

GENOMIC EVOLUTION AND DIVERSITY IN ARTIODACTYLA

Silja Kostia

Department of Biosciences, Division of Genetics,
University of Helsinki

Academic Dissertation in Genetics

To be presented for public criticism, with the permission of the Faculty of Science,
University of Helsinki, in the auditorium of the Department of Ecology and Systematics at
Pohjoinen Rautatiekatu 13, on May 31st, 2000, at 12 o'clock noon.

Helsinki 2000

ISSN 1239-9469
ISBN 952-91-2116-4 (nid.)
ISBN 952-91-2117-2 (PDF version)
Helsingin yliopiston verkkojulkaisut
Helsinki 2000

To Finnish Sisu

Cover figure: White Finnsheep and a black one. Photo: Dr Juha Kantanen.

Table of contents

ORIGINAL PUBLICATIONS	1
ABBREVIATIONS	2
I. INTRODUCTION AND AIMS OF THE STUDY	3
II. REVIEW OF LITERATURE	4
1. REPETITIVE DNA AS A SOURCE FOR GENOMIC AND EVOLUTIONARY MARKERS	4
1.1. Short interspersed nuclear elements (SINEs)	4
1.1.1. The origin of SINEs	4
1.1.2. Bov-tA and Bov-A2 SINEs and Bov-B LINE	4
1.1.3. Distribution of SINEs and LINEs in ruminants	6
1.2. Simple sequence repeats (SSR)	7
1.2.1 The origin of SSRs	7
1.2.2. Microsatellite evolve through errors during the replication	7
1.2.3. Bovine microsatellites	8
1.3. Interspersed repetitive sequence (IRS)-PCR	8
1.3.1. Preparative IRS-PCR	8
1.3.2. IRS-PCR generate multilocus fingerprints	9
1.4. SINE insertion sites are powerful phylogenetic markers; relationships among cetartiodactyls as an example	9
2. NUMTS; PSEUDOGENES OF MITOCHONDRIAL ORIGIN	10
2.1. The origin of numts	10
2.1.1. A DNA or RNA mediated process?	10
2.1.2. Integration by end-joining mechanism	10
2.2. The genomic organization of numts	11
2.2.1. Several numts of different age	11
2.2.2. Several recent integrations in closely related taxa	11
2.2.3. Amplification of numts	12
2.3. Evolution of mitochondrial DNA	12
2.4. The ETAS domain include the putative terminaton signal for mtDNA replication	13

3. POLYMORPHISM IN THE MAJOR HISTOCOMPATIBILITY COMPLEX (MHC)- DRB GENE OF DOMESTIC CATTLE AND SHEEP	14
3.1.MHC of cattle and sheep	14
3.2.Origin and maintenance of MHC diversity	16
3.2.1. Trans-species origin of MHC alleles	16
3.2.2. Intraexonic sequence exchanges (gene conversions)	16
3.2.3. Balancing selection is operating on MHC loci	17
3.3.DRB typing methodology	17
3.3.1. Before the PCR era	17
3.3.2. Exon 2 based typing	18
3.3.3. Microsatellite based <i>DRB</i> typing	18
3.4.Association of MHC haplotypes and resistance	18
3.5.MHC and conservation of endangered populations	19
III. MATERIALS AND METHODS	20
IV. RESULTS AND DISCUSSION	21
1. SINE TARGETING OF BOVINE MICROSATELLITE MARKERS	21
2. INTER-SINE AND INTER-SSR FINGERPRINTS IN SIX ARTIODACTYL SPECIES	21
3. DISTRIBUTION AND EVOLUTION OF BOV-tA SINE IN THE ARTIODACTYLA	23
4. MULTIPLE INTEGRATIONS OF THE MITOCHONDRIAL CONTROL REGION INTO THE NUCLEAR GENOME OF RUMINANTS	23
4.1. Integration history of five numts of mitochondrial control region	23
4.2. Artnumt pseudogene evolution	24
4.3. Structural features of the ETAS domain	25
5. MHC CLASS II DRB DIVERSITY OF FINNCATTLE AND FINNSHEEP	25
5.1. Heterozygosity excess in Finnish cattle breeds	25
5.2. <i>Ovar-DRB1</i> diversity of Finnsheep	26
5.3. <i>DRB</i> diversity and endangered native breeds	26
5.4. SSCP analysis in <i>DRB</i> typing	26
V. SUMMARY AND CONCLUDING REMARKS	28
ACKNOWLEDGEMENTS	30
REFERENCES	31

ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, referred to in the text by their Roman numerals.

- I. Kostia S., Vilkki J., Pirinen M., Womack J.E., Barendse W. and Varvio S.-L. (1997). SINE targeting of bovine microsatellites from bovine/rodent hybrid cell lines. *Mammalian Genome* 8: 365-367.
- II. Kostia S., Ruohonen-Lehto M., Väinölä R. and Varvio S.-L. (2000). Phylogenetic information in inter-SINE and inter-SSR fingerprints of the Artiodactyla and evolution of the Bov-tA SINE. *Heredity* 84:37-45.
- III. Kostia S., Varvio S.-L. and Väinölä R. (2000). Multiple integrations of the mtDNA control region into the nuclear genome of ruminants. Manuscript.
- IV. Kostia S., Kantanen J., Kolkkala M. and Varvio S.-L. (1998). Applicability of SSCP analysis for MHC genotyping: fingerprinting of *Ovar-DRB1* exon 2 alleles from Finnish and Russian breeds. *Animal Genetics* 29: 453-455.
- V. Kostia S., Kantanen J., Kulju K. and Varvio S.-L. (2000). Extensive *BoLA-DRB3* diversity in endangered native Finncattle, and heterozygosity excess in Finnish cattle breeds revealed by SSCP and sequence analysis of exon 2. Submitted.

ABBREVIATIONS

ARE1P, -2P	Artiodactyl repeat element in porcine
ARS	Antigen recognition site
BCS	Bovine consensus sequence
BDDF	Bovine dimer-driven family
BDF	Bovine dimeric family
BMF	Bovine monomeric family
BoLA	Bovine leucocyte antigen system
bp	Base pairs
CCR	Conserved central region
Chr	Chromosome
CSB	Conserved sequence block
D-loop	Displacement loop
ETAS1, -2	Extended termination associated sequence
H	Heavy strand
H-2	Mouse leucocyte antigen system
HLA	Human leucocyte antigen system
IRS-PCR	Interspersed repetitive sequence PCR
kDa	Kilo dalton
LINE	Long interspersed nuclear element
Mbp	Mega base pairs
MHC	Major histocompatibility complex
Mhc-Ovar	Sheep leucocyte antigen system
MtDNA	Mitochondrial DNA
Myr	Million years
NJ	Neighbor joining
Numt	Nuclear mitochondrial DNA segment
O _H	Origin of heavy strand replication
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PSO	Polymorphism specific oligonucleotide
PRE-1	Porcine repetitive element 1
QTL	Quantitative trait loci
RFLP	Restriction fragment length polymorphism
RTase	Reverse transcriptase
SINE	Short interspersed nuclear element
SSCP	Single strand conformational polymorphism
SSR	Simple sequence repeat
TAS	Termination associated sequence
YAC	Yeast artificial chromosome

I. INTRODUCTION AND AIMS OF THE STUDY

Mammalian genome diversity and underlying evolutionary processes are objects of modern comparative genomics. A dominant portion of the 3000 Mbp of hereditary information in the mammalian genome consists of non-coding DNA. The genes, i.e. protein coding DNA only covers about 3 % of the genome. A large part of the non-coding DNA is repetitive. Simple sequence repeats (SSRs) or microsatellites are tandemly organised, while short interspersed nuclear elements, SINEs, are the best known group of interspersed repetitive DNA. Both SSRs and SINEs offer powerful tools for genome analysis by providing several classes of single and multilocus DNA markers and approaches.

Pseudogenes are another class of non-coding DNA. Nuclear mitochondrial DNA segments, numts, are pieces of mitochondrial genome that have been moved from the cytoplasm to the nucleus and integrated into the chromosomal DNA. In addition to providing molecular markers for different purposes, numts as ‘molecular fossils’ can give information about the evolution of the mitochondrial genome.

Non-coding DNA can be considered as selectively neutral hereditary material. In some cases, changes in the protein coding DNA can be favoured by selection. Major histocompatibility complex (MHC) genes encode proteins involved in pathogen resistance and maintain an unusually high genetic diversity. MHC genes are sources of

highly polymorphic molecular markers.

This review of the literature forms an introduction to the origin and evolution of SINE, SSR, numt and MHC diversity (Fig. 1) and analysis of genomic variation both in intra- and interspecific level utilizing these DNA markers.

The aims of the present thesis are the following:

1. to assess the applicability of the SINE targeting approach for isolation of bovine microsatellite markers from complex DNA sources (publication I)
2. to study the applicability of interspersed repetitive sequence (IRS)-PCR generated multilocus fingerprints in gaining phylogenetic information (publication II)
3. to analyse the integration history of mtDNA sequences into the nuclear genome of ruminants, to study their evolution as nuclear pseudogenes and the information that these ‘molecular fossils’ can provide about structural evolution of the mitochondrial control region (publication III)
4. to examine MHC *DRB* diversity of Finnsheep, endangered Finncattle and two commercial breeds by SSCP and sequence analysis of exon 2 (publications IV and V)

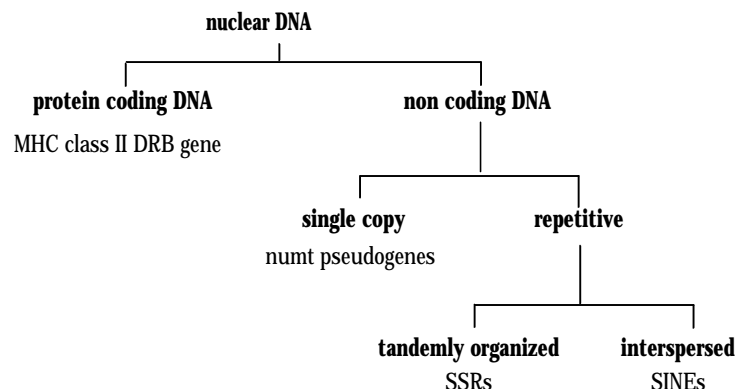


Figure 1. Mammalian genome organization and DNA markers introduced

II. REVIEW OF THE LITERATURE

1. REPETITIVE DNA AS A SOURCE FOR GENOMIC AND EVOLUTIONARY MARKERS

1.1. Short interspersed nuclear elements (SINEs)

1.1.1. The origin of SINEs

SINEs are from 100 bp to 500 bp in length and often present in more than 10^5 copies per genome of multicellular animals from invertebrates to mammals (e.g. Okada 1991; Deininger and Batzer 1993). SINEs have originated via reverse transcription of an RNA intermediate and integration back into the chromosomal DNA, a process which is known as retroposition (Rogers 1985; Weiner et al. 1986). Typical features for SINEs are an internal RNA polymerase III promoter, a poly A- or simple repeat 3'-tail, and a variable length direct repeat at insertion site (e.g. Deininger and Batzer 1993).

Most of the SINEs reported to date, from sources as diverse as mammals and plants, are fusion products of a tRNA-derived and an unrelated sequence. Primate *Alu* and rodent B1 families originate from 7SL RNA (reviewed by Okada et al. 1991; Deininger and Batzer 1993). Recently, Oshima et al. (1996) have proposed that tRNA-related SINEs may have arisen with the aid of long interspersed nuclear elements, LINES. SINEs do not encode the reverse transcriptase (RTase) needed for processing a complementary DNA (cDNA) from the RNA intermediate but may have utilized RTase of the corresponding LINES for retroposition. Turtle and salmon SINEs share a 60-80 bp segment at their 3' end with a particular LINE, in which it most probably represents the recognition site for the RTase

(Eickbush 1992; Smit 1996). Further confirmation for the above described hypothesis (reviewed by Shedlock and Okada 2000) have come from cattle (Okada and Hamada 1997) and from cichlids (Terai et al. 1998) as well as from mammalian-wide interspersed repeat (MIR) SINEs (Smit 1996; Terai et al. 1998)

1.1.2. Bov-tA and Bov-A2 SINEs and Bov-B LINE

Bov-tA and Bov-A2 are the main SINE families of cattle. Bov-tA is a construct of a 115 bp Bov-A monomer and a 73 bp part noted as a tRNA derivative (Rogers 1985; Sakamoto and Okada 1985) and identified as 85 bp by Kaukinen and Varvio (1992). Bov-tA SINE often possess a $(CA)_n$ simple sequence repeat tail (Kaukinen and Varvio 1992). The Bov-A2 includes two Bov-A elements connected by a 27 bp linker sequence comprising hexanucleotide $(CACTTT)_n$ repeats and contains a $(AGC)_n$ trinucleotide repeat tail (Kaukinen and Varvio 1992) (Fig. 2). The A-monomer has also been detected alone (Kaukinen and Varvio 1992). From database entries Lenstra et al. (1993) have estimated that Bov-tA and Bov-A2 are present in 285 000 and 220 000 copies, respectively, in the haploid genome of cattle. Both Bov-A2 (Bov-A) and Bov-tA were first identified by Watanabe et al. (1982), who did not give any specific names for them, but later several names have been used (Table 1). Here, I will use those suggested by Lenstra et al. (1993).

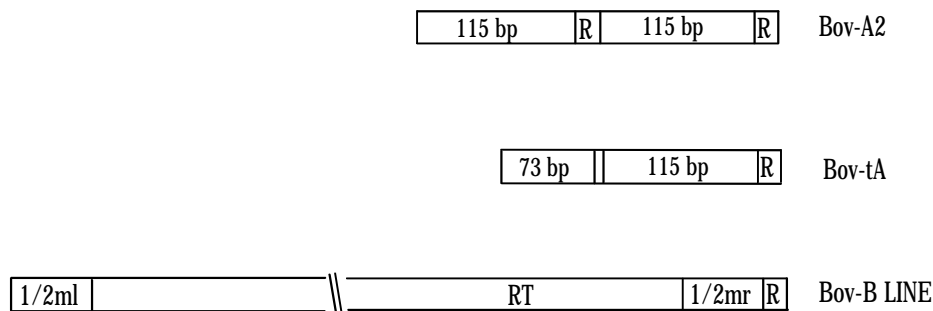


Figure 2. Bov-A2 and Bov-tA SINEs and Bov-B LINE. The 115 bp element is the Bov-A monomer and the 73 bp block a tRNA related part (noted as 85 bp by Kaukinen and Varvio 1992) of the Bov-tA. See II/1.1.2. details regarding repeats (R). Adapted from Okada and Hamada (1997) with modifications.

The Bov-B has been originally described as a SINE (art2, Pst, see Table 1) of 560 bp or less in length constructed by a *Pst*I repeat and a 78 bp segment homologous to Bov-A at the 3' end (Duncan 1987; Majewska et al. 1988). Bov-B was estimated to be present in 60 000 copies in the haploid genome of cattle (Lenstra et al. 1993). Characterization of the first full-length (3.1 kbp) Bov-B (bovine dimer-driven family, BDDF, see Table 1) led to its reclassification as a LINE (Szemraj et al. 1995). Malik and Eickbush (1998) have confirmed that instead of being target sites for Bov-B insertions as proposed by Szemraj et al. (1995), Art2 and Pst are deletions of full length LINES.

The origin of Bov-A has been a mystery since no promoter for polymerase III typical for SINEs has been identified (Rogers 1985; Weiner et al. 1986). According to the hypothesis presented by Okada and Hamada (1997) Bov-A has been generated by deletion of the central part of the Bov-B LINE (step 1 in Fig. 3), and Bov-A2 has arisen by duplication of Bov-A (step 2' in Fig. 3). Then, the Bov-tA has been originated by combining a tRNA pseudogene with Bov-A (step 2 in Fig. 3). Then, the Bov-tA has been originated by combining a tRNA pseudogene with Bov-A (step 2 in Fig. 3).

Table 1. The nomenclature of Bov-A, Bov-A2 and Bov-tA SINEs and Bov-B LINE

SINE/ LINE ⁱ	Name	Reference
Bov-A	D and A	Schon et al. 1981
	A	Rogers 1985
	BCS	Spence et al. 1985
Bov-A2	BDF	Skowronski et al. 1984
	A1-A2	Rogers 1985
Bov-tA	C-D/A	Schon et al. 1981
	BMF	Skowronski et al. 1984
	C-A3	Rogers 1985
	C-BCS	Spence et al. 1985
Bov-B	art-2	Duncan 1987
	Pst	Majewska et al. 1988
	BDDF	Szemraj et al. 1995

ⁱ= names according to Lenstra et al. (1993)

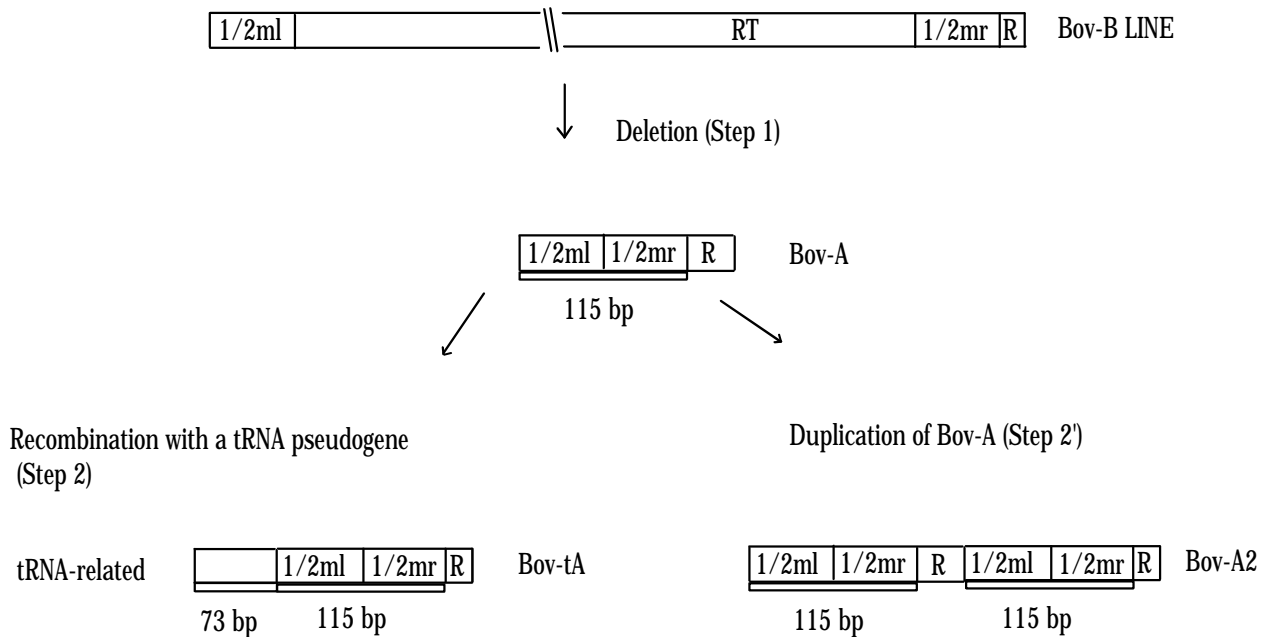


Figure 3. Hypothetical origin of Bov-A and Bov-tA SINEs from Bov-B LINE as presented by Okada and Hamada (1997). Adapted from Okada and Hamada with modifications.

1.1.3. Distribution of SINEs and LINES in ruminants

The Bov-tA SINE has been found in the pecoran and tragulid ruminants (Fig. 4) determined by dot blot hybridization (Shimamura et al. 1999). The Bov-A2 and Bov-B have been detected from the same species as Bov-tA by Southern or dot blot hybridization of genomic DNA (Jobse et al. 1995; Buntjer et al. 1997; Shimamura et al. 1999). The presence of Bov-A as well as Bov-B like sequences in the *Tragulus* was confirmed by PCR amplification of the elements (Buntjer et al. 1997). However, from their hybridization studies Modi et al. (1996) have come to the conclusion that Bov-B is not present in chevrotain (Tragulidae) but must instead have inserted after the divergence of Tragulina and Pecora, which is not in accordance with observations of other authors and the hypothesis by Okada and Hamada (1997) concerning the origin of Bov-A from a Bov-B LINE.

Kordis and Gubensek (1995) have described a 620 bp segment in the viper (*Vipera ammodytes*) genome with a high homology to the cattle art2 SINE (truncated Bov-B LINE) suggesting horizontal transfer between vertebrate classes. However, Malik

and Eickbush (1998) have proposed that a more probable explanation for this nucleic acid homology is that these SINEs encode the carboxyl end of the Bov-B open reading frame (ORF) containing both apurinic-apyrimidic endonuclease and reverse-transcriptase domains characteristic for LINES.

ARE-1P and ARE-2P (Artiodactyl Repetitive Element 1,2-Porcine) SINEs (or LINES, see Nikaïdo et al. 1999) have been recently described in pig, and a search of database entries and hybridization studies have revealed that these elements occur as rare components in the bovine and sheep genomes (Alexander et al. 1995). This was confirmed by Buntjer et al. (1997), who showed that ARE1 and ARE2 are abundantly present in the genomes of artiodactyl species by Southern hybridization of genomic DNA. Other SINEs in the ruminant genomes are CHR-1 (for Cetacea, hippopotamus and Ruminantia) and CHR-2 (Shimamura et al. 1997; 1999).

In addition to enabling estimation of the integration time of a particular SINE family in the artiodactyl lineage, presence/absence of a single SINE and LINE insertions has been used as phylogenetic markers to infer relationships of the cetartiodactyl lineage (see II/1.4).

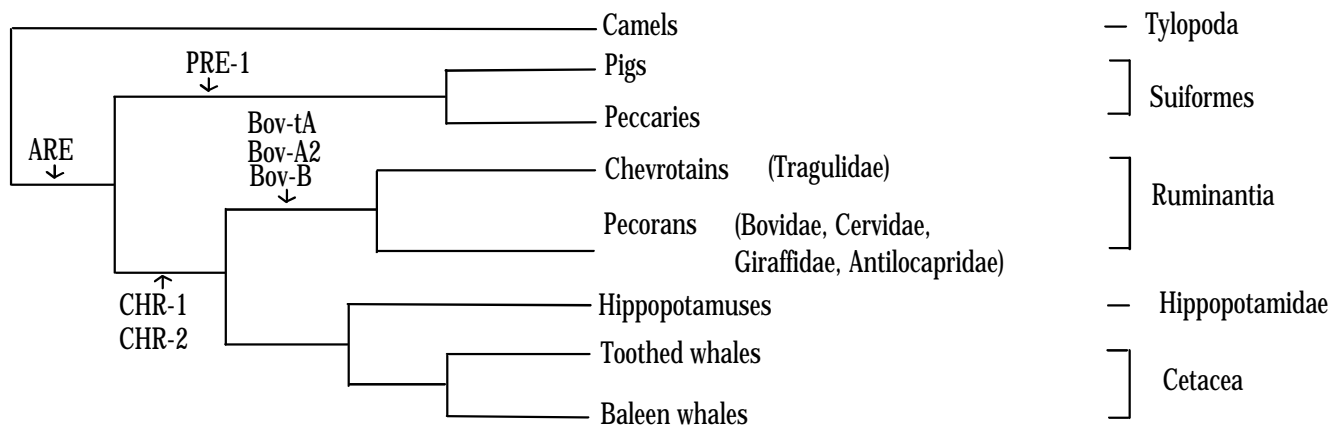


Figure 4. Distribution of Bov-A2, Bov-tA, Bov-B, CHR-1, CHR-2, ARE and PRE-1 elements in the Cetartiodactyl lineage. Adapted from Nikaido et al. (1999) with modifications.

1.2. Simple sequence repeats (SSRs)

1.2.1. The origin of SSRs

Simple sequence repeats or microsatellites are monotonous repetitions of very short (1-6 bp) nucleotide motifs (e.g. Tautz 1989). The (CA)_n/(GT)_n represents the most frequent microsatellite repeat of the mammalian genome, outnumbered only by poly (A)/(T) monomers (Beckmann and Weber 1992). The mammalian copy number estimates for (CA)_n/(GT)_n repeats based on hybridization studies range from 30 000 in cattle to 100 000 in mouse (Hamada et al. 1982; Hamada and Kakunaga 1982). Analysis of a large number of human cosmid clones and database sequences have given the average of one (CA)_n/(GT)_n repeat per 30 kb in the human DNA (Stallings et al. 1991; Beckmann and Weber 1992).

The process that leads to microsatellite formation is not yet established (Levinson and Gutman 1987a; Tautz 1989). It has been suggested that microsatellites may arise from cryptically simple sequences which are microsatellite-like and more common in genomes than expected by chance (reviewed by Hancock 1999). Messier et al. (1996) have presented evidence concerning substitutions, which most probably have lead to different microsatellites in distinct primate lineages: one substitution has resulted in the (ATGT)₂ tetranucleotide repeat from the 'basic'

ATGTGTGT sequence, which has then expanded to (ATGT)₅ in one lineage. Similarly, another substitution has created (GT)₅, which has expanded by one repeat in one lineage. Since 3' ends of several SINE families contain a polymorphic microsatellite, for example human *Alus* (e.g. Economou et al. 1990), PRE-1 of pig (Miller and Archibald 1993; Ellegren 1993) and ruminant Bov-tA (Kaukinen and Varvio 1992), they have been proposed as one candidate for origin of microsatellites (Kaukinen and Varvio 1992; Arcot et al. 1995).

1.2.2. Microsatellites evolve through errors during the replication

Microsatellites are highly polymorphic (Litt and Luty 1989; Weber and May 1989; Tautz 1989). In general, their mutation rate is high, estimates ranging from 10⁻² to 10⁻⁵ per locus per generation (e.g. Weber and Wong 1993 and references therein; Ellegren 1995). Slipped strand mispairing (slippage) during DNA replication (Levinson and Gutmann 1987a; Schlötterer and Tautz 1992) has been suggested to be a mechanism for generating diversity in microsatellite loci. It occurs when the nascent strand of tandemly repeated DNA dissociates from the template strand and reanneals out of phase with neighboring repeat. This leads either to insertion or deletion of a DNA segment, depending on whether the slippage

occurs in the 5'→3' direction or in the opposite one. In addition to slippage strand mispairing, recombination by crossing over or by gene conversions, more typical for minisatellite evolution, can be responsible for part of the new variation (Hancock 1999).

Several lines of evidence support the slippage hypothesis. First, most of the length mutations at microsatellites represent gains or losses of a single repeat unit although also more complex alterations have been identified (e.g. Weber and Wong 1993; Di Rienzo et al. 1994; Primmer et al. 1996). Second, interruption of microsatellites results in reduced polymorphism (Weber 1990) and reduced mutation rates (Bichara et al. 1995; Chong et al. 1995; Kunst et al. 1997) consistent with the greater difficulty of forming slipped intermediates in the presence of sequence interruptions. In addition, *in vivo* studies have shown that microsatellite mutation rates increase with array length in both *E.coli* (Levinson and Gutmann 1987b; Murphy et al. 1989) and yeast (Wierdl et al. 1997).

1.2.3. Bovine microsatellites

In domestic animals, like cattle, the main purpose for generating polymorphic DNA markers is for constructing genetic linkage maps, which in turn are developed for identifying regions of the genome that influence economically important traits. The majority of the traits selected in livestock production are quantitative traits: the individual phenotype reflects the action of several genes, confounded by environmental effects. Mapping of underlying genes or quantitative trait loci (QTL) should allow marker assisted selection (MAS), which is expected to increase the rate of genetic progress (Georges et al. 1993a). Examples of cattle QTL mapped by microsatellite analysis are horn development (Georges et al. 1993b), weaver disease (Georges et al. 1993a) and milk production (Georges et al. 1995).

Soon after the first reports about microsatellite polymorphisms in humans (Litt and Luty 1989; Weber and May 1989), Fries et al. (1990) showed that dinucleotide blocks may be an abundant source of DNA polymorphism

in cattle. Vaiman et al. (1994) suggested that the cattle genome possesses less (CA)_n/(GT)_n sequences than other mammalian species and gave an estimate of from a few thousand to 30 000 as the total number of microsatellites. The small and large scale development of random microsatellite markers (e.g. Vaiman et al. 1994; Moore et al. 1994 and references therein) resulted in genetic linkage maps for cattle (Barendse et al. 1994; 1997; Bishop et al. 1994) of which the most recent one is constructed of 1231 polymorphic microsatellite loci, two other DNA markers and 14 erythrocyte antigen and serum proteins (Kappes et al. 1997). At present, the generation of additional markers is concentrated to specific genomic regions containing QTL (Kappes et al. 1997).

1.3. Interspersed repetitive sequence (IRS)-PCR

1.3.1. Preparative IRS-PCR

IRS-PCR approach is based on utilization of repetitive sequences as priming sites for amplification of genomic segments. When two priming sites are in opposite orientation, 3'-ends facing each other, and close enough for conventional PCR, amplification can occur.

The roots of the IRS-PCR are in the studies of human Alu elements, which were the first priming sites for the approach named at that point Alu-PCR (Nelson et al. 1989). The goal was to develop a method to isolate human-specific sequences from hybrid cells containing regions of the human genome in rodent cell backgrounds utilizing the narrow taxonomic distribution of SINEs. Similarly, the approach was useful in isolation of fragments from yeast artificial chromosome (YAC) clones as well as for probe preparation in bacteriophage vectors (Nelson et al. 1989). With some modifications to this procedure, IRS-PCR (Ledbetter et al. 1990) has been used for example for isolation of chromosome and subchromosome -specific genomic regions (e.g. Cotter et al. 1990; Ledbetter et al. 1990) and DNA markers (e.g. Guzzetta et al. 1992; Brooks- Wilson et al. 1992). IRS-PCR has made it possible to search microsatellites (see II/1.2) from complex DNA

sources and from specific chromosomal areas (Pereira de Souza et al. 1994; Davies et al. 1994). Genomic fragments isolated by IRS-PCR has also been utilized for YAC contig assembly (Hunter et al. 1994; Liu et al. 1995) as well as for filter-based genotyping (McCarthy et al. 1995; Elango et al. 1996).

1.3.2. IRS-PCR generate multilocus fingerprints

Sinnet et al. (1990) have suggested Alu morphs detected by Alu-specific primers as promising markers for genome mapping. Their dominant nature (presence/absence of a genomic fragment flanked by Alu repeats) was suggested to be compensated because multiple loci can be analysed simultaneously. Multiplex mapping has also been performed by IRS-PCR with primers specific for mouse SINEs and LINEs (Cox et al. 1991). Zietkiewicz et al. (1992) have shown that 'alumorphs' can serve as markers in linkage studies.

Kaukinen and Varvio (1992) have characterized multilocus mendelian polymorphisms in a sample of cattle and sheep with four primers derived from a consensus sequence of the Bov-tA element. They named the approach SINE-PCR. Similarly, Miller (1994) have detected polymorphisms with primers annealing to pig SINE and LINE.

Polymorphisms revealed by IRS-PCR are mainly dominant, presence-absence of a fragment. They result from base mutations or structural changes in an annealing site, preventing the amplification or generating length variability in the intervening sequence because of an alternate priming site. With primers derived from SINEs, length polymorphisms can be detected due to the occurrence of polymorphic microsatellites at their 3' end (see II/1.2.1).

Various approaches utilizing the concept described above have been introduced, named with diverse sets of names and used as multilocus profiling techniques in both eukaryotes and prokaryotes (see Caetano-Anolles and Gresshoff 1998). In general, IRS-PCR have been mostly utilized for generating intraspecific polymorphisms. However, the method is similarly suitable for interspecific

studies. Zietkiewicz et al. (1994) have suggested this application using primers derived from CA microsatellite repeats.

1.4. SINE insertion sites are powerful phylogenetic markers; relationships among cetartiodactyls as an example

Single SINE loci analysed by a presence/absence approach with primers flanking the insertion have turned out to be promising markers for phylogenetic purposes (reviewed by Cook and Tristem 1997; Hillis 1999). A clade marker approach assumes that two host species share a particular element because of an insertion occurred in the germ line of their common ancestor. It relies on assumption that SINEs can insert essentially anywhere in the genome, which implies that convergence, i.e. that insertion will occur in identical positions in two distinct lineages, is unlikely and that the insertion events are irreversible (Cook and Tristem 1997). SINE insertion analysis is excellent for tree topology but not for branch length estimation, because the generation of new insertions may well be episodic rather than clock-like (Cook and Tristem 1997). Okada and coworkers have used SINE insertions as 'temporal landmarks of evolution' in molecular systematics of salmonid fish (Murata et al. 1993; Murata et al. 1996) as well as cetartiodactyls (e.g. Nikaido et al. 1999). Similarly, Tatout et al. (1999) have used SINE insertions as clade markers in analysis of phylogenetic relationships of wild crucifer species (genus *Brassica*).

Persisting questions in cetartiodactyl evolution have been whether hippopotamus is a close relative of pig, and whether whales are closer relatives of artiodactyls than of the other ungulates (Fig. 4). Attempts to solve this have been done both by morphological and molecular studies (reviewed by Nikaido et al. 1999). Recently, distribution of single SINE and LINE insertions have been utilized to find answers to these questions (Shimamura et al. 1997; Nomura et al. 1998; Nomura and Yasue 1999; Nikaido et al. 1999). Shimamura et al. (1997) showed that CHR-1 SINE is present in whales, hippopotamus and ruminants but not in

pig and camels by using presence/absence approach for nine SINE insertions. Nomura et al. (1998) isolated 30 SINE sequences from hippopotamus, which they found to be genetically related to CHR-2 described from whale by Shimamura et al. (1997). The dot blot hybridization experiments revealed that CHR-2 was more abundant in whale and hippopotamus genomes than in other artiodactyls. Further analysis of three CHR-2 insertions revealed that ruminants were first separated from a common ancestor of hippopotamus, ruminants and cetaceans. After these events hippopotamus and cetaceans were separated

(Nomura and Yasue 1999). Recently, Nikaido et al. (1999) have characterized more SINE and LINE insertion sites and confirmed the previous results. Yasue and Wada (1996) and Sulanderi et al. (1997) have previously shown that PRE-1 SINE sequences are distributed in the genomes of pig and its close relatives, but not in hippopotamus and cattle. The perfect correspondence between the reported SINE/LINE insertions (Nikaido et al. 1999) and the established phylogeny of artiodactyls (Gatesy et al. 1999) have supported their usefulness for phylogenetic inference.

2. NUMTS; PSEUDOGENES OF MITOCHONDRIAL ORIGIN

2.1. The origin of numts

2.1.1. A DNA or RNA mediated process?

The formation of a numt involves several steps: the generation of a mtDNA fragment to be transferred, transposition of it from the mitochondria to the nucleus and integration into the chromosomal DNA. The genomic organization of numts in different species suggests that there may be several strategies for formation of numt pseudogenes.

The generation of numts is assumed to be a DNA rather than an RNA mediated process, because control region sequences, which represent the untranscribed part of the mitochondrial genome have been isolated from many species (e.g. Lopez et al. 1994; Zischler et al. 1995; 1998; Lemos et al. 1999). However, in some plants, transfer of mtDNA sequences to the nucleus have been suggested to involve an RNA intermediate (e.g. Nugent and Palmer 1991; Grohmann et al. 1992; Blanchard and Schmidt 1995).

Zischler et al. (1995) have proposed that the displacement loop, the third DNA strand during the mtDNA replication (see II/2.4) may have a tendency to escape from mitochondria and integrate into the nuclear genome. Lopez et al. (1994) have suggested a

mechanism analogous to the 'petite' mtDNA mutations in yeast for generation of a cat numt. The mitochondrial genome has been fragmented and then reunited by intramolecular recombination, and this intermediate has transposed into the nucleus (Lopez et al. 1994 and references therein). Similar indications have been found also in a yeast numt (Farrelly and Butow 1983).

Release of mitochondrial DNA or RNA into the cytoplasm due to weakening of mitochondrial membranes in stress conditions has been suggested to enhance insertions of mitochondrial fragments into nuclear genome (e.g. Hadler et al. 1983; Kamimura et al. 1989; Shay and Werbin 1992). Thorsness et al. (1993) have isolated yeast mutants with elevated rates of transfer of mitochondrial DNA to the nucleus possibly due to leakage of nucleic acids through the membranes.

2.1.2. Integration by end-joining mechanism

Several lines of evidence have suggested that mitochondrial sequences have integrated into the nuclear genome by a non-homologous integration mechanism (e.g. Nomiya et al. 1985; Fukuda et al. 1985; Kamimura et al. 1989; Blanchard and Schmidt 1996; Zischler et al. 1995). In agreement with this, the flanking

sequences in human numts, have not shown significant homology to corresponding regions of the human mtDNA at the junctions of the mtDNA-like and the nuclear DNA sequences (Nomiyama et al. 1985; Fukuda et al. 1985; Kamimura et al. 1989). A further confirmation for the hypothesis has recently come from an experimental study in yeast (Ricchetti et al. 1999).

Remnants of transposable or viral elements have been found in proximity of yeast (Farrelly and Butow 1983), locust (Gellissen et al. 1983), human (Nomiyama et al. 1985; Fukuda et al. 1985; Kamimura et al. 1989), protist (Ossario et al. 1991) and rat (Zullo et al. 1991) numts. Analysis of human and yeast numts from databases has suggested that mtDNA movement and the integration process appear to be independent in retroelement insertions (see Fig. 1, 2, and 3 in Blanchard and Schmidt 1996). Similarly, neither direct nor inverted repeats or duplication of the nuclear DNA sequences have been found in the flanking sequences of human numts (Nomiyama et al. 1985; Fukuda et al. 1985; Kamimura et al. 1989), and although some LINE sequences have been co-isolated with human numts, sequence analysis has revealed that mitochondrial sequences were first taken up at various sites in the nuclear DNA, after which a retroelement was translocated within one of the insertions, and thus did not have a role in the movement of mtDNA sequences (Nomiyama et al. 1984).

2.2. The genomic organization of numts

2.2.1. Several numts of different age

MtDNA fragments representing protein-coding, tRNA and rRNA genes and D-loop region sequences seem to have been continuously integrated into the nuclear genome during the evolution of primates (Fukuda et al. 1985; Hu and Thilly 1994; van der Kuyl et al. 1995; Zischler et al. 1995; 1998). Based on library screening, Fukuda et al. (1985) have estimated that human nuclear DNA contains at least several hundred copies of mtDNA-like sequences. When a number of

primate numts have been compared, lack of sequence homology between the nuclear target sites (Nomiyama et al. 1985; Fukuda et al. 1985; Kamimura et al. 1989), phylogenetic analysis of numts (e.g. van der Kuyl et al. 1995) and the differences in homology to contemporary mitochondrial sequences (Hu and Thilly 1994) have indicated separate integrations. Recently, sequence comparisons and phylogenetic analysis of orthologous mtDNA from four opossum species (genus *Didelphis*) and paralogous nuclear sequences have suggested that mtDNA migration to the nuclear genome has occurred more than once in the evolution of *Didelphis* (Lemos et al. 1999).

In some cases numts have duplicated after the nuclear integration. For example, Collura and Stewart (1995) have sequenced two cytochrome b nuclear pseudogenes from orangutan cellular DNA, which most probably have originated by duplication after the integration into the nuclear DNA. Similarly, in rat, flanking region sequences of numts share 88 % identity suggesting a single integration event, followed later by a duplication (Zullo et al. 1991).

2.2.2. Several recent integrations in closely related taxa

Sorenson and Fleischer (1996) have described six independent recent transpositions of a mitochondrial control region in seven closely related taxa of diving ducks (tribe Aythyini). PCR amplification with primers specific for mitochondrial DNA and nuclear integrations, respectively, revealed only one numt, which was present in two species, while the other ones were found only from one species. Phylogenetic analysis confirmed the independent transposition events by placing each nuclear sequence as a close relative of the mtDNA haplotype of the species from which it was isolated.

A similar situation has been described by Sunnucks and Hales (1996), who found seven numts having a high proportion of unique characters and sharing a few derived ones and thus evidently representing seven independent integration events in aphids (genus *Sitobion*). Only one of the numts was more

closely related to another pseudogene than to contemporary mtDNA from the same species. Recently, Vaughan et al. (1999) have localized mitochondrial sequences to chromosomal DNA of grasshoppers (orthoptera), which revealed different nuclear locations of numts in different genera. Bensasson et al. (2000) have sequenced 87 distinct numts from 12 grasshopper individuals representing different subfamilies. The distance tree revealed that they result mainly from events which have occurred since these grasshopper species diverged from each other.

2.2.3. Amplification of numts

A dramatic example of a numt has come from domestic cat (Lopez et al. 1994). A 7.9 kb mitochondrial segment has recently integrated into a specific chromosomal location, become amplified 38-76 times and now occurs as a tandem repeat macrosatellite with multiple-length alleles segregating in cat populations (Lopez et al. 1994). Hu and Thilly (1995) have reported that one human numt is present at least five copies in each cell and that amplification has occurred very recently.

2.3. Evolution of mitochondrial DNA

Mitochondrial DNA has been reported to evolve from five to ten times the rate of single copy nuclear DNA (reviewed by Gray 1989). The substitution rate is strongly dependent on the region of mitochondrial genome considered (e.g. Pesole et al. 1999). The 16-17 kb mammalian mitochondria includes 13 protein coding genes, 22 tRNA and 2 rRNA genes as well as the control, or D-loop region (Fig. 5). This control region is highly structured and has been divided into three domains: the central region (CCR) and two peripheral ones, 5' left and 3' right (Saccone et al. 1991) or as recently named ETAS and CSB domains (Sbisa et al. 1997) (Fig. 6). Nonsynonymous sites of mitochondrial protein coding genes, the CCR and tRNA and rRNA genes evolve much more slowly than synonymous sites and both the peripheral domains of the control region (Pesole et al. 1999). Based on comparisons of several pairs of species from the Mammalian

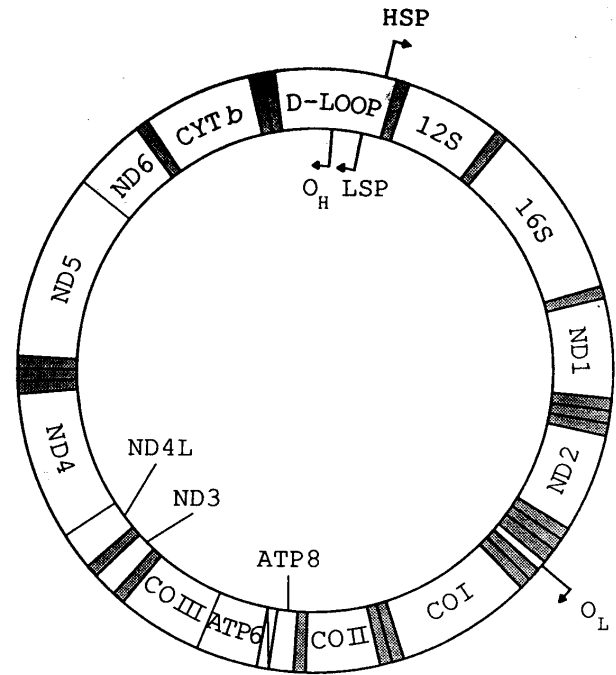


Figure 5. Representation of vertebrate mtDNA. The shaded areas represent the 22 tRNA genes. The 12s and 16s rRNA genes are shown. O_H and O_L are the respective origins of H- and L-strand synthesis. HSP and LSP are the respective promoters for transcripts copies from the H and L template strands. Arrows denote the directions of synthesis. For protein coding genes, see Clayton (1991). Adapted from Clayton (1991).

orders Primates, Carnivora, Cetacea and Perissodactyla, substitution rates for the ETAS domain and the CCR of the control region were 19 and 4×10^{-9} substitutions per nucleotide site per year, respectively (see standard errors in Table 2 in Pesole et al. 1999)

In addition to high substitution rates in the ETAS and CSB domains, they can also show large structural differences due to long and short repeats (LR and SR, respectively), which have been found in a diverse set of species (e.g. Fumagalli et al. 1996 and references therein). In addition, both domains also contain indels of different size and have been found difficult to align even between two closely related species (Pesole et al. 1999). Saitou and Ueda (1994) have compared nuclear and mitochondrial noncoding sequences in primates and found that insertions and deletions occur at the rate 2.0/kb/Myr in mitochondria while the corresponding estimate for nuclear DNA was ten times lower.

Base composition varies between the various mitochondrial regions and between different species. For example, in cattle and sheep 70 and 64 % of bases in the ETAS domain of the control region are A or T, respectively, while the AT and GC contents of the CCR domain are approximately equal (Sbisa et al. 1997). At fourfold degenerate third codon positions, the proportion of GC in cattle has been measured to be 33 % while the corresponding estimate in human is 47 % (Perna and Kocher 1995).

2.4. The ETAS domain include the putative termination signal for mtDNA replication

Replication of the mitochondrial genome begins by initiation of the heavy (H) strand synthesis from the origin of heavy strand replication (O_H) (Fig. 6) located in the CSB domain, utilizing as primer an RNA transcribed from the light (L) strand promoter (reviewed by Clayton 1982; Clayton 1991). DNA synthesis proceeds unidirectionally, and is optionally aborted to form the displacement loop (D-loop) triplex replication intermediate (Doda et al. 1981). Light strand synthesis starts at a separate origin, only after 2/3 of the H strand has been completed and thus productive

replication involves read-through of the terminator (Clayton 1982).

The 3' end of the displacement loop has experimentally been mapped to the ETAS domain 500–1000 bp downstream the initiation site, for example in human and mouse (Doda et al. 1981), in pig (MacKay et al. 1986) and in cattle (Madsen et al. 1993). By sequence comparisons, the termination points have been characterized also in sheep (Wood and Phua 1996) and in several cervid species (Douzery and Randi 1997). Since conserved sequence blocks have been preserved in the otherwise highly divergent ETAS domain, they have been suggested as candidates for termination signal/s.

Termination associated sequences (TASs) are conserved elements of 12-15 bp described first in human, mouse (Doda et al. 1981) and pig (MacKay et al. 1986) upstream of the mapped 3' ends of the displacement loop. Because the number of TAS and 3' ends corresponded, Doda et al (1981) and MacKay et al. (1986) concluded that termination of H strand synthesis is associated with these sequences. The accumulation of more mitochondrial sequence data have revealed features which have lead some authors to question their functional role (Foran et al. 1988; Sbisa et al. 1997). For example, the

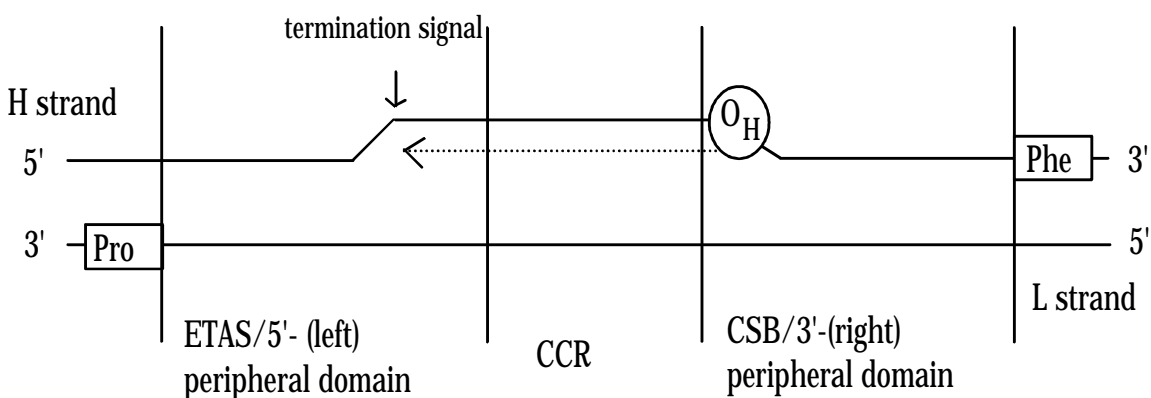


Figure 6. Schematic representation of the mitochondrial control region bounded by the tRNA genes for proline (Pro) and phenylalanine (Phe). The origin of heavy (H) strand replication O_H , the displacement (D) loop and the place of a putative termination signal are presented. Adapted partly from Douzery and Randi (1997).

conservation of some of the TAS is low, they occur in variable number which can exceed the number of 3' termini experimentally mapped and in addition they are located at variable distance upstream from the termination sites (e.g. Foran et al. 1988). In cattle, seven TAS have been identified (Madsen et al. 1993), in sheep Zardoya et al. (1995) have reported 10 TAS while Sbisa et al. (1997) have mentioned 16. In cervids Douzery and Randi (1997) have described four sequence blocks showing high homology to cattle and sheep TAS.

Another candidate for termination signal has been secondary structures accompanied by negative free-energy changes and thus likely to form in a non-duplex molecule sections. Dunon-Bluteau and Brun (1987) and Brown et al. (1986) have identified secondary structures near the mapped termination point of the D-loop in some mammalian species (but see Foran et al. 1988). Dunon-Bluteau and Brun (1987) noticed a conserved pentanucleotide (5' TACAT 3') base-paired in the secondary structures. Saccone et al. (1991) have also reported that these short mirror symmetries, TACAT/ATGTA (or GYRCAT by Douzery and Randi 1997) repeated several times in the mammalian mtDNA are able to form stable hairpin-loops. An 8 bp conserved element with a mirror sequence has been identified in sheep

and a 22 bp element in cervids, both noted to be able to form stable secondary structures (Hiendleder et al. 1998; Douzery and Randi 1997).

By functional analysis Madsen et al. (1993) have demonstrated that in cattle, a ~ 48-kDa protein binds specifically to a single TAS element (TAS-A) located 58 bp upstream of the D-loop 3' end thus being a trans-acting factor in D-loop formation. TAS-A has been identified by sequence homology in sheep (Wood and Phua 1996) and in cervids (Douzery and Randi 1997). Suzuki et al. (1996) have identified an approximately 97-kDa protein named Mt5-binding protein (Mt5BP), which binds to the Mt5 element including the 3' end of the displacement loop in some mammalian species including cattle.

Sbisa et al. (1997) have recently characterized two blocks, each about 60 bp, named ETAS 1 and ETAS 2, which are conserved in 26 analysed species representing 10 mammalian orders. Sbisa et al. (1997) have suggested that ETAS1 could contain the recognition signal (primary and secondary structural elements) for the termination of the nascent DNA or RNA chain, while ETAS2 could contain the binding sites for termination factor(s). In general, ETAS1 and ETAS2 included all the TASs and secondary structures previously identified and described above.

3. POLYMORPHISM IN THE MAJOR HISTOCOMPATIBILITY COMPLEX (MHC)-DRB GENE OF DOMESTIC CATTLE AND SHEEP

3.1. MHC of cattle and sheep

Bovine leucocyte antigen system (*BoLA*) has been located in chromosome 23 and represents one of the best characterized MHC regions surpassed only by the human leucocyte antigen (*HLA*) system and *H2* complex of mice (Lewin 1996). For sheep MHC (*Mhc-Ovar*), which has been mapped to chromosome 20, available data suggest a structure similar to the *BoLA*

complex (Andersson and Davies 1994; Schwaiger et al. 1996). MHCs of human and mice have been divided into three areas: class I, class II and Class III (Trowsdale 1995). The class II MHC genes of cattle appear to be further divided into two distinct regions (Fig. 7; Andersson and Davies 1994; Lewin 1996). For presentation of organization of MHC genes in vertebrates, see Trowsdale (1995).

Class I and II MHC membrane glycoproteins bind foreign antigens and present them to T lymphocytes. Class I molecules exist on most nucleated cell types whereas class II molecules are restricted primarily to B-cells and macrophages. Class II MHC molecules are composed of two chains (α and β) (Fig. 8), encoded by separate genes, A and B (Fig. 7). The *BoLA-DRA* gene has been found to be monomorphic, *DRB3* is highly expressed, *DRB2* is transcribed at low level and *DRB1* is a pseudogene. For a presentation of expression of other BoLA class II α and β genes see Fig. 7 and Davies et al. (1997). Compared to other domesticated species *Mhc-Ovar* is poorly characterized. There is indication of existence of only one *Ovar-DRA* gene, while several *DRB* genes exist of which *Ovar-DRB1* has been found to be highly polymorphic and *Ovar-DRB2* is a pseudogene (reviewed by Andersson and Davies 1994; Schwaiger et al. 1996).

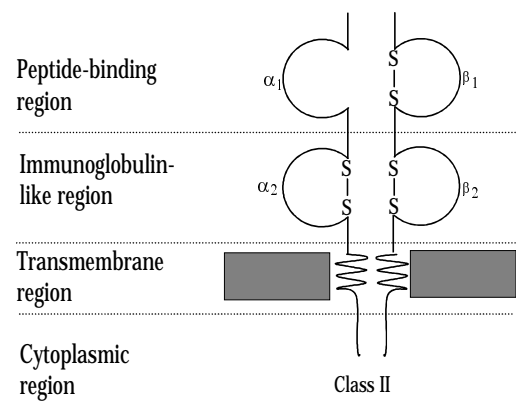


Figure 8. Schematic representation of the MHC class II molecule. Adapted partly from Andersson and Davies (1994).

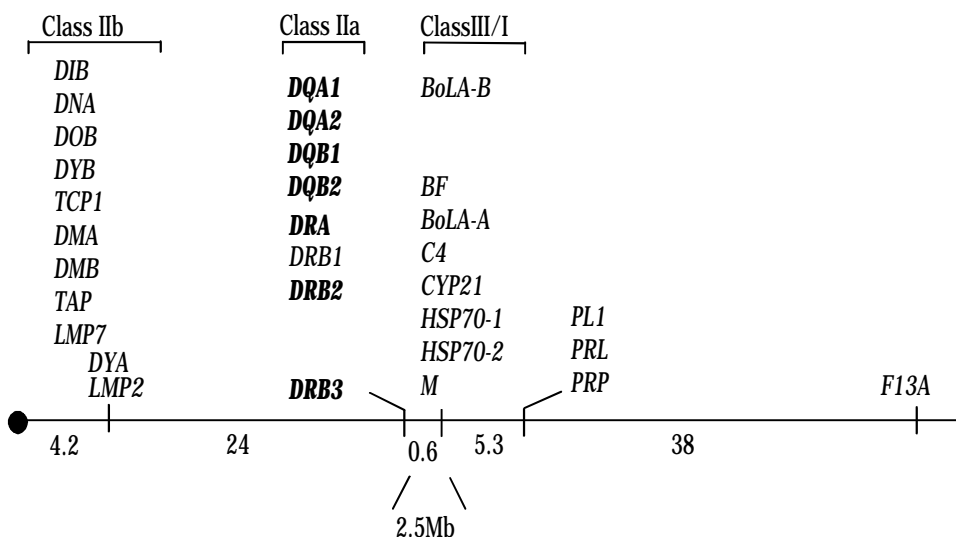


Figure 7. Genetic map of structural genes located on bovine chromosome 23. Map distances are expressed in cM. Adapted mainly from Lewin (1996) but also from centimorgans Davies et al. (1997). For details, see Lewin (1996). Expressed class II genes are bolded.

3.2. Origin and maintenance of MHC diversity

3.2.1. *Trans-species origin of MHC alleles*

First evidence of MHC polymorphism pre-dating the speciation of two species has come from rats and mice (e.g. Arden and Klein 1982; Figueroa et al. 1988), and human and chimpanzee (e.g. Lawlor et al. 1988; Mayer et al. 1989; Gyllensten and Erlich 1989). In general, the trans-species evolution of MHC allelic polymorphism (Klein 1987) has been described for all species investigated.

Due to common ancestry, related MHC alleles from different species cluster by lineages rather than by species when phylogenetic trees are constructed. An allelic lineage is a group of alleles that are more similar to one another than they are to members of other such groups. A lineage usually encompasses genes from different but related species, whereby all the genes of the group are derived from a common ancestral gene that is not the same as the ancestral gene of other lineages (Klein et al. 1993).

Although the allelic lineages are old, the alleles can have a more recent origin (Gyllensten et al. 1991; 1996). The human leucocyte antigen system (*HLA*)-*DRB1* alleles have been divided into thirteen lineages corresponding to serological *DR* specificities. Lack of intron variation within allelic lineages have proposed that the vast majority (greater than 90 %) of the contemporary *HLA-DRB1* alleles have a mean age of 250 000 years (Bergström et al. 1998). Evidence of new alleles described for isolated human populations (Titus-Trachtenberg et al. 1994; Mack and Erlich 1998) have also been shown as support for the more recent (10 000-20 000 years) origin of part of the MHC diversity.

3.2.2. *Intraexonic sequence exchanges (gene conversions)*

Exchange of segments between alleles and, more rarely, between loci due to gene conversion or other recombinational processes have generated MHC diversity both in class I and class II loci of a variety of species. Gene conversion means transfer of genetic

information from a donor to an acceptor gene without the donor being changed in the process. Gyllensten et al. (1991) have suggested that allelic polymorphism at the *HLA-DRB1* locus may have been generated in part by combining different variants of the two structural domains, β pleated sheet and α helical domain (see Fig. 9).

Several authors have reported that in artiodactyls, sequence exchange plays a role in generating diversity in class II genes similar to that reported by Gyllensten et al. (1991). Schwaiger et al. (1993b; 1994) have identified *DRB* exon 2 sequences in goat and sheep, which most probably have originated via double-recombination and/or sequence conversion events, and there is also evidence that single crossover have occurred. They concluded that, in addition to exchange between α helical and β sheet encoding regions, sequence exchange in class II genes appear rather randomly as has also been noticed by She et al. (1991) for murine class II genes. In the case of cattle, Mikko and Andersson (1995a) have published a large number of new *BoLA-DRB* alleles but mentioned that only a few 'new' sequence motifs were identified, and concluded that much of the allelic diversity is due to the presence of various combinations of shared sequence motifs. Similarly, in moose and in American bison sequence analysis of *DRB* alleles have indicated that intraexonic sequence exchange (intraexonic recombination/interallelic recombination) has contributed significantly to the generation of allelic diversity at this locus (Andersson and Mikko 1995; Mikko et al. 1997). The importance of gene conversions in generating variability in cattle class II genes has been also suggested by Ohta (1995).

Most of the evidence about intraexonic exchanges have come from sequence comparisons of exon 2 alleles. In addition, Belich et al. (1992) and Watkins et al. (1992) have described *HLA-B* alleles from isolated Native American populations, which most probably have originated through single microrecombination between pre-existing alleles. Direct indication has come from analysis of human sperm indicating that about

1/10,000 sperm represents a new *HLA-DPB1* allele sequence generated by gene conversion within the second exon (Zangenberg et al. 1995). Some authors have criticised strongly the intraexonic exchange hypothesis and instead suggested that a long temporal existence of MHC alleles (tens of millions of years; Klein et al. 1993) have made possible the accumulation of parallel mutations in different allelic lineages under the direction of positive selection, i.e. convergent evolution (e.g. reviewed by Klein and U’huigin 1995; U’huigin 1995).

3.2.3. *Balancing selection is operating on MHC loci*

The most direct evidence of balancing (overdominant) selection operating in MHC genes has come from comparisons of substitution patterns in antigen recognition sites (ARS) and the other regions of the human and mouse class I and II MHC genes (Hughes and Nei 1988; 1989). Hughes and Nei (1988; 1989) compared the number of mutations that alter the amino acid in a particular site, nonsynonymous substitutions, (dn) and synonymous substitutions (ds), which do not change the amino acid in ARS, to other regions of the genes. The frequency of dn was significantly higher than ds in ARS while in other regions, the reverse was true indicating purifying selection. Under neutrality the dn/ds ratio is expected to be more or less equal. Similar comparisons in other species also have revealed an enhanced rate of nonsynonymous nucleotide substitution in the ARS (e.g. Hughes and Hughes 1995).

The large number of alleles at MHC loci showing a relatively even frequency distribution and a lower level of homozygosity than expected under neutral theory has suggested the action of balancing selection (Hedrick and Thompson 1983). Heterozygosity excess has been found in many isolated Native American populations (Black and Salzano 1981; Markow et al. 1993; Black and Hedrick 1997) suggesting that selection is acting in present generations.

Overdominant and frequency-dependent selection, or heterozygote and rare allele advantage, respectively, are selection mechanisms based on host-pathogen interaction. The possibility that MHC polymorphism could be maintained by heterozygote advantage was suggested already by Doherty and Zinkernagel (1975), who argued that a heterozygote having two different MHC molecules would be more resistant to infectious diseases than a homozygote with only one type of MHC molecule. According to the frequency-dependent selection model, host individuals carrying a recently arisen mutant allele have a selective advantage because pathogens will not have had time to adapt to infecting cells carrying a new mutant antigen (Bodmer 1972). Hughes and Nei (1988), and Takahata and Nei (1990) have pointed out that the frequency dependent selection model can not explain the high degree of polymorphism nor the long persisting polymorphism at the MHC loci but would lead to constant turnover of alleles in the population since old alleles lose resistance to pathogens. For further biological support for overdominant hypothesis, see review by Nei and Hughes (1991).

MHC-based mating preference and selective abortion are selection mechanisms associated with reproduction and have been documented in mice and human populations (reviewed in Edwards and Potts 1996; Edwards and Hedrick 1998).

3.3 *DRB typing methodology*

3.3.1. *Before the PCR era*

Before the introduction of PCR, cattle and sheep MHC class II polymorphism were mostly characterized by serology, by isoelectric focusing and by restriction fragment length polymorphism (RFLP) and Southern hybridization analysis of genomic DNA. With all these typing methods, used alone or in collection, it was possible to detect haplotypes, i.e. a collection of alleles of different loci (e.g. Davies et al. 1992; Andersson and Davies 1994; Escayg et al. 1996).

3.3.2. Exon 2 based typing

The second exon of the *DRB* gene encodes the antigen binding groove of the MHC class II molecule (Fig. 8) and maintains virtually all the polymorphism. At present, 63 different *BoLA-DRB3* exon 2 alleles have been named (Davies et al. 1997; Russel et al. 1997; <http://www2.ri.bbsrc.ac.uk/bola/>) of which a large part has been identified using exon 2 specific primers and cloning and/or sequencing of the amplified segments (e.g. Sigurdardottir et al. 1991; Gelhaus et al. 1995; Mikko and Andersson 1995a; Maillard et al. 1999).

PCR amplified second exons have been also examined by RFLP analysis (van der Eijk et al 1992). However, Gelhaus et al (1995) concluded that although PCR-RFLP analysis is suitable for inbred populations, with outbred population some information could be missed. Maillard et al. (1999) have performed the RFLP analyses for cloned PCR products, which allowed the isolation of both alleles from each animal and made the approach more sensitive for detecting novel alleles. Sitte et al. (1996) have done genotyping by sequence specific oligonucleotide (SSO) typing for presence or absence of a deletion at the codon 65 of *BoLA-DRB3*2A* allele (*BoLA-DRB3*0201* in present nomenclature; Davies et al. 1997).

3.3.3 Microsatellite based *DRB* typing

The second intron of the cattle *DRB* gene possesses a highly polymorphic microsatellite repeat (Ammer et al. 1992; Ellegren et al. 1993). In addition, basically the same motif (gt)n(ga)m has been sequenced from several artiodactyl species such as sheep, goat, gazelle and giraffe (Schwaiger et al. 1993b; Schwaiger et al. 1994). Ellegren et al. (1993) have suggested that strong association between the microsatellite length polymorphism and exon2 sequence polymorphism can be utilized for *DRB* typing but noticed that the microsatellite polymorphism did not distinguish all the known *DRB3* alleles. van Haeringen et al. (1999) have shown that the resolution of microsatellite based *DRB3* typing is much better when the length polymorphism of

another microsatellite in corresponding location in *DRBP1* pseudogene is included.

Schwaiger et al. (1993a, b) have described an oligonucleotide typing method for *DRB* genes, in which PCR fragments, including the second exon and the adjacent intron are first separated in polyacrylamide gel based on length variations of the microsatellite repeat, then hybridized with probes for both the intronic repeat and exonic sequence. This polymorphism-specific oligonucleotide typing (PSO) has been utilized especially for *Mhc-Ovar DRB1* typing (Schwaiger et al. 1996).

3.4. Association of MHC haplotypes and resistance

So far, several documented cases exist where specific MHC haplotypes or genotypes provide resistance to parasites. One classic example comes from chicken, in which animals with the haplotype B21 have much higher resistance to Marek's disease (MD), a type of viral leukemia (Briles et al. 1977). If chickens are selected for resistance to this fatal disease, the frequency of the B21 haplotype increases rapidly (Gavora et al. 1986). In cattle, Xu et al. (1993) have reported an association of resistance to persistent lymphocytosis (PL) caused by bovine leukemia virus and a *BoLA-DRB3* exon 2 polymorphism and Zanotti et al. (1996) have found association of class II haplotypes and resistance to the same disease. Paterson et al. (1998) have described association of MHC variation with juvenile survival and parasite resistance in unmanaged sheep population. Although the data was more consistent with the frequency dependent selection model, the authors proposed that different MHC alleles may exhibit different associations at different stages during the Soya sheep's life leading to heterozygotes showing the highest overall fitness.

In humans, Hill et al. (1991) have provided evidence that in West Africa specific class I and class II alleles are associated with resistance to *Plasmodium falciparum* malaria and suggested that their data favour the hypothesis of frequency dependent selection

rather than overdominant selection (but see Hughes and Nei 1992). Heterozygote advantage have been reported for *HLA* in hepatitis B virus infection (Thursz et al. 1997) and human immunodeficiency virus-type-1 (HIV-1) (Carrington et al. 1999).

3.5. MHC and conservation of endangered populations

The level of MHC diversity is expected to be connected to welfare and survival of a population. MHC polymorphism is believed to increase the ability to bind peptides of different origin, and thus facilitate the recognition of many different pathogens. However, there is no unambiguous link between low intraspecific MHC variation and disease susceptibility (Caro and Laurenson 1994). The best-known example of a species maintaining low MHC diversity and documented health problems is cheetah; nonetheless, there is low variability throughout the cheetah genome, and other loci could be responsible for the numerous susceptibilities to disease recorded (reviewed by Edwards and Potts 1996). A number of cases have been found where a limited amount of MHC

polymorphism has been characterized without showing any indication of lower viability (reviewed in Table 14-2 in Edwards and Potts 1996). Moose has provided an additional example of a viable population with restricted MHC polymorphism probably due to a population bottleneck or limited amount of pathogens due to a rather solitary lifestyle (Mikko and Andersson 1995b; Ellegren et al. 1996).

Based on the overdominance selection model, it has been suggested that all captive breeding programs for endangered vertebrate species should be designed with the preservation of MHC allelic diversity as the main goal (Hughes 1991). Another way of choosing animals for captive breeding programmes is based on a frequency dependent model, and favour rare alleles or haplotypes probably having higher fitness. However, any captive breeding scheme designed around a linkage group will erode genome-wide diversity drastically compared to a model based on preservation of overall levels of genetic diversity (reviewed by Edwards and Potts 1996).

III. MATERIALS AND METHODS

Materials

DNA samples (isolated from blood and muscle with standard methods) from six species representing two ruminant families were included into the five studies.

Family Bovidae:

Cattle (*Bos taurus*) (I, II, III, V)

Sheep (*Ovis aries*) (II, III, IV)

Family Cervidae:

Moose (*Alces alces*) (II, III)

Reindeer (*Rangifer tarandus*) (II, III)

White-tailed deer (*Odocoileus virginianus*) (II)

Red deer (*Cervus elaphus*) (III)

In addition, whale (*Balaenoptera acutorostrata*; Cetacea), pig (*Sus scrofa*; Suina) and horse (*Equus caballus*; Perissodactyla) DNA samples were used (II, III).

Polymerase chain reaction (PCR) amplification with

1. primers derived from consensus sequence of Bov-tA SINE (Kaukinen and Varvio 1992) (I, II)
2. SSRanchored primers (II)
3. universal mtDNA control region primers L15905 and H16517 (Gottelli et al. 1994) (III)
4. primers specific for four numts (III)
5. primers LA31 and LA32 (Sigurdardottir et al. 1991) (IV, V)

Gel electrophoresis

1. low-melting agarose (I, II, III, IV, V)
2. polyacrylamide (PAGE) (II, III)
3. single strand conformation polymorphism (SSCP) (III-V)

PCR fragment purification, cloning, library construction and sequencing

1. purification of PCR fragment from low-melting agarose (I, II, III, IV, V) and from PAGE (III)
2. cloning of purified PCR fragment (I, II, III, IV, V)
3. plasmid library construction and hybridization with oligonucleotide probes (I, II)
4. allele fingerprinting of replicate clones (III, IV, V)
5. plasmid isolation and manual sequencing (I, II, III, IV, V)
6. automated cycle sequencing (V)

Phylogenetic analysis

1. by maximum parsimony (II, III)
2. by neighbor-joining (NJ) (II, III)

Other analysis

1. test for Hardy-Weinberg proportions and f estimates (V)
2. comparative sequence analysis of various structural and sequence evolution features in mitochondrial DNA and numt pseudogenes (III)

IV. RESULTS AND DISCUSSION

1. SINE TARGETING OF BOVINE MICROSATELLITE MARKERS (publication I)

At present, when a large number of bovine microsatellite markers exist the major goal is to generate new markers from specific genomic regions and fill the gaps in the genetic linkage map (Kappes et al. 1997; see II/1.2.3). In the present study, SINE targeting by IRS-PCR and primers specific to the Bov-tA element (called C-A element in publication I; see Table 1) was utilized for isolation of bovine microsatellites from hybrid cell genomes containing fragments of bovine chromosomes in rodent cell background.

Seven different libraries were constructed (for details, see publication I) from the two bovine/rodent hybrid cell lines, BO 8.1C and BO 8.1R (Table 1, publication I). The hybridization with labelled (TG)₁₀ and (TC)₁₀ probes resulted in 102 clones with positive signal. Inserts from these were isolated, resolved in agarose gels and their lengths compared. Altogether 47 positive clones were sequenced resulting in 27 unique genomic fragments, 15 duplicates of them and 5 false positives (inter-SINE PCR fragments with no microsatellite repeat).

The libraries from the two cell lines were treated separately. Their genomic origins were partly in the same chromosomal regions, which provides one explanation for the high number of sequenced duplicates. Interestingly, no duplicates were found between libraries which were constructed using IRS-PCR

fragments generated with primer SINE2 (c1, c2, r1) and with primer SINE10 together with primer/s from the A-subelement (c3, c4, r2, r3). This indicates that these primers may anneal to different Bov-tA subfamilies.

Nine long repeats ($n \geq 12$) were found (1/3 of the unique genomic segments) of which one was a previously characterized microsatellite locus, UWCA9 (Sun et al. 1994) (Table 2, publication I). With one exception (b2) they were located as 3' tails of the Bov-tA element. Primer/s were designed for eight repeats of which five detected useful polymorphism, and their genotype data for international reference families were merged with the Cattle Genotypic Database (CGD). Four of the microsatellites were assigned to bovine Chromosome (Chr) 11 and one to Chr 9 (Fig. 1, publication I).

SINE targeting proved to be a useful method for isolation of bovine microsatellite markers from complex DNA sources. Using several primers derived from different parts of the element could be a workable strategy for SINE targeting. The unique PCR primers, which were used together with a SINE primer in genotyping (Table 2, publication I) were designed carefully avoiding homologies to any repeated sequences. All of them worked well which further supports the usability of the approach in developing markers from genomic regions of interest.

2. INTER-SINE AND INTER-SSR FINGERPRINTS IN SIX ARTIODACTYL SPECIES (publication II)

Applications of IRS-PCR have mostly been concentrated to intraspecific polymorphisms (e.g. Sinnet et al. 1991; Kaukinen and Varvio 1992; Miller 1994). However, IRS fingerprints

should also carry information on the evolutionary relationships of related species and thus provide characters for systematic studies. This was suggested by Zietkiewicz et

al. (1994) when showing the applicability of inter-simple sequence repeat (SSR)-PCR for genomic fingerprinting at the interspecies level. In the present study, IRS-PCR was utilized to generate fingerprints in six artiodactyl species: cattle, sheep, moose, reindeer, white-tailed deer and pig. Since their phylogenetic relationships are quite well established (see discussion of publication II), they serve as a good model to examine the phylogenetic signal of IRS-PCR fingerprints.

Three primers derived from the Bov-tA SINE were used. Of these SINE10, was from the tRNA related part while two, SINE11 and SINE2, were from the A-subelement. Similarly, SSR-anchored primers, (CA)₈GG, (CA)₈AG and (CA)₈AT in three different combinations were applied in IRS-PCR fingerprinting experiments. To combine the SSR primers was based on the idea that more information can be generated in a single experiment.

Polyacrylamide gel electrophoresis of IRS-PCR products revealed primer and species-specific fingerprint profiles, with a low level of intraspecific polymorphisms (Fig. 2 and Table 1, publication II). In some cases, the variability of inter-SINE PCR fragments was related to differences in available priming sites. Interestingly, SINE11 generated twice as many IRS-PCR fragments in cervids than in bovids. This may indicate a subfamily structure similar to that found in the priming site of SINE10, where an indel of 9 bp most probably has prevented the amplification in a large number of Bov-tA SINEs.

The fragments of IRS-PCR fingerprint profiles were treated as binary characters (presence/absence), and species relationships were assessed both by character- and distance-based methods, i.e. by maximum parsimony and NJ-clustering, respectively. For the character-based approach, a fragment was scored as present if found in any individual of a given species. For the phenetic approach, a band-sharing similarity was calculated for each

pair of individuals as the number of shared fragments divided by the average total number in the two individuals.

Changes in inter-IRS fingerprint patterns may be caused by length variations in the inter-element segments, and by processes of primer site evolution. New annealing sites for SINE primers are created by retroposition (II/1.1.1.) and lost by point mutations or indels in the primer site. Degeneration through mutation or collapse of the SSR length destroys priming sites of SSRanchored primers (II/1.2.2.).

In most of the six experiments the IRS-PCR fingerprint data recover the correct phylogeny with parsimony and clustering analysis. In general, on time scales < 30 Myr the IRS-PCR fingerprinting approach seems to reveal a considerable amount of true information on mammalian relationships. A prominent example comes from the cervid species, whose relationships have been contradictory (see discussion of publication II). The ((reindeer, wt deer) moose) hierarchy was supported by the combined data from inter-SINE and inter-SSR fingerprints and from four of the six individual experiments. Recent data from nuclear and mitochondrial sequences have supported the same topology (Randi et al. 1998; Polziehn and Strobeck 1998).

Here, we show that IRS-PCR can provide a “quick-and-dirty” method for phylogenetic analysis. Since the phylogenetic signal varied between individual experiments but the combination of them results in a robust tree, a good strategy may be to use several primers for fingerprinting experiments and then combine the data. Especially the SSRs which are known to be ubiquitous components of all the eukaryotes can provide universal priming sites for IRS-PCR. The narrow taxonomic distribution of SINEs (e.g. Deiniger and Batzer 1993) as well as the fact that sequence information is required for primer synthesis limit their usefulness in inter-IRS fingerprinting.

3. DISTRIBUTION AND EVOLUTION OF Bov-tA SINE IN THE ARTIODACTYLA (publication II)

Interspersed sequences independently described in two bovids, cattle (Watanabe et al. 1982) and goat (Schon et al. 1981) were shown to represent the same SINE family (Rogers 1985).

We offer several lines of evidence that supports the presence of Bov-tA in cervid species. Primers derived from Bov-tA PCR amplified a comparable amount of fragments from cervids, which indicates that, as well as in bovids, this SINE family is a ubiquitous component of a cervid genome (Fig. 2 and Table 1, publication II). Similarly, the Bov-tA sequences generated by inter-SINE PCR indicated a similar structure of Bov-tA in cervids (Fig. 1, publication II). The phylogenetic analysis did not reveal clustering with taxonomic origin (Fig. 5, publication II). Sequence analysis of paralogous sequences of

bovids and cervids indicated divergence over 25 MYR ago. Recently, Shimamura et al. (1999) have confirmed by dot blot hybridization that Bov-tA is distributed into genomes of pecoran and tragulid ruminants (see Fig. 4).

Using a primer derived from the tRNA related part of the Bov-tA, Kaukinen and Varvio (1992) have shown that amplification occurs not only in bovids (sheep and cattle), but also in pig. The primer site showed a high homology to porcine SINE (Fig. 4 in Lenstra et al. 1993), which could explain the successful amplification. Here, the SINE 2 and -11, both derived from the A-subelement amplified 2 and 4 fragments from pig DNA, respectively. These fragments most probably represent random priming.

4. MULTIPLE INTEGRATIONS OF THE MITOCHONDRIAL CONTROL REGION INTO THE NUCLEAR GENOME OF RUMINANTS (publication III)

An attempt to examine mitochondrial variation of Finncattle with universal PCR primers for the mitochondrial control region gave us four mtDNA related sequences. In addition, a fifth sequence was isolated from moose genomic DNA with the same pair of primers. These five sequences are the first nuclear integrations of mtDNA described in Artiodactyla.

The five numts were characterized in detail. First, their integration history was constructed. Second, their evolution as nuclear pseudogenes was considered. Third, primary and secondary structural features of the numts and contemporary mtDNA were identified and compared.

The mitochondrial DNA fragment in question contains four segments showing different evolutionary dynamics in the mitochondrial environment (Fig. 1, publication III). In the analysis of structural features, sequence evolution and phylogenetic

information in the data, these domains were treated separately.

4.1. Integration history of five numts of mitochondrial control region

Segments transferred before the divergence of two species will usually be present in the nuclear genome of both. Where there is a well-established independent species phylogeny, the approximate date of transfer can be inferred by the phylogenetic distribution of the pseudogene (Perna and Kocher 1996).

Sequence divergences of the marginal conserved domains (tRNA and CCR; see Fig. 1 in publication III) gave a working hypothesis with five different integrations named *arntnumt1-5* (Fig. 3 in publication III). This hypothesis was experimentally tested with seven new primers, which were designed for targeted amplification of the *arntnumt2*, -3, -4 and -5 (Fig. 1 and Table 1, publication III).

In the case of *artnumt2* and -3, results were in accordance with the hypothesis since PCR experiments resulted in amplification products from cattle and sheep, but not from cervids. According to the working hypothesis the *artnumt4* and -5 should be present in the genomes of whale, pig and (horse). However, *artnumt4* and -5 were amplified from bovids and cervids but not from the outgroup species (Table 1, publication III). The most probable explanation for this is primer site degeneration in the *artnumt4* and -5, since several indels were found in these sequences, particularly in the *artnumt5* (Table 2, publication III).

Some of the PCR experiments resulted in more than one type of sequences from a single individual per inferred numt locus, differing from each other by 0.2 to 1.9 %. They can represent alleles or PCR artefacts of a single locus, or alternatively secondary nuclear duplication/multiplication (e.g. Lopez et al. 1994; Hu and Thilly 1995; see II/2.2). The single strand conformation polymorphism (SSCP) analysis of *artnumt2* and -3 (Fig. 4a and b, publication III) confirmed that they are independent single-copy genes and located *artnumt2* to the chromosome X in cattle. In addition, 'fingerprinting' and sequencing of replicate clones gave an estimate about sequence divergence for the allele sequences (0.5% to 0.8%) and the artefact clones (0.3% to 1%). Using these figures, the extra sequences described from the *artnumt4* and -5 were interpreted as PCR artefacts and both of them were determined to represent a single locus.

Although the phylogenetic analysis based on the marginal conserved domain (tRNA, CCR) sequences displays *artnumt4* & 5 as sister lineages (Fig. 2a, publication III), the data provides no statistical support for this relationship (i.e. common integration followed by nuclear duplication). A more ancient origin of *artnumt5* is supported by structural features. The marginal domains neither do not resolve the time of *artnumt2* and -3 integrations with respect to the cervid/bovid and the bovine/ovine radiation (20-25 Myr ago; > 30 Myr according to Hiendleder et al. 1998). The patterns of divergence and the identical size of *artnumt2* and -3 suggests that their integartions

were practically simultaneous, and although it does not seem likely, the possibility of a single nuclear integration followed by duplication within the nuclear genome cannot be excluded

4.2. *Artnumt* pseudogene evolution

When inserted into the nuclear genome, mitochondrial sequences begin to evolve like pseudogenes at a constant rate. Pseudogenes are subject to no functional constraints and thus would accumulate nucleotide substitutions at a rate equal to the mutation rate.

Estimates of nuclear pseudogene divergence were obtained from comparisons of homologous sequences among the five taxa. The rates of numt pseudogene evolution at different loci generally appear similar (c. 0.25%/Myr) but lower than corresponding estimate (1.2%/Myr) in other mammalian numts (Lopez et al. 1997). The sequence relationships correspond to the established phylogeny (Fig. 2b, publication III). The average rate of estimated pseudogene divergence between cattle and sheep in our study (c. 9%, see Fig. 2b and Table 2 in publication III) was remarkably similar to the 9.1 % found earlier for cattle $\psi^3\beta$ and goat $\psi^x\beta$ and $\psi^2\beta$ globin pseudogenes (Li 1997). The approximately similar diversity of four *artnumt* loci of different age and globin pseudogenes does not provide support for the hypothesis of Zischler et al. (1998) suggesting that due to differences in the basic evolutionary dynamics of mitochondrial and nuclear sequences pseudogenes of mitochondrial origin will undergo more rapid sequence evolution immediately after the integartion than later on.

Comparison of divergence at the various nuclear loci and the four domains of the studied segment (tRNA, CCR, 5'ETAS, 3'ETAS) in cattle and sheep suggested that the divergence is relatively uniform among loci and segments (Table 2, publication III). However, for some reason, the tRNA segment seemed to have diverged consistently less (average 3.6 %) than the control region sequence (average 8.2 %). The hypervariable 5'ETAS showed the most prominent compositional biases and also differences in lineages. The rate of insertions and deletions

since bovine/ovine divergence (Table 2, publication III) was in accordance with estimate (0.2/kb/Myr) presented by Saitou and Ueda (1994).

4.3. Structural features of the ETAS domain

The 3'ETAS region is of relatively constant size in mitochondria and includes elements and sequences which have a putative functional relevance in the replication of mitochondrial DNA (see II/2.4.). An essential part of these sequences are short conserved motifs GYRCAT/TACAT/ATGTA. The array of the short elements was conserved in ruminants and well preserved in *artnumt1*, -2 and -3. Most motifs were also identifiable even in *artnumt4* and -5, but with more degeneration (Fig. 5a, publication III). The pairwise comparison of ruminant mtDNA illustrated the markedly slower evolution in the conserved elements than in the intervening parts, in support of their implied functional roles. Similar comparisons in pseudogenes showed homogenisation of the rate and the relaxation of the functional constraint (Fig 6, publication III). An attempt

to demonstrate the relaxation through degeneration of the potential secondary structures was not as illustrative (Fig. 5b, publication III).

The 5'ETAS domain differs greatly in size among taxa (from 163 to 212 bp in ruminants) and especially the 3' end has been involved in major rearrangements. Interestingly, a degenerated 53-bp repeat of the 3'ETAS sequence was recognized in the adjacent 5'ETAS domain of the three youngest *artnumt* pseudogenes (Fig. 7, publication III). Traces of this duplicated sequence were seen in two cervid mtDNAs, but not in bovids. At *artnumt2* & 3, the degeneration was markedly stronger in 5'ETAS than in the 3'ETAS domain. It may be that after the cervid-bovid split and the practically simultaneous fossilization of *artnumt2* & 3 some 25 Myr ago, the degeneration in the mtDNA 5'ETAS replicate continued both in the cervid and ovid lineages, but to different degrees in different taxa.

5. MHC CLASS II DRB DIVERSITY OF FINNCATTLE AND FINNSHEEP (publications IV and V)

5.1. Heterozygosity excess in Finnish cattle breeds

DRB3 diversity among the five Finnish breeds were compared using four parameters: 1) total number of exon 2 alleles, 2) number and combined frequency of private alleles (i.e. found only in one Finnish breed), 3) number and combined frequency of new (unpublished) alleles and 4) observed and expected heterozygosity (H_{obs} and H_{exp} , respectively).

The amount of *DRB3* diversity found in Finnish cattle breeds (Fig. 1; publication V) were comparable to that reported in other breeds (e.g. Mikko and Andersson 1995a; van Haeringen et al. 1999). The test for observed and expected heterozygosity revealed heterozygosity excess in both commercial breeds, Finnish Ayrshire and Holstein-Friesian

and, in Western Finncattle (Table 1, publication V) indicating balancing selection in spite of intensive artificial selection. Recently, van Haeringen et al. (1999) have observed a highly significant trend towards a more even allele frequency distribution than expected under neutrality. Similarly, Sigurdardottir et al. (1988) have reported that the observed homozygosity at the cattle *DQ* locus was significantly lower than the expected homozygosity in a sample of 197 breeding bulls. Evidence for balancing selection has also been found in a free-living sheep population (Paterson 1998).

In general, MHC diversity of all farm animals studied is high (see Shook and Lamont 1996). Domesticated animals, like cattle and sheep, live in close herds which is a pathogen rich environment. That is why selection models

based on host-pathogen interaction are the most probable mechanism for maintaining the MHC diversity in these species (e.g. Mikko and Andersson 1995a).

5.2. *Ovar-DRB1* diversity of Finnsheep

The SSCP and sequence analysis of 31 sheep representing five different breeds from Finland and Russia resulted in 19 *Ovar-DRB1* alleles (Table 1, publication IV). Our sample size was too small for any conclusion about *DRB1* diversity of sheep in general, which by RFLP studies has been found to be either low and moderate (Grain et al. 1993; Escayg et al. 1996) or high (Boyce et al. 1996), and by sequencing and oligonucleotide typing to be high (Schwaiger et al. 1993ab; Schwaiger et al. 1994; Schwaiger et al. 1996). Schwaiger et al. (1996) have proposed that many more *Ovar-DRB1* alleles remain to be discovered, which I can agree with. In our sample of 31 sheep, seven new alleles were detected. As in the case for cattle (publication V), the native sheep breeds may carry unique MHC diversity to be discovered.

Exchange of short sequence motifs by some form of recombination event(s) has been proposed to generate diversity at the *DRB* locus for example in primates (e.g. Gyllensten et al. 1991) and in artiodactyls (e.g. Schwaiger et al. 1993; Mikko and Andersson 1995a). Gyllensten et al. (1991) have suggested that part of the allelic polymorphism in primates may have been generated by combining different variants of two structural domains, the β -pleated sheet and α -helical domain. For two of the seven new *Ovar-DRB1* sequences, *DRB1*0204* and *DRB1*0322*, described from Finnish and Russian breeds, 'ancestors' of both β -pleated sheets and α -helical domains were found among the published alleles, and for four of them, *DRB1*0115*, *DRB1*0116*, *DRB1*0205* and *DRB1*0323*, the ancestors of β -pleated sheets (Fig. 9). For *DRB1*0324* the most similar β -pleated sheet sequence was *DRB1*0306* and the α -helical domain of goat *Caae-DRB1*0202* differed by only one bp supporting the transspecies hypothesis.

5.3. *DRB* diversity and endangered native breeds

Livestock breeds are recognized as important components of world biodiversity because the genes and gene combinations they carry may be useful to agriculture in the future (e.g. Hall and Bradley 1995). Native breeds are adapted to their local environments during several generations of natural selection. Today many of them, having once been economically important, are rare or even endangered because of displacement by commercial breeds.

Eastern, Northern and Western Finncattle are native cattle breeds in Finland originally named on the basis of their geographic breeding areas. They have been used mainly for milk production but have been replaced during last decennia by more productive Finnish Ayrshire and Holstein-Friesian cattle, which are efficiently selected, imported breeds. At present, Eastern and Northern Finncattle are considered endangered, with only 150 and 120 cows left, respectively. They were threatened with extinction in the 1970s and 1980s, but are now being conserved in two gene bank herds. The Western Finncattle consists of 6000 cows.

The SSCP and sequence analysis of *BoLA-DRB3* exon 2 diversity in Finnish cattle breeds revealed extensive MHC diversity especially in Eastern Finncattle. This observation was in accordance with previous studies based mainly on neutral microsatellite loci indicating that native Finncattle maintain an equal amount and even slightly more genetic diversity than the commercial breeds (Varvio and Kaukinen 1993; Kantanen 1999).

5.4. SSCP analysis in *DRB* exon 2 typing

SSCP analysis allows detection of nucleotide changes in short DNA fragments due to mobility differences of a single-stranded molecule (Orita et al. 1989a, b). In general, the short double-stranded DNA fragments are first denatured by heating and then loaded in non denaturing polyacrylamide gels. The two complementary strands of DNA will migrate

	β-pleated sheet				α-helical domain					
	11	21	31	41	51	61	71	81		
	EQ	SKSECHFFNG	TERVRF	LDTRY	FYNQEEYVRF	DSDVGEYRAV	TELGRPDAEY	WNSQKDLLEQ	RRAAVDTYCR	HNYGVGESFT
DRB1*0115	-Y	T-K--R-S--	-----	---G---A--	---W-----	A---RS---	-----	-EI--R	K-----	-----
DRB1*0116	-Y	T-K--R-S--	-----	---G---A--	---W-----	A---RS---	-----	-EI--R	--TE-----	-----I---S
<i>DRB1*0105</i>	-Y	T-K--R-S--	-----	---G---A--	---W-----	A-----	-----	-EI---	T-----	-----
DRB1*0204	-Y	-T-----	-----	---G-----	---W-----	A---QS--H	-----	-E---R	---E-----	-----I---S
<i>DRB1*0201</i>	-Y	-T-----	-----	---G-----	---W-----	A-----K-	-----	-EI--R	--TE-----	-----I---S
<i>DRB1*0303</i>	-Y	H-----	-----	---G-----	---W--F--	A---QS--H	-----	-E---R	---E-----	-----I---S
DRB1*0205	-Y	-T-----	-----	---G-----	---W-----	A-----K-	-----	-EI--R	--TE-----	-----F----
<i>DRB1*0201</i>	-Y	-T-----	-----	---G-----	---W-----	A-----K-	-----	-EI--R	--TE-----	-----I---S
DRB1*0322	-Y	H-----	-----	---G-----	---W--F--	A-----K-	-----	-F--S	--T-----	-----
<i>DRB1*0303</i>	-Y	H-----	-----	---G-----	---W--F--	A---QS--H	-----	-E---R	---E-----	-----I---S
<i>M93432</i>	-Y	-----	-----L-E--	---G-----	---W--F--	A-----K-	-----	-F--S	--T-----	-----
DRB1*0323	-Y	H---R-S--	-----Y----	---G-----	-N-W-----	A---RS---	-----	-F---	T-E-----	-----I---S
<i>DRB1*0321</i>	-Y	H---R-S--	-----Y----	---G-----	-N-W-----	A-----K-	-----	-F---	T-TE-----	-----I---S
DRB1*0324	-Y	H---R-S--	-----	---G---A--	---W-----	A-----	-----	-E---R	--TE-----	-----
<i>DRB1*0306</i>	-Y	H---R-S--	-----Y----	---G-----	---W-----	A-----	-----	-F--R	K--N-----	-----
<i>DRB1*0202**</i>	-Y	-T-----S--	-----L----	---G--TL-Y	---W-----	A-----	-----	-E---	--TE-----	-----

Figure 9. Amino acid sequences of new *MhcOvar*-DRB1 alleles (bolded) and the published β-pleated sheet and α-helical domain sequences compared to the *HLA-DRB* consensus sequence. ** indicates *Caae-DRB1*0202*.

differently and will therefore separate during gel electrophoresis. Homozygous individuals produce two fragments, while in heterozygotes the SSCP pattern consist of four fragments. Although the basic idea of the approach is simple, i.e. that single stranded molecules adapt their conformation due to their primary sequence, the actual mechanisms is not known. SSCP analysis has been applied in analysis of polymorphism in MHC *DRB* and *DQ* loci in moose (Mikko and Andersson 1995b; Ellegren et al. 1996), in bison (Mikko et al. 1997) and in sheep (Snibson et al. 1998)

The approximately 300 bp (including primers) PCR amplified *DRB* exon 2 fragments of cattle and sheep were resolved in SSCP gels. The analyses were used both for visual comparison of *DRB* polymorphism as well as for 'allele fingerprinting' of cloned PCR products (publication IV). SSCP analysis has been reported to be highly sequence dependent and most effective for short PCR fragments (e.g. Glavac and Dean 1993; Hayashi and Yandell 1993).

SSCP was powerful enough to separate even one bp differences between two *DRB* exon 2 fragments, which was confirmed by sequencing a large number of replicate clones. Hayashi (1991) have estimated that with SSCP single base pair changes will be detected in 99 % of the cases for 100-300 bp fragments. Thus, it is a more sensitive method than denaturing gradient gel electrophoresis (DGGE) (Aldridge et al. 1998), which was recently introduced as a promising method for *DRB* typing with a remark that certain allele combinations may be difficult to separate if two alleles have similar melting temperatures. Glavac and Dean (1993) have concluded that most of the information in SSCP comes from the purine rich strand. The purine: pyrimidine ratio of exon 2 strands is 3:2, which could explain the observed SSCP patterns where one of the two fragments seems to be more informative than the other (for example, Fig. 2a in publication IV).

V. SUMMARY AND CONCLUDING REMARKS

The five studies included in the present thesis focus on intra- and interspecific genomic evolution and diversity of ruminants in the order Artiodactyla (Mammalia).

Repetitive DNA, which constitutes a large part of hereditary material in the mammalian genome, is a source of genetic and evolutionary markers. By interspersed repetitive sequence (IRS)-PCR with primers derived from the ruminant Bov-tA SINE (short interspersed nuclear element) five new bovine microsatellite markers (*HELMTT41-45*) were isolated from bovine/rodent hybrid cell lines. SINE targeting approach thus provide an efficient tool to isolate microsatellites from complex DNA sources involving genomic regions of interest. Based on the results, I suggest that a useful strategy for SINE targeting of microsatellites could be to use several primers derived from various parts of the element resulting in amplification from maximum number of elements in different genomic regions.

The usefulness of the IRS-PCR approach for interspecific studies was examined by fingerprinting two bovid (cattle and sheep) and three cervid (moose, reindeer and white-tailed deer) species with primers derived from Bov-tA SINE and (CA) simple sequence, or microsatellite repeat. Another artiodactyl species, pig, was included as an outgroup in the experiments. Inter-IRS fingerprints revealed a considerable amount of true information on mammalian relationships on time scales < 30 Myr and thus could provide a universal “quick-and-dirty” method for phylogenetic studies. The Bov-tA SINE family was shown to be widespread already in the common ancestor of the bovids and cervids as evidenced by a comparable number of inter-SINE fragments and by analysis of paralogous Bov-tA SINE sequences.

The third survey presents nuclear integrations of mitochondrial sequences in Artiodactyla isolated by “universal” primers from cattle, sheep and moose genomic DNA. The integration history of the five numts (*artnumt1-5*) representing the mitochondrial control region (5' end and adjacent tRNAs) was reconstructed both experimentally and by comparative sequence analysis. The youngest integration, *artnumt1*, was estimated to be approximately 5 Myr while the oldest ones, *artnumt4* and *artnumt5* are probably older than the suid-artiodactyl divergence (65 Myr). The single locus status of *artnumt2* and -3 (about 20-25 Myr old integrations) was confirmed by SSCP analyses that also indicated X-linkage of *artnumt2* in cattle. The rate of *artnumt* pseudogene evolution (0.25%/Myr) was comparable to other ruminant pseudogenes. *Arnumts* provided ‘molecular fossils’ of different age, that have evolved since integration like nuclear pseudogenes and that reveal information about ancestral structure and evolution of the highly variable part of the control region. Short conserved elements (e.g. GYRCAT) were mapped in the mitochondrial DNA and in the *artnumt* pseudogenes. Comparison of sequence diversity and secondary structural features revealed degeneration of conservation in *artnumts* supporting the suggested functional role of the elements.

The fourth and fifth studies focused on MHC class II *DRB* diversity of two domestic animals, cattle and sheep. The second exon of the *DRB* gene is a source of highly polymorphic marker. In the fourth study the applicability of SSCP and sequence analysis for *DRB* typing was examined in a small sample of sheep. An ‘allele fingerprinting’ approach was developed to aid the *DRB* typing. The PCR-SSCP approach resolved even one bp differences between two PCR amplified exon 2 fragments and was thus found appropriate for *DRB* typing and able to discriminate novel allele sequences. In the fifth study *DRB* diversity of endangered native Finncattle and in comparison, two commercial breeds, Finnish Ayrshire and Holstein-Friesian was examined with the same approach.

The heterozygosity excess found in the commercial breeds and in the Western Finncattle indicates balancing selection. Deficiency of heterozygosity compared to Hardy-Weinberg proportions observed in Eastern Finncattle is most probably due to a subdivided population structure. The Eastern Finncattle maintained most different *DRB3* alleles, and similarly most private and new *DRB3* diversity. I can conclude, that although the Eastern and Northern Finncattle are considered highly endangered, consisting of only 150 and 120 cows, respectively, they carry considerable amount of unique MHC diversity, which should be conserved.

ACKNOWLEDGEMENTS

This work was carried out at the Department of Biosciences, Division of Genetics (former Department of Genetics), University of Helsinki. I thank professor Olli Halkka and professor Tapio Palva for providing excellent working facilities and a for a liberal attitude towards my work.

The leader of the research group, Dr Siru Varvio, I want to thank for the possibility to work in her group and for co-operation. During these years we have experienced a lot.

Dr Johanna Vilkki and Dr Antti Sajantila I want to thank for critically reviewing my thesis and for interesting discussions.

Dr Risto Väinölä has contributed a lot to finishing my thesis of which I am sincerely grateful to him. Dr Juha Kantanen and Dr Johanna Vilkki I want to thank for co-operation and for enthusiasm. My closests colleagues both in labwork and teaching, Dr Marja Ruohonen-Lehto and MSc Minna Pekkinen (Pirinen) I want to thank for stimulating teamwork and sharing both the good and the bad moments. My 'younger' colleagues, Kristo Kulju, Mikko Kolkkala and Pekka Katajisto have helped me whenever I needed especially with all my computer problems of which I am grateful to them. MSc Jaana Tähtinen (Kaukinen) I want to thank for all the help at the very beginning of my career and MSc Jukka Palo for beeing a good colleague. All the other past and present members of the group are also acknowledged.

For excellent technical asistance and for beeing such a nice person I want to thank for Sari Hellsten. Marja-Leena Ahonen, Sven Kevin and Pirkko Winter are also acknowledged for much appreciated techical support.

My family; my parents Arja and Heikki, my sisters Elisa and Paula, by brothers in law Arto Lahenius and Henri Takkinen and and nephews Tuukka and Aapo Lahenius and my closests family; Martin and Albert Romantschuk I want to thank for their love, support and empathy during all these years. Martin is also thanked for providing language revision and xerox copying service.

All my friends and mates I want to thank for relaxing company and all kind of support during these years. Marja Ruohonen-Lehto is especilly thanked for her support during the most blue moments of this project and Birgitta Andersson for providing her home for my working place when I needed.

This work has been financially supported by the 350th Anniversary Fund of the University of Helsinki, the Finnish Cultural Foundation, the Emil Aaltonen foundation, the Oskar Öflund Foundation and the Academy of Finland.

Helsinki, May 2nd, 2000

REFERENCES

- Aldridge BM, McGuirk SM, Clark RJ, Knapp LA, Watkins DI & Lunn DP** (1998). Denaturing gradient gel electrophoresis: a rapid method for differentiating *BoLA-DRB3* alleles. *Animal Genetics* 29: 389-394.
- Alexander LJ, Rohrer GA, Stone RT & Beattie CW** (1995). Porcine SINE-associated microsatellite markers: evidence for new artiodactyl SINEs. *Mammalian Genome* 6: 464-468.
- Ammer H, Schwaiger F-W, Kammerbauer C, Gomolka M, Arriens A, Lazary S & Epplen JT** (1992). Exonic polymorphism vs intronic simple repeat hypervariability in *MHC-DRB* genes. *Immunogenetics* 35: 332-340.
- Andersson L & Davies CJ** (1994). The major histocompatibility complex. In: *Cell-Mediated immunity in ruminants* (eds. BM Goddeeris & WI Morrison), pp. 37-57. CRC press, Boca Raton, FL, USA.
- Andersson L & Mikko S** (1995). Generation of MHC class II diversity by intra- and intergenic recombination. In: *Origin of major histocompatibility complex diversity*. *Immunological Reviews* (ed G Möller), No. 143, pp. 5-12. Munksgaard, Copenhagen, Denmark.
- Arcot SS, Wang Z, Weber JL, Deininger PL & Batzer MA** (1995). *Alu* repeats: A source for the genesis of primate microsatellites. *Genomics* 29: 136-144.
- Arden B & Klein J** (1982). Biochemical comparison of major histocompatibility complex molecules from different subspecies of *Mus musculus*: evidence for *trans*-specific evolution of alleles. *Proc Natl Acad Sci USA* 79: 2342-2346.
- Barendse W, Vaiman D, Kemp SJ, Sugimoto Y, Armitage SM, Williams JL et al.** (1997). A medium-density genetic linkage map of the bovine genome. *Mammalian Genome* 8: 21-28.
- Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM et al.** (1994). A genetic linkage map of the bovine genome. *Nature Genetics* 6: 227-235.
- Beckmann JS & Weber JL** (1992). Survey of human and rat microsatellites. *Genomics* 12: 627-631.
- Bensasson D, Zhang D-X & Hewitt GF** (2000). Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol Biol Evol* 17: 406-415.
- Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Petzler ML & Parham P** (1992). Unusual *HLA-B* alleles in two tribes of Brazilian Indians. *Nature* 357: 326-329.
- Bergström TF, Josefsson A, Erlich HA & Gyllensten U** (1998). Recent origin of *HLA-DRB1* alleles and implications for human evolution. *Nature Genetics* 18: 237-242.
- Bichara M, Schumacher S & Fuchs RPP** (1995). Genetic instability within monotonous runs of CpG sequences in *Eschericia coli*. *Genetics* 140: 897-907.
- Bishop MD, Kappes SM, Keele JW, Stone RT, Sunden SLF, Hawkins, GA et al.** (1994). A genetic linkage map for cattle. *Genetics* 136: 619-639.
- Black FL & Hedrick PW** (1997). Strong balancing selection at HLA loci: evidence from segregation in South Amerindian families. *Proc Natl Acad Sci USA* 94: 12452-12456.
- Black FL & Salzano FM** (1981). Evidence for heterosis in the HLA system. *Am J Hum Genet* 33: 894-899.
- Blanchard JL & Schmidt GW** (1996). Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution pattern. *Mol Biol Evol* 13: 537-548.
- Blanchard JL & Schmidt GW** (1995) Pervasive migration of organellar DNA to the nucleus in plants. *J Mol Evol* 41:397-406.
- Bodmer WF** (1972). Evolutionary significance of the HL-A system. *Nature* 237: 139-145.
- Boyce WM, Hedrick PW, Muggli-Cockett NE, Kalinowski S, Penedo MCT & Ramey II RR** (1996). Genetic variation of major histocompatibility complex and microsatellite loci: a comparison in bighorn sheep. *Genetics* 145: 421-433.
- Briles WE, Stone HA & Cole RK** (1977). Marek's disease: effects of *B* histocompatibility alloalleles in resistant and susceptible chicken lines. *Science* 195: 193-195.
- Brooks-Wilson AR, Smailus DE, Weier H-U & Goodfellow PJ** (1992). Human repeat

- element-mediated PCR: cloning and mapping of chromosome 10 DNA markers. *Genomics* 13: 409-414.
- Brown GG, Gadaleta G, Pepe G, Saccone C & Sbisà E** (1986). Structural conservation and variation in the D-loop containing region of vertebrate mitochondrial DNA. *J Mol Biol* 192: 503-511.
- Buntjer JB, Hoff IA & Lenstra JA** (1997). Artiodactyl interspersed DNA repeats in Cetacean genomes. *J Mol Evol* 45: 66-69.
- Caetano-Anolles G & Gresshoff PM** (1998). DNA markers: protocols, applications, and overviews. Wiley-Liss, Inc., New York, USA.
- Caro TM & Laurenson MK** (1994). Ecological and genetic factors in conservation: a cautionary tale. *Science* 263: 485-486.
- Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ et al.** (1999). *HLA* and HIV-1: Heterozygote advantage and *B*35-Cw*04* disadvantage. *Science* 283: 1748-1752.
- Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR & Zoghbi HY** (1995). Gametic and somatic tissue-specific heterogeneity of the expanded *SCA1* CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics* 10: 344-350.
- Clayton DA** (1982). Replication of animal mitochondrial DNA. *Cell* 28: 693-705.
- Clayton DA** (1991). Replication and transcription of vertebrate mitochondrial DNA. *Annu Rev Cell Biol* 7: 453-478.
- Collura RV & Stewart C-B** (1995). Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 378: 485-489.
- Cook JM & Tristem M** (1997). 'SINEs of the times' - transposable elements as clade markers for their hosts. *Trends in Ecology and Evolution* 12: 295-297.
- Cotter FE, Hampton GM, Nasipuri S, Bodmer WF & Young BD** (1990). Rapid isolation of human chromosome specific DNA probes from somatic cell hybrid. *Genomics* 7: 257-263.
- Cox RD, Copeland NG, Jenkins NA & Lehrach H** (1991). Interspersed repetitive element polymerase chain reaction product mapping using a mouse interspecific backcross. *Genomics* 10: 375-384.
- Davies CJ, Andersson L, Ellis SA, Hensen EJ, Lewin HA, Mikko S et al.** (1997). Nomenclature for factors of the BoLA system 1996: report of the ISAG BoLA nomenclature committee. *Animal Genetics* 28: 159-168.
- Davies CJ, Andersson L, Joosten I, Mariani P, Gasbarre LC & Hensen EJ** (1992). Characterization of bovine MHC class II polymorphism using three typing methods: serology RFLP and IEF. *Eur J of Immunogenetics* 19: 253-262.
- Davies W, Høyheim B, Chaput B, Archibald AL & Frelat G** (1994). Characterization of microsatellites from flow-sorted porcine chromosome 13. *Mammalian Genome* 5: 707-711.
- Deininger PL & Batzer MA** (1993). Evolution of retroposons. In: *Evolutionary Biology* (eds. MK Hect, RJ MacIntyre & MT Clegg), Vol. 27, pp. 157-196. Plenum Press New York, NY, USA.
- Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M & Freimer NB** (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA* 91: 3166-3170.
- Doda JN, Wright CT & Clayton DA** (1981). Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. *Proc Natl Acad Sci USA* 78: 6116-6120.
- Doherty PC & Zinkernagel RM** (1975). Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. *Nature* 256: 50-52.
- Douzery E & Randi E** (1997). The mitochondrial control region of Cervidae: evolutionary patterns and phylogenetic content. *Mol Biol Evol* 14: 1154-1166.
- Duncan CH** (1987). Novel Alu-type repeat in artiodactyls. *Nucleic Acids Research* 15: 1340.
- Dunon-Bluteau DC & Brun GM** (1987). Mapping at the nucleotide level of *Xenopus laevis* mitochondrial D-loop H strand: structural features of the 3' region. *Biochem Int* 14: 643-657.
- Economou EP, Bergen AW, Warren AC & Antonarakis SE** (1990). The polydeoxyadenylate tract of *Alu* repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci USA* 87: 2951-2954.
- Edwards SV & Hedrick PW** (1998). Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends in Ecology and Evolution* 13: 305-311.

- Edwards SV & Potts WK** (1996). Polymorphism of genes in the major histocompatibility complex (MHC): implications for conservation genetics of vertebrates. In: *Molecular genetic approaches in conservation* (eds. TB Smith & RK Wayne), pp. 214-237. Oxford University Press, Oxford, UK.
- Eickbush TH** (1992). Transposing without ends: the non-LTR retrotransposable elements. *New Biology* 4: 430-440.
- Elango R, Riba L, Housman D & Hunter K** (1996). Generation and mapping of *Mus spretus* strain-specific markers for rapid genomic scanning. *Mammalian Genome* 7: 340-343.
- Ellegren H** (1995). Mutation rates at porcine microsatellite loci. *Mammalian Genome* 6: 376-377.
- Ellegren H** (1993). Variable SINE 3' poly(A) sequences: an abundant class of genetic markers in the pig genome. *Mammalian Genome* 4: 429-434.
- Ellegren H, Davies CJ & Andersson L** (1993). Strong association between polymorphisms in an intronic microsatellite and in the coding sequence of the *BoLA-DRB3* gene: implications for microsatellite stability and PCR-based *DRB3* typing. *Animal Genetics* 24: 269-275.
- Ellegren H, Mikko S, Wallin K & Andersson L** (1996). Limited polymorphism at major histocompatibility complex (MHC) loci in the Swedish moose *A. alces*. *Molecular Ecology* 5: 3-9.
- Escayg AP, Hickford JGH, Montgomery GW, Dodds KG & Bullock DW** (1996). Polymorphism at the ovine major histocompatibility complex class II loci. *Animal Genetics* 27: 305-312.
- Farrelly F & Butow RA** (1983). Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* 301: 296-301.
- Figuroa F, Günther E & Klein J** (1988). MHC polymorphism pre-dating speciation. *Nature* 335: 265-267.
- Foran DR, Hixson JE & Brown WM** (1988). Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop synthesis. *Nucleic Acids Research* 16: 5841-5861.
- Fries R, Eggen A & Stranzinger G** (1990). The bovine genome contains polymorphic microsatellites. *Genomics* 8: 403-406.
- Fukuda M, Wakasugi S, Tsuzuki T, Nomiya H, Shimada K & Miyata T** (1985). Mitochondrial DNA-like sequences in the human nuclear genome. *J Mol Biol* 186: 257-266.
- Fumagalli L, Taberlet P, Favre L & Hausser J** (1996). Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Mol Biol Evol* 13: 31-46.
- Gatesy J, Milinkovitch M, Waddell V & Stanhope M** (1999). Stability of cladistic relationships between Cetacea and higher-level Artiodactyl taxa. *Syst Biol* 48: 6-20.
- Gavora JS, Simenson M, Spencer JL, Fairfull RW & Gowe RS** (1986). Changes in the frequency of major histocompatibility haplotypes in chickens under selection for both high egg production and resistance to Marek's disease. *Journal of Animal Breeding* 103: 218-226.
- Gelhaus A, Schnittger L, Mehlitz D, Horstmann RD & Meyer CG** (1995). Sequence and PCR-RFLP analysis of 14 novel BoLA-DRB3 alleles. *Animal Genetics* 26: 147-153.
- Gellissen G, Bradfield JY, White BN & Wyatt GR** (1983). Mitochondrial DNA sequences in the nuclear genome of a locust. *Nature* 301: 631-634.
- Georges M, Nielsen D, Mackinnon M, Mishra A, Okimoto R, Pasquino AT et al.** (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139: 907-920.
- Georges M, Dietz AB, Mishra A, Nielsen D, Sargeant LS, Sorensen A et al.** (1993a). Microsatellite mapping of the gene causing weaver disease in cattle will allow the study of an associated quantitative trait locus. *Proc Natl Acad Sci USA* 90: 1058-1062.
- Georges M, Drinkwater R, King T, Mishra A, Moore SS, Nielsen D et al.** (1993b). Microsatellite mapping of a gene affecting horn development in *Bos taurus*. *Nature Genetics* 4: 206-210.
- Glavač D & Dean M** (1993). Optimization of the single-strand conformation polymorphism (SSCP) technique for detection of point mutations. *Human Mutation* 2: 404-414.
- Gottelli D, Sillero-Zurbini C, Applebaum GD, Roy MS, Girman DJ, Garcia-Moreno J et al.** (1994). Molecular genetics of the most endangered canid: the Ethiopian wolf *Canis simensis*. *Molecular Ecology* 3: 301-312.

- Grain F, Nain M-C, Labonne M-P, Lantier F, Lechopier P, Gebuhrer L et al.** (1993). Restriction fragment length polymorphism of DQB & DRB class II genes of the ovine major histocompatibility complex. *Animal Genetics* 24: 377-384.
- Gray MW** (1989). Origin and evolution of mitochondrial DNA. *Annu Rev Cell Biol* 5: 25-50.
- Grohmann L, Brennicke A & Schuster W** (1992). The mitochondrial gene encoding ribosomal protein S12 has been translocated to the nuclear genome in *Oenothera*. *Nucleic Acids Research* 20: 5641-5646.
- Guzzetta V, Montes de Oca-Luna R, Lupski JR & Patel PI** (1991). Isolation of region-specific and polymorphic markers from chromosome 17 by restricted *Alu* polymerase chain reaction. *Genomics* 9: 31-36.
- Gyllensten UB & Erlich HA** (1989). Ancient roots for polymorphism at the HLA-DQ α locus in primates. *Proc Natl Acad Sci USA* 86: 9986-9990.
- Gyllensten UB, Sundvall M & Erlich HA** (1991). Allelic diversity is generated by intraexon sequence exchange at the *DRB1* locus of primates. *Proc Natl Acad Sci USA* 88: 3686-3690.
- Gyllensten U, Bergström T, Josefsson A, Sundvall M & Erlich HA** (1996). Rapid allelic diversification and intensified selection at antigen recognition sites of the MHC class II DPB1 locus during hominoid evolution. *Tissue Antigens* 47: 212-221.
- Hadler HI, Dimitrijevic B & Mahalingam R** (1983). Mitochondrial DNA and nuclear DNA from normal rat liver have a common sequence. *Proc Natl Acad Sci USA* 80: 6495-6499.
- Hall SJG & Bradley DG** (1995). Conserving livestock breed biodiversity. *Trends in Ecology and Evolution* 10: 267-270.
- Hamada H & Kakunaga T** (1982). Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* 298: 396-398.
- Hamada H, Petrino MG & Kakunaga T** (1982). A novel repeated element with Z-DNA-forming potential is widely found in evolutionary diverse eukaryotic genomes. *Proc Natl Acad Sci USA* 79: 6465-6469.
- Hancock JM** (1999). Microsatellites and other simple sequences: genomic context and mutational mechanisms. In: *Microsatellites, evolution and applications* (eds. DB Goldstein & C Schlötterer), pp. 2-9. Oxford University Press Inc., New York, USA.
- Hayashi K** (1991). PCR-SSCP: A simple and sensitive method for detection of mutations in the genomic DNA. *PCR Methods and Applications* 1: 34-38.
- Hayashi K & Yandell DW** (1993). How sensitive is PCR-SSCP? *Human Mutation* 2:338-346.
- Hedrick PW & Thomson G** (1983). Evidence for balancing selection at HLA. *Genetics* 104: 449-456.
- Hiendleder S, Lewalski H, Wassmuth R & Janke A** (1998). The complete mitochondrial DNA sequence of the domestic sheep (*Ovis aries*) and comparison with the other major ovine haplotype. *J Mol Evol* 47: 441-448.
- Hill AVS, Allsopp CEM, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA et al.** (1991). Common West African antigens are associated with protection from malaria. *Nature* 352: 595-600.
- Hillis DM** (1999). SINEs of the perfect character. *Proc Natl Acad Sci USA* 96: 9979-9981.
- Hu G & Thilly WG** (1995). Multi-copy nuclear pseudogenes of mitochondrial DNA reveal recent acute genetic changes in the human genome. *Curr Genet* 28: 410-414.
- Hu G & Thilly WG** (1994). Evolutionary trail of the mitochondrial genome as based on human 16S rDNA pseudogenes. *Gene* 147: 197-204.
- Hughes AL** (1991). MHC polymorphism and the design of captive breeding programs. *Conservation Biology* 5: 249-251.
- Hughes AL & Hughes MK** (1995). Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics* 42: 233-243.
- Hughes AL & Nei M** (1988). Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 335: 167-170.
- Hughes AL & Nei M** (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc Natl Acad Sci USA* 86: 958-962.
- Hughes A & Nei M** (1992). Maintenance of MHC polymorphism. *Nature*: 355: 402-403.
- Hunter KW, Ontiveros SD, Watson ML, Stanton Jr VP, Gutierrez P, Bhat D et al.** (1994). Rapid and efficient construction of yeast artificial chromosome contigs in the mouse genome with interspersed repetitive sequence PCR (IRS-PCR): generation of a 5-cM, >5

- megabase contig on mouse chromosome I. *Mammalian Genome* 5: 597-607.
- Jobse C, Buntjer JB, Haagsma N, Breukelman HJ, Beintema JJ & Lenstra JA** (1995). Evolution and recombination of bovine DNA repeats. *J Mol Evol* 41: 277-283.
- Kantanen J** (1999). Genetic diversity of domestic cattle (*Bos taurus*) in North Europe. PhD thesis, University of Joensuu, publications in sciences, No 52. Joensuu, Finland.
- Kamimura N, Ishii S, Liandong M & Shay JW** (1989). Three separate mitochondrial DNA sequences are contiguous in the human genomic DNA. *J Mol Biol* 210: 703-707.
- Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS, Smith TPL et al.** (1997). A second-generation linkage map of the bovine genome. *Genome Research* 7: 235-249.
- Kaukinen J & Varvio S-L** (1992). Artiodactyl retroposons: association with microsatellites and use in SINemorph detection by PCR. *Nucleic Acids Research* 20: 2955-2958.
- Klein J** (1987). Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunology* 19: 155-162.
- Klein J & O'hUigin C** (1995). Class II *B Mhc* motifs in an evolutionary perspective. In: Origin of major histocompatibility complex diversity. *Immunological reviews* (ed G Möller), no. 143, pp. 89-112. Munksgaard, Copenhagen, Denmark.
- Klein J, Satta Y, O'hUigin C, Takahata N** (1993). The molecular descent of the major histocompatibility complex. *Annu Rev Immunol* 11: 269-295.
- Kordis D & Gubensek F** (1995). Horizontal SINE transfer between vertebrate classes. *Nature Genetics* 10: 131-132.
- Kunst CB, Leeflang EP, Iber JC, Arnheim N & Warren ST** (1997). The effect of FMR1 CGG repeat interruptions on mutational frequency as measured by sperm typing. *J Med Gen* 34: 627-631.
- Lawlor DA, Ward FE, Ennis PD, Jackson AP & Parham P** (1988). *HLA-A* and *B* polymorphism predate the divergence of humans and chimpanzees. *Nature* 335: 268-271.
- Ledbetter SA, Nelson DL, Warren ST & Ledbetter DH** (1990). Rapid isolation of DNA probes within specific chromosome regions by interspersed repetitive sequence polymerase chain reaction. *Genomics* 6: 475-481.
- Lemos B, Canavez F & Moreira MAM** (1999). Mitochondrial DNA-like sequences in the nuclear genome of the opossum genus *Didelphis* (Marsupialia: Didelphidae). *J Hered* 90: 543-547.
- Lenstra JA, van Boxtel JAF, Zwaagstra KA & Schwerin M** (1993). Short interspersed nuclear element (SINE) sequences of the *Bovidae*. *Animal Genetics* 24: 33-39.
- Levinson G & Gutman GA** (1987a). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4: 203-221.
- Levinson G & Gutman GA** (1987b). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Research* 15: 5323-5338.
- Lewin HA** (1996). Genetic organization, polymorphism, and function of the bovine major histocompatibility complex. In: The major histocompatibility complex region of domestic animals species (eds. LB Schook & SJ Lamont), pp. 65-98, CRC Press, Boca Raton, FL, USA.
- Li W-H** (1997). Molecular evolution. Sinauer Associates, Inc., Sunderland, MA, USA.
- Litt M & Luty JA** (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am J Hum Genet* 44: 397-401.
- Liu J, Stanton VP, Fujiwara TM, Wang J-X, Rezonzew R, Chumley MJ et al.** (1995). Large-scale cloning of human chromosome 2-specific yeast artificial chromosomes (YACs) using interspersed repetitive sequences (IRS)-PCR approach. *Genomics* 26: 178-191.
- Lopez JV, Culver M, Stephens JC, Johnson WE & O'Brien SJ** (1997). Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in Mammals. *Mol Biol Evol* 14: 277-286.
- Lopez JV, Yuhki N, Masuda R, Modi W & O'Brien SJ** (1994). *Numt*, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J Mol Evol* 39: 174-190.
- Mack SJ & Erlich HA** (1998). HLA class II polymorphism in Ticuna of Brazil. Evolutionary implications of the DRB1*0807 allele. *Tissue Antigens* 51: 41-50.
- MacKay SLD, Olivo PD, Laipis PJ & Hauswirth WW** (1986). Template-directed arrest of

- mammalian mitochondrial DNA synthesis. *Molecular and Cellular Biology* 6: 1261-1267.
- Madsen CS, Ghivizzani SC & Hauswirth WW** (1993). Protein binding to a single termination-associated sequence in the mitochondrial DNA D-loop region. *Molecular and Cellular Biology* 13: 2162-2171.
- Maillard J-C, Renard C, Chardon P, Chantal I & Bensaid A** (1999). Characterization of 18 new *BoLA-DRB3* alleles. *Animal Genetics* 30: 200-203.
- Majewska K, Szmraj J, Plucienniczak G, Jaworski J & Plucienniczak A** (1988). A new family of dispersed, highly repetitive sequences in bovine genome. *Biochimica et Biophysica Acta* 949: 119-124.
- Markow T, Hedrick PW, Zuerlein K, Danilovs J, Martin J, Vyvial T & Armstrong C** (1993). HLA polymorphism in the Havasupai: evidence for balancing selection. *Am J Hum Genet* 53: 943-952.
- McCarthy L, Hunter K, Schalkwyk L, Riba L, Anson S, Mott R et al.** (1995). Efficient high-resolution genetic mapping of mouse interspersed repetitive sequence PCR products, towards integrated genetic and physical mapping of the mouse genome. *Proc Natl Acad Sci USA* 92: 5302-5306.
- Malik HS & Eickbush TH** (1998). The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINES. *Mol Biol Evol* 15: 1123-1134.
- Mayer WE, Jonker M, Klein D, Ivanyi P, van Seventer G & Klein J** (1989). Nucleotide sequences of chimpanzee MHC class I alleles: evidence for *trans*-species mode of evolution. *EMBO J* 7: 2765-2774.
- Messier W, Li S-H & Stewart C-B** (1996). The birth of microsatellites. *Nature* 381:483.
- Mikko S, Spencer M, Morris B, Stabile S, Basu T, Stormont C & Andersson L** (1997). A comparative analysis of MHC *DRB3* polymorphism in the American Bison (*Bison bison*). *J of Heredity* 88: 499-503.
- Mikko S & Andersson L** (1995a). Extensive MHC class II *DRB3* diversity in African and European cattle. *Immunogenetics* 42: 408-413.
- Mikko S & Andersson L** (1995b). Low major histocompatibility complex class II diversity in European and North American moose. *Proc Natl Acad Sci USA* 92: 4259-4263.
- Miller JR & Archibald AL** (1993). 5' and 3' SINE-PCR allows genotyping of pig families without cloning and sequencing steps. *Mammalian Genome* 4: 243-246.
- Miller JR** (1994). Use of porcine interspersed repeat sequences in PCR-mediated genotyping. *Mammalian Genome* 5: 629-632.
- Modi WS, Gallagher DS & Womack JE** (1996). Evolutionary histories of highly repeated DNA families among the Artiodactyla (Mammalia). *J Mol Evol* 42: 337-349.
- Moore SS, Byrne K, Berger KT, Barendse W, McCarthy F, Womack JE & Hetzel DJS** (1994). Characterization of 65 bovine microsatellites. *Mammalian Genome* 5: 84-90.
- Murata S, Takasaki N, Saitoh M, Tachida H & Okada N** (1996). Details of retropositional genome dynamics that provide a rationale for a generic division: the distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Atlantic salmon and trout (*Salmo*). *Genetics* 142: 915-926.
- Murata S, Takasaki N, Saitoh M & Okada N** (1993). Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINES) as temporal landmarks of evolution. *Proc Natl Acad Sci USA* 90: 6995-6999.
- Murphy GL, Connell TD, Barrit DS, Koomey M & Cannon JG** (1989). Phase variation of conococcal protein II: regulation of gene expression by slippage-strand mispairing of a repetitive DNA sequence. *Cell* 56: 539-547.
- Nei M & Hughes AL** (1991). Polymorphism and evolution of the major histocompatibility complex in mammals. In: *Evolution at the molecular level* (eds. RK Selander, AG Clark & TS Whittam), pp. 222-247. Sinauer Associates, Sunderland, MA, USA.
- Nelson DL, Ledbetter SA, Corbo L, Victoria MF, Ramirez-Solis R, Webster TD et al.** (1989). *Alu* polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc Natl Acad Sci USA* 86: 6686-6690.
- Nikaido M, Rooney AP & Okada N** (1999). Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements. Hippopotamuses are the closest extant relatives of whales. *Proc Natl Acad Sci USA* 96: 10261-10266.

- Nomiyama H, Fukuda M, Wakasugi S, Tsuzuki T & Shimada K** (1985). Molecular structures of mitochondrial DNA-like sequences in human nuclear DNA. *Nucleic Acids Research* 13: 1649-1658.
- Nomiyama H, Tsuzuki T, Wakasugi S, Fukuda M & Shimada K** (1984) Interruption of a human nuclear sequence homologous to mitochondrial DNA by a member of the *KpnI* 1.8 kb family. *Nucleic Acids Research* 12: 5225-5234.
- Nomura O & Yasue H** (1999). Genetic relationships among hippopotamus, whales, and bovine based on SINE insertion analysis. *Mammalian Genome* 10: 526-527.
- Nomura O, Lin Z-H, Muladno, Wada Y & Yasue H** (1998). A SINE species from hippopotamus and its distribution among animal species. *Mammalian Genome* 9: 550-555.
- Nugent JM & Palmer JD** (1991). RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66: 473-481.
- Ohshima K, Hamada M, Terai Y & Okada N** (1996). The 3' ends of tRNA derived short interspersed repetitive elements are derived from the 3' ends of long interspersed repetitive elements. *Molecular and Cellular Biology* 16: 3756-3764.
- Ohta T** (1995). Gene conversions vs point mutation in generating variability at the antigen recognition site of major histocompatibility complex loci. *J Mol Evol* 41: 115-119.
- Okada N** (1991) SINEs: short interspersed repeated elements of the eukaryotic genome. *Trends in Ecology and Evolution* 6: 358-361.
- Okada N & Hamada M** (1997). The 3' ends of the tRNA-derived SINEs originated from the 3' ends of LINEs: a new example from the bovine genome. *J Mol Evol* 44: S52-S56.
- Orita M, Iwahana H, Kanazawa H, Hayashi K & Sekiya T** (1989a). Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms. *Proc Natl Acad Sci USA* 86: 2766-2770.
- Orita M, Suzuki Y, Sekiya T & Hayashi K** (1989b). Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. *Genomics* 5: 874-879.
- Ossario PN, Sibley LD & Boothroyd JC** (1991). Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. *J Mol Biol* 222: 525-536.
- Paterson S** (1998). Evidence for balancing selection at the major histocompatibility complex in a free-living ruminant. *J Heredity* 89: 289-294.
- Paterson S, Wilson K & Pemberton JM** (1998). Major histocompatibility complex variation associated with juvenile survival and parasite resistance in a large unmanaged ungulate population (*Ovis aries L.*). *Proc Natl Acad Sci USA* 95: 3714-3719.
- Polzieln RO & Strobeck C** (1998). Phylogeny of wapiti, red deer, sika deer, and other North American cervids as determined from mitochondrial DNA. *Molecular Phylogenetics and Evolution* 10: 249-258.
- Pereira de Souza A, Allamand V, Richard I, Brenguier L, Chumakov I, Cohen D & Beckmann JS** (1994). Targeted development of microsatellite markers from inter-*Alu* amplification of YAC clones. *Genomics* 19: 391-393.
- Perna NT & Kocher T** (1996). Mitochondrial DNA: molecular fossils in the nucleus. *Current Biology* 6: 128-129.
- Perna NT & Kocher TD** (1995). Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol* 41: 353-358.
- Pesole G, Gissi C, De Chirico A & Saccone C** (1999). Nucleotide substitution rate of mammalian mitochondrial genomes. *J Mol Evol* 48: 427-434.
- Primmer CR, Ellegren H, Saino N & Moller AP** (1996). Directional evolution in germline microsatellite mutations. *Nature Genetics* 13: 391-393.
- Randi E, Mucci N, Pierpaoli M & Douzery E** (1998). New phylogenetic perspectives on the Cervidae (Artiodactyla) are provided by the mitochondrial cytochrome *b* gene. *Proc R Soc Lond B* 265: 793-801.
- Ricchetti M, Fairhead C & Dujon B** (1999). Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* 402: 96-100.
- Rogers J** (1985). The origin and evolution of retroposons. *Int Rev of Cytol* 93: 187-279.
- Russel GC, Davies CJ, Andersson L, Ellis SA, Hensen EJ, Lewin HA et al.** (1997). BoLA class II nucleotide sequences 1996: report of the ISAG BoLA nomenclature committee. *Animal Genetics* 28: 169-180.

- Saccone C, Pesole G & Sbisà E** (1991). The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. *J Mol Evol* 33: 83-91.
- Saitou N & Ueda S** (1994). Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Mol Biol Evol* 11: 504-512.
- Sakamoto K & Okada N** (1985). Rodent type 2 Alu family, rat identifier sequence, rabbit C family and bovine or goat 73-bp repeat may have evolved from tRNA genes. *J Mol Evol* 22: 134-140.
- Sbisà E, Tanzariello F, Reyes A, Pesole G & Saccone C** (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. *Gene* 205: 125-140.
- Schlötterer C & Tautz D** (1992). Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20: 211-215.
- Schon EA, Cleary ML, Haynes JR & Lingrel JB** (1981). Structure and evolution of goat γ , β^C - and β^A -globin genes: three developmentally regulated genes contain inserted elements. *Cell* 27: 359-369.
- Schwaiger F-W, Maddox J, Ballingall K, Buitkamp J, Crawford AM, Dutia BM et al.** (1996). The ovine major histocompatibility complex. In: *The major histocompatibility complex region of domestic animals species* (eds. LB Shook & SJ Lamont), pp. 121-176. CRC Press, Boca Raton, FL, USA.
- Schwaiger FW, Weyers E, Buitkamp J, Ede AJ, Crawford A & Epplen JT** (1994). Interdependent *MHC-DRB* exon-plus-intron evolution in artiodactyls. *Mol Biol Evol* 11: 239-249.
- Schwaiger FW, Buitkamp J, Weyers E & Epplen JT** (1993a). Typing of artiodactyl *MHC-DRB* genes with the help of intronic simple repeated DNA sequences. *Molecular Ecology* 2: 55-59.
- Schwaiger FW, Weyers E, Epplen C, Brün J, Ruff G, Crawford A & Epplen JT** (1993b). The paradox of *MHC-DRB* exon/intron evolution: α -helix and β -sheet encoding regions diverge while hypervariable intronic simple repeats coevolve with β -sheet codons. *J Mol Evol* 37: 260-272.
- Shay JW & Werbin H** (1992). New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutation Research* 275: 227-235.
- She JX, Boehme SA, Wang TW, Bonhomme F & Wakeland EK** (1991). Amplification of major histocompatibility complex class II gene diversity by intraexonic recombination. *Proc Natl Acad Sci USA* 88: 453-457.
- Shedlock AM & Okada N** (2000) SINE insertions: powerful tools for molecular systematics. *BioEssays* 22: 148-160.
- Shimamura M, Abe H, Nikaido M, Ohshima K & Okada N** (1999). Genealogy of families of SINEs in Cetaceans and Artiodactyls: the presence of a huge superfamily of tRNA^{Glu}-derived families of SINEs. *Mol Biol Evol* 16: 1046-1060.
- Shimamura M, Yasue H, Ohshima K, Abe H, Kato H, Kishiro T et al.** (1997). Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* 388: 666-670.
- Shook LB & Lamont SJ** (1996). The major histocompatibility complex region of domestic animal species. CRC Press, Boca Raton, FL, USA.
- Sigurdardóttir S, Borsch C, Gustafsson K & Andersson L** (1991). Cloning and sequence analysis of 14 DRB alleles of the bovine major histocompatibility complex using the polymerase chain reaction. *Animal Genetics* 22: 199-209.
- Sigurdardóttir S, Lundén A & Andersson L** (1988). Restriction fragment length polymorphism of DQ and DR class II genes of the bovine major histocompatibility complex. *Animal Genetics* 19: 133-150.
- Sinnett D, Deragon J-M, Simard LR & Labuda D** (1990). Alu-morphs-human DNA polymorphisms detected by polymerase chain reaction using *Alu*-specific primers. *Genomics* 7: 331-334.
- Sitte K, East IJ & Jazwinska EC** (1996). Detection of a common BoLA-DRB3 deletion by sequence specific oligonucleotide typing. *Animal Genetics* 27: 271-273.
- Skowronski J, Plucienniczak A, Bednarek A & Jaworski J** (1984). Bovine 1.709 satellite. Recombination hotspots and dispersed repeated sequences. *J Mol Biol* 177: 399-416.
- Smit AFA** (1996). The origin of interspersed repeats in the human genome. *Current*

- Opinions in Genetics and Development 6: 743-748.
- Snibson KJ, Maddox JF, Fabb SA & Brandon MR** (1998). Allelic variation of ovine MHC class II *DQA1* and *DQA2* genes. *Animal Genetics* 29: 356-362.
- Sorenson MD & Fleischer RC** (1996). Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. *Proc Natl Acad Sci USA* 93: 15239-15243.
- Spence SE, Young RM, Gerner KJ & Lingrel JB** (1985). Localization and characterization of members of a family of repetitive sequences in the goat β globin locus. *Nucleic Acids Research* 13: 2171-86.
- Stallings RL, Ford AF, Nelson D, Torney DC, Hildebrand CE & Moyzis RK** (1991). Evolution and distribution of (GT)_n repetitive sequences in mammalian genomes. *Genomics* 10: 807-815.
- Sulanderi S, Muladno, Harumi T, Yanai S, Wada Y & Yasue H** (1997). Localization of swine PRE-1 homologues in 13 loci of *Phacochoerus aethiopicus* and *Tayassu tajacu* genomes, and their sequence divergence. *Animal Genetics* 28: 210-215.
- Sun HS, Huff BM, Dentine MR & Kirkpatrick BW** (1994). Polymorphic bovine microsatellites UWCA5, UWCA7 and UWCA9. *Animal Genetics* 25: 59.
- Sunnucks P & Hales DF** (1996). Numerous transposed sequences of mitochondrial cytochrome oxidase I-II in aphids of the genus *Sitobion* (Hemiptera: Aphididae). *Mol Biol Evol* 13: 510-524.
- Suzuki H, Suzuki S, Sakurai T, Kumar S & Ozawa T** (1996). A bovine mtDNA-binding protein to a conserved sequence adjacent to the termination associated sequence in the vertebrate mitochondrial displacement loop region. *Biochem Mol Biol Internat* 38: 275-283.
- Szemraj J, Plucienniczak G, Jaworski J & Plucienniczak A** (1995). Bovine *Alu*-like sequences mediate transposition of a new site-specific retroelement. *Gene* 152: 261-264.
- Takahata N & Nei M** (1990). Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* 124: 967-978.
- Tatout C, Warwick S, Lenoir A & Deragon J-M** (1999). SINE insertions as clade markers for wild crucifer species. *Mol Biol Evol* 16: 1614-1621.
- Tautz D** (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463-6471.
- Terai Y, Takahashi K & Okada N** (1998). SINE cousins. The 3'-end tails of the two oldest and distantly related families of SINEs are descended from the 3' ends of LINEs with the same genealogical origin. *Mol Biol Evol* 15: 1460-1471.
- Thorsness PE, White KH & Fox TD** (1993). Inactivation of *YME1*, a member of the *ftsH-SEC18-PAS1-CSC48* family of putative ATPase-encoding genes, causes increased escape of DNA from mitochondria in *Saccharomyces cerevisiae*. *Mol Cell Biol* 13: 5418-5426.
- Thursz MR, Thomas HC, Greenwood BM & Hill AVS** (1997). Heterozygote advantage for HLA class-II type in hepatitis B virus infection. *Nature Genetics* 17: 11-12.
- Titus-Trachtenberg EA, Rickards O, De Stefano GF & Erlich HA** (1994). Analysis of HLA class II haplotypes in the Cayapa Indians of Ecuador: a novel *DRB1* allele reveals evidence for convergent evolution and balancing selection at position 86. *Am J Hum Genet* 55: 160-167.
- Trowsdale J** (1995). 'Both man & bird & beast': comparative organization of MHC genes. *Immunogenetics* 41: 1-17.
- U'hUigin C** (1995). Quantifying the degree of convergence in primate MHC-DRB genes. In: *Origin of major histocompatibility complex diversity*. Immunological reviews (ed G Möller), no. 143, pp. 123-140. Munksgaard, Copenhagen, Denmark.
- Vaiman D, Mercier D, Moazami-Goudarzi K, Eggen A, Ciampolini R, Lepingle A et al.** (1994). A set of 99 cattle microsatellites: characterization, synteny mapping and polymorphism. *Mammalian Genome* 5: 288-297.
- van der Eijk MJT, Stewart-Hayenes JA & Lewin HA** (1992). Extensive polymorphism of the BoLA-DRB3 gene distinguished by PCR-RFLP. *Animal Genetics* 23: 483-496.
- van der Kuyl AC, Kuiken CL, Dekker JT, Perizonius WRK & Goudsmit J** (1995). Nuclear counterparts of the cytoplasmic mitochondrial 12S rRNA gene: a problem of ancient DNA and molecular phylogenies. *J Mol Evol* 40: 652-657.

- van Haeringen WA, Gwakisa PS, Mikko S, Eythorsdottir E, Holm LE, Olsaker I et al.** (1999). Heterozygosity excess at the cattle *DRB* locus revealed by large scale genotyping of two closely linked microsatellites. *Animal Genetics* 30: 169-176.
- Varvio S-L & Kaukinen J** (1993). Bovine microsatellites: racial differences and association with SINE-elements. In: *DNA Fingerprinting: State of the Science* (eds. DJ Pena, R Chacabarty, JT Epplen & AJ Jeffreys), pp. 437-443. Birkhäuser Verlag, Basel, Switzerland.
- Vaughan HE, Heslop-Harrison JS & Hewitt GM** (1999). The localization of mitochondrial sequences to chromosomal DNA in orthoptera. *Genome* 42: 874-880.
- Watanabe Y, Tsukada T, Notake M, Nakanishi S & Numa S** (1982). Structural analysis of repetitive DNA sequences in the bovine corticotropin- β -lipotropin precursor gene region. *Nucleic Acids Research* 10: 1459-1469.
- Watkins DI, McAdam SN, Liu X, Strang CR, Milford EL, Levine CG et al.** (1992). New recombinant *HLA-B* alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature* 357: 329-333.
- Weber JL** (1990). Human DNA polymorphisms based on length variations in simple-sequence tandem repeats. In: *Genome Analysis Vol. 1: Genetic and Physical Mapping* (eds. KE Davies & SM Tilghman), pp. 159-181. Cold Spring Harbor Lab. Press, New York, USA.
- Weber JL & May PE** (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am J Human Genet* 44: 388-396.
- Weber JL & Wong C** (1993). Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123-1128.
- Weiner AM, Deininger PL & Efstratiadis A** (1986). Nonviral retroposon: genes pseudogenes and transposable elements generated by the reverse flow of genetic information. *Ann Rev Biochem* 55: 631-661.
- Wierdl M, Dominska M & Petes TD** (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* 146: 769-779.
- Wood NJ & Phua SH** (1996). Variation in the control region sequence of the sheep mitochondrial genome. *Animal Genetics* 27: 25-33.
- Xu A, van Eijk MJT, Park C & Lewin HA** (1993). Polymorphism in *BoLA-DRB3* exon 2 correlates with resistance to persistent lymphocytosis caused by bovine leukemia virus. *J Immunol* 151: 6977-6985.
- Yasue H & Wada Y** (1996). A swine SINE (PRE-1 sequence) distribution in swine-related animal species and its phylogenetic analysis in swine genome. *Animal Genetics* 27: 95-98.
- Zangenberg G, Huang M-M, Arnheim N & Erlich H** (1995). New HLA-DPB1 alleles generated by interallelic gene conversion detected by analysis of sperm. *Nature Genetics* 10: 407-414.
- Zanotti M, Poli G, Ponti W, Polli M, Rocchi M, Bolzani E et al.** (1996). Association of BoLA class II haplotypes with subclinical progression of bovine leukaemia virus infection in Holstein-Friesian cattle. *Animal Genetics* 27: 337-341.
- Zardoya R, Villalta M, López-Pérez MJ, Garrido-Pertierra A, Montoya J & Bautista JM** (1995). Nucleotide sequence of the sheep mitochondrial DNA D-loop and its flanking tRNA genes. *Curr Genet* 28: 94-96.
- Zietkiewicz E, Labuda M, Sinnott D, Glorieux FH & Labuda D** (1992). Linkage mapping by simultaneous screening of multiple polymorphic loci using *Alu* oligonucleotide-directed PCR. *Proc Natl Acad Sci USA* 89: 8448-8451.
- Zietkiewicz E, Rafalski A & Labuda D** (1994). Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics* 20: 176-183.
- Zischler H, Geisert H & Castresana J** (1998). A hominoid-specific insertion of the mitochondrial D-loop: implications of reconstructing ancestral mitochondrial sequences. *Mol Biol Evol* 15: 463-469.
- Zischler H, Geisert H, von Haeseler A & Pääbo S** (1995). A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. *Nature* 378: 489-492.
- Zullo S, Sieu LC, Slightom JL, Hadler H & Eisenstadt JM** (1991). Mitochondrial D-loop sequences are integrated in the rat nuclear genome. *J Mol Biol* 221: 1223-1235.
