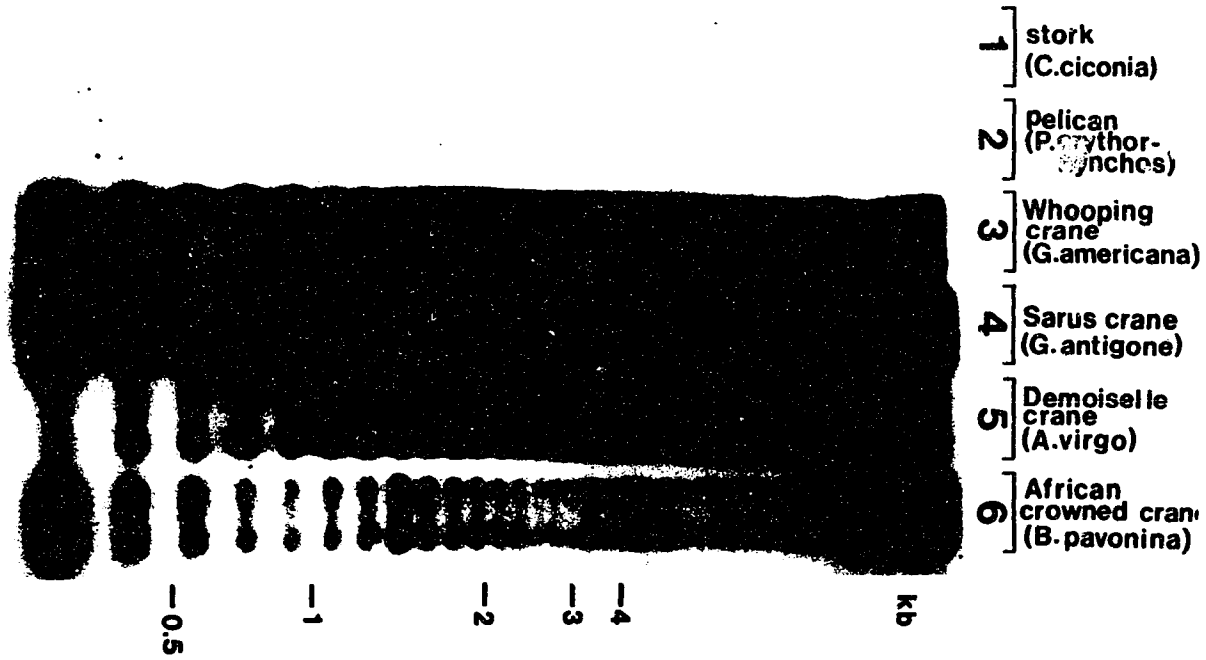

pSC2
G.antigone
µg
0.1–
0.2–
0.4–
0.8–
µg
– 1.25
– 2.5
– 5.0
–10.0

of the haploid genome complement. Since the repeat units are contiguous, the *Pst*I family extends over 1760 kilobases. Quite clearly, these estimates of the family size are dependent upon the stringency of the hybridization conditions. Since these were high (see Materials and Methods) our estimates are necessarily lower limits.

Cross hybridization with other avian species

We attempted to discover how widespread the *Pst*I family was in other avian species by obtaining genomic DNA from several other members of the crane family: the whooping crane, demoiselle crane (*Anthropoides virgo*) and African crowned crane (*Balearica pavonina*) and from two species of different orders: white pelican (*Pelecanus erythrorhynchos*) and white stork (*Ciconia ciconia*). The DNA samples were digested with *Pst*I, fractionated on agarose and transferred to Gene Screen Plus. The blot was hybridized with the pSC2 probe. As shown in Figure II-5, all the crane species have sequences homologous to the probe in their genomes whereas neither the pelican nor the stork DNA samples exhibited detectable hybridization. Further studies (data not shown) revealed no sequence homology was present in a species of the same order: the sora rail (*Porzana carolina*) and species from two additional orders; Passeriformes (black-billed magpie, *Pica pica*) and Strigiformes (long-eared owl, *Asio otus*, short-eared owl, *Asio flammeus*); however, a weak Pst

Figure II-5. Autoradiographs of a Southern hybridization of ^{32}P-labelled pSC2 plasmid DNA with *Pst*I digested genomic DNAs from 1, *C. ciconia*; 2, *P. erythorhynchos*; 3, *G. americana*; 4, *G. antigone*; 5, *A. virgo*; and 6, *B. pavonina*.


1 stork (C.ciconia)
2 Pelican (P.erythor-rynchos)
3 Whooping crane (G.americana)
4 Sarus crane (G.antigone)
5 Demoiselle crane (A.virgo)
6 African crowned cran (B.pavonina)
0.5
1
2
3
4
kb

ladder was present in both a Falconiformes (Andean condor) and a Galliformes (Japanese quail, *Coturnix japonica*). A repetition of the experiment using two Jananese quails from a different inbred population failed to demonstrate the presence of a *PstI*-like family. Therefore, additional samples will be required to clarify this descrepancy. Further samples of Andean condor would also be required to confirm the preliminary observation noted above.

Chromosomal localization of the *Pst*I elements

In general, the number and morphology of sarus crane chromosomes are similar to those of whooping crane described earlier (Biederman *et al.*, 1982). Six pairs of macro-autosomes and the sex chromosome Z are readily distinguished from the micro-chromosomes. Macro-autosomes 1 to 3 can be unequivocally identified without using chromosome banding techniques. After *in situ* hybridization with the ^{3}H-labelled SC2 fragment, clusters of silver grains were consistently observed in the centromeric region of chromosome 2 of the sarus crane (Fig II-6). Among the 20 metaphase spreads in which hybridizing grains were found, 17 (85%) had silver grains clustered at the centromeric region of both homologues, the remaining 3 cells had grains over just one of the homologues. The number of grains per centromeric region varied from 1-8 with an average of 3-4 grains. In four metaphases, a single silver grain was found on the centromeric region of one of the homologues of chromosome 1.

Figure II-6. Cytological localization of the Pst family in *G. antigone*. Autoradiograph of a representative metaphase spread from a female sarus crane hybridized with ^{3}H-labeled Pst I fragment from pSC2. The macrochromosomes 1, 2, and 3 are indicated. Clusters of silver grains are localized over the centromeric regions of both chromosomes 2 (indicated by arrows). Two additional partial metaphase spreads also showed silver grains over the centromeric region of both homologous chromosomes 2 (inset).


2
2
3
3
2
2
2
10 µ

Only a few silver grains appeared over the microchromosomes. Therefore, the majority, if not all, of the DNA sequences hybridizing to the cloned SC2 fragment are located in the centromeric region of chromosome 2.

Similar observations were made for the whooping crane chromosome preparations. In metaphase spreads where hybridization grains were observed, the grains were found to be located over the centromeric region of chromosome 2 exclusively (Fig II-7). However, only 6 cells out of 32 (19%) examined had both homologues labelled. The majority of cells analyzed had only 1-2 grains in the centromeric region of the chromosome. The chromosome preparations of the whooping crane that were used in this study were approximately two years old which may have contributed to the lesser hybridization signal observed.

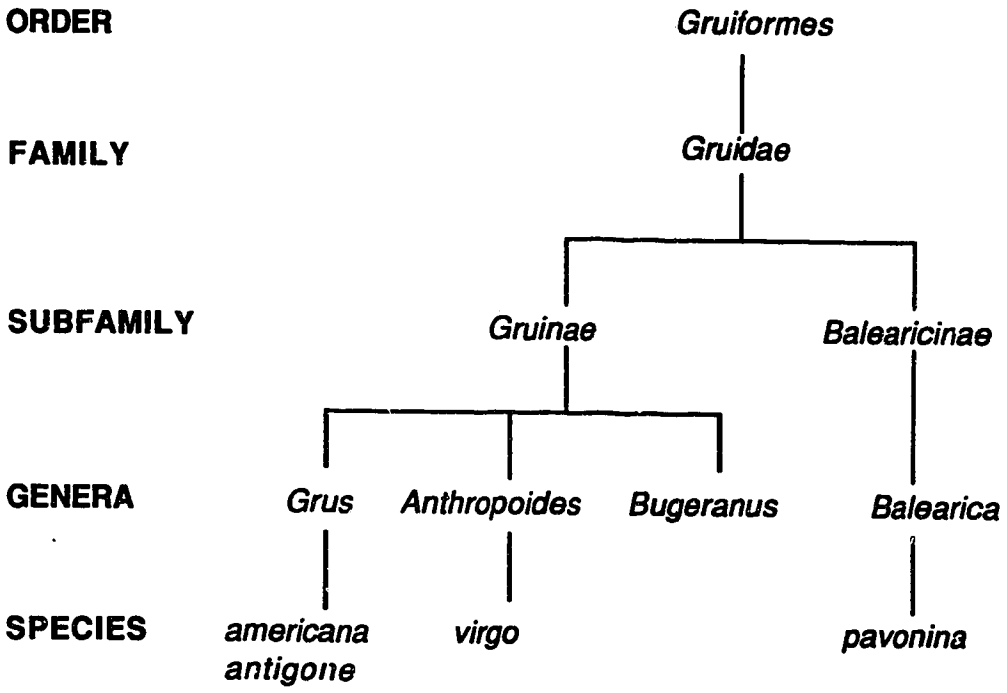
DISCUSSION

We have described a repetitive component of the crane genome which exhibits many of the properties of satellite DNA as defined by Singer (1982). The tandem arrangement of the repeating units was established by restriction digestion using *Pst*I. This enzyme cut once within a majority of the units and a long ladder of fragment sizes resulted. The *Pst*I family is highly conserved within the crane family, although, quite clearly, the distribution within the array of units containing the *Pst*I site differs between the species surveyed (Fig. II-5). Taxonomists have grouped Grus and Anthropoides

Figure II-7. Cytological localization of the Pst family in *G. americana*. Autoradiograph of a metaphase spread (a) and partial metaphase spreads (b-d) from a female whooping crane hybridized with the pSC2 insert fragment. Macro-chromosome 1 to 3 are indicated. (a, b) silver grains were found on the centromeric regions of both homologous chromosomes 2 (indicated by arrows). (c, d) only one homologue of chromosome 2 showed silver grains over its centromeric region (indicated by arrow).


10μ
a

Figure II-8. Taxonomy of *Gruidae* family (after Walkinshaw, 1973). Only those species discussed in the text are included.



ALL THE FAMILIES IN THIS ORDER

Diatryma, Phororhacos, Mesites, Bustards, Cariamas, Hemipodes, Cranes, Limpkins, Trumpeters, Kagu Sun Bitterns, Finfeet, Gallinules, Coots, Rails

into one subfamily, Gruinae, while Balearica falls into the subfamily Balearicinae (Fig II-8). Based on this classification, one would expect that Anthropoides would have diverged less from Grus than Balearica would have. Indeed, if one uses the intensity of the hybridization signal of the entire Pst ladder as a measure of relatedness, it is clear from our data (Fig. II-5) that Balearica has diverged more from Grus than has Anthropoides.

Interestingly, the Pst array was absent from the sora rail, a member of a different family within the same order as the cranes. Further screening of species within this order will allow instructive comparisons to be made with current taxonomic classifications based on whole genome DNA hybridization data (Sibley *et al.*, 1988). The preliminary data that shows the *Pst*I-like array may be present in two other orders (Falconiformes, Galliformes) would allow similar analyses to be conducted.

An interesting feature of the *Pst*I family is its location at a unique site in the crane genome - the centromeric heterochromatin of chromosome 2 (Figs. II-6, II-7). The satellite described by Brown and Jones (1972) in the Japanese quail, for example, is located predominantly in the micro-chromosomes. Virtually all of the vertebrate satellites described to date are found in several to many heterochromatic locations. The general picture that emerges is that centromeric domains frequently, if not always, contain tandemly repeated DNA sequences, but that the arrays

differ on non-homologous chromosomes. We are struck by the observations of Zhang and Horz (1984), Linxweiler and Horz (1985) as well as Zhang *et al.* (1983), all of whom show quite clearly that nucleosomes can be arranged in only a limited number of "phases" on such arrays. This is a point on which we shall elaborate below, but in the present context, such phasing might permit a close juxtapositioning of the two DNA helices comprising the sister chromatids. A close association of the two DNA molecules throughout the domain of centromeric heterochromatin may be an important feature of chromosome stability.

The *Pst*I family appears to include blocks as large as 17 monomer units within which no *Pst*I site is found. However, it is highly unlikely that random mutational events at up to 17 consecutive *Pst*I sites generated these large blocks (Hodgetts and Strobeck, 1989), since DNA sequence comparisons between the 5 cloned monomer units (Fig. II-3) revealed only slight deviations from the consensus sequence. It is far more likely that saltatory amplification of a much smaller domain in which *Pst*I sites were missing generated the larger blocks. The commonly accepted mechanism for such amplification is unequal crossing over (Smith, 1973; Tartof, 1973) and evidence exists that shows it occurs within tandemly repeated arrays (Tartof, 1973; Petes, 1980).

The repeat length of the monomeric unit is 187 bp, a feature which characterizes the repeats in many plant and animal species (see Table I-1). Although this conservation

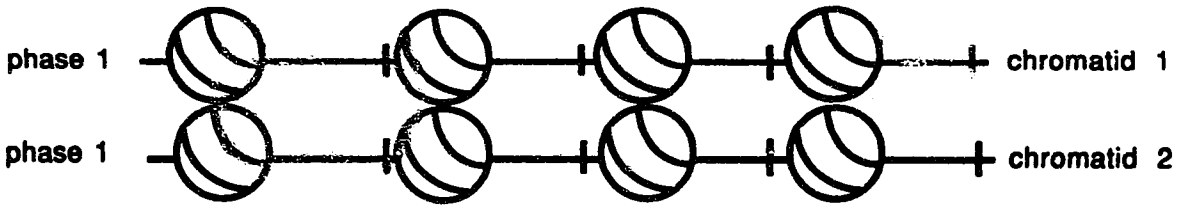
of length is striking, there is no detectable sequence homology between the repeats of distantly related species. Martinez-Zapater *et al.* (1986) note the coincidence between the conserved size of the repeat units and the length of DNA found in a mononucleosome (core + linker), and they suggested that the nucleosomes may determine the size of the unit to be amplified. We would like to extend this notion further with the suggestion that unequal crossing over can only occur within tandem arrays when the nucleosomes are maximally out of phase on the two DNA molecules involved. While it should be emphasized at this point that the nucleosomal organization of DNA prior to a recombinant event remains to be established, it has been well documented that the nucleosomes are phased on the 234 bp mouse satellite (Linxweiler and Horz, 1985) and on the African green monkey satellite (Zhang *et al.*, 1983). We suggest that interference may occur between "in-phase" nucleosomes as shown very schematically in Figure II-9a, and prevent the molecular juxtapositioning required for a recombination event. However, the phase established on one homologue may differ from that established on the other, since (for example) on the African green monkey satellite there are eight alternative phase positions (Zhang *et al.*, 1983). We postulate that if the nucleosomes are aligned in maximally different phases such that interdigitation of the two nucleosomal arrays as shown in Fig II-9b can occur, the recombinational events leading to amplification of the monomer units would be favoured. Since

Figure II-9. The role of nucleosome phasing in the amplification of tandem arrays.In

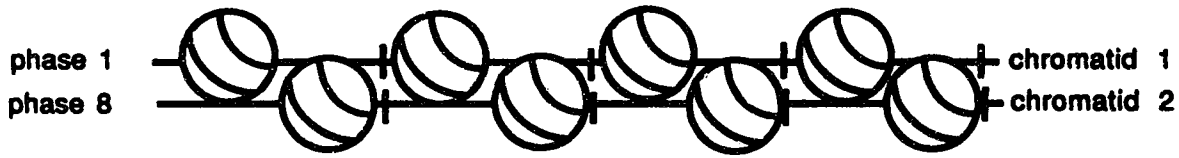
a) the two chromatids cannot pair closely because the nucleosomes are in identical phases on the tandemly repeated monomer units (whose ends are represented by the short vertical lines). However, in

b) the nucleosomes are maximally out of phase on the two arrays which allows a close juxtapositioning of the two homologues. The choice of phases 1 and 8 is based on the African green monkey satellite data (Zhang *et al.*, 1983) where eight phasing possibilities exist and is meant to imply that these are the two extremes. The chromatids 1 and 2 could be either sisters or non-sisters.

a)


phase 1
phase 1
chromatid 1
chromatid 2

b)


phase 1
phase 8
chromatid 1
chromatid 2

the spacing of the nucleosomes (which is to be distinguished from their phasing) is roughly every 200 bp, only repeat units of this size (or larger) could provide phasing differences which would permit interdigitation of the nucleosomes and thus amplification of the monomer units. To restate this argument in another way, tandem arrays with repeat units shorter than 200 base pairs would only evolve into a significant component of the genome to the extent that mechanisms other than recombination (replication?) are operating.

While tandemly repeated arrays may have evolved as a common solution to the requirement for phased nucleosomes in the centromeric domains of higher eukaryotes, the yeast *S. cerevisiae* apparently has evolved another mechanism. Evidence supports the existence of phased nucleosomes in the centromeric domains (Bloom and Carbon, 1982) but there is no indication that this occurs on tandem arrays.

To summarize, we postulate that there is a requirement for close juxtapositioning of the two sister chromatids to insure chromosome stability at metaphase or movement at anaphase. This may require differentially phased nucleosomes on sister chromatids which introduces a selection pressure for tandem arrays in the centromere region. Such arrays provide a simple solution to the problem of phasing. The physical constraints in the amount of DNA that comprises a nucleosome, leads to the conclusion that large scale amplification of such arrays beyond the early duplicative

events, (which, presumably occur by a different mechanism), cannot occur if the repeat units are less than the size of the mononucleosome and its linker (200 bp). This, we suggest, explains the rarity of satellites smaller than this size which are located mainly in the centromeric heterochromatin.

REFERENCES

Barsacchi-Pilone, G., Batistoni, R., Andronico, F., Vitelli, L., and Nardi, I. 1986. Heterochromatic DNA in *Triturus* (Amphibia, Urodela). Chromosoma 93: 435-446.

Biederman, B.M., and Lin, C.C. 1982. A leukocyte culture and chromosome preparation technique for avian species. In vitro 18: 415-418.

Biederman, B.M., Lin, C.C., Kuyt, E., and Drewien, R.C. 1982. Genome of the whooping crane. J. Heredity 73: 145-146.

Biggin, M.D., Gibson, T.J., and Hong, G.F. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA 80: 3963-3965.

Bloom, K.S., and Carbon, J. 1982. Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromes. Cell 29: 305-317.

Brown, J.E., and Jones, K.W. 1972. Localisation of satellite DNA in the microchromosomes of the Japanese quail by *in situ* hybridization. Chromosoma 38: 313-318.

Comings, D.E., and Mattoccia, E. 1970. Studies of microchromosomes and G-C rich DNA satellite in the quail. Chromosoma 30: 202-214.

Datta, U., Dutta, P., and Mandal, R.K. 1988. Cloning and characterization of a highly repetitive nucleotide sequence. Gene 62: 331-336.

Feinberg, A.P., and Vogelstein, B. 1984. Addendum of "A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity". Anal. Biochem 137: 266-267.

Harper, M.E., and Saunders, G.G. 1981. Localization of single copy DNA sequences on G-banded human chromosomes by *in situ* hybridization. Chromosoma 83: 431-439.

Hodgetts R. and Strobeck C. 1989. Mechanisms of evolution of tandemly repeated DNA sequence arrays located in centric heterochromatin. The Genetics Society of Canada Bulletin 20, Supm 1: 35-35

Ish-Horowicz, D., and Burke, J.F. 1981. Rapid and efficient cosmid vector cloning. Nucleic Acids Res. 9: 2989-2998.

Lin, C.C., Draper, P.N., and De Braekeleer, M. 1985. High resolution chromosomal localization of the b-gene of the human b-globin gene complex by *in situ* hybridization. Cytogenet Cell. Genet. 39: 269-274.

Linxweiler, W., and Horz, W. 1985. Reconstitution experiments show that sequence-specific histone-DNA interactions are the basis for nucleosome phasing on mouse satellite DNA. Cell 42: 281-290.

Maio, J.J., Brown, F.L. and Musich, P.R. 1981. Towards a molecular paleontology of primate genomes. Chromosoma 93: 103-125.

Maniatis, T., Fritsch, E.F., and Sambrook, J. 1982. Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martinez-Zapater, J.M., Estelle, M.A., and Somerville, C.R. 1986. A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol. Gen. Genet. 204: 417-423.

McDonnell, M.W., Simon, M.N., and Studier, F.W. 1977. Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels. J. Mol. Biol. 110: 119-146.

Miklos, G.L.G. 1982. Sequencing and manipulating highly repeated DNA. In: Genome Evolution. Edited by F.B. Flavell. Academic Press, London.

Petes, T.D. 1980. Unequal meiotic recombination within tandem arrays of yeast ribosomal DNA genes. Cell 19: 765-774.

Sanger, F., Nicklen, S., Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463-5467.

Sibley, C.G., Ahlquist, J.E., Monroe, B.L. 1988. A classification of the living birds of the world based on DNA-DNA hybridization studies. The Auk 105: 409-423.

Singer, M.F. 1982. Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76: 67-112.

Smith, G.P. 1973. Unequal crossover and the evolution of multigene families. Cold Spring Harbor Symp. Quant. Biol. 38: 507-513.

Tartof, K.D. 1973. Unequal mitotic sister chromatid exchange and disproportionate replication as mechanisms regulating ribosomal RNA gene redundancy. Cold Spring Harbor Symp. Quant. Biol. 38: 491-500.

Tone, M., Nakano, N., Takao, E., Narisawa, S., and Mizuno, S. 1982. Demonstration of W chromosome-specific repetitive DNA sequences from the domestic fowl, *Gallus g. domesticus*. Chromosoma 86: 551-569.

Walkinshaw, L. 1973. Cranes of the world. Winchester Press, New York, pp. 9-14.

Wu, C., Bingham, P.M., Livak, K.J., Holmgren, R., and Elgin, S.C.R. 1980. The chromatin structure of specific genes: I. Evidence for higher order domains of defined DNA sequence. Cell 16: 797-806.

Yu, L.-C., Lowensteiner, D., Wong, E.F.-K., Sawada, I., Mazrimas, J., Schmid, C. 1986. Localization and characterization of recombinant DNA clones derived from the highly repetitive DNA sequences in the Indian muntjac cells: Their presence in the Chinese muntjac. Chromosoma 93: 521-528.

Zabel, B.U., Naylor, S.L., Sakaguchi, A.Y., Bell, G.I., and Shows, T.B. 1983. High-resolution chromosomal localization of human genes for amylase, propiomelanocortin, somatostatin, and a DNA fragment (D357) by *in situ* hybridization. Proc. Natl. Acad. Sci. USA 80: 6932-6936.

Zhang, X.Y., and Horz, W. 1984. Nucleosomes are positioned on mouse satellite DNA in multiple highly specific frames that are correlated with a diverged repeat of nine base-pairs. J. Mol. Biol. 176: 105-129.

Zhang, X.Y., Fittler, F., and Horz, W. 1983. Eight different highly specific nucleosome phases on α-satellite DNA in the African green monkey. Nucleic Acids Res. 11: 4287-4306.

CHAPTER III

Sequence conservation in avian CR1: an interspersed repetitive DNA family evolving under functional constraints

---

A version of this chapter has been submitted for publication: Chen Z.-Q., C.C. Lin and R.B. Hodgetts 1990. Sequence conservation in avian CR1: an interspersed repetitive DNA family evolving under functional constraints.

INTRODUCTION

Repetitive DNA families have proven to be useful for studies on molecular evolution (Britten, 1986). In the pursuit of our interest in the repetitive component of the avian genome, we have described, and speculated upon the function of, a tandemly repeated, centromeric DNA family in cranes (*Gruidae*) (Chen et. al., 1989). In the present study, we explore the characteristics of an interspersed repetitive DNA family. The eukaryotic interspersed DNA families studied to date appear to be confined to closely related species, with only one exception: the L1 family found in a wide range of mammalian species (Voliva et. al., 1983). This narrow distribution of interspersed DNA families has led some to speculate that they comprise a non-functional (junk) component of the genome (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). However, since their discovery by Britten and Kohn (1968), many functional roles have been suggested, most notable of which are those involved in gene regulation (Britten and Davidson, 1969; Davidson and Britten, 1979). If any functionally significant, interspersed DNA families exist in birds, we hypothesized that they should be highly conserved. Therefore, we undertook a screen of the crane genome hoping to identify repetitive sequence elements that were present in many other avian species. The first clone we identified turned out to be a member of the interspersed

repetitive DNA family, CR1, previously characterized in the chicken (*Gallus domesticus*) genome (Stumph et. al., 1981).

The chicken CR1 family has been the subject of extensive study. It is comprised of approximately 7000 to 20000 repeats per haploid genome, the lengths of individual members varying from 160 to 850 base pairs (Stumph et. al., 1981; Hache and Deeley, 1988). Recently, it was found that a CR1 member was present in the chicken vitellogenin pseudogene (ΨVTGIII) but absent in the ancestral VTGIII gene, indicating that some of the CR1 members are, or were, mobile (Silva and Burch, 1989). Interestingly, arbitrarily assigned 5' ends of the known CR1 members are all truncated at various positions whereas the 3' ends are relatively conserved. Some members have short direct repeats at both ends and occasionally, an open reading frame is found throughout most of the CR1, coincident with the assigned 5' to 3' orientation. These features have prompted the speculation that CR1 is a novel retroposon family (Silva and Burch, 1989).

Additional functional properties have been ascribed to CR1. Chromatographic columns containing CR1 DNA were used to isolate specific nuclear proteins from chicken oviduct tissue, and the fractionated proteins were found to bind at a 3' region of the CR1 sequence (Sanzo et al., 1984). Secondly, in dissecting the upstream region of the chicken lysozyme gene, Baniahamad et al. (1987) used a reporter gene to define a silencer element in the central portion of a CR1 member found there.

Preliminary evidence has shown that the CR1 family exists in duck and peacock (Schip et. al., 1987) and in this paper, we provide evidence showing that the CR1 DNA family exists in a wide range of avian genomes. Our data include a characterization of clones obtained from the emu (*Dromaius novaehollandiae*), one of the bird species known to bear the most primitive characters (Cracraft, 1974). The persistence of CR1 throughout avian evolution argues strongly for a functional role and this is supported by our finding that two regions are highly conserved: one known to interact with nuclear protein(s) and the other known to contribute to the regulation of gene expression.

MATERIALS AND METHODS

Specimens

Avian species used in this study include emu, cassowary (*Casuarius casuarius*), pelican (*Pelecanus erythrorhynchos*), stork (*Ciconia ciconia*), condor (*Vulture gryphus*), shelduck (*Tadorna tadornoides*), quail (*Coturnix japonica*), crane (*Grus antigone*), owl (*Asio otus*) and magpie (*Pica pica*). Blood samples from these birds were kindly provided by Dr. R.M. Cooper and Dr. M. Mainka at the Calgary city zoo (Alberta, Canada). Newt (*Triturus alpestris*), frog (*Xenopus laevis*), fish (*Carassius auratus*) and lobster (*Homarus vulgaris*) genomic DNAs were kindly provided by Dr. R. Sasi in the Department of Pathology, University of Alberta.

DNA Extraction, Southern and dot blot hybridization

The protocol for extracting genomic DNA from avian blood samples and the DNA hybridization conditions have been described previously (Chen et. al., 1989).

Isolation of conserved repetitive DNA families

A library of *Eco*RI restriction fragments from genomic DNA of the sarus crane was constructed in the vector pUC19. The bacterial strain JM83 was transformed with an aliquot of the library and transformants which contained repetitive DNA sequences were identified by the intensity of their signals after colony hybridizations with crane genomic DNA. The repetitive nature of these clones was confirmed by isolating and labelling the plasmids and hybridizing them to Southern blots of restricted crane genomic DNA. The extent of sequence conservation was determined by dot hybridization to a test panel of DNA samples from a pelican, condor, crane, chicken, owl, magpie, mouse and human.

PCR and subsequent cloning

The primers used for the PCR were 5'-CAGGACAAGGGGTAATGG GT-3' and 5'-CATAGAATGGTTTGGGTTGG-3'. Reactions contained 100 ng of each primer, 30 ng of the template DNA and 1 unit of Taq polymerase (Pharmacia) in a reaction buffer containing 50 mM Tris (pH 9.0), 10 mM $MgCl_2$, 15 mM $(NH_3)_2SO_4$, 0.1 µg bovine serum albumin and 2 mM each of the four

deoxyribonucleotides. The reactions were carried out in an automated PCR machine (Tyler Research). Initial incubation was at 95^0C for 4 minutes, followed by 16-30 cycles at 60^0C for 1 minute, at 73^0C for 3 minutes and at 91^0C for 5 seconds. Aliquots of the amplification reactions were mixed with 50% glycerol and loaded onto 2% agarose gels. After electrophoresis, the products were visualized under UV light and photographed. If subsequent cloning was undertaken, an aliquot of the PCR mixture was incubated briefly with T4 DNA polymerase and additional nucleotides to fill in ends of the fragments. Afterwards, the fragments were purified and concentrated with Geneclean (Bio/Can Scientific Inc.), and then ligated into blunt ended pUC19 vector DNA (Maniatis et. al., 1982) and transformed into DH5α competent cells (GIBCO/BRL).

DNA sequencing

Cloned plasmids were isolated from their bacterial hosts by a mini procedure (Maniatis et. al., 1982), and a T7 sequencing kit (Pharmacia) was employed for double stranded DNA sequencing. The ^{35}S-labelled reaction products were fractionated on 6% acrylamide gels under denaturing conditions and visualized by autoradiography of the dried gels.

RESULTS

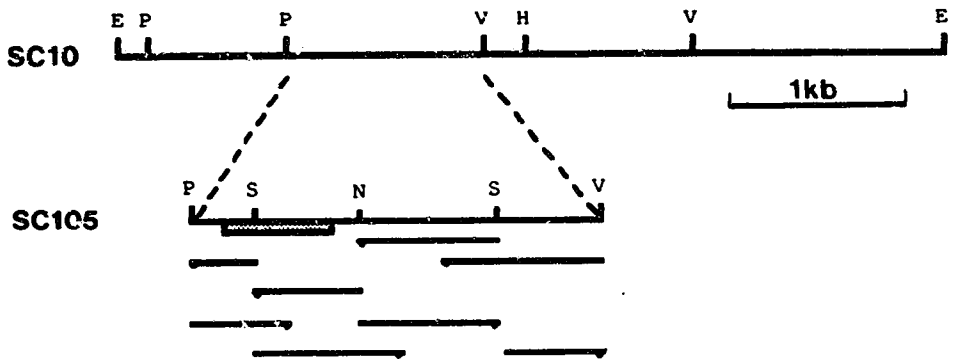
Isolation and restriction mapping of SC10

From the plasmid library constructed of crane DNA, a repetitive clone, designated as pSC10, was selected for study because it showed homology to all the six avian DNA samples, but it did not hybridize to any of the mammalian samples used in the test panel (data not shown). The pSC10 has a 4.6 kb insert of crane DNA and the repetitive component in the clone was characterized by restriction analysis on the insert SC10 and a 1.25 kb *Pst*I/*Pvu*II subclone, SC105. The results are summarized in Figure III-1a. Individual restriction fragments isolated from SC10 were used as probes for Southern hybridizations with *Hin*dIII digested crane genomic DNA. As shown in Figure III-1b, the *Pst*I and *Pvu*II fragments hybridized to a corresponding band or bands on the Southern blots (lanes 1 and 3, respectively), indicating their single copy nature. By contrast, the *Pst*I/*Pvu*II fragment found homologous sequences throughout the genome (lane 2). These results demonstrated the interspersed nature of the repetitive DNA in clone SC10 and suggested that an entire repeat unit was confined to the *Pst*I/*Pvu*II fragment.

DNA sequencing of the repetitive unit in SC10

Double stranded sequencing of the repeat unit was facilitated by subcloning the *Pst*I/*Pvu*II fragment from pSC10 into *Pst*I/*Hinc*II restricted pUC19. The sequencing strategy carried out on the resulting plasmid, pSC105, is summarized in Figure III-1a. The DNA sequence of SC105 was then compared with sequences stored in the MicroGenie (Beckman)

Figure III-1. a. Restriction maps of SC10 and SC105. The restriction enzymes used were *Eco*RI (E), *Pst*I (P), *Pvu*II (V), *Hin*dIII (H), *Sau*3a (S) and *Nco*I (N). The hatched box indicates the CR1 element. Lines beneath the SC105 map show the sequencing strategy employed. b. Autoradiographs of Southern hybridizations of *Hin*dIII digested crane genomic DNA (5 µg) probed with ^{32}P labeled *Pst*I fragment (Lane 1), *Pst*I/*Pvu*II fragment (Lane 2), *Pvu*II fragment (Lane 3). Sizes of fragments were determined from DNA markers (GIBCO/BRL).



a

b

data bank. The search revealed a significant homology between this DNA sequence and members of the CR1 family. A comparison of the consensus DNA sequence of the chicken CR1 family members (Stumph et. al., 1984) with homologous sequences of the crane SC105 is shown in Figure III-2. There is approximately 80% homology between these two sequences. The differences between them were comprised primarily of small deletions and base substitutions. Homology between the two sequences extended for about 300 base pairs.

Conservation of the CR1 homologous sequence in other avian species

To determine the extent of conservation of the CR1 homologous sequence in avian species, SC105 was labelled and hybridized to a Southern blot containing restricted genomic DNAs from emu, cassowary, pelican, stork, condor, quail, crane, owl and magpie. As shown in Figure III-3a, the probe detected smears of different intensity in each of the samples, except for the emu and cassowary, where hybridization was barely detected. A very pronounced band, about 2.8 kb, characterizes the stork DNA and suggests that a significant number of the CR1 units are clustered in a tandem array in the genome of this species. The same blot was stripped and rehybridized with the *Hin*dIII/*Eco*RI fragment of pSC10 which does not include any CR1 sequence (see Figure III-1a). The discrete bands observed (Figure III-3b) prove that the smears seen in Figure III-3a were not due to incomplete digestion nor degradation of the DNA samples.

Figure III-2. Nucleotide sequence of CR1 elements. The CR1 consensus sequence of chickens, CHICK, is taken from Stumph et. al. (17) and is shown on the top lines; Y indicates a pyrimidine. Differences from this consensus sequence as they exist in SC105 (CRANE), and five emu clones (EMU 1, 2, 3, 4 and 5) are shown below. Nucleotides are numbered (increasing in 3' to 5' orientation) according to Silva and Burch (11). The silencer (1) and binding domain of the nuclear protein (2) are shown by the heavy lines. The two primers used for PCR are indicated by the thin lines and match the CRANE sequence.

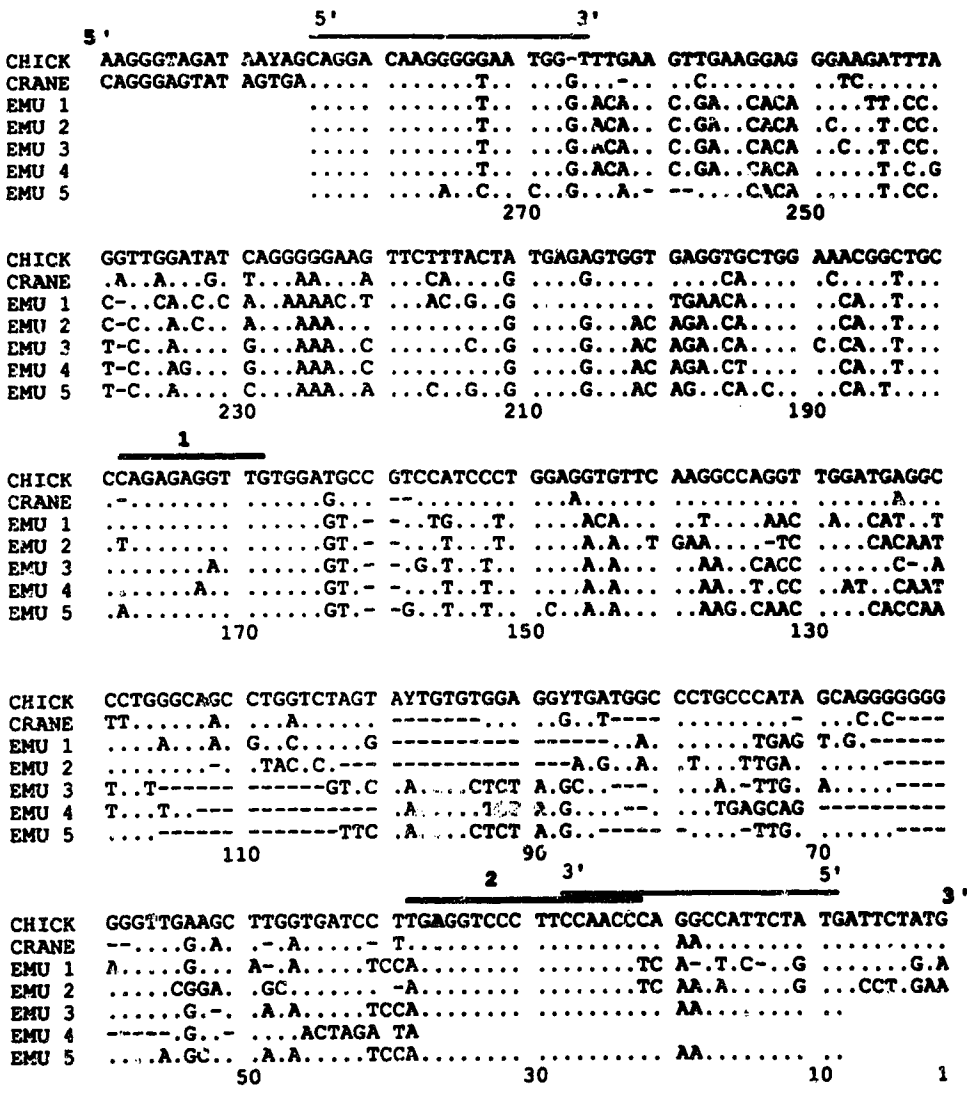
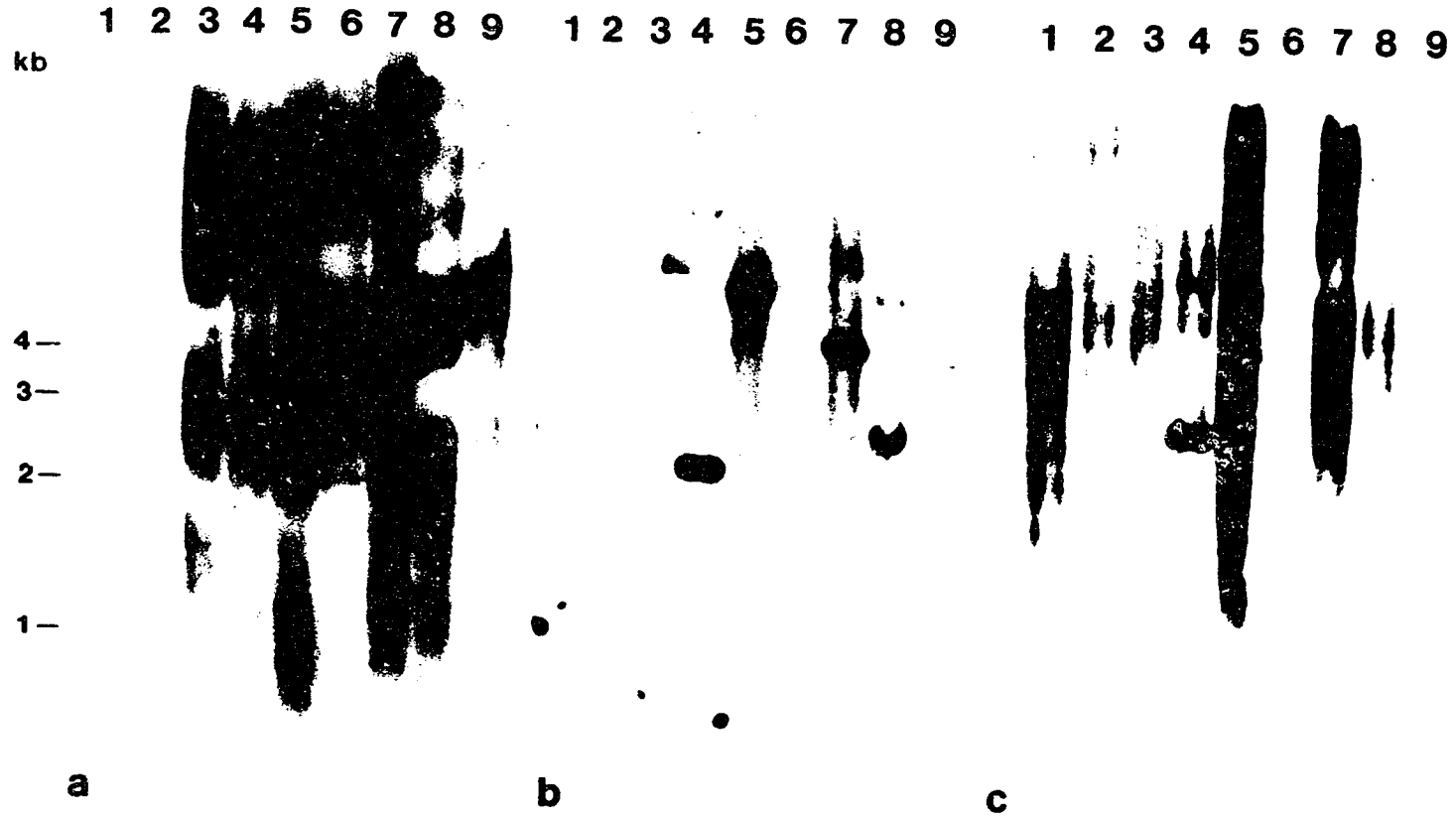

5'
3'
5'
CHICK AAGGGTAGAT AAYAGCAGGA CAAGGGGGAA TGG-TTTGAA GTTGAAGGAG GGAAGATTTA
CRANE CAGGGAGTAT AGTGA..... ....T.. ...G...-.. ..C..... ..TC.....
EMU 1 ....T.. ...G.ACA.. C.GA..CACA ....TT.CC.
EMU 2 ....T.. ...G.ACA.. C.GA..CACA .C...T.CC.
EMU 3 ....T.. ...G.ACA.. C.GA..CACA ..C..T.CC.
EMU 4 ....T.. ...G.ACA.. C.GA..CACA .....T.C.G
EMU 5 ....A..C.. C..G...A.- --....CACA .....T.CC.
270 250
CHICK GGTTGGATAT CAGGGGGAAG TTCTTTACTA TGAGAGTGGT GAGGTGCTGG AAACGGCTGC
CRANE .A..A...G. T...AA...A ...CA....G ....G..... ....CA.... .C....T...
EMU 1 C-..CA.C.C A..AAAAC.T ...AC.G..G .... TGAACA.... ..CA..T...
EMU 2 C-C..A.C.. A...AAA... ....G ....G...AC AGA.CA.... ..CA..T...
EMU 3 T-C..A.... G...AAA..C ......C..G ....G...AC AGA.CA.... C.CA..T...
EMU 4 T-C..AG... G...AAA..C ....G ....G...AC AGA.CT.... ..CA..T...
EMU 5 T-C..A.... C...AAA..A ...C..G..G ....G...AC AG..CA.C.. ..CA.T....
230 210 190
1
CHICK CCAGAGAGGT TGTGGATGCC GTCCATCCCT GGAGGTGTTC AAGGCCAGGT TGGATGAGGC
CRANE .-..... ......G... --..... ...A..... ....A...
EMU 1 ....GT.- -..TG...T. ....ACA... ..T....AAC .A..CAT..T
EMU 2 .T..... ......GT.- -...T...T. ....A.A..T GAA....-TC ....CACAAT
EMU 3 .....A. ......GT.- -.G.T..T.. ....A.A... ..AA..CACC ......C-.A
EMU 4 .....A.. ......GT.- -...T..T.. ....A.A... ..AA..T.CC ..AT..CAAT
EMU 5 .A..... ......GT.- -G..T..T.. .C..A.A... ..AAG.CAAC ....CACCAA
170 150 130
CHICK CCTGGGCAGC CTGGTCTAGT AYTGTGTGGA GGYTGATGGC CCTGCCCATA GCAGGGGGGG
CRANE TT......A. ...A..... -------... ..G..T---- .....-- ...C.C----
EMU 1 ....A...A. G..C.....G ---------- -------..A. ......TGAG T.G.------
EMU 2 .....-. .TAC.C.--- ---------- ---A.G..A. .T...TTGA. .....-----
EMU 3 T..T------ ------GT.C .A....CTCT A.GC..---. ...A.-TTG. A.....----
EMU 4 T...T..--- ---------- .A....1C2 A.G....--. ...TGAGCAG ----------
EMU 5 ....------ -------TTC .A....CTCT A.G..----- -....-TTG. ......----
110 90 70
2
3'
5'
3'
CHICK GGGTTGAAGC TTGGTGATCC TTGAGGTCCC TTCCAACCCA GGCCATTCTA TGATTCTATG
CRANE --....G.A. .-.A.....- T..... ..... AA..... .....
EMU 1 A.....G... A-.A.....TCCA..... .....TC A-.T.C-..G .....G.A
EMU 2 .....CGGA. .GC..... -A..... .....TC AA.A.....G ...CCT.GAA
EMU 3 ......G.-. .A.A.....TCCA..... ..... AA..... ..
EMU 4 -----.G..- ....ACTAGA TA
EMU 5 ....A.GC.. .A.A.....TCCA..... ..... AA..... ..
50 30 10 1

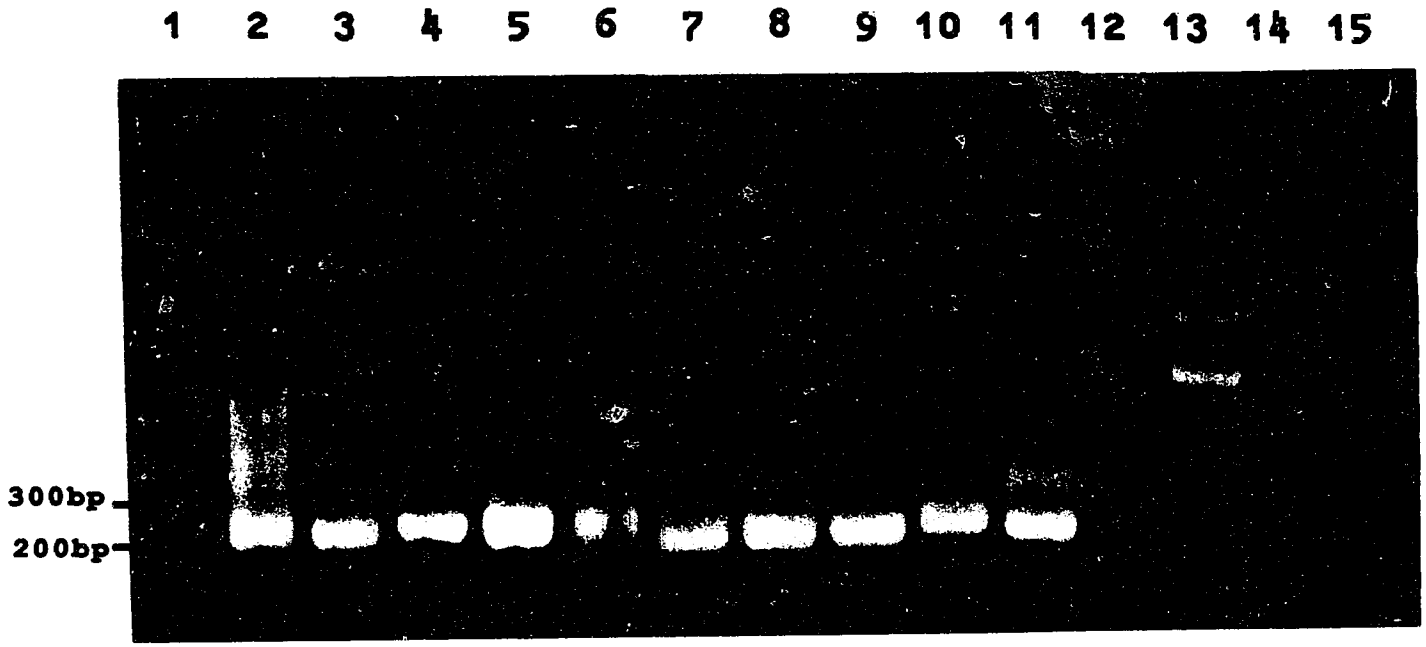
Figure III-3. a) Autoradiograph of a Southern hybridization of ^{32}P labeled SC105 with *Eco*RI digests of 5 µg genomic DNA samples isolated from emu (Lane 1), cassowary (2), pelican (3), stork (4), condor (5), quail (6), crane (7), owl (8) and magpie (9). b) the blot in a) was stripped and re-probed with the *Hin*dIII/*Eco*RI fragment of SC10 (see Figure III-1) which does not include the CR1 element, and c) the blot in b) was then stripped and probed with labelled DNA fragments isolated from the major band of the emu PCR (see Figure III-4).


1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9
1 2 3 4 5 6 7 8 9
kb
4
3
2
1
a
b
c

When the SC105 probe was hybridized to a Southern blot containing human, mouse, newt, frog and fish DNA samples, no homology was detected in any of these cases (data not shown).

The failure to detect cross homology between the crane CR1 repeat and a comparable sequence in either the emu or cassowary might simply reflect the high stringency of the hybridization conditions (Chen et. al., 1989). Since low stringency hybridizations using repetitive DNA probes are liable to give false positives, we felt that the application of PCR might reveal the presence of the CR1 family in these primitive birds. Two nucleotide oligomers, homologous to the crane CR1 sequence (see Figure III-2), with appropriate G-C content and spanning most of the element, were synthesized and used as primers. PCR reactions were performed on DNA samples of ten avian species and four non-avian species, and the amplification products were examined on 2% agarose gels. As shown in Figure III-4, all the avian samples (lanes 2 to 11), including the emu and cassowary, had a predominant band of about 250 bp and several other discrete bands. However, the four non-avian samples (three lower vertebrates and an invertebrate) all lacked the major band (lanes 12 to 15). Discrete amplification products in addition to the major band were produced in the avian samples as well as in the non-avian samples. While several explanations present themselves, further analysis is required to distinguish among them.

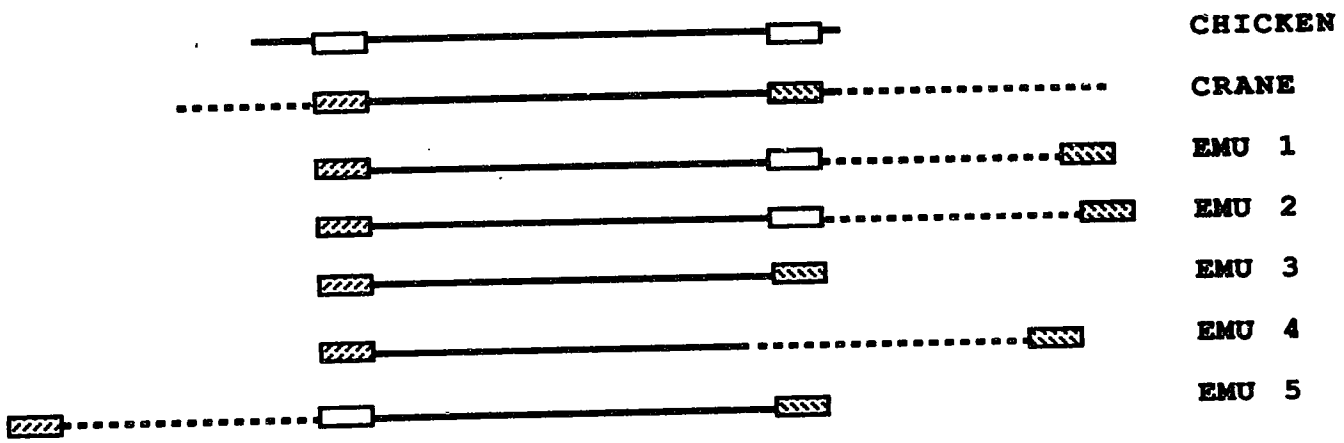
Figure III-4. An ethidium bromide stained agarose gel (2%) on which aliquots of PCR products using various templates were loaded. Lane 1 contains 1 kb ladder DNA (GIBCO/BRL) as a molecular weight marker. Lanes 2 to 15 contain samples obtained from, in order, quail, shelduck, cormorant, cassowary, emu, condor, magpie, owl, pelican, stork, newt, frog, fish and lobster.


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
300bp
200bp

Conclusive evidence that the major bands seen in the avian PCR samples were genuine CR1 amplification products comes from Southern hybridizations and DNA sequencing. When the DNA fragments in the emu's major band (as shown in Figure III-4) were used to reprobe the Southern blot presented in Figure III-3a and 3b, a strong hybridization signal was seen in both the emu and cassowary lanes (Figure III-3c). Homology was also detected between the emu probe(s) and the other avian species with the exception of the magpie. The CR1 homology between the emu and the crane, shown in Figure III-3c, was not apparent in Figure III-3a. Thus, the CR1 element in pSC10 has apparently diverged significantly from emu elements, although among the emu elements comprising the PCR generated probe, there exist some with detectable homology to some crane CR1 members. Since the emu probe only represents a subset of CR1 family members in the emu genome, the implication of its failure to detect homology in the magpie DNA requires further study.

The mixture of PCR amplified fragments from the emu was cloned into pUC19 and DNA sequences were obtained from five randomly chosen clones. Sequences of clone 3 were confirmed on both strands, while sequences of the other four clones were obtained on one strand only. Figure III-5 shows a schematic comparison of these DNA sequences with the crane and chicken CR1. Emu clone 3 is 258 bp in length, and both its ends are located within a CR1 domain that is comparable to the crane and chicken CR1. Emu clones 1, 2, 4 and 5 are

Figure III-5. Schematic comparison of the chicken CR1 with the crane CR1, 3C105, and five emu clones. The solid lines represent CR1 elements. The broken lines indicate non-CR1 sequences. A hatched box indicates sequences matching perfectly with one of the primers used for PCR, whereas an open box shows some divergence from the primer.


CHICKEN
CRANE
EMU 1
EMU 2
EMU 3
EMU 4
EMU 5

each about 600 bp in size. Though they all have a region of about 250 bp similar to clone 3, priming has occurred within the CR1 homologous region at only one of the two sites. The cloned DNA fragments extend beyond the mismatched sites to a suitable template in flanking DNA with which the primer could form a stable hybrid. The regions in these clones beyond the CR1 share no homology to each other, nor to the crane or chicken CR1, indicating that the emu CR1 units were resident at different sites in the genome.

In comparing DNA sequences of the emu CR1 elements (Figure III-2) with the crane and chicken CR1s, two highly conserved regions are evident: the first between positions 23-38 and the second between 165-194. Elsewhere, substantial divergence has occurred in the form of base substitutions and, particularly between positions 80-115, as deletions of nucleotides in the emu elements relative to the other two.

DISCUSSION

We have applied PCR to detect interspecific homology of a repetitive DNA family. Our results, some of which were unexpected, point to its potential use in the study of genome evolution. For each of the avian samples we have examined, the amplification products formed a pronounced band 200 to 300 bp long (Figure III-4, lanes 2-11). By contrast, this band was absent in the amplification products from the non-avian samples (lanes 12-15). The CR1-like templates in the emu and cassowary DNA were not predicted from the Southern

hybridization experiment (Figure III-3a), which failed to detect cross homology between these species and the crane CR1 probe. Because of the discrepancy between the results from these two experiments, it was essential to confirm that the amplification band from the emu and cassowary samples was indeed CR1-related. When the DNA fragments in the major emu amplification band were labeled and used to reprobe a Southern blot which had previously been hybridized with the crane CR1 probe (Figure III-3a), both emu and cassowary (Figure III-3c, lanes 1, 2) showed a strong signal. This contrasts with the weak signal seen in these samples when the crane CR1 repeat was used as a probe. Secondly, the cloning and sequencing of individual amplified DNA sequences from the emu sample provide direct evidence that these amplified products belong to the CR1 family (Figure III-2).

All five emu clones were randomly chosen and were CR1-like. However, the Southern analysis (Figure III-3a) failed to detect a conclusive hybridization signal in the emu sample. This is likely due to the limited sequence homology between emu CR1 elements and the crane CR1 probe. PCR, by comparison to the routine DNA-DNA hybridization, was clearly a more sensitive method for detecting interspecific sequence homology. The reason for this is that the probability of there being at least one amplifiable template among the collection of CR1 units in the emu is very high, given the length of the two crane primers that were used. By contrast, the average composition of the family members diverges so

substantially between the crane and the emu that the homology was not observed in a Southern using a single CR1 unit as a probe.

A comparison of the CR1 sequences obtained from the emu genome with the ones obtained from the genomes of the chicken and crane revealed two highly conserved regions within the CR1: a region between positions 23-38, and one between positions 165-194 (Figure III-2). By contrast, the remaining part of the CR1 diverged a great deal between these species. It is very unlikely that this high degree of sequence conservation occurred by chance alone in such distantly related species as the emu and chicken. In fact, it has already been pointed out that these regions were also well conserved among the chicken CR1 members (Stumph et. al., 1984; Silva and Burch, 1989). Most interestingly, a nuclear protein(s) binds specifically at the region between positions 23-38 (Sanzo et. al., 1984). Moreover, an *in vitro* study has shown that the region between positions 165-194 can serve as a silencer of transcription (Banlahamad et. al., 1987). It is most likely that the conserved domains of each CR1 member have been monitored by functional constraints, and sequences elsewhere diverge between family members and between different species. These features of the CR1 family, the high degree of divergence of each member from the consensus sequence concomitant with interspecifically conserved regions within these DNA sequences that interact with specific

proteins, may shed light on the question of why the CR1 has been maintained in the avian genome.

In fact, the conserved regions are very reliable identification tags for the CR1 family, and their existence is of practical significance: since PCR products might contain artefacts generated during amplification (Dunning et. al., 1988; Eckert and Kunkel, 1990), probes made from these conserved regions should permit direct isolation of CR1 members from the avian genome without the use of PCR. Because of the lack of well preserved avian fossils, avian taxonomy has been a very challenging task. It is not uncommon that a species has been placed in different orders of the class Aves by different taxonomists (Hendrickson, 1969). Clearly, the CR1 family, with thousands of copies in each genome and its apparently ubiquitous distribution in all avian species, is very promising system with which to study avian speciation.

The fact that the CR1 family exists in a wide variety of avian genomes including the primitive species, emu, indicates that this family has experienced a long evolutionary history. An ancient origin of this family is also revealed by the high degree of DNA sequence divergence (other than the conserved regions) among CR1 members within a species (Stumph et. al., 1984) as well as between species (Figure III-2). These differences can be quantitated (Li et. al., 1981), and with the accumulation of DNA sequence data from different species, it will be possible to re-examine existing phylogenic

relationships, assuming sequence variation at these positions is under neutral selection (Maeda et. al., 1988).

REFERENCES:

Baniahamad A., Muller M., Steiner C. and Renkawitz R. 1987. Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. EMBO J. 6: 2297-2303.

Britten R. J. 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science. 231: 1393-1398.

Britten R. J. and Davidson E. H. 1969. Gene regulation for higher cells: a theory. Science. 165: 349-357.

Britten R. J. and Kohn D. E. 1968. Repeated sequences in DNA. Science. 161: 529-540.

Chen Z.-Q., Lin C. C. and Hodgetts R. B. 1989. Cloning and characterization of a tandemly repeated DNA sequence in the crane family (Gruidae). Genome. 32: 646-654.

Cracraft J. 1974. Phylogeny and evolution of the ratite birds. Ibis 116: 494-521.

Davidson E. H. and Britten R. J. 1979. Regulation of gene expression; possible role of repetitive sequences. Science 204: 1052-1059.

Doolittle W. F. and Sapienza C. 1980. Selfish genes, the phenotype paradigm and gene evolution. Nature. 284: 601-603.

Dunning A. M., Talmud P. and Humphries S. E. 1988. Errors in the polymerase chain reaction. Nucleic Acids Res. 16: 10393-10393.

Eckert K. A. and Kunkel T. A. 1990. High fidelity DNA synthesis by the Thermus aquaticus DNA polymerase. Nucleic Acids Res. 3739-3744.

Hache R. J. G. and Deeley R. G.. 1988. Organization, sequencing and nuclease hypersensitivity of repetitive elements flanking the chicken apo VLDII gene: extended sequence similarity to elements flanking the chicken vitellogenin gene. Nucl. Acid. Res. 16: 97-113.

Hendrickson H. T. 1969. A comparative study of the egg white proteins of some species of the avian order gruiformes. Ibis. 11: 80-91.

Li W. H., Gojobori T. and Nei M.. 1981. Pseudogenes as a paradigm of neutral evolution. Nature. 292: 237-239.

Maeda N., Wu C. I., Bliska J. and Reneke J. 1988. Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. Mol. Biol. Evol. 5: 1-20.

Maniatis T., Fritsch E. F. and Sambrook J. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Orgel L. E. and Crick F. H. C. 1980. Selfish DNA: the ultimate parasite. Nature. 284: 604-607.

Sanzo M., Stevens B., Tsai M. J. and O'Malley B. W. 1984. Isolation of a protein fraction that binds preferentially to chicken middle repetitive DNA. Biochem. 23: 6491-6498.

Schip F. V. H., Samallo J., Meijilink F., Gruber M. and Geert A. B. 1987. A new repetitive element of the CR1 family downstream of the chicken vitellogenin gene. Nucl. Acid. Res. 15: 4193-4202.

Silva R. and Burch J. B. 1989. Evidence that chicken CR1 elements represent a novel family of retroposons. Mol. Cell. Biol. 9: 3563-3566.

Stumph W. E., Hodgson C. P., Tsai M. J. and O'Malley B. W. 1984. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. Proc. Natl. Acad. Sci. USA. 81: 6667-6671.

Stumph W. E., Kristo P., Tsai M. J. and O'Malley B. W. 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. Nucl. Acid. Res. 9: 5383-5397.

Voliva C. F., Jahn C. L., Comer M. B., Hutchison III C. H. and Edgell M. H. 1983. The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. Nucl. Acids. Res. 11: 8847-8859.

CHAPTER IV. DISCUSSION

In this thesis, I have described two DFs in the crane genome, the tandemly clustered *Pst*I family and the interspersed CR1 family. Their conservation in other avian species was also examined. I have shown that the *Pst*I family comprises about 0.1% of the crane genome. The estimated size of the CR1 family is about 7000-20000 members per haploid chicken genome (Stumph et. al., 1981; Hache and Deeley, 1988). If these numbers are also applicable in cranes, then it equals about 0.1-0.3% of the crane genome. These two families likely represent only a small proportion of the DFs in the crane genome. As described previously (Rosen et. al., 1973), approximately 30% of the avian genome is comprised of repetitive DNA. Numerous questions concerning DFs in the avian genome remain to be explored.

The *Pst*I family

In Chapter II, I have reported the cloning and characterization of the *Pst*I family. The existence of this family in the crane genome was initially revealed by a ladder of bands shown on a background smear after gel electrophoresis of *Pst*I digests of crane genomic DNA. By Southern hybridization of the digests with cloned *Pst*I family members, I demonstrated that these bands comprised the *Pst*I family. According to Singer (1982), a DF would appear on a

gel as a ladder only if it comprised more than 0.5% of a genome, and the intensity of the Pst ladder shown on the agarose gel (Figure II-1) seems to be consistent with the statement. However, my estimation on the *Pst*I family based on DNA-DNA hybridization data (Figure II-4) was only 0.1% of the crane genome. One possibility to account for this significant discrepancy is that my estimation of the *Pst*I family size was a lower limit, since under the high stringency condition used, those elements which diverged from the sequence of the probe would have given a reduced hybridization signal. Another possible explanation is that this Pst ladder may contain undefined DFs which also have monomer units of about 200 bp in size and conserved *Pst*I sites among some of their members. This can be further studied by screening the mini library, constructed from DNA fragments isolated from the ladder, with the existing *Pst*I elements as probes. An alternative way to determine the size of the Pst array is by large scale restriction mapping (Willard, 1989). A rare-cutting restriction enzyme (such as *Not*I) could be found that digests the Pst array into a few large fragments. These could be resolved by pulse field electrophoresis and visualized on a Southern blot by probing the blot with a PstI family member (such as SC2). The physical size of the array could be estimated from the cumulative size of these bands.

Of the five *Pst*I elements I have examined, SC1, 2, 6 and 8 share a high degree of homology with the consensus

sequence, whereas SC9 diverges from it by more than 25%. Apparently, SC9 represents a subgroup of the *Pst*I family which differs from the one including SC1, 2, 6 and 8. The chromosome *in situ* hybridization results showed that the SC2-like sequences are clustered at the centromeric region of chromosome 2 of the sarus crane. It would be interesting to discover where SC9-like sequences reside, as variants of a given satellite DNA family may be located on different chromosomes, exemplified by the human alphoid satellite DNA (Devilee et. al., 1986; Jabs and Persico, 1987).

Centromeric satellite sequences can be highly polymorphic, and consequently the polymorphism may be used to determine the degree of genetic relatedness among individuals in a population. Recently, it has been reported that, with *Hind*III restriction alone, a falcon centromeric tandem repeat probe could distinguish 16 of a random sample of 18 individuals on the basis of unique hybridization patterns (Longmire et. al., 1988). A similar study on human samples showed that with two restriction enzymes and pulse field gel electrophoresis, all 24 unrelated human samples were distinguishable on the basis of their hybridization patterns with an alphoid satellite DNA probe (Jabs et. al., 1989). Both studies also demonstrated that such DNA RFLPs behaved as codominant Mendelian alleles. It will be interesting to see whether the cloned *Pst*I family elements will detect DNA RFLPs among a crane population, since solving paternal and maternal relationships in the population is very important for

wildlife management and behavioural studies (Burke and Bruford, 1987; Wetton et. al., 1987; Quinn et. al., 1987), especially on endanged species such as the whooping crane.

My preliminary results showed that the *Pst*I-like family is present in condor and quail, species distantly related to crane, but absent in rail, a species closely related to crane. This is the only case reported in avian species where the distribution of satellite DNA showed a discrepancy with traditional classification. However, a similar observation has been reported in plants, where the "350-family", a repetitive DNA family prominent in *Agropyron cristatum* (the P genome), is found to be absent from a number of the Agropyron species, but this DNA family was originally discovered in the distantly related rye (*Secale cereale*), the R genome (Xin and Appels, 1988). These observations were explained by the saltatory theory which states that an existing DF results from the amplification of randomly chosen DNA sequences from a set of ancestral candidates. This theory explains how two genomes of distantly related species may have picked up similar sequences to generate DNA families by chance alone, but it is hard to imagine how the homologous sequences in other species have also been completely lost by chance. I suggest that horizontal transmission might be the cause of these phenomena. A possible mechanism could be a retroviral infection of distantly related genomes, although this has not been investigated as yet. It is noteworthy that a region in calf 1.706 satellite sequence has several features

reminiscent of transposable elements (Calos and Miller, 1980) including terminal direct and inverted repeats (Streeck, 1981).

In developing our ideas about a possible function for satellite DNA, we were struck by the fact that nucleosomes are phased on satellite DNA and the lengths of a large number of the tandemly arrayed DNA families are conserved. We speculated that there is a requirement for phased nucleosomes in centromeric domains of higher eukaryotic chromosomes to ensure chromosome stability at metaphase. At the time of publication of our paper, we were not aware that there was some evidence supporting our notion. Vig and Zinkowski (1985) showed a clear correlation between the amount of C-banding (containing mainly satellite DNA) in mouse chromosomes and the time required for sister chromatid separation. Furthermore, direct evidence came from Lica et al. (1986) who, in an electron microscopic study on a mouse marker chromosome, observed that each of the sister chromatid contact sites is accompanied by the presence of mouse satellite DNA. They also discovered that such sister chromatid contacts were disrupted when the satellite DNA was degraded by restriction enzymes.

The CR1 family

In Chapter III, I have reported the cloning and characterization of the CR1 family in the crane. The CR1

containing clone SC10 was isolated from a crane DNA library by a systematic screening procedure. At present CR1 is the sole DNA family known to be widely conserved in the avian genome, but it does not exclude the possibility that there exist other conserved DFs in the genome. The screening procedure described in Chapter III, which utilized a test panel of several avian samples to identify clones containing conserved repeats, provides a feasible approach to answer the question.

My DNA sequencing data revealed two highly conserved regions among the CR1 elements found in different species, which suggested functional constraints in their evolution. One of the regions can function as a silencer to down regulate adjacent gene expression: it exerts a strong repressing effect on weak promotors, but has a weak effect on strong transcription units (Baniahmad et. al., 1987). If a considerable fraction of the CR1 family members serve such a function, then they create a repressive genomic environment with respect to transcription, and thus contribute to a genome in which most of the genes are turned off during most of development.

The second conserved region is known to bind a nuclear protein in chicken. Such binding may affect gene expression through alteration of local chromatin configuration (Sanzo et. al., 1984), or may well be the binding between scaffold protein and matrix attachment site (Phi-Van and Stratling, 1988). In any case eventual understanding of the CR1

function requires the isolation and further characterization of this CR1 binding nuclear protein (Sanzo et. al., 1984).

My data showed that other than the two conserved regions, DNA sequences among some CR1 members (such as SC105 and the emu CR1 elements) diverge to such an extent that they fail to hybridize in a Southern analysis. This differs from what has been observed with the Alu family in higher primates: the Alu family members diverge only 4-14% from the consensus sequences of the primates (Deininger and Daniels, 1986). The different degree of divergence seen in these two DNA families may indicate that CR1 has had a much longer history as a repeat family than the Alu. The Alu family is mainly confined to primates and a related Alu-like family in rodents (Krayev et. al., 1980; Haynes et. al., 1981). Perhaps the generation of the Alu family happened long after the mammals had diverged. By comparison, the appearance of CR1 in the avian genome may have occurred before or during the early stage of the divergence of the class Aves. My speculation on the ages of these two DFs is consistent with the morphological data which suggest that the divergence of primates from rodents took place about 60-70 million years ago, whereas the early speciation of the class Aves occurred about 80-90 million years earlier (Nei, 1987).

CR1 has the potential of being a useful DNA RFLP system for genotyping avian species. It is likely that DNA sequence from a highly variable region in the CR1, such as between positions 80 to 115 (see Figure III-2), would only hybridize

to a limited number of CR1 elements in a genome. Together with carefully chosen restriction enzymes which have no recognition sites within CR1, a set of discrete bands with diagnostic significance might be expected.

It has been speculated that CR1 is a novel transposon, and may code for a protein product (Silva and Burch, 1990). With current knowledge on CR1 sequences, one should be able to isolate full length CR1 members and /or "master copies" by, for example, uni-directional "chromosome walking". Only then will one be able to address the questions of whether CR1 does encode a protein, and whether this protein is involved in the transposition of CR1 elements.

REFERENCES

Baniahamad A., Muller M., Steiner C. and Renkawitz R. 1987. Activity of two different silencer elements of the chicken lysozyme gene can be compensated by enhancer elements. EMBO J. 6: 2297-2303.

Biederman B. M., Lin C. C., Kuyt E. and Drewien R.C. 1982. Genome of the whooping crane. J. Heredity 73: 145-146.

Burk T. and Bruford M. W. 1987. DNA fingerpringting in birds. Nature 327: 149-152.

Calos M. P. and Miller J. H. 1980. Transposable elements. Cell 20: 579-595.

Deininger P. L. and Daniels G. R. 1986. The recent evolution of mammalian repetitive DNA elements. Trends Genet. 2: 76-80.

Devilee P., Cremer T., Slagboom P., Bakker E., Scholl H. P., Hager H. D., Stevenson A. F. G., Cornelisse C. J. and Pearson P. L. 1986. Two subsets of human alphoid repetitive DNA show distinct preferential localization in the pericentric regions of chromosomes 13, 18, and 21. Cytogenet. Cell. Genet. 41: 193-201.

Hache R. J. G. and Deeley R. G. 1988. Organization, sequencing and nuclease hypersensitivity of repetitive elements flanking the chicken apo VLDII gene: extended sequence similarity to elements flanking the chicken vitellogenin gene. Nucl. Acid. Res. 16: 97-113.

Haynes S. R., Toomey T. P., Leinwand L. and Jelinek W. R. 1980. The chinese hamster Alu-equivalent sequence: a conserved, highly repetitious interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposabnle element. Mol. Cell. Biol. 1: 573-583.

Jabs E. W. and Persico M. G. 1987. Characterization of human centromeric regions of specific chromosomes by means of alphoid DNA sequences. Am. J. Hum. Genet. 41: 374-390.

Jabs E. W., Goble C. A. and Cutting G. R. 1989. Macromolecular organization of human centromeric regions reveals high-frequency, polymorphic macro DNA repeats. Proc. Natl. Acad. Sci. USA 86: 202-206.

Krayev A. S., Kramerov D. A., Skryabin K. G., Ryskov A. P., Bayev A. A. and Georgiev G. P. 1980. The nucleotide

sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back1 RNA. Nucleic Acids Res. 8: 1201-1215.

Lica L. M., Narayanswami S. and Hamkalo B. A. 1986. Mouse satellite DNA, Centromere structure, and sister chromatid pairing. J. Cell Biol. 103: 1145-1151.

Longmire J. L., Lewis A. K., Brown N. C., Buckingham J. M., Clark L. M., Jones M. D., Meincke L. J., Meyne J., Ratliff R. L., Ray F. A., Wagner R. P. and Moyzis R. K. 1988. Isolation and molecular characterization of a highly polymorphic centromeric tandem repeat in the family falconidae. Genomics 2: 14-24.

Nei M. 1987. In: Molecular evolutionary genetics. pp8-11. Columbia University Press. New York.

Phi-Van L. and Stratling W. H. 1988. The matrix attachment regions of the chicken lysozyme gene co-map with the boundaries of the chromatin domain. EMBO J. 7: 655-664.

Quinn T. W., Quinn J. S., Cooke F. and White B. N. 1987. DNA marker analysis detects multiple maternity and paternity in single broods of the lesser snow goose. Nature 326: 392-394.

Rosen J. M., Liarakos C. D. and O'Malley B. W. 1973. Effect of estrogen on gene expression in the chicken oviduct. I. renaturation studies. Biochemistry 12: 2803-2809.

Singer, M. F. 1982. Highly repeated sequences in mammalian genomes. Int. Rev. Cytol. 76: 67-112.

Sanzo M., Stevens B., Tsai M. J. and O'Malley B. W. 1984. Isolation of a protein fraction that binds preferentially to chicken middle repetitive DNA. Biochemistry 23: 6491-6498.

Streeck R. E. 1981. Inserted sequences in bovine satellite DNA's. Science 213: 443-445.

Stumph W. E., Kristo P., Tsai M. J. and O'Malley B. W. 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. Nucl. Acid. Res. 9: 5383-5397.

Vig B. K. and Zinkowski R. P. 1985. Sequence of centromere separation: influence of pericentric heterochromatin in *Mus*. Genetica (The Hague). 67: 153-159.

Wetton J. H., Carter R. E., Parkin D. T. and Walters D. 1987. Demographic study of a wild house sparrow population by DNA fingerprinting. Nature 327: 147-149.

Xin Z.-Y. and Appels R., 1988. Occurrence of rye (Secale cereale) 350-family DNA sequences in Agropyron and other Triticeae. Pl. Syst. Evol. 160: 65-76.