Pioneering studies on the gene order, DNA sequence and evolution of the MHC class-I region in the new world primate *Callithrix jacchus*

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von
NATALIA REGINA MESA HERRERA
aus Medellin
(Kolumbien)

Göttingen 2007

D 7

Referent: Prof. Dr. rer. nat. Jürgen Wienands

Korreferent: Prof. Dr. rer. nat. Ernst A. Wimmer

Tag der mündlichen Prüfung: am 5. Juli 2007

……To my Family and Martin, who always were behind me,

…… To my Friends, who encouraged me,

…… To my Professors, who believed in me.

Table of contents

		Page
1.	Aims and scopes…………………………………………………………	1
2.	Introduction……………………………………………………………....	3
2.1	Common marmoset *CALLITHRIX JACCHUS*………………………	3
2.2	Taxonomy………………………………………………………………….	4
2.3	The major histocompatibility complex (MHC)……………………….	5
2.3.1	MHC class I molecules………………………………………………...	7
2.3.2	MHC class II molecules………………………………………………..	9
2.3.3	MHC in *Callithrix jacchus* and others New World Monkeys (NWM)	10
2.3.4	MHC in other primates ………………………………………………..	12
2.3.5	Evolution………………………………………………………………….	13
2.4	Framework genes hypothesis…………………………………………	14
2.4.1	ATP6V1G2………………………………………………………………	15
2.4.2	BAT1……………………………………………………………………..	15
2.4.3	TCF19……………………………………………………………………	16
2.4.4	POU5F1…………………………………………………………………	16
2.4.5	TRIM family……………………………………………………………..	16
2.4.6	TCTEX5…………………………………………………………….........	18
2.4.7	MOG …...……………………………………………..………………	18
3	Materials…………………………………………………………………	19
3.1	Chemicals………………………………………………………………	19
3.2	Buffers and other solutions……………………………………………	20
3.3	Media...	23
3.4	Commercials kits……………………………………………………….	24
3.5	Enzymes…………………………………………………………………	24
3.6	Oligonucleotides primers for PCR: sequences and annealing temperatures applied………………………………………………..	25
3.7	DNA ladders and other standards for comparison of molecular size………………………………………………………………………	26
3.8	Vectors…………………………………………………………………..	27
3.9	BAC-clones used for genomic analysis………………………………	27
3.10	The gene sequence of *Callitrix jacchus* used as MHC class I probe…………………………………………………………………….	29

3.11 Computer programs……… 29
3.12 Plastic articles……… 29
3.13 Radioactive substrate……… 30
3.14 Lab equipment……… 30
3.15 Addresses of manufacturers……… 32
4 Methods……… 34
4.1 Preparation of materials……… 34
4.2 Filters of carrying bacterial artificial chromosome (BAC) clones for screening……… 35
4.2.1 Source and origin of BAC bank CHORI of common marmoset (*Callithrix jacchus*)……… 35
4.2.2 Organization of BAC bank CHORI 259 from of the new world monkey common marmoset (*Callithrix jacchus*) on microtiter plates and on nylon filters……… 35
4.2.3 Procedure for hybridization on high density filters……… 35
4.3 Preparation of a BAC clone cultures……… 37
4.4 Isolating plasmid DNA (*Miniprep)*……… 38
4.5 Quantification of DNA……… 39
4.5.1 UV-absorption spectroscopy method……… 39
4.5.2 Minigel method……… 40
4.6 Restriction digestion of plasmid DNA ……… 41
4.7 Separation of large DNA fragments by agarose gel electrophoresis……… 42
4.8 Extraction of DNA fragments from agarose gels……… 43
4.9 Random primer method……… 46
4.10 Analysis of plasmid DNA by blotting and hybridization……… 46
4.10.1 Southern blotting……… 46
4.10.2 Hybridization analysis of DNA……… 48
4.11 Autoradiography……… 49
4.12 Polymerase chain reaction (PCR)……… 49
4.13 Inverse PCR……… 51
4.14 DNA sequencing ……… 52
5 Results……… 54

5.1	Identification of BAC clones from *Callithrix jacchus* containing MHC class-I region genes …………………………………………	55
5.1.1	Contig 1, the interval BAT1 - TCF19………………………………	58
5.1.2	Contig 3, the intervals TRIM39 - TRIM26…………………………	59
5.1.3	Contig 4, interval TCTEX4 – MOG…………………………………	61
5.2	Genomic analysis by Southern blot hybridization …………………	62
5.2.1	Contig 1, the interval BAT1 – TCF19………………………………	62
5.2.1.1	Southern blot analysis of contig 1…………………………………	62
5.2.1.2	Evaluation of hybridization results (contig 1)………………………..	77
5.2.1.3	Construction of a contig 1 map……………………………………	81
5.2.2	Contig 3, the interval TRIM26 – TRIM39……..……………..…….....	84
5.2.2.1	Southern blot analysis of contig 3…………..……………………….	84
5.2.2.2	Evaluation of hybridization results (contig 3)…..……………………	87
5.2.2.3	Contribution to a contig 3 map..……………………………………	87
5.2.3	Contig 4, the intervall TCTEX4 - MOG …..…………………………	89
5.2.3.1	Southern blot analysis of contig 4…………………………………	89
5.2.3.2	Evaluation of hybridization results (contig 4)………………………..	91
5.2.3.3	Contribution to a contig 4 map..……………………………………	91
5.3	PCR and sequence analysis of BAC clone inserts….……………...	92
5.3.1	Polymerase chain reaction (PCR) analysis………………………	92
5.3.2	Approximative alignment of sequences to homologoues DNA of Hs and other mammals by BLAST…………….………………………..	96
5.3.3	Sequences of contig 1………………………………………………	99
5.3.4	Sequences of contig 3………………………………………………	103
5.3.5	Sequences of contig 4……..………………………………………	106
5.4	Identification of the *Callithrix jacchus* sequences by comparative alignment………………………………………………………………..	106
5.4.1	Programs used for identification of sequences by localization and gene assignment…………………………………………………….....	106
5.4.2	Collection and presentation of sequence data……………………	113
5.4.3	Demonstration of individual sequences identified by localization and possibly gene assignment in the Hs genome or proteome sequence K19161C12-TS2……………………………………………	114

5.4.4	K28-347D1-TCF19, nucleotide sequence…………………………..	116
5.4.5	MOG sequences………………………………………………………..	118
5.5	Evolution of class-I sequences and framework genes found in *Callithrix jacchus* compared to their homologous in other primates and mammals………………………………………………………..	120
5.5.1	Gene trees class-I of genes…………………………..……………….	120
5.5.2	Gene trees of framework genes………………………………………	122
5.6	TS1: teachings and challenges………………………………………	124
5.6.1	A gene map of TS1 and its homologous genomic vicinities in the duplicons B and C of Hs……………………..………………………..	124
5.6.2	The two new class I pseudogenes present in the HLA-B and -C duplicons of Hs, Patr and other mammals, also in the 14 Mamu-B duplicons………………………………………………………………...	127
5.6.3	A MHC ferritin heavy chain pseudogene in *Callithrix jacchus*……..	134
5.6.4	Characterization and possible interpretations of “Vorspann” v, the sequence next to which a ferritin transposon has been inserted….	141
6	Discussion……………………………………………………………	142
6.1	Physical map of the intervals from MHC class-I…………………….	144
6.1.1	Physical map of the BAT1 – TCF19 interval (contig 1)…………….	144
6.1.2	Physical map of the intervals CAT56 – TRIM39 and TRIM 39 – TRIM26 (contig 2 and contig 3 respectively)………………………...	145
6.1.3	Physical map of interval TCTEX4 and MOG (contig 4)……………	145
6.2	Sequence analysis of BAC clones……………………………………	146
6.2.1	BAT1 gene in *Callithrix jacchus*………………………………………	146
6.2.2	TCF19 gene in *Callithrix jacchus*……………………………………	146
6.2.3	MOG gene in *Callithrix jacchus*………………………………………	146
6.3	On alignment of sequences in general and in particular those of the *Callithrix jacchus*…………………..……………………………….	147
6.3	On pseudogenes in the MHC class I region…………………………	147
6.4	On duplicons in the MHC class I region…………..…………………	148
6.5	On transposon especially the pseudo ferritin found in MHC Caja transposon in the evolution of MHC…………………..…………...	148
6.6	On potential value of sequence TS1 for molecular immunogenetics……………………………………………………...	148

7 Summary………………………………………………………………... 150
8 References..………………………………………………..….......... 152
9 Annex: The 55 *Callithrix jacchus* sequences……………………….. 175
Acknowledgements
Lebenslauf

……If you cannot study function, study structure!

Francis Crick

1. Aims and scopes

The object of these pilot studies is to contribute to the structural analysis of the MHC class I region in a lower primate, the new world monkey *Callithrix jacchus* (Caja).

The aims are:

- Selection of BAC clones containing inserts apt to genomic and sequence analysis of the MHC class I region;
- Establishing the contigs belonging MHC class I in *Callithrix jacchus* of inserts for obtaining a restriction map and hybridization with class I probe and framework gene, this framework gene is flanking the MHC class I region.
- Typifying BAC clone inserts selecting various primers by PCR. This knowledge should be useful for a complete walking chromosome sequencing of the entire class I region in Caja having been started recently by Prof. T. Shiina;
- Identification and localization of the 55 sequences obtained by alignment in the Hs genome and other species.
- Using the special opportunity of studying one sequence, TS1 containing apparently three different kinds of pseudogenes, interesting by an arrangement on both DNA strands, much more however, as they seem to concern three unsolved basic problems in immunogenetics, i.e. gene clustering by duplication, control of gene expression by promotor regions (5'FR), and transposition of (pseudo)genes.

The scope of these pilot investigations in some parts is methodology, like

- testing an unusual way of concluding a gene order,
- analyzing partial information about MHC class I region in Caja,
- supporting it by localization and assignment of 55 Caja sequences obtained.

Basically, intentionally and successfully, however, the scope of these studies is t h e o r y. This becomes obvious in studies of, and suggested by the Caja sequence K161C12-TS1:

(1) By its two class I pseudogenes, it seems to offer a useful key for the analysis of the known gene clustering by duplication, may be even of the descendence and

evolution of their neighbors, the HLA genes B, C and others in Hs and Patr - likewise the 14 B genes in *Macaca mulatta*

(2) Second, a principle for the insertion of transposons might be exemplified structurally by the rather young ferritine pseudogene apparently unique to Caja being

(3) inserted next to a sequence v that by comparison with other 5'FR may be suspected of being an ancient promotor region. Strangely enough, read in the opposite direction, on its complementary strand, it turns out to be highly identical and almost completely coincident with the Hs class I pseudogene.

2. Introduction

2.1 Common marmoset *Callithrix jacchus*

The word "marmoset" is an old French word for "grotesque figure". The South American primate genus *Callithrix* comprises two groups of species, namely, the subgenera of *jacchus* and *argentata*. Five taxa have been proposed for the jacchus group: *C.j. jacchus, C.j. aurita C.j., penicillata, C.j.kuhlii,* and *C.j. geoffroyi;* these, however, at present still under discussion (Nagamachi *et al.*, 1997). The *jacchus* group lives in the Atlantic forest of eastern Brazil, in the states of Maranhão and Ceará northeast of São Paulo, see Figure 2.1 (Nagamachi *et al.*, 1997). The animals are inhabitants of the canopy of secondary rain forests and gallery forests.

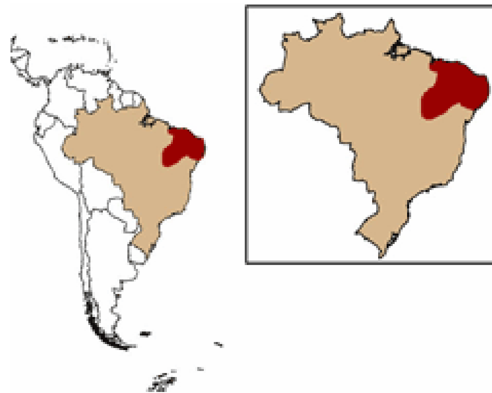


Figure 2.1: Area of distribution of Callithrix jacchus in the Atlantic forest of eastern Brazil, the states of Maranhão and Ceará northeast of São Paulo (http://pin.primate.wisc.edu/factsheets/entry/common_marmoset/taxon)

They are the smallest higher primates, characterized by a small body of about 50 cm length head to tail and 350 – 450 g average weight (Warren *et al.*, 1997). This species is easily identified by the (two) white tufts of hair on both sides of the head, and a white blaze on the forehead. The body is covered with soft thin hair, grayish in color with darker stripes. The tail has darker-colored rings. Females are a little larger than males. Common marmosets have large chisel-like incisors used for gnawing holes in trees, the canines are shaped more like incisors (Groves, 2001). The dental formula of common marmoset is 2:1:3:2 on both the upper and lower jaw (Martin, 1990). The labium major of the female resembles the scrotum of the male. There is little sexual dimorphism in this species. The common marmoset has no opposable thumbs. It is omnivorous, feeding on a variety of food-stuffs, including fruits, insects,

snails, and small vertebrates such as lizards. To eat gum it gouges holes in tree trunks with its large incisors and laps the exudates from the hole.
(www.damisela.com/zoo/mam/primates/callitrichidae/jacchus/taxa.html)
Its main predators are hawks and small carnivorous mammals such as the tayra (Stevenson and Rylands, 1998) (http://www.saudeanimal.com.br/sagui.html).

Figure 2: Marmoset (*Callithrix jacchus*).
(*Photo courtesy of Raimond Spekking, Wikimedia Commons*, www.genome.gov.).

Callithrix jacchus monkeys begin their activities early in the morning until the dusk; they live on the branches of the trees using their claws like adapted nails. Their territory is marked by the scents that unfold from the glands of the chest, abdomen, the genitals and urine, these marks allow among them the identification of the species, relationship, and receptivity sex. The groups generally include 8 to 12 animals and their social structure is very varied. The mating pairs are monogamous. The time of gestation is 140 to 150 days, and of one to three individuals are born (http://www.univie.ac.at/zoologie/theo/marmoset/calli.html)

2.2 Taxonomy

Common Name: white-tufted-ear marmoset [English]
Kingdom: **Animalia**
Subkingdom: **Eumetazoa**
Branch: **Bilateria**
Phylium: **Chordata**
Sub-phylium: **Vertebrata**
Superclass: **Gnathostomata**

Class: **Mamalia**

Subclass: **Eutheria**

Order: **Primates**

Suborder: **Haplorrhini**

Infra-order: **Simiiformes**

Section: **Platyrrhini**

Family: **Callitrichidae**

Subfamily: **Callitrichinae**

Genus: ***Callithrix***

Species: ***Callithrix jacchus***

(http://www.damisela.com/zoo/mam/primates/callitrichidae/jacchus/taxa.html)

(http://www.itis.usda.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=572915)

2.3 The major histocompatibility complex (MHC)

MHC is a group of polymorphic genes or multigene family of immune system genes that code cell surface glycoproteins which are found on the surface of cells which present peptides to circulating T cells and which help the immune system to recognize foreign substances (Go *et al*., 2003). In 1940, the geneticist George D. Snell performed experiments of grafts of skin in mice, demonstrating that an immune reaction of the animal against the grafted weave cause the rejection (Snell, 1948; lein, 2001). The term histocompatibility comes from the Greek word *histos* = tissue and the word "compatibility" in order to talk about the molecules with a function in the reaction of transplants although this is not its true physiological role (Snell, 1948; Snell, 1951).

MHC proteins are found in all higher vertebrates, from fish and birds to non-human primates up to the human. They have been shown to display the levels of genetic polymorphism and further the molecular evolutionary processes which generate the diversity required to the immunity. The complex is called in the human the human leukocyte antigen (HLA) system, for non-human species LA is used the prefixed species name, e.g. BoLA (The MHC Sequencing Consortium, 1999).

In the human, the MHC spans almost four megabases (4.000.000 bp) on the short arm of chromosome 6, band p21.3, it includes approximately 224 coding and non-coding sequences, of which about half have known immunological functions. It also

includes 96 pseudogenes, non-functional gene remnants (Kulski *et al.*, 2005; Beck *et al.*, 2000; Go *et al.*, 2003; The MHC Sequencing Consortium, 1999).

The MHC has been associated with a role in influencing propensity for known autoimmune diseases such as insulin-dependent diabetes mellitus, multiple sclerosis, systemic lupus erythematosus, myasthenia gravis, and rheumatoid arthritis (Wong *et al.*, 2005; Bach, 2005; Lie, 2005; Reveille, 2000; Azizah *et al.*, 2004). The MHC contains genes contributing to several other hereditary disorders that are either not autoimmune in nature or in which the role of autoimmunity is uncertain. These include ankylosing spondylo arthropathies, where there is a clear association with the class I allele HLA-B27, and narcolepsy, equally clearly associated with certain class II alleles (Luthra-Guptasarma and Singh, 2004). The MHC also includes the genes for steroid 21-hydroxylase and hemochromatosis (Nardi *et al.*, 2003). Hemochromatosis is one of the most common simple Mendelian disorders of man, it has been found a 250-kilobase region more than three megabases telomeric of the major histocompatibility complex (MHC). It has been identified related with HLA-H containing a two missense alterations (Feder *et al.*, 1999; Weitkamp *et al.*, 1994). Recently, two major quantitative trait loci for dyslexia have been mapped distal to the MHC (Stein, 2001) and also for Parkinson's disease (Eslamboli, 2005).

The MHC is divided into three subgroups called MHC class I, MHC class II and MHC class III regions, each containing groups of genes with related functions. The MHC has shown conservation of paralogous genes in the class I, II and III regions, but differences in position, type and number of immunoglobulin-related genes (Amadou *et al.*, 1999; The MHC sequencing consortium, 1999; Daza-Vamenta *et al.*, 2004; Hurt *et al.*, 2004; Anzai *et al.*, 2003; Neff, 2005). The MHC, particularly in the human, displays strong allelic diversity therein, especially among the nine classical HLA genes. The most conspicuously diverse loci are present in human HLA-A, HLA-B, and HLA-DRB1, with roughly 472, 805, and 256 known alleles, respectively (http://www.ebi.ac.uk/imgt/hla/stats.html).

The molecular genetic structure of the MHC is considered to be an evolutionary result of selective pressure imposed by infectious microorganisms (Ploegh, 1998). Several studies have compared the genomic organization of the MHC in (other) organisms, as mouse, chicken, rat, cat, zebra fish and Fugu fish (Trowsdale, 1995; Timon *et al.*, 1998; Graser *et al.*, 1999; Yuhki *et al.*, 2003) as well as in more primitive members of the chordates.

The way of “inherited in block” is used by the MHC producing haplotypes, although with rare exception there are events of recombination (Trowsdale, 1995).
A haplotypical association is usually stronger and more meaningful than an allelic association. The co-dominant expression and haplotypical transmission have an important consequence: within a family, HLA-identical sibling frequency should be 25% according to Mendelian expectations (Alper *et al.*, 1992; Degli-Esposti *et al.*, 1992; Gaudieri; *et al.*, 1997).
This means that certain alleles tend to occur together in the same haplotype rather than randomly segregating together. This is called linkage disequilibrium (LD) and is quantitated by a Δ value (Mattiuz *et al.*, 1971 and Begovich *et al.*, 1992).

2.3.1 MHC class I molecules

The MHC class I encodes heterodimeric peptide-binding glycoproteins as well as antigen-processing molecules such as TAP and tapasin (Leonhardt *et al.*, 2005; Herberg *et al.*, 1998; Pamer and Cresswell, 1998). The molecules contain two non-covalently linked polypeptide chains, i.e., one MHC class I encoded heavy chain, called the $\propto$ chain of 44 to 47 kD with the domains ($\propto$1, $\propto$2 and $\propto$3) and one non-MHC encoded subunit, a β_2-microglobulin of 12 kD. The molecular complex has four domains, three formed from the MHC-encoded $\propto$ chain, and one of the β_2-microglobulin (Jones *et al.*, 1988).
Three quarters of the total polypeptide - the $\propto$ chain - form an extracellular chain, a short segment is the trans-membrane, the carboxy terminal end extends into the cytoplasm, see Figure 3.
The peptide-binding site has a size of approximately 25 Å x 10 Å x 11 Å. It is formed by two amino terminal segments $\propto$1 and $\propto$2 of the $\propto$ chain, both approximately 90 residues long. These interact to form an eight-stranded, antiparallel β-pleated sheet, the bottom supporting the two parallel $\propto$-helical strands. The variations among different class I alleles, important for peptide binding and T cell recognition, are contributed by the polymorphic or variable amino acid residues of class I proteins confined to the $\propto$1 and $\propto$2 domains. Peptides that bind to the groove of an MHC class I molecule are usually 8 – 10 amino acid residues long, and interact with their back bone by a series of hydrogen bonds and ionic interactions at both ends. A cluster of tyrosine residues, common to all MHC class I molecules, form hydrogen bonds to the amino terminus of the peptide bound, while binding to the carboxy end

terminus of the peptide back bone in addition to the carboxy terminus itself (Flutter and Gao 2005; Koch and Tampe, 2006; Sullivan *et al.*, 2006).

The ∝3 chain folds into an Ig domain, a part of the amino acid sequence, which is conserved among all class I molecules. This segment contains a loop serving as binding site for CD8. Near the carboxy terminal end is a stretch of approximately 25 hydrophobic amino acids that traverses the lipid bilayer of the plasma membrane. Next to this there are 30 residues located in the cytoplasm, a cluster of basic amino acids that interact with the phospholipid head groups of the inner sheet of the lipid bilayer, thereby, anchoring the MHC molecule in the plasma membrane (Flutter and Gao, 2005; Koch and Tampe, 2006; Sullivan *et al.*, 2006), see Figure 3.

The β_2-microglobulin encoded by a gene outside the MHC is the light chain combining with a class I molecule, interacting non-covalently with the ∝3 segment. The β_2-microglobulin is the invariant component of all class I molecules.

The purpose of the second-mentioned surveillance system is to identify abnormal body cells, such as those infected with viruses, or those which have turned malignant. When such cells display unfamiliar peptide antigens, e.g. fragments of viral proteins, they are attacked and destroyed (Flutter and Gao, 2005; Germain, 1995; Hamilos, 1989;Jones *et al.*, 1988).

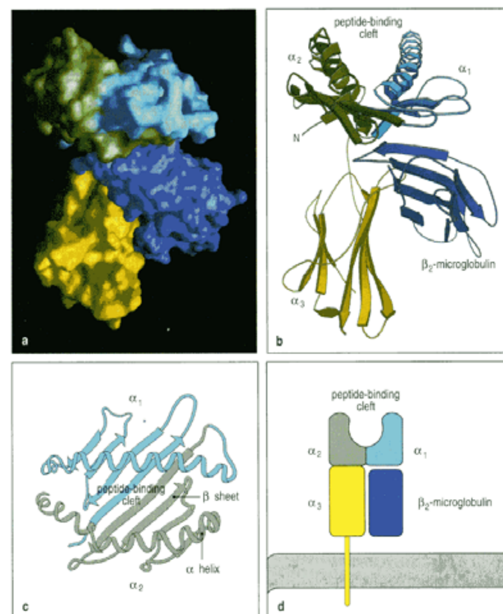


Figure 3. The structure of MHC class I protein with bound to β_2 microglobulin as determined by X-ray crystallography.

Panel (a) shows a computer graphic representation of the two subunit-comples, of human MHC class I molecule, panel (b) a ribbon diagram. Panel (c) gives a view of its tertiary structure, looking down on the molecule from above; the sides of the cleft are formed by the inner faces of the two ∝ helices. Panel (d) shows a schematic representation of the extracellular parts of an MHC class I molecule (Janeway *et al.*, 1999).

2.3.2 MHC class II molecules

The molecules of the MHC class II constitute a heterodimer that contains an ∝ chain and a β chain. The two portions alpha helixes are on a beta-pleated, both form a cleft, that allows the union specifies of peptides. This portion presents a variable amino acid sequence with respect to the rest of the molecule (Castellino, 1997).

The molecules of the MHC class II are DP, DQ and DR, although the highly polymorphic peptides are recognized by receptors of CD4 T cells. The pathogens and extracellular proteins phagocyted in intracellular vesicles derive the peptide antigens that present these molecules class II (Castellino, 1997).

MHC class II molecules expressed in the thymus also have a vital role in the intrathymic maturation of CD4 T cells. MHC class II molecules are expressed constitutively on antigen-presenting cells, including B lymphocytes, macrophages, and dendritic cells (Lundberg, 1997).

They are also expressed on the epithelial cells of the thymus and their expression can be induced on other cells, principally by the cytokine interferon-γ. T cells also express MHC class II molecules when they are activated. Expression of the genes encoding the ∝ and β chains of MHC class II molecules must be strictly coordinated and is under complex regulatory control. The regulation of MHC class II gene expression is not fully understood as it involves the action of transcription factors that are defined only in part. The existence of these transcription factors and a means of identifying then were first suggested by the study of patients with MHC class II deficiency. Many diseases are associated with the MHC class II region (Jones *et al.*, 2006; Germain, 1995; Hamilos *et al.*, 1989).

The structure of class II differs from that of class I in the absence of remainders of tyrosines in the end of the cleft. These are replaced by remainders of glycine and valine that have so large minor (Castellino, 1997; Lundberg and McDevitt 1992).

2.3.3 MHC in *Callithrix jacchus* and others New World monkeys (NWM)

The New World Monkey (NWM) and Old World simians radiated about 58 millions years ago (Ciochon and Chiarelli, 1980). The comparison with the 'classical' (HLA-A, B and C) and three conserved 'non-classical' (HLA-E, F and G) MHC class I genes of apes and monkeys have shown in their differention been preserved during primate evolution. These six human genes have orthologues with African apes. Several *MHC* class I genes of Asian apes have shown significant differences from the human genes which continues with the Old World monkeys, and even more so in the New World monkeys, where *E* and *F* are the only human gene orthologues. The locus C is confined to humans and is the result of a recent duplication of the B locus (Boyson *et al*., 1993; Boyson *et al*., 2001; Knapp *et al*., 1998).

Common marmoset has an evolutionary distance from humans of 55 million years (Bontrop *et al*., 1995). Immunological studies have shown that its immune system is a particularly good model when compared to other primates for testing antibody specificity and recognition (Genian and Hauser, 2001). Others NWM have been studied as animal models in multiples diseases, such as *Aotus*, *Saguinus*, *Saimiri*, principally, and other species (Cadavid *et al*., 1997; Patarroyo *et al*., 1987; Diaz *et al*., 2000; Middleton *et al*., 2004).

Homologies of MHC class I genes are present in all Catarrhini (old world apes and the human). However, the Callitrichinae (tamarins and common marmoset) are exceptions to the rules of MHC stability. This specie *Callithrix jacchus* (common marmoset) has Caja-G 01 until Caja-G 05 and Caja-E genes MHC class I, these sequences gene have been compared with others NWM which form separated clusters in the phylogenetic tree (Cadavid *et al*., 1997). Their cDNAs indicate that there is no orthology between MHC class I loci in the genera of this phyletic group, furthermore, their MHC class exhibits limited variability, probably the result of the recent origin of these loci (Cadavid *et al*., 1997).

MHC class II region genes have limited polymorphism, it has been expanded by reactivation of pseudogene segments as a result of exon shuffling (Doxiadis *et al*., 2006). This limited variability may contribute its susceptibility to particular bacterial infections (klebsiella, bordetella, clostridium and shigella) (Potkay, 1992). These encode the evolutionary equivalents of human HLA-DR and DQ molecules (Bontrop *et al*., 1999), but the Caja DP region has been inactivated (Antunes *et al*., 1998). The

Caja DR region contains only three loci: Caja-DRB*W12, DRB*W16 and DRB1*03 (Wu *et al*., 2000; Antunes *et al*., 1998; Prasad *et al*., 2006).

Several *Aotus* species (*nancymaae, nigriceps* and *vociferans*) have been studied for MHC class II. This work has revealed a high homology between MHC-DR and HLA-DRB (Suarez *et al*., 2006; Niño-Vasquez *et al*., 2006). The sequence of MHC class I of exon 2 and 3 which was divided into two groups Aog1 and Aog2, their analyses showed that displays to similar characteristics to Catarhini's classical loci. This sequence was found as a processed pseudogene (Aona PS2) (Cadavid *et al*., 1997).

Most genetic and functional work on class I genes has been on the cotton top tamarin (*Saguinus Oedipus*) (Cadavid *et al*., 1997; Shneider *et al*., 1993); this species has revealed only 11 different expressed class I molecules. The selective pressure on MHC class I molecules and the peptide binding of these molecules may drive the generation of MHC class I polymorphism, furthermore, the gene conversion occurring in genetic exchange not only occurrs between alleles of locus but also between loci (Watkins *et al*., 1991; Watkins *et al*., 1991; Watkins *et al*., 1996). Alvarez-Tejado *et al*. (1998) found data suggesting that sequences of MHC (Mhc-Saoe CR*01 and CR*02) may be related to the Mhc-C locus. However, they share only a few of the conserved residues (from gorilla and human) which suggests that the relationship to an ancestor of the MHC-C lineage is very distant or is a product of convergent evolution to perform a C locus related to function, and the locus C is the result of a duplication of the B locus in great apes and human (Boyson *et al*., 1996). The pseudogenes found in these species, PS1 and PS2 have been characterized from class I mRNA not functionally expressed template. PS2 has been found in *Callithrix jacchus* and *Aotus trivirgatu* (Cadavid *et al*., 1996). Other processed have found likely true pseudogenes (Gp12, Gp13, Gp14) and partial pseudogenes (Gp15, Gp16, Gp17, Gp18, Gp19 and Gp20) (Cadavid *et al*., 1999).

The characterization and phylogenetic analysis of sequences of the MHC class I of new world monkey *Saimiri sciureus* showed that they are related to HLA class1 genes (HLA A and G), and the structure and the organization of one clone was similar to HlA-A_2 (Pascalis *et al*., 2003).

2.3.4 MHC in other primates

Several non-human primates of the Old World monkey (OWM) have been used in biomedical research as a model, for multiples diseases: (see following Table 2.3.5 Adams and Parham, 2000). Some of these diseases (immuno-related) have needed the characterization of their MHC system (Otting *et al.*, 2002).

Table 2.3.5 Adams and Parham, 2001.

"The relationship of non-human primate *MHC* class I loci to those expressed in humans. Orthologous loci are indicated by blue shading, homologous loci are indicated by yellow shading. Hatched coloration indicates the presence of an additional homologous locus within that species".

	Species	Common name	*HLA-A*	*HLA-B*	*HLA-C*	*HLA-E*	*HLA-F*	*HLA-G*	*HLA-H*	Unrelated
African Apes	*Pan troglodytes*	Common chimpanzee								?
	Pan paniscus	Bonobo				n.d	n.d.	n.d.		?
	Gorilla gorilla	Gorilla								?
Asian Apes	*Pongo pygmaeus*	Orangutan					n.d.		?	?
	Hylobates Lar	Gibbon			?	n.d.	n.d.	n.d.	?	?
Old World Monkeys (Catarrhini)	*Macaca mulatta*	Rhesus macaque			?				?	?
	Macaca fascicularis	Crab-eating macaque or Cynomologous monkey	n.d.		?		n.d.	n.d.	?	?
	Macaca arctoides	Stump-tailed macaque	n.d.		?	n.d.	n.d.	n.d.	?	?
	Papio cynocephalus	Yellow baboon			?		n.d.	n.d.	?	?
	Papio cynocephalus (hamadryas) anubis	Olive baboon			?	n.d.	n.d.	n.d.	?	?
	Cercopithecus aethiops	African green monkey	?	?	?	n.d.	n.d.		?	n.d.
New World Monkeys (Platyrrhini)	*Ateles belzebuth*	Long-haired spider monkey	?	?	?		n.d.	?	?	
	Ateles fusciceps	Brown-headed spider monkey	?	?	?		n.d.	?	?	n.d.
	Pithecia pithecia	White-faced saki	?	?	?		n.d.	?	?	
	Saimiri sciureus	Common squirrel monkey	?	?	?	n.d.	n.d.	?	?	
	Aotus trivirgatus	Owl monkey, Douroucouli	?	?	?		n.d.	?	?	
	Saguinus mystax	Moustached tamarin	?	?	?	n.d.	n.d.	?	?	
	Saguinus geoffroyi	Geoffrey's tamarin	?	?	?	n.d.	n.d.	?	?	
	Saguinus fuscicollis	Brown-headed tamarin	?	?	?	n.d.	n.d.	?	?	
	Saguinus oedipus	Cotton-top tamarin	?	?	?			?	?	
	Leontopithecus rosalia	Golden lion tamarin	?	?	?	n.d.	n.d.	?	?	
	Callithrix jaccus	White-tufted-ear marmoset	?	?	?		n.d.	?	?	

Macaca mulatta monkey is known like rhesus monkey, their MHC class I region has been described to have a length of 3.28 Mb and 64 genes: 23 are expressed and 41 pseudogenes, however, have been difficult to define if has identified orthologous of the human HLA A, B, E, F and G genes (Boyson *et al.*, 1996; Boyson *et al.*, 1995; Go *et al.*, 2002; Otting *et al.*, 2002; Daza-Vementa *et al.*, 2004; Otting *et al.*, 2005; Shiina *et al.*, 2006). The rhesus monkeys, and possibly the Old World monkeys in general diverged from humans 27–30 Million years (Myr) ago. Rhesus monkeys were found to not have the pair of MHC-B and MHC-C but many repeated genes similar to MHC-

B. These results support the inference that MHC-B and MHC-C duplicated after the divergence between apes and Old World monkeys (Fukami-Kobayashi *et al*., 2005; Kulski *et al*., 2005; Kulsi *et al*., 2004; Kulski *et al*., 2002).

In Chimpanzee (*Pan troglodytes*) has been found the structural and/or functional orthologues for all human *HLA* genes (*HLA-A*/*B*/*C*/*E*/*F*/*G* vs. *Patr-A*/*B*/*C*/*E*/*F*/*G*) (Anzai *et al*., 2003; Adams and Parham, 2001). However, it has been assumed to be closely linear to that of human, although, the genomic architecture of chimp MHC is unknown (Anzai *et al*., 2003).

The work carried out on chimpanzee, rhesus monkey and human has revealed that MHC-B and MHC-C duplicated 22.3 Myr ago, and the ape MICA and MICB duplicated 14.1 Myr ago (Fukami-Kobayashi *et al*., 2005; Kulski *et al*., 2005; Kulski *et al*., 2002).

2.3.5 Evolution

The diversity was created and maintained by evolutionary forces: selection, mutation, genetic drift and migration. The frequency dependent selection and heterozygote advantage were two types of balancing selection that have been suggested to explain MHC allelic diversity (Harlt, 2001).

The MHC shows a high degree of polymorphism (100 times greater than the genome average, i.e. a 10 % difference between any two unrelated individuals). One hypothesis holds that the MHC class I originated first as a result of a recombination between an immunoglobulin-like C-domain and the peptide-binding domain of an HSP70 heat-shock protein (Flajnik *et al*., 1991). A phylogenetic analysis supports a relationship between the MHC class II-alpha chain and β_2 microglobulin and between the MHC class II beta-chain and the class I alpha chain (Hughes *et al*., 1993).

About 370 million years or more have been suggested for the physical linkage of MHC regions because of the presence of all three classes of I, II and III genes in the amphibian Xenopus (Kaufman *et al*., 1990).

There is no single definite candidate for a primordial MHC gene. According to one hypothesis, the MHC class II evolved first, whereas to a second hypothesis the MHC loci do not always exist in a single linked cluster as they do in mammals, but can be found in two or multiple clusters (Miller *et al*., 1994; Bingulac-Popovic *et al*., 1997). In vertebrates, the immune system is an adaptive defense system with its components MHC, TCR and immunoglobulin (Ig) genes. In invertebrates, however, the innate

immune system is the only defensive system. These main components of the adaptive immune system are missing not only in invertebrates but also in primitive 'jawless' vertebrates (Matsunaga *et al.*, 1998; Klein *et al.*, 1998).

MHC class I genes do not show an orthologous (i.e., homologous by descent from a common ancestral locus) relationship between mammals of different orders whereas orthologous relationships have been found, among mammalian class II loci 6 (Huges and Nei, 1990). The HLA-C locus has been found only in the human, gorilla and chimpanzee but not in other monkeys (Boyson *et al.*, 1996). The MHC class I proteins however, have been shown to be present in *Callithrix jaccus* (Cadavid *et al.*, 1997).

New genes are created by repeated gene duplication and some duplicate genes are maintained in the genome for a long time, while others are deleted or become non-functional by deleterious mutations. This concept disagrees with the earlier idea that MHC diversity and evolution are governed by concerted evolution of the multigene families of the major histocompatibility complex (MHC) genes and the immunoglobulin (Ig) genes. The alleles seem to have a fast turnover rate. The lack of correspondence between the alleles in human and chimpanzee suggests that five million years of separation have been sufficient to reconfigurate MHC alleles. This means that the alleles are constantly undergoing modifications during their trans-species evolution (Parham and Otta, 1996).

2.4 Framework genes hypothesis

The framework hypothesis can explain the non-orthologous class I sequences are occupy homologous locations to the conserved genes; the identification of non-class I genes allows a comparative map to be drawn, which shows the orthologous class I region. That was showed in the class I region between human and mouse, but the definition of the class I framework applies to all mammals. The class I framework could help define the ancestral MHC, by trancing the conserved genes earlier in the phylogeny and their linkage with MHC genes (Amadou, 1999).

The proximal to the distal part of the class I region, conserved genes mapped in several species (Human, mouse, rat, chimpanzee, rhesus monkey, lemur, pig, chicken) are: BAT1, POU5F, TCF19, GNL1, TRIM26, TRIM39, TCTEX5 and MOG.

2.4.1 **ATP6V1G2**: ATPase, H+ transporting, lysosomal 13kDa, V1 subunit G2

Gene aliases: NG38; ATP6G; VMA10; ATP6G2

A vacuolar ATPase (V-ATPase) is encoded by the ATP6V1G2 gene, it is an enzyme that mediates acidification of intracellular compartments of eukaryotic cells. It is necessary for intracellular processes such as protein sorting, zymogen activation, receptor-mediated endocytosis, and synaptic vesicle proton gradient generation. The composition of V-ATPase protein is a cytosolic V1 domain and a trans-membrane V0 domain. The V1 domain has three A and three B subunits, two G subunits plus the C, D, E, F, and H subunits and contains the ATP catalytic site. The V0 domain consists of five different subunits: a, c, c', c'', and d. Vacuolar type ATPases are involved in bone resorption, glycosylation in the Golgi, degradation of cellular debris in lysosomes, and the processing of endocytosed receptor-ligand complexes. By sequence analysis, the ATP6G2 gene was mapped to chromosome 6p21.3, approximately 1 kb telomeric to NFKBIL1 and centromeric to BAT1. The full-length 118-amino acid protein is 82 % similar to the cow protein. The other isoform encodes a 77-amino acid protein. RT-PCR analysis detected expression in lymphocytic but not monocytic or macrophage-like cell lines (Neville and Campbell, 1999).

2.4.2 **BAT1**: HLA-B associated transcript 1

Gene aliases: D6S81E, UAP56

This gene encodes the protein by a member of the DEAD protein family of ATP-dependent RNA helicases. By chromosome walking with overlapping cosmids isolated a 435-kb DNA segment that was centromeric to HLA-B. This gene lies between TNF and HLA-B. The gene contains 10 exons spanning about 10 kb of genomic DNA and encodes a 428-amino acid protein detected with three different length mRNAs (4.1, 17, and 0.9 kb) in all tissues analyzed, although at different relative levels. Its cellular functions include initiation of translation, RNA splicing, and ribosome assembly. UAP56 (BAT1) is an essential splicing factor that is recruited to the pre-mRNA dependent on U2AF65 and is required for the U2 snRNP-branch point interaction. UAP56 is a member of the DEAD box family of RNA-dependent ATPases, which mediate ATP hydrolysis during several steps of pre-mRNA splicing. Proteins of this family have nine conserved amino acid motifs but differ at their amino and carboxyl ends. From studies of other family members, the first block is involved in ATP binding, the fifth block may be an ATPase, the sixth block is needed for RNA

helicase activity, and the ninth block is involved with ATP hydrolysis-independent RNA interactions during unwinding (Spies *et al.*, 1989; Peelman *et al.*, 1995; Fleckner *et al.*, 1997; Allcock *et al.*, 2001; Price *et al.*, 2004).

2.4.3 TCF19: transcription factor 19 (SC1)

Gene aliases: SC1; SC1-1

The complete genomic structure of a 5.5 kb DNA comprises three exons, generating a 2.5 kb transcript. The TCF19 gene spans a 3.2 stretch of DNA between POU5F1 and S to a 0.2-Mb region between HLA-C, about 130 kb telomeric of HLA-C and about 600 bp from each other. It is mammal-specific and preferentially expressed in the G1-s. This gene encodes a 359 amino acid protein, and is a possible trans-activating factor for the later stage of cell cycle progression (Ku *et al.*, 1991; Krishnan *et al.*, 1995; Teraoka *et al.*, 2000).

2.4.4 POU5F1: POU domain, class 5, transcription factor 1

Gene aliases: OCT3; OTF3; OTF4; Oct4; MGC22487

The genomic span of the POU5F1 gene is 45 kb, and the localization is the 6 kb HLA fragment. This localization was confirmed by linkage of a Random Fragment Lenthg Polymorphic (RFLP) in 9 CEPH families, indicating tight linkage to HLA-A, -B, -C, and -DR. This gene is encoded the protein by the members of the POU (representing a homeodomain protein family of the founder members which are Pit-1, Oct-1/2 and Unc-86) homeodomain protein OCT/Oct-3 (where OCT stands for octamer binding protein) is an embryonic transcription for expressed in oocytes, embryonic stem and embryonic carcinoma cells. The octamer cis-acting transcriptional regulatory motif (ATGCAAAT) is found in enhancers and promoters of many genes which are expressed either ubiquitously or in tissue-specific fashion (Takeda *et al.*, 1992; Sylvester *et al.*, 1994).

2.4.5 TRIM family: The tripartite motif (TRIM) protein family.

The TRIM is composed of three zinc-binding domains, a RING (R), a B box type (B1) and a B-box type 2 (B2), followed by a coiled coil (CC) region. Their genes are implicated in a variety of processes, such as development and cell growth and are involved in several human diseases. However, little is known about the biological and molecular mechanisms mediated by the TRIM genes. (Reymond *et al.*, 2001; Reddy

et al., 1992; Borden, 1998). In this study, TRIM10, TRIM15, TRIM26 and TRIM39 were found.

2.4.5.1 TRIM10

Gene aliases: RNF9; HERF1; RFB30

By screening a human chromosome 6-specific library with a B30.2 domain-encoding exon that had been mapped to 6p21.3 as the probe, a cDNA encoding RFB30 was obtained. Sequence analysis predicted that the 481-amino acid protein contains a RING finger-B box domain encoded by exon 1, a coiled-coil domain encoded by exons 2 through 6, and a B30.2 domain encoded by exon 7. In embryonic mice, expression was detected on day 11.5 at the beginning of erythropoiesis. It was determined that the TRIM10 gene contains at least seven exons (Henry *et al.*, 1997; Harada *et al.*, 1999).

2.4.5.2 TRIM39 Gene aliases: TFP; RNF23; MGC32984

The deduced 519-amino acid protein, 98 % identical to the 489-amino acid mouse sequence, contains a RING finger B-box coiled-coil (RBCC) domain and a C-terminal B30.2 domain. Northern blot analysis revealed nearly ubiquitous expression of a 3.4-kb transcript, with strongest expression in the testis. By genomic sequence analysis, it was determined that the RNF23 gene contains at least 8 exons (Orimo *et al.*, 2000).

2.4.5.3 TRIM26 Gene aliases: AFP; RNF95; ZNF173

Positional cloning of short fragment cDNA sequences from the class I region of the human major histocompatibility complex (MHC) was performed using a hybridization selection approach. Full-length cDNA clones were isolated and also partial genomic clones encoding a protein with two domains rich in cysteine and histidine similar to those characteristic of metal-dependent DNA-binding proteins. The predicted protein also contains a domain thought to form a coiled-coil, and possibly to promote dimerization. A third feature of the predicted protein is a polyglutamic acid region near the carboxyl terminus. Because of these properties, the gene product was named acid-finger protein, AFP (a designation used originally for alpha-fetoprotein AFP). Although the biological role of AFP is unknown, one potential function is the

binding of nucleic acids. The gene is expressed in multiple tissues and conserved among mammals (Chu *et al.*, 1995).

2.4.6 TCTEX5

Gene aliases: HCGV; HCG-V; TCTE5; TCTEX5; MGC125741; MGC125742; MGC125743, PPP1R11.

The gene is located within the major histocompatibility complex class I region on chromosome 6. This gene encodes a specific inhibitor of protein phosphatase-1 (PP1) with a differential sensitivity toward the metal-independent and metal-dependent forms of PP1. Alternative splicing results in two transcript variants encoding different isoforms. Other alternatively spliced transcripts have been described, but their full length sequences have not been determined (Yoshino et al., 1998).

2.4.7 MOG Myelin-oligodendrocyte glycoprotein

Gene aliases: MGC26137

The primary nuclear transcript of the human MOG gene, extending from the putative start of transcription to the site of poly (A) addition, is 15,561 nucleotides in length. The gene contains 8 exons, separated by 7 introns. The introns vary in size from 242 to 6,484 bp and contain numerous repetitive DNA elements, including 14 Alu sequences within 3 introns. The human MOG gene lies 60 kb telomeric to HLA-F in a head-to-head orientation. The product of this gene is a membrane protein expressed on the oligodendrocyte cell surface and the outermost surface of myelin sheaths. Due to this localization, it is a primary target antigen involved in immune-mediated demyelination. This protein may be involved in completion and maintenance of the myelin sheath and in cell-cell communication. Myelin-oligodendrocyte glycoprotein is found on the surface of myelinating oligodendrocytes and external lamellae of myelin sheaths in the central nervous system. The N-terminal, extracellular region of MOG has characteristics of an immunoglobulin variable domain and strong homology with the N-terminus of butyrophilin, a protein expressed in the lactating mammary gland. Alternatively spliced transcript variants encoding different isoforms have been identified (Pham-Dinh *et al.*, 1995; Roth *et al.*, 1995).

3 Materials

3.1 Chemicals

Reagents	Source
- Acetic acid glacial	Roth
- Agar-agar	Roth
- Agarose LE seakem	Biozym
- Albumin, bovine	Sigma
- Bacto-yeast-extract	Gibco BRK GhbH
- Bact- peptone	Gibco BRK GhbH
- Boric acid	Roth
- Bromphenol blue	Servo
- BSA fraction V	Sigma
- Carbon dioxide	Merck
- N,N-dimethyl formamide	Roth
- Ethylene diamino tetraacetic acid (EDTA)	Sigma
- Ethanol	Roth
- Ethidium bromide solution 1 %	Roth
- Formaldehyde 37 %	Roth
- Formamide, deionized	Roth
- Glucose	Merck
- Glycogen 20 mg/ml	Merck
- Hydrocloric acid 37 %	Roth
- Isopropanol	Roth
- Isopropyl-ß-D-pyranoside thiogalacto (IPTG)	Biomol
- Magnesium chloride	Merck
- Magnesium sulfate	Merck
- 2-Propanol	Roth
- Polyvinyl pyrrolidone	Merck
- Potassium acetate	Roth
- Potassium chloride	Merck
- Sodium acetate $\cdot 3\ H_2O$	Roth
- Sodium chloride	Roth
- Sodium hydroxide	Roth
- Sodium iodide	Roth

- Sodium dodecyl sulfate (SDS) ultrapure	Roth
- Tri-sodium citrate $\cdot 2H_2O$	Roth
- Tris base 99.9 %	Roth
- Tryptone/peptone of casein	Roth
- 5-Brom-4-chlor-3-indolyl-ß-D- galactoside (X-gal)	Roth
- Xylene cyanol	Merck
- Yeast extract	Roth

3.2 Buffers and other solutions

3.2.1 Ammonium acetate 10 M

385.4 g ammonium acetate dissolved in 150 ml H_2O, H_2O added to 500 ml

3.2.2 Denaturation solution

1.5 M NaCl

0.5 M NaOH

3.2.3 Denhardt´s solution 100x

10 g Ficoll 400

10 g polyvinyl pyrrolidone

10 g bovine serum albumin (BSA fraction V), H_2O to 500 ml

Filters sterilized and stored at -20°C as 25 ml aliquots in 1 % BSA fraction V

3.2.4 DNA-lysis buffer

50 mM Tris-HCl pH 8.0

100 mM EDTA pH 8.0

0.5 % SDS

3.2.5 EDTA 0.5M pH 8.0

186g Na_2 EDTA $\cdot 2H_2O$ dissolved in 700ml H_2O

pH adjusted to 8.0 with 10 M NaOH

H_2O added to 1 liter

3.2.6 HCl 1 M

91.38 ml H_2O

86.2 ml concentrated HCl

3.2.7 KCl 1 M

74.6 g KCl

H_2O to 100 ml

3.2.8 Loading for buffer

0.25 % Bromophenol blue

0.25 % Xylene cyanol

15 % Ficoll 400

50 mM EDTA pH 8.0

3.2.9 Lysis solution I P1

50 mM Glucose

25 mM Tris-HCl pH 8.0

10 mM EDTA pH 8.0

3.2.10 Lysis solution II P2

0.2 M NaOH

1 % SDS

3.2.11 Lysis solution III P3

3 M Potassium acetate

Adjusted with concentrated HCl to pH 5.5

3.2.12 $MgCl_2$ 1M

20.3 g $MgCl_2 \bullet 6\ H_2O$

H_2O to 100 ml

3.2.13 $MgSO_4$ 1M

24.6 g $MgSO_4$

H_2O to 100 ml

3.2.14 NaCl 5M

292 g NaCl

H_2O to 1 liter

3.2.15 NaOH

400 g NaOH dissolved in 450 ml H_2O

H_2O added to 1 liter

3.2.16 Pre-hybridization solution

20X SSC 75 ml

50X Denhardt´s 25 ml

10 % SDS 25 ml

H_2O added to 250 ml

3.2.17 Sodium acetate 3 M

408 g sodium acetate dissolved 3 ·H_2O in 800 ml H_2O, H_2O added to 1 liter, pH adjusted to 4.8 or 5.2 with 3 M acetic acid

3.2.18 SSC (sodium chloride/sodium citrate), 20 x

3 M NaCl (175g/liter)

0.3 M Sodium citrate 2 H_2O (88 g/liter)

3.2.19 Sodium Iodine

121 g NaI

H_2O added to 250 ml

3.2.20 20x SSC (Sodium chloride sodium citrate)

3 M NaCl

0.3 M sodium acetate

Adjust with HCl concentrated to pH 7.0

3.2.21 TBE (tris/borate/EDTA) electrophoresis buffer

10X stock solution, 1 liter:

108 g tris base (890 mM)

55 g boric acid (890 mM)

40 ml 0.5 M EDTA, pH 8.0

3.2.22 TrisHCl [Tris (hydroxymethyl) aminomethano], 1 M

121 g tris base dissolved in 800 ml H_2O

Adjust to desired pH with concentrated HCl

Mix and add H_2O to 1 liter

Approximately 70 ml of HC is needed to achieve pH 7.4 solution and approximately 42 ml for 8.0.

IMPORTANT NOTE: The pH of tris buffer changes significantly with temperature, decreasing approximately 0.0028 pH units per 1°C. Tris-buffered solutions should be adjusted to the desired pH at the temperature used. Because the pK_a of tris is 8.08, it should not be used as a buffer below pH ~ 7.2 or above pH ~ 9.0.

3.2.23 10x Tris-phosphate-buffer

108 g tris base

15.5 ml phosphoric acid (85 %, 1.679 g/ml)

40 ml 0.5M EDTA pH 8.0

3.3 Media

3.3.1 LB Medium (Luria-Bertani):

1 % Pepetone 140

0.5 % Yeast extracts

1 % NaCl

1.5 % Agar to solid medium

3.3.2 SOB Medium

2 % Select peptone 140

0.5 % Yeast extract

10 mM NaCl

2.5 mM KCl

10 mM $MgCl_2$

10 mM $MgSO_4$

3.3.2 SOC Medium

SOB Media

20 mM Glucose

3.3.3 Antibiotic solutions

3.3.2.1 Ampicillin

50 mg Ampicillin to 1 liter medium

Before dilution, the ampicillin solution must be sterilized by filtration

3.3.2.2 Chloramphenicol

20 mg to 1 liter medium

Before dilution, the chloramphenicol solution must be sterilized by filtration

3.4 Commercials kits

• Big Dye Terminator version 1.1, Applied Biosystems	4337450
• Cycle Sequencing Kit, Applied Biosystems	4337450
• Megaprime DNA Labelling, Promega	RPN1605
• Prime-a-Gene Labeling System, Promega	U1100
• QIAamp DNA Mini Kit, Qiagen	51306
• QIAquick Gel Extraction Kit, Qiagen	28706
• QIAquick PCR Purification Kit, Qiagen	28104
• System, dCTP, Amersham	RPN1607
• Taq PCR Care Kit, Qiagen	201223

3.5 Enzymes

• BamHI	10.000 U	New England Biolabs (NEB)
• Biotherm Taq-Polymerase	1000 U	Genecraft
• *EcoRI*	10.000 U	NEB
• HindIII	10.000 U	NEB
• Not I	10.000 U	NEB
• PstI	10.000 U	NEB

- Ribonuclease A Roth
- RNAase A Roth
- Sequenace Perkin Elmer
- T4-DNA Ligase 20.000U NEB

3.6 Oligonucleotides primers for PCR: sequences and annealing temperatures applied.

Primer	Primer sequence	Annealing temperature °C
T7	5'aatacgactcactataggg 3'	52 -(touchdown)
pTARBAC2	5'cttacgcagggcatccatt 3'	52 -(touchdown)
pTARBAC3	5'acatttaggtgacactatag 3'	52 -(touchdown)
pTARBAC4	5'atacaaagaaacgtacggc 3'	52 -(touchdown)
T7.29	5'gccgctaatacgactcactatagggagag 3'	60
SP6.26	5'ccgtcgacatttaggtgacactatag 3'	60
K18T3[1]	5'agtacttataggaattattacca 3'	52
K18T3[1]	5'gttgctttattgtgtcactagtc 3'	52
POU5F1[2] F	5'atggcgggacacctggcttcgg 3'	58
POU5F1[2] R	5'ctcctccgggttttgctccagct 3'	58
MOG exon 2 F	5'caggacagttcagagtgataggacc 3'	60
MOG exon 2 R	5'attgctgcctcctcttggtaaga 3'	60
Caja-G and E F	5'gctcccactccatgaggtat 3'	55
Caja-G	5'tgagaggaggagagccta 3'	55
Caja-E R	5'cacgtgtcctccaggta 3'	55
Klex exon 2-4on 2 consenso F	5' gctcccactccatgaggtatt 3'	53
Klex exon 2-4on 2 consensoR	5' cgcccacttctggaaggttc 3'	53
HLA B 326H24-1 A	5'cttgtgtgccctccctcccc 3'	52
HLA B 326H24-1 B	5'tgtttcctccccagtcatctttcct 3'	52
HLA B 297I7-4 A	5'aagggacaagaacaatggaacagtgaa 3'	52
HLA B 297I7-4 B	5'gctgtgtgtaagtggtgggggtg 3'	52
HLA B 391L07-1 A	5'taaagaaaggcaccaacaggattactatga 3'	52
HLA B 391L07-1 B	5'catgcttcctcctccacagttctacttt 3'	52
G2482F ALU+470	5'ctgaaactggttctggttcatgtgac 3'	53
G2483R ALU-166	5' gttagggatcgtttcctcag 3'	53
M13 rev	5'ggaaacagctatgaccatga 3'	50
M13 for	5'ttgtaaaacgacggccagtg 3'	50
RFB30-5'	5' gctcagttctcctcaaaatgg 3'	50
RFB30-3'	5' gcccaaagaagggaatgacc 3'	50
TCF19-5'	5' gactttgctgccattaccatc 3'	50
TCF19-3'	5' gctcactctcatcatccagt 3'	50
TC4-5'	5' gtccagttcaaacttgtattgt 3'	50
TC4-3'	5' cacaggtcatcatcctcatc 3'	50

3.6.1 Hybridization probes for screening of CHORI BAC-bank and Southern blot and hybridization

Name of probe	Specification	Length of fragment
Caja G- class I	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	750
K18T3	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	150
ATPV6g2	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	
BAT1	Mini-prep, BamHI and Hind III digestion, loading for electrophoresis, DNA extraction from gel	1300
TCF19	Mini-prep, Digestion BamHI and Hind III loading for electrophoresis, DNA extraction from gel	
HRC	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	
CAT 56	Mini-prep, *EcoRI* digestion , loading for electrophoresis, DNA extraction from gel	1000
TRIM 39	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	1000
TRIM 26	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	500
PPP1R11	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	550
TCTEX5	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	400
TCTEX4	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	500
MOG	Mini-prep, *EcoRI* digestion, loading for electrophoresis, DNA extraction from gel	800
HLA-B	PCR fragment	500
ALU	PCR fragment	450

3.7 DNA ladders and other standards for comparison of molecular size

Standard sample/ enzyme	**Supplier**	**Ordering No.**
• Alkaline phosphatase (CIAP) 1000U	Promega	M1821
• DNA from fish sperm	Roche	10223646001
• DNA Ladder 100 bp Plus	Fermentas (MBI)	SM0322
• Gene Ruler DNA Ladder Plus100 bp	Fermentas	SM0321
• Lambda DNA 500 bp	Fermentas	SD0011
• Lamda DNA/ *HindIII* + *EcoRI*	Fermentas	SM0191
• pUC 19 DNA 0,5mg/ml	Fermentas	SD0061
• dNTP-Set 1	Roth	K039.1

3.8 Vectors

pDrive Quiagen

pTARBAC2.1 – BACPAC Resources Center

3.9 BAC-clones used for genomic analysis

To analyze MHC class I regions, clones from the BAC-bank library CHORI259 of the common marmoset were used. 42 selected clones were bought from BAC-PAC Resources Center at the Children's Hospital Oakland747 - 52th Street, Oakland CA 94609 USA, which is directed by Dr. Pieter deJong.

Table 3.9.1: Lab clone assignment to each GenBank BAC clone CHORI-259 of common marmoset.

The clones were named in lab clone 1 until clone 42, to practical reason.Each clone has the corresponds coordinate of GenBank BAC clone CHORI-259 of Common marmorset

Clone	Coordinates
Clone 1	120D3
Clone 2	271C5
Clone 3	18G14
Clone 4	282L12
Clone 5	204C3
Clone 6	127E3
Clone 7	239N18
Clone 8	193P12
Clone 9	261L9
Clone 10	282015
Clone 11	245C6
Clone 12	234L16
Clone 13	463N5
Clone 14	436B4
Clone 15	19I16
Clone 16	44G8
Clone 17	510K19
Clone 18	217M17
Clone 19	161C12
Clone 20	279K21
Clone 21	325P21
Clone 22	277A2
Clone 23	348C21
Clone 24	171K8

Clone 25	174N1a
Clone 26	224H20
Clone 27	347K7
Clone 28	347D1
Clone 29	379F14
Clone 30	425A13
Clone 31	459O20
Clone 33	169G2
Clone 34	123H17
Clone 37	99L19
Clone 38	334C4
Clone 39	329M5
Clone 40	379M5
Clone 41	174N1b
Clone 42	485C4

Table 3.9.1 CHORI-259 BAC clone GenBank.

CHORI-259 Segment 1 High Density Filter set (4x4). Filters 01I – 06 I, plate range between 0001 - 0288.

Filter-Set	Manufacture dates	Library
007874	01/27/03 – 01/30/03	CHORI-259 seg 1

Filter No.	Barcode	BAC-bank	Replica	Plate range
01I	40819	CHORI- 259	R2	0001 0048
02I	40689	CHORI- 259	R2	0049 0096
03I	40713	CHORI- 259	R2	0097 0144
04I	40713	CHORI- 259	R2	0145 0192
05I	40761	CHORI- 259	R2	0193 0240
06I	40785	CHORI- 259	R2	0241 0288

Table 3.9.2 CHORI-259 BAC clone GenBank.

CHORI -259 Segment 2 High Density Filter set (4x4) Filters 06G- 011G, plate range between 0289 -.0528

Filter-Set	Manufacture date	Library
007897	01/30/03-02/04/03	CHORI- 259 seg 2

Filter No.	Barcode	BAC-bank	Replica	Plate range
07G	40841	CHORI- 259	R2	0289 0336
08G	40865	CHORI- 259	R2	0337 0384
09G	40889	CHORI- 259	R2	0385 0432
010G	40913	CHORI- 259	R2	0433 0480
011G	40937	CHORI- 259	R2	0481 0528

3.10 The gene sequence of *Callitrix jacchus* used as MHC class I probe

The sequence was obtained and determined by Ulrike Geisler at the DPZ. The isolation of common marmoset MHC class I cDNA sequences from liver was carried out by reverse transcriptase polymerase chain reaction (RT-PCR). The primers were designed from primate consensus MHC class I sequences (Shufflebotham and Watkins, 1997; Cadavid *et al*, 1997; Cadavid *et al*, 1999). Sequence analysis of several clones indicated the presence of two different MHC class I gene, transcripts, eventually derived from a Caja-G and a Caja-E class I gene, respectively. She found one group of clones to be identical with the Caja-G*04 allele (GenBank accession number U59640), whereas the other group displayed 99 % sequence identity with the Caja-E*02 allele (AF004920).

3.11 Computer programs

The DNA sequence was analyzed in internet with Program BLAST (Altschul *et al*., 1997, www.ncbi.nlm.nih.gov/blast. The program BLAST 2 SEQUENCE (www.ncbi.nlm.nih.gov/gorf/bl2.html) compared the similitude between two sequences. The DNA repeats in one sequence were searched by the program Repeat Masker (www.repeatmasker.org/cgi-bin/WEBRepeatMasker). For alignment analysis was used the programs BioEdit version 7.0.5 (Hall, 1999), Swissprot, AUGUSTUS and Genruner version 3.05. The construction Phylogenetic tree was made by the program Clustal X (Version 1.83) and MEGA 3.1.

3.12 Plastic articles

Article	Company
• Centrifuge tubes 13 ml and 50 ml	Falcon
• Combitips plus 5ml, 10ml	Eppendorf
• Cryo-tube vials 1.8 ml	Nunc
• Electroporation cuvettes 2mm	Peqlab
• Non-skirted	Abgene
• Nylon membrane 200 x 200	Amersham
• Paper Whatman 3MM	Schleicher and Schuell
• Petri box ∅ 9 cm	Sarstedt
• PCR containers 0.2 ml	Eppendorf
• PCR plates 96 well, thermo fast	Sarstedt

- 96 V- soil microplates Biozym
- PP-tubes sterile, 15 ml Greiner
- PP-tubes, 170/77 10 ml Greiner
- PP-test tubes, 50 ml Greiner
- Pump-spray bottle Roth
- PE-spray bottle Neolab
- Safeseal tips, 2.5 µl Biozym
- Safeseal tips, 20 µl Biozym
- Safeseal tips, 100 µl Biozym
- Safeseal tips, 1000 µl Biozym
- Tips 2.5 ml Biozym
- Tips 20 ml, glass Biozym

3.13 Radioactive substrate

α^{32}P-dCTP250µCi Amersham

3.14 Lab equipment

- Autoclave type A40/45 Webco
- Autoradiographic cassettes, Dupont de Nemours
- Balance BP 3105 Sartorius
- Developing machine,
  Kodak M35 X-OMAT processor, Kodak
- Centrifuge:
  - Mikro 22 Hettich
  - Mikroliter 2025 Hettich
  - Sigma 3k30, Hettich
  - Sepatech Varifuge 3.2 RS Heraeus
  - Labofuge GL Heraeus
  - Minifuge GL Heraeus
- CO_2 Oven, B 5060, Heraeus
- Electrophoresis camera (DNA) B1A, B2
- Electrophoresis power supply Standard Power Pack P25 Biometra
- Electroporation- pulse generator EPI2500

- Freezers -20°C and -80°C — Liebherr, Premium
- Fume hood: Kojair — BioFlow Technik,
- Geiger counter — Berthold LB 122
- Gel documentation — Jet Imager 2000, Intas
- Heating blocks thermomixer — Eppendorf
- Hybridization oven Großer: — Saur Laborbedarf
- Hybridization oven 6/12 — UNIEQUIP, UNITHERM
- Incubator — Heraeus
- Incubator/shaker 3033 — GFL
- Laminar flow cabinets, — Heraeus
- Light box Cronex-Kassetten, — DuPont de Nemours
- Magnetic stirrers MR 3001 — K. Heidolph
- Microwave oven — AEG Micromat
- Parafilm — American National Can.
- PCR Machine — PCR System 2700 BioSystems
- Pipettes — Eppendorf
- Power supplies — Standard Power Pack P25 (Biometra)
- pH meter — Micropressor pH Meter, WTW, Schütt
- Photometer GeneQuant pro, — Amersham Bioscience
- Refrigerators — Privileg
- Sequence automat, ABI 3100- — Avanta Genetic Analyzer Applied Biosystems
- Speedvac evaporator — Helmut Saur
- UV Stratalinker 2400 — Stratagene
- Vortex mixer — L46 (GLW)
- Vortex — Genie 2, Scientific Industries
- Water baths — Schütt GFL 1083
- Water purifiction equipment — Biocel MilliQ , Millipore Purification

3.15 Addresses of manufacturers

- Amersham Bioscience, Munziger Str 9, 79111 Freiburg
- Applied Biosystems Applera, Frankfurter Str 12 b, 64293 Darmstadt
- Bachofer GmbH, Postfach 7089, 72770 Reutlingen
- Bauknecht Haushaltsgeräte GmbH, Gottlob-Bauknecht Str.1-11, 73614 Schondorf
- Berthold Technologies & Co KG, Calmbacher Str. 22, 75323 Bad Wildbach
- Biochrom AG, Postfach 460309, 12213 Berlin
- Biomol GmbH, Waidmann Str. 35, 22769 Hamburg
- Biozym Scientific GmbH, Postfach 63, 31833 Hessisch Oldendorf
- Calbiochem-Novobiochem GmbH, Postfach 116, Bad Soden
- Carl-Roth GmbH & CoKG, Schoemperlen Str. 1 - 5, 76185 Karlsruhe
- Dupont de Nemours GmbH, Sägewerk Str. 3, 83395 Freilassing
- Eppendorf AG, Friedens Str. 116, 51145 Köln
- Fischer, Dr.K, Schneidemühl Str. 9, 69115 Heidelberg
- Fluka Chemie AG, Industrie Str. 25 CH-9471 Buchs
- Fröbel Laborgeräte, Alwinder Str. 4, 88131 Lindau
- GeneCraft GmbH, Reiffeisten Str. 12, 59348 Lüdinghausen
- Genomed GmbH, Wieland Str. 28a, 32545 Bad Oeynhausen
- Gilson International Deutschland, Otto-Hahn Str. 17, 65520 Camberg
- Hereus, Kendro Laboratory Products GmbH, 63450 Hanau
- Helmut Saur Laborbedarf, Carl-Zeiss Str. 58, 72770 Reutlingen
- Hettich GmbH & Co KG, Industrie Str. 2 10, 31311 Uetze/ Hänigsen
- Hewlett-Packard GmbH, Postfach 1430, 71004 Böblingen
- Intas Science Imaging Instruments GmbH, Florenz-Sartorius Str. 14, 37079 Göttingen
- Invitrogen GmbH, Technologiepark Karsruhe, Emmy-Noether Str. 10, 761331 Karlsruhe
- Janke & Kunkel GmbH & Co KG, Janke & Kunkel Str. 10, 79219 Stauffen
- Köttermann GmbH & Co KG, Industrie Str. 2-10, 31311 Uetze/Hänigsen
- Küjner, Adolf, Dinkelberg Str. 1, CH-4172 Birsfelden
- Laboratorium Prof. Dr. Berthold, Siemens Str. 4, 30173 Hannover
- MBI Fermentas GmbH, Opel Str. 9, 68789 St. Leon-Rot

- Membrane Pure GmbH, Am Kuemmerling Str 37, 55294 Bodenheim
- Merck KGaA, Frankfurter Str. 250, 64293 Darmstadt
- MWG Biotech AG., Azinger Str. 7a, 85560 Ebersberg
- NEB, New England Biolabs GmbH, Brüning Str. 50, 5926 Frankfurt am Main
- Peqlab Biotechnologie GmbH, Carl-Thiersch Str. 2b, 91052 Erlangen
- Polaroid GmbH, Sprendlinger Lands Str. 109, 63069 Offenbach
- Promega GmbH, Schildkröt Str. 15, 68199 Mannheim
- Quiagen GmbH Max-Volmer Str. 4, 407024 Hilden
- Renner GmbH, Ried Str. 6, 67125 Darmstadt
- Roche Diagnostic GmbH, Sandhofer Str. 116, 68305 Mannheim
- Thermo Haake, Diesel Str.4, 76227 Karlsruhe
- W. Krannich GmbH & Co KG, Elliehäuserweg Str17, 37079 Göttingen
- Sarstedt AG & Co, Postfach 1220, 51582 Nümbrecht
- Sartorius AG, Weender Lands Str. 94-108, 37075 Göttingen
- Schleicher & Schuell BioScience GmbH, Postfach 1111, 38582 Dassel
- Schott-Geräte-GmbH, Hattenberg Str.10, 55122 Mainz
- Schütt Labortechnik GmbH, Rudolf-Wissel Str. 11, 37079 Göttingen
- Serva Electrophoresis GmbH, Carl-Benz Str. 7, 69115 Heidelberg
- Sigma-Aldrich Chemie GmbH, Eschen Str. 5, 82024 Taufkirchen
- Sorvall, Kendro Laboratory Products GmbH, Robert-Bosch Str.1, 63505 Langenselbold
- Vesper GmbH, Industrie Str. 16, 49079 Göttingen
- Webeco GmbH & Co KG, Mühlen Str. 38, 23611 Bad Schwartau
- Whatman Biometra GmbH, Rudolf-Wissel Str. 30, 37079 Göttingen

4 Methods

4.1 Preparation of materials

The preparation of buffers, reagents and plastic materials used in the manipulation of nucleic acids was as follows: For all stock solutions, water was used which had been both deionized and bidistilled. All reagents were of the highest grade available.

The sterilization of materials was performed either by

- filtration through a 0.22 µm filter or by
- autoclaving of solutions for 15 min (min) at 121 °C = 15 bars.

There are toxic, carcinogenic, mutagenic and teratogenic reagents. For this reason, the user must proceed with precaution; the protective equipment includes eye protection, laboratory coat and gloves.

For the use of radioactive isotopes it is very important to know the applicable regulations and approved procedures. The levels of radioactivity e.g. from ^{32}P in the working area should be checked with a hand-held Geiger mini-monitor, both near the containers and in the working area.

The strength of radioactivity in a source is defined by an international SI unit, the Becquerel (Bq).

1 Becquerel (Bq) = 1 disintegration per second (e.g within a radioactive specimen).

Definitions in Curie (Ci) may be converted into Bq or disintegrations per min (dpm) by the equations

1 Ci = 3.7×10^{10} Bq;

1 Ci = 2.2×10^{12} (dpm);

1 microcurie (µCi) = 3.7×10^{4}Bq = 2.22×10^{6} dpm.

The relative centrifugal force (RFC) is measured in x g, it is used to specify forces applied in centrifugation. It is proportional to the speed in rpm (revolutions per min) of a centrifuge and depends on the rotor model (with the maximum spinning radius of tubes therein).

The relationship between RCF and speed (rpm) is determined by the following equation:

$RCF = 11.12\, r\, (rpm/1000)^2$

where r is the radius between the rotating particle and the axis of rotation.

4.2 Filters of carrying bacterial artificial chromosome (BAC) clones for screening

4.2.1 Source and origin of BAC bank CHORI of common marmoset (*Callithrix jacchus*)

The BACPAC Resource Center (BPRC) at Children's Hospital Oakland Research Institute (CHORI) in Oakland, California (United States) has prepared BAC and hybridization (PAC) genomic DNA libraries and provides copies thereof with high-density colony hybridization filters.

In Pieter De Jong's Laboratory at the Children's Hospital Oakland Research Institute, BACPAC Resources Center, a BAC library CHORI-259 of common marmoset (*Callithrix jacchus*, male) was constructed by Doriana Misceo and Dr. Baoli Zhu using the cloning techniques developed in the same laboratory (Osoegawa *et al.*, 1998). Dr. Suzette Tardif, Associate Director of the Southwest National Primate Research Center in San Antonio, TX provided the marmoset sample to make the library. Cells were taken from the kidney of a male marmoset (SNPRC #17081). The DNA was isolated and then partially digested with a combination of *EcoRI* restriction enzyme and *EcoRI* methylase. The DNA fragments were ligated into the pTARBAC2.1 vector between the *EcoRI* sites. The ligation products were transformed into DH10B (T1 resistant) electrocompetent cells (Invitrogen, 1999).

4.2.2 Organization of BAC bank CHORI 259 from of the new world monkey common marmoset (*Callithrix jacchus*) on microtiter plates and on nylon filters

The whole library was arrayed into 528 384-well microtiter plates and gridded also onto eleven nylon filters high-density (22 x 22 cm) as tools for screening by probe hybridization. Each hybridization membrane represents over 18.000 distinct BAC clones, stamped in duplicate using a code of arrangements. It indicates as amend event supplementing the two coordinate numbers up to the full bank number of the clone.

4.2.3 Procedure for hybridization on high density filters

The filters were soaked in pre-hybridization solution and then placed in a bottle; 20 ml of the pre-hybridization solution were added to pre-hybridize for 30 min at 60 °C. Thereafter, a labeled probe was added to hybridize overnight (16-18 hours) at 60 °C.

The filters were washed with wash I solution at 60 °C for 30 min and this step was repeated twice.

The filters were wrapped individually in plastic wrap and placed in a large cassette (35 x 43 cm^2), two in each cassette. A light overlapping of the filters occurred, but this did not give rise to carry problems in the reading of data, see figure 4.1. For the choice of sensitivity, the exposition for autoradiography was done for 2 to 36 hours at -80° C. The position of the positive clones was viewed by the use of a transparent net (overlay template, see Figure 4.2) which allows definition of the coordinates (as an orientation guide for the choice) of t he interesting clones.

The filters were stripped off the differently labeled DNA probes by boiling them in 0.1 % SDS for 15 min. Stripped filters were preserved in plastic wrap at -20 °C.

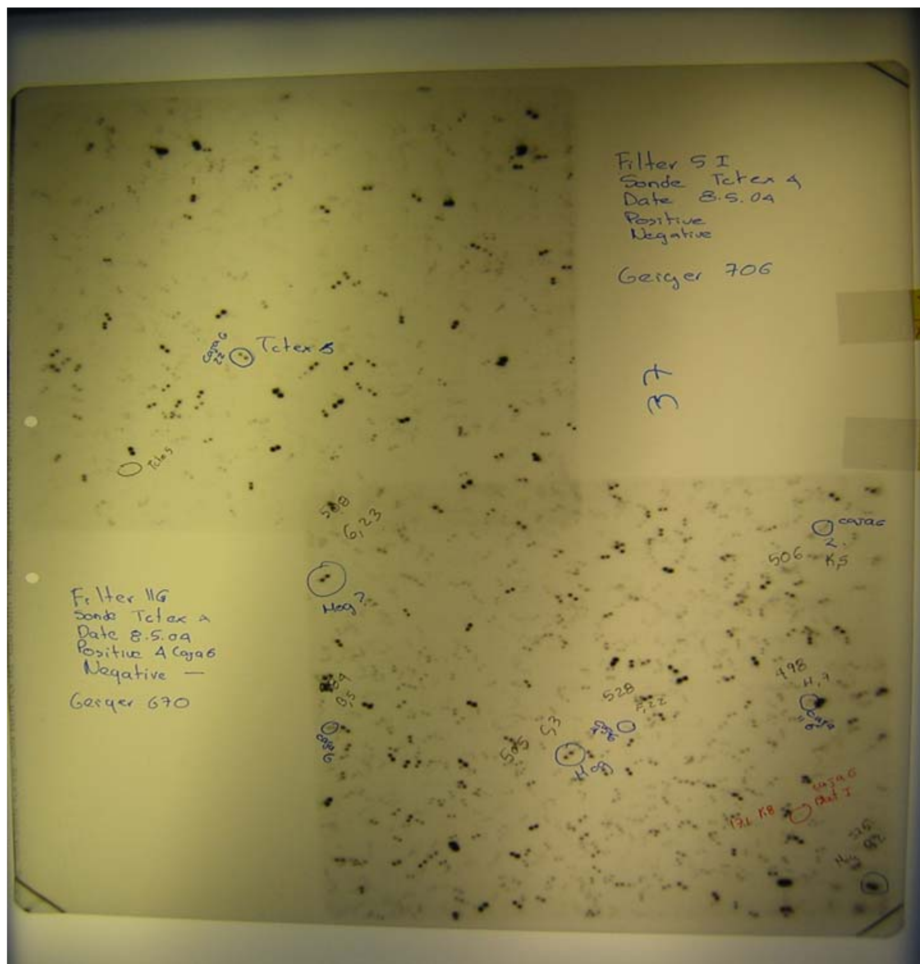


Figure 4.1: Example of placing two BAC bank filters in one cassette for autoradiography. The BAC clone filter 10G and 11G were hybridized with TCTEX4 probe the centre, a small overlap of screened filters is seen, this presented no problem in the reading of the data, and the two points indicated the double positive control for each clone.

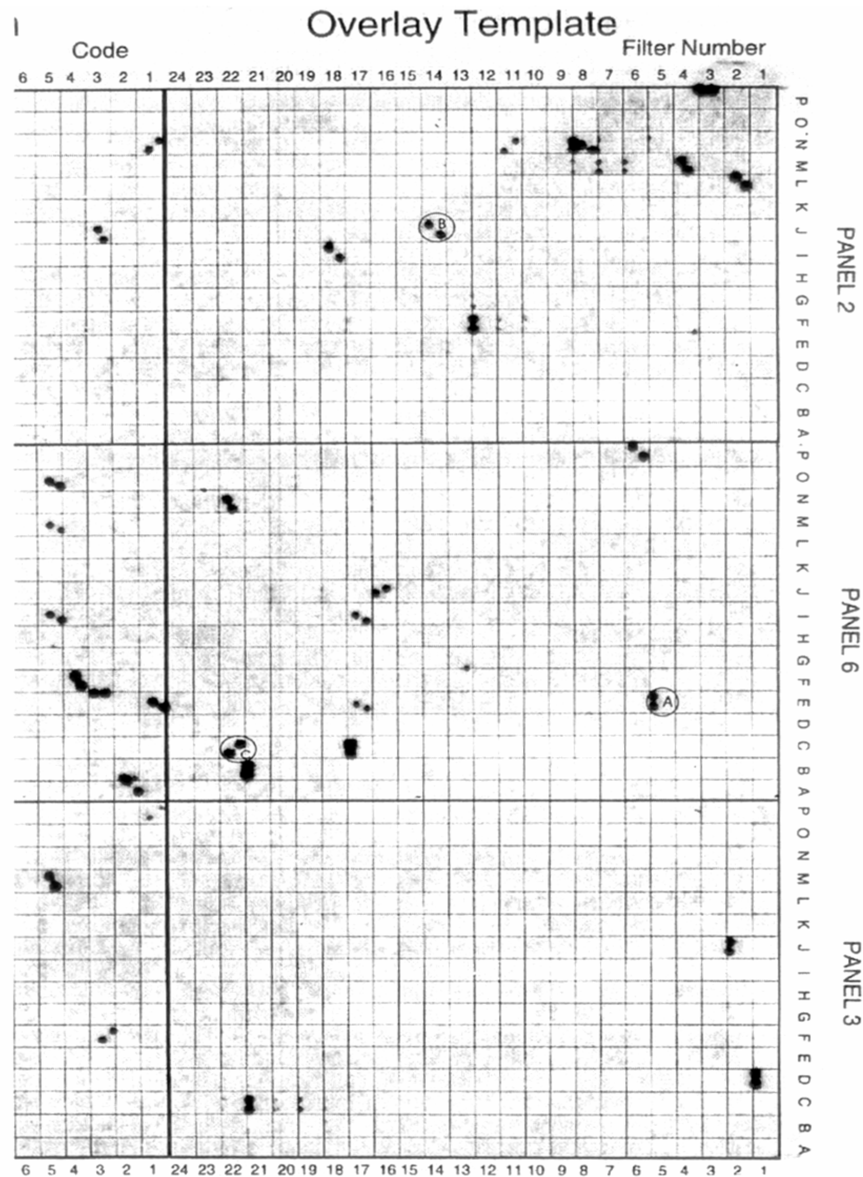


Figure 4.2: Example of a positive clone localization in originals 384 well plates. Here, the overlay template is placed over the autoradiography identifying clones positive in screening with the hybridization probe (http://bacpac.chori.org).

4.3 Preparation of a BAC clone cultures

Each clone was received as a bacterial LB agar stab culture. The DH10 *E.coli* host had been placed into LB agar containing 12.5 µg/ml chloramphenicol for BAC clones. The DH10 *E.coli* host harbors a number of identical plasmids consisting of vector insert DNA BAC clone (Figure 4.3). This culture has a finite life time at 4 °C.

To inoculate a single isolated bacterial colony from the stab end fare received into 2 ml LB medium supplemented with 20 µg/ml chloramphenicol in a sterile toothpick was used. The culture was grown for 3 hours by shaking at 5-6 g at 37 °C. This 2 ml culture

was transferred into a 50 ml snap-cap polypropylene tube with 10 ml LB medium supplemented with chloramphenicol and grown overnight (up to 16 h) by shaking at 5 - 6 g at 37° C.

1 ml of glycerol stocks of cells was prepared as follows: 1 ml of a freshly grown overnight culture was supplemented with 40 % of sterile glycerol, mixed well, transferred to two freezer vials and immediately placed into a dry ice/ethanol bath or into a box in the -80° C freezer.

4.4 Isolating plasmid DNA (*Miniprep)*

This is a rapid alkaline lysis miniprep method. The culture was centrifuged at 400 g for 10 min. The supernatant was discarded, the pellet resuspended in 0.3 ml P1 solution plus 0.3 ml of P2 solution by shaking the tube gently to mix contents. For hydrolysis, it was left at room temperature for 5 min. Thereafter, 0.3 ml of P3 solution was added slowly, the tubes shaken being gently during addition. After having been placed in ice for at least 5 min, they were centrifuged at 5.000 g for 15 min at 4 °C. Afterwards, the supernatant was collected using a pipette; this extraction step was repeated three times, then, 2 ml of ice-cold isopropanol was added to the combined supernatants any white precipitating material being avoided. The tubes were mixed by inverting a few times, placed on ice for at least 5 min and centrifuged with maximum velocity in a cold microcentrifuge for 15 min. The supernatant was discarded and 0.5 ml of 70 % Etanol was added to wash the DNA pellets by inverting them several times. Thereafter, they were collected by centrifugation in a cold microcentrifuge at 10.000 g for 5 min. The supernatant was removed as much as possible. Occasionally, pellets become dislodged from the tube therefore, it is better to carefully aspirate off the supernatant rather than to pour it off. The DNA pellets were dried at room temperature and resuspended in 40 µl of H_2O with shaking.

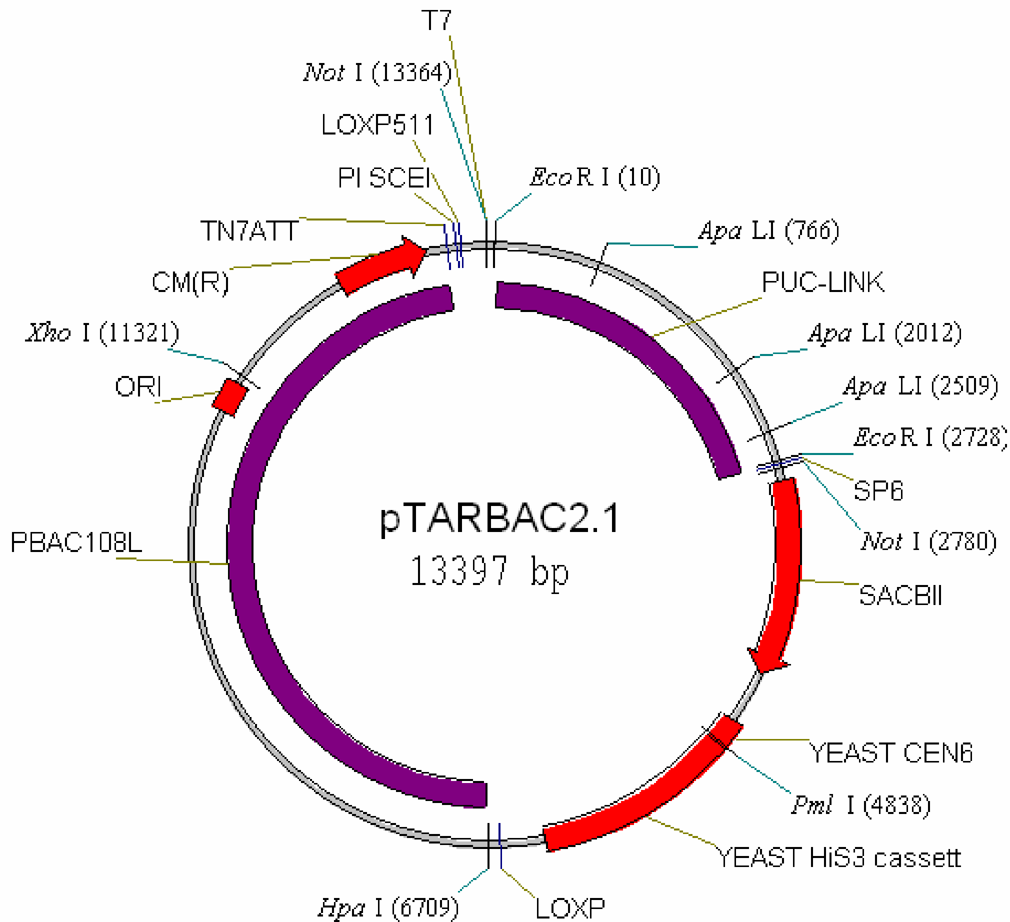


Figure 4.3: Map of the plasmid pTARBAC 2.1 (vector) (http://bacpac.chori.org/ptarbac21.htm).

This is the cloning vector of CHORI 259, the BAC bank of the common marmoset. It contains, among many others, a gene encoding resistance to chloramphenicol and shows various restriction enzymes, for example *EcoRI*, where to cut a DNA fragment.

4.5 Quantification of DNA

4.5.1 UV-absorption spectroscopy method

The concentration of a DNA solution is estimated by measuring its absorption at several wavelengths (260 nm, 280 nm and 320 nm) which also provides an indicator of the purity of the preparation. The absorption spectrum of DNA shows a maximum at 260 nm. In the case of double-stranded DNA one A_{260} or (OD_{260}) unit (measured by a light path of 1 cm) corresponds to a concentration of 50 µg/ml.

Significant absorption at 230 nm indicates contamination by phenolate ions and other organic compounds (Stulnig and Amberger, 1994), elevated absorption at 280 nm

indicates the presence of protein, whereas absorption at 320 nm indicates the presence of particulate matter. In general, highly pure DNA has an A_{260}/A_{280} ratio > 1.8.
The DNA was dissolved in water, and water was used as a blank or reference sample. The DNA concentrations were calculated using the following equation:

Sample DNA concentration

C= A_{260} • 50 µg DNA/ml • dilution vol/sample vol • cm path $^{-1}$ (µg DNA/ml); light path (cm)

A_{260} = A_{260} corrected for turbidity

50 µg DNA/ml = DNA conc. specific absorption factor

dilution vol/sample vol • = dilution factor

in short C= A_{260} corrected • 50 µg DNA/ml • V• f^{-1} (µg DNA/ml)

C = concentration (µg/ml),

A_{260} = UV absorption at wave length 260 nm

V = dilution factor = dilution volume/sample volume;

f = cuvette length (light path) in cm (can be neglected, 1 cm- cuvettes were used.

4.5.2 Minigel method

This is a rapid and convenient method to measure the quantity of DNA and to analyze its intactness. Into a slot of a 0.8 % agarose minigel containing ethidium bromide (0.5 µg/ml) 2 µl of the DNA solution, mixed with 1 µl gel loading buffer 6 x and 3 µl water was applied. As DNA standards, two samples of plasmid DNA pUC19 vector (25 µg/µl and 50 µg/µl) were run in parallel. Electrophoresis was carried out for 20 min at 100 volts in TBE 1X buffer. The gel shown in Figure 4.4 was photographed using a length of wave length UV irradiation of 260 nm.

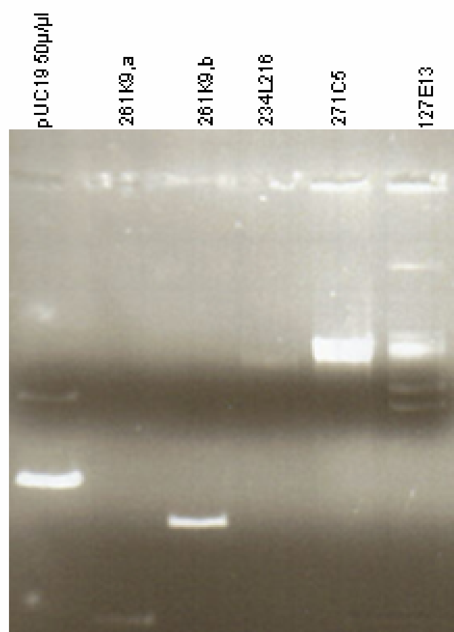


Figure 4.4: Quantitative staining of DNA an agarose gel containing ethidium bromide.

Image of a 1 % agarose gel stained with 20 mg/ml ethidium bromide. DNA fragment extraction from single bands performed with NaI as described in the text. pUC19 50 µ/µl was the probe control to quantification of DNA by comparation.

4.6 Restriction digestion of plasmid DNA

The treatment of plasmid DNA with a restriction endonuclease produces a series of precisely defined fragments which are separated according to size by gel electrophoresis.

1 unit (U) of restriction enzyme is defined as the amount of enzyme required to digest 1 µg of lambda DNA in 1 hour at 37 °C in 50 µl of assay buffer.

The digestion protocol of DNA requires in a 20 µl volume of final reaction mixture 2 µg DNA for every 2 U of enzyme, the adequate amount of buffer, and for some enzymes 10 mg/ml BSA. The mixture is incubated at 37°C the incubation time depends on the enzyme, ranging between 1.5 – 3 hours. The enzyme is inactivated at 65 °C, figure 4.5.

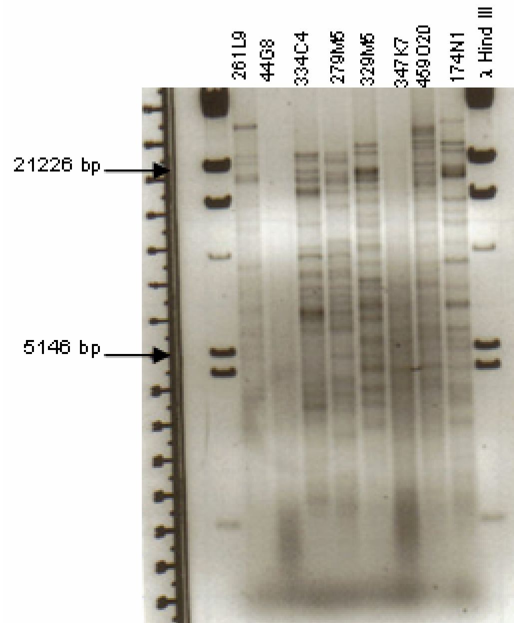


Figure 4.5: Fragments obtained by restriction digestion separated by agarose gel electrophoresis.

The treatment of DNA plasmid clones 9 (261L9), 16 (44G8), 38 (334C4), 40 (379M5), 39 (329M5), 27 (347K7), 31 (459O2) and 41 (174N1) with restriction endonuclease *EcoRI* and lambda vector digested Hind III produced a series of precisely defined fragments. These were separated according to size by electrophoresis in 0.8 % agarose gel containing 20 mg/ml ethidium bromide. The figure shows DNA by its UV absorption. Note: the picture is a negative image.

4.7 Separation of large DNA fragments by agarose gel electrophoresis

Agarose gel electrophoresis is an effective standard method for separating, identifying and purifying DNA fragments in the range of 0.5 to 25 kb. The gel is prepared with an appropriate agarose concentration. Voltage and running time are chosen for optimal separation. The DNA and its fragments, stained by ethidium bromide contained in the gel, are visualized directly upon illumination with UV light. DNA molecules exposed to an electric field (defined by length of gel and a potential difference at its ends) migrate towards the anode due to the negatively charged phosphate groups along the DNA backbone.

The gels were prepared with an adequate volume of electrophoresis buffer (TBE or TPE 1X) with the desired amount of electrophoresis-grade agarose to a volume of electrophoresis buffer sufficient for constructing the gel. The agarose was melted in a

microwave oven at 55 °C. For fingerprints and DNA concentration gel it typically contained 0.8 %, or for the other purposes 1.5 %. Generally a 20 µg/ml ethidium bromide solution was added so that a final concentration of 0.5 µg/ml was obtained.

After the gel had hardened, it was placed into an electrophoresis chamber, with sufficient electrophoresis buffer to cover the gel. The DNA sample was prepared with 6x loading buffer. The samples were run in parallel with a DNA ladder or other DNA as molecular size standard.

The progress of separation at 50 - 120 volts within a limiting time was monitored by vision using UV light to illuminate the migrating stained DNA (see Figure 4.5).

4.7.1 Fluorescence photography of DNA in agarose gels

The DNA in agarose gels stained with ethidium bromide was photographed by illumination with UV light (>2500 µW/cm^2) in a complete photographic documentation system (UV transilluminator with a Polaroid camera and a digital photograph system), see figure 4.6.

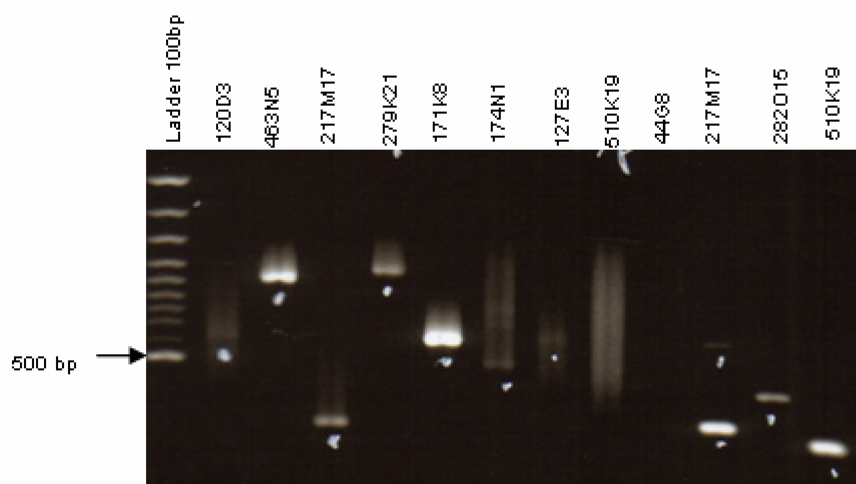


Figure 4.6: Fluorescence photograph of a 1 % agarose gel.

After electrophoresis of various DNA samples from BAC clones obtained by PCR with primer pTARBAC 3 in a 1 % agarose gel stained with 20 mg/ml ethidium bromide; ladder marker sizes 100 bp.

4.8 Extraction of DNA fragments from agarose gels

This method was used for different purposes:

- To isolate fragments of the inserts from different plasmids containing probes, e.g. framework genes, MHC class I Caja-G gene and others;

- to isolate fragments as hybridization probes;
- to isolate fragments after inverse PCR for sequencing.

The segment of a gel containing the DNA fragment of interest is excised (approximately 100 mg) and incubated with 6 M sodium iodide solution (NaI) at 55°C until the agarose has melted (Figure 4.7a and 4.7b). Thereafter, 20 µl of silica gel are added and the sample is incubated at 55°C for 10 min. It is centrifuged for one min at 21.000 rpm and the supernatant discarded.

The pellet is washed with 300 µl of the NaI solution and, thereafter, with 70 % ethanol. It is dried, resuspended in 20 µl of water and incubated for 20 min. Thereafter, the mixture is centrifuged at maximum speed for one min.

The supernatant contains the DNA, which is transferred in a new tube and the concentration and quality of the DNA is checked by agarose gel electrophoresis; see Figure 4.7c.

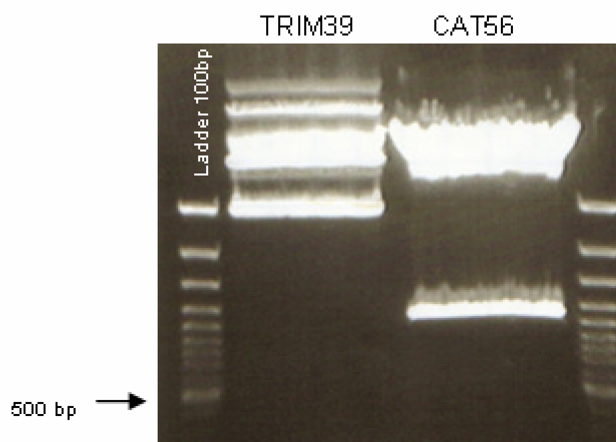


Figure 4.7 a: Electrophoretic separation of fragments after restriction digestion of two clones with TRIM 39 and CAT56 framework genes, respectively.

Two clones with TRIM 39 and CAT 56 framework genes were digested with restriction endonucleases *BamHI* and *HindIII*. These were separated according to size by electrophoresis in 0.8 % agarose gel containing 20 mg/ml ethidium bromide, ladder marker sizes 100 bp. The figure shows DNA by its UV absorption.

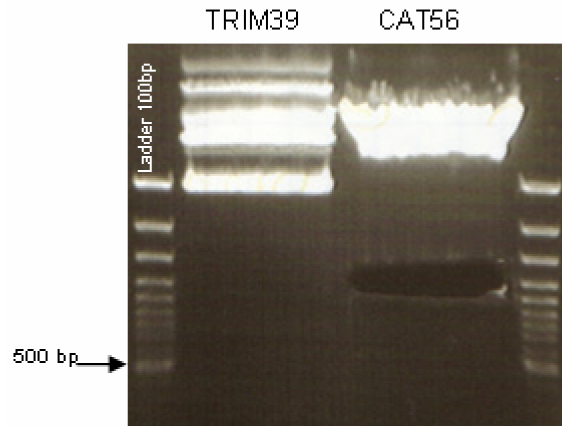


Figure 4.7 b: Cutting a fragment from a clone containing the CAT56 framework gene.

The 950 bp fragment of this clone with the CAT 56 framework gene was separated by gel electrophoresis in 0.8 % agarose gel containing 20 mg/ml ethidium bromide, ladder marker sizes 100 bp. Then excised for extraction of DNA from the gel sample by NaI method.

Figure 4.7 c: Agarose gel electropherogram for control of intactness of a fragment, CAT56.

The DNA gel extraction was done with NaI, and then was checked by electrophoresis in 1 % agarose gel stained with 20 mg/ml ethidium bromide, ladder marker sizes 100 bp.

4.9 Random primer method

Various specifically labeled oligonucleotide primers can assist by DNA polymerase in the initiation of DNA synthesis on single-stranded templates. Products of this primed DNA synthesis are radiolabeled by the use of one [α^{-32}P]dCTP and three unlabeled dNTP's as substrates.

Random priming, i.e. with a commercially available mixture of different primers was employed to label probes for screening of BAC clones and for analyzing plasmid DNA by hybridization after blotting. The probes were obtained by digestion of clones containing the fragments of interest. The DNA fragment was extracted from an agarose gel after electrophoresis.

For random priming, the commercial kit Megaprime DNA labeling system from Amersham or Promega was used. The protocol for the polymerase reaction is as follows: First, 25 ng of template DNA, 5 µl of 10 mM primer and 30 µl of bidistilled water are incubated at 100°C for 5 min. After the tube is on ice, the following reagents are added to a final reaction volume of 50 µl:10 µl of reaction buffer, 0.5 µl of 2 mM dNTPs (A, G, T), 2 µl of Klenow-DNA-polymerase and 50 µCi [α^{32}P]dCTP. The mixture was incubated at 37°C for 30 min, when the Amersham kit was used, and for one hour at room temperature when the Promega kit was used. The labeled probe was denatured at 100°C for 5 min and, after being chilled could be added to the hybridization solution. Labelled probes were stored at 4°C (Feinberg and Vogelstein, 1983).

4.10 Analysis of plasmid DNA by blotting and hybridization

4.10.1 Southern blotting

Southern blotting is the transfer of DNA fragments from an electrophoresis gel to a nylon membrane support. DNA fragments are immobilized as result of the transfer and of subsequent treatment. The membrane thus contains a semi-permanent reproduction of the banding pattern of the gel.

The gel containing the *EcoRI* digested plasmid DNA for every clone and the appropriate DNA size marker (100 bp) was stained with ethidium bromide and blotted onto a nylon membrane, positively charged by its chemical nature. To identify the positions of bands in the gel, it was photographed with a ruler laid along its side (Figure 4.8). The results obtained by this methodology are used for sequencing and mapping genes.

The protocol is divided into three stages: 1. preparation of the gel, 2. set up of the transfer and 3. immobilization of DNA.

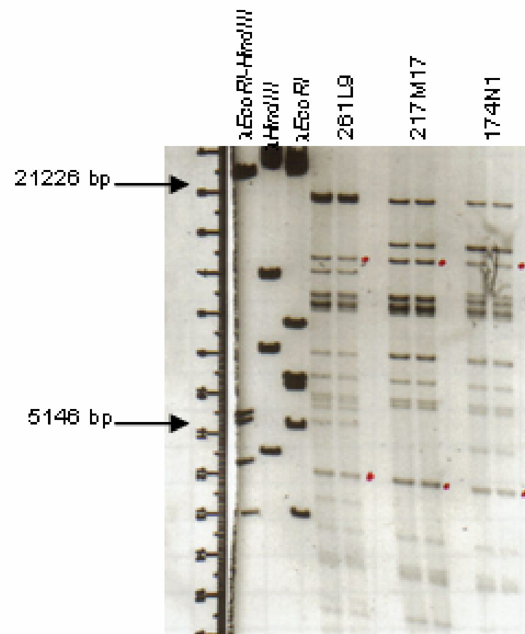


Figure 4.8: A gel before the Southern blotting.

In this 0.8 % agarose gel containing ethidium bromide, fragments from the clones 9(261L9), 18(217M17) and 25(174N1) have been subject to electophoresis, size markers, lambda DNA digested with *EcoRI*, *EcoRI-HindIII* and Hind III, respectively. On the left is a ruler to identify the fragments independently of any size changes of the picture. Note: the picture is a negative image.

4.10.1.1 Preparation of the gel

The agarose gel was pretreated by soaking in a series of solutions which depurinate, denature and neutralize the DNA.

The gel was rinsed in distilled water for 30 min on a platform shaker with slow agitation at room temperature. The water was discarded and the gel placed in 10 gel volumes of 0.25 M HCl to result in a partial depurination of the DNA fragments.

The HCl was poured off and the gel rinsed in distilled water. Finally 10 gel volumes of denaturation solution were added, and agitation continued for 20 min.

The denaturation solution was poured off and the gel rinsed with distilled water.10 volumes of neutralization solution were added, and agitation continued for 20 min. At this stage, the gel was ready to start its transfer onto the nylon membrane.

4.10.1.2 Set up of the transfer

The gel was placed on the plane surface of a reservoir and submerged half with transfer buffer 20x SSC.

A piece of nylon membrane just large enough to cover the exposed surface of the gel, was cut, laid into the gel and squeezed to disperse air bubbles by rolling over it a glass pipette over the surface.

Several pieces of Whatman 3M filter paper were put over the nylon membrane. Finally, a glass plate was laid on top of the pile, and a weight (0.2 - 0.4 kg) was placed on top to hold everything in place. The transfer was overnight.

4.10.1.3 Immobilization of the DNA

Before immobilization of DNA, the filter paper was removed and the membrane recovered. Beforehand, the position of the wells on the membrane was marked with a soft pencil to ensure that the up-down and back-front orientations remained recognizable.

When the membrane was air dried, the cross-linking was achieved by UV ligth for 30 sec at 1200 µJ. This leads to covalent attachment of DNA and enables the membrane to be reprobed several times.

4.10.2 Hybridization analysis of DNA

Hybridization analysis is sensitive and permits detection of even a single copy of a gene within a complex genome. The principle of this hybridization analysis is that single stranded DNA molecules of defined sequence (the probe) can base pair to a second DNA molecule that contains a complementary sequence (the target).

Hybridization was started with incubation of the membrane in pre-hybridization solution. It contained reagents which block non-specific DNA binding sites on the membrane surface, reducing background hybridization for 30 min at 60°C or 68°C. Hybridization was continued after replacement of the pre-hybridization solution by fresh hybridization buffer containing ^{32}P-labelled probe. The probe hybridized during an overnight incubation.

The membrane was washed twice with wash solution for 30 min at 60°C or 68°C. This step gradually removes bound probe molecules until only highly matched hybrids remain.

The membrane was wrapped in plastic foil for setting up autoradiography, see following next chapter.
After radioautography, the membrane was stripped by boiling for 15 min in 0.1 % SDS; this stripping step was perfomed several times until no radioactivity was left. The membrane was wrapped again and kept at -20 °C until the next use.

4.11 Autoradiography

In a darkroom, the membrane in plastic wrap was placed on an intensifying screen of calcium tungstate (Dupont). This is a light-tight X-ray film holder to increase the efficiency with which high-energy particles can be detected. The membrane was covered with a sheet of X-ray film from Biomax Kodak.
The film was exposed at -70 °C for an appropriate length of time. Thereafter, it was removed for developing in a dark room. The X-ray film was developed in an automatic X ray film processor.

4.12 Polymerase chain reaction (PCR)

In vitro enzymatic amplification of a specific DNA segment can be rapidly obtained by the so-called Polymerase Chain Reaction (PCR): A stretch of double-stranded DNA can be amplified provided two single-stranded oligonucleotide primers are flanking it, i.e. bound to the 5-ends of its two thermo-resistant DNA polymerase single strands, is used requiring the 4 deoxyribonucleoside triphosphates (dNTPs) as substrates, a buffer and salts including $MgCl_2$. An enhancer can be used to increase yield, specificity, and to overcome difficulties encountered with regions of a high GC content or with very long templates. These enhancers can be Triton, Tween 20, BSA or DMSO. The PCR reagents have been standardized and appropriate conditions such as temperature and concentrations have been defined for certain primer amplifications.
To obtain a specific good field of the final product, PCR makes use of the exponentially increasing number of templates. Thus, after 30 cycles of replication, a 2^{28} –fold (270 million-fold) amplification of the discrete product should result.
Each one of these amplification cycles requires a step of 1. denaturation of DNA molecules at 95°C, 2. annealing which means hybridization of DNA primers in a temperature range between of 40 °C – 68 °C, and 3. extension, the synthesis (replication) of new DNA by DNA polymerase (which catalyzes growth of new strands

from the 5` → 3` ends) extending across the segments of original DNA between the new primers at 72 °C.

4.12.1 Priming conditions and melting temperature

The primer hybridizations and polymerase reaction should be carried out under temperature conditions favoring the requirements of the enzyme plus a high degree of specificity at a minimum of background; the optimal annealing temperature is at or even above the calculated T_m (melting temperature) of the primers making their binding more discriminatory, therefore, more specific. The following formula was used to approximate the T_m of primers:

$$T_m = (C+G) \times 4°C + (A+T) \times 2 °C$$

Table 4.12.1: Components of master mix find concentrations for PCR reaction. The mixture of components for the PCR were prepared according to recipes given under Materials

Components	Final concentration
10 x PCR buffer	1 x
10 µM Primer A	1 µM
10 µM Primer B	1 µM
25 mM 4 dNTPs mix	0.2 mM
25 mM $MgCl_2$	1.5 mM
100 x Enhancer[d]	1 %
Taq polymerase	2.5 U/µl
Template DNA	100 - 500 ng/ml
H_2O up to total volume	-

[d] Enhancer agents (Triton, Tween 20 , BSA, DMSO)

Table 4.12.2 PCR cycling parameters. These cycling parameters were used to perform both PCR and inverse-PCR.

	Temperature	Step	Time	Cycle
1.	95 °C	Denaturation	2 min	1 cycle
2.	94 °C	Denaturation	1 min	30 cycles
3.	4 °C under T_m primers	Annealing	30 sec	
4.	72 °C	Extension	1 - 3 min	
5.	72 °C	Extension	7 min	1 cycle
6.	4 °C	Cool-stop	unlimited	-

To analyze the product, an aliquot of or the total amount of each reaction mixture was loaded onto an agarose gel of the appropriate concentration.

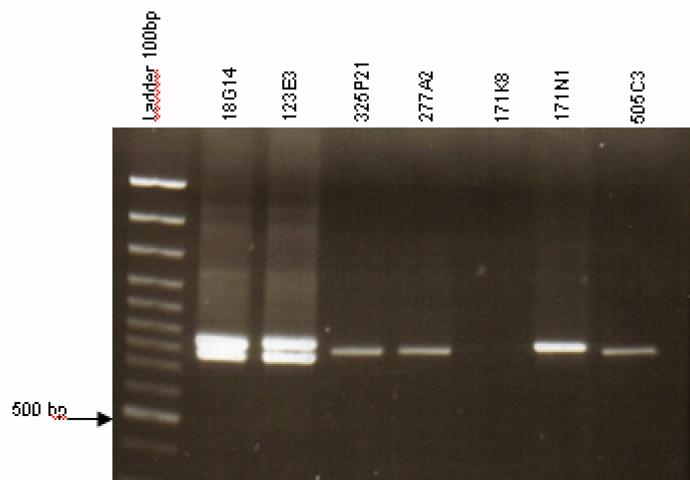


Figure 4.10: Analysis of PCR products by agarose gel electrophoresis.
PCR was carried out with primer HLA-B 29713 to the DNA plasmid of BAC clones samples were loaded into a 1 % agarose gel containing 20 mg/ml ethidium bromide; ladder marker sizes 100 bp.

4.13 Inverse PCR

To amplify DNA, an unknown clone that flanking one end of a known DNA sequence, for which no primers are available inverse PCR or inverted or inside-out PCR is used. The goal of inverse PCR is to generate PCR fragments that contain the end of a BAC insert which can be sequenced. The DNA plasmid was digested by a restriction enzyme, the individual restriction fragments were converted into circles by ligation. The resulting circularized DNA was used as template for inverse PCR. The primers were bound specifically to known sequences in the vector, pointing in opposite directions. Starting at the 3'-end of the bound primer, the subsequent unknown sequence was replicated and amplified in repeated cycles.
For inverse PCR, plasmids were digested with restriction enzyme PstI and incubated for two hours at 37°C. The enzyme was inactivated at 65°C for 20 min.
Ligation was achieved with T4 Ligase at 16°C overnight. The inverse PCR by produced DNA amplification containing the end of the insert just like a normal product of PCR carried out under the same conditions. The primers, used alternatively for inverse PCR were: ptarbac2, ptarbac3, ptarbac4, T7.29 or SP6.

4.14 DNA sequencing

Sequences were found by the Sanger method using the dideoxy-mediated chain reaction coupling with the commercial sequencing kit of the APPLIED BIOSYSTEMS company. A synthetic oligonucleotide primer was annealed to a single stranded DNA template. The reaction mixture contained a small portion of a 2´, 3´-dNTP which carried a 3´-H atom on the deoxyribose moiety, rather than the 3´-OH group. This molecule marked ddNTP was incorporated into a growing DNA chain, preventing the formation of a phosphodiester bond with the succeeding dNTP. The resulting populations of oligonucleotides were loaded onto a capillary-based automated DNA sequencer (ABI 3700) with the new generation Bigdye. This system incorporates dichlororhodamine dye terminators sequenced with AmpliTaq polymerase resulting in a very complex peak pattern. Therefore, the analysis is automated.

The sequencing reaction occurs in a 20 µl final volume including 0.5 - 1.0 µg DNA, 2 µl of 10 pmolar primer, 2 µl BigDye, 5 x sequencing buffer and water. The conditions are shown in Table 4.14.1.

Table 4.14.1: Cycling parameters for DNA sequencing. These cycling parameters were used to subsequently perform sequence analysis by ABI 3700.

	Temperature	Step	Time	Cycle
1.	96°C	Denaturation	2 min	1 cycle
2.	96°C	Denaturation	30 sec	25 cycles
3.	50°C	Annealing	15 sec	
4.	60°C	Extension	4 min	

The sequencing reaction was ended by precipitation with 250 µl 100 % ethanol and 10 µl 3M sodium acetate pH 5.5. The mixture was centrifuged at maximal speed for 15 min. Supernatant was discarded with a pipette; the pellet was washed in 200 µl 70 % ethanol and recentrifuged at maximal speed for 5 min. Finally, it was dried in the dry-pump machine.

Sequencing electropherograms were viewed with EditView 1.0.1 software and Bioedit program. The sequence analyses were realized with Clustal X version1.8 for alignments and sequence identity (see Figure 4.11).

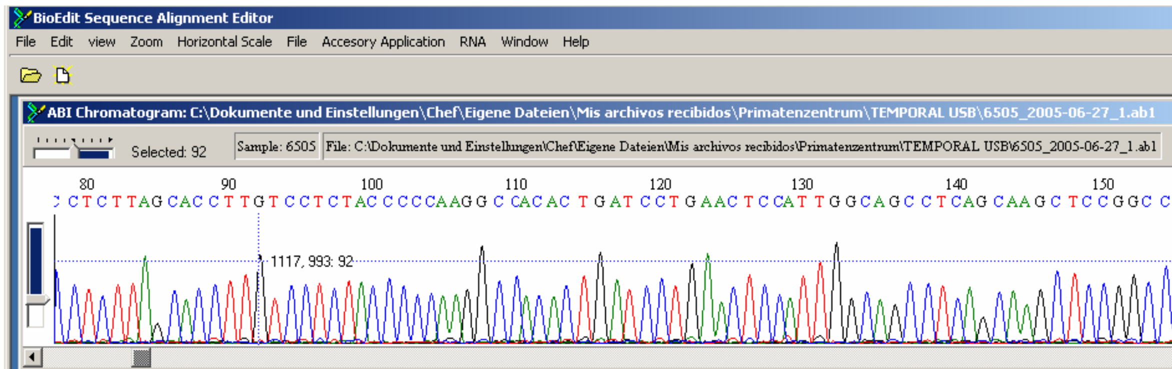


Figure 4.11: Sequencing electropherogram of a DNA fragment from a BAC clone. The sequence was performed in the ABI 3100. Sequence analysis was viewed with bioedit equipped with editView 1.0.1 software from Applied Biosystems.

The sequences were translated in BLAST and were analyzed in Swiss-prot. To resolved some difficult in the sequence analysis was used AUGUSTUS software. The evolution analysis with sequences was realized with Clustal X, Genrunner version 3.05 and MEGA 3.1.

5 Results

The results of these pilot studies on the MHC class I region in *Callithrix jacchus,* on gene order and DNA sequences, their location, assignment and eventually interpretation are reported essentially in the sequence of their production. A table of contents indicating page numbers is placed behind the title on this thesis.

This technical note may help to read the Results chapters: A solution of the layout problem of combining a fairly brief text with documents showing the results, i.e. numerous figures and tables each filling often a whole page or more, shall be attempted by giving in the beginning of chapter a short introductory text providing a survey of the subjects to be delft with. Detailed yet brief explanations will then be given in the legends for enabling study of the data presented right next to these. A Conclusions paragraph may be added, where adequate, at the end of the chapter. Only discussion proper, of problems, criticisms, observations seeming of general interest, and ideas hopefully leading further, is left for the discussion section.

Although it is needed as seen from various aspects mentioned, no information is available on the number, order and clustering of genes in the MHC class I region of the Common marmoset. Therefore, the first aim was to construct a physical map showing a sequence of genes in the MHC class I region based upon fragment analysis of clones from the BAC bank CHORI-259.

The work finally resulted in a partial description of the order of MHC class I genes, especially in contig 1(Figure 5.2.1.3.1). In consideration of the much greater value of sequence data - compared to the indirect evidence of such fragment-gene assignments, - and in view of the facilities existing to do sequence analysis and had decided in time to start a partial sequencing of the MHC class I region. BAC clones selected (5.1) for making the first aim (5.2) could thus be tested also for usefulness in sequencing by PCR, i.e. following the second aim (5.3). The partial sequences obtained were localized by comparative alignment in a Hs MHC class I region map drawn to bp-scale (5.4); both complemented the results of project 1 and allowed a study on evolution of sequences (5.5). Clustal X and MEGA 3.1 programs for constructing phylogenetic trees of genes could be used to reversely identify unknown class I pseudogenes within unassigned DNA sequences, indicating the nearest relatives of these genes (5.6). Knowledge of the sequences alignment and gene contents may serve, furthermore, as milestones on the way of a complete sequencing of the MHC class I region of *Callithrix jacchus.*

Sequence TS1 turned out to be, however, the most instructive and challenging one. It might be once considered a “Rosette stone” for immunogenetics, a key to the evolution of the classical class I gene duplicons.

5.1 Identification of BAC clones from *Callithrix jacchus* containing MHC class I region gene

The BAC clones containing class I information were identified e.g with Caja-G a probe of cDNA from *Callithrix jacchus* prepared in P^{32} labelled form: MHC class I Caja-G a fragment of 734bp of clone 15 of liver A389 cDNA. It contains exon 2 with 272bp (polymorphic), exon 3 with 275bp (polymorphic) and exon 4 with 187bp (conserved). This probe is named Caja-G had been kindly provided by Ulrike Geisler.
The MHC class I gene complex has shown homology between human, rat, mouse, pig, cat, monkeys like Rhesus, horse and other mammals (Amadou *et al*., 1999; Trowsdale, 1999; Daza-Vamenta *et al*., 2004; Hurt *et al*, 2004; Beck *et al*., 2005; Chardon *et al*., 2000; Tallmadge *et al*., 2005). The MHC class I probe Caja-G was chosen in recognition of the fact that MHC class I proteins have homology not only between species but also with other members of the HLA family of genes, and to a lesser extent with their pseudogenes found within gene clusters of class I region. Caja-G has a high degree of identity with the other class I probe used, called HLA-B, i.e. taken from *Macaca mulatta* (Rhesus monkey), both monitoring the same fragments in hybridization. MICA, a third class I probe, does not cross-react with both Caja-G and HLA-B in hybridization.
For a physical mapping, all clones of BAC Bank CHORI-259 were screened. This bank contains the complete genome of *Callithrix jacchus* incorporated at an 11.2-fold genomic representation into pTARBAC 2.1 vector clones of *E.coli* (Materials, 3.9.).
An initial screening of the library with the Caja-G probe yielded 256 positive BAC clones. Framework gene sequences within the isolated BAC clones were demonstrated also by hybridization. The probes used contain so-called framework genes, conserved sequences belonging to the MHC organization in humans and in rats from which they had been obtained (Amadou *et al*., 1995). Altogether, they are flanking all four intervals containing class I and other genes. In Hs these intervals are arranged in the following manner: BAT1 - POU5F1, TCF19 - CAT56, GNL1 - (TRIM39, TRIM26) – TCTEX5 and TCTEX4 - MOG (MHC class I map, text and Figure 5.1.1). A probe POU5F1 was included, because the TCF19 framework probe

did not hybridize with small restriction fragments nor in the final Southern blots - in contrast to the whole clones which did so. A probe of K18T3 was generated, furthermore, in an attempt of closing a gap within contig 1. Of course, it has a unique specificity in hybridization, not cross-reacting with the class I probes.

153 BAC clones were found by hybridization to contain MHC class I gene(s) and MIC gene(s), respectively. For establishing a genomic map for the MHC class I region of *Callithrix jacchus,* 42 clones were allocated to the four class I intervall, most of these (31) belonging to contig 1. See of their probe-hybridizations are summarized in table 5.2.1.2.

Figure 5.1: Map of the MHC class I region in Hs showing to bp-scale the arrangement of four intervals called contigs 1 – 4.

The class I protein genes are clustered (scarcely in contigs 2 and 3) between conserved framework genes (marked green in figure 5.4.1). Figure made by Prof. Johannes Heinrich Matthaei.

5.1.1 Contig 1, the interval BAT1 - TCF19

The interval between BAT1 and TCF19 is polymorphic. Polymorphic applies to regions with genetic alterations between species, e.g. in respect to the MHC class I region with the genes HLA-B, HLA-C and MIC A/B in several mammalian species. Double positive screening of BAC clones was registered between MHC class I gene Caja-G* and the following probes: ATP6V1g2, BAT1, TCF19, HLA-B, POU5F1, MICA and K18T3 (see table 5.1.1.1).

*) Following a lab routine, in this text, the probe is named occasionally MHC class I instead of MHC-class I-Caja-G or just Caja-G that is precise and sufficient.

Table 5.1.1.1: Screening of the interval between BAT1 and TCF19

The 11 filters of BAC-bank CHORI-259 were screened with the BAT1, TCF19, POU5F1, MICA, K18T3, HLA-B and MHC class I probes, respectively. Numbers of the clones matching with the indicated groups of these probes, tested individually, are listed.

Clones positive to probes	No. of clones
TCF19, MHC class I	11
BAT1, MHC class I	24
TCF19, BAT1	45
TCF19, BAT1, MHC class I	18
POU5F1, MHC class I, TCF19	5
POU5F1, BAT1, MHC class I	4
POU5F1, BAT1	2
POU5F1, HLA-B, MHC class I	2
POU5F1, TCF19, MHC class I	5
POU5F1, BAT1, MHC class I	4
K18T3, MHC class I	1
K18T3, BAT1, TCF19, MHC class I	1
K18T3, BAT1, MICA, HLA-B, MHC class I	1
K18T3, HLA-B, TCF19	1
K18T3, BAT1	1
K18T3, MICA, HLA-B, MHC class I	1
K18T3, MICA	1
K18T3, HLA-B	6
K18T3, HLA-B, MHC class I	21
K18T3, HLA-B, TCF19, MHC class I	1
K18T3, HLA-B, BAT1, MHC class I	4

From the BAC clones screening with BAT1, TCF19 and MHC class I probes, respectively, 31 were selected for characterization of their fragments by Southern blot fingerprinting and sequence analysis. These are listed in table 5.1.1.2.

Table 5.1.1.2: Screening into the interval between BAT1 and TCF19 by means of BAC-bank CHORI-259.BAC-clones are characterized by their filter number, plate and coordinates. All probes from other mammals tested here, showed complementarily with BAC clone DNA from *Callithrix jacchus*.

Filter	Plate and coordinates	Positive probe	Positive probe	Positive probe	Positive probe
03I	120D3	MHC class I	TCF19	BAT1	
06I	271C5	MHC class I	TCF19	BAT1	
03I	127E13	MHC class I	TCF19	BAT1	
06I	261L9	MHC class I	TCF19	BAT1	
01I	18G14		TCF19	BAT1	
07G	282L12		TCF19	BAT1	
05I	204C3		TCF19	BAT1	
05I	239N18		TCF19	BAT1	
05I	193P12		TCF19	BAT1	
06I	282O15		TCF19	BAT1	
01I	44G8	MHC class I		BAT1	
011G	510K19		TCF19	BAT1	MICA
05I	217M17	MHC class I			MICA
04I	161C12			BAT1	MICA
04I	174N1	MHC class I			MICA
08G	347K7	MHC class I	TCF19	HLA-B	POU5F1
08G	347D1	MHC class I	TCF19	HLA-B	POU5F1
08G	379F14	MHC class I	TCF19		POU5F1
09G	425 A13	MHC class I	TCF19		POU5F1
10G	459O20	MHC class I	TCF19		POU5F1
06I	261K9	MHC class I		BAT1	POU5F1
04I	169G2	MHC class I		BAT1	POU5F1
03I	123H17	MHC class I		BAT1	POU5F1
03I	123E17	MHC class I		BAT1	POU5F1
04I	150M18			BAT1	POU5F1
03I	99L19			BAT1	POU5F1
07G	334C4	MHC class I	K18T3	HLA-B	
07G	329M5	MHC class I	K18T3	HLA-B	
08G	379E3	MHC class I	K18T3	HLA-B	
04I	174N1	MHC class I	K18T3	HLA-B	MICA
011G	485C4		K18T3		MICA

5.1.2 Contig 3, the intervals TRIM39 - TRIM26

The interval TRIM39 - TRIM26 of the MHC class I region is polymorphic in human. Their existence in *Callithrix jacchus* was demonstrated by double positive screening of the MHC class I probe Caja-G, with probes CAT56, TRIM39 and TRIM26 probes, respectively (see Figure 5.2.2.3). In this screening of the interval CAT56-TRIM39, 20 positive clones were found (Table 5.1.2.1). For the MHC class I-TRIM39-TRIM26 interval, 21 positives clones were detected (Table 5.1.2.2).

Table 5.1.2.1: Screening for clones containing Caja-G and frameworks genes of CAT56 - TRIM39 (part of contig 3).

Group of probes individually matching BAC clones	No. of BAC clones positive with each probe of the group	Clones found on filters *
CAT56, Caja-G	9	03I, 04I, 05I, 06(3), 07G (2), 09G
TRIM39, CAT 56	3	06I(2), 07G
TRIM39, Caja-G	9	01I, 05I(3), 06I, 07(3), 09G, 010G(3)
CAT56, TRIM39, Caja-G	1	07G
TRIM39, TRIM26, CAT 56	3	06I(2), 07G

* In parentheses: number of clones on filter mentioned

One BAC clone, number 325P21 on filter 7 (07G), was found to be screened with four framework probes, MHC class I Caja G, CAT56, TRIM39 and TRIM26.
BAC clone 325P21 was selected for characterization by Southern blot fingersprints.
All other framework probes (7) flanking the MHC class I region were positive.
8 BAC clones were chosen for screening with TRIM39, TRIM26 and MHC class I in order to characterize their matching DNA by Southern blot fingersprints, see table 5.1.2.3.
For control, Southern blot analysis was done also with nylon membranes carrying clones screening for the BAT1-TCF19 interval. All these clones were negative (Table 5.1.2.3).
Of the BAC clones positive in screening to TRIM39, TRIM26 and MHC class I chosen eight were chosen for characterization by Southern blot fingerprints; clone 325P21 was not included.
In Southern blot analysis, membranes with clones containing interval BAT1-TCF19 were tested for hybridization with probes TRIM39, TRIM26, POU5F1, HLA-B, CAT56, TCTEX5, TCTEX4 and MOG. All of these controls were negative (Table 5.1.2.3).

Table 5.1.2.2 Screening for clones containing Caja-G and the interval TRIM39 - TRIM26 (part of contig 3).

Group of probes individually matching BAC clones	No. of BAC clones positive with each probe of the group	Clones found on filters *
TRIM39, TRIM26	14	01I, 05I(2), 06I(5), 07G(2),010G(2), 011G(2)
TRIM26, Caja-G	11	03I, 04I(2), 05I(2), 06I, 07G, 010G(4)
TRIM39, Caja-G	9	01I, 05I(2), 06I, 07G(2), 09G, 010G(2)
TRIM26, TRIM39, Caja-G	7	01I, 05(2), 06I, 07G, 010G(2)

Table 5.1.2.3: Screening into the intervals CAT56 - TRIM39 and TRIM39 - TRIM26 by means of BAC-bank-bank CHORI-259: Each BAC-clone is characterized by its filter number and plate coordinates. All the probes from other mammals tested showed complementarily, with BAC clone DNA from *Callithrix jacchus*.

BAC clones	Filter No.	Positive with probe	Positive with probe	Positive with probe	Positive with probe
245C6	06I	Caja-G	TRIM26	TRIM39	---
234L16	05I	Caja-G	TRIM26	TRIM39	---
463N5	010G	Caja-G	TRIM26	TRIM39	---
19I16	01I	Caja-G	TRIM26	TRIM39	---
325P21	07G	Caja-G	TRIM26	TRIM39	CAT56
279K21	06I	---	TRIM26	TRIM39	---
277A2	06I	---	TRIM26	---	---
436B4	010G	Caja-G	TRIM26	---	---

5.1.3 Contig 4, interval TCTEX4 – MOG

The interval was investigated using 15 BAC clones for multiple positive screenings with probes Caja-G, HLA-B, TCTEX4 and MOG (Table 5.1.3.1). The contig was constructed by means of fragments from two of these15 clones (see Figure 5.2.3.3.1).

Table 5.1.3.1: Screening of the interval TCTEX4 - MOG:

Group of probes individually matching BAC clones	No. of BAC clones positive with each probe of the group	Clones found on filters (No. of clones)
TCTEX4, Caja-G	8	01I, 03I, 04I, 07G,010G,011G(3)
MOG, Caja-G	3	01I, 04I, 08G
MOG, TCTEX4	5	06I, 08G, 011G(3)
Caja-G, MOG, TCTEX4	1	01I

Table 5.1.3.2: Screening into the interval TCTEX4 - MOG by means of BAC-bank CHORI-259 clones with probes TCTEX4, MOG, Caja-G, HLA-B and MICA: Each BAC-clone is characterized by its plate coordinates and filter number. All the probes from other mammals showed hybridization with BAC clone DNA from *Callithrix jacchus*.

BAC clones	Filter No.	Positive with probe	Positive with probe	Positive with probe	Positive with probe
348C21	08G	Caja-G	TCTEX4	MOG	HLA-B
171K8	04I	Caja-G	TCTEX4	MOG	HLA-B

5.2 Genomic analysis by Southern blot hybridization

To determine with BAC clones the arrangement of genes within contigs of the MHC class I region, *EcoRI* digestion fingerprints of DNA from the 42 selected BAC clones were probe-hybridized in Southern blots for the analysis of gene content of the fragments. With the relatively few clones used, only three contigs were found in the MHC class I regional organization using 10 different framework specific probes, the MHC class I probes Caja-G, HLA-B from Mamu and MICA, furthermore, a probe for the gene K18T3. All hybridization results are collected in table 5.2.1.2.2 summarizing the data from all 42 clones used for representation of contigs 1, 3 and 4. It contains the clones of fragments split from the clone's inserts by the *EcoRI* restriction enzyme, and their probe-screening.

5.2.1 Contig 1, the interval BAT1 - TCF19

5.2.1.1 Southern blot analysis of contig 1

The genes in this contig turned out to be the same as in human, mouse, rat and chimpanzee. As intended initially, this contig should have included the interval between the framework genes BAT1 and TCF19. TCF19 is localized next to POU5F1. 10 clones by screening results covering BAT1 – TCF19 were identified by hybridization on figerprintes. These BAC clones are: 510K19, 261L9, 204C3, 161C12, 44G8, 171N, 217M17, 485C4, 334C3 and 329M5. With these, 13 probe Caja-G hybridizing fragments were found for contig 1 between BAT1 and POU5F1. But it seemed difficult to assemble all of these into a map. Only the four fragments of 4.2, 9, 10 and 11 kb could be positioned. The HLA-B probe (Mamu-B) was found positive with 7 *EcoRI* fragments. Of these, the fragments of 6 and 7 kb did not seem to fit in. At least 8 class I genes were expected to be located in the region between

BAT1 and POU5F1. The MICA probe hybridized to an *EcoRI* fragment with length 17 kb in the clones 510K19, 204C3, 261C12, 161C12, 44G8, 217M17, 171N1, 485C14, 334C4 and 329M5.

Initial screening of contig 1 presented the problem of TCF19 hybridizing in clone screening, however, not in *EcoRI* fragments of the clones. TCF19 is located between POU5F1 and HCR. A Southern blot analysis for framework gene HCR was negative in all clones, too. To close this gap between contig 1 and contig 2, screening with a POU5F1 probe was done. 10 clones were found POU5F1-positive: 379F14, 425A14, 459O20, 261K9, 169G2, 123H17, 150M18, 99L19, 317K7 and 347D1. Eigth of these clones were bought and found POU5F1-positive also in Southern blot analysis: 379F14, 425A14, 459O20, 169G2, 123H17, 99L19, 317K7 and 347D1. However, there was no clone covering HLA-B and POU5F1. In the PUO5F1-screening clones, two fragments of 10 and 11 kb were found both of which were positive for Caja-G and HLA-B probes. The assumption was that HLA-B/Caja-G-related and therefore hybridizing genes would be located only in or near the gene HLA-B. In the aim to close a gap between HLA-B and PUO5F1, a primer called K18T3 was prepared from clone 217M17. With this primer, a probe was made and used for screening the 11 filters of the BAC bank. In this screening, 5 clones were found: 174N1, 217M17, 485C4, 329M5 and 334C4. Southern blot analysis of these clones showed a 17 kb fragment positive for MICA. Therefore, the following clones overlap both K18T3 and MICA: 174N1, 485C4, 329M5 and 334C4. Overlapping to HLA-B was indicated by a fragment of 9 kb being positive for both Caja-G and Mamu-B probes. Clone 329M5 had two fragments, of 9 and 2.3 kb, both positive for Caja-G and Mamu-B probes. The clones 334C14 and 379E3 overlapped with fragments of 4.2 or 2.2 kb, respectively, with K18T3; with fragments of 11 kb (in addition, of 4.2 kb in 379E3 and 18G149) positive for Caja-G and Mamu-B probes. That means, these clones are in the cluster containing fragments positive for Caja-G/Mamu-B and POU5F1. The clones positive for probe K18T3 screened to close the gap between HLA-B and the framework gene POU5F1.

All fragment - gene-probe - assignments obtained for the clones used for restriction mapping are collected in table 5.2.1.2.1 providing the basis for construction of maps, mainly for the interval BAT1 - POU5F1, contig 1.

A final construction of the contig 1 map led to the conclusion that class I (=Caja-G) positive fragments must be assumed to occur on both sides of K18T3. This probe finally, turned out to be useful, but not absolutely necessary for obtaining the map.

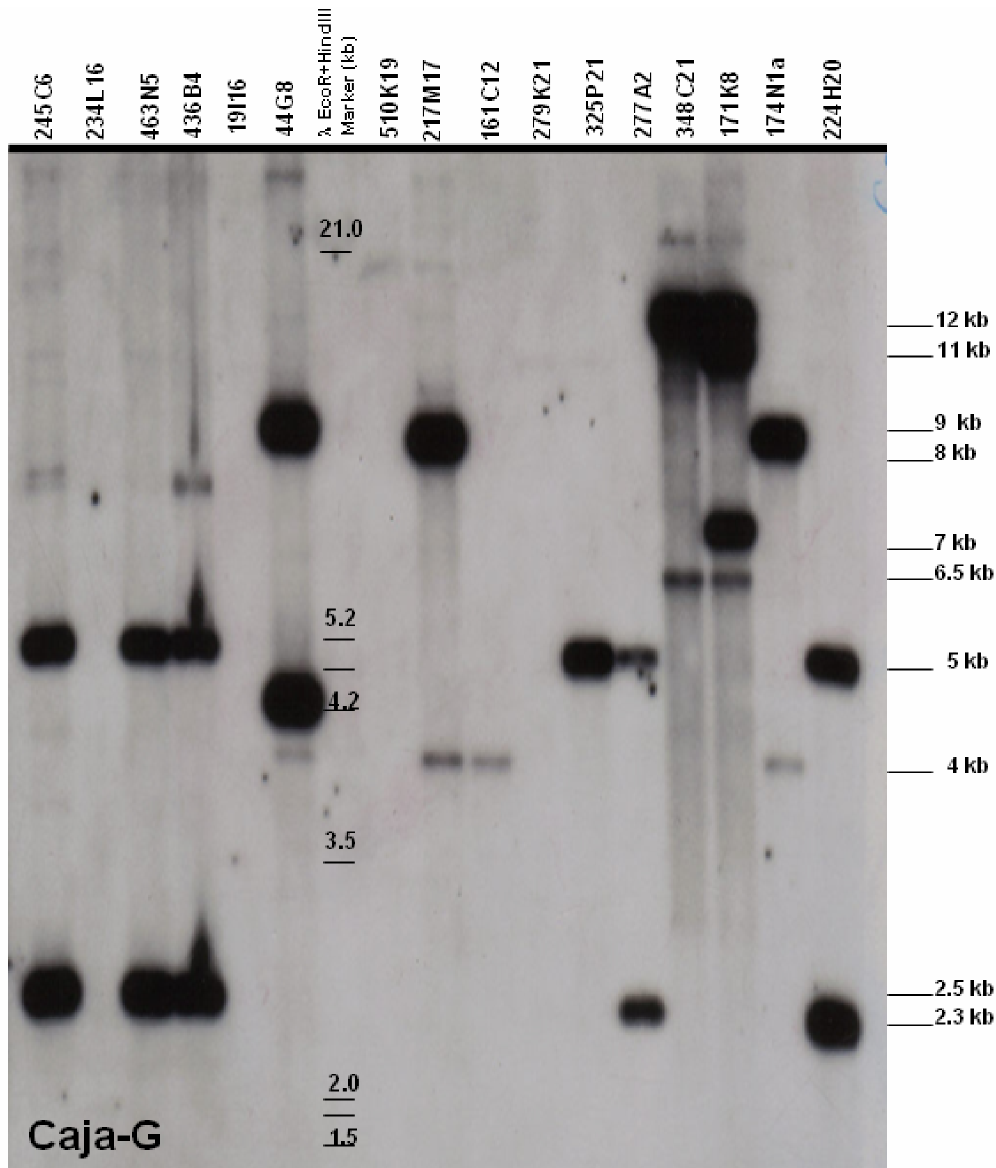


Figure 5.2.1.1.1: Autoradiograph after Southern blot hybridization of Caja-G probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

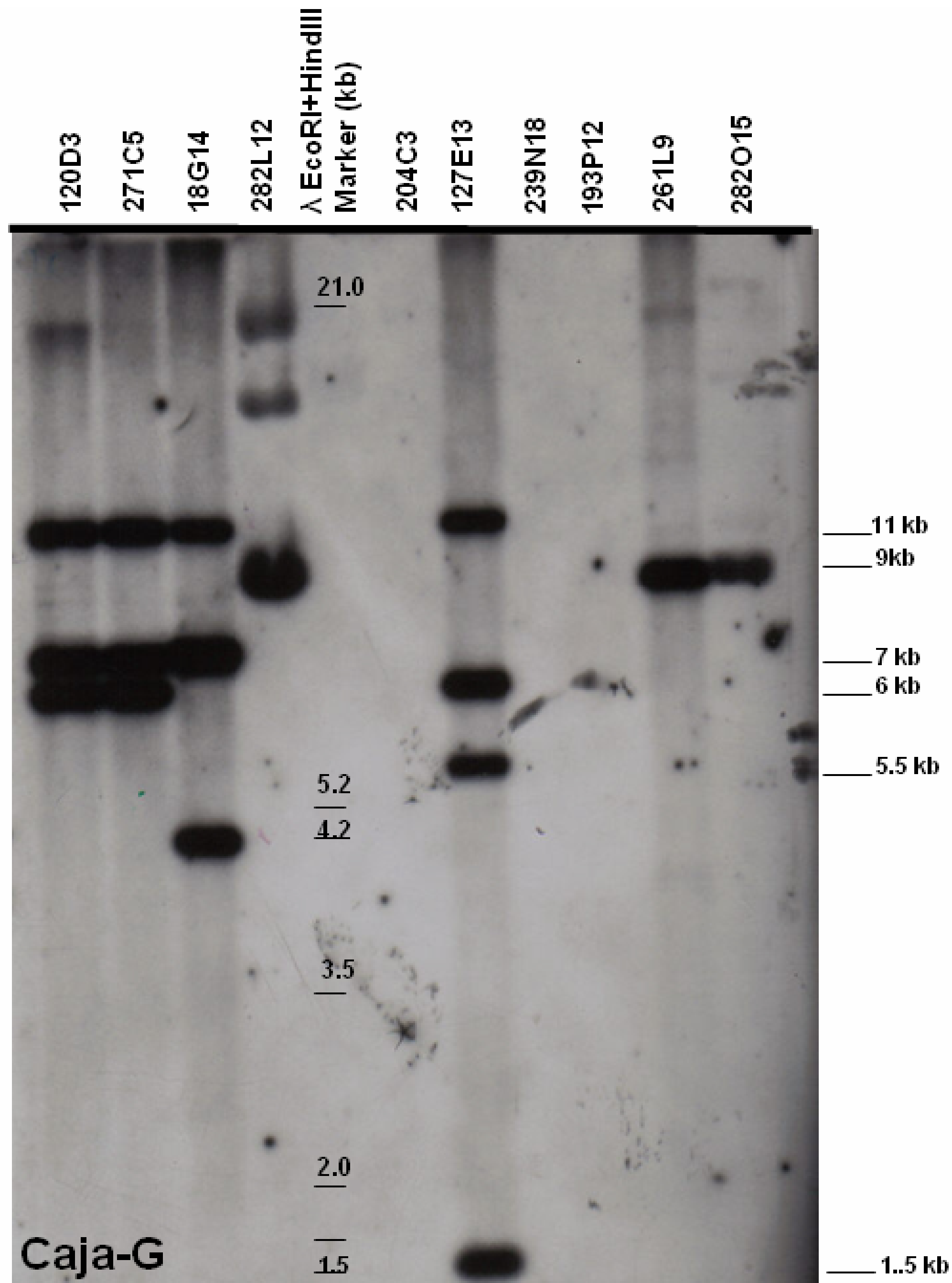


Figure 5.2.1.1.2: Autoradiograph after Southern blot hybridization of Caja-G probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence

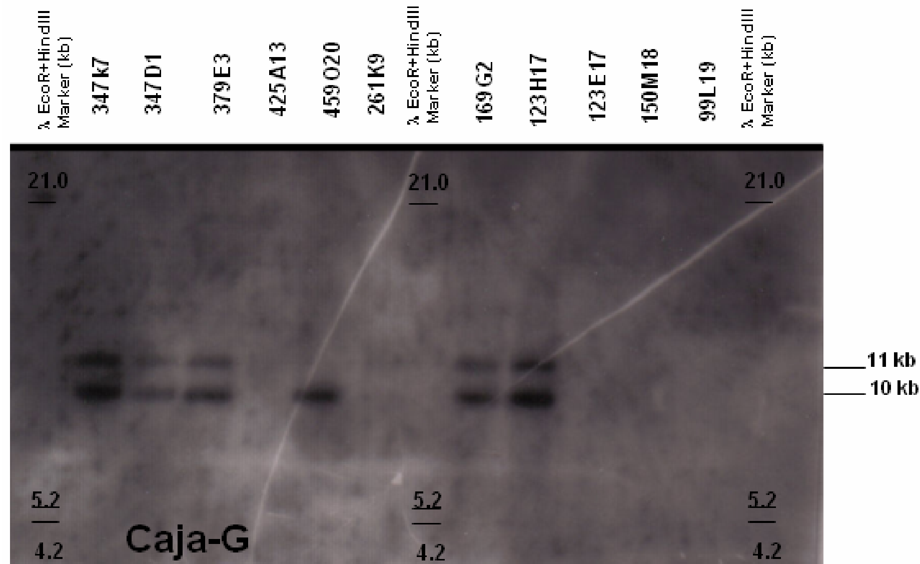


Figure 5.2.1.1.3 Autoradiograph after Southern blot hybridization of Caja-G probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI*+*HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

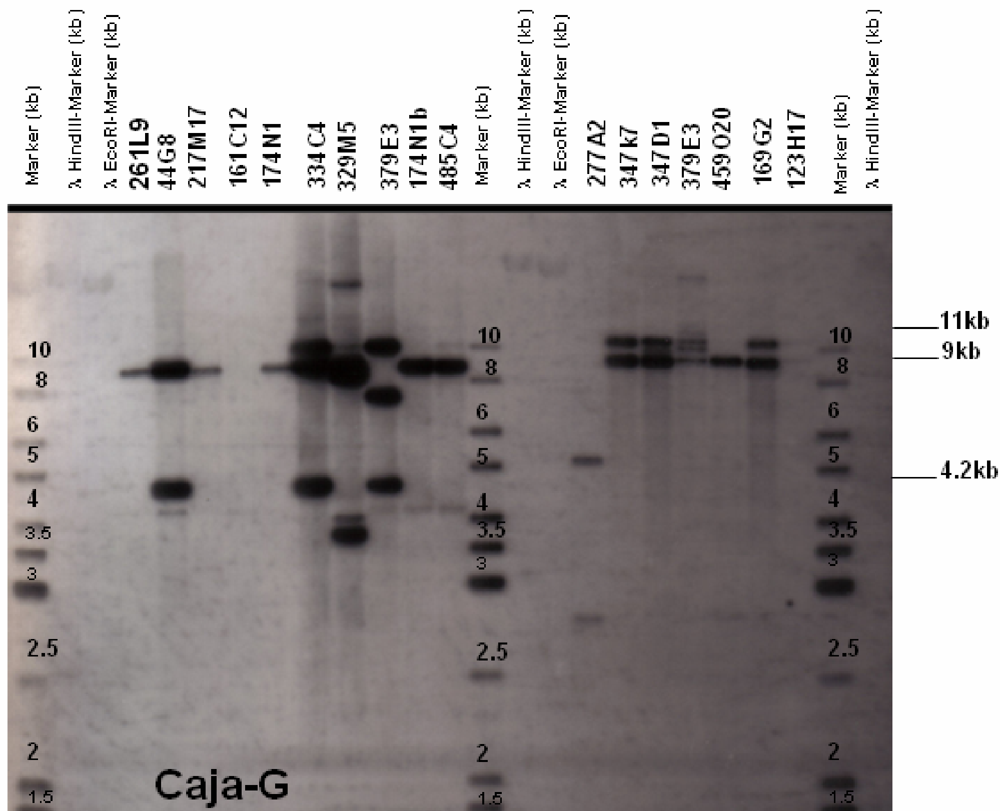


Figure 5.2.1.1.4: Autoradiograph after Southern blot hybridization of Caja-G probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

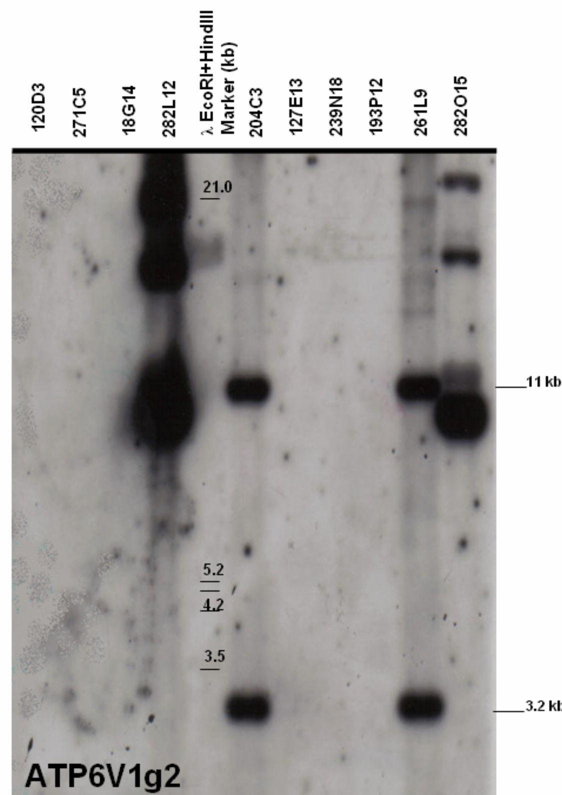


Figure 5.2.1.1.5: Autoradiograph after Southern blot hybridization of ATP6V1g2 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

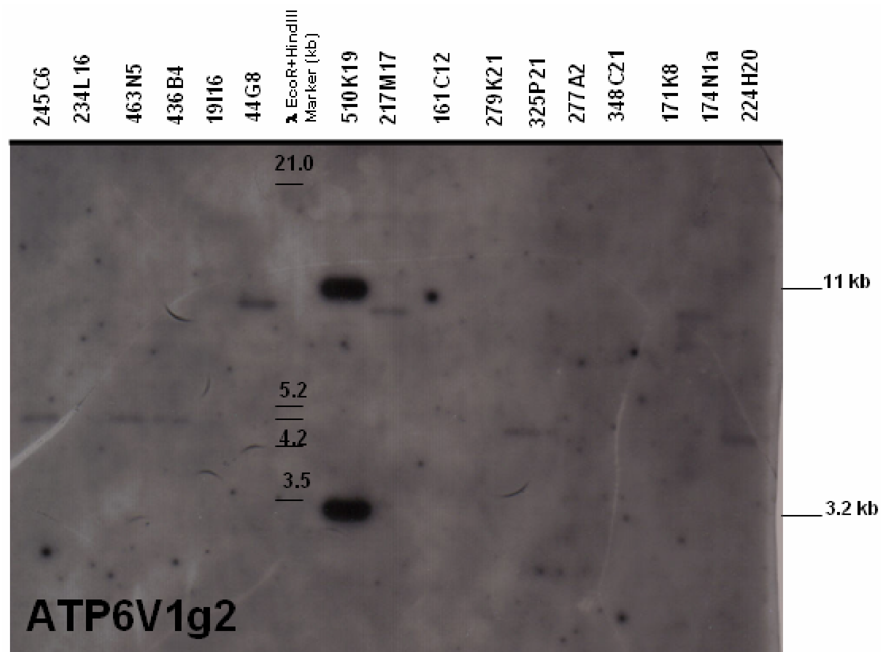


Figure 5.2.1.1.6: Autoradiograph after Southern blot hybridization of ATP6V1g2 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

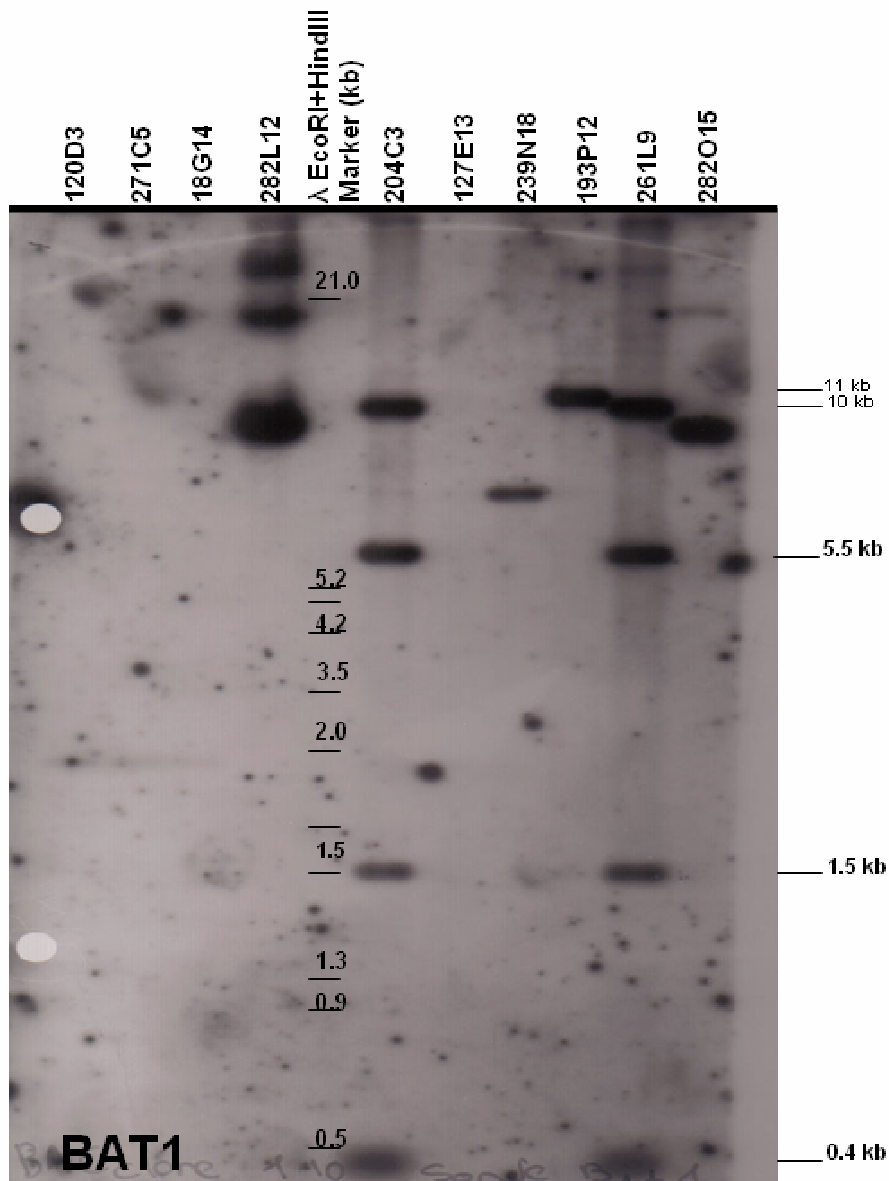


Figure 5.2.1.1.7: Autoradiograph after Southern blot hybridization of BAT1 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

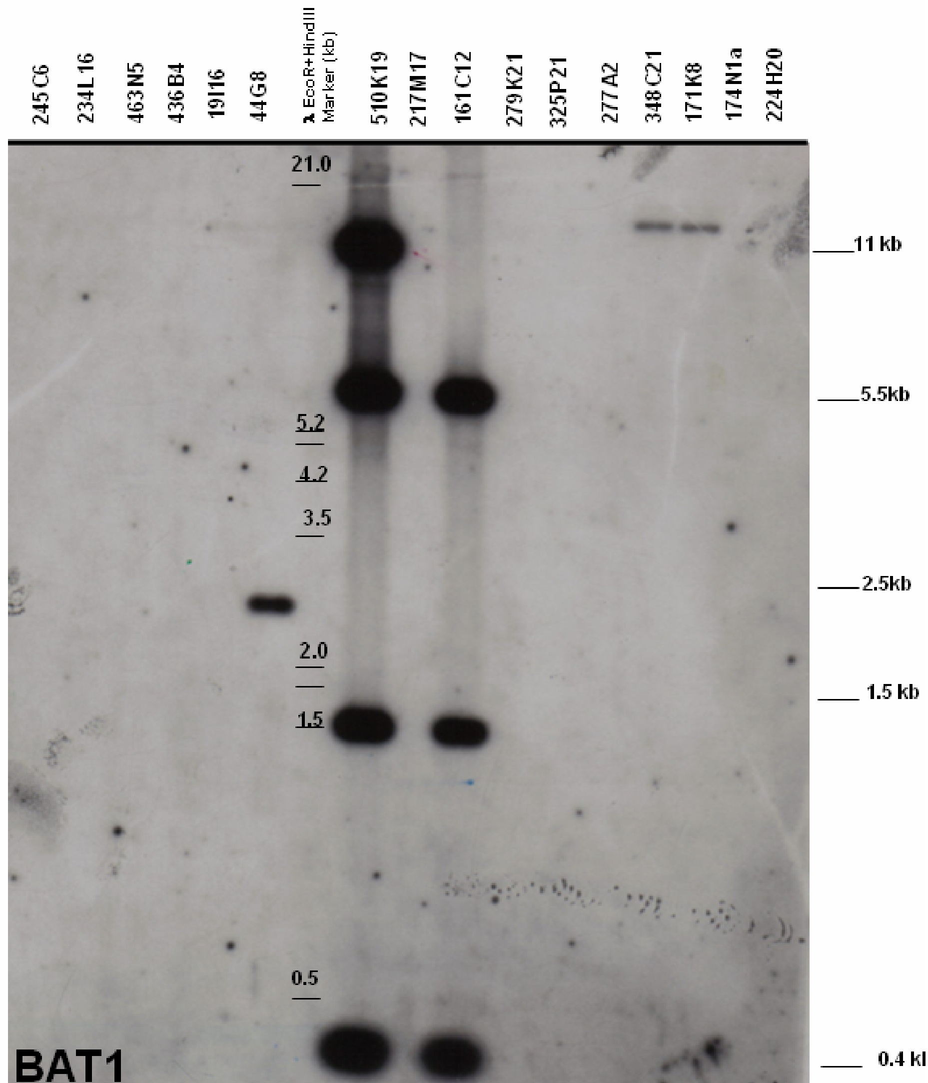


Figure 5.2.1.1.8: Autoradiograph after Southern blot hybridization of BAT1 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

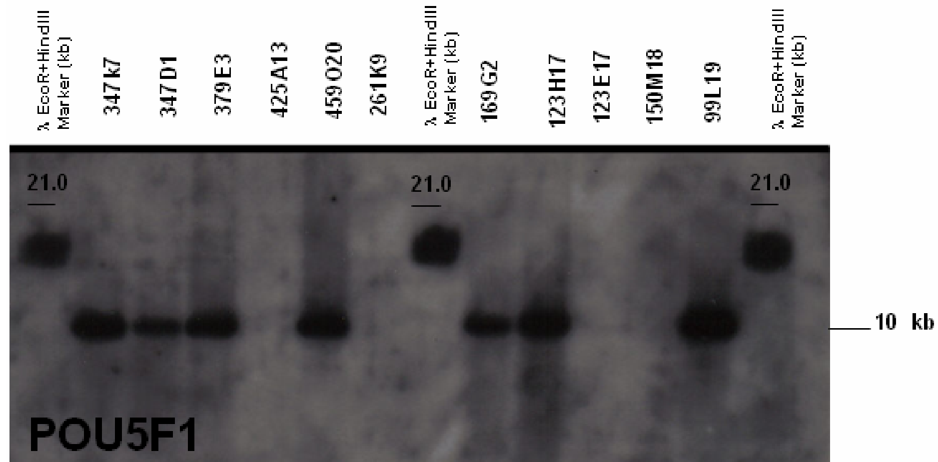


Figure 5.2.1.1.9: Autoradiograph after Southern blot hybridization of POU5F1 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI*+*HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

Figure 5.2.1.1.10: Autoradiograph after Southern blot hybridization of K18T3 probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI*+*HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

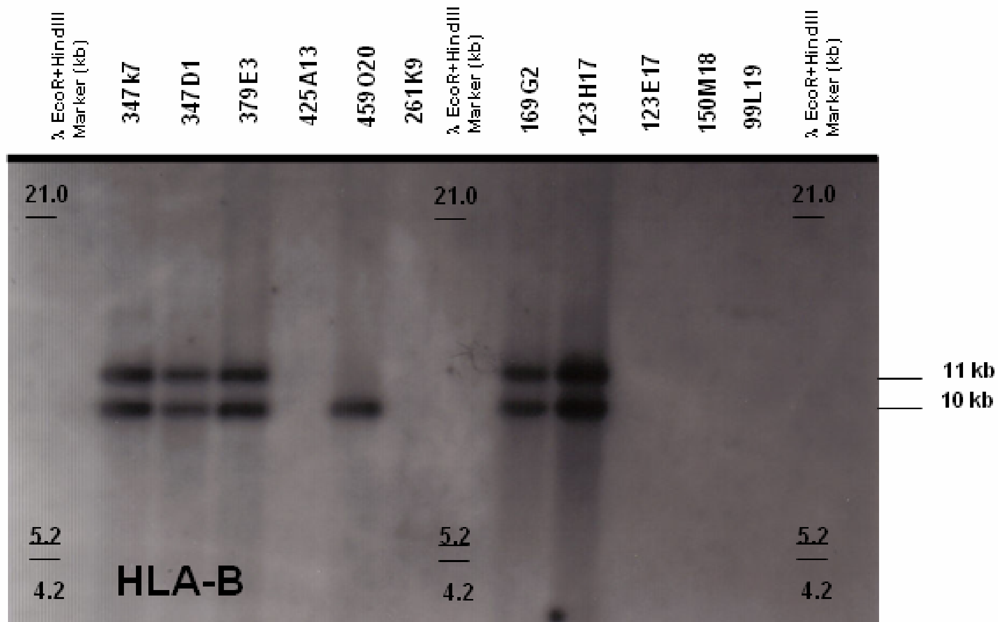


Figure 5.2.1.1.11: Autoradiograph after Southern blot hybridization of HLA-B probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

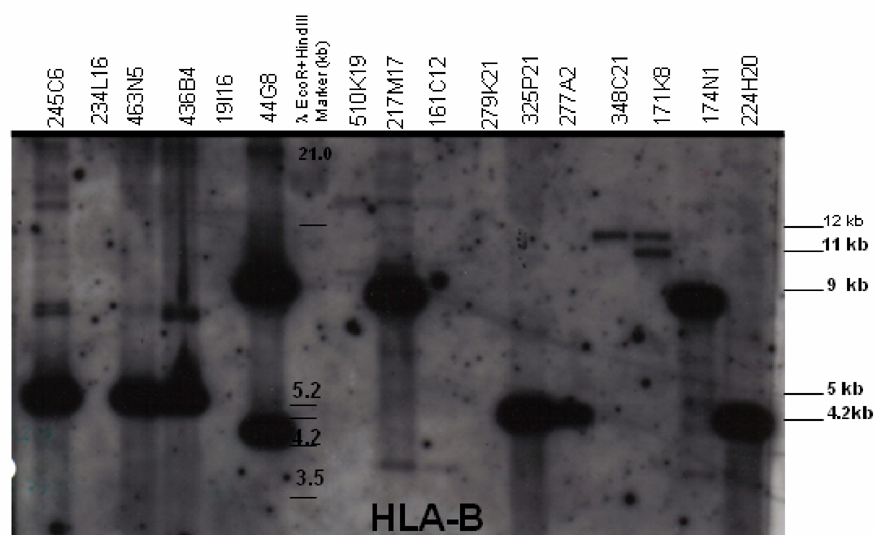


Figure 5.2.1.1.12: Autoradiograph after Southern blot hybridization of HLA-B probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme EcoRI. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) EcoRI+HindIII used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

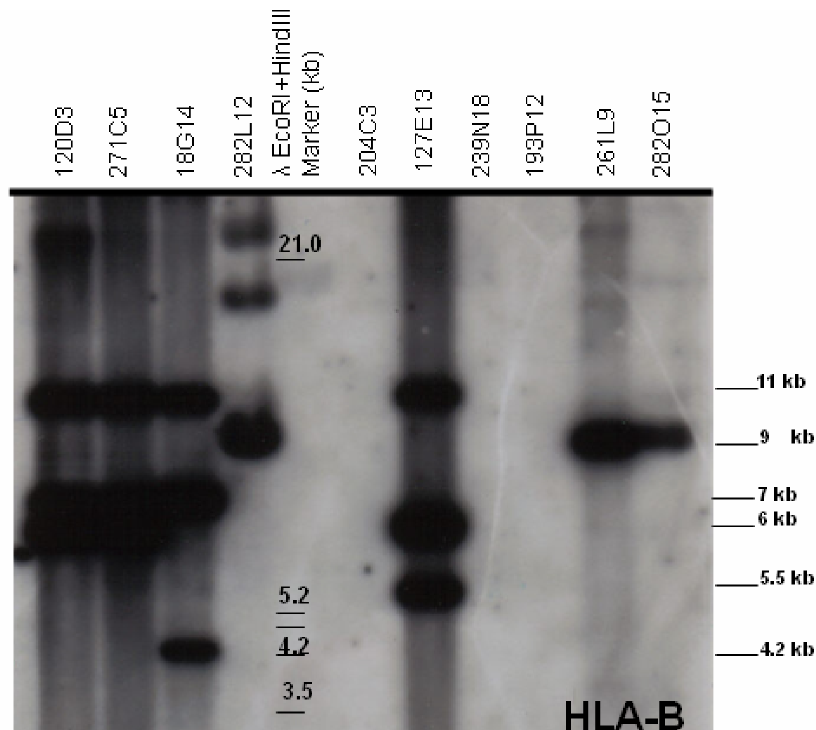


Figure 5.2.1.1.13: Autoradiograph after Southern blot hybridization of HLA-B probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme EcoRI. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) EcoRI+HindIII used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

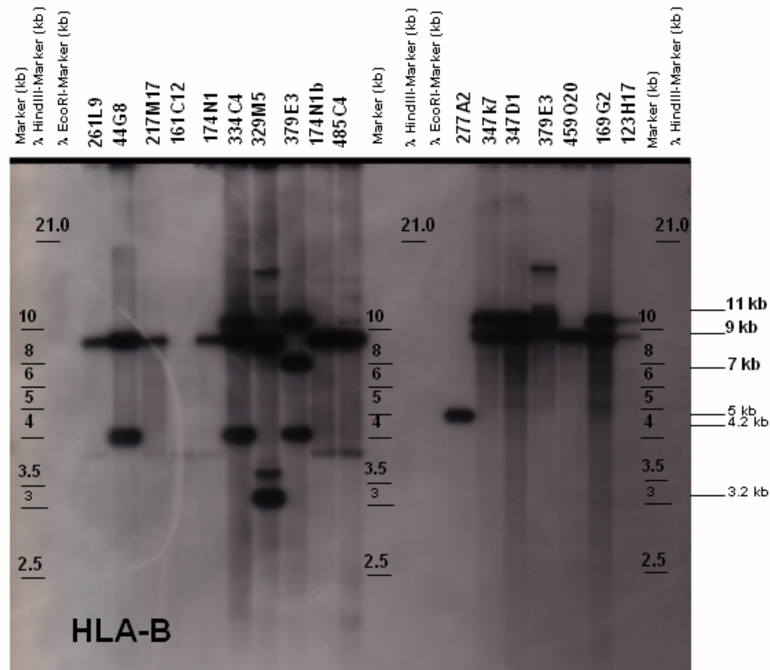


Figure 5.2.1.1.14: Autoradiograph after Southern blot hybridization of HLA-B probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) EcoRI+HindIII used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

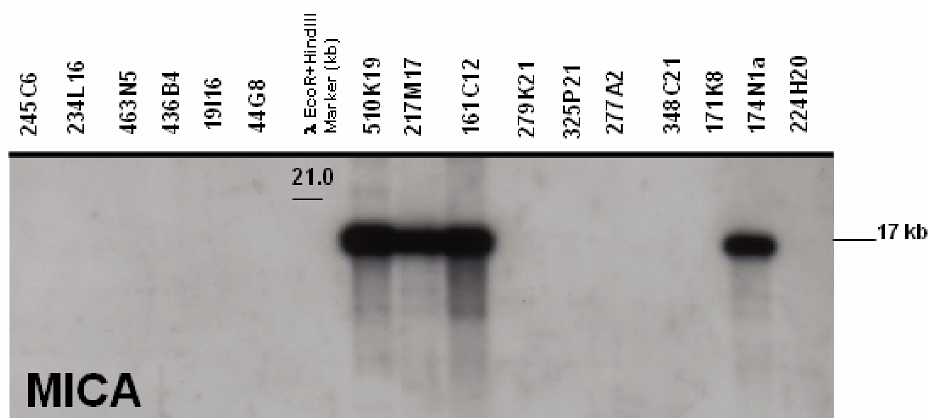


Figure 5.2.1.1.15: Autoradiograph after Southern blot hybridization of MICA probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene

probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

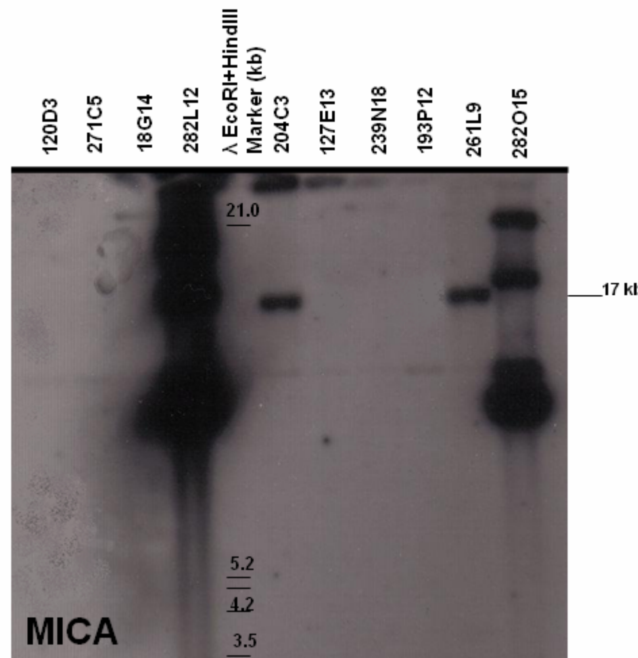


Figure 5.2.1.1.16: Autoradiograph after Southern blot hybridization of MICA probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

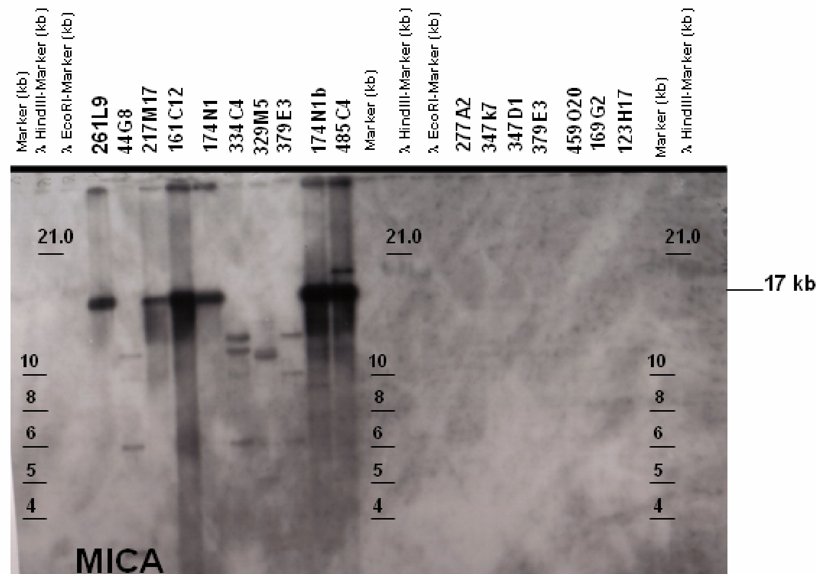


Figure 5.2.1.1.17: Autoradiograph after Southern blot hybridization of MICA probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.
MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

5.2.1.2 Evaluation of hybridization results (contig 1)

Nine of the 31 clones identified were used for did hybridize fragment analysis. 12 *EcoRI* fragments of the nine clones (Table 5.2.1.1):

- **MHC-class I-Caja-G:** This probe found six fragments of 11 kb, 10 kb, 9 kb, 7 kb, 4.2 kb and 3.5 kb. These 15 clones appeared to be overlapping: 261L9, 44G8, 174N1, 217M17, 174N1, 485C4, 334C4, 329M5, 379E3, 347K7, 347D1, 379F14, 123H17, 169G2 and 459O20 (Figures 5.2.1.1.1- 5.2.1.1.4, respectively).
- **ATP6Vg2:** This framework gene probe found two fragments of 11 kb and 3.2 kb. Three clones 204C3, 510K19 and 261L9 appeared to be overlapping (Figures 5.2.1.1.5 and 5.2.1.1.6, respectively).
- **BAT1:** This framework gene probe found four fragments, of 11 kb, 5.5 kb, 1.5 kb and 0.4 kb. Four clones appeared to be overlapping: 204C3, 510K19,

261L9, and 161C12. The clone 161C12 was cleaved into the fragments 5.5 kb, 1.5 kb and 0.4 kb only (Figures 5.2.1.1.7 and 5.2.1.1.8, respectively).

- **POU5F1:** This framework gene probe found one fragment of 10 kb. These six clones appeared to be overlapping: 347K7, 379F14, 459O20 123H17, 169G2, and 99L19 (Figure 5.2.1.1.9).
- **K18T3:** This probe found two fragments of 2.5 kb and 2.3 kb. These eight clones appeared to be overlapping: 44G8, 174N1, 217M17, 174N1, 485C4, 334C4, 329M5 and 379E3 (Figure 5.2.1.1.10).
- **MHC-class I-HLA-B:** This probe found six fragments, the same as found by the MHC-class I-probe plus one fragment of 3.2 kb. These 15 clones appeared to be overlapping: 261L9, 44G8, 174N1, 217M17, 174N1, 485C4, 334C4, 329M5, 379E3, 347K7, 347D1, 379F14, 123H17, 169G2 and 459O20 (Figures 5.2.1.1.11 - 5.2.1.1.14, respectively).
- **MICA:** This gene probe found one fragment of 17 kb. The following seven clones appeared to be overlapping: 510K19, 204C3, 261L9, 161C12, 174N1, 217M17 and 485C4 (Figures 5.2.1.1.15 - 5.2.1.1.17, respectively).
- **TCF19:** This framework gene probe was negative in hybridization with the fragments of all the BAC clones checked.

In respect to the organization of framework genes, MHC-class I, HLA-B and MICA hybridization of fragments was done to achieve a schematic representation of contig 1. Contig 1 appears to be represented by the following 19 BAC clones: 204C3, 510K19, 261L9, 161C12, 44G8, 174N1, 217M17, 174N1, 485C4, 334C4, 329M5, 379E3, 347K7, 347D1, 379F14, 123H17, 169G2, 459O20 and 99L19 (Figure 5.2.18). Screening of BAC bank was done with a MIC-A probe from rhesus monkey. Results are in agreement with those obtained in rhesus monkey and human for contig 1; however, the other contigs were not found MICA positive, in contrast to findings in the intervals GNL1 - TRIM39 (with MICC) and TCTEX 4 - MOG (with MIC D,G, and F) in rhesus monkey and human (The MHC sequencing consortium 1999; Seo *et al.*, 2001), chimpanzee (Anzai *et al.*, 2003) and *Microcebus murinus* (Neff, 2005), see MHC class I map, Figure 5.1.1.

Both MIC genes A and B may be assigned to five BAC clones of *Callithrix jacchus* as the band 17 kb separated by electrophoresis may well contain two co-migrating fragments of similar length. This interpretation shown in Figure 5.2.2 is supported by

two corresponding long *EcoRI* fragments in Hs. In the ethidiumbromide stained electropherograms, larger relative amounts of DNA can be seen for the clones K17 and 19 as compared to k 18, 25, 41 and 42. For publication, densitometry should confirm these visual results.

The interval between BAT1 and TCF19 is polymorphic; polymorphic applies to regions with genetic alterations between species compared in respect to the MHC class I region, HLA-B, HLA-C and MIC genes in several mammalian species. A double positive screening of BAC clones was included with MHC class I gene Caja-G and the following probes: ATP6V1g2, BAT1, TCF19, HLA-B, POU5F1, MICA and K18T3.

The clones 510K19, 204C3 and 261L9 were overlapping in ATP6V1g2 *Eco-RI* fragments 3.2 kb and 11 kb, this framework is localized next to BAT1. The clones 510K19, 204C3, 261L9 and 161C12 were overlapping in BAT1 with the *EcoRI* fragments 0.4 kb, 1.5 kb, 5.5 kb, and 11 kb (in clone 44G8 found only 2.5 kb); however clone 239N19 had a fragment 8 kb and clone193P12 had 12 kb. Both clones were not included in contig 1 because uncertainty of their potential overlapping with BAT1 and MHC class I genes (Figure 5.2.1.3).

From the BAC clones screening with BAT1, TCF19 and MHC class I probes, respectively, 31 were selected for characterization of their fragments by Southern blot fingerprint and sequence analysis (Table 5.2.1.2.1).

Table 5.2.1.2.1: *Eco-RI* fragments of Hs MHC class I contigs:

Data used for constructing *Callithrix jacchus* gene maps (Figure 5.4.1). T39= TRIM39, T26=TRIM26, T5= TCTEX5, T4= TCTEX4, M= MOG, P= POU5F1 and ATP= ATP6Vg1, see next page.

Clone		12	10	9	7	6.5	6	5	5.5	4.2	1.5	2.3	17	11	3.2	2.3	2.2	4	11	16	8	5	9	7	1.5	10	10	11	10	9	7	6.5	6	5	5.5	4.2	1.5	2.3	2.5	5.5	1.5	0.4	11	6	12
		HLA-B											MICA	ATP		k18			T39	T26		T5		T4	M	P		CAJA-G											BAT1						
K 1	120D3																																												
K 2	271O5																																												
K 3	18G14																																												
K 4	282L12																																												
K 5	204C3																																												
K 6	127E13																																												
K 7	239N18																																												
K 8	193P12																																												
K 9	261L9																																												
K 10	282O15																																												
K 11	245O6																																												
K 12	234L16																																												
K 13	463N5																																												
K 14	436B4																																												
K 15	19I16																																												
K 16	44G8																																												
K 17	510K19																																												
K 18	217M17																																												
K 19	161C12																																												
K 20	279K21																																												
K 21	325P21																																												
K 22	277A2																																												
K 23	348C21																																												
K 24	171I8																																												
K 25	174N1																																												
K 26	224H20																																												
K 27	347K7																																												
K 28	347D1																																												
K 29	379F14																																												
K 30	425A13																																												
K 31	459O20																																												
K 33	169G2																																												
K 34	123H17																																												
K 37	99L19																																												
K 38	334C4																																												
K 39	329M5																																												
K 40	379E3																																												
K 41	174N1																																												
K 42	485C4																																												

5.2.1.3 Construction of a contig 1 map

A gene map of contigs in *Callithrix jacchus* was attempted using these hybridization results. Unfortunately, even in contig 1 - with its biggest share of clones selected -, the amount of information - for reasons discussed - turned out to be too small for constructing a consistent contig of BAC clone inserts. In view of the redundancy (clustering) of similar genes in this region, all hybridizing nominally different class I probes alike, fragment-probe hybridization assignment as usually done for using just their information for mapping requires a sequence of shorter and more fragments all having different lengths and distinctively defined gene contents of individual genes. Fortunately, a second source of information was found that could help out in a comparative interpretation of the data: the well-documented human MHC class I DNA with its many known gene loci. The computer was made to do an *EcoRI* splitting of the Hs contig 1 DNA and to deliver a proper long list of all 123 fragment lengths in bp, furthermore, an assignment of all genes known for this section, and even the definitions of both fragment and gene loci by their 1st and last bp numbers (according to both, Shiina's and Venter's bp scales).

With this ancillary, comparative information, at least 14 genes of contig 1 as written in the headline of figure 5.2.1.3.1 could be identified which by their order and fragment length(s) did allow the arrangement of all the BAC clone fragments obtained, according to their probe specifications, into the gene map shown in Figure 5.2.1.3.1.

Consequently, a minimal contig of clones for an eventual sequencing of contig 1 could be composed of clones 510K19, 334C4, 120D3 and e.g. 347K7. Their inserts, taken together, should be covering almost the entire contig 1, i.e. 418 kb or more in Hs. A table of the *EcoRI* fragments in Hs, generally fitting those found in *Callithrix jacchus*, plus a map that illustrates the locations and lengths of both fragments and genes assigned shall be given in a subsequent publication. It suggests great similarity of *EcoRI* splitting in Hs and *Callithrix;* differences in a few cleavage sites are suggested, however, by some of the hybridization results. A map of this kind is a prediction having no absolute certainty. Of preliminary nature, it will be confirmed or corrected by the DNA sequencing that it shall support.

Contig 1 genes: ATP6V1g2 | BAT1 | MICB | HLA-X | MICA | HCGIV-01 + HLA-B | K18T3 | HLA-C | HCGIX-3.1 + HCGII-2 | HCG-27 | HCGIX-3.2 | HCGII-11 | POU5F1 | TCF19

Fragments in kb: 3.2 11 = 11 1.5 0.4 5.5 17 9? 17 2.3 + 6.5=8.8 2.3 4.2 7 11 6 2 11 10 = 10

BAC clones

K5:204C3 3.2 11 = 11 1.5 0.4 5.5 17 8 17 ?

K9:261L9 3.2 11 = 11 1.5 0.4 5.5 17 8 17 9?

K17:510K19 3.2 11 = 11 1.5 0.4 5.5 17 8 17 9 2.3

K19:161C12 1.5 0.4 5.5 17 8 17 ? 2.3

K16:44G8 2.5 17 8 17 9 2.3 ? 4.2

K25:174N1a 6.5 17 9 2.3

K18:217M17 17 9 2.3

K42:485C4 17 9 2.3 4.2

K41:174N1b 17 9 2.3 4.2

K39:329M5 9 2.3 4.2 2.3

K38:334C4 9 2.3 4.2 ? 11

K40:379 E3 2.2 7 11 4.2

K3:18G14 7 11 4.2

K6:127E13 5.5 11 6 1.5

K2:271C5 7 11 6

K1:120D3 7 11 6 2.3

K27:347 K7, K28:347 D1, K29:379 F14, K31:459 O20, K33:169 G2 and K34:123 H17 equally show 11? ? ? 11 10 = 10

K37:99 L19 9.7

Several months after finishing this map, we have learnt from an e-mail by Prof. T. Shiina the reason why comparison of these Caja fragments with those of Hs is rather inadequate: the contig 1 or HLA-B/C sections of the two primates are too different in their organization that is complicated in Caja by the existence of 10 B/C duplicons compared to 1 B and 1 C duplicon in Hs - expanding contig 1 from 418 to some 600 kb. Yet, gene order on both sides of the duplicons might still be quite alike.

Between the frameworks genes BAT1 and POU5F1-TCF19 have been localized contig 1 as part of the MHC class I region, in humans with a length of 418 kb (The MHC consortium, 1999), as H2-complexes in mouse with 300 kb, as RT1-complex of rat with 450 kb (Hurt, *et al.*, 2004) and of chimpanzee (Patr) (Fukami-Kobayashi *et al.*, 2005). These framework genes were found to be orthologoues. In humans this interval contains the MICA and MICB genes (Bahram *et al.*, 1994), HLA-B and HLA-C. The MHC class I genes homologoues to HLA-B as expressed in this New World primate are not orthologous to any of the classical MHC class I loci of the Catarrhini (*A*, *B*, or *C* loci); instead, they are most similar to the human non-classical *HLA-G* (Watkins *et al.*, 1990).
In conclusion, the contig 1 has been found between BAT1 and POU5F1 by fragments and specific hybridizations of probes Caja-G, MICA, Mamu-B and respective framework genes (including ATP6Vg1 and TCF19) display an order of genes not different from humans and other primates.
The restriction map obtained here suggests locations of these Caja genes very similar to the human ones. The number of class I genes in contig 1 of *Callithrix* presumably equals that in Hs, where 14 are known so far: three HLA genes + nine pseudogenes + two MIC genes (cf. class I map, Figure 5.2.1.3.1). Of these, the map obtained displays evidence for nine or perhaps ten. A further gene, next to POU5F1, called HCGII-11 is predicted to possibly explain the observation that six of the BAC clones rendered a fragment of approximately 10 kb that hybridizes with POU5F1, Caja-G and HLA-B (see Table 5.2.1.2.2). It shall be searched for its eventual alignment and be commented in the sequences chapter 5.5.1.

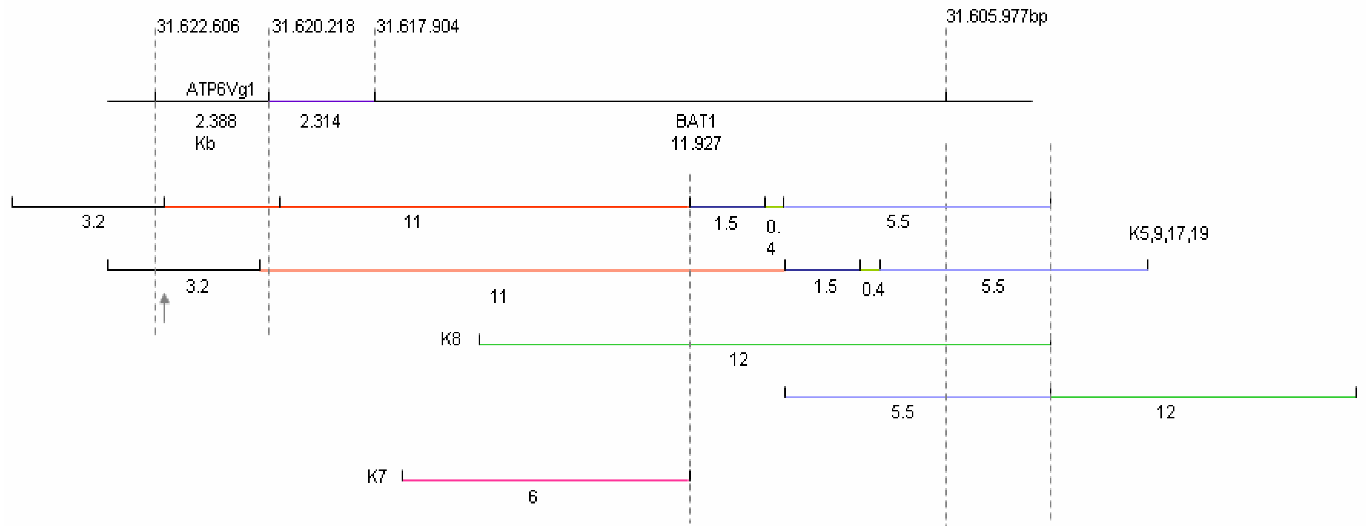


Figure 5.2.1.3.2: Alignment of BAC-clone insert fragments to ATP6Vg1 and BAT1 to the Hs bp scale by Shina ID:BA000025.

All four BAT1 postive fragments, of 11, 1.5, 0.4 and 5.5 kb from clones k5, k 9, k17 and k19 fit in as shown. Fragments from clones k8, (12kb) and K7 (6kb) are neither helpful nor against the map. Possibly these clones are *EcoRI* cleaved in a modified way.

Another cross reaction is seen in a fragment of 11kb that hybridizes with the neighbored genes ATP6Vg1 and BAT1. This fragment may bridge an intergenic gap as shown graphically in figure 5.2.1.3.2. In the endeavor of explaining the genetic data obtained for *Callithrix* during the experimental part of this work, *thinking in terms of genomic lengths* was needed. It led to the construction of a map to bp-scale for the entire MHC class I region in Hs, the primate chosen for reference (5.2.1.3.1).

5.2.2. Contig 3, the interval TRIM26 – TRIM39

5.2.2.1 Southern blot analysis of contig 3

Southern blot analysis of BAC clones for screening with framework probes of MHC class I region, Caja-G, HLA-B, TRIM26, TRIM39 and TCTEX5, included seven clones belonging to contigs as defined here in analogy to the human MHC class I contig 3 (245C6, 463N5, 436B4, 277A2, 224H20, 325P21 and 279K21). This contig presented also two class I positive *EcoR-I* fragments, of 2.5 kb and 5 kb. The 2.3 kb fragments hybridized with both probes, Caja-G and Mamu-B, the 5 kb fragment only with Caja-G. This discrepancy seems unimportant, but usually, these two probes

hybridized equally. Clone 325P21, however, rendered only the fragment of 5 kb. This indicates presence of MHC class I gene(s) in or near the interval TRIM26 – TRIM39 had run observed in rat, mouse, chimpanzee and human by Hurt *et al*., 2004; Anzai *et al*., 2003; Reymond, 2001; Wheeler *et al*., 2001; Orimo *et al*., 2000; Rahman, 1998; Yoshino *et al*., 1998; Chu *et al*., 1995 and Wei, (unpublished).

Callithrix jacchus did not show MICA positive *EcoRI* fragments as have also not been found in human (Bahram *et al*., 1994). The MIC genes expected in contigs 3 and 4 are pseudogenes. (see Figure 5.1.1).

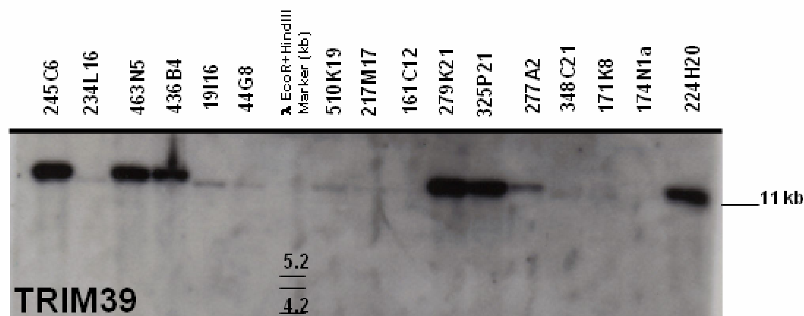


Figure 5.2.2.1.1: Autoradiograph after Southern blot hybridization of TRIM39 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus)*.

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

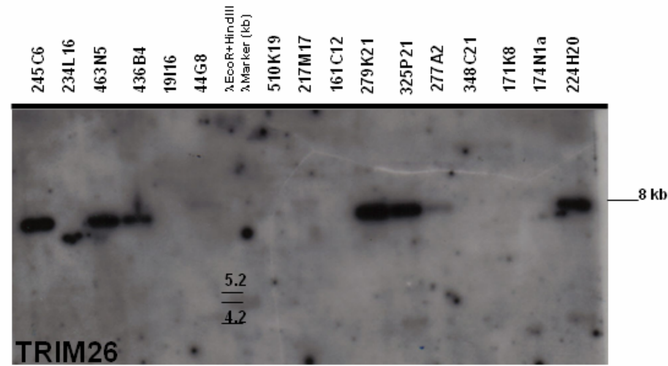


Figure 5.2.2.1.2: Autoradiograph after Southern blot hybridization of TRIM26 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

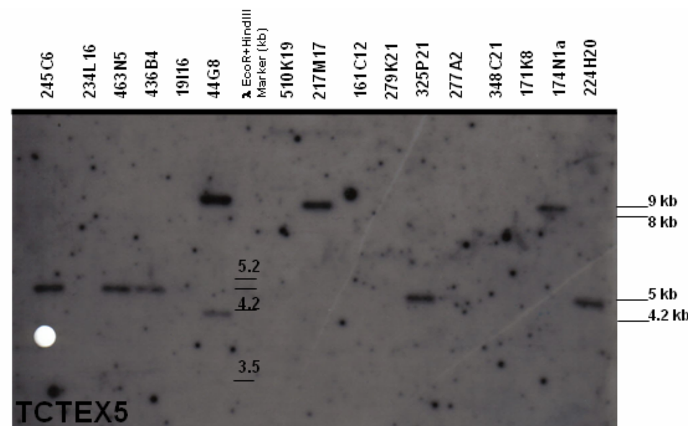


Figure 5.2.2.1.3: Autoradiograph after Southern blot hybridization of TCTEX5 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

5.2.2.2 Evaluation of hybridization results (contig 3)

From seven BAC clones of the *Callithrix jacchus* library, a total of 10 *EcoRI* (insert) fragments were obtained and used for screening of all BAC clones in the library:

- **TRIM39:** This framework gene probe detected one fragment of 11 kb. These clones appeared to be overlapping: 245C6, 463N5, 436B4, 279K21, 325P21, 277A2, 503C3. The clones 234L16, 19I16, 44G8, 510K9, 217M17, 348C21, 171K8 and 174N1 were negative (Figure 5.2.2.1.1).
- **TRIM26:** This framework gene probe detected one fragment of 8 kb. These clones appeared to be overlapping: 245C6, 463N5, 436B4, 279K21, 325P21, 277A2 and 503C3. The clones 234L16, 19I16, 348C21, 171K8, and 174N1 were negative (Figure 5.2.2.1.2).
- **TCTEX5:** This framework gene probe detected one fragment of 5 kb. These clones appeared to be overlapping: 245C6, 463N5, 436B4, 325P21, 503C3, the clone 44G8, however presented two fragments 9 kb and 4.2 kb, and the clones 217M17 and 174N1 found one fragment: 8 kb (Figure 5.2.2.1.3).
- **CAT56:** This framework gene probe was negative in hybridization for each of the clones checked.
- **MHC class I I-Caja-G**: This probe detected two fragments of 2.5 kb and of 5 kb. These clones appeared to be overlapping: 245C6, 463N5, 436B4, 325P21, 277A2 and 503C3, where as clone 279K21 had only the 4.2 kb fragment (Figure 5.2.1.1.1).
- **MHC-class I-HLA-B:** This probe found one fragment of 5 kb. These clones appeared to be overlapping: 245C6, 463N5, 436B4, 325P21, 277A2, 503C3, but the clone 279K21 was negative (Figure 5.2.1.1.12).
- **MICA:** This probe was negative in the intervals 2 and 3 for each of the seven clones tested (Figure 5.2.1.1.15).

5.2.2.3 Contribution to a contig 3 map

The intervals CAT56 - TRIM39 and TRIM39 - TRIM26 of the MHC class I region have polymorphic margins in humans. They both belong to contig 3 in the human, where contig 2 is the longest of the four contigs, however almost devoid of class I genes (see map, Figure 5.1.1). The small numbers of clones investigated in this area did not really allow constructing any one of the contigs 2 - 4, i.e. a share of stepwise overlapping clone inserts with sufficiently many, short and well discernible fragments

covering the segment between its framework borders. The intervalls expected by comparison with Hs and other mammals seem to be existing, however, as evidenced by positive screening of clones with the probes Caja-G, CAT56, TRIM39 and TRIM26 (see Figure 5.2.2.1.1 and 5.2.2.1.2). Of course, their existence in *Callithrix jacchus* is proven quite clearly by sequences done and localized as seen in map, figure 5.4.1. These sequences confirm 30 different genes and pseudogenes of the Hs contig 4, several of contig 3 and 2, many of contig 1.

The Interval CAT56 - TRIM39 contains class I gene(s) between the conserved framework genes CAT56 and TRIM39, with approximately of 500 kb in case of the RT1-complex of rat (Hurt *et al.*, 2004).

The second intervall contains class I gene(s) between the conserved framework genes TRIM39 and TRIM26 with approximately of 200 kb in the RT1-complex. Near TRIM26, towards contig 4, are TRIM15 and TRIM10 (Hurt, *et al.*, 2004). For practical reasons, the framework gene TCTEX5 located near TCTEX4 within less than 50 kb, was included in contig 4 presumed by comparison. TCTEX5 considered as end of a contig 3 detected clones with the framework genes TRIM 26 and TRIM39. So, there is probably no gap in *Callithrix* between its presumable contigs 3 and 4 (Hurt *et al.*, 2004; Jones *et al.*, 1999; Zhang *et al.*, 1998; Amadou *et al.*, 1999).

Arrangement of framework genes, Caja-G and HLA-B studied by hybridization of fragments is shown in figure 5.2.1.1.1 and table 5.2.1.1.12, the order being given by homology with other mammals.

Contig 3

	HLA-E (Caja-G) 2.3 5	TRIM 39 11	TRIM26 8	TCTEX5 5
K11:245C6	2.3 + 5	11	8	5
K13:463N5	2.3 + 5	11	8	5
K14:436B4	2.3 + 5	11	8	5
K26:224H20	2.3 + 5	11	8	5
K22:277A2	2.3 + 5	11	?	
K21:325P21	5	11	8	5?
K20:279k21		11	8	

Figure 5.2.2.3: The overlapping BAC clone fragments drawn according to the order of framework genes known for contig 3 (intervals CAT56 - TRIM39 and TRIM39 - TRIM26) in other mammals (mouse, rat, chimpanzee and human). Contig 3 represented in this figure is a part within the MHC class I region as shown in the maps, Figures 5.1.1 and 5.4.1.

5.2.3 Contig 4, the intervall TCTEX4 - MOG

5.2.3.1 Southern blot analysis of contig 4

In Southern blot analysis, the clones 348C21 and 171K8 presented four *EcoRI* fragments positive with Caja-G and HLA-B. Therefore, *Callithrix* also may have MHC class I gene(s) between TCTEX4 and MOG is well known for human and chimpanzee (Shiina *et al.*, 2003).

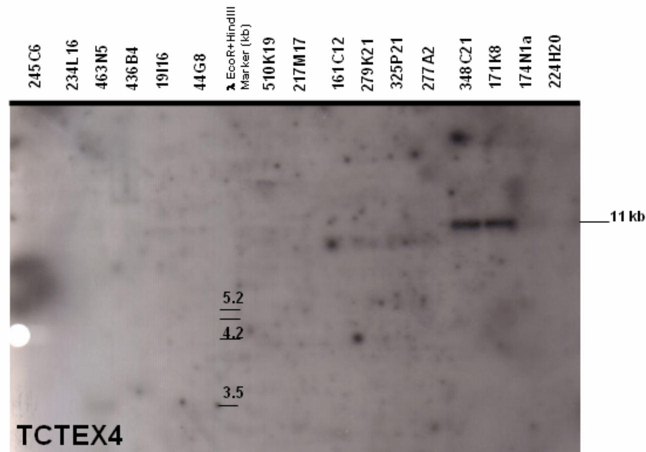


Figure 5.2.3.1.1: Autoradiograph after Southern blot hybridization of TCTEX4 framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I-positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI*+*HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

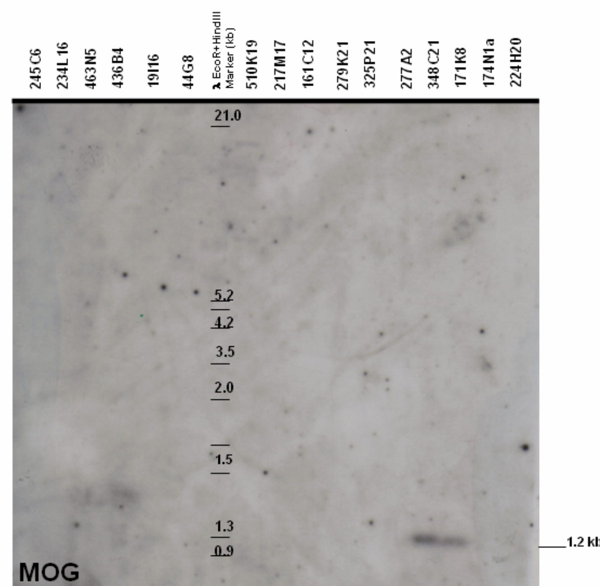


Figure 5.2.3.1.2: Autoradiograph after Southern blot hybridization of MOG framework gene probe with BAC-clone fragments (CHORI-259, *Callithrix jacchus*).

MHC-class I positive BAC-clones were digested with restriction enzyme *EcoRI*. After gel electrophoresis and fingerprinting, their fragments were hybridized with said gene probe. DNA sample Lambda (λ) *EcoRI+HindIII* used as ladder marker. kb values on the right margin calculated relative to marker DNA fragments which were seen by means of ethidium bromide fluorescence.

5.2.3.2 Evaluation of hybridization results (contig 4)

Here is shown an identical hybridization with of contig 4 and class I probes in these two clones suggesting them to contain the interval TCTEX4 - MOG of common marmoset.

- **MOG:** This framework gene probe detected one fragment of 1.2 kb. These clones appeared to be overlapping: 348C21 and 171K8 (Figure 5.2.3.1.1).
- **TCTEX4:** This framework gene probe detected two fragments, both of 11 kb. These clones appeared to be overlapping: 348C21 and 171K8 (Figure 5.2.3.1.2).
- **MHC class I-Caja-G:** This framework gene probe detected two fragments, of 12 kb and 6 kb. The same two clones matching also probes MOG and TCTEX4 appeared to be overlapping: 348C21 and 171K8 (Figure 5.2.1.1.1).
- **MHC class I HLA-B:** This probe detected two fragments of 11 kb and 12 kb in clone 171K8, and only one of 12 kb in clone 348C21 (Figure 5.2.1.1.12).
- **MICA**: This probe was negative for this contig (Figure 5.2.1.1.15).

5.2.3.3 Contribution to a contig 4 map

Contig 4 contains between the conserved frameworks genes TCTEX4 and MOG many MHC class I genes and pseudogenes intercalated with many other genes (see MHC class I maps to bp-scale, Figures 5.1.1 and 5.3.1.1). The interval has remarkably different lengths in the mammals investigated, approximately 50 kb in the rat RT1-complex (Hurt *et al.*, 2004) 400 kb in contig 4 of Hs. This interval is conformed by two clones, however the sequence analysis strongly and extensively supports that this contig exist in *Callithrix jacchus.*

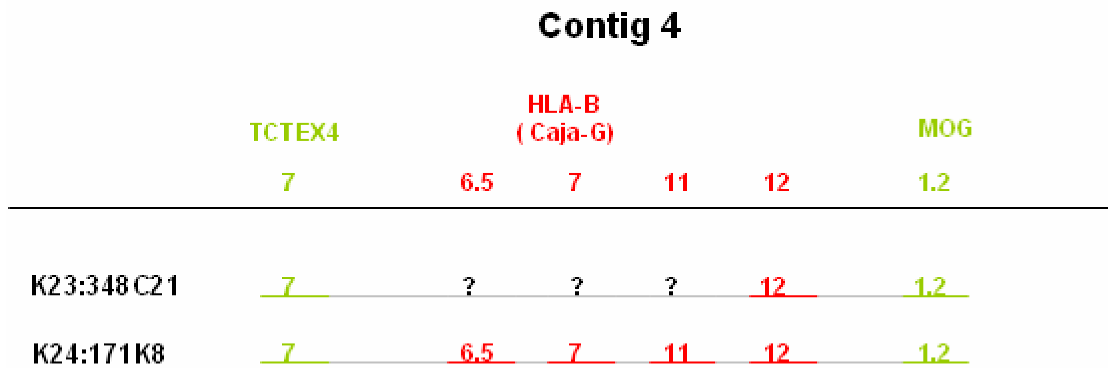


Figure 5.2.3.3.1: Two overlapping BAC clone inserts with framework genes of contig 4.

The order of the class I containing fragments 6.5, 7 and 11 kb remains an enigma. Nevertheless, it may be concluded, that these two clones must contain at least four HLA-B and Caja-G positive class I genes or pseudogenes, however, not necessarily between TCTEX4 and MOG.

Again, sequences established lateron, localized and as possible gene assigned in comparison to the Hs genome strongly suggest many homologoues Caja genes to exist in the interval of contig 4.

Complete evidence for class I gene maps based on gene assignment by comparison shall be coming from a complete sequencing that has been started already by Prof. T. Shiina using the clones selected by the screening described in chapter 5.1 of this thesis.

5.3 PCR and sequence analysis of BAC clone inserts

5.3.1 Polymerase chain reaction (PCR) analysis

Locus-specific PCR was used to characterize the clones by primer specific fragments which were sequenced partially or totally (see next chapter, 5.4). Oligonucleotide primers to POU5F1, MOG, K18, HLB 32, 29, 39, Caja-G and Caja-E were designed with primer-3 on the Web (http//www.genome.wi.mi.edu/cgi-bin/primer/primer3.cgi).

- For **POU5F1** were found 7 positives clones (347K7, 347D1, 379F14, 459O20, 169G2, 123H17 and 99L19) containing a fragment of 450 bp (see Figure 5.3.1).
- For **K18T3** was found one clone (217M19) containing a fragment of 150 bp.

- For **Caja-G** were found 5 positives clones (120D3, 271C5, 127E13, 44G8 and 171K8) containing a fragment of 650 bp.
- For **Caja-E** were found eight positive clones (271C5, 18G14, 204C3, 127E13, 325P21, 277A2, 348C21 and 171K8) containing a fragment of 500 bp.
- For **HLAB-32** were found 13 positive clones (120D3, 271C5, 127E13, 463N5, 325P21, 277A2, 224H20, 347K7, 347D1, 379F14, 459O20, 169G2 and 123H17) containing a fragment 700 of bp.
- For **HLAB-29** were found 12 positive clones (120D3, 271C5, 18G14, 261L9, 463N5, 161C12, 347K7, 347D1, 379F14, 459O20, 169G2 and 123H17) containing a fragment of 800 bp.
- For **HLAB-39** were found three positive clones (261L9, 325P21, 277A2) with a fragment 850 bp.
- For **Klexon 2 MHC consensus** were found two positive clones (120D3 and 127E13) containing a fragment 700 of bp.
- **MOG** was negative in all clones (see Figure 5.3.1.1).

Acknowledgement: The primers HLA-B 29, 32 and 39 were kindly given to the author by Cornelia Rosner; the primer ALU was a gift of Martin Osther and a MICA DNA sample by Anne Averdam, all working in the DPZ, Göttingen.

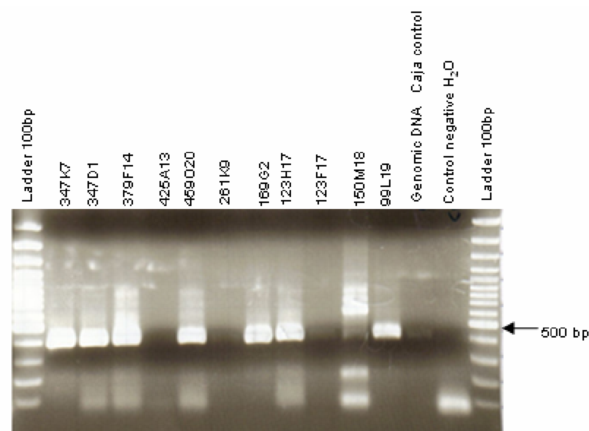


Figure 5.3.1.1 PCR with primer POU5F1.

Image of a 1 % agarose gel stained with ethidium bromide 20 mg/ml. Only clones positive with primer POU5F1 show fragments. The control with nuclear DNA from *Callithrix jacchus* (sample on the right, next to ladder).

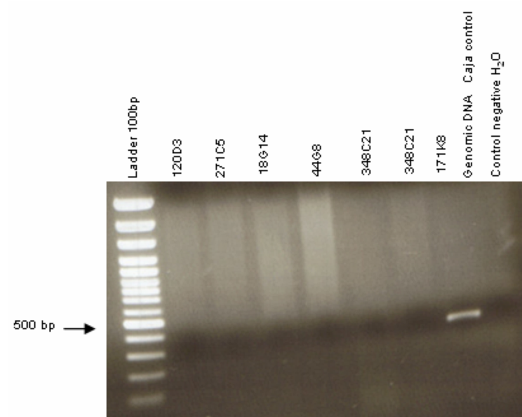


Figure 5.3.1.2 PCR with primer MOG.

Fluorescence photograph of a 1 % agarose gel stained by ethidium bromide 20 mg/ml. Only the nuclear DNA from *Callithrix jacchus* was positive with MOG.

Table 5.3.1.1: Genetic typification of BAC clone inserts by PCR. The table contains all clones tested with different primers from framework, Caja-G and HLA genes.

Clone	Caja G 650 bp	Caja-E 500 bp	Klexon2 consenso	TRIM39 1500 bp	TCF19 350 bp	TRIM10 400 bp 900 bp	MOG 350bp	HLA-B-32 700 bp	HLA-B-29 800 bp	HLA-B-39 850 bp	POU5F1 450 bp	K18T3 150 bp
120E3	+	-	+	-	-	+	-	+	+	-	-	-
271C5	+	+	-	-	-	-	-	+	+	-	-	-
18G14	-	+	-	-	+	-	-	-	+	-	-	-
282L12	-	-	-	-	+	-	-	-	-	-	-	-
204C3	-	+	-	-	-	-	-	-	-	-	-	-
127E13	+	+	+	-	-	+	-	+		-	-	-
239N18	-	-	-	-	-	-	-	-	-	-	-	-
193P12	-	-	-	-	-	+	-	-	-	-	-	-
261L9	-	-	-	-	-	+	-	-	+	+	-	-
282015	-	-	+	-	-	-	-	-	-	-	-	-
245C6	-	-	+	-	-	+	-	-	-	-	-	-
234L16	-	-	+	-	-	+	-	-	-	-	-	-
463N5	-	-	+	-	-	+	-	+	+	-	-	-
436B4	-	-	+	-	-	+	-	-	-	-	-	-
19I16	-	-	+	-	-	+	-	-	-	-	-	-
44G8	+	-	+	+	-	-	-	-	-	-	-	-
510K19	-	-	-	-	-	-	-	-	-	-	-	-
217M17	-	-	-	-	-	+	-	-	-	-	-	+
161C12	-	-	-	-	-	-	-	-	+	-	-	-
279K21	-	-	-	+	-	+	-	-	-	-	-	-
325P21	-	+	-	+	-	-	-	+	-	+	-	-
277A2	-	+	-	-	-	-	-	+	-	+	-	-
348C21	-	+	-	-	-	+	-	-	-	-	-	-
171K8	-	+	-	-	-	-	-	-	-	-	-	-
174N1	-	-	-	-	-	-	-	-	-	-	-	-
224H20	-	-	-	+	-	+	-	+	-	-	-	-
347K7	-	-	-	+	+	-	-	+	+	-	+	-
347D1	-	-	-	-	+	-	-	+	+	-	+	-
379F14	-	-	-	-	+	-	-	+	+	-	+	-
425A13	-	-	-	-	-	-	-	-	-	-	+	-
459O20	-	-	-	+	+	-	-	+	+	-	+	-
169G2	-	-	-	-	-	-	-	+	+	-	+	-
123H17	-	-	-	+	+	-	-	+	+	-	+	-
99L19	-	-	-	-	+	-	-	-	-	-	+	-
334C4	-	-	-	+	-	-	-	-	-	-	-	-
329M5	-	-	-	-	-	-	-	-	-	-	-	-
379M5	-	-	-	-	-	-	-	-	-	-	-	-
174N1	-	-	-	-	-	-	-	-	-	-	-	-
485C4	-	-	-	-	-	-	-	-	-	-	-	-
Caja DNAg	+	+	+	+	+	+	+	+	+	+	+	+

5.3.2 Approximative alignment of sequences to homologoues DNA of Hs and other mammals by BLAST

In an attempt to facilitate insight into unknown basic functions to be expected in he highly conserved framework, by an example from the MHC even from years to many (Amadou *et al*., 1995) ways for the analysis of their structure, i.e. the DNA sequence, were investigated. The fragments from BAC clones obtained by inverse PCR were excised from the gel, and extracted by NaI. They were sequenced e.g. by means of the primers T7.29, ptarbac2, ptarbac3, ptarbac4, TCF19 and HLA-B29, respectively.
Several clones could be sequenced partially. The clone 161C12 has been sequenced completely by Dr. Takashi Shiina at Tokyo University School of Medicine in Tokyo, Japan.
After automated sequencing of fragments recovered from electropherograms, sequencing results were inspected in order to identify errors and ambiguities. These corrected sequences were transferred to BIOEDIT for correction of mistakes, and furthermore to BLASTVECTOR for removal of vector sequences. Genebank database was then searched using the BLASTIN algorithm in order to identify significant nucleotide matches. A value for similarity or homology of *Callithrix jacchus* sample sequences with database sequences of several species was calculated referring to the BAC-banks of these species. Significant matches were arbitrarily defined, first as those which showed sequence identity (homology) greater than 80 % over a segment of DNA of not less than 30 bp in length, and which gave a BLASTIN high scoring point of 300 or above at a significant p value. To perform sample sequencing, by the use of pairs of primers, sequencing in two directions of reading was performed.

Table 5.3.2.1: Gene assignment of MHC class I sequences by comparison with Hs and other mammals using BLAST giving % identity for a best aligning part of a sequence.

ψ designates pseudogenes otherwise abbreviated ps.

Clone	Contig	Primer	Sequence (bp)	Species: in or near gene; homology %
204C3	1	ptarbac3	438	Hs: pseudogene 88 % Mamu: MIC2 89 %
		T7.29	619	Hs: BAT2; NCR3 87 % Mamu: BAT2; NCR3 88 %
510K19	1	ptarbac3	211	Patr:Patr-C; 88 % Hs: ψ before OTF3; 87 % Mamu: Mamu-B18; 86 %
217M17	1	ptarbac3	208	Hs: MICA; 88 % Patr: MIC; 87 % Mamu: MIC; 89 %
		T7.29	383	Hs: ψ HCGIIX4; 90 % Patr: MICA-B; 90 % Mamu: MIC3; 90 %
347K7	1	ptarbac3	515	Hs: SEEK1 exon 1; 84 % Patr: SEEK1 exon 1; 85 % Mamu: C6orf15; 85 %
		ptarbac4	363	Hs: HCGIX4; 90 % Patr: MICB; 90 % Mamu: MIC 3; 90 %
347D1	1	ptarbac3	518	Hs: SEEK1 exon 1; 86 % Patr: SEEK1 exon 1; 86 % Mamu: C6orf15 87%
		ptarbac4	534	Hs: SEEK1 exon 1; 84 % Patr: SEEK1; 84 % Mamu: C6orf15; 85 %
		TCF19	2314	Hs: TCF19 exon 2; 90 % Patr: TCF19; 90 % Mamu: TCF19; 90 % Rano: TCF19; 89 %
459O20	1	ptarbac3	624	Hs: 3.8-1-4; 85 % Patr: ψ 3.8-1-4; 85 % Mamu: 3.8-1-4; 88 %
		ptarbac4	511	Hs: 3.8-1-4; 87 % Patr: ψ 3.8-1-4; 86 % Mamu: ψ3.8-1-4; 88 %
		T7.29	121	Hs: SEEK1; 89 % Patr: SEEK1; 89 % Mamu: SEEK1; 88 %
		TCF19	258	Patr: TCF19 exon 2; 92 % Mamu: TCF19; 91 % Rano: TCF19; 90 %
169G2	1	T7.29	207	Hs: HLA-B; 94 % Patr: Patr-B exon 8; 95 % Mamu: Mamu-B13; 94 %
		ptarbac3	361	Hs: class I; 95 % Patr: class I; 95 % Mamu: class I; 94 %
		ptarbac4	328	Hs: TCF19; 89 % Hs: SEEK1; 95 % Patr: SEEK1; 95 %
123H7	1	T7.29	527	Hs: TCF19; 89 % Patr: TCF19; 89 % Mamu: TCF19; 89 %
		TCF19	567	Hs: TCF19; 89 % Mamu: TCF19; 89 % Patr: TCF19; 89 %
99L19	1	TCF19	377	Hs: TCF 19 exon 2; 89 % Patr: TCF19 exon 1 and 2; 92 % Mamu: TCF19; 89 % Rano: TCF19; 89 %
161C12	1	Unknown primer of Dr. Shiina	3000*	Hs: BAT1 exon 5 ; 90 % Patr: BAT1 exon 4 ; 88 % Mamu: BAT1; 88 % Susc: BAT1 exon 4; 90 %

				Bota: BAT1; < 70 % Mumu: BAT1a; < 70 % Rano: BAT1a; < 70 % Aime: BAT1a; < 70 %
		Unknown primer of Dr. Shiina	2994*	Patr: pseudogene HCGIV-2; 88 % Atbe: Atbe-Be3; 91 %
261L9	1	HLA-B 29	207	Hs: HLA-B; 73 % Patr: Patr-B; 74 %
Clone	**Contig**	**Primer**	**bp**	**Species: in or near gene; homology %**
245C6	3	T7.29	280	Hs: TRIM15 exon 2; 92 % Patr: TRIM15 exon 5; 94 % Mamu: TRIM15; 93 %
463N5	2 or 3	HLA-B	358	Hs: HLA-B; 91 % Patr: Patr-B; 92 % Mamu: Mamu-B13; 88 %
		ptarbac3	365	Hs: HLA-E; 86 % Patr: Patr E; 86 % Mamu: Mamu-E; 85 %
436B4	2 or 3	T7.29	247	Hs: HLA-B; 94 % Hs: HLA-B27; 94 % Patr: Patr-B; 95 % Mamu: Mamu-B10; 96 %
325P21	2 or 3	T7.29	323	Hs: TRIM10 exon 6; 91 % Mamu: TRIM10; 88 % Susc: TRIM10; 92 %
		ptarbac3	171	Hs: HLA-G exon 3; 92 % Patr: Patr-G; 94 %
279K21	2 or 3	T7.29	482	Caja: Caja-E; 97 %
224H20	3	T7.29	159	Hs: TRIM10 exon 7; 93 % Patr: TRIM10-TRIM15;94 % Mamu: TRIM10; 93 %
Clone	**Contig**	**Primer**	**bp**	**Species: in or near gene; homology %**
348C21	4	T7.29	485	Hs: MOG Intron 3; 90 % Susc: MOG; 93 %
171K8	4	T7.29	637	Hs: MOG; 88 %
		ptarbac2	567	Hs: MOG; 88 %
		Caja-G	362	Hs: HLA-C exons 2-3; 90 % Hs: HLA-C exons 2-3; 89 % Hs: HLA-C exons 1-8; 89 % Hs: HLA-C; 89 % Saoe: Saoe-G G*12; 91 %
120D3	Outside contigs 1-4	HLA-B 32 b	235	Hs: HLA-B; 89 % Hs: HLA-BC; 89 %
		HLA-B 29 b	235	Hs: HLA-B27; 93 % Hs: HLA-B; 92 %
		HLA-B 32a	431	Hs: HLA-Bw57; 89 % Hs: HLA-C; 69 %
		Caja-G 2273	666	Hs: HLA-C; 85 % Hs: HLA-C exons 1, 2 and 3; 85 % Saoe: ψ Saoe-G*19; 93 %. Saoe: ψ Saoe-G*12; 92 %
		Caja-E	363	Hs: HLA-C; 88 % Hs: HLA-Cw-0202 allele, intron 2;88 % Saoe: ψ Saoe-G*19; 95 %
		MIC	142	Hs: -DR class I ψ; 87 % Hs: HLA-B*48 exons 1; 86 % Hs: HLA-B gene exon 4; 86 % Sasc:Sasc-G; 86 %
		Klex2	84	Saoe: ψ Saoe-G*20; 93 % Saoe: ψ Saoe-G*19; 91 % Saoe: ψ Saoe-G*12; 91 % Caja: Caja-G*03; 93 % Caja: Caja-G*02; 93 %
127E13		HLA-B 29 a	280	Hs: HLA-B27; 93 % Hs: HLA-B; 92 % Hs: HLA-GHKAJ; 93 %
		HLA-B 29 b	175	Hs: HLA-B27; 89 % Hs: HLA-B27; 89 %
193P12		T7.29	693	Hs: clone RP11-512E16; 86 %
		T7.29	511	Hs: clone RP11-512E16; 86 %
282L12		HLA-B 29	272	Hs: MICA; 89 % Patr: TNF∝; 88 % Mamu: MIC2; 89 %

18G14		ptarbac3	279	Vector ptarbac 2.1; 96 %
19I16		T7.29	229	Hs: RP 11-765; 96 %
174N2		T7.29	202	Hs: CCNE1; 86 %
271C5		T7.29	170	Vector ptarbac 2.1; 92 %
		T7.29	232	Mamu:Mamu-L; 91 %
		HLAB-29	575	Hs: HLA-B27; 92 % Patr: Patr-B exon 8; 98 % Mamu: Mamu-B1; 94 %
			27153	

* Unpublished results of Dr. T. Shiina communicated to Dr. L.Walter

5.3.3 Sequences of contig 1

In the MHC class I contig number 1, 21 insert sequences were obtained from 11 BAC clones. These may be summarized as follows: 19 sequences were achieved partially, i.e. 14 by the primers T7.21, ptarbac3, ptarbac2 and ptarbac4, belonging to the ptarbac 2.1 vector, four sequences by a TCF19 specific primer and one sequence by the primer HLA-B 29. Two sequences of clone 161C12 could be established (Table 5.3.3.2). All DNA sequences established (55) are presented in the annex. In this chapter they are briefly described. For an approximate localization they have been first aligned by BLAST program reporting a localization at or near a gene and the percentage identity found (for a stretch of best alignment - may be only some 50 - 150 bp). These data have been colleted in table 5.3.3.2.

Table 5.3.3.1: References to species compared

Species	Abbreviation	References
Homo sapiens	Hs	Shiina *et al.*, 2001 Iris *et al.*, 1993 Hirakawa *et al.*, 2001 Beck *et al.*, 2000 Rhode *et al.*, 1999 Mizuki *et al.*, 1997 Robinson *et al.*, 2000 Roth *et al.*, 1995; Gonzales *et al.*, 1996.
Pan troglodites	Patr	Anzai *et al.*, 2003.
Maccaca mulata	Mamu	Kulski *et al.*, 2004; Fukami-Kobayashi *et al.*, 2005
Mus musculus	Mumu	Strausberg *et al.*, 2002
Rattus norvegicus	Rano	Hurt *et al.*, 2004 Ku *et al.*, 1991 Reymond *et al.*, 2001
Ateles belzebuth	Atbe	Sawai *et al.*, 2004 Daza-Vamenta *et al.*, unpublished
Danio reiro	Dare	Strausberg *et al.*, 2002
Sus scrofa	Susc	Peelman *et al.*, 1995 Ando *et al.*, 2005
Ailuropoda melanoleuca	Aime	Liu unpublished
Bos tauros	Bota	Strausberg *et al.*, 2002

Table 5.3.3.2: Sequenced clones from contig 1

Primers	Clones					
parbac3	347K7	459O20	217M17	347D1	204C3	510K19
ptarbac4	347K7	459O20	-	347D1	-	169G2
T7.29	123H7	459O20	217M17	-	204C3	261L9
TCF19	123H7	459O20	-	347D1	99L19	-
HLA-B29	-	-	-	-	-	261L9
Dr. Shiina	-	-	-	-	-	161C12

5.3.3.1 Clone 204C3

One partial insert sequence (1) of clone 204C3 obtained has a length of 438 bp. It was sequenced using primer ptarbac3. With the vector sequence tested it contains no significant similarities. According to Genebank database search, the sequence aligns in *Homo sapiens* (Hs), *Macaca mulatta* (Mamu) and *Pan troglodytes* (Patr) near the MIC gene. The respective homologies are 88 % in Hs, 89 % in Mamu, and 87 % in Patr.

A second sequence (2) obtained by primer T7.29 has a length of 619 bp. It contains no significant similarities with the vector sequence. It aligns in Hs near the BAT2 region at a homology of 87 %, in Mamu, however, near the Natural Cytotoxicity Triggering Receptor 3 (NCR3) at 88 % homology.

5.3.3.2 Clone 510K19

One partial insert sequence (3) obtained of clone 510K19 has a length of 211 bp. It was sequenced by means of primer ptarbac3, the vector sequence has a strong match at one end, the positions 157 - 207, and Hs, Mamu and Patr the sequence aligns with regions near the MIC gene. The respective homologies are: 88 % in Hs, 87 % in Patr, and. 89 % in Mamu.

5.3.3.3 Clone 217M17

One partial insert sequence (4) of clone 217M7 obtained has a length of 208 bp. It was sequenced by means of primer ptarbac3. At one end, the positions 157 - 207 It has a strong match with the vector sequence. The sequence aligns in both Hs and Patr to regions near the HLA complex group 2 pseudogene (HCGII-2). The respective homologies are 86 % in Hs and 88 % in Patr.

A second sequence (5) obtained using primer T7.29 has a length of 383 bp. It contains no significant similarities with the vector sequence. It aligns in Hs near the

HCGIX-4, in Patr near the MICA-B, and in Mamu near the MIC3, its homology appears 90 % in Hs, 90% in Patr and 90 % in Mamu, respectively.

5.3.3.4 Clone 347K7

One partial insert sequence (6) clone 347K7 obtained has a length of 515 bp. It was sequenced using primer ptarbac3. It has a strong match with the vector sequence at one end, the positions 464 - 515. This sequence aligns both in Hs and Patr near SEEK1 exon 1, in Mamu, however, near C6orf15. The respective homologies are 84 % in Hs, 85 % in Patr and 85 % in Mamu.

A second sequence (7) obtained by means of primer ptarbac4 has a length of 363 bp. It shows no significant similarities with the vector sequence. It aligns, in Hs near the HCGIX-4, in Patr near the MICA-B and in Mamu near the MIC3. Homologies are of 90 % in Hs, 90 % in Patr and 90 % in Mamu.

5.3.3.5 Clone 347D1

One partial insert sequence (8) of clone 347D1 obtained has a length of 518 bp. It was sequenced by means of primer ptarbac3. This sequence has a weak match with the vector means (weak here 1 random match expected in 40 queries of a length of 350 kb), namely, a terminal match with score 16 to 18, an internal match with score 23 to 24 at one end, the positions 468 - 491. The sequence aligns in Hs near the SEEK1 exon 1, in Patr near the SEEK1 and in Mamu near the C6orf15. The respective homologies are 86 % in Hs, 86 % in Patr and 87 % in Mamu.

A second sequence (9) obtained by primer ptarbac4 has a length of 534 bp. It has has a strong match one end, the vector sequence in its positions 505 - 531, with a weak match in its positions 7 - 30. The sequence aligns in Hs near the SEEK1 exon 1, in Patr near the SEEK1 exon 1 and in Mamu near the C6orf15. The respective homologies are 84 % in Hs, 84 % in Patr and 85 % in Mamu.

A third sequence (10) obtained using primer TCF19 from humans has a length of 2314 bp. The sequence aligns in Hs to the TCF19 gene exon 2, and in Patr, Mamu and *Rattus norvegicus* (Rano) also somewhere to the TCF 19 gene. The respective homologies are of 90 % in Hs, 90 % in Patr, 90 % in Mamu and 89 % in Rano.

5.3.3.6 Clone 459O20

One partial insert sequence (11) of clone 459O20 obtained has a length of 624 bp. It was sequenced by means of primer ptarbac3. The sequence aligns in Hs pseudogene 3.8-1.1, in Patr pseudogene 3.8-1.2 and in Mamu near the pseudogene 3.8-1-09. The respective homologies are 85 % in Hs 88 %, 85 % in Patr and in Mamu.

A second sequence (12) obtained with primer ptarbac4 has a length of 511 bp. This sequence has moderate match the vector in positions 1 - 29. The sequence aligns in Hs to pseudogene 3.8-1-04, in Patr to pseudogene 3.8-1-04 and in Mamu to pseudogene 3.8-1-04. The respective homologies are 87 % in Hs 86 %, in Patr and 88 % in Mamu.

A third sequence (15) obtained using primer TCF19 has a length of 258 bp. This sequence aligns to the TCF19 gene in Hs, in Mamu and in Rano; in Patr, however, to the TCF19 gene exon 2. The respective homologies are 92 % in Hs, 92 % in Patr, 91 % in Mamu and 90 % in Rano.

5.3.3.7 Clone 169G2

The partial insert sequence (16) of clone 169G2 obtained has a length of 207 bp. It was sequenced using primer T7.29. It contains no significant similarities with the vector sequence. The sequence aligns in Hs at HLA-B, in Patr near the Patr-B exon 8 and in Mamu near the Mamu-B13. Its respective homologies are 94 % in Hs, 95 % in Patr and 94 % in Mamu.

5.3.3.8 Clone 123H7

One partial insert sequence (19) of clone 123H7 obtained using primer T7.29 has a length of 527 bp. It contains no significant similarities with the vector sequence. The sequence aligns near the TCF19 gene in Hs and Patr, in Mamu, however, in the TCF19 gene*). Its respective homologies are 89 % in Hs, 89 % in Patr and 89 % in Mamu. *) Such alledged differences should be checked by alignment via a program like Clustal. BLAST gives rather a first approximation.

A second sequence (20) obtained using primer TCF19 has a length of 567 bp. This sequence aligns to the TCF19 gene in Hs, Mamu and Rano, yet in Patr to the TCF19 gene exon 2. The respective homologies are 91 % in Hs, 90 % in Patr, 90 % in Mamu and 86 % in Rano.

5.3.3.9 Clone 99L19

A partial insert sequence (21) obtained of 99L19 using primer TCF19 has a length of 377 bp. This sequence aligns to the TCF19 gene in Hs, Mamu, and Rano. It aligns in Patr to the TCF19 gene exon 1 and 2. The respective homologies are 89 % in Hs, 92 % in Patr, 89 % in Mamu and 89 % in Rano.

5.3.3.10 Clone 161C12

One sequence obtained of clone 161C12 has a length of 2994 bp. It contains no significant similarities with the vector sequence. The sequence aligns in *Ateles belzebuth* (Atbe) with Atbe-B2 and Atbe-B3, in Hs and Patr, however, near the HCG IV-2. The respective homologies are 88 % in Hs and Patr, 91% in Atbe and 89 % in Mamu.

A second sequence obtained has a length of 3000 bp. It contains no significant similarities with the vector sequence. The sequence aligns in Hs with the BAT1 exon 5; with the BAT1 exon 4 in Patr and in *Sus scrofa* (Sucr); with the BAT1 in Mamu, *Ailuropoda melanoleuca* (Aime) and *Bos tauros* (Bota); with the BAT1a in Rano, *Danio reiro* (Dare) and *Mus Musculus* (Mumu). The respective homologies are 90 % in Hs, 90 % in Patr, 89 % in Mamu, 90 % in Sucr, in Bota, Mumu, Rano and Aime data for % are not available.

5.3.3.11 Clone 261L9

The partial insert sequence (22) of clone 261L9 obtained using primer HLA-B.29 has a length of 207 bp. It contains no significant similarities with the vector sequence. The sequence aligns in Hs near the HLA-B exon 7, in Patr near the TRIM39 gene and in Mamu the Mamu-B13 gene. The respective homologies are 94 % in Hs, 95 % in Patr and 94 % in Mamu.

5.3.4 Sequences of contig 3

In the contigs 2 and 3, 8 partiality sequences were obtained. These were sequenced by means of primers belonging to the ptarbac 2.1 vector. In summary: five sequences of the clones 245C6, 463N5, 436B4, 325P21 and 224H20 were obtained by means of the primer T7.29 and 3 sequences of the clones 279K21, 463N5 and 325P21 were achieved using the primer ptarbac3.

Sequence analysis of clones belonging to contig 3

The sequence analysis to clone is interest because indicates to find framework genes, to confirm the presence of MHC class I gene and HLA-B.

The clones 325P21 and 224H20 showed the presence of the framework gene TRIM10. The clone 245C6 presented the framework gene TRIM15.

The organization of this frameworks is very similar to that described for mouse, rat and humans (Fukami-Kobayashi *et al.*, 2005; Hurt *et al.*, 2004; Shiina *et al.*, 2004; Kulski *et al.*, 2004).

The clones 224H20, 463N5 and 436B4 showed the presence of the HLA-B gene that is consistent with the Southern blot analysis, the primer T7.29 indicates the position of this HLA-B genes are in direction of CAT56, however is not easy to place in the map.

The clone 224H20 showed the presence of HLA-G, this primer T7.29 indicates the position of this HLA-G gene is in direction TRIM39, that is interest to clarify because this gene is present in the interval TCTEX4-MOG in humans (Pham-Dinh *et al.*, 1993).

5.3.4.1 Clone 245C6

The partial insert sequence (29) of clone 245C6 obtained has a length of 280 bp. It was sequenced by means of primer T7.29, and showed no significant similarities with the vector sequence. The sequence aligns with the TRIM15 exon 2 in Hs and in Patr; however with the TRIM15 gene in Mamu and in Rano. The respective homologies are 84 % in Hs, 94 % in Patr, 93 % in Mamu and 86 % in Rano.

5.3.4.2 Clone 463N5

One partial insert sequence (30) of clone 463N5 obtained has a length of 358 bp. It was sequenced using primer T7.29, and contains no significant similarities with the vector sequence. The sequence aligns in Hs near the HLA-B exon 8, in Patr the Patr-B exon 8 and in Mamu near the Mamu-B13. The respective homologies are 91 % in Hs, 92 % in Patr and 88 % in Mamu.

A second sequence (31) obtained has a length of 365 bp. It was sequenced using primer ptarbac3, and showed no significant similarity with the vector sequence. The sequence aligns in Patr near the Patr-E exon1 and in Mamu near the Mamu-E gene. The respective homologies are 86 % in Patr and 85 % in Mamu.

5.3.4.3 Clone 436B4

The partial insert sequence (32) of clone 436B4 obtained has a length of 247 bp. It was sequenced using primer T7.29, and showed no significant similarities with the vector sequence. The sequence aligns in Hs near the HLA-B, in Patr near the Patr-B exon 8 and in Mamu the pseudogene Mamu-B10. The respective homologies are 94 % in Hs, 95 % in Patr and 96 % in Mamu.

5.3.4.4 Clone 325P21

One partial insert sequence (33) of clone 325P21 obtained has a length of 323 bp. It was sequenced using primer T7.29, and showed no significant similarity with the vector sequence. This sequence aligns the TRIM10 exon 6 in Hs, in Patr and in Mamu, likewise, the TRIM10 gene in Rano and in Mumu, near the TRIM10 gene in Sucr. The respective homologies are 91 % in Hs, 89 % in Patr, 88 % in Mamu, 92 % in Susc, 97 % in Rano and 94 % in Mumu.

A second sequence (34) obtained has a length of 171 bp. It was sequenced using primer ptarbac3. It contains no significant similarities with the vector sequence. Quite differently the sequence aligns in Hs near the HLA-G gene and in Patr near the Patr-G exon 3 and in Mamu near the pseudogene MamuG-2. The respective homologies are 94 % in Hs, 95 % in Patr and 96 % in Mamu.

5.3.4.5 Clone 279K21

The partial insert sequence (35) of clone 279K2 obtained has a length of 482 bp. It was sequenced using primer T7.29, and showed no significant similarities with the vector sequence. The sequence aligns in Hs near the HCGII-4 and in Mamu near the pseudogene human LOC282911 (HCG11-4). The respective homologies are 98 % in Hs and 83 % in Mamu. The respective homologies are 94 % in Hs and 96 % in Mamu.

5.3.4.6 Clone 224H20

The partial insert sequence (36) of clone 224H20 obtained has a length of 159 bp. It was sequenced using primer T7.29, and showed no significant similarities with the vector sequence. This sequence aligns in Hs near the TRIM10 exon 7 and in Mamu near the TRIM10 gene. The respective homologies are 93 % in Hs and 93 % in Mamu.

5.3.5 Sequences of contig 4

Sequence analysis of clones belonging to contig 4

In this contig number 4, 3 partial sequences were obtained. Two of these replicated were by means of primers belonging to the ptarbac 2.1 vector: two different sequences (although of equal bp size) of the clones 348C21 and 171K8, respectively. One further sequence was obtained of the clone 171K8 with the primer ptarbac2.

5.3.5.1 Clone 348C21

The partial insert sequence (37) of clone 348C21 obtained has a length of 485 bp. It was sequenced using primer T7.29. It contains no significant similarities with the vector sequence. The sequence aligns in Hs near the MOG exon 3 and intron 3, and in Sucr near the MOG. The respective homologies are 93 % in Hs and 93 % in Sucr.

5.3.5.2 Clone 171K8

One partial insert sequence (38) of clone 171K8 obtained has a length of 637 bp. It was sequenced using primer T7.29, and showed no significant similarities with the vector sequence. The sequence aligns in Hs near the MOG gene and has a homology of 88 % in Hs (Roth *et al*., 1995; Gonzales *et al*., 1996).

A second sequence (39) has a length of 567 bp. It was sequenced using primer ptarbac2. It showed a strong match at one end, the positions 5 - 39 with the vector sequence. This sequence aligns in Hs near MOG. The respective homology is 88 % in Hs (Roth *et al*., 1995; Gonzales *et al*., 1996).

5.4 Identification of the *Callithrix jacchus* sequences obtained by comparative alignment

5.4.1 Programs used for identification of sequences by localization and gene assignment

First step in the search for alignment was searching for homology and percent identity of the sequences by means of the BLAST program. This step gave information on the sequence type (location at or near a gene or intergenic sequence), the species in which the homology was found, the percentage of a maximum identity and the genomic position in the human genome compared. BLAST algorithms

program shows a specific part within a sequence having the best score (Altschul *et al.*, 1997). BLAST, however, gives no information for parts of lower identity.
Second step was to take information for other species and the query sequence and to align both sequences in Clustal X. This step allowed complete alignment between gene sequences and comparative DNA. Some sequences needed a modified approach, in i.e. using the Gene runner program, enabling correct alignment with the reverse or complementary sequence.
The third step was to calculate manually comparative genomic bp positions in the human genome.
Six of the 55 sequences were formally translated into amino acid sequences by means of the Swiss-prot program: K19-161C12-TS1, K19-161C12-TS2, K28-347D1-TCF19, K24-171K8ptarbac2-38, K24-171K8ptarbac2-39 and K7-239N18T729-41 were informative. These sequences were thereupon examined by the AUGUSTUS program for gene prediction, and by Clustal X for amino acid alignment.
Some of the problems remaining resolved or critical in sequence aligning and evaluation will be mentioned in the discussion.

Figure 5.4.1: Gene map of the Hs MHC class I region showing aligment of the BAC clone sequences as found by BLAST and / or Clustal X. bp-scale map of the Hs MHC class I region: 1st and last bp of genes and sequences calculated for entry into a bp scale of the Venter genome, chromosome 6. For a more convenient display of all known Hs genes of the region and comparative alignment of the 55 *Callithrix jacchus* BAC clone sequences, the scale has been locally expanded from 10 to 1 or 0.1 kb per graduation line, respectively. See next three pages.

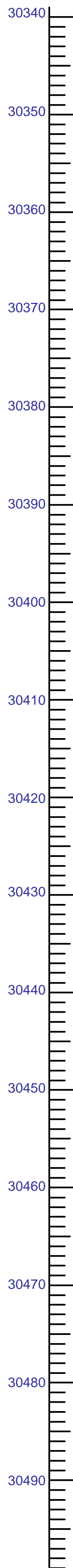
Figure 5.4.1.: bp-scale map of the Hs MHC class I region (legend on preceding page)

29600000
29700
29710
29720
29730
29740
MOG
732788
748128
743737
744314(38)
743747
744316(39)
743775
744384(41)
746705
747350(37)
29750
29760
29770
29780
29790
29800
HLA-F
799920
802262
29810
29820
HCG9P5
824007
824403
29830
29840
HCP5P14
846303
848642

29850
29860
HCG4P10 866336
867225
HCG4 866795
868829
29870
HCP5P13 868890
870002
29880
29890
MICG
888146
888448
29900
904397
904559(34)
HCG4P8
901886
902871
HLA-G
903497
906859
29910
HCGVIII-2
909405
910684
914075
914824(12)
29920
29930
MICF
927943
928242
29940
3.8- 1.4Ψ
941672
942840
943498
944248(11)
29950
HCP5P10
947101
949757
001119
011349(47)
018399
018502(48)
052489
052209(28)
29960
050881
051331(5)
HCG4P7 961867
962764
HLA-H 963508
966246
HLA-16 972197
974706
29970
HCG2P7 974785
978417
HCP5P6 991509
993871
HCG4P6 000348
001407
HLA-K 002215
005466
29980
HLA21 009520
010636
HCG4P5 016669
017558
HLA-A 018310
021633
HCG4P4 023982
025281
29990
HCP5P3 030962
031590
HLA-80 031590
034814
HCG2P6 036086
043178
MICD 046128
051501
30000
HCGIX 054159
056871
HCP5P2 076761
079028
HCG4P3 080602
081585
HLA-J 080602
081585

30010
30020
30030
30040
30050
30060
30070
30080
30090
30100
30110
30120
30130
30140
PPP1R11
146089
142968
30150
30160

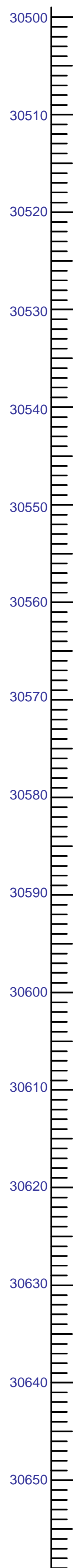
30180
30190
30200
30210
216321
216669(29)
TRIM40
212489
223570
30220
30230
231061
231414(33)
TRIM10
227701
236690
30240
TRIM15
238970
248452
30250
30260
30270
TRIM26
260211
289132
30280
30290
30300
30310
30320
30330
HLA-L
335592
336273(45)
334627
339458



HCG18
366555
402424

TRIM39
403019
419485

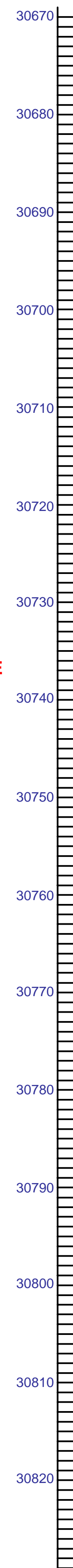
MICC
490466
495087



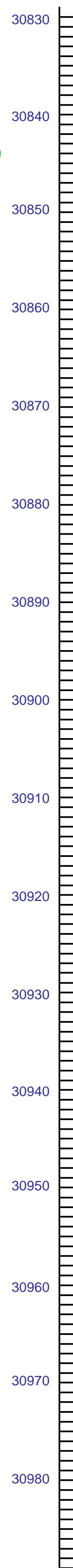
HLA-E
565250
569077

611549
611056(35)

PRR3
79600
79881



PPP1R10
676161
692999



30990
31000
31010
31020
31030
31040
31050
31060
31070
31080
31090
31100
31110
31120
31130
31140

143435
143623(16)

31150
31160
31170
31180
31190
31200
31210
31220
31230
31240
31250
31260
31270
31280
31290
31300

159494
159368(53)
161329
160923(18)
169767
170547(8)
170512
171130(9)
210425
210960(6)
210856
211299(7)
231061
231414(33)
233525
236024(10)
237043
233594(13)
238331
237860(15)
238250
237661(19)
238196
237456(20)
238204
237717(21)
246432
247028(1)

HCG22
157810
159820

SEEK1
190602
215816

TCF19
235226
238517

POU5F1
246432
240107

HCGIX-3.2
273622
274213

HCG27
273516
279723

31310
31320
31330
31340
31350
31360
31370
31380
31390
31400
31410
31420
31430
31440
31450
31460

316588
316308(3)
341823
343925(TS1)
345337
346057(45)
345549
345164(40)
344583
344998(46)
348224
348399(27)
348224
348399(42)
348481
348224(24)
348578
348215(30)
348578
348292(32)
348539
348299(44)
348537
348331(22)
349041
348759(26)
374599
375006(31)
423162
422987(42)
427248
426864(40)
427033
429135(TS1)
429344
429048(32)
429699
428582(27)
429441
429752(26)
430451
431175(45)
433437
433178(24)
433510
433271(44)
433521
433186(30)
433519
433243(32)
433518
433304(22)
464397
464701(4)

HCGII-2
322353
323267

HCGIX-3.1
324262
324845

HLA-C
344508
347874

348230
348479(43)

HCG2P1
414185
415926

HLA-B
429628
432914

HCGIV-1
433041
433421

HLA-17
453292
454224

P5-8
455532
456872

429427
429677(43)

31470
31480
31490
31500
31510
31520
31530
31540
31550
31560
31570
31580
31590
31600
31610
31620
31660

HCGIX-2
479704
480113

MICA
479350
491069

HLA-X
538410
537765

HCP5
538938
541461

HCGIX-1
575055
574658

MICB
605975
617904

607147
610237(TS2)

BAT1
605975
617904

ATP6GV1
620218
622606

663226
664386(2)

Table 5.4.1.1: Comparative definition of the first 55 MHC class I sequences of *Callithrix jacchus* by BAC clone and primer used, length in bp obtained by PCR, first bp on the scale of Shiina, Venter genome numbers of sequence, % identity of aligning part within or near a Hs gene.

#	Clone-contig	Primer	Sequ. length (bp)	First bp on Shiina scale	First Venter genome bp number of sequence / 1st – last bp aligned[1]: % identity with Hs[2]	Gene hit or near[3]: defined by its first - last Venter genome numbers[1]
1.	204C3-1	ptarbac3	438	744235	31247042 /247028 – 246432: 88	HCGIX-3.2: 31247042 - 31274213
2.	204C3-1	T729	619	349000	31662883 / 663226 – 664386: 87	1C7: 31664932 - 31323267
3.	510K19-1	ptarbac3	211	696000	31315718 / 316308 – 316588: 88	HCGII-2: 31320349 - 31323267
4.	217M17-1	ptarbac3	208	621000	31464017 / 464397 – 464701: 86	HCGII-1: 31414185 - 31415110
5.	217M17-1	T729	383	560000	30050871 / 050881 – 051331: 90	HCG-IX: 30054159 - 30056871
6.	347K7-1	ptarbac	491	847000	31210397 / 210425 – 210960: 84	SEEK1: 31190602 - 31215816
7.	347K7-1	ptarbac4	363	847000	31210397 / 210856 – 211299: 90	SEEK1: 31190602 - 31215816
8.	347D1-1	ptarbac3	518	845500	31169307 / 169767 – 170547: 86	HCG22: 31157810 - 31159820
9.	347D1-1	ptarbac4	534	847000	31170107 / 170512 – 171130: 84	HCG22: 31157810 - 31159820
10.	347D1-1	TCF19	2314		31236549 / 236024 – 233525: 92	TCF19: 31235226 - 31238517
11.	459O20-1	ptarbac3	624	472000	29942715 / 943498 – 944248: 85	3.8 – 1.4: 29941672 - 29942840
12.	459O20-1	ptarbac4	511	166000	29913972 / 914075 – 914824: 86	MICF: 29928242 - 29927943
13.	459O20-1	T729	121		31237214 / 237043 – 233594	TCF19: 31235226 - 31238517
14.	459O20-1	T729	159		No alignment in Hs localized	
15.	459O20-1	TCF19	258		31235226 / 238331 – 237860: 92	TCF19: 31235226 - 31238517
16.	169G2-1	T729	181		31140485 / 143435 – 143623:94	HCG22: 31157810 - 31159820
17.	459O20-1	ptarbac4	121		No alignment in Hs localized	
18.	169G2-1	ptarbac4	328	882000	31160924 / 160923 – 161329: 94	HCG22: 31157810 - 31159820
19.	123H7-1	T729	527		31237984 / 237661 – 238250: 89	TCF19: 31235226 - 31238517
20.	169G2-1	TCF19	567		31237876 / 237456 – 238196: 94	TCF19: 31235226 - 31238517
21.	99L19-1	TCF19	377		31237479 / 237717 – 238204: 89	TCF19: 31235226 - 31238517
1TS	161C12-1	?	2994	583800 668236	31432323 / 432291 – 429329: 88	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
2TS	161C12-1	?	3000		31607147 / 607147 – 610237: 90	BAT1: 31605975 - 31617904
22.	261L9-1	HLA-B 29	207	590669 675217	31430382 / 430382 – 429958: 94	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
23.	271C5-1	T7.29	170		ptarbac2.1 = vector	
24.	271C5-1	HLA-B 32A	575	590206	31429698 / 430105 – 429698: 95	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
25.	271C5-1	T7.29	170		ptarbac2.1 = vector	
26.	127E13-1	HLA-B29	280	590910 675173	31432636 / 432681 – 433030: 93	HLA-B: 31429628 – 31432914 HLA-C: (after HLA-C)
27.	127E13-1	HLA-B29B	175	590070 674636	31432914 / 432071 – 431868: 89	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
28.	282L12-1	ptarbac3	272		30052489 / 052489 – 052209: 89	HCG-9: 30054159 - 30056871
29.	245C6-3	T729	280		30216316 / 216321 – 216669: 84	TRIM40: 30212489 - 30223570
30.	463N5-3	HLA-B32	332	590551 675100	31432914 / 432798 – 432580: 92	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
31.	463N5-3	ptarbac3	365		31374599 / 374599 – 375006: 85	HCGII-3: 31376001 - 31379540
32.	436B4-3	T7.29	273	590608 675178	31432914 / 429344 – 429048: 94	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
33.	325P21-3	T7.29	323		30227701 / 231061 – 231414: 91	TRIM10: 30227701 - 30236690
34.	325P21-3	ptarbac3	171		29903497 / 904397 – 904559: 94	HLA-G: 29903497 - 29906859
35.	279K21-3	ptarbac3	482		30611717 / 611549 – 611056: 98	HLA-E: 30565250 - 30569077
36.	224H20-3	T7.29	159			
37.	348C21-4	T7.29	485		29732788 / 746705 – 747350: 93	MOG: 29732788 - 29748128
38.	171K8-4	T7.29	637		29732788 / 743737 – 744314: 88	MOG: 29732788 - 29748128
39.	171K8-4	ptarbac2	567		29732788 / 743747 – 744316: 88	MOG: 29732788 - 29748128
40.	171K8-4	Caja G	362	587515 672050	31347834 / 347713 – 344005: 90	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
41.	239N18-4	T7.29	319		29732788 / 743775 – 744384: 88	MOG: 29732788 - 29748128
42.	120D3-1	HLA-B32B	175	590071 674636	31429628 / 433575 – 433744: 89	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
43.	120D3-1	HLA-B32B	244	590086 674651	31429628 / 433593 – 433843:89	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
44.	120D3-1	HLA-B29	235	590874 675422	31425377 / 425251 – 425012: 93	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
45.	120D3 -1	Caja-G	666	587217 671748	30334627 / 347882 - 348645: 85	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834

46.	120D3-1	Caja-E	363	587910 672448	31347834 / 344583 – 344998: 88	HLA-B: 31429628 – 31432914 HLA-C: 31347834 - 31347834
47.	120D3-1	MIC	142		30018310 / 011199 – 011349: 86	HLA-A: 30018310 - 30021633
48.	120D3-1	Klex2	84		30018310 / 018399 – 018502: 93	HLA-A: 30018310 - 30021633
49.	193P12-1	T7.29	693		No alignment in Hs localized	
50.	18G14-1	ptarbac2	279		ptarbac2.1 = vector	
51.	19I16-1	T729	221		No alignment in Hs localized	
52.	174N1-3	T7-29	202		ptarbac2.1 = vector	
53.	169G2-1	T729	121		31159820 / 159494 – 159368:	HCG22: 31157810 - 31159820

1: Insertions in *Callithrix jacchus* omitted for calculation of Venter genome bp numbers

2: Insertions/deletions in *Callithrix jacchus* disregarded for calculation of % identity with Hs

3: For Venter genome bp numbers of genes see map to bp-scale of the MHC class I region in Hs, figure 5.4.1.

The 55 sequences mapped and tabulated mostly belong to the intervalls MOG - TCTEX4 and ATP6gV1 - POU5F1 (see 5.2.3.3.1 and 5.2.1.3.1) called for short contigs 4 and 1, respectively.

The sequences of contig 4 were only localized and possibly gene assigned. Deeper considerations could be given only to some items of contig 1. Especially a number of sequences aligning in the HLA-B/C section were evaluated for their identity with a homologous part of the Hs B versus the C duplicon. As the results displayed in table 5.4.2 show, all of these HLA sequences by their alignment proved to be more similar to duplicon B rather than to C.

On the other hand, TS1 is also aligning in both duplicons B and C at high identities. With its unique and absolutely rare insertion of a rather young ferritin pseudogene (see below), a TS1 containing this rarity showed have been sequenced from one of two probable sequences, either the B or in the C duplicon. Because of a high identity in these duplicons, a critical evaluation is required.

To avoid a background of mutations irrelevant for the decision, only those point mutations were counted that affected only one stretch either B or C. Most of the mutations between TS1 and its homologous sequences in Hs and Patr duplicons B and C were disregarded. In other words the similarties in these crucial bases between TS1 and either strand B or C were counted. Table 5.4.3 shows the results: generally higher similarities with duplicon C in all three partial sequences of TS1-ps1, -ps2 and -v fragments in both Hs and Patr (see 5.6.1). Therefore, TS1 seems to be a homologous copy sequenced from BAC clone DNA stemming from a Caja duplicon C. However, more homologues of TS1–pseudogenes ps1, ps2 and v shall be compared, e.g. those of the 14 Mamu-B duplicons; results will be assed by statistical evaluation. A total sequencing of the HLA B/C section started by Prof. Takashi Shiina should present the final answer soon.

From these results i.e. the sequences localized near HLA-B/C all having higher identities with Hs duplicon B, and TS1 having more specific similarities with Hs duplicon C, we may conclude that *Callithrix jacchus* has both duplicons, B and C, in agreement with Hs, Patr and other primates (see below, 5.6.2 and Discussion, 6.4).
In conclusion, these sequences within the for time available could all be defined, but evaluated only in some parts seeming of more general interest, as for their evolution (5.5), TS1, however, for its special challenges for molecular immunogenetics (5.6).

Table 5.4.1.2: Similarity of TS1-ps1, TS1-ps2 and TS1-v with homologous DNA in duplicons B and C of Hs and Patr. For lowering background by not counting mutations occurred just in TS1 or even in all three strands, only single base exchanches in either B or C have been compared with TS1 bases that rement unchanged, i.e. equal with one base in either strand C or B. The procedure allows to better weigh relative similarity of TS1 with both duplicons, B and C. Further data should be collected, e.g. from the 14 Mamu-B duplicons in order to have greater numbers for a statistical evaluations. These preliminary data favor TS1 being taken from a Caja duplicon C.

Mutation points rement equal in	TS1-ps1	TS1-ps2	TS1-v
Hs duplicon B	35	19	2
Hs duplicon C	36	31	9
Patr duplicon B	34	18	5
Patr duplicon C	40	29	9

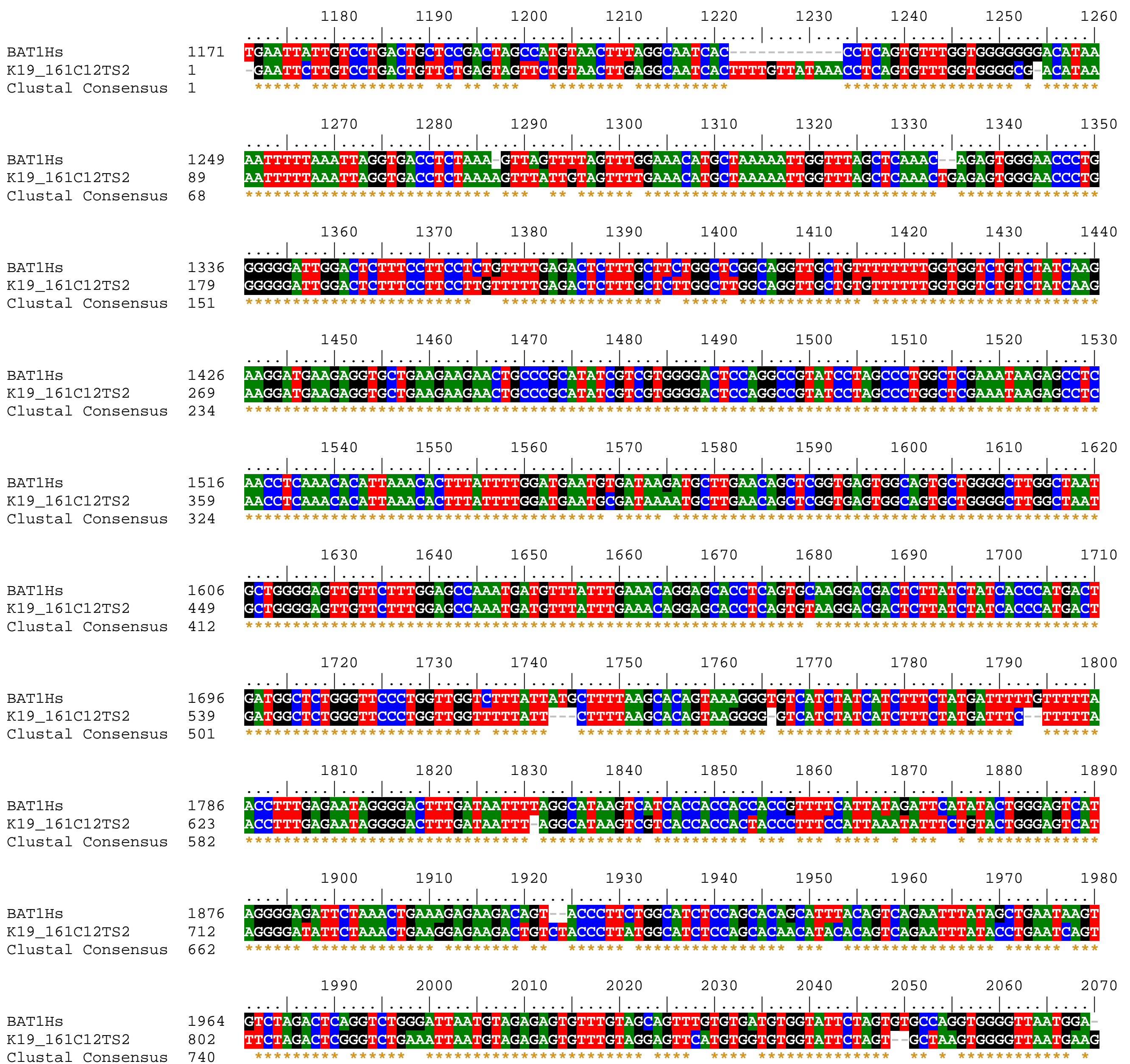
5.4.2 Collection and presentation of sequence data

Nearly all of the 55 pilot sequences obtained by PCR on Caja BAC clones have been localized and where possible gene assigned by comparative alignment in the Hs genome. These results are shown in a map to bp-scale, Figure 5.4.1.1. The alignment data of the sequences were calculated in bp numbers both on the scale by T. Shiina and of the Venter genome. The alignment was defined by first and last aligning bp and shown in table 5.4.1.1
The results of localization are visualized in the map shall be completed with all genes known in the Hs MHC class I region. All loci of the class I genes in Hs, and for comparison those of the *Callithrix jacchus* sequences found, have been defined and listed in table 5.4.1.1.

The data of localization and assignment by homologous alignment in the genomes of primates and other mammals are collected in table 5.4.2.

5.4.3 Demonstration of individual sequences identified by localization and possibly gene assignment in the Hs genome or proteome sequence K19161C12-TS2

This sequence contains 3000 bp. The AUGUSTUS program indicated a gene between positions 236 and 2259. The alignment was compatible with BAT1 gene in human (Allcock *et al.*, 1999). The total sequence aligned. In the BAT1 gene, it is from intron 4 until intron 6, see Figure 5.4.2.



```
                                  2080      2090      2100      2110      2120      2130      2140      2150      2160
BAT1Hs              2052   --AGATTTTTCTGTAAGAATTGAATCTTGGTGAATGAGAGTGGGGTTGGACATAGGCCCCATA----AGTCATTACAAATGATCTTTGGC
K19_161C12TS2       890    AAAGATTTTTCTGTCAGAATTGAATCTTGGTAAATGAGAGTGGGGATGGACATAGGCTCCAGAGAAAAGTTATTACAAATGATCTTTGGC
Clustal Consensus   816      ************ *************** ************ *********** *** *    *** *******************

                                  2170      2180      2190      2200      2210      2220      2230      2240      2250
BAT1Hs              2137   AATTCTATATGGTGAGCTATAAAGGTGGGCTCCAGGTAGGGATGTCATATTTGCCTGACTTGATAGAAAAGTAATCCAGAGAGTCATAGA
K19_161C12TS2       980    AATTCAATATGGCGAGCTATAAAGGTGGGCTCCAGGTAGAGATGTCTTATTTGTCCAACTGGATAGAAAAGTGATCCAGAGGGTCATAGA
Clustal Consensus   895    ***** ****** ************************* ****** ****** *  *** *********** ******** ********

                                  2260      2270      2280      2290      2300      2310      2320      2330      2340
BAT1Hs              2227   TGGACTCTGATATCTGGAATATAATATGTGCTTGATATTTGTAGTCTGCTGAAGGCTGGCTGGGGCTTGGGCAGGAAAGGGTCGGGAGAA
K19_161C12TS2       1070   TGAACTTTGATAGCTGGAATGTAATATGTGCTTGATATTTGTAGTCTGCTGAAGGCTGGCCTGGGCTTGGGCAGGAAAGAGTTGGGAGAA
Clustal Consensus   975    ** *** ***** ******* ************************************** ***************** ** *******

                                  2350      2360      2370      2380      2390      2400      2410      2420      2430
BAT1Hs              2317   GGTCCCATAAAGCATGTTTAGAAGGCCTT------GAGAGCCTCTGCACTGGGCTTTATCCCCATTTCATAGTTGGGAACTTTGGGGTTT
K19_161C12TS2       1160   TGGCCTGTAAAGCATGTTTTGAAGGCCTTATGCCTGAGAGTATCTGCACAGGGCTTTGTCCCCATTTATAGTTGTGAACTTCGGGGTTT
Clustal Consensus   1056    * **  ************ *********       *****  ******* ******* ********* ******* ****** *******

                                  2440      2450      2460      2470      2480      2490      2500      2510      2520
BAT1Hs              2401   TACCTTATTTCTTGCTTGGTTAAAACAAACAGATGGAATCTGATCCCACTTCTTGATTCCAAGTCCATTGCTCTTTCCATTGTGTTGTTA
K19_161C12TS2       1250   T---TATCTTTCTTGCTT---------AAACAGTTGGAATATAATTCCAATTCTTGATTCCAAGTTCATTGC--------------------
Clustal Consensus   1129   *    * **********         ****** ****** * ** *** ************** ******

                                  2530      2540      2550      2560      2570      2580      2590      2600      2610
BAT1Hs              2491   CTATTTCCAGCAATCTTCACCTCACTGGGAAGTCTACCTCTAATCTTTGTTTATCATACCTGCTTATTTTCTCCTACAATTTTTTTCCTT
K19_161C12TS2       1309   -----------AGTCTTCACCTCACTGGGAGGTCTACCTTTAATCTTT----ACTATACCTGCTTATTTTCTCCTACAATTCTTTTC-TT
Clustal Consensus   1179              * **************** ******** ********     * ************************* ***** **

                                  2620      2630      2640      2650      2660      2670      2680      2690      2700
BAT1Hs              2581   GTTCTTGTAGACATGCGTCGGGATGTCCAGGAAATTTTTCGCATGACCCCCNACGAGAAGCAGGTCATGATGTTCAGTGCTACCTTGAGC
K19_161C12TS2       1384   GTTCTCATAGACATGCGTCGGGATGTCCAGGAAATTTTTCGCATGACCCCCCATGAGAAGCAGGTCATGATGTTCAGTGCTACCTTGAGC
Clustal Consensus   1248   *****  ******************************************* * ***********************************

                                  2710      2720      2730      2740      2750      2760      2770      2780      2790
BAT1Hs              2671   AAAGAGATCCGTCCAGTCTGCCGCAAGTTCATGCAAGATGTAAATACCCTTCTACCTTCTCTCCCTCCACTCCCCGCCCGCTGCCTCCTC
K19_161C12TS2       1474   AAAGAGATCCGTCCAGTCTGCCGCAAGTTCATGCAAGATGTAAATACCCTTCTACCTTCTCTCCCTCCACTCCCCGCCCGCTGCCTCCTC
Clustal Consensus   1334   ******************************************************************************************

                                  2800      2810      2820      2830      2840      2850      2860      2870      2880
BAT1Hs              2761   CCCTTCCTCGCCCTCTTCCTC-AGACTCCCTTGTCATTCAAGTGCCAAGAAGGCGGCTTGTGCCCAACTGGGAGTAATGACTCCTTGAAG
K19_161C12TS2       1564   CCCTTCCTCGCCCTCTTCCTCCAGACTCCCTTGTCATTCAAGTGCCAAGAAGTCGGCTTATGCCCAGCTGGGAGTAACGACTCCTTGAAG
Clustal Consensus   1424   ******************** ****************************** ****** ****** ********** ************

                                  2890      2900      2910      2920      2930      2940      2950      2960      2970
BAT1Hs              2850   AGACATACAGAAGCAGAGACAGCTAGTGTTAGGGCCTGCGCGGGTGCCAGGGAAACTCCGGAAGACTTGGTCGGGTTAATGTGAGAGCGG
K19_161C12TS2       1654   AGACATACAGAAGCAGAGACAGCTAGTGTTAGGGTCTGCGCGGGTGCCAGGGAAACTCCGGAAGACTTGGTCGGGTTAATGTGAGAGCGG
Clustal Consensus   1509   ********************************** *******************************************************

                                  2980      2990      3000      3010      3020      3030      3040      3050      3060
BAT1Hs              2940   GTAGTGTTCGACTTTTTC--ATAAATCACAACATTTTTGAACCTCTTCTCCCTTCGGGGGAGGGCAGGATTTTTCTGCCCTACCACCCAC
K19_161C12TS2       1744   GTAGTGTTCGACTTTTTCTCAAAAATCGCAACATTTTTGAACCTCTTCTCCTTTTGGGGGAGGGCAGGATTTTTCTGCCCTACCACCCAC
Clustal Consensus   1598   ******************  * ***** ********************** ** ************************************

                                  3070      3080      3090      3100      3110      3120      3130      3140      3150
BAT1Hs              3028   CCATCCATCGTCTCTTACATGCACCCTACAGCCACGCACCCTCAAGGTGGCATCGAGCATACAGCTGGAGCCTTCTGCTCACCAAAACTC
K19_161C12TS2       1834   CCATC----GTGTCTTACGTGCCCCCTACAGCCACGCATCCTCAAGGTGGCATCGAGCATACAGCTGGAGCCTTCTGCTCACCAAAACTC
Clustal Consensus   1682   *****    ** ****** *** *************** ******************************************************

                                  3160      3170      3180      3190      3200      3210      3220      3230      3240
BAT1Hs              3118   CTACTTCCCGGTGGCAGGAGAGCAAGAGAGGGACAGACAGATGGCAGGGCATGTCCAAAAGAAGAGCATCAGCACAAATGAATCCTCCCC
K19_161C12TS2       1920   CTACTTCCCAGTGGCAGGAGAGCAAGAGAGGGACAGACAGATGGCAGGGCATGTCCAAAAGAAGAGCATCAGCACAGATGAATCCTCCCC
Clustal Consensus   1764   ********* ******************************************************************** *************
                                  3250      3260      3270      3280      3290      3300      3310      3320      3330
BAT1Hs              3208   TTCCCCACCTCCAGGGGTGGGGGCCTTTGGCACCTCAATCCCCGATACCCTACTCCTTCCCACCCACATCTCCTTGCACCCATCTGGAAC
K19_161C12TS2       2010   TTCCCCACCTCCAGGGGTGGGGGCCTTTGGCACCTCAATCCCCGATACCCTACTCCTTCCCACCCACATCTCCTTGCACCCATCCGGAAC
Clustal Consensus   1852   *************************************************************************************** *****

                                  3340      3350      3360      3370      3380      3390      3400      3410      3420
BAT1Hs              3298   CTCGGTTGATGTGAGCCGGCAACAGAGAAGCACCGTGGCGCGGCGAGGGAATGCAGACGGCACCCAGCGGTGGATGGCGGCAGCGGAGGC
K19_161C12TS2       2100   CTCGGTTGATGTGAGCCGGCAACAGAGAAGCACCGTGGCGCGGCGAGGGAATGCGGACGGCACCCAGCGGTGGATGGCGGCAGCGGAGGC
Clustal Consensus   1941   ****************************************************** ***********************************
```

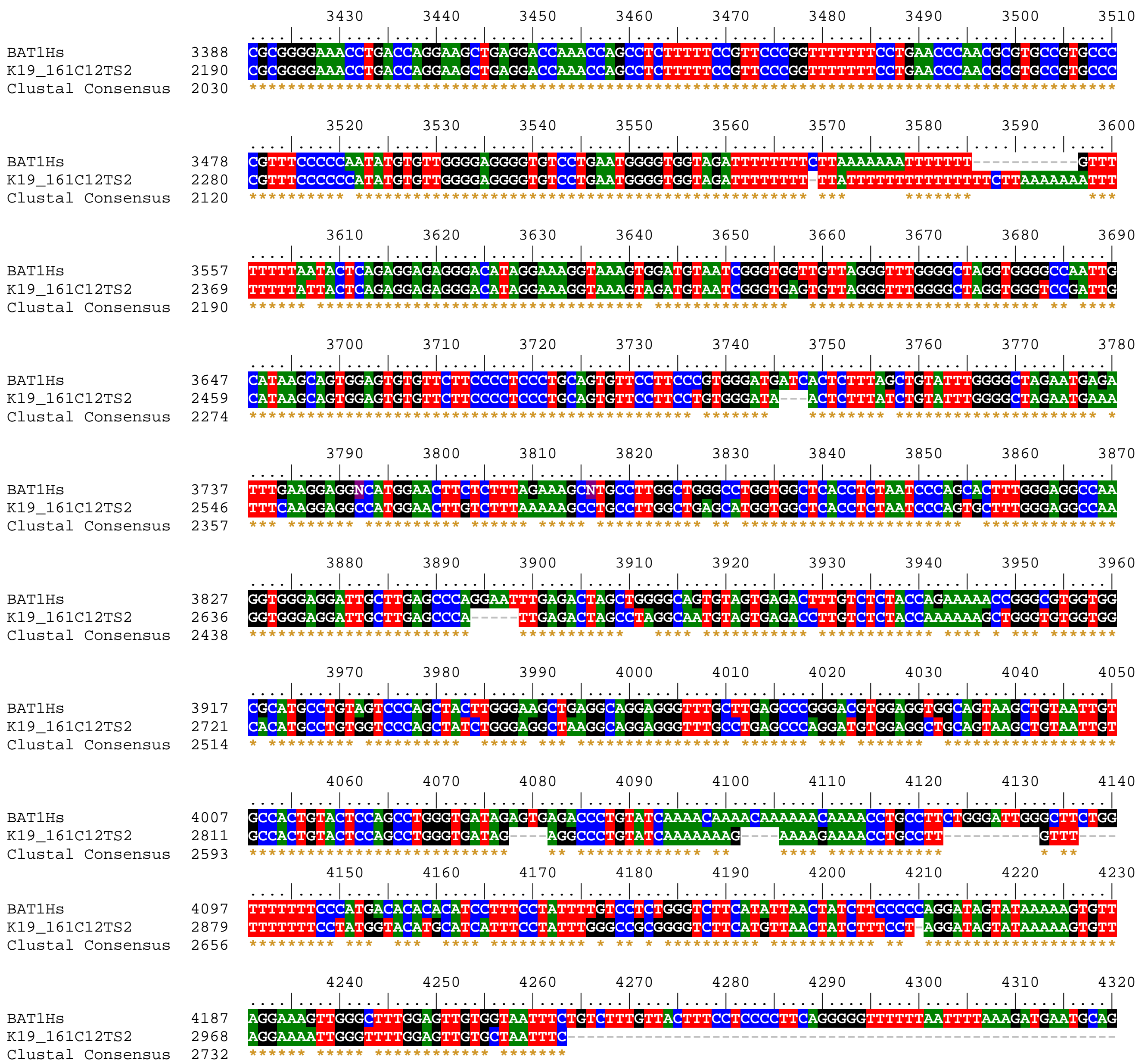


Figure 5.4.2: K19-161C12-TS2: part of BAT1 gene

Alignment has been performed between K19-161C12-TS2 and BAT1 gene of Hs (ID:NM 004640) by Clustal X, into intron 4 until 6 in position 1172 until 4263.

5.4.4 K28-347D1-TCF19, nucleotide sequence

This sequence contains 2314 bp. The AUGUSTUS program indicated a gene between positions 532 and 1968. The alignment was compatible with a part of gene TCF19, however, the nucleotide sequence showed only a small fragment high identity (see Figure 5.4.2.1). Nevertheless, this sequence was translated by Swiss-prot program; thereupon the amino acid sequence was aligned in Clustal X. The sequence had a homology with the human sequence between positions 158 and

239, (see Figure 5.4.2.2) again only a small part of the whole gene product (protein). Although, such homology may deserve consideration, e.g. from a comparative evolutionary aspect, we have generally refused such finding as being in sufficient for both localization and gene assignment of a sequence.

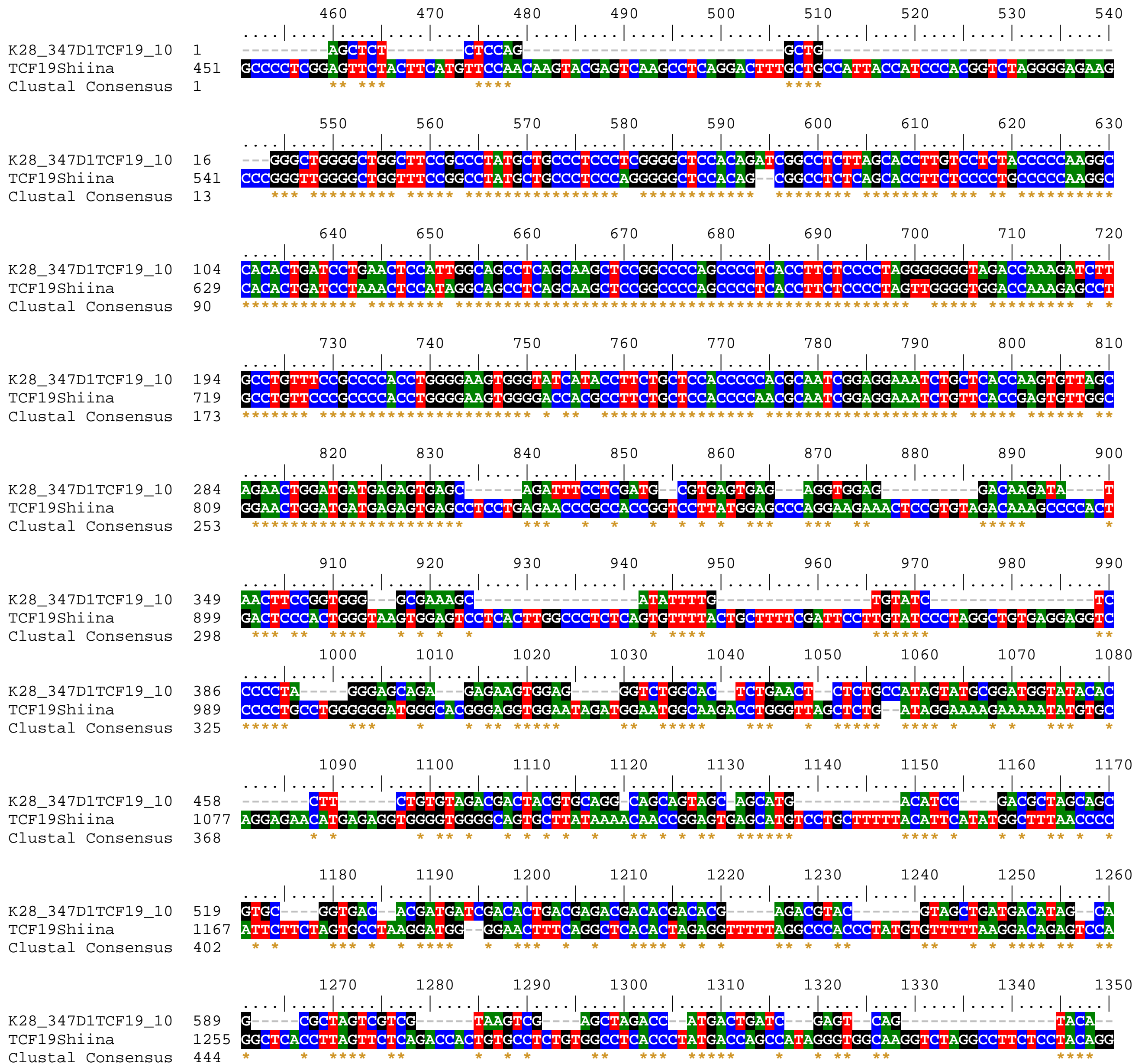


Figure 5.4.4.1: K28-347D1-TCF19 , nucleotide sequence

Alignment has been performed between K28-347D1-TCF19 and TCF19 gene of Hs (ID: NM 001077511) by Clustal X, in position 460 until 1350.

```
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
K28_347D1TCF19_10  1    -------LSPGWGWG-----WLPPYAA------------------------------------------------------------LPR
HsTCF19            1    MLPCFQLLRIGGGRGGDLYTFHPPAGAGCTYRLGHRADLCDVALRPQQEPGLISGIHAELHAEPRGDDWRVSLEDHSSQGTLVNNVRLPR
Clustal Consensus  1           *  * * *     : **  .*                                                            ***

                               100       110       120       130       140       150       160       170       180
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
K28_347D1TCF19_10  18   ------------------------------------------------------------------GSTDRPLSTLSSTPKATLILNSIG
HsTCF19            91   GHRLELSDGDLLTFGPEGSPGTSPSEFYFMFQQVRVKPQDFAAITIPRSRGEARVGAGFRPMLPSQGAPQRPLSTFSPAPKATLILNSIG
Clustal Consensus  11                                                                     *:.:*****:*.:***********

                               190       200       210       220       230       240       250       260       270
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
K28_347D1TCF19_10  43   SLSKLRPQPLTFSPRGGRPKILPVSAPPGEVGIIPSAPPPRNRRKSAHQVLAELDDESE-------------------------------
HsTCF19            181  SLSKLRPQPLTFSPSWGGPKSLPVPAPPGEVGTTPSAPPQRNRRKSVHRVLAELDDESEPPENPPPVLMEPRKKLRVDKAPLTPTGNRRG
Clustal Consensus  34   **************  * ** ***.*******  ***** ******.*:**********

                               280       290       300       310       320       330       340
                        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
K28_347D1TCF19_10  101  ------------------------------------------------------QISSMRE--------------
HsTCF19            271  RPRKYPVSAPMAPPAVGGGEPCAAPCCCLPQEETVAWVQCDGCDVWFHVACVGCSIQAAREADFRCPGCRAGIQT
Clustal Consensus  83                                                         .*.: **
```

Figure 5.4.4.2: K28-347D1-TCF19, amino acid sequence

Translated nucleotide sequence was performed with AGUSTUS and BLAST. Alignment has been performed between K28-347D1-TCF19 and TCF19 protein of Hs (ID: NM 001077511) by Clustal X, the homology is localized in position 157 until 239.

5.4.5 MOG sequences

K24-171K8ptarbac2-38 and K24-171K8ptarbac2-39 were compatible alignment with MOG gene. The human MOG gene had a length of 15340 with eight exons (ID: 4340).

K24171K8ptarbac2-38 sequence presented a homology 86 % in the position 10933 until 11573, with four insertions and two deletions, this region corresponds to intron 1-2 (89-8942), see Figure 5.4.5.1.

K24171K8ptarbac2-39 sequence presented a homology 88 % in the position 6010 until 6579, with 3 insertions and 3 deletions, this region corresponds to intron 1-2 (89-8942), see Figure 5.4.5.2.

Both sequences not presented homology with MOG gene sequence from *Callitrhix jacchus* (gene ID: AY566834).

```
                              10810     10820     10830     10840     10850     10860     10870     10880     10890
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
K24_171K8T729_38   1     ------------------------------------------------------------------------------------------
MOGHs              10801 AGCAGGAAGAGTCTTCACTCCTCCCAATGCCCCACCTCCCATTTGTTACTAAGAGGAGAGGAGAAAGTAGCAAGGAGGGTATGGGGAATG
Clustal Consensus  1

                              10900     10910     10920     10930     10940     10950     10960     10970     10980
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
K24_171K8T729_38   1     ------------------------------------------TGACTTCATAGGGAATTCATCCCAGATGCTTGCTTGTCTATTTCTTC-
MOGHs              10891 TTCTGGGGGAATGGGTGTTGGTGCGATCAACAACAAAGTCCTTTCTCTCACCTTGAATTCATCCCAGATGCCTGCTTGTTTACTTCTTCC
Clustal Consensus  1                                               *    ***    ***************** ******* ** ******

                              10990     11000     11010     11020     11030     11040     11050     11060     11070
                         ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
K24_171K8T729_38   48    ACATGAAAAA-GGCTTTCAGCCCACCTGGCTGAGCAGAAAGAATCTGAATGTTAGAGTTAGGTAGCCTGGGTTTGAATTCCATCTCGGGT
MOGHs              10981 ACACAAAAAAAGGCCTTCAGCCCTCATGGCTGAGCAGAAAGAATCTGAATGTTAGAGTCAGGCAGCCTGGGTTTGAATTCCATCTCAGGT
Clustal Consensus  37    ***  ***** *** ******** * ******************************** *** *********************** ***
```

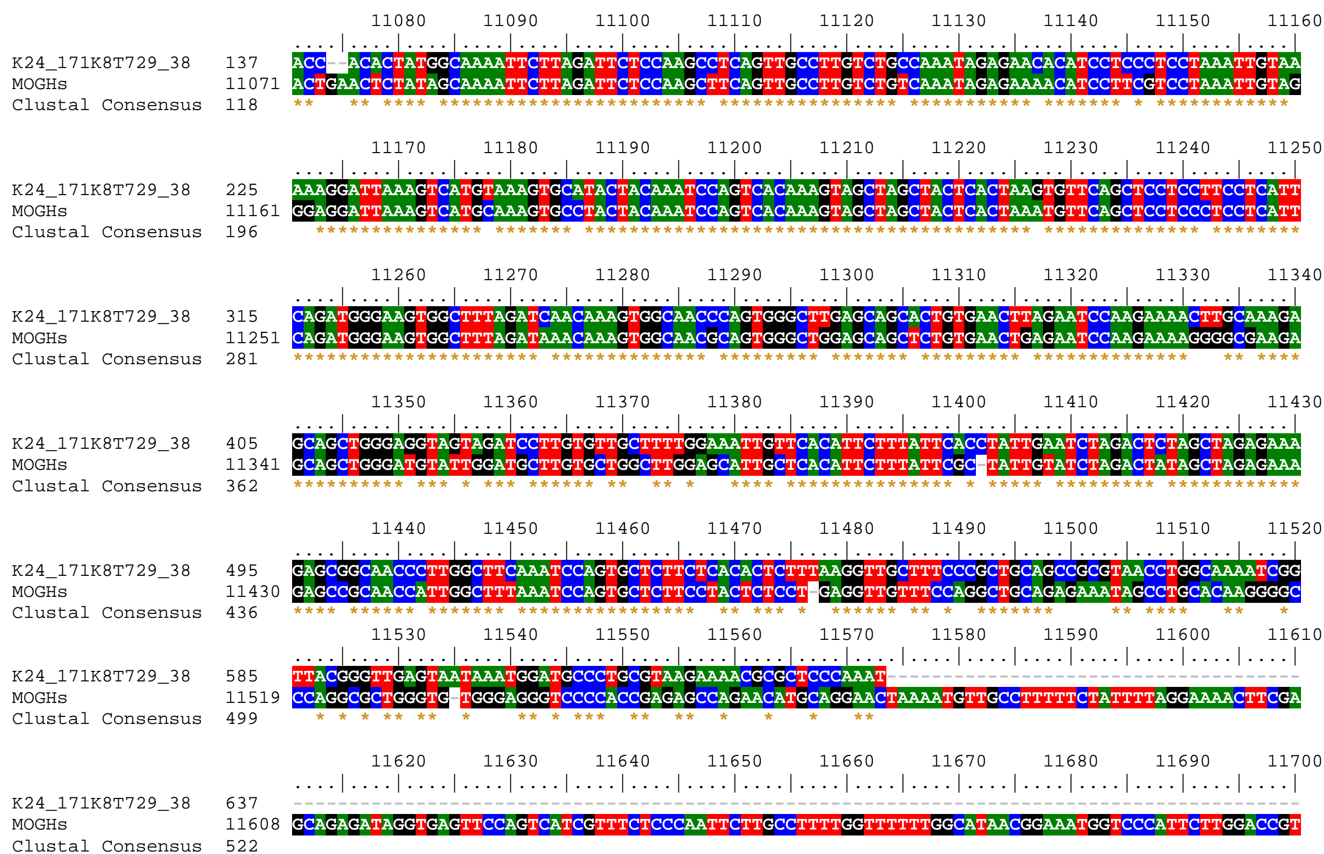
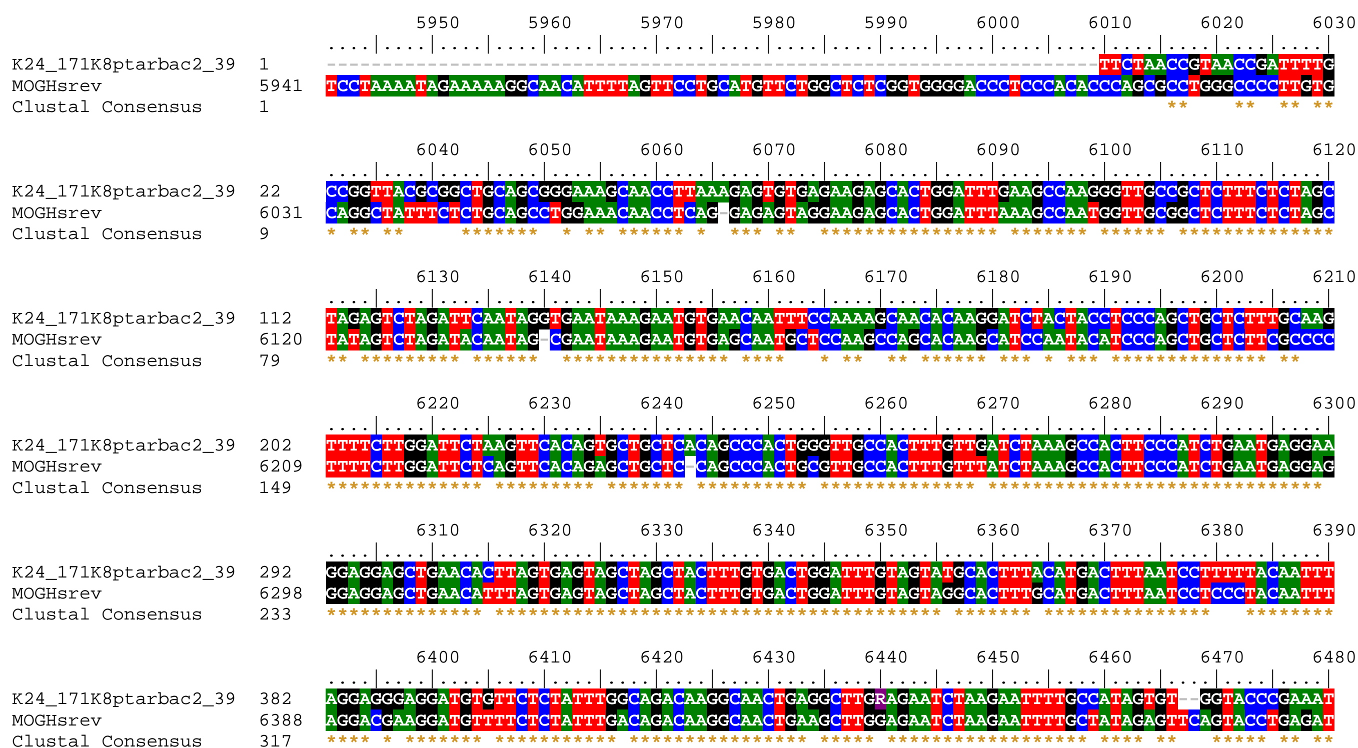
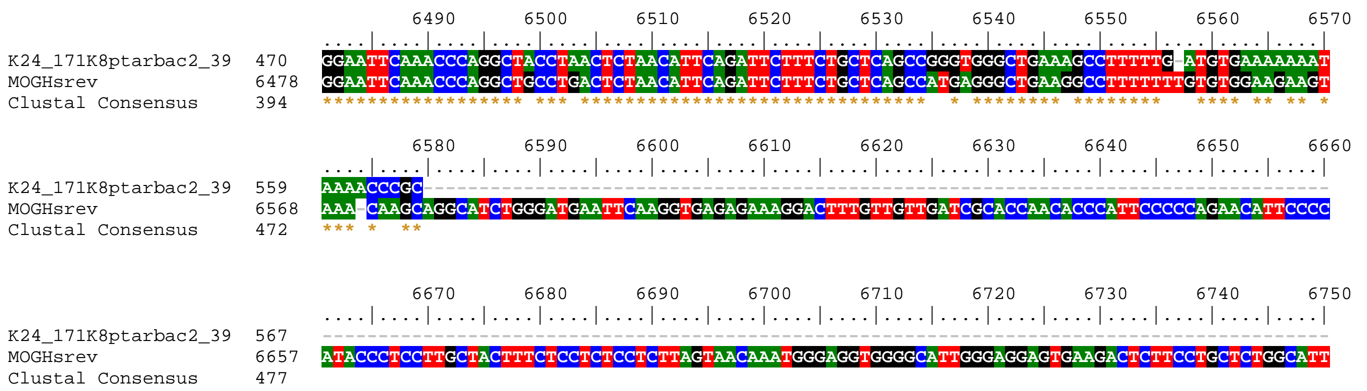


Figure 5.4.5.1: Sequence K24_171K8ptarbac2_38

Alignment has been performed between K24_171K8ptarbac2_38 and MOG gene of Hs (ID: AC000049) by Clustal X, into intron 1-2 in position 10933 until 11573.





5. 4.5.2 Sequence 24_171K8ptarbac2_39

Alignment has been performed between K24_171K8ptarbac2_39 and MOG gene of Hs (ID: AC000049) by Clustal X, into intron 1-2 in position 6010 until 6579.

5.5 Evolution of class I sequences and framework genes found in *Callithrix jacchus* compared to their homologous in other primates and mammals

The molecular evolution of four *Callithrix jacchus* genes is inspected here by gene trees, two for the Caja-G genes and comparison of Caja-G with the class Ia and class Ib HLA genes, and the one for the framework genes BAT1 and TCF19 as hit by the sequencing (Figures 5.5.1 - 5.5.5). Several other gene trees for the evolution of genes found in the Caja sequence TS1 are shown in the chapter 5.6 as TS1.

When selecting the data for computer calculations, we tried to insist on the principle of comparing only homologous parts of the sequences, both of *Calithrix* and other animals. Sequences were aligned with the help of the Clustal X program (Thompson *et al.*, 1994) and thereupon used to make a phylogenetic tree including other animals that had shown identity in BLAST. The trees were constructed by the neighbor-joining method (Saitou and Nei, 1987) using the MEGA version 3.1 program (Kumar *et al.*, 20001) employing Tamura and Nei´s model. The reliability of each node was assessed by the bootstrap method (Felsenstein, 1985) using 1000 replications.

5.5.1 Gene trees of class I genes

The sequence of K1:120D3Caja-G:45 and the sequences of Caja-G 01, 03, 04 and 05 as reported by Cadavid *et al.* (1997) are compared. The tree shows that Caja-G´s genes formed a cluster. One gene the sequence K1-120D3-CajaG45, seems to represent the most ancient branch, closest to Caja-G4. It fell outside of this cluster during their evolution.

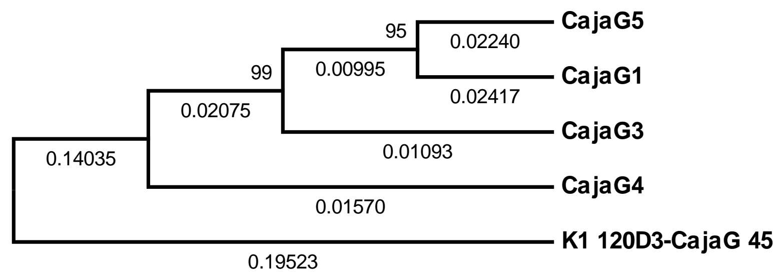


Figure 5.5.1.1: Gene tree for different alleles of Caja-G genes = HLA genes of *Callithrix jacchus* and the sequence K1_120D3-CajaG_45.

The tree is based on numbers of nucleotide substitutions per site, constructed by the neighbor-joining (NJ) method. Numbers without decimals represent the percentage of bootstrap samples supporting the node using the maximum-parsimony method.

Caja-G4 is closest to HLA-G suggesting the greatest structural similarity between these two orthologoues genes. A major cluster is separated from HLA-F, a HLA-pseudogene having evolved in early HLA gene evolution. (Figure 5.5.1.2 and also 5.6.2.3). Upon alignment of DNA sequences, HLA genes show very high levels of identity with Caja-G. Therefore, the choice of Caja-G plausibly provided a probe hybridizing all HLA genes.

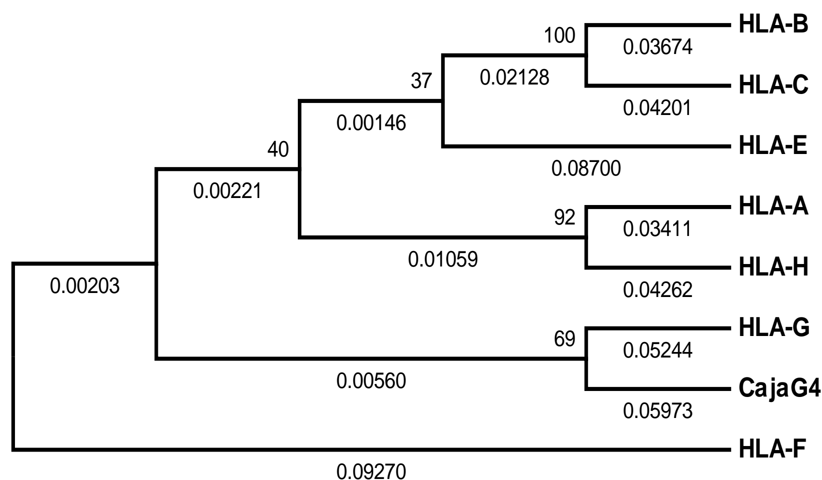


Figure 5.5.1.2: Gene tree for HLA genes and Caja-G4 of *Callithrix jacchus*.

The tree is based on numbers of nucleotide substitutions per site, constructed by the neighbor-joining (NJ) method. Numbers without decimals represent the percentage of bootstrap samples supporting the node using the maximum-parsimony method.

5.5.2 Gene trees of framework genes

The topology of this tree of an orthologoues gene, essential and, therefore, very much conserved, shows the same evolutionary branching as seen for the following framework gene TCF19.

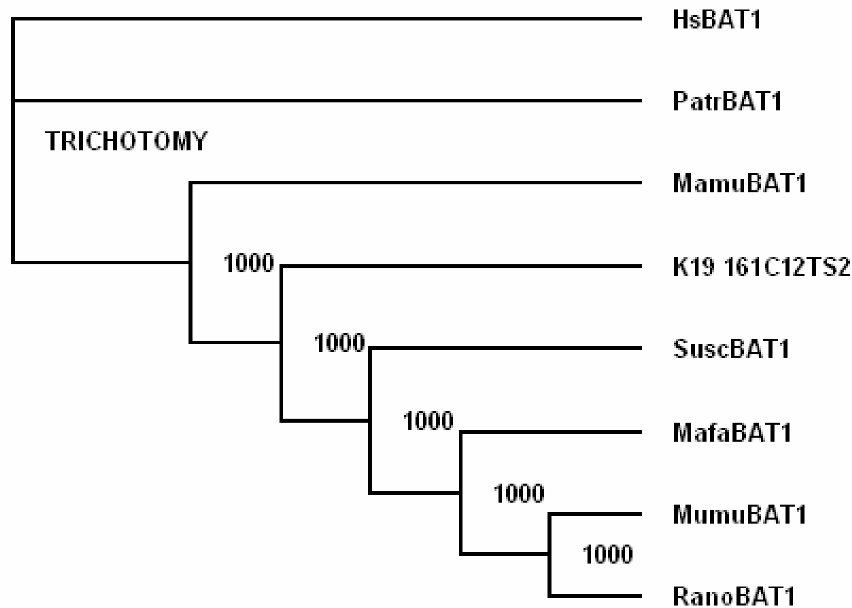


Figure 5.5.2.1: Phylogenetic tree of the framework gene BAT1 of the sequence K19 161C12-TS2 and other mammals.

Bota: *Bos Taurus*, Mumu: *Musculus musculus,* Rano: *Rattus norvergicus,* Patr: *Pan troglodytes* and Hs: *Homo spiens*. The tree is based on numbers of nucleotide substitutions per site, constructed by the neighbor-joining (NJ) method. Numbers without decimals represent the percentage of bootstrap samples supporting the node using the maximum-parsimony method.

Figure 5.5.2.2 contains the sequence K28-347D1TCF19-10 and the sequence of TC19 gene in other species. The tree shows that between this species have coexisted long enough for concerted evolution.

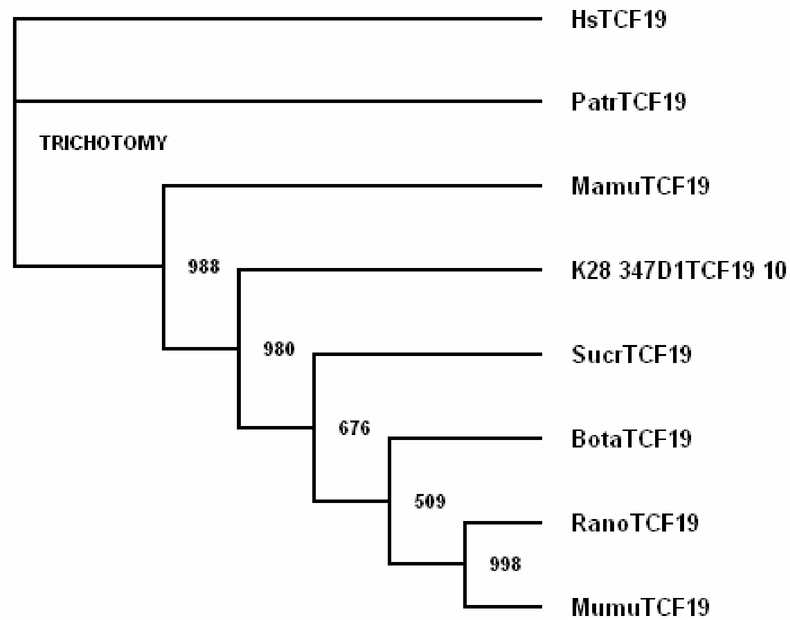


Figure 5.5.2.2: Phylogenetic tree of the framework genes TCF19 of the sequence K28347D1TCF19-10 and other mammals.

Note this tree was constructed with amino acidic sequences. Sucr: *Sus scrofa*, Mamu: *Macaca mulatta,* Patr: *Pan troglodytes,* Hs: *Homo spiens,* Mumu: *Musculus musculus,* Rano: *Rattus norvergicus* and Bota: *Bos Taurus*. The tree is based on numbers of nucleotide substitutions per site, constructed by the neighbor-joining (NJ) method. Numbers without decimals represent the percentage of bootstrap samples supporting the node using the maximum-parsimony method.

It seems easy to obtain with the programs used a branching order of trees for the evolution of genes and their (partial) DNA sequences. The phylogenetic results shown here seem to be in agreement with the overall phylogeny of the species considered - exept, however, that rodents Mumu and Rano are closer to primates than Bota and Sucr. The general phylogenetic order like needs not to be repeted in the evolution of every individual gene: a justification for doing trees.

5.6 TS1: teachings and challenges

Dr. Takashi Shiina, when visiting our lab in autumn of 2004, had chosen from my selection of BAC clones the clone K19-161C12 for a pilot sequencing in the MHC of *Callithrix jacchus*. In January 2005 he sent already two sequences, approx. 3000 bp each, to Dr. Lutz Walter, my supervisor at the time. He gave those to me. Later we called them TS1 and TS2 so we could distinguish them from the 53 sequences that I had done by myself.

Whereas TS2 was assigned to a fragment of BAT1, TS1 turned out to be a quite unusual sequence containing within 2994 bp for four different pseudogenes. These are not expressed any more, as we think, but may be of value for the study of evolution, i.e. of clustering by duplication, of regulatory sequences like promotor regions, and of occasional transposon insertion, and of a consensus sequences possibly important for DNA replication.

5.6.1 A gene map for TS1, and for homologous genomic vicinities in the duplicons B and C of Hs

To study the load of information contained in TS1 was not easy. Written in both directions and containing a ferritin pseudogene probably never seen in any MHC, it had to be taken apart into its natural parts that are arranged as shown in the gene map worked out by alignment with the two homologous regions in Hs DNA, one near HLA-B, the other near HLA-C, i.e. within the duplicons B and C, respectively, see Figure 5.6.1. Clustal X aligned the two class I pseudogenes called TS1-ps1 and -ps2 at very high identity. But furtheron through to the end of the sequence, it showed nothing but random base pairings, likewise in all primates and other mammals' DNA tested. Clustal X would just not recognize the fact that behind a non-aligning insertion of more than 950 bp was located an at least formal continuation of ps2, a highly identical TS1 sequence called v r-c, that on this strand is complementary to the so-called "Vorspann" v located next to the 3' end of the pseudo-ferritin gene(psfth1) on the other side. In aligning, Clustal X does not gap such a long stretch as it routinely does for shorter distances in order to achieve maximum alignment.

The enigma of the upper part was solved, formally at least, by means of the AUGUSTUS program that predicts (parts of) genes by looking for criteria like start and stop signals for translation. It gave a first hint to a Hs ferritin heavy chain gene that upon direct questioning the Hs FTH1 (of the 11th chromosome) did very well

align with TS1 entered "reverse-complementary" thus detecting a ferritin pseudogene psfth1 of *Callithrix jacchus*, see Figure 5.6.3.1.

For comparison with TS1, its homologous genomic vicinities in the Hs duplicons B and C are defined in the comparative map, figure 5.6.1. Data are presented according to the class I region map established by T. Shiina and his coworkers (2001).

In conclusion of these experiences: For study of alignment it has proven very useful to take sequences apart, possibly into natural subunits. These were aligned separately as well as joined again like ps1+ps2+v r-c for entry as a query. TS1 thus deprived of its ferritine gene insertion aligns totally, including v r-c, at more or less high identities with homologous DNAs of almost all primates and other mammals tested. No one besides *Callithrix jacchus* seemed to have a homologue of TS1-psfth1.

Note: We must distinguish "reverse-complementary" as a term defining the direction for reading by computer programs. E.g. TS1 must be entered r-c for alignment of ferritin genes, see figure 5.6.3. This commands reading the left hand strand (complementary for the program) reversely. However, genes located on this strand, whereas possible read in opposite direction to those on the right hand strand (read forward in the sense of the computer), are not given a "r-c" as illustrated in figure 5.6.1.1. They may be assigned a r-c, however, as read in alignment of their antisense reversely like v in ps1+ps2+v r-c. Please consider figure 5.6.1.1 and in particular the discussion on possible senses of v and of v r-c. In duplicon C, the v r-c homologue seems to be known as the pseudogene HCGIV-02. This unusual but conceivable double-sense, a rather different complementary-palindromic sense on the two complementary DNA strands, is making, of course, v a difficult example. This sequence v r-c seems even more remarkable as, moreover, a part of it occurs in various locations like in a 3'-terminal part of TS1-ps1 (Figure 5.6.2.5); it might be a matrix-association-region (MAR) or rather some other *essential sequence* as it is conserved even in a ancient pseudogene that other wise is very much altered by mutations (see chapter 5.6.5).

Another controversy must be understood: The Shiina scale base pair numbering for the class I region of the MHC in Hs and other primates runs towards telomere, the Venter genome bp numbering, however, in the opposite direction, throughout the

whole chromosome. Both scales are used to identify all 55 sequences in table 5.4.1.1.

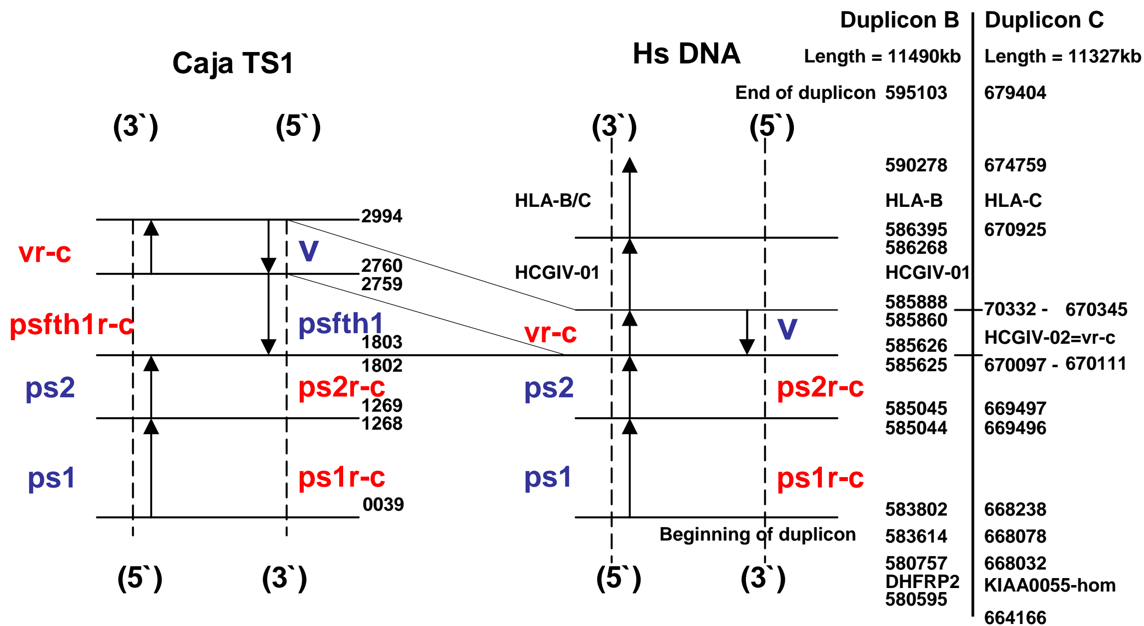
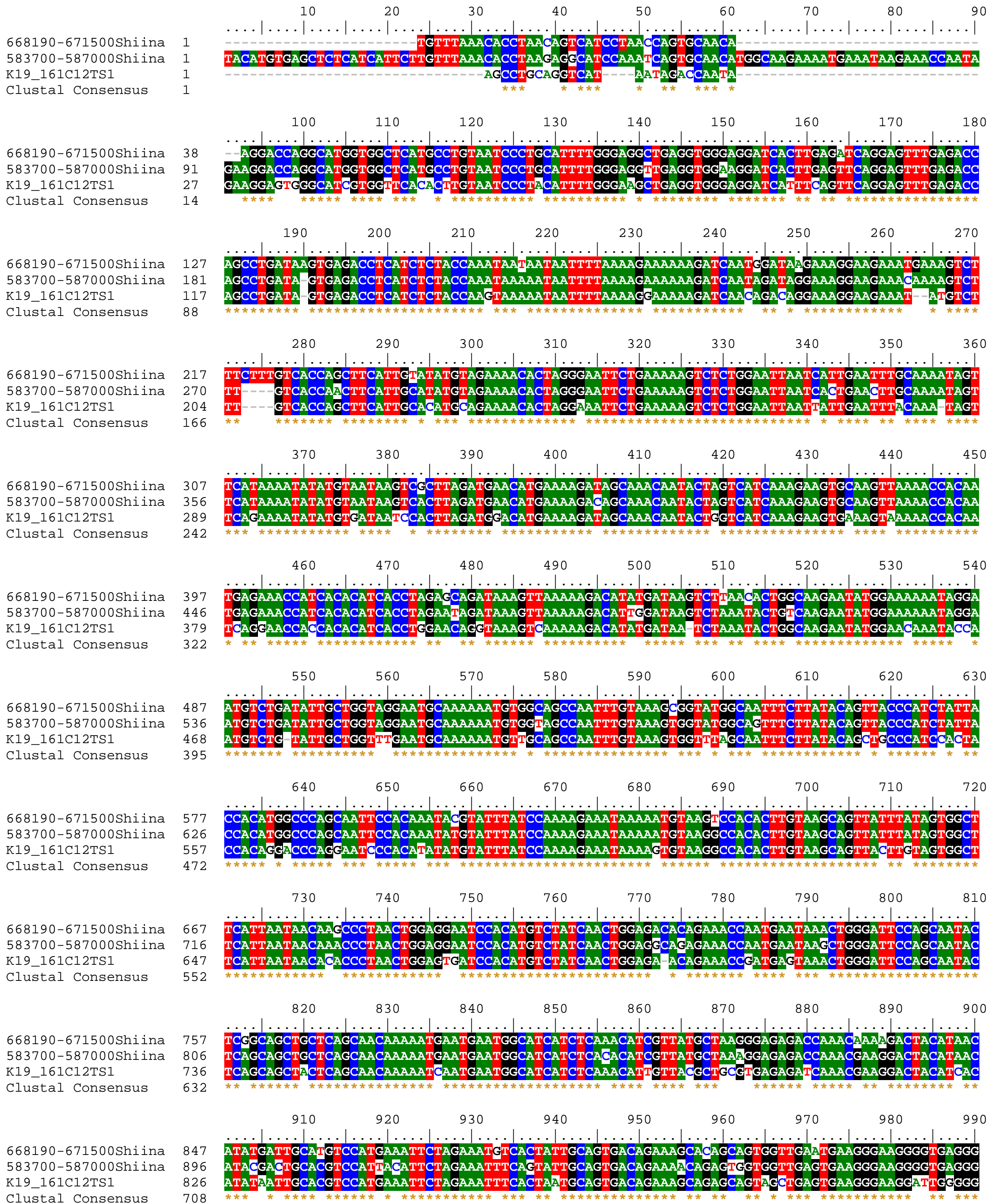


Figure 5.6.1.1: Gene maps for comparison of Caja TS1 with its homologous Hs DNA stretches showing the extension of the two highly identical duplicons B and C in the HLA-B/C section of the class I interval BAT1 – POU5F1. The insertion of a ferritin heavy chain pseudogene (psfth1) in TS1 - as marked by arrows - is reading on the left hand strand with decreasing Shiina bp numbers, i.e. opposite to the new class I pseudogenes TS1-ps1 and TS1-ps2. Therefore, psfth1 runs towards centromere. v (for Vorspann) is a preliminary lab name. v could be a sort of (pseudo)5'FR next to which the transposon psfth1 might have been inserted in *Callithrix* (or a near ancestor). Alternatively, v r-c could be just the 5' terminal part of a longer ps2 dissected for the insertion. In any case, the Caja sequence v read reverse – complementarily (r-c), forward on the right and strand, is almost completely coincident and of very high identity with the known Hs pseudogene HCGIV-02 (figure 5.6.4.1), located ub duplicon C, from which TS1 seems to have been copied according to a higher identity in single that is considered to be part of duplicon C. a5'-terminal part of v r-c aligns at high identity in ps1 and other locations in and outside class I, thus suspicion of an essential consensus sequence (see Figure 5.6.2.5 and chapter 5.6.5).

5.6.2 The two new class I pseudogenes present in the HLA-B and -C duplicons of Hs, likewise - not shown - of and Patr, other mammals Mamu, i.e. its 14 B duplicons



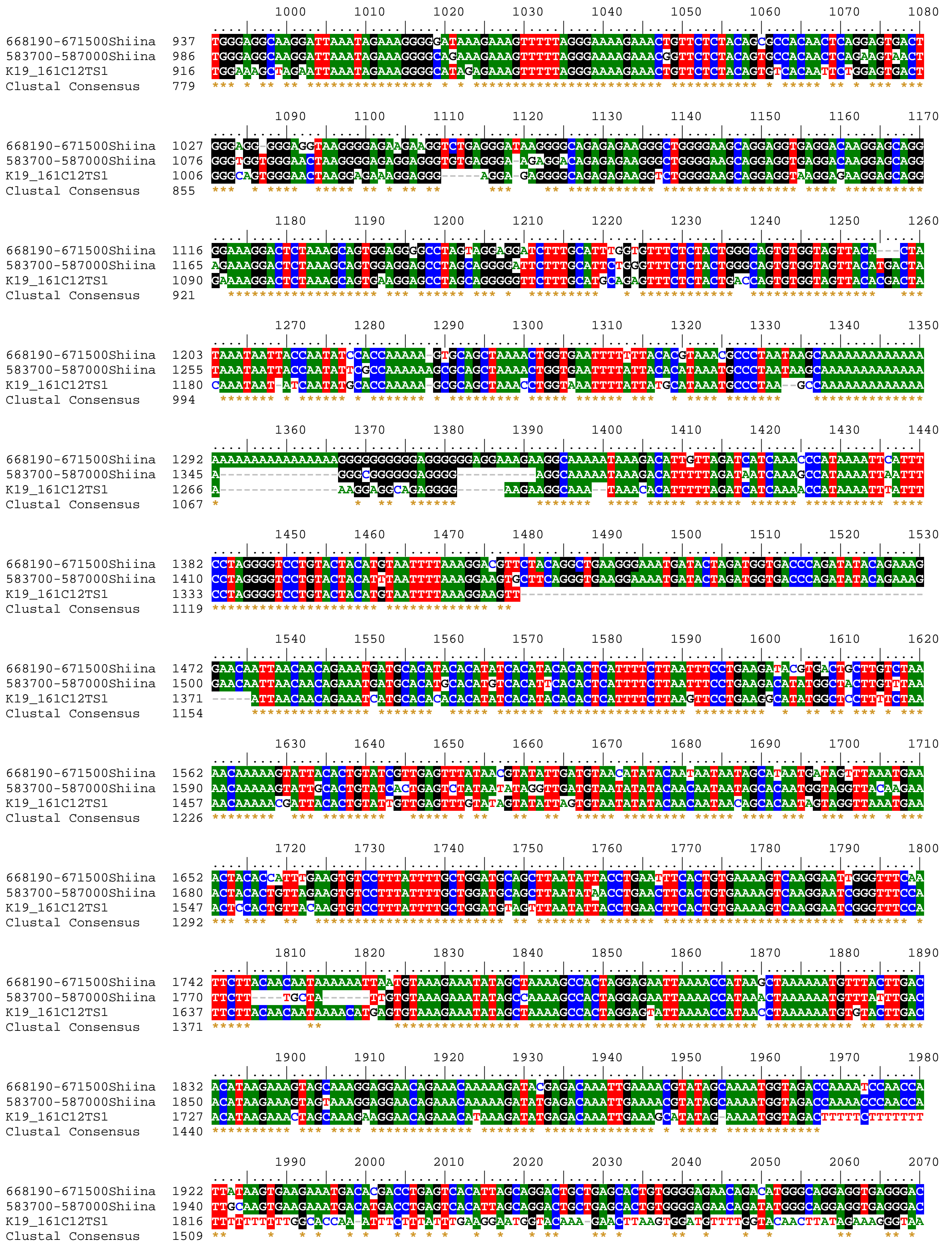


Figure 5.6.2.1: Alignment of TS1, -pseudogenes ps1 and ps2, with their homologous DNA sequences in Hs duplicons B (583700-) and C (668190-).

Alignment has been performed by Clustal X. The nucleotide sequence of Hs to duplicon B and C has ID: BA000025. TS1-ps1 is one of the rarer *complete* class I pseudogenes. It shows a remainder of its former start codon changed to ATC (the Hs ps1's homologues in both duplicons still have their ATG) and a poly A tail. TS1-ps2 seems to be a fragment pseudogene, possibly interrupted by the insertion of a pseudoferritin gene (psfth1) as shown in Figure 5.6.3.1.

TS1-ps1- formally translated 5' to 3' in RF 1 5'3' Frame 1

X X P A G H N R P I E G V G I V V H T C N P Y I L G S **Stop** G G R I I S V Q E F E T S L I V R P H L Y Q V K I I L K G K R S T D R K G R N **Met** S L S P A S L H **Met** Q K T L G N S E K V S G I N Y **Stop** I Y K **Stop** F R K Y **Met Stop Stop** S T **Stop Met** D **Met** K R **Stop** Q T I L V I K E V K V K T T I R N H H T S P G T G K V K K T Y D N L N T G K N **Met** E Q I P **Met** S V L L V **Stop Met** Q K **Met** L Q P I C K V V **Stop** Q F L I Q L P I H Y H R T Q E S H I Y V F I Q K K **Stop** K C K A T L V S S Y L **Stop** W L H **Stop Stop** H T L T G V I H **Met** S I N W R T E T D E **Stop** T G I P A I L S S Y S A T K I N E W H H L K H C Y A A **Stop** E I K R R T T S H I **Met** H V H E I L E I S L **Met** Q **Stop** Q K A E Q **Stop** L S E G K D W G W K A R I K **Stop** K G A **Stop** R K F L G K R N C S L Q C H N S G V T G Q W E L R R K E G G E G Q R E G L G K Q E V R R R S R E K D S K A V K E P S R G F F A C R V S L L T S V V V T R L Q I I S I C T K K R S **Stop** T W **Stop** I L L C I N A L S Q K K K K

TS1-ps1- 5'3' Frame 2

X S L Q V I I D Q **Stop** K E W A S W F T L V I P T F W E A E V G G S F Q F R S L R P A **Stop Stop Stop** D L I S T K **Stop** K **Stop** F **Stop** K E K D Q Q T G K E E I C L C H Q L H C T C R K H **Stop** E I L K K S L E L I I E F T N S S E N I C D N P L R W T **Stop** K D S K Q Y W S S K K **Stop** K **Stop** K P Q S G T T T H H L E Q V K S K R H **Met** I I **Stop** I L A R I W N K Y Q C L Y C W F E C K K C C S Q F V K W F S N F L Y S C P S T T T G P R N P T Y **Met** Y L S K R N K S V R P H L **Stop** A V T C S G F I N N T P **Stop** L E **Stop** S T C L S T G E Q K P **Met** S K L G F Q Q Y S A A T Q Q Q K S **Met** N G I I S N I V T L R E R S N E G L H H I **Stop** C T S **Met** K F **Stop** K F H **Stop** C S D R K Q S S S **Stop** V K G R I G G G K L E L N R K G H R E S F **Stop** G K E T V L Y S V T I L E **Stop** L G S G N **Stop** G E R R E E R G R E K V W G S R R **Stop** G E G A G K R T L K Q **Stop** R S L A G G S L H A E F L Y **Stop** P V W **Stop** L H D Y K **Stop** Y Q Y A P K S A A K P G K F Y Y A **Stop Met** P **Stop** A K K K K K

TS1-ps1- 5'3' Frame 3

X A C R S **Stop Stop** T N R R S G H R G S H L **Stop** S L H F G K L R W E D H F S S G V **Stop** D Q P D S E T S S L P S K N N F K R K K I N R Q E R K K Y V F V T S F I A H A E N T R K F **Stop** K S L W N **Stop** L L N L Q I V Q K I Y V I I H L D G H E K I A N N T G H Q R S E S K N H N Q E P P H I T W N R **Stop** S Q K D I **Stop Stop** S K Y W Q E Y G T N T N V C I A G L N A K N V A A N L **Stop** S G L A I S Y T A A H P L P Q D P G I P H I C I Y P K E I K V **Stop** G H T C K Q L L V V A S L I T H P N W S D P H V Y Q L E N R N R **Stop** V N W D S S N T Q Q L L S N K N Q **Stop Met** A S S Q T L L R C V R D Q T K D Y I T Y N A R P **Stop** N S R N F T N A V T E S R A V A E **Stop** R E G L G V E S **Stop** N **Stop** I E R G I E K V F R E K K L F S T V S Q F W S D W A V G T K E K G G R R G A E R R

S G E A G G K E K E Q G K G L **Stop** S S E G A **Stop** Q G V L C **Met** Q S F S T D Q C G S Y T T T N N I N **Met** H Q K A Q L N L V N F I **Met** H K C P K P K K K K K

Figure 5.6.2.2: Formal amino acid (aa) translations for TS1-ps1 as given by Swiss-prot in three reading frames (RF) from the (mRNA) 5'-end.

RF I is starting at third, RF II at first, RF III at second base of the DNA sequence (shown in the Annex). In each of these RF, many mutations have spoiled a potential original sense, as recognised by many Stop and Met codons ruining potential coding for polypeptides of any useful (and meaningful) length. Moreover, chances for transcription and translation are destroied by mutation of signals required for initiation of transcription in the 5'FR (TATA box, CAG and other signals). Consequent non-translatability by itself has removed natural selection pressure against deleterious mutations in these former genes, whereas general mechanisms for conservation of unchanged genetic information have kept such molecular fossils still good enough for investigation of the evolution of MHC class I molecules. Moreover, there may be still some essential sequence(s) that even DNA of pseudogenes has to maintain, as observed and discussed below (Figure 5.6.2.5 and chapter 5.6.5).

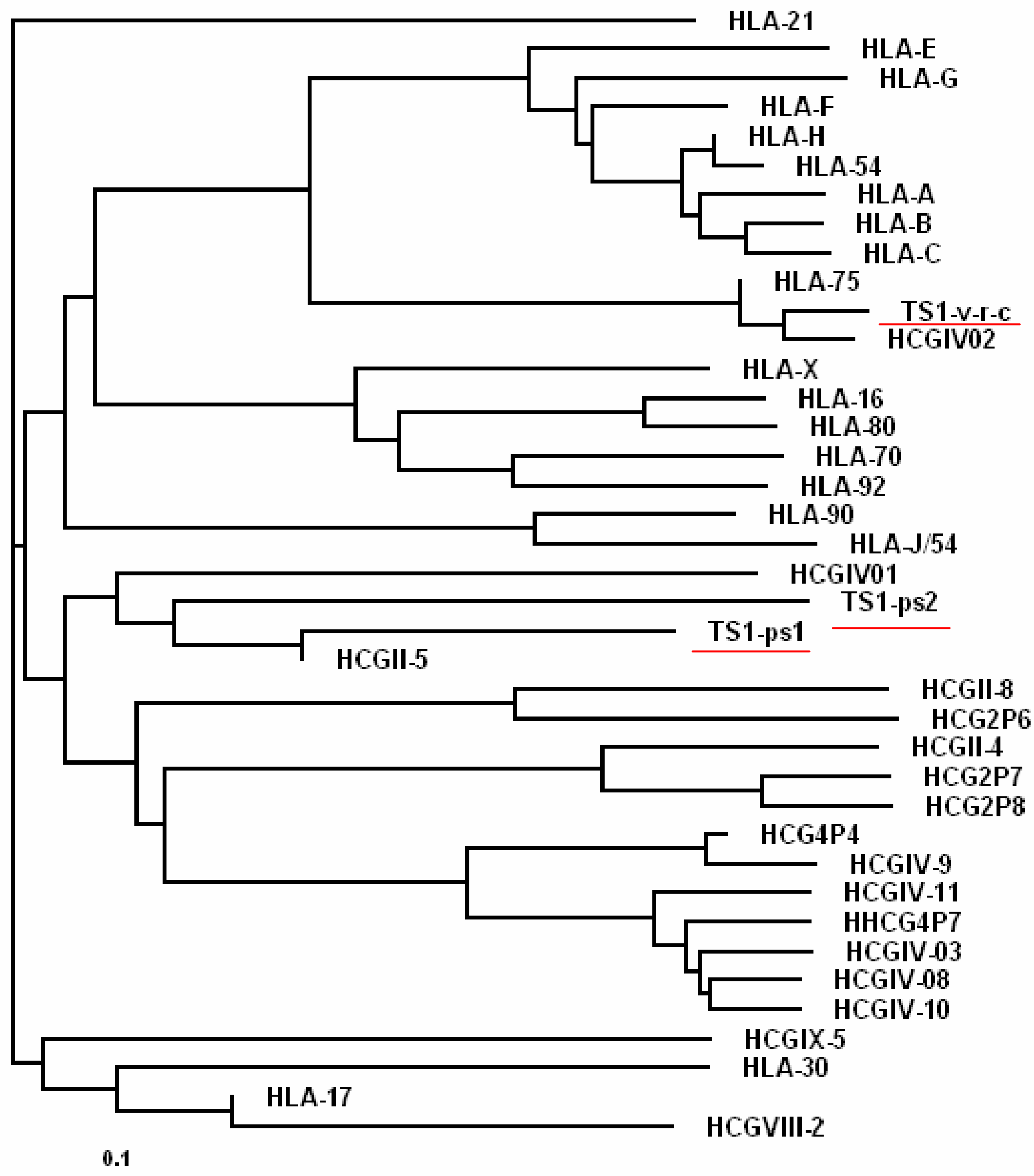


Figure 5.6.2.3: Evolutionary tree considering the classical and non-classical HLA genes and the class I pseudogenes of Hs giving a chance to the Caja TS1-ps1, -ps2 and -v r-c for finding their place within the known clusters.

Con-clustering of TS1-v r-c together with HCGIV-02 is expected by their nearly identical sequences. Their showing up in the HLA class Ia and Ib-cluster, however, may be worth further study, in particular with respect of 5'FR. TS1-ps1 and -ps2 seem most related to HCGIV-01 related to the clusters II and IV. This result must be studied in more detail, looking carefully at alignment of different parts, segments of the sequences. It has been found that three thirds of TS1-ps1 receive different classifications when aligned separately and evaluated for the tree. The tree is based on numbers of nucleotide substitutions per site, constructed by the neighbor-joining

(NJ) method. Numbers without decimals represent the percentage of bootstrap samples supporting the node using the maximum-parsimony method.

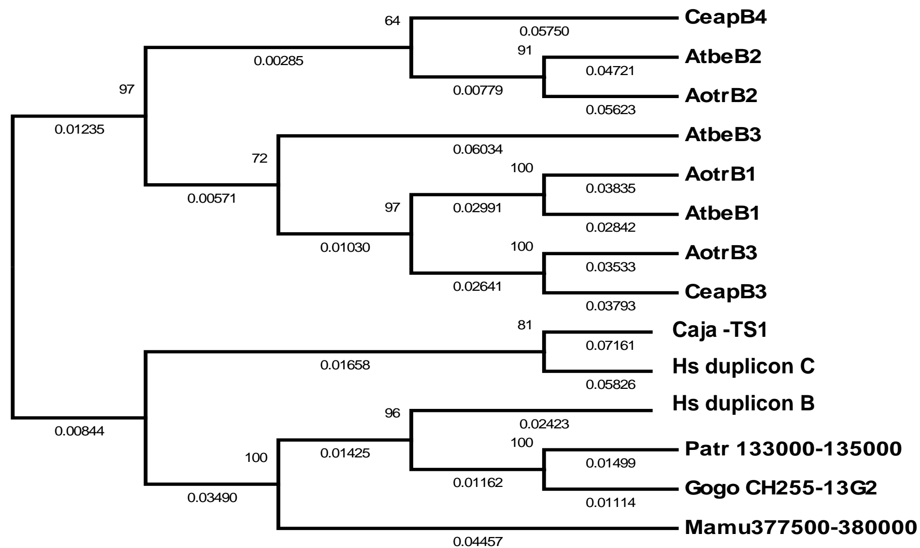


Figure 5.6.2.4: Phylogenetic tree of the sequence K19-161C12-TS1 and its homologues in Hs and other primates.

Note the tree was constructed with amino acidic sequences obtained formal translations by Swiss-prot. Atbe: Ateles belzebuth, Aotr: *Aotus trivirgatus*, Ceap: *Cebus apella*, Patr: *Pan troglodytes,* Gogo: *Gorilla gorilla* and Mamu: *Macaca mulatta.* The tree is based on numbers of nucleotide substitutions per site, constructed by the neighbor-joining (NJ) method. Numbers without decimals represent the percentage of bootstrap samples supporting the node using the maximum-parsimony method.

This preliminary study of relationships between TS1-ps1+ps2+v r-c homologous sequences in the B duplicon of various primates by a phylogenetic tree confirms B duplicon in all primates tested. Surprisingly, TS1 homologous Hs duplicon B DNA turns out to be nearest to a phylum of Patr, Gogo and Mamu (represented by for its B4 duplicon); Hs duplicon C DNA, however, is nearest to TS1-ps1+ps2+v r-c. This is a third piece of evidence obtained in favor of TS1 stemming from a Caja C rather a than B duplicon.

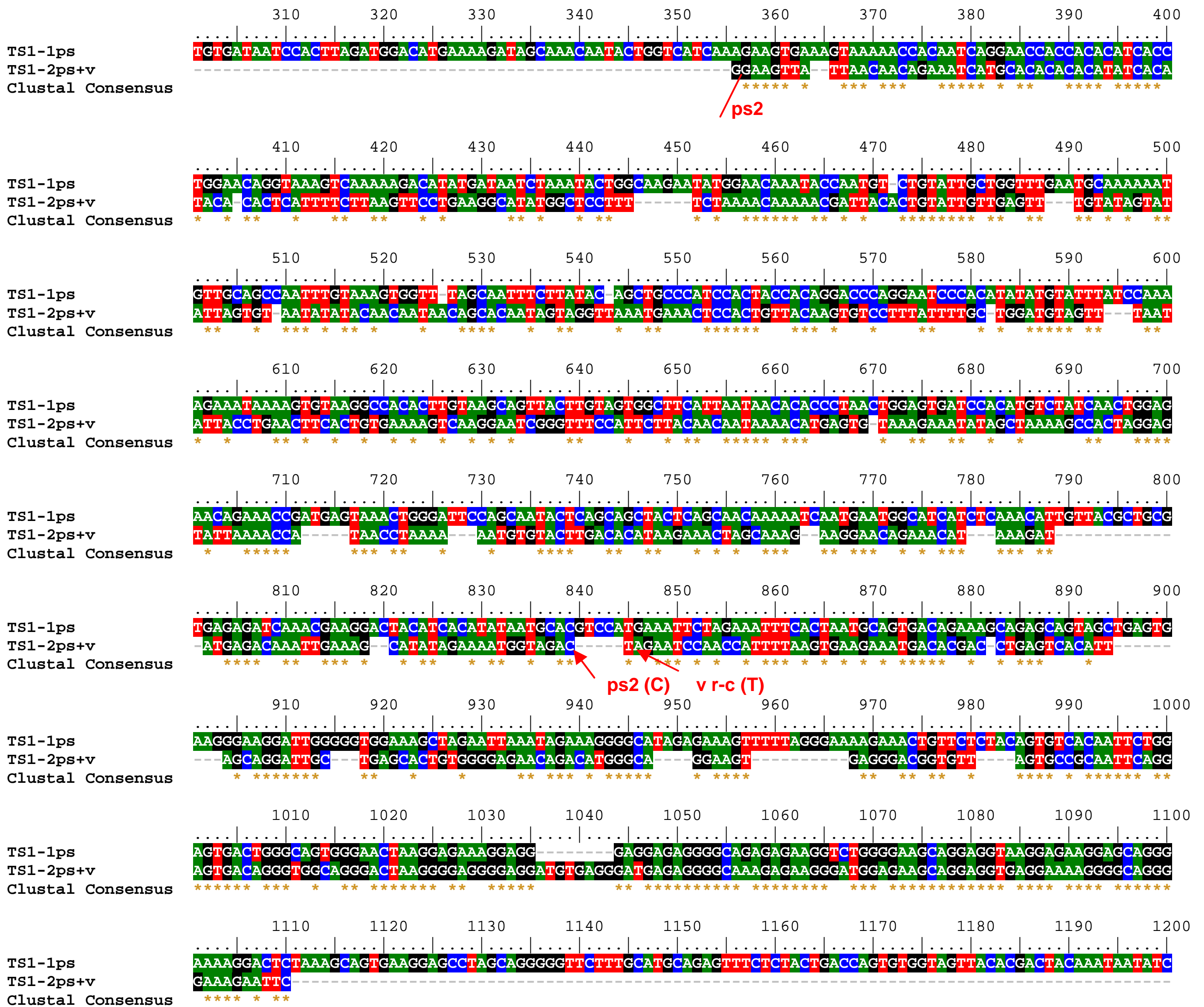


Figure 5.6.2.5: Observation of high identity alignment of the 5'-terminal part of TS1-v r-c with a section 5' terminal end, near the TS1-ps1.

Before this stretch, TS1-ps2 and a 3' - terminal part of v r-c, also aligned for comparison, show a rather normal random distribution of base identities. Alignment has been performed between TS1-p1, TS1-2ps+v by Clustal X.

TS1-ps1 and -ps2 were aligned also with each other. Their degree of identity might be slightly above expectation for a random distribution. It should be calculated for many ps1 and ps2 homologues apparent in class I duplicons of primates.

Quite safe, however, seemed a high identity between alignment of TS1-ps1 and TS1-v r-c. A part of v r-c, bases 100 to 235 on the TS1 map, showed 59 % identity with bases 985 to 1110 of TS1-ps1, i.e. near the 5' terminal end of ps1. This observation

contrasts to the many mutations as seen e.g. by stop and Met codons in most parts of the formal translations of the no more expressed pseudogene (Figure 5.6.2.2).

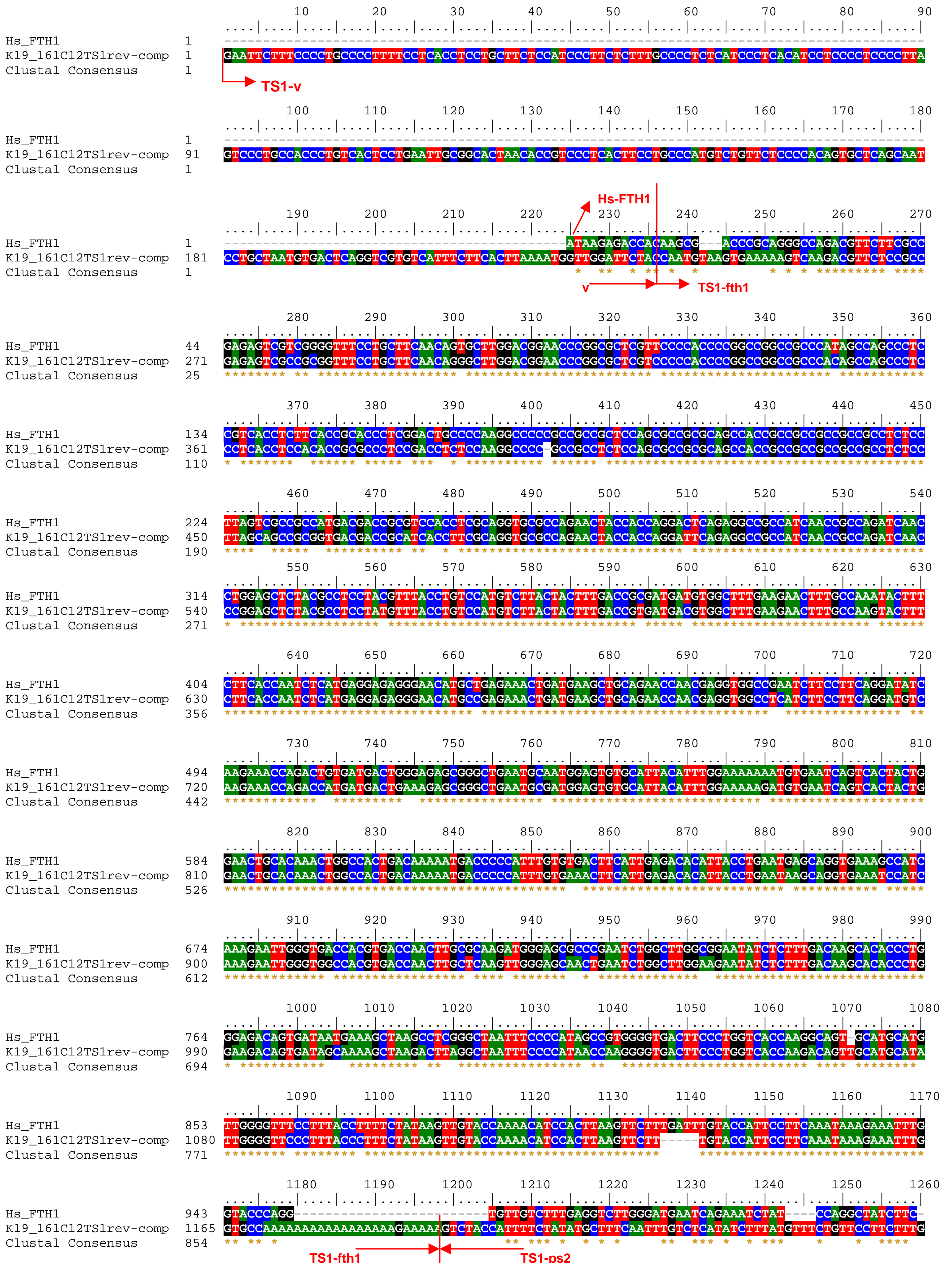
5.6.3 A ferritin heavy chain pseudogene in the MHC of *Callithrix jacchus*

All living cells need apoferritin, its light and heavy chain protein subunits, 24 of which (at different ratios in different tissues) combine to form a package for storage of 4500 Fe^{2+} ions. Each subunit's tertiary structure contains 4 α-helices composed of many very conservative amino acids. Primates and other mammals usually have one expressed ferritin gene from which during evolution have descended altogether a few pseudo ferritin genes, six heavy chain pseudogenes in Hs (Constanzo *et al.*, 1986).

We have taken a closer look at sequence and evolution of the pseudoferritin gene TS1-psfth1 found and localized as shown in the map, figure 5.6.1.1. TS1-psfth1 seems unique for the MHC as far as we can enquire in sequences known today. It was not found in Hs nor any other primate or mammal. For many animals, mRNAs have been sequenced only no complete DNA as is needed for further investigation.

We shall look how psfth1 aligns with the human FTH1 gene as seen in figure 5.6.3.1. Thereafter, we see both aligned together with all Hs heavy chain pseudogenes. A gene tree will suggest then a place in evolution for psfth1 (Figure 5.6.3.2). On the protein level, where mutated genes look more conserved compared than in their nucleotide sequence, we look back from Hs and Caja over a selection of animals down to *Caenorhabditis*. So we go far below the first animals having an MHC i.e. teleosts (*Dare* is included). If psfth1 would be such an old transposon we could not possibly trace it to such an old ancestor, because of no complete DNA sequences being available, also of insufficient knowledge on early immune proteins and MHC organizations.

Figure 5.6.3.2 shows a very high degree of identity particularly in the large middle part of the molecules most of which were coding for the α-helical parts of ferritin once, i.e. in the still expressed evolutionary state of the ferritin gene at the time of their origin. Note, only the expressed gene H11 = FTH1 has a TATAA box, the CAG signal for transcription start, and alike some of the pseudogenes and TS1-psfth1, the 28 nucleotide signal in the 5'FR before the translation start, ATG. The 28 nucleotide signal is nearly identical in Hs, rat and several other mammals including Caja (Murray *et al.*, 1987).


Hs_FTH1
K19_161C12TS1rev-comp
Clustal Consensus
TS1-v
Hs-FTH1
TS1-fth1
TS1-ps2

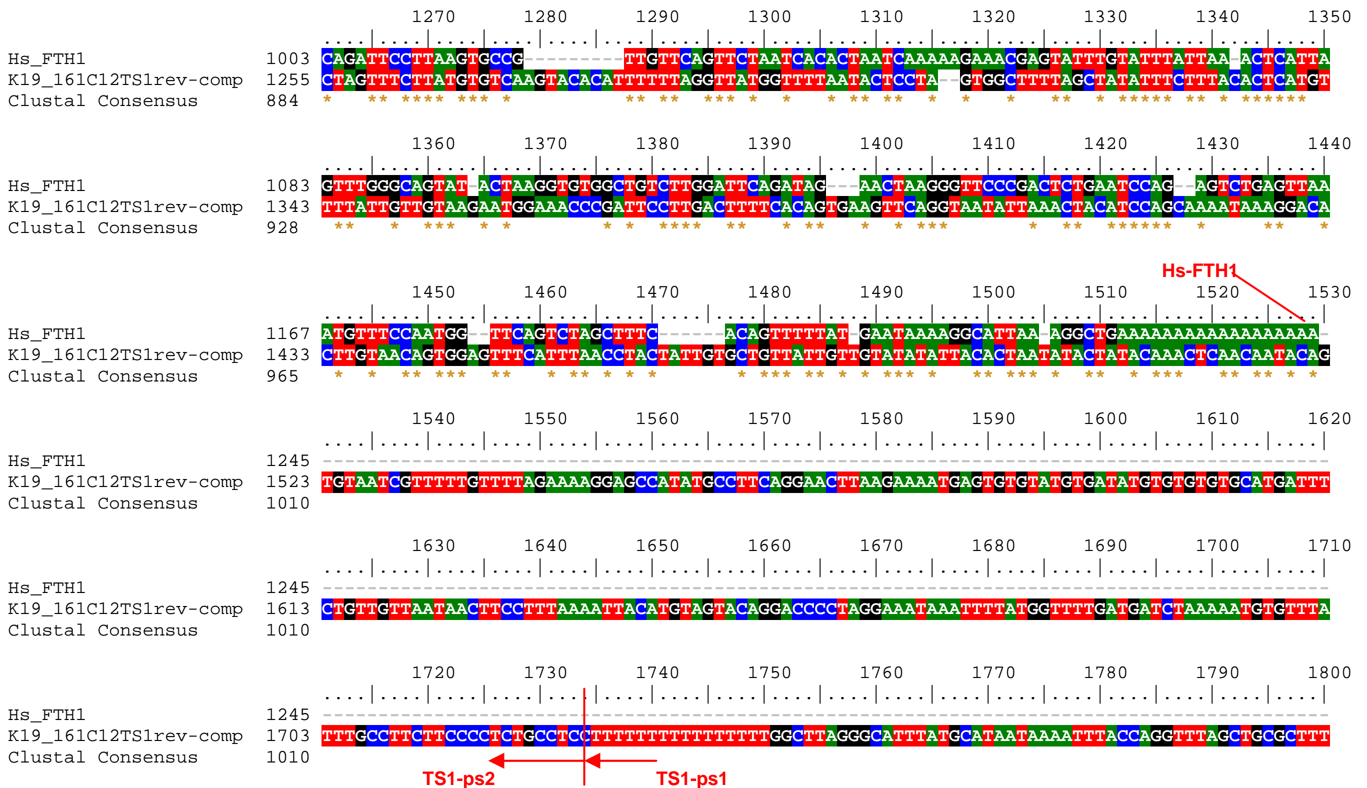
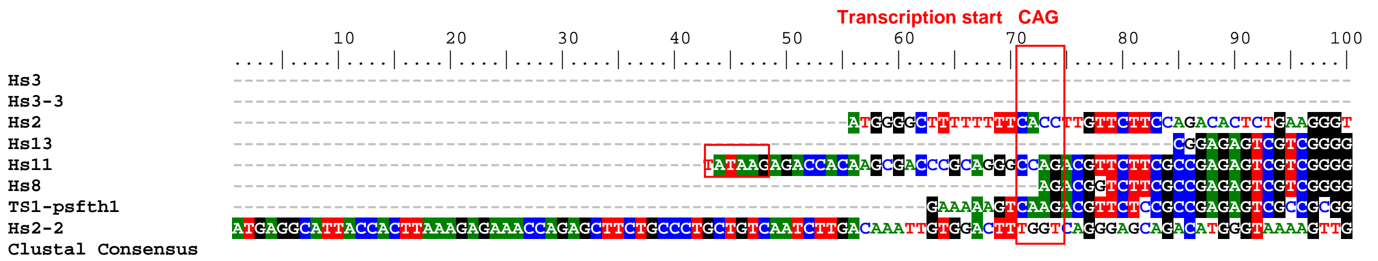
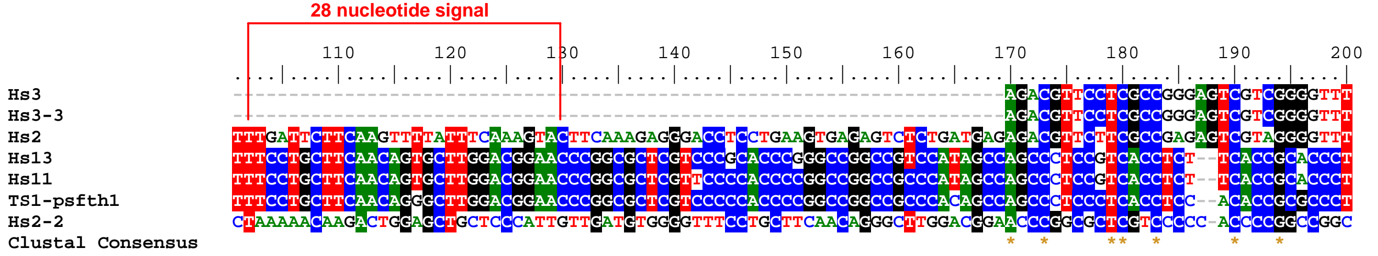
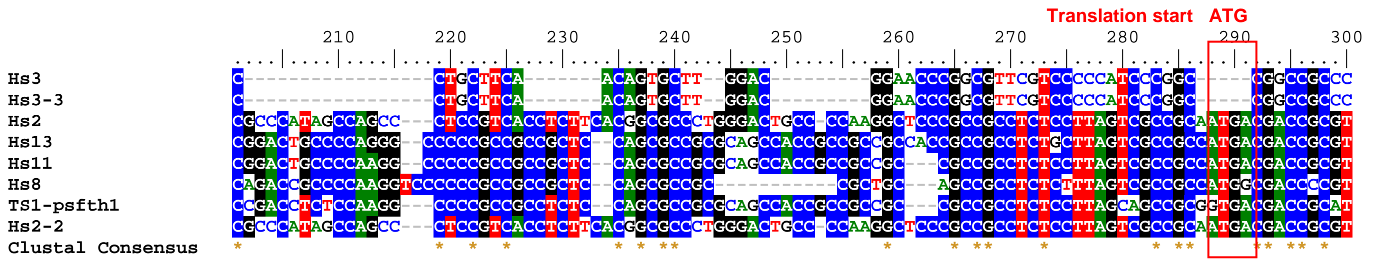


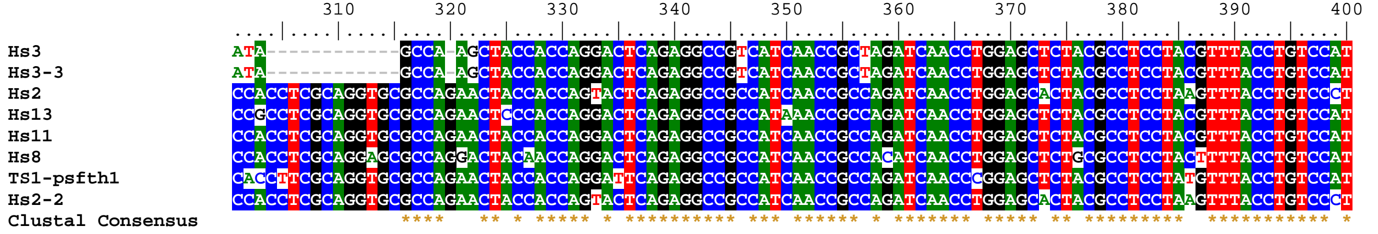
Figure 5.6.3.1: Alignment of TS1 psfth1 with the only expressed Hs ferritin gene FTH1 located on chromosome 11.

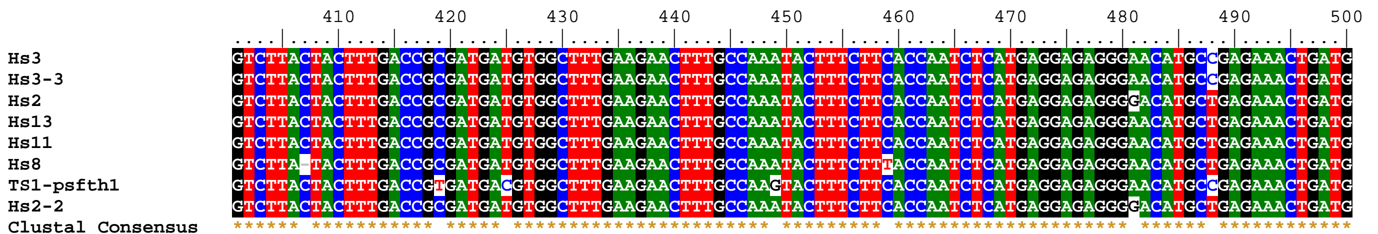
TS1 entered r-c, at its 3' end shows v as known from figure 5.6.1.1. TS1-psfth1 aligns with FTH1, in their beginnings with ± random identity, from bp 257 on this reverse scale, however, at very high identity up to shortly before beginning of as through to the poly A tail of TS1-psfth1. Thereafter, Hs FTH1 shoes only random identity, because DNA contains ps2 there (on its complementary strand). Alignment has been performed between K19_161C12TS1rev-comp and FTH1 gene of Hs (ID: NT033927) by Clustal X.

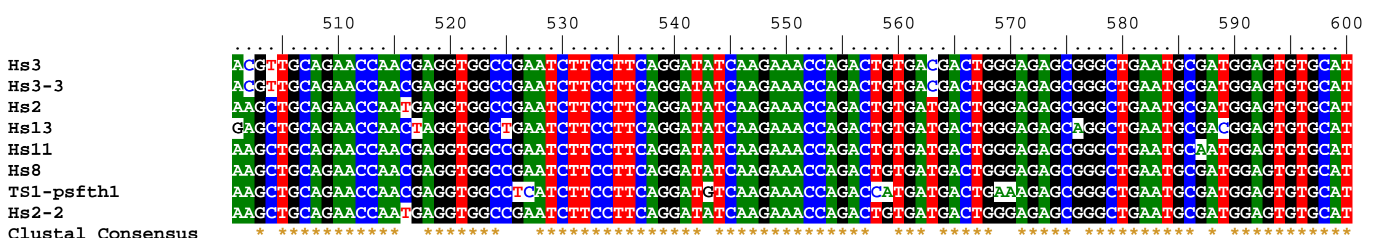

Transcription start CAG
Hs3
Hs3-3
Hs2
Hs13
Hs11
Hs8
TS1-psfth1
Hs2-2
Clustal Consensus

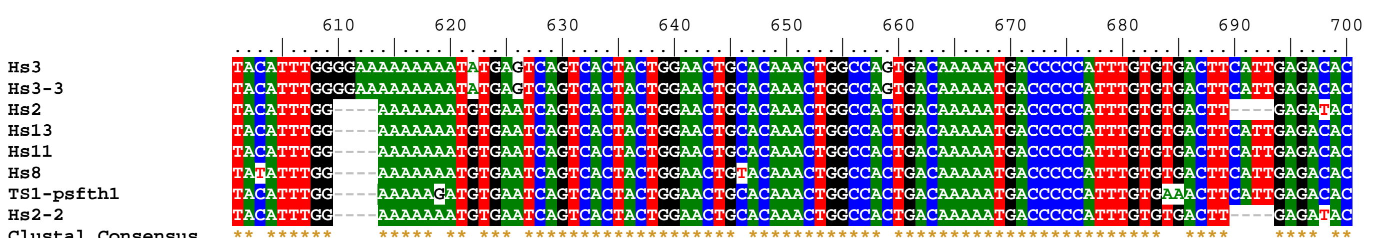

28 nucleotide signal
Hs3
Hs3-3
Hs2
Hs13
Hs11
TS1-psfth1
Hs2-2
Clustal Consensus


Translation start ATG
Hs3
Hs3-3
Hs2
Hs13
Hs11
Hs8
TS1-psfth1
Hs2-2
Clustal Consensus


Hs3
Hs3-3
Hs2
Hs13
Hs11
Hs8
TS1-psfth1
Hs2-2
Clustal Consensus


Hs3
Hs3-3
Hs2
Hs13
Hs11
Hs8
TS1-psfth1
Hs2-2
Clustal Consensus


Hs3
Hs3-3
Hs2
Hs13
Hs11
Hs8
TS1-psfth1
Hs2-2
Clustal Consensus


Hs3
Hs3-3
Hs2
Hs13
Hs11
Hs8
TS1-psfth1
Hs2-2
Clustal Consensus

```
                             710       720       730       740       750       760       770       780       790       800
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hs3                 ATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGGCCACGTGACCAACTTGTGCAAGATGGGAGCACCCGAATCTGGCTCGGCAGAATATCT
Hs3-3               ATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGGCCACGTGACCAACTTGTGCAAGATGGGAGCACCCGAATCTGGCTCGGCAGAATATCT
Hs2                 ATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGACCACGTGACCAACTTGCGCAAGATGGGAGCGCCCGAATCTGGCGTGGGAGAATATCT
Hs13                ATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGATCAAGTGACCAACTTGCGCAAGATGGGAGCACCCGAATCTGGCTTGGCAGAATATCT
Hs11                ATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGACCACGTGACCAACTTGCGCAAGATGGGAGCGCCCGAATCTGGCTTGGCGGAATATCT
Hs8                 ATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGAGTGACCACGTGACCAACTTGCGCAAGATGGGAGCGCCCGAATCTGGCTCGGCAGAATATCT
TS1-psfth1          ATTACCTGAATAAGCAGGTGAAATCCATCAAAGAATTGGGTGGCCACGTGACCAACTTGCTCAAGTTGGGAGCAACTGAATCTGGCTTGGAAGAATATCT
Hs2-2               ATTACCTGAATGAGCAGGTGAAAGCCATCAAAGAATTGGGTGACCACGTGACCAACTTGCGCAAGATGGGAGCGCATTACTTCCA-GAAAAGGCCTGTCA
Clustal Consensus   *********** *********** ************** ***  ** ************  **** *******     * *           *  * **

                             810       820       830       840       850       860       870       880       890       900
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hs3                 CTTTGACAAGCACACGCTGGGAGACAGTGATAACGAAAGCTAAGCCTCAGGCTAATTTCCCCATAGCCATGGGGTGACTTCCCTGGTCACCAAGGCAGT-
Hs3-3               CTTTGACAAGCACACGCTGGGAGACAGTGATAACGAAAGCTAAGCCTCAGGCTAATTTCCCCATAGCCATGGGGTGACTTCCCTGGTCACCAAGGCAGT-
Hs2                 CTTTGACAAGCACACTCTGGGAGACAGTGATAATGAAAGCTAAGCCTCAGGCTAATTTCCCCATAGCCATGGGGTGACTTCCCTGGTCACCAAGGCAGT-
Hs13                TTTTGACAAGCACATCCTGGGAGACAGTGATAATGAAAGCTAAGCCTCAGGCTAATTTCCCCATAGCCGTGGGGTGACTTCCCTGGTCACCAAGGCAGT-
Hs11                CTTTGACAAGCACACCCTGGGAGACAGTGATAATGAAAGCTAAGCCTCGGGCTAATTTCCCCATAGCCGTGGGGTGACTTCCCTGGTCACCAAGGCAGT-
Hs8                 CTTCGACAAGCACACCCTGGGAGACAGTGATAATGAAAGCTAAGTCTCAGGCTAATTTCCCCATAGCCGTGGGGTGCCTTCCTTGGTCACCAAGGCAGT-
TS1-psfth1          CTTTGACAAGCACACCCTGGAAGACAGTGATAGCAAAAGCTAAGACTTAGGCTAATTTCCCCATAACCAAGGGGTGACTTCCCTGGTCACCAAGACAGTT
Hs2-2               ACCTGA-AAATACATGA-GCCAACCAACTATCACATTAAC-ACGATTCAG--TAG---------------------------------------------
Clustal Consensus       ** **  ***    *  *  **   **      * * * *  *  *  **

                             910       920       930       940       950       960       970       980       990      1000
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hs3                 GCATGCATGTTGGGGTTTCCTTTACTTTTTCTATAAGTTGCACCACAACATCCA----------------------------------------------
Hs3-3               GCATGCATGTTGGGGTTTCCTTTACTTTTTCTATAAGTTGCACCACAACATCCA----------------------------------------------
Hs2                 GCATGCATGTTGGGGTTTCCTTTACCTTTTCTATAAGTTGTACCAAAACATCCACTTAAGTTCTTTGATTTGTACCATTCCTTCAAATAAAGAAATTTGG
Hs13                GCATGCATGTTGGGGTTTCCTTTACCTTTTCTATAAGTTGTACCAAAACATCCACTTAAATTCTTTGATTTGTACCATTCCTTCAAATAAAGAAATTTGG
Hs11                GCATGCATGTTGGGGTTTCCTTTACCTTTTCTATAAGTTGTACCAAAACATCCACTTAAGTTCTTTGATTTGTACCATTCCTTCAAATAAAGAAATTTGG
Hs8                 GCGTGCACGTTAGGGTTTCCTTTACCTTTTCTATAAGTTGTACCAAAACATCCACTTAAGTTCTTTGATTTGTACCATTCCT------------------
TS1-psfth1          GCATGCATATTGGGGTTCCCTTTACCCTTTCTATAAGTTGTACCAAAACATCCACTTAAGTTCT-----TTGTACCATTCCTTCAAATAAAGAAATTTGG
Hs2-2               ----------------------------------------------------------------------------------------------------
Clustal Consensus

                            1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
                    ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Hs3                 ----------------------------------------------------------------------------------------------------
Hs3-3               ----------------------------------------------------------------------------------------------------
Hs2                 TACCCAAAAAAAAAAGAGTC--------------------------------------------------------------------------------
Hs13                TACCC-------------------------------------------------------------------------------to polyAAA tail-
Hs11                TACCCAGGTGTTGTCTTTGAGGTCTTGGGATGAATCAGAAATCTATCCAGGCTATCTTCCAGATTCCTTAAGTGCCGTTGTTCAGTTCTAATCACACTAA
Hs8                 ----------------------------------------------------------------------------------------------------
TS1-psfth1          TGCCAAAAAAAAAAAAAAAAAGAAAAA-------------------------------------------------------------------------
Hs2-2               ----------------------------------------------------------------------------------------------------
Clustal Consensus
```

Figure 5.6.3.2: Alignment of TS1-psfth1 with the expressed FTH1 called also Hs11 of Hs, and the pseudogenes derived from it.

There is very high identity particularly in the middle part of the molecules behind the start codon ATG, eventually coding for the 4α-helical parts. But only H11, the Hs FTH1 has TATA-box, CAG trancription start and, like some of the pseudoferritins including Caja TS1-psfth1, the 28 nucleotide signal in the 5'FR before the translation start, ATG. Alignment has been performed between TS-psfth1**,** FTH1 gene, and pseudogenes Fth1 of Hs, arabic number indicates the chromosome which the pseudogene is located (Hs11: FTH1: ID: NT033927, Hs3: ID:NT022459.1, Hs3-3: ID:NW:921651.1 , Hs2: ID:, Hs13:ID:NR002202.1, Hs-8:ID:,Hs2-2:ID:) by Clustal X.

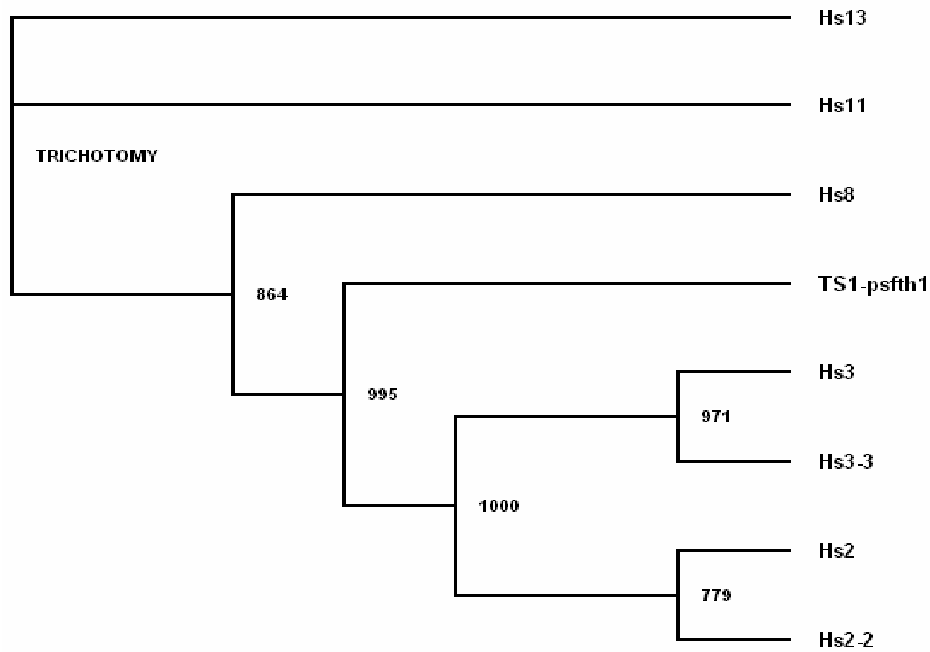


Figure 5.6.3.3: Gene tree for TS1-psfth1, the expressed FTH1 (Hs11) and the Hs pseudogenes derived from it, or a common ancestor.

(Method applied as under figure 5.5.1.1). This tree shows TS1-psfth1 closely related to Hs11, Hs13 and Hs8. Inspection of their 5'- and 3'-ends, however, shows greatest similarity between TS1-psfth1 and Hs2 (Figure 5.6.3.2).

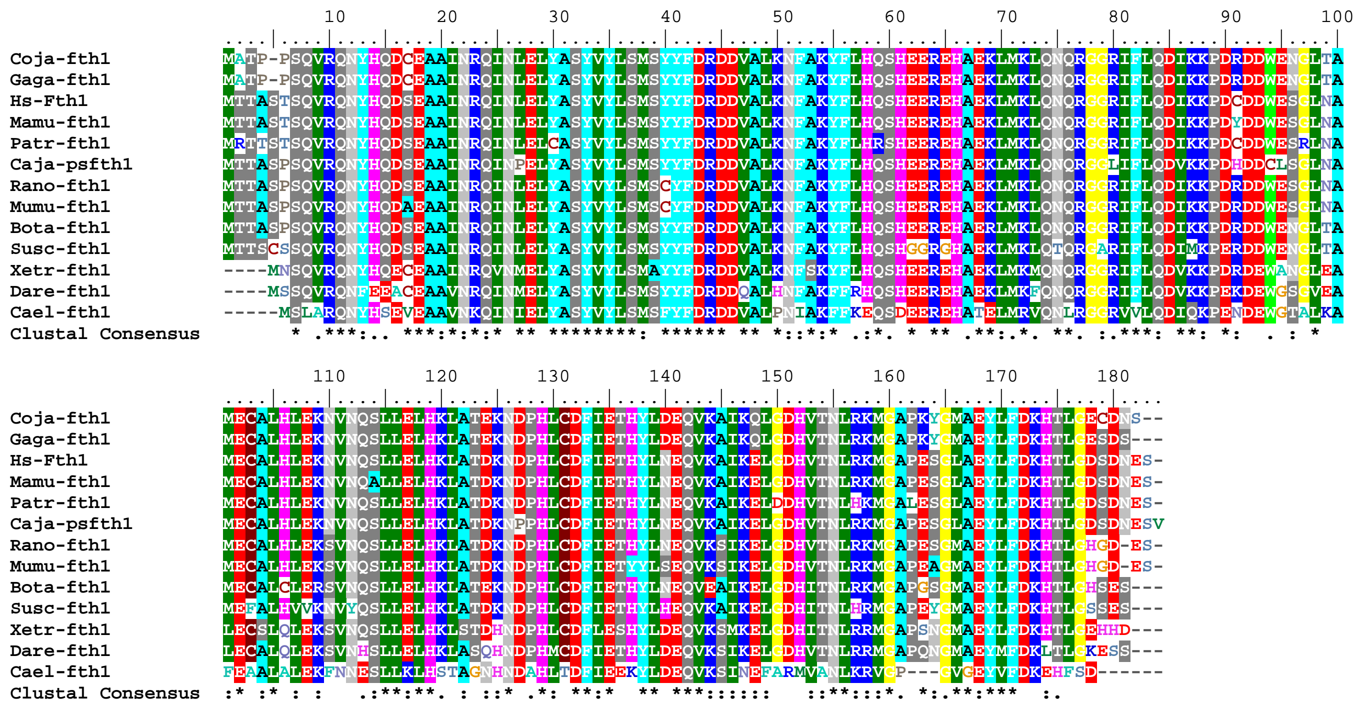


Figure 5.6.3.4: Alignment of selected apoferritin proteins, several species.

As the consensus shows (*), many apparently essential amino acids are conserved during evolution, some less when *Drosophila* apoferritine had been included. Aligning even apoferritins of yeast and bacterium, no amino acids in common to all are seen. Therefore, the ferritin gene of metazoa seems to have an origin different from the one of lower eukaryotes and prokaryotes; their gene product looks quite different. Note: Hs: ID: AAH70494.1, Patr: ID: XP509574.2, Mumu: ID: NP034369.1, Rano: ID: AAH78892.1, Mamu: ID: XP001104405.1, Bota. ID: NP776487.1, Susc:ID: NP999140.1, Coja: *Coturnix japonica* ID: AAT01287.1, Gaga: *Gallus gallus* ID:NP990417.1, Xetr: *Xenopus tropicalis* ID:NP001005135.1, Dare: *Danio rerio* ID:NP571660.1, Cael: *Caenorhabditis elegans* ID:NP491198.1.

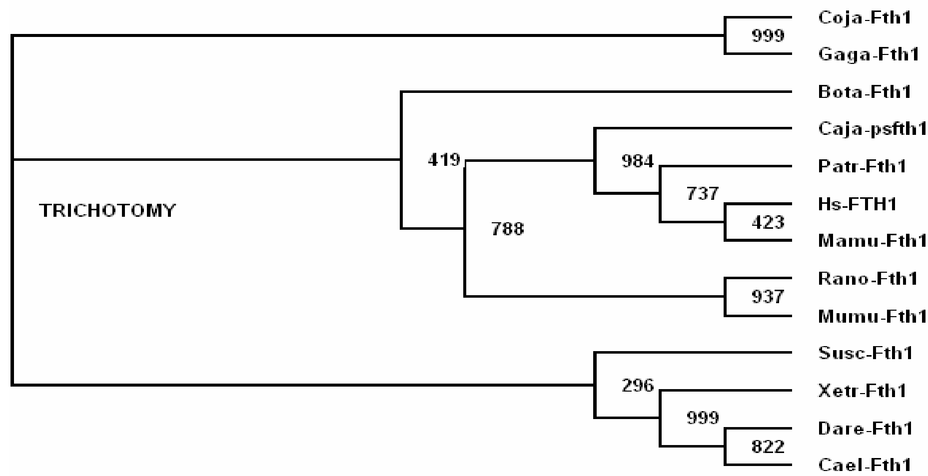


Figure 5.6.3.5: Gene tree for the evolution of animal apoferritins.

Note: the tree was constructed with amino acidic sequences as coded by expressed apoferritin genes; only a Caja pseudoferritin gene as of TS1 was compared in its *formal* amono acid translation produced by Swiss-prot (Method applied as under Figure 5.5.1.1).

5.6.4 Characterization and possible interpretations of "Vorspann" v, the sequence next to which a ferritin transposon has been inserted

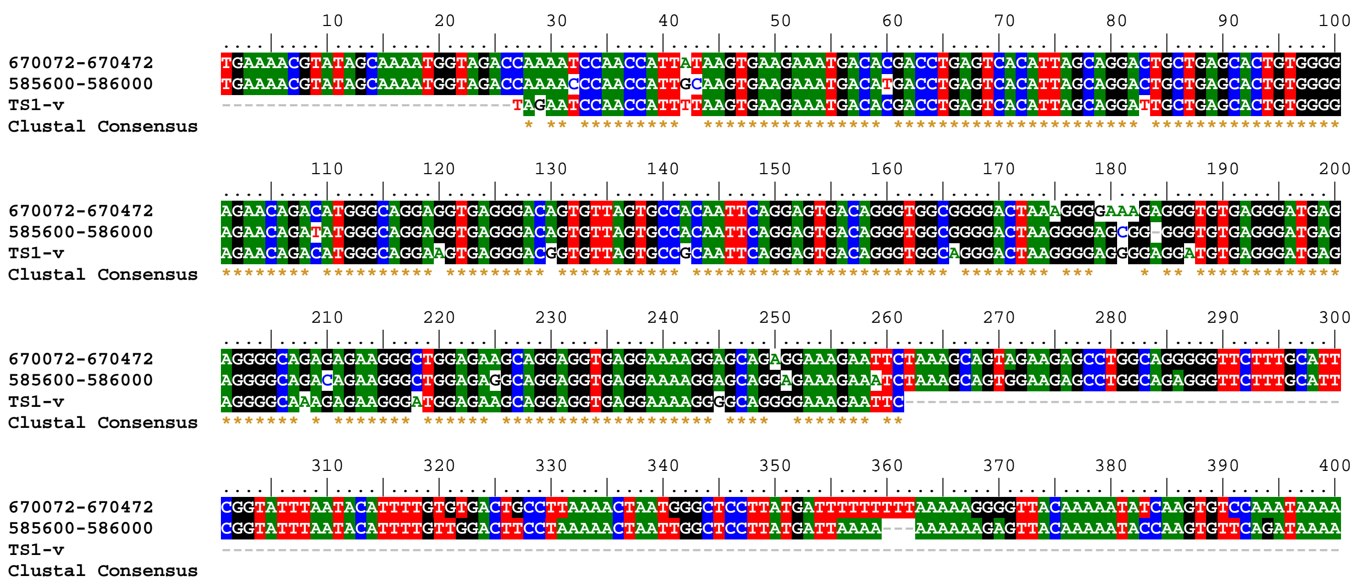


Figure 5.6.4.1: Alignment of TS1-v r-c and HCGIV-02 with the homologous Hs DNA strands of duplicons B (585 600 - 585 900) and C (670 072 - 670 432) (Hs: ID BA000025).

TS1-v r-c has less single differences with its homologue in duplicon C rather than B DNA. Alignment has been performed by Clustal X.

6 Discussion

The aim of this study was to describe the gene map of *Callithrix jacchus'* major histocompatibility complex (MHC) class I region and to elucidate its evolutionary relationship with other species. In order to establish the localization of the class I regions into the gene map BAC-Bank "CHORI-259" class I specific gene and framework probes were used to identify the clones.

The MHC class I gene complex is regarded as evolutionarily conserved because it genes have shown homology between different species: human, rat, mouse, rhesus monkey, chimpanzee and lemur (Amadou *et al.*, 1999; The MHC sequencing consortium, 1999; Daza-Vamenta *et al.*, 2004; Hurt *et al.*, 2004; Anzai *et al.*, 2003; Neff, 2005), however, *Callitrhix jacchus* is not known,. An improved description and understanding of the knowledge dynamics and mapping of this gene system is important for the comprehension the multiple diseases that are associated with MHC class I (Luthra-Guptasarma and Singh, 2004; Nardi *et al.*, 2003; Feder *et al.*, 1999; Weitkamp *et al.*, 1994; Stein, 2001; Eslamboli, 2005)

The aim of this work was to describe a gene map of the major histocompatibility complex (MHC) class I region on *Callithrix jacchus* and the evolutionary relationship with other species.

In order to establish the localization of the class I regions into the gene map the BAC-Bank "CHORI-259" was used with specific probes of class I gene and frameworks genes to identify the clones.

To elucidate the genomics structure of MHC of *Callithrix jacchus* is very important for the use of this animal like a model for diseases associated with the MHC (Wong *et al.*, 2005; Pedra *et al.*, 2005; Bach, 2005; Schulz, 2005; Lie, 2005; Reveille, 2005; Azizah *et al.*, 2004; Shinomiya *et al.*, 2004).

The analysis of the MHC class I in *Callitrhix jacchus* was constructed with three contigs represented in the BAC-clones were overlapping. The construction of the map needed the screenings of the BAC-Bank of Resources Center. The BAC-Bank contains the complete genome of *Callitrhix jacchus.* The genome was digested with restriction enzyme *EcoRI* and was cloned in the vector pTARBAC2.1. In the screening were found 256 clones class I positive of which 45 clones were selected. In them were characterized the MHC class I region and the framework genes that they are flanking this region, framework hypothesis (Amadou, 1999).

153 BAC clones were found by hybridization to contain MHC class I gene(s) and/or MIC gene(s).

To determinate a genomic map for the MHC class I region of *Callithrix jacchus,* 42 clones were allocated to the three class I contigs.

The MHC class I gene clones were identified with a specific probes of class I and probes of framework genes. The class I probes were Caja-G04 of *Callitrhix jacchus* and Mamu-B of *Macaca mulatta.* The framework probes used came from rat and human.

The organization between class I region and framework gene showed a great similarity as has been reported in human (The MHC sequencing consortium, 1999), rat (Lambract-Washington *et al*., 2000; Günther and Walter, 2001), mouse (Hanson and Trowsdale, 1991), rhesus monkey (Daza-Vementa *et al*., 2004), chimpanzee (Kulski *et al*., 2005) and lemur (Neff, 2005).

The contig 1 was represented by region between framework genes BAT1 and TCF19 included class I genes. The contig 2 was not found, because there are not CAT56 positive in the hybridization analysis; however, these results are not concluding that this contig does not exist. The contig 3 was represented by region between framework genes TRIM39 and TRIM26 included class I genes. The contig 4 was represented by region between framework genes PPP1R11 and MOG.

The screening included a search for MIC genes. The major histocompatibility complex class I chain related gene A (MICA) is located near HLA-B on chromosome 6 (Bahram *et al*., 1994) and is expressed by keratinocytes and epithelial cells and interacts with gamma-delta T cells (Christmas *et al*., 1993) and is implicated in the induction of stress (Groh *et al*., 1996). MICA encodes molecules similar to MHC class I antigens and may share the capacity to bind peptides or other short ligands (Koldovsky *et al*., 1995; Bahram *et al*., 1994). The MICA is in the lignand of NKG2D receptor of the NK cells (Bauer *et al*., 1999). The MIC genes have been found in rhesus monkey (Seo *et al*., 1994 and Seo *et al*., 2001) and chimpanzee (Kulski *et al*., 2005).

The probes Caja-G and HLA-B have proven, quite similarly, to be very useful for a screening of DNA containing class I genes and their pseudogenes: in *Callithrix jacchus*, *Microcebus murinus* and other mammals. For this very reason, however, these probes are not very suited for mapping an order of several or even many of these similar sequences of class I genes. The major shortage is a too small number

of discernible fragments; probably, more clones should also have been used for more steps in their lengths. So, we needed ancillary evidence for the identification of HLA-positive fragments. We took fragments from other primates, Hs, where gene contents of the fragments may be concluded from knowledge of the order and bp defined locations of genes along the DNA. All fragments found could be assigned in the map, some unique lengths interpreted as ends of the inserts cleaved.

These results are in agreement with the expectation of a similar organization of the MHC class I region in *Callithrix*, rat and human (Hurt *et al*., 2004 and Shiina *et al*., 2003), including the rule of clusters of class I genes being placed between non-class I framework genes.

These framework genes had not been mapped before in the Common marmoset: ATP6V1g2, BAT1, TCF19, POU5F1, TRIM39, TRIM26, TRIM10, TRIM15, TCTEX5, TCTEX4 and MOG. Four contigs or intervals, respectively, showed clearly three segments containing class I genes: BAT1 - TCF19, CAT56 - TRIM39 and TCTEX4 - MOG.

6.1 Physical map of the intervals of MHC class-I

6.1.1 Physical map of the interval BAT1 – TCF19 (contig 1)

Between the framework genes BAT1 and POU5F1-TCF19 has been localized contig 1 as part of the MHC class I region, in human with a 418 kb (The MHC consortium 1999), H2-complexes in mouse with 300 kb, as RT1-complex in rat with 450 kb (Hurt, *et al*., 2004) and in Patr chimpanzee (Fukami-Kobayashi, *et al*., 2005). These framework genes were found to be orthologous. In human, this interval contains the MICA and MICB genes (Bahram *et al*., 1994), HLA-B and HLA-C. The MHC class I genes expressed in this New World primate are not orthologous to any of the classical MHC class I loci of the Catarrhini (*A*, *B*, or *C* loci), instead, they are most similar to the human non-classical *HLA-G* (Watkins, *et al*., 1990). This MHC class I region between BAT1 and POU5F1 has a length of approximately 350 kb.

The contig 1 was established between BAT1 and POU5F1 by *EcoRI* fragments with specific hybridization of probes Caja-G, MICA, Mamu-B and respective framework genes (including ATP6Vg1 and TCF19).

6.1.2 Physical map of the intervals CAT56 - TRIM39 and TRIM39 - TRIM26 (contig 2 and contig 3 respectively)

The intervals CAT56 - TRIM39 and TRIM39 - TRIM26 of the MHC class I region have polymorphic margins in human. They both belong to contig 3 in the human, where contig 2 is the longest of the four contigs, however, almost devoid of class I genes. The small numbers of clones investigated in this area did not really allow constructing any one of the contigs 2 - 4, i.e. a share of stepwise overlapping clone inserts with sufficiently many, short and well discernible fragments covering the segment between its framework borders. The intervalls expected by comparison with Hs and other mammals seem to be existing, however, as evidenced by positive screening of clones with the probes Caja-G, CAT56, TRIM39 and TRIM26. In this screening of the intervall CAT56 - TRIM39, 20 positive clones were found. For intervall TRIM39 - TRIM26, 21 positive clones were detected.

The Interval Cat56 - TRIM39 contains class I gene(s) between the conserved framework genes CAT56 and TRIM39, with approximately of 500 kb in case of the RT1-complex of rat (Hurt *et al*., 2004).

The second intervall contains class I gene(s) between the conserved framework genes TRIM39 and TRIM26 with approximately of 200 kb in the RT1-complex. Near TRIM26, towards contig 4 are TRIM15 and TRIM10 (Hurt, *et al*., 2004). For practical reasons, the framework gene TCTEX5 located near TCTEX4 within less than 50 kb, was included in contig 4 presumed by comparison. TCTEX5 considered as end of a contig 3 detected clones with the framework genes TRIM 26 and TRIM39. So, there is probably no gap in *Callithrix* between its presumable contigs 3 and 4 (Hurt, *et al*., 2004; Jones, *et al*., 1999; Zhang, *et al*., 1998; Amadou, *et al*., 1999).

6.1.3 Physical map of interval TCTEX4 and MOG (contig 4)

Contig 4 contains many MHC class I genes and pseudogenes intercalated with many other genes between the conserved frameworks genes TCTEX4 and MOG. The interval has remarkably different lengths in the mammals investigated, approximately 50 kb in the rat RT1-complex (Hurt *et al*., 2004) and 400 kb in contig 4 of Hs. This interval is conformed by 2 clones, however, the sequence analysis support that this contig exist in *Callithrix jacchus.*

6.2 Sequence analysis of BAC clones

Sequence analysis and hybridization DNA of BAC-clones of *Callithrix jacchus* have demonstrated that the organization of MHC class I gene and region are similar at other mammals for instance Hs, Mamu, Rano, Patr, etc. However, we have found some particular difference, like ferritine sequence, inverted sequence and pseudogene sequence.

6.2.1 BAT1 gene in *Callithrix jacchus*.

K19161C12-TS2 has shown alignment with BAT1 gene from Hs, this gene is coding to an essential splicing factor (Fleckner *et al*., 1997). BAT1 gene contains 10 exons (aprox 10 kb) (Peelman *et al*., 1995), Caja sequence has shown from intron 4 until intron 6, the sequences differences between Hs and Caja include nucleotide substitutions and insertions/deletions; 7 insertions and 14 deletions (Figure 5.4.2). The localization of BAT1 gene is conserved in Hs and Susc (pig) (Spies, *et al*., 1989), K19161C12-TS2 sequence suggests is also conserved in *Callithrix jacchus*. Phylogenetic analysis performed between several species showed that this gene is conserved, however, the most homology was between Hs and Patr; Caja was localized nearly to Mamu and Susc (Figure 5.5.2.1).

6.2.2 TCF19 gene in *Callithrix jacchus*

The hybridization analysis was negative to the sonde of TCF19, however, the K28-347D1-TCF19 sequence showed homology with human TCF19 gene into the nucleotide and amino acid sequence (Figure 5.4.4.1 and 5.4.4.2). The gene TCF19 encoded a 359 amino acid polypeptide; into K28-347D1-TCF19 amino acid sequence 72 amino acid in position 157 until 239 have shown a high homology. The transcript of this gene has seen in Mumu, Rano, Susc, Bota and Patr (Ku, *et al*., 1991). Phylogenetic analysis has shown that this gene is conserved between these species, however, the most homology was between Hs and Patr; Caja was localized nearly to Mamu and Susc (Figure 5.5.2.2).

6.2.3 MOG gene in *Callithrix jacchus*

The human myelin oligodendrocyte has been found in several species. The complete nucleotide sequence and structural characterization is located in chromosome 6 p22 - p21.3, with eight exons and approximately 100 kb, telomeric to HLA-F (Amado,u *et*

al., 1995; Pham-Dinh *et al*., 1993; Roth *et al*., 1995). MOG gene organization is very similar in both Hs and Mumu (Pham-Dinh *et al*., 1995). The MOG gene sequence of *Callithrix jacchus* reported here suggests a possible similarity with other species and not many evolutionary changes seem to have taken place.

6.3 On alignment of sequences in general and particular those of *Callithrix jacchus*

Most of the 55 sequences seem to be localized as listed in table 5.4.1.1. The rest although more difficult, should be solved, too. One of apparently few differences between Hs and Caja in class I region was the MHC ferritin pseudogene. The question of Caja with duplicon B and C is answered by two linear of evidences: Caja - HLA-B related sequences were all found more similar with duplicon B sequences in Hs; TS1 sequences, however, are generally more similar to C sequences. Therefore, Caja should have both, a –B and a –C duplicon.

6.4. On duplicons in the MHC class I region

Duplicon in the MHC class I region have been studied recently by Kulsi, *et al*., 2004, concentrating on the α-block (HLA section in contig 4). Although the youngest class I duplicon B and C in many Mamu instead 14 B duplicons are known (Fukami-Kobayashi *et al*., 2005). The duplicon B of Hs contains HCGIV-01 pseudogene and the duplicon C contains HCGIV-02 pseoudogene (Shiina *et al*., 2001). TS1 was aligned into MHC class I duplicon B and C with a high homology. We are assigned ps1 and ps2 like pseudogenes, because they fulfill some of the criteria proposed by Zang *et al*., 2004. However, it is difficult to decide whether the ps1 and ps2 specifically belong at type 1, 2 or 3, but they have the characteristics of class I pseudogenes, which are: complete genes are prevented by frameshifts, premature stop codons, fragments lacking or other defects (Hughes, *et al*,. 1995).

So in this area, the very young of class I duplicons, Caja contains the pseudogenes known from TS1 (ps1 and ps2) which should be useful indicators. More careful investigation must decide, if these two ancient genes might not have been originated by a duplication.

6.5 Transposons, particularly the pseudo ferritin found in MHC Caja transposon and its role in MHC evolution

TS1-ps fth1 should have been transposed by an mRNA from the expressed ferritin gene from Caja, may its reincorporation into the genome reversion of the transcription process. (Perhaps its reincorporation into the genome was some sort of a reversion of a transcription process).

TS1-ps fth1 was aligned with all mRNA (gene and pseudogenes) from FTH1 of Hs. The TS1ps fth1 has the start codon and poly A end, but does not have introns, which suggests that is a processed pseudogene type 2 (Zhang *et al*,. 2004). Gatti *et al.*, (1987) found pseudogenes of FTH1 in chromosome 1 - 6, 8, 9, 11, 13, 14, 17 and X. Strangely, it does contain the 28 nucleotide signal of ferritin genes but in a fragmented fashion (Murray *et al.*, 1987). This observation requires further investigation, particularly in other locations expected to be homologous. On the other hand, there is a part of v r-c that does seem to be more frequently conserved, even within pseudogenes.

A simpler hypothesis that can be proposed is that the insert is placed at random. In this case, TS1-ps2 and v r-c would be part of the whole cluster of class I pseudogenes aligning with v r-c = HCGIV-02, suggesting the possibility that much of these is ancient 5'FR, as is actually suggested by an ATG general placed towards their 3' ends.

6.6 On the potential value of sequence TS1 for molecular immunogenetics

A couple of class I pseudogenes have being completely sequenced, both of which are representative examples of the many duplicons that exist in the B/C section in various primates (Kulski *et al.*, 2004). The earliest B and C duplicons contain highly identical DNA. Should the principles of duplication and mutation have been shaping the MHC from its very beginning, eventually, these probes would have to show a lot more duplicons. However, due to the degree of percentages with much less identity present in this study, this requires more sophisticated statistical evaluations beyond the scope of the present work. In any case, this underscores one of the potential uses of ps1 and ps2.

The analysis of the duplicon structure and organization suggests the presence of block duplicon model.

Comparative analysis of MHC class I processed pseudogene sequence from new world primates: *Saguinus oedipus* (Saoe-PS1: ID: U52113) and *Saguinus mystax* (Samy-PS2 ID: U52121) (Cadavid *et al*., 1997) did not show identity with *Callithrix jacchus* ps1 and ps2.

This study presented the pioneer work about the organization of the MHC class I complex in the New World monkey *Callithrix jacchus*. In addition, several sequences which have never been reported before for this species have been identified. Such is the case of a ferritin pseudogene, the ps1 and ps2 pseudogenes and an inversion sequence.

Throughout the development of the present investigation, it has become apparent that *Callithrix jacchus* should be an excellent candidate animal model for studies of MHC complex related diseases.

7 Summary

Analysis of MHC class I region in common marmoset (*Callithrix jacchus*) was made with 42 BAC clones. These clones containing class I region genes were identified for the purpose of both restriction mapping and PCR sequencing. Contigs 1, 3 and 4 of BAC clones were obtained for genomic analysis by comparison of *EcoRI* fragments characterized by Southern blot fingerprint hybridization with framework and class I gene probes. A gene map for the interval BAT1 - PUO5F1 was constructed by comparison of the Caja *EcoRI* fragments so characterized with the computer-split *EcoRI* fragments of the Hs class I region having known gene assignments. Ancillary evidence needed for identification of a fragment obtained from six BAC clones was obtained by analysis of sections of the Hs DNA sequence near POU5F1 for class I pseudogenes using a collection of all Hs class I pseudogene sequences for comparison of their alignment by the neighbour-joining program. Seven new pseudogenes were thereby detected within only 50 kb from POU5F1, each one classified by its greatest similarity with a known pseudogene. Adequately refined, this method might be useful for finding and classifying all of the many class I pseudogenes.

A genetic typification of 42 BAC clone inserts by PCR with various primers was performed. This knowledge was used for pilot sequencing in the class I region of Caja. The resulting 55 sequences were localized and potentially gene-assigned by comparative alignment, generally with DNA of Hs, in many cases also of other primates and mammals. These sequence localizations and assignments were defined by their first and last bp numbers on the scales of Shiina and also of the Venter genome. For survey, they were both tabulated and entered into a to-bp-scale map drawn for the Hs class I region.

TS1, one of the two Caja sequences established by Prof. Takashi Shiina using a BAC clone selected by the author, was shown to contain within its 2994 bp four different pseudogenes of peculiar interest: on the right hand strand, two class I pseudogenes (ps1 and ps2); on the left hand strand, reading anti-parallel, a rather young ferritine heavy chain pseudogene (psfth1) apparently unique to Caja; and next to it exist a “Vorspann“ v, potentially an ancient 5’FR (flanking region) which in its reverse-complementary, the right hand strand turns out to be the known pseudogene HCGIV-02, as a very highly identical, almost completely coincident sequence. The two new pseudogenes, ps1 and ps2, were recognized as being useful indicators for

the duplicated DNA as found in the two duplicons B and C of Hs and Patr that contain the ps1, ps2 and v homologues near the classical HLA genes B and C, respectively. Alignment of ps1, ps2 and v could be observed as well in all of the 14 Mamu B duplicons of the rhesus monkey *Macacca mulatta*. High identity alignment of ps1, ps2 and v is being searched for in the vicinities of other classical HLA genes being conditions for duplicons.

Physical mapping and sequence analysis of the MHC class I region of *Callithrix jacchus* is to supply new genomic basic data, which further analyses of the MHC region for evolution of MHC, disease susceptibility, immune reactivity and transplantation repulsion of the MHC to make possible. Additionally the map offers the basis for following Sequencing of the BAC clones of the individual contigs.

8 References

Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P. and Inoko, H. 2002. Evidence of *en bloc* duplication in vertebrate genomes. Nat. Genetic. 31, 100-105.

Adams, E.J. and Parham, P. 2001. Species-specific evolution of MHC class I genes in the higher primates. Immunol Rev. 183, 41-64.

Allcock, R.J., Price, P., Gaudieri, S., Leelayuwat, C., Witt, C.S. and Dawkins, R.L. 1999. Characterization of the human central MHC gene, BAT1: genomic structure and expression. Exp. Clin. Immunogenet. 16, 98-106.

Allcock, R.J., Williams, J.H. and Price, P. 2001. The central MHC gene, BAT1, may encode a protein that down-regulates cytokine production. Genes Cells 6, 487-494.

Alper, C.A., Awdeh, Z. and Yunis, E.J. 1992. Conserved extended MHC haplotypes. Experimental & Clinical Immunogenetics 9, 58-71.

Alvarez, M., Martinez-Laso, J., Garcia-Torre, C., Varela, P., Perez-Blas, M., Moreno, M.A., Gomez-Casado, E., Arnaiz-Villena, A. 1998. Description of a new kind of MHC DNA sequence in Saguinus oedipus (cotton-top tamarin). Eur. J. Immunogenet. 25, 287-292.

Alvarez, M., Martinez-Laso, J., Varela, P., Diaz-Campos, N., Gomez-Casado, E., Vargas-Alarcon, G., Garcia-Torre, C. and Arnaiz-Villena, A. 1997. High polymorphism of Mhc-E locus in non-human primates: alleles with identical exon 2 and 3 are found in two different species. Tissue Antigens 49, 160.

Alvarez-Tejado, M., Martinez-Laso, J., Garcia-de-la-Torre, C., Varela, P., Recio, M.J., Allen, L., Gomez-Casado and E., Arnaiz-Villena, A. 1998. Description of two Mhc-C-related sequences in the New World monkey Saguinus oedipus. Eur. J. Immunogenet. 25, 409-417.

Amadou, C. 1999. Evolution of the Mhc class I region: the framework hypothesis. Immunogenetics; 49 4:362-367.

Amadou, C., Ribouchon, M.T., Mattei, M.G., Jenkins, N.A., Gilbert, D.J., Copeland, N.G., Avoustin, P. and Pontarotti, P. 1995. Localization of new genes and markers to the distal part of the human major histocompatibility complex (MHC) region and comparison with the mouse: new insights into the evolution of mammalian genomes. Genomics 26, 9-20.

Ando, A., Shigenari, A., Kulski, J.K., Renard, C., Chardon, P., Shiina, T. and Inoko, H. 2005. Genomic sequence analysis of the 238-kb swine segment with a cluster of TRIM and olfactory receptor genes located, but with no class I genes, at the distal end of the SLA class I region. Immunogenetics 57, 864-873.

Anholts, J.D.H. 2002. Four new HLA alleles found, and conformation of one Class I and two Class II alleles. Thesis Department of Immunohematology and Bloodtransfusion, University Hospital of Leiden, Leiden, Netherlands.

Antunes, S.G., de Groot, N.G., Brok, H., Doxiadis, G., Menezes, A.A., Otting, N. and Bontrop, R.E. 1998. The common marmoset: a new world primate species with limited Mhc class II variability. Proc. Natl. Acad. Sci. USA 95, 11745-11750.

Anzai, T. 2003. Comparative sequencing of human and chimpanzee MCH class I regions unveils insetions /deletions as the major path to genomic divergence. Proc. Natl. Acad. Sci USA 100, 7708-7713.

Anzai, T., Shiina, T. and Inoko, H. 2004. Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. Mol. Biol. Evol. 21, 2079-2091.

Anzai, T., Shiina, T., Kimura, N., Yanagiya, K., Kohara, S., Shigenari, A., Yamagata, T., Kulski, J.K., Naruse, T.K., Fujimori, Y., Fukuzumi, Y., Yamazaki, M., Tashiro, H., Iwamoto, C., Umehara, Y., Imanishi, T., Meyer, A., Ikeo, K., Gojobori, T., Bahram, S., Inoko, H. 2003. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. Proc. Natl. Acad. Sci. USA 100, 7708-7713.

Arnaiz-Villena, A., Martinez-Laso, J., Castro, M.J., Morales, P., Moscoso, J., Varela, P., Gomez-Casado, E. and Allende, L.M. 2001 The evolution of the MHC-G gene does not support a functional role for the complete protein. Immunol Rev. 183, 65-75.

Arnaiz-Villena, A., Morales, P., Gomez-Casado, E., Castro, M.J., Varela, P., Rojo-Amigo, R. and Martinez-Laso, J. 1999. Evolution of MHC-G in primates: a different kind of molecule for each group of species. J. Reprod. Immunol. 43,111-125.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman J.G., Smith J.A. and Struhl, K. 1994-2005. Current protocols in molecular biology. 1 - 5 vols. Different Chapter.

Azizah, M.R., Kuak, S.H., Ainol, S.S., Rahim, M.N., Normaznah, Y. and Norella, K. 2004. Association of the tumor necrosis factor alpha gene polymorphism with susceptibility and clinical-immunological findings of systemic lupus erythematosus. Asian Pac. J. Allergy Immunol. 22, 159-163.

Bach, J.F. 2005. [Infections and autoimmunity]. Rev. Med. Interne. 1, 32-34.

Beck, S. and Trowsdale, J. 2000. The human major histocompatability complex: lessons from the DNA sequence. Annu. Rev. Genomics Hum. Genet. 1, 117-137.

Beck, T. 2005. The feline major histocompatibility complex is rearrenged by an inversion with a breakpoint in the distal class I region. Immunogenetics 56, 702-709.

Begovich, A.B., McClure, G.R. and Suraj, V.C. 1992. Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. J. Immunology 148, 249-258.

Bergstrom, T.F., Erlandsson, R., Engkvist, H., Josefsson, A., Erlich, H.A. and Gyllensten, U. 1999. Phylogenetic history of hominoid DRB loci and alleles inferred from intron sequences. Immunol. Rev. 167, 351–365.

Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W.S., Johnson, S.L., Gates, M., Postlethwait, J.H. and Klein, J. 1997. Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, Danio rerio. Immunogenetics 46, 129-134.

Bowen, M.A., Bajorath, J., D'Egidio, M., Whitney, G.S., Palmer, D., Kobarg,J., Starling,G.C., Siadak, A.W. and Aruffo, A. 1997. Characterization of mouse ALCAM (CD166): the CD6-binding domain is conserved in different homologs and mediates cross-species binding. Eur. J. Immunol. 27, 1469-1478.

Boyson, J.E., Iwanaga, K.K., Urvater, J.A., Hughes, A.L., Golos, T.G. and Watkins, D.I. 1999. Evolution of a nonclassical MHC class I locus in two Old World primate species. Immunogenetics 49, 86-98.

Boyson, J.E., Shufflebotham, C., Cadavid, L.F., Urvater, J.A., Knapp, L.A., Hughes, A.L. and Watkins, D.I. 1996. The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. J. Immunol. 156, 4656-4665.

Brok, H.P., Uccelli, A., Kerlero De Rosbo, N., Bontrop, R.E., Roccatagliata, L., de Groot, N.G., Capello, E., Laman, J.D., Nicolay, K., Mancardi, G.L., Ben-Nun, A., Hart, B.A. 2000. Myelin/oligodendrocyte glycoprotein-induced autoimmune encephalomyelitis in common marmosets: the encephalitogenic T cell epitope pMOG24-36 is presented by a monomorphic MHC class II molecule. J. Immunol. 165, 1093-1101.

Bronson, S.K., Pei, J., Taillon-Miller, P., Chorney, M.J., Geraghty, D.E. and Chaplin, D.D. 1991. Isolation and characterization of yeast artificial chromosome clones linking the HLA-B and HLA-C loci. Proc. Natl. Acad. Sci. USA 88, 1676-1680.

Cadavid, L.F., Hughes, A.L., Watkins, D.I. 1996. MHC class I-processed pseudogenes in New World primates provide evidence for rapid turnover of MHC class I genes. J. Immunol. 157, 2403-2309.

Cadavid, L.F., Mejia, BE, Watkins DI. 1999. MHC class I genes in a New World primate, the cotton-top tamarin (Saguinus oedipus), have evolved by an active process of loci turnover. Immunogenetics 49, 196-205.

Cadavid, L.F., Shufflebotham, C., Ruiz, F.J., Yeager, M., Hughes, A.L. and Watkins, D.I. 1997. Evolutionary instability of the major histocompatibility complex class I loci in New World primates. Proc. Natl. Acad. Sci. USA 94, 14536-14541.

Caggiari, L., De Re, V. and Rehermann, B. 2006. Identification of novel chimpanzee MHC I and II alleles using an improved sequence-based typing strategy. Hum. Immunol. 67, 63-72.

Canavez, F.C., Moreira, M.A., Ladasky, J.J., Pissinatti, A., Parham, P. and Seuanez H.N. 1999. Molecular phylogeny of new world primates (Platyrrhini) based on beta2-microglobulin DNA sequences. Mol. Phylogenet. Evol. 12, 74-82.

Cardenas, P.P., Suarez, C.F., Martinez, P., Patarroyo, M.E. and Patarroyo, M.A. 2005. MHC class I genes in the owl monkey: mosaic organization, convergence and loci diversity. Immunogenetics 56, 818-832.

Castellino, F. 1997. Antigen presentation by MHC class II molecules: invariant chain function, protein trafficking, and the molecular basis of diverse determinant capture. Hum. Immunol. 54, 159-169.

Chardon, P., Renard C., Gaillard, C.R. and Vaiman, M. 2000. The porcine major histocompatibility complex and related paralogous region. Genet. Sol. Evol. 32, 109-128.

Christmas, S.E., Brew, R., Deniz, J. and Taylor, J.J. 1993. T cell receptor heterogeneity of gamma delta T cell clones from human female reproductive tissues. Immunology 78, 436-442.

Chu, T.W., Capossela, A., Coleman, R., Goei, V.L., Nallur, G. and Gruen, J.R. 1995. Cloning of a new 'finger' protein gene (ZNF173) within the class I region of the human MHC. Genomics 29, 229-239.

Coimbra-Filho, A.F. 1984. Situacao Atul Dos Calitriquideos que Ocorrem no Brasil (Callitrichidae-Primates). in *A Primatologia no Brasil*. ed. M.T. de Mello. Sociedade Brasileira de Primatologia: Brasilia.

Costanzo, F., Colombo, M., Staempfli, S., Santoro, C., Marone, M., Frank, R., Delius, H. and Cortese, R. 1986 Structure of gene and pseudogenes of human apoferritin H. Nucleic Acids Res. 14, 721-736.

Dagenais, S.L. and Nakamura, I. 1997. A physical map of the mouse H2 Bat5/Db interval. Mamm. Genome 8, 39-41.

Daza-Vamenta, R., Glusman, G., Rowen, L., Guthrie, B. and Geraghty, D.E. 2004. Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Res. 14, 1501-1515.

Degli-Esposti, M.A., Leaver, A.L., Christiansen, F.T., Witt, C.S., Abraham, L.J. and Dawkins, R.L. 1992. Ancestral haplotypes: conserved population MHC haplotypes. Human Immunology 34, 242-252.

Diaz, D., Naegeli, M., Rodriguez, R., Nino-Vasquez, J.J., Moreno, A., Patarroyo, M.E., Pluschke, G. and Daubenberger, C.A. 2000. Sequence and diversity of

MHC DQA and DQB genes of the owl monkey Aotus nancymaae. Immunogenetics 51, 528-537.

Digby, L. 1995. Infant Care, Infanticide, and Female Reproductive Strategies in Polygynous Groups of Common Marmosets (*Callithrix jacchus*). Behav. Ecol. Sociobol. 37, 51-61.

Doxiadis, G.G., van der Wiel, M.K., Brok, H.P., de Groot, N.G., Otting, N., 't Hart, B.A., van Rood, J.J. and Bontrop, R.E.. 2006. Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species. Proc. Natl. Acad. Sci. USA 103, 5864-5868.

Dunham, I., Sargent, C.A., Dawkins, R.L., Campbell, R.D. 1989. An analysis of variation in the long-range genomic organization of the human major histocompatibility complex class II region by pulsed-field gel electrophoresis. Genomics 5, 787-796.

Epple, G. 1970. Maintenance, breeding, and development of marmoset monkeys (callitrichidae) in captivity. Folia Primat. 12, 56-76.

Epple, G., Belcher, A.M., Kuderling, I., Zeller, U., Scolnick, L., Greenfield, K.L., Smith III, A.B. 1993. Making sense out of scents: species differences in scent glands, scent-marking behaviour, and scent-mark composition in the callitrichdae. In marmosets and tamarins: systematics, behaviour, and ecology. Ed. Anthony B. Rylands, Oxford University Press.

Epple, G.1968. Comparative Vocalization in Marmoset Monkeys. Folia Primat. 8 , 1-40.

Eslamboli A. 2005. Assessment of GDNF in primate models of Parkinson's disease: comparison with human studies. Rev Neurosci; 16, 303-310.

Evans, D.T., Knapp, L.A., Jing, P., Piekarczyk, M.S., Hinshaw, V.S. and Watkins, D.I. 1999. Three different MHC class I molecules bind the same CTL epitope of the influenza virus in a primate species with limited MHC class I diversity. J. Immunol. 162, 3970-3977.

Evans, D.T., Piekarczyk, M.S., Allen, T.M., Boyson, J.E., Yeager, M., Hughes, A.L., Gotch, F.M., Hinshaw, V.S. and Watkins, D.I. 1997. Immunodominance of a single CTL epitope in a primate species with limited MHC class I polymorphism. J. Immunol. 159, 1374-1382.

Evans, D.T., Piekarczyk, M.S., Cadavid, L., Hinshaw, V.S. and Watkins, D.I. 1998. Two different primate species express an identical functional MHC class I allele. Immunogenetics 47, 206-211.

Evans, S. 1983. The Pair-bond of the Common Marmoset, *Callithrix jacchus jacchus*: An Experimental Investigation. Animal Behavior 31, 651-658.

Evans, S. and Poole, T.B. 1984. Long-term Changes and Maintenance of the Pair-bond in Common Marmosets *Callithrix jacchus jacchus*. Folia Primat. 42, 33-41.

Feder, J.N. 1999. The hereditary hemochromatosis gene (HFE): a MHC class I-like gene that functions in the regulation of iron homeostasis. Immunol. Res. 20, 175-185.

Flajnik, M.F., Canel, C., Kramer, J. and Kasahara, M. 1991. Which came first, MHC class I or class II? Immunogenetics 33, 295.

Fleckner, J.; Zhang, M.; Valcarcel, J.; Green, M. R. 1997. U2AF(65) recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. Genes Dev. 11, 1864-1872.

Flügge, P., Zimmermann, E., Hughes, A.L., Günther, E. and Walter, L. 2002. Characterization and phylogenetic relationship of prosimian MHC class I genes. J. Mol. Evol. 55, 768-775.

Fukami-Kobayashi, K., Shiina, T., Anzai, T., Sano, K., Yamazaki, M., Inoko, H. and Tateno, Y. 2005. Genomic evolution of MHC class I region in primates. Proc. Natl. Acad. Sci. USA 102, 9230-9234.

Gaudieri. S., Leelayuwat, C., Tay GK, Townend DC, Dawkins RL. 1997. The major histocompatability complex (MHC) contains conserved polymorphic genomic sequences that are shuffled by recombination to form ethnic-specific haplotypes. J. Mol. Evolution 45, 17-23.

Germain RN. 1995. The biochemistry and cell biology of antigen presentation by MHC class I and class II molecules. Implications for development of combination vaccines. Ann. NY Acad. Sci. 754, 114-125.

Giffon, T., Lepourcelet, M., Pichon, L., Jezequel, P., Bouric, P., Carn, G., Pontarotti, P., Gall, J.Y. and David, V. 1996. Cloning of a human homologue

of the mouse Tctex-5 gene within the MHC class I region. Immunogenetics 44, 331-339.

Go, Y., Satta, Y., Kawamoto, Y., Rakotoarisoa, G., Randrianjafy, A., Koyama, N. and Hirai, H. 2003. Frequent segmental sequence exchanges and rapid gene duplication characterize the MHC class I genes in lemurs. Immunogenetics 55, 450-461.

Goei, V.L., Parimoo, S., Capossela, A., Chu, T.W. and Gruen, J.R. 1994. Isolation of novel non-HLA gene fragments from the hemochromatosis region (6p21.3) by cDNA hybridization selection. Am. J. Hum. Genet. 54, 244-251.

Gonzalez, P., Pinto, F., Nogales, M., Jimenez-Asensio, J., Hernandez, M. and Cabrera, V.M. 1996. Phylogenetic relationships of the Canary Islands endemic lizard genus Gallotia (Sauria: Lacertidae), inferred from mitochondrial DNA sequences. Mol. Phylogenet. Evol. 6, 63-71.

Goujet-Zalc, C., Faille, A., Teillaud, J.L., Turmel, P., Mahouy, G., Charron, D.J. 1984. Molecular characterization of two Ia-like antigens in marmoset. Immunogenetics 19, 155-161.

Graser, R., Vincek, V., Takami, K. and Klein, J. 1998. Analysis of zebrafish Mhc using BAC clones. Immunogenetics 47, 318-325.

Groh, V., Bahram, S., Bauer, S., Herman, A., Beauchamp, M. and Spies, T. 1996. Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. Proc. Natl. Acad. Sci. USA 93, 12445-12450.

Groves C. 2001. Primate taxonomy. Washington DC: Smithsonian Inst Pr. 350 p

Groves, C.P. 2001. Primate Taxonomy. Smithsonian Institution Press. Washington y Londres. 350 pages.

Gruen, J.R., Nalabolu, S.R., Chu, T.W., Bowlus, C., Fan, W.F., Goei, V.L., Wei, H., Sivakamasundari, R., Liu, Y., Xu, H.X., Parimoo, S., Nallur, G., Ajioka, R., Shukla, H., Bray-Ward, P., Pan, J. and Weissman, S.M. 1996. A transcription map of the major histocompatibility complex (MHC) class I region. Genomics 36, 70-85.

Grundschober, C., Labonne, M.P., Javaux, F., Steiner, Q.G., Gebuhrer, L.and Tiercy, J.M. 1998. Sequence of four new HLA-Cw alleles: a possible role of interallelic recombination. Tissue Antigens. 51, 72-79.

Ha, H., Howard, C.A., Yeom, Y.I., Abe, K., Uehara, H., Artzt, K. and Bennett, D. 1991. Several testis-expressed genes in the mouse t-complex have expression differences between wild-type and t-mutant mice. Dev. Genet. 12, 318-332.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 41, 95-98.

Handel-Fernandez, M.E., Kurimoto, I., Streilein, J.W. and Vincek, V. 1999. Genetic mapping and physical cloning of UVB susceptibility region in mice. J. Invest. Dermatol. 113, 224-229.

Hansen, A. 2004. Bioinformatik ein Leitfaden für Naturwissenschaftler 2. überarbeitete und erweiterte Auflage. Birkhäuser. Germany.

Harada, H., Harada, Y., O'Brien, D.P., Rice, D.S., Naeve, C.W. and Downing, J.R. 1999. HERF1, a novel hematopoiesis-specific RING finger protein, is required for terminal differentiation of erythroid cells. Mol. Cell. Biol. 19, 3808-3815.

Hartl, D.L. 2000. A primer of population genetics. 3rd ed. Sinauer Associates, USA.

Hauptmann, G. and Bahram, S. 2004. Genetics of the central MHC. Curr Opin Immunol. 16, 668-672.

Heeney, J.L., van Els, C., de Vries, P., ten Haaft, P., Otting, N., Koornstra, W., Boes, J., Dubbes, R., Niphuis, H., Dings, M., Cranage, M., Norley, S., Jonker, M., Bontrop, R.E. and Osterhaus, A. 1994. Major histocompatibility complex class I-associated vaccine protection from simian immunodeficiency virus-infected peripheral blood cells. J. Exp. Med. 180, 769-774.

Helou, K., Yan, Q., Yuan, X.J., Kunz, H.W., Levan, G. and Gill, T.J. III. 1999. Cytogenetic localization of the growth and reproduction complex (Grc) in the rat and in the mouse and its position in relation to RT1.EC and other loci in the rat MHC. Hereditas 130, 105-109.

Henry, J., Ribouchon, M., Depetris, D., Mattei, M., Offer, C., Tazi-Ahnini, R. and Pontarotti, P. 1997. Cloning, structural analysis, and mapping of the B30 and

B7 multigenic families to the major histocompatibility complex (MHC) and other chromosomal regions. Immunogenetics 46, 383-395.

Herberg JA, Sgouros J, Jones T, Copeman J, Humphray SJ, Sheer D, Cresswell P, Beck S and Trowsdale J. 1998. Genomic analysis of the Tapasin gene, located close to the TAP loci in the MHC. Eur. J. Immunol. 28, 459-467.

Hickson, R.E. and Cann, R.L. 1997. Mhc allelic diversity and modern human origins. J. Mol. Evol. 45, 589-598.

http://www.bacpac.chori.org

http://www.damisela.com/zoo/mam/primates/callitrichidae/jacchus/taxa.html

http://www.ebi.ac.uk/emboss/transeq/

http://www.ensembl.org/index.html

http://www.expasy.org/

http://www.genome.wi.mi.edu/cgi-bin/primer/primer3.cgi.

http://www.infobiogen.fr/doc/ClustalW/clustalw.html

http://www.itis.usda.gov/servlet/SingleRpt/SingleRpt?search_topic=TSN&search_value=572915

http://www.ncbi.nlm.nih.gov/BLAST/

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?CMD=search&DB=pubmed

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

http://pin.primate.wisc.edu/factsheets/entry/common_marmoset/taxon

http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker

http://www.saudeanimal.com.br/sagui.html

http://www.zoo.ru/Moscow/monk_e8.html

Hughes, A.L. and Nei, M. 1993. Evolutionary relationships of the classes of major histocompatibility complex genes. Immunogenetics 37, 337-346.

Hughes, A.L. and Nei, M. 1990. Evolutionary relationships of class II major-histocompatibility- complex genes in mammals. Molecular Biology & Evolution 7, 491-514.

Hughes, A.L. and Yeager, M. 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. 32, 415-435.

Hurt, P., Walter, L., Sudbrak, R., Klages, S., Mueller, I., Shiina, T., Inoko,H., Lehrach, H., Guenther, E., Reinhardt, R. and Himmelbauer, H. 2004. The genomic sequence and comparative analysis of the rat major histocompatibility complex. Genome Res. 14, 631-639.

Ioannidu, S., Walter, L., Dressel, R. and Gunther, E. 2001. Physical map and expression profile of genes of the telomeric class I gene region of the rat MHC. J. Immunol. 166, 3957-3965.

Iris, F., Bougueleret, L., Prieur, S., Caterina, D., Primas, G., Perrot, V., Jurka, J., Rodriguez-Tome, P., Claverie, J., Cohen, D. and Dausset, J. 1993. Dense Alu clustering and a potential new member of the NF kappa B family within a 90 kilobase HLA class III segment. Nat. Genet. 3, 137-145.

Janeway, C.A., Travers, P., Walport, M. and Capra, J.D. 1999. Immunobiology. 4th ed. Elservier Science London.

Jones, E.P., Kumanovics, A., Yoshino, M. and Fischer Lindahl, K. 1999. Mhc class I and non-class I gene organization in the proximal H2-M region of the mouse. Immunogenetics 49, 183-195.

Jones, E.Y., Fugger, L., Strominger, J.L., Siebold, C. 2006. MHC class II proteins and disease: a structural perspective. Nat. Rev. Immunol. 6, 271-282.

Jones, J.S. and Partridge, L. 1983. Tissue rejection: the price for sexual acceptance. Nature 304, 484-485.

Jones, R.A., Scott, C.S. and Child, J.A. 1988. Expression of MHC class I and class I-like gene products on the cell membrane of mature and immature T cells. Leuk. Res. 12, 799-804.

Kahloun, A., Chauvel, B., Mauvieux, V., Dorval, I., Jouanolle, A.M., Gicquel, I., Le Gall, J.Y. and David, V. 1993. Localization of seven new genes around the HLA-A locus. Hum. Mol. Genet. 2, 55-60.

Kanki, J.P., Chang, S. and Kuwada, J.Y. 1994. The molecular cloning and characterization of potential chick DM-GRASP homologs in zebrafish and mouse. J. Neurobiol. 25, 831-845.

Kaufman, J., Skjoedt, K. and Salomonsen, J. 1990. The MHC molecules of nonmammalian vertebrates. Immunological Reviews 113, 83-117.

Klein, J. and Sato, A. 1998. Birth of the major histocompatibility complex. Scan. J. Immunology 47, 199.

Klein, J. 2001. George Snell's first foray into the unexplored territory of the major histocompatibility complex. Genteric 159, 435-439.

Klein, J. 1976. Evolution and function of the major histocompatibility complex: facts and speculations. In: Gotze D, ed. The Major Histocompatibility System in Man and Animals, New York: Springer-Verlag, 339-378.

Knapp, L.A., Cadavid, L.F. and Watkins, D.I. 1998. The MHC-E locus is the most well conserved of all known primate class I histocompatibility genes. J. Immunol. 160, 189-196.

Koch, J. and Tampe, R. 2006. The macromolecular peptide-loading complex in MHC class I-dependent antigen presentation. Cell. Mol. Life Sci. 63, 653-662.

Koldovsky, P., Haas, I., Beir, H. and Ganzer, U. 1995. Immunohistochemica examination of 11 cell lines derivaded from human head and nech saqumous cell carcinomas, their recurrence or metastases. Europ. Arch. Oto-Rhino-Laryngo 252, 359-365.

Krebs, K.C., Jin, Z.Y., Rudersdorf, R., Hughes, A.L. and O'connor, D.H. 2005. Unusually high frequency MHC class I alleles in mauritian origin cynomolgus macaques. J. Immunol. 175, 5230-5239.

Kriener, K., O'hUigin, C, and Klein, J. 2001. Independent origin of functional MHC class II genes in humans and New World monkeys. Hum. Immunol. 62, 1–14.

Krishnan, B.R., Jamry, I. and Chaplin, D.D. 1995. Feature mapping of the HLA class I region: localization of the POU5F1 and TCF19 genes. Genomics 30, 53-58.

Ku, D.H., Chang, C.D., Koniecki, J., Cannizzaro, L.A., Boghosian-Sell, L., Alder, H. and Baserga, R. 1991. A new growth-regulated complementary DNA with the sequence of a putative trans-activating factor. Cell Growth Differ. 2, 179-186.

Kulski, J.K., Anzai, T. and Inoko, H. 2005. ERVK9, transposons and the evolution of MHC class I duplicons within the alpha-block of the human and chimpanzee. Cytogenet. Genome Res. 110, 181-192.

Kulski, J.K., Anzai, T., Shiina, T. and Inoko, H. 2004. Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. Mol. Biol. Evol. 21, 2079-2091.

Kumar, S., Tamura, K., Jakobsen, I.B. and Nei, M. 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Lafont, B.A.P., Buckler-White, A., Plishka, R., Buckler, C. and Martin, M.A. 2003. Characterization of pig-tailed macaque classical MHC class I genes: implications for MHC evolution and antigen presentation in macaques. J. Immunol. 171, 875-885.

Lambracht-Washington, D. and Fischer Lindahl K. 2002. Does the rat have an H2-D orthologue next to Bat1? Immunogenetics 53, 1039-1046.

Lawlor, D.A, E., Taylor, P., and Parham, P. 1991. Gorilla class I major histocompatibility complex alleles: comparison to human and chimpanzee class I. J. Exp. Med. 174, 1491-1509.

Lazaro AM, Moraes ME, Marcos CY, Moraes JR, Fernandez-Vina MA, Stastny P. 1999. Evolution of HLA-class I compared to HLA-class II polymorphism in Terena, a South-American Indian tribe. Hum Immunol. 60, 1138-1149.

Lehner, B., Semple, J.I., Brown, S.E., Counsell, D., Campbell, R.D. and Sanderson, C.M. 2004. Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. Genomics 83, 153-167.

Leonhardt, R.M., Keusekotten, K., Bekpen, C. and Knittler, M.R. 2005. Critical role for the tapasin-docking site of TAP2 in the functional integrity of the MHC class I-peptide-loading complex. J. Immunol. 175, 5104-5114.

Lepourcelet, M., Andrieux, N., Giffon, T., Pichon, L., Hampe, A., Galibert, F. and Mosser, J. 1996. Systematic sequencing of the human HLA-A/HLA-F region: establishment of a cosmid contig and identification of a new gene cluster within 37 kb of sequence. Genomics 37, 316-326.

Lie, J.T., Kobaysshi, S., Tokano,Y. and Hashimoto, H. 1995. Systemic and cerebral vasculitis coexisting with disseminated coagulopathy in systemic lupus erythematosus associated with antiphospholipid syndrome. J. Rheumatol. 22, 2173-2176.

Liu W, Zhao Y, Liu Z, Zhang Y, Lian Z, Li N. 2006. Construction of a 7-fold BAC library and cytogenetic mapping of 10 genes in the giant panda (*Ailuropoda melanoleuca*).BMC Genomics. 17, 1-8.

Lundberg AS, 1992. Evolution of major histocompatibility complex class II allelic diversity: direct descent in mice and humans. Proc. Natl. Acad. Sci. USA 89, 6545-6549.

Luo, M.-J.; Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M. and Reed, R. 2001. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. Nature 413, 644-647.

Luthra-Guptasarma, B. and Singh, B. 2004. HLA-B27 lacking associated beta2-microglobulin rearranges to auto-display or cross-display residues 169-181: a novel molecular mechanism for spondyloarthropathies. FEBS Lett. 575, 1-8.

Martin, R.D. 1990. Primate Origins and Evolution: A Phylogenetic Reconstruction. Princeton University Press: Princeton, N.J.

Matsunaga, T. and Rahman, A. 1998. What brought the adaptive immune system to vertebrates? Immunological Reviews 166, 177-186.

Mattiuz, P.L., Ihde, D., Piazza, A., Ceppelini, R., Bodmer, W.F. 1970. New approaches to the population genetic and segregation analysis of the HLA system. In: Terasaki P, ed. Histocompatibility Testing, Copenhagen: Munksgaard, 1971, 193-205.

Miller, M.M., Goto, R., Bernot, A., Zoorob, R., Auffray, C., Bumstead, N. and Briles, W.E. 1994. Two Mhc class I and two Mhc class II genes map to the chicken Rfp-Y system outside the B complex. Proc. Natl. Acad. Sci. USA 91, 4397-4401.

Milner, C.M. and Campbell, R.D. 2001. Genetic organization of the human MHC class III region. Front. Biosci. 6, D914-926.

Mizuki, N., Ando, H., Kimura, M., Ohno, S., Miyata, S., Yamazaki, M., Tashiro, H., Watanabe, K., Ono, A., Taguchi, S., Sugawara, C., Fukuzumi, Y., Okumura, K., Goto, K., Ishihara, M., Nakamura, S., Yonemoto, J., Kikuti, Y.Y., Shiina, T., Chen, L., Ando, A., Ikemura, T. and Inoko, H. 1997. Nucleotide sequence analysis of the HLA class I region spanning the 237-kb segment around the HLA-B and -C genes. Genomics 42, 55-66.

Nagamachi, C.Y., Pieczarka, J.C., Schwarz, M., Barros, R.M., Mattevi, M.S..1997. Chromosomal similarities and differences between tamarins, Leontopithecus and Saguinus (Platyrrhini, Primates). Am. J. Primatol. 43, 265-276.

Nair, S., Dey, R., Sanford, J.P. and Doyle, D. 1992. Molecular cloning and analysis of an eIF-4A-related rat liver nuclear protein. J. Biol. Chem. 267, 12928-12935.

Nardi, G., Cadiz, C., Lachman, J. and Cornelio, C. 2003. Advances in hereditary hemochromatosis. Acta Gastroenterol. Latinoam. 33, 103-107.

Neff, J. 2005. Genomische und evolutionäre Analyse von MHC-Klasse-I-Genen bei einem Halbaffen (*Microcebus murinus*). Tiermedizin Hochschule Hannover. Dissertation.

Nei, M., and Gojobori, T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418-442.

Neville, M. J. and Campbell, R. D. 1999. A new member of the Ig superfamily and a V-ATPase G subunit are among the predicted products of novel genes close to the TNF locus in the human MHC. J. Immun. 162, 4745-4754.

Norcross, J.L. and Newman, J.D. 1993. Context and Gender-specific Differences in the Acoustic Structure of Common Marmoset (*Callithrix jacchus*) Phee Calls. Am. J. Primatology 30, 37-54.

Ohta, T. 1991. Role of diversifying selection and gene conversion in evolution of Major Histocompatibility Complex loci. Proc. Natl. Acad. Sci. USA 15, 6716-6720.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M. and Hamada, H. 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. Cell 60, 461-472.

Orimo, A., Yamagishi, T., Tominaga, N., Yamauchi, Y., Hishinuma, T., Okada, K., Suzuki, M., Sato, M., Nogi, Y., Suzuki, H., Inoue, S., Yoshimura, K., Shimizu, Y. and Muramatsu, M. 2000. Molecular cloning of testis-abundant finger Protein/Ring finger protein 23 (RNF23), a novel RING-B box-coiled coil-B30.2 protein on the class I region of the human MHC. Biochem. Biophys. Res. Commun. 276, 45-51.

Osoegawa, K., Vessere, G.M., Li Shu, C., Hoskins, R.A., Abad, J.P., de Pablos, B., Villasante, A., and de Jong, J. 2007. BAC clones generated from sheared DNA. Genomics 89, 291-299.

Ota, M., Katsuyama, Y., Kimura, A., Tsuchiya, K., Kondo, M., Naruse, T., Mizuki, N., Itoh, K., Sasazuki, T. and Inoko, H. 2001. A second susceptibility gene for developing rheumatoid arthritis in the human MHC is localized within a 70-kb interval telomeric of the TNF genes in the HLA class III region. Genomics 71, 263-270.

Otting, N., Heijmans, C.M., Noort, R.C., de Groot, N.G., Doxiadis, G.G., van Rood, J.J., Watkins, D.I. and Bontrop, R.E. 2005. Unparalleled complexity of the MHC class I region in rhesus macaques. Proc. Natl. Acad. Sci. USA 102, 1626-1631.

Pamer, E. and Cresswell, P. 1998. Mechanisms of MHC class I--restricted antigen processing. Annu. Rev. Immunol. 16, 323-358.

Parham, P. and Ohta, T. 1996. Population biology of antigen presentation by MHC class I molecules. Science 272, 67.

Pascalis, H., Heraud, J.M., Fendel, R., Lavergne, A. and Kazanji, M. 2003. Molecular characterization of major histocompatibility complex class 1 (MHC-I) from squirrel monkeys (Saimiri sciureus). Immunogenetics 55, 633-639.

Peelman, L.J., Chardon, P., Nunes, M., Renard, C., Geffrotin, C., Vaiman, M., Van Zeveren, A., Coppieters, W., van de Weghe, A., Bouquet,Y. 1995. The BAT1 gene in the MHC encodes an evolutionarily conserved putative nuclear RNA helicase of the DEAD family. Genomics 26, 210-218.

Pellet, P., Vaneensberghe, C., Debre, P., Sumyuen, M.H. and Theodorou, I. 1999. MIC genes in non-human primates. Eur. J. Immunogenet. 2, 239-241.

Pham-Dinh, D., Gaspera, B.D., De Rosbo, N.K. and Dautigny, A. 1995. Structure of the human myelin/oligodendrocyte glycoprotein gene and multiple alternative spliced isoforms. Genomics 29, 345-352.

Pham-Dinh, D., Jones, E.P., Pitiot, G., Della Gaspera, B., Daubas, P., Mallet, J., Le Paslier, D., Lindahl, C.F. and Dautigny, A. 1995. Physical mapping of the human and mouse MOG gene at the distal end of the MHC class Ib region. Immunogenetics 42, 386-391.

Pham-Dinh, D., Mattei, M. G., Nussbaum, J. L., Roussel, G., Pontarotti, P., Roeckel, N., Mather, I. H., Artzt, K., Lindahl, K. F. and Dautigny, A. 1993. Myelin/oligodendrocyte glycoprotein is a member of a subset of the immunoglobulin superfamily encoded within the major histocompatibility complex. Proc. Nat. Acad. Sci. USA 90, 7990-7994.

Ploegh, H. 1998. Viral strategies of immune evasion. Science 280, 248-253.

Rahman, A., Friedman, D.S. and Goldstein, L.S. 1998. Two kinesin light chain genes in mice. Identification and characterization of the encoded proteins. J. Biol. Chem. 273, 15395-15403.

Reveille, J.D. 2000. The changing spectrum of rheumatic disease in human immunodeficiency virus infection. Semin Arthritis Rheum. 30, 147-166.

Reymond, A., Meroni, G., Fantozzi, A., Merla, G., Cairo, S., Luzi, L., Riganelli, D., Zanaria, E., Messali, S., Cainarca, S., Guffanti, A., Minucci, S., Pelicci, P.G. and Ballabio, A. 2001. The tripartite motif family identifies cell compartments. EMBO J. 20, 2140-2151.

Rhodes, D.A. and Trowsdale, J. 1999. Genetics and molecular genetics of the MHC. Rev. Immunogenet. 1, 21-31.

Robinson, J. and Marsh, S.G. 2000. The IMGT/HLA sequence database. Rev. Immunogenet. 2, 518-531.

Rolleke, U., Flugge, G., Plehm, S., Schlumbohm, C., Armstrong, V.W., Dressel, R., Uchanska-Ziegler, B., Ziegler, A., Fuchs, E., Czeh, B. and Walter, L. 2006. Differential expression of major histocompatibility complex class I molecules in the brain of a New World monkey, the common marmoset (Callithrix jacchus). J. Neuroimmunol. 176, 39-50.

Roth, M.P., Malfroy, L., Offer, C., Sevin, J., Enault, G., Borot, N., Pontarotti, P. and Coppin, H. 1995. The human myelin oligodendrocyte glycoprotein (MOG) gene: complete nucleotide sequence and structural characterization. Genomics 28, 241-250.

Rowe, N. 1996. The pictorial guide to the living primates. East Hampton (NY): Pogonias Pr. 263 p.

Rylands, A.B., Coimbra-Filho, A.F., and Mittermeier, R.A. 1993. Systematics, Geographic Distribution, and Some Notes on the Conservation Status of the

Callitrichidae. In: Marmosets and Tamarins: Systematics, Behaviour, and Ecology. Ed. A.B. Rylands. Oxford University Press: Oxford.

Saitou, N. and Nei, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406-425.

Sawai, H., Kawamoto, Y., Takahata, N. and Satta, Y. 2004. Evolutionary relationships of major histocompatibility complex class I genes in simian primates. Genetics 166, 1897-1907.

Sawai, H., Kawamoto, Y., Takahata, N. and Satta, Y. 2004. Evolutionary relationships of major histocompatibility complex class I genes in simian primates. Genetics 166, 1897-1907.

Scholer, H.R., Ruppert, S., Suzuki, N., Chowdhury, K. and Gruss, P. 1990. New type of POU domain in germ line-specific protein Oct-4. Nature 344, 435-439.

Schrader, L. and Todt, D. 1993. Contact Call Parameters Covary with Social Context in Common Marmosets, *Callithrix j. jacchus*. Animal Behaviour 46, 1026-1028.

Seo, J.; Walter, L. and Günther E. 2001. Genomic analysis of MIC genes in rhesus macaques. Tissue Antigens 58, 159-206.

Seo, J.W.; Bontrop, R. Walter, L. and Günther E. 1999. Major histocompatibility complex linked MIC genes in rhesus macaques and other primates. Immunogenerics 50, 358-362.

Shiina, S., Tamiya, G., Oka, A. and Inoko, H. 2001. Published Only in Database. Homo sapiens 2,229,817bp genomic DNA of 6p21.3 HLA class I region.

Snell, G.D. 1948. Methods for the study of histocompatibility genes. J. Genet. 49, 87-108.

Snell, G.D. 1951. A fifth allele at the histocompatibility-2 locus of the mouse as determined by tumor transplantation. J. Natl. Cancer Inst. 11, 1299-1305.

Snoek, M., Groot, P.C., Spies, T., Campbell, R.D. and Demant, P. 1991. Fine mapping of the crossover-sites in the C4-H-2D region of H-2 recombinant mouse strains. Immunogenetics 34, 409-412.

Spies, T., Bresnahan, M. and Strominger, J.L. 1989. Human major histo-compatibility complex contains a minimum of 19 genes between the complement cluster and HLA-B. Proc. Natl. Acad. Sci. USA 86, 8955-8958.

Stein, J. 2001. The magnocellular theory of developmental dyslexia. Dyslexia 7, 12-36.

Stephen, A.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389-3402.

Stevenson, M.F. and Poole, T.B. 1976. An Ethogram of the Common Marmoset (*Callithrix jacchus jacchus*): General Behavioural Repertoire. Animal Behaviour 24, 428-451.

Stevenson, M.F. and Rylands, A.B. 1988. The marmosets, genus *Callithrix*. In: Mittermeier RA, Rylands AB, Coimbra-Filho AF, da Fonseca GAB, editors. Ecology and behavior of neotropical primates, Volume 2. Washington DC: World Wildlife Fund. p 131-222.

Stewart, C. A., Horton, R., Allcock, R.J.N., Ashurst, J.L., Atrazhev, A.M., Coggill, P., Dunham, I., Forbes, S., Halls, K., Howson, J.M.M., Humphray, S.J., Hunt, S., Mungall, A.J., Osoegawa, K., Palmer, S., Roberts, A.N., Rogers, J., Sims, S., Wang, Y., Wilming L.G., Elliott, J.F., de Jong, P.J., Sawcer, S., Todd, J., Trowsdale, J. and Beck, S. 2004. Complete MHC Haplotype Sequencing for Common Disease Gene Mapping. Genome Research 14, 1176-1187.

Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D., Altschul,S.F., Zeeberg,B., Buetow, K.H., Schaefer, C.F., Bhat, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F., Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J., Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J., McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S., Worley,K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W., Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,

Bouffard,G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smailus,D.E., Schnerch, A., Schein, J.E., Jones, S.J. and Marra, M.A. CONSRTM Mammalian Gene Collection Program Team. 2002. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. Proc. Natl. Acad. Sci. USA 99, 16899-16903.

Suarez, C.F., Cardenas, P.P., Llanos-Ballestas, E.J., Martinez, P., Obregon, M., Patarroyo, M.E. and Patarroyo, M.A. 2003. Alpha1 and alpha2 domains of Aotus MHC class I and Catarrhini MHC class Ia share similar characteristics. Tissue Antigens. 61, 362-373.

Sullivan, L.C., Hoare, H.L., McCluskey, J., Rossjohn, J. and Brooks, A.G. 2006. A structural perspective on MHC class Ib molecules in adaptive immunity. Trends Immunol. 27, 413-420.

't Hart, B.A., van Meurs, M., Brok, H.P., Massacesi, L., Bauer, J., Boon, L., Bontrop, R.E. and Laman, J.D. 2000. A new primate model for multiple sclerosis in the common marmoset. Immunol Today 21, 290-297.

Takeda, J., Seino, S. and Bell, G. I. 1992. Human Oct3 gene family: cDNA sequences, alternative splicing, gene organization, chromosomal location, and expression at low levels in adult tissues. Nucleic Acids Res. 20, 4613-4620.

Tallmadge, R.L., Lear, T.L. and Antczak, D.F. 2005. Genomic characterization of MHC class I genes of the horse. Immunogenetics 57, 763-774.

Teraoka, Y., Naruse, T.K., Oka, A., Matsuzawa, Y., Shiina, T., Iizuka, M., Iwashita, K., Ozawa, A. and Inoko, H. 2000. Genetic polymorphisms in the cell growth regulated gene, SC1 telomeric of the HLA-C gene and lack of association of psoriasis vulgaris. Tissue Antigens 55, 206-211.

The MHC sequencing consortium. Complete sequence and gene map of a human major histocompatibility complex. 1999. Nature 401, 921-923.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence

weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Timon, M., Elgar, G., Habu, S., Okumura, K. and Beverley, P.C. 1998. Molecular cloning of major histocompatibility complex class I cDNAs from the pufferfish Fugu rubripes. Immunogenetics 47, 170-173.

Trowsdale J. 1995. "Both man & bird & beast": comparative organization of MHC genes. Immunogenetics 41, 1-17.

Trowsdale, J. 1993. Genomic structure and function in the MHC. Trends Genet. 9, 117-122

Trtkova, K., Kupfermann, H., Grahovac, B., Mayer, W.E., O'hUigin, C., Tichy, H., Bontrop, R., and Klein, J. 1993. Mhc-DRB genes of platyrrhine primates. Immunogenetics 38, 210-222.

Trtkova, K., Mayer, W.E., O'Huigin, C. and Klein, J. 1995. Mhc-DRB genes and the origin of New World monkeys. Mol. Phylogenet. Evol. 4, 408-419.

Warren, R.D. and Crompton, R.H. 1997. Locomotor ecology of Lepilemur edwardsi and Avahi occidentalis. Am. J. Phys. Anthropology 104, 471-486.

Watkins, D.I., Chen, Z.W., Hughes, A.L., Evans, M.G., Tedder, T.F. and Letvin, N.L. 1990. Evolution of the MHC class I genes of a New World primate from ancestral homologues of human non-classical genes. Nature 346, 60-63.

Watkins, D.I., Garber, T.L., Chen, Z.W., Toukatly, G., Hughes, A.L. and Letvin, N.L. 1991. Unusually limited nucleotide sequence variation of the expressed major histocompatibility complex class I genes of a New World primate species (Saguinus oedipus). Immunogenetics 33, 79-89.

Watkins, D.I., Hodi, F.S. and Letvin, N.L. 1988. A primate species with limited major histocompatibility complex class I polymorphism. Proc. Natl. Acad. Sci. USA 85, 7714-7718.

Watkins, D.I., Kannagi, M., Stone, M.E. and Letvin NL. 1988. Major histocompatibility complex class I molecules of nonhuman primates. Eur. J. Immunol. 18, 1425-1433.

Watkins, D.I., Letvin, N.L., Hughes, A,L. and Tedder, T.F. 1990. Molecular cloning of cDNA that encode MHC class I molecules from a New World primate

(Saguinus oedipus). Natural selection acts at positions that may affect peptide presentation to T cells. J. Immunol. 144, 1136-1143.

Watkins, D.I. 1995. The evolution of major histocompatibility class I genes in primates. Crit. Rev. Immunol. 15, 1-29.

Weiss, E.H., Kuon, W., Dorner, C., Lang, M. and Riethmuller, G. 1985. Organization, sequence and expression of the HLA-B27 gene: a molecular approach to analyze HLA and disease associations. Immunobiology 170, 367-380.

Weitkamp, L.R., Moss, A.J., Lewis, R.A., Hall, W.J., MacCluer, J.W., Schwartz, P.J., Locati, E.H., Tzivoni, D., Vincent, G.M. and Robinson, J.L. 1994. Analysis of HLA and disease susceptibility: chromosome 6 genes and sex influence long-QT phenotype. Am. J. Hum. Genet. 55, 1230-1241.

Wheeler, D.L., Church, D.M., Lash, A.E., Leipe, D.D., Madden, T.L., Pontius, J.U., Schuler, G.D., Schriml, L.M., Tatusova, T.A., Wagner, L. and Rapp, B.A. 2001. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 29, 11-16.

Wong, Y.T., Del-Rio-Martin, J., Jaques, B., Shaw, J.A. and Talbot, D. 2005. Audit of diabetes in a renal transplant population. Transplant Proc. 37, 3283-3285.

Wroblewski, J.M., Kaminsky, S.G. and Nakamura, I. 1994. Bat-1 genes and the origin of multiple class I loci in the H-2D region. Immunogenetics 39, 276-280.

Wroblewski, J.M., Kaminsky, S.G., Milisauskas, V.K., Pittman, A.M., Chaplin, D.D., Spies, T. and Nakamura, I. 1990. The B144-H-2Db interval and the location of a mouse homologue of the human D6S81E locus. Immunogenetics 32, 200-204.

Wu, M.S., Tani, K., Sugiyama, H., Hibino, H., Izawa, K., Tanabe, T., Nakazaki, Y., Ishii, H., Ohashi, J., Hohjoh, H., Iseki, T., Tojo, A., Nakamura, Y., Tanioka, Y., Tokunaga, K. and Asano, S. 2000. MHC (major histocompatibility complex)-DRB genes and polymorphisms in common marmoset. J. Mol. Evol. 51, 214–222.

Yoshino, M., Xiao, H., Jones, E.P. and Fischer Lindahl, K. 1998. BAC/YAC contigs from the H2-M region of mouse Chr 17 define gene order as Znf173-Tctex5-mog-D17Tu42-M3-M2. Immunogenetics 47, 371-380.

Yuhki, N., Beck, T., Stephens, R.M., Nishigaki, Y., Newmann, K. and O'Brien, S.J. 2003. Comparative Genome Organization of Human, Murine, and Feline MHC Class II Region. Genome Research 13, 1169-1179.

Zhang, J., Zhang, L., Zhao, S. and Lee, E.Y. 1998. Identification and characterization of the human HCG V gene product as a novel inhibitor of protein phosphatase-1 T. Biochemistry 37, 16728-16734.

9 Annex

1 - 55: The nucleotide sequences obtained from the MHC class-I region of *Callithrix jacchus.* Sequences determined by the author with two exceptions (TS1 andTS2) stated in the text

1. 204C3 ptarbac3

AAATGATCTTCAACGGGAATTCCTGTTCCACAGCGGACAGCGGTGTGAGCGCCCCAAAGCTGGAAATCAGAACCCTAAAG
GCAGCTCGGCCCTCCCCAGCCCACAGCGCCGTTATTCCGTTTCTATATCGGTAAACACGCTTCATTTTCCATAGCCCAGGG
CGGGGTGACGGGTGATCCCAGTCTTCGCAGTGAACTGCGGGCAGTAGAATTCAAAAAGCTTGCGGGCAAAGCGGGGCGC
GGTGGCCCACGCCTGTAATGCCAGTACGTTTGGGAGGTCGAGGCTGGCGGATCACCTGAGGTAGGGAGTTCGAGATCAG
CCTGACCAACATGGAGAAACCCTGTCTCTACTAAAAATACAAAATTACCGGGCATGGGGCAACTGCCTGTAATCCCCCTATT
CGTAGCTGGCGGCAGGAAAAGGGTTGCCCGAAGGCAA

2. 204C3 T729

TCTACTTCTGAGTAGCTGGTATTATTGGCACCGCTGCCACCATGTCCGGCTAATTTTAAATTTTTTTGTAGAGACAGGGTCTT
GTCATGTTGTCCAGGCTGGTCTCAAACTCCTGGCCTCAAGCAATTCTCTAGCCTTGGCCTCCCAAAGTGCTGGGATTACAA
GTGTGAGCCATTGTGCCTAGCCAGTTTTTTGAATATTTTAAGCCTACTGTTGAGGCCTACTGCTCTGAAAAATAAAATTTCTT
TAAAATGATTACTGCTCATTGACAATGTACATGGTTACCCAAGAGCCTCTAGCTCTGATGAAGATGTCCAAAGAGATGTATG
TTTGTTTTCATGCCTGCTAACATAACATCCAGCCTGTACGCCCATGGTTTAAGGAGTAATTTAGACTTTCAAGTATTATTATTT
AAGAAATACATTTTATAAGGTTGTATCTTCCATAGATAGTGAATCCTTTGATGGATCTGGGCAATGTAAACTGAAAGCCTCTG
GAAGAGATTTTACCCAGATTCTAATTGCTCACATTTAGGACACATTTCACACACTGCCTGGGGCAATTGGGCGAGCCACCAC
ATGGGCTTGTTAATCCCCAGAGCCCTTTTGGCCAGGGCACA

3. 510K19 ptarbac3

AGTACTTCTAGGGAATTCATTACCACCAGCACCTGCATTACGAGAACTCCTGAAGGAAGCACTAAATTTGGAAAGGAAAGAT
AGTTACCAGCCACTGTGAAAACACACTCAAGCAAGTACAGACTAGTGACACAATAAAGCAACCACTTAAGCAAGTCTGCAG
CATATCATGGCGTGTAATATGAGAGACGCCGTACGTTTCTTTGTTACC

4. 217M17 T3

AGTACTTATAGGGAATTATTACACCAGAACCTGCATTACGAGAACTCCTGAAGGAAGCACTAAATTTGGAAAGGAAAGATAG
TTACCAGCCACTGTGAAAACACACTCAAGCAAGTACAGACTAGTGACACAATAAAGCAACCACTTAAGCAAGTCTGCAGCAT
ATCATGGCGTGTAATATGAGAGACGCCGTACGTTTCTTTGTATA

5. 217M17T729

CGTCTTCATAGTTGGAATTCCGAGTCCTGGCTGGAGCGCGATCTGGAATCTGACTCGCTTGAATTCTGACTCGGTGGATTC
GGATCCGGGTGAGTAGGGAAATGCACCTCAGCCCCTCCCACGGGCGCCCACGGATTCTTGGATCCGAAAACGCTTCCTGC
TGCTCAGTCACCCCAGGAAGGCAGCGCCCGCCTCTGGGCTGTTCTGATGGAAACAGGCTCCGCCGCCCGCAGGAAAACC
CACAACTAAGGGGCCAGGAAAAAGCCCCTCAGGATCTCGCCACTTCAGTGAGGATCCTAATTTACACCCTCCATCAAAGAC
TGCAGCCGCGTAACCTGGCAAAATCGGTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAG

6. 347K7 ptarbac3

TAAGTACTTCATAGGGAATTCCTTGACCTCAAGTGACCAGTCCGCCTTGGCCTCCCAAAGTACTGGGATTACAGGCAGGAA
CCACCGCACCTGGCCACTAAACACGCTTTAACTTCTTTCCTCCTTCCGTCCCTCTTACTGTATCCTTCTAGGATTCCCCAAG
TTCTGTTAAAAGCTTCTGGAACTGGAATCTAGCAAATGAATGTCCCATTTAATAGGCAGAGAAGGAGGAGGTGGGTGAATCA
CAGAACCAAAGTGCATGTGGTGACAAGCTCCAGTTGTGCTCAGGGTGGGGTAGCTGGCAGGGCCCTGATGCATTCAGTCA
TTGGTTCTGCAAATGTTTGCATAGTTCCACCAGCAGCATCTGCCCGGCAGAGAGCAGGACCAGGGCGAGGGTGAGGACAA
GAGAGAGCACAGAGCCTGGATGGGGTAGTCAGCCACACAGGAACAACATGTGTAGGCTGAGTCATATCATGGCGTTGTAA
TATGAG

7. 347K7 ptarbac4

GCTCCTCAATTACACGCCTGTATATGCTGCAGCCTACACATGTTGTTCCTGTGTGGCTGACTACCCCATCCAGGCTCTGTCT
CTCTCTTGTCCTCACCCTCGCCCTGGTCCTGCTCTCTGCCGGGCAGATGCTGCTGGTGGAACTATGCAAACATTTGCAGAA
CCAATGACTGAATGCATCAGGGCCCTGCCAGCTACCCCACCCTGAGCACAACTGGAGCTTGTCACCACATTGCACTTTGGT
TCTGTGATTCACCCACCTCCTCCTTCTCTGCCTATTAAATGGGACATTCATTTGCTAGATTCCAGTTCCAGAAGCTTTTAACA
GAACTTGGGGAATCCTAGAAGGATACGTTGAGGGAC

8. 347D1 ptarbac3

GGTACTTCTAGGGAATTCCTGGACTCTCAGTGACTCACGTTCTGGCCTCCCAAAGTACTGGGATTACAGGCAGGAACACAC
CGCACCTGGCCAGCTAAACACGCTTTAACTTCTTTCCTCCTTCCGTCCCTCATTACTGTATCCTTCTAGGATTCCCCAAGTTC
TGTTAAAAGCTTCTGGAACTGGAATCTACGCAAATGAATGTCCCATTTAATAGGCAAAGAAGGAGGAGGTGGGTGAATCAC
AGAACCAAAGTGCATTGGTGGTGACAAGCTCCAGTTGTGCTCAGGGTGGGGTAGCTGGCAGGGCCCTGATGCATTCAGAT
CATTGGTTTCTGCAAATGTTTGCATAGTTCCACCAGCACCATCTGCCCGGCAGAGAGCAGGACCAGGGCGAGGGTGAGGA
CAAGAGAGAGACAGAGCCTGGATGGGGTAGTCAGCCACACAGGAACAACATGTGTAGGCTGCAGCATATCATGGCGTGTA
ATATGAGAGACGCCGTACGTTTCTTTGTATAGC

9. 347D1 ptarbac4

TCTCTCTATTACACGCCTGATATGCTGCAGCCTACACATGTTGTTCCTGTGTGGCTGACTACCCCATCCAGAGCTCTGTCTC
TCTCTTGTCCTCACCCTCGCCCTGGTCCTGCTCTCTGCCGGCGCAGCATGCTGCTGGTGGAACTATGCAAACATTTTGCAC
GCAACCAATGACTGATTGCATCAGGGCCCTGCCAGCTACCCCACCCTGAGCACAACTGGAGCTTGTACCCAGCATTGCACT
TTGGTTCTGATGATTCACCCACCTCCTCCTTCTCTGCCTATTAAATGGGACATTCATTTGCTTAGATTCCAGTTCCAGAAGCT
TTTAACAGAACTTGGGGAATCCTAGCGCAAGGATACAGTAAGAGGGACGGAAGGAGGAAAGAATGTTAAAGCGTGTTTAGT
GTGGCCAGGTGCCGCGTGGTTCCTGCCTGTAATCCCAGCTAGCTTTGGGAGGCCAAGGCGGACTGGCTCACTCGAGTGT
CAGGAATTCCCTATGAAGATCCTTCTATAGTGTCACCTAAATGTAAT

10. 347D1 TCF19

AGCTCTCTCCAGGCTGGGGCTGGGGCTGGCTTCCGCCCTATGCTGCCCTCCCTCGGGGCTCCACAGATCGGCCTCTTAG
CACCTTGTCCTCTACCCCCAAGGCCACACTGATCCTGAACTCCATTGGCAGCCTCAGCAAGCTCCGGCCCCAGCCCCTCA
CCTTCTCCCCTAGGGGGGGTAGACCAAAGATCTTGCCTGTTTCCGCCCCACCTGGGGAAGTGGGTATCATACCTTCTGCTC
CACCCCCACGCAATCGGAGGAAATCTGCTCACCAAGTGTTAGCAGAACTGGATGATGAGAGTGAGCAGATTTCCTCGATGC
GTGAGTGAGAGGTGGAGGACAAGATATAACTTCCGGTGGGGCGAAAGCATATTTTGTGTATCTCCCCCTAGGGAGCAGAG
AGAAGTGGAGGGTCTGGCACTCTGAACTCTCTGCCATAGTATGCGGATGGTATACACCTTCTGTGTAGACGACTACGTGCA
GGCAGCAGTAGCAGCATGACATCCGACGCTAGCAGCGTGCGGTGACACGATGATCGACACTGACGAGACGACACGACAC
GAGACGTACGTAGCTGATGACATAGCAGCGCTAGTCGTCGTAAGTCGAGCTAGACCATGACTGATCGAGTCAGTACATCAT
GTCACGACTGCTACGACTGACATCGACTGACGTAGTGCGCGCATGTACGACTGCGACTGATCACATGCATGTACGCAGTA
GCGAGCATGCATGCATGCTACGTCGACAGCTTGCATGGTAGCTACGATCAGCTGATAGCGATCGTACGATAGCTATAGATG
CTCAGATCTAGTCGAGAGCATACGTACGTACTGCAGGCTAGAAGCTAGCTATGCGATGCATACGATGCATGAGATGCTAGC
TCGCATGACGCGTACGACTACGGACTAGTCATGTAGATGACTCGAGTACGTACTGACTATCGATCGATATGATCAGATGCT
AGTGTAGTATCATCAGATCAGCTATCACATTAGTAGATGTAGTCGACGTATCGATGATCATGTAGACTGATGACTGATAGAT
CGATATACGATCGACGACTGACGATCGTCGGATACGTACGTAGCAGTCAGAGCCTCGCTGCTAGCAGTAGCGCACGCGAG
AGATGCATGCTTCAACGTAGTACGATCAGTCATCGACTGCACTGCGTCGATGACGATCGATGCACGATCAGTAGCTGCTGT
GAGCGATCGTATCAGCAGTACTAGTCAACGACGCAGACATCGTACGATCACAGTCGTACGTAGAGAGCAGTGATGTACGTA
CTGACGTAGCTAGTAGTAGCATGATCGTACTGACTGATCGATGATGCTATGTGCATGACTGCATCTACGTCAGCACGTAGC
AGTACGTCATGAGATGYCAGTCACTGACTGACGATGTCGACTGATGCATGACTGTAGTACTGATGAGCAGCTAGTGATCTG
TACGCACATGCAGCAGTACGATGACGTACAGCTGCCTAGATCGATCGATCAGATCATAGACGCATATCTCAGAGTCATCAC
GTACGATCGACTGACGAGATGATGATAGCAGCATGCAGAGTACGACGATACGATCACTAGTAGATGAGTCAGTGCATGCTA
GCACGTGTAGCATGACGATCGATGACTGATCGTCAGTCACTGATGAGAGAATACGTGTCAGTCATGCATGTAGTTCAGTAT
GAATCAGTTACCGTCATTGACATGTCAGTCAGCACTGCACTAGTGAGATCACGTAGTCAGCGATGATGTAGTGACTCAGCC
TGCTAGCTATCGACGAGCTAGCAGTACGTTGCAGACTGTACGTACTACGTGTCAGCGTAGTCAGTCTACGTCAGTCATACT
GCAGTCAAGCTACTCTGACTAGTCGATCGACTGGCTACGTCAGTACGTACGCATGCTGCATGCTATCGTGCAGCTAGTGTC
GACTGACGTCAGCATGCAGTCAGTCAGATGATCATCGAGATGCATCGGTCATCAGATCAGCTAAGCTAGCAGATCGACTGA
GACTGAGACGCAGCACTGATGACCTGACGATGCATGACTGACGTACGTCGATGCATGTCGATCGAGCTAGCACAGTCGCA

GTAACCTATACGTACGTAGTCAACGATCCTGAATGACTATGCACTGCAGTCAGTCAGTCAGCTAGCTAACGTACTGAGTCAG
ACGTACGTAGCTAGCTAGCGTACGTACTGAGTCACGATTCATGCTCATGATCATGCTAAGACGAGTACTAGCTCATTACGTG
ATGCGATCGACGACGTACGACTACTGACGTACGACTGCTAGTCGTACAGCTG

11. 459O20 ptarbac3

AGTACTTCATAGGGAATTCTTCAGCCCATGATCCGCAAAACCTCCACATCCTCTCTCGGGATGTTCCTGCCCCTCGCAGCT
CCAGCAGGAACCTGGGTTTCCCGGAAGACACAGACCCTTCGGGAGTCCTCCCACCTGGAGAGTTTCCCCAGTGACTTACA
CCTCTGGACTCACTTCTCACTGTGGTTTTGAGAACTTTCTCTCTCCCTCCTCCCTAACCTCCCTAAGCTGTCATCAGATTAG
GGCCCCACTCCCCTCATTGTAGCCATTCCTTGGGGGCCCCAGACCATTCCTCTCAATCCTGACTCTTGTAGCTCCTGGTTC
ACTGGCTCGCTCTCCAGCAGTGCTGTCTCCTTCATCTTGATGACTTCAACATATGCAGATGTGGTGGGCTGAGTGATGGTC
TCCACATAAGTTCTGTGTCAGTCCTTAGAACCTGTGAACAGGCTGAATTACATGGCAAAAGGGACACCGCTCATGTAATGAA
GATTAGGGAACTTAAAATAGGGAGATTATCCTGGACTATTCGCATGGGCCCAGTCAAATGACATGAGCCATTAAAAGCAGA
GAACCAACTCTGGCTGGAGTCAGAGATGTCGCGGAAGAGGGAGGGAAGAAGAGACA

12. 459O20 ptarbac4

TCATATTACACGCCATGTATATGCTGCAGCTCGTAGGGGCAGGTGCAGTGTAGGCAGCGGTGGAATCCACCATCTGTGTAG
TTTTGCCAAGTAAGTGTGACAACACAAGGGAGAGGGGAGGGAGGGATGGAGATTATTTACCATAGAATCCAAATCCGGGG
AGGAGGGAGTAGAGGACACCCAGGGCTGAGGGACAATGGGGAGATGGCAGGATCACTAGATTGGGAATCCTGGGGATTG
GGTGTGTCRAAGGATTGCTGGAATTGATGTACTACAAGGTGGGCACCAAGTGTGAAATACAGGTACCTACAAAATGAATTAT
TTGCTGATATCAAATTACAGCATATGGTAACATGACAGTATCTGTGTCCTTAAGAGCCTGTGGCCACCTTGCAAGGTGTGTA
TGAGGGGGAAGATGGTCATGAGAGTGGAAGTGTGGATACTGTCCAGCTACTCTTCTAGAGGGGCTTGGGGTTCATTGGGC
CCTGAAGGGAGCGCCTAGGATCGATTC

13. 459O20 T729

TGTACTTCATAGTGGAATTCCATCAGTGCAGACTCAGGGACCACTGAGAACTACGTCACTGCAGCCGCGTAACCTGGCAAA
ATCGGTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAG

14. 459O20 T729

ATTGCAGCTCGTAGGGGCAGGTGCAGTGTAGGCAGCGGTGGAATCCACCATCTGTGTAGTTTTGCCAAGTAAGTGTGACA
ACACAAGGGAGAGGGGAGGGAGGGATGGAGATTATTTACCATAGAATCCAAATCCGGGGAGGAGGGAGTAGAGGACACC

15. 459O20 TCF19

CGAGCTCACTAGGAGGAGAGCTCAGGCTAGAGAGCTGGGGCTGGCTTCCGCTCCTATGCTGCCCTCACCACTACAGCGCT
CCACAGCGGCCTCTTAACACCTTGTCCTCTACCCCCAAGGCCACACTGATCCTGAACTCCATTGGCAGCCTCAGCAAGCTC
CGGCCCCAGCCCCTCACCTTCTCCCCTAGTTGGGGTAGACACAAAGATCTTGCCTGTATCCGCCCCACCTGGGGAAGTGG
GTATCATACCTTCTGCT

16. 169G2 T729

TGTACTTCTAGGGAATTCCAGGAACAAAGCAACCCAGTGAAGCTGAATCACAAAATCTGAGTATAAATGATTGAAATATTTGA
TCAATCTCTAAATTACACAGATGCAGGATACACGCCTGCAGCCGCGTAACCTGGCAAAATCGGTTACGGTTGAGTAATAAAT
GGATGCCCTGCGTAAG

17. 459O20 T729

TGTACTTCATAGTGGAATTCCATCAGTGCAGACTCAGGGACCACTGAGAACTACGTCACTGCAGCCGCGTAACCTGGCAAA
ATCGGTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAG

18. 169G2 ptarbac4

TGCTCTCATATTACACGCCATGAATGCTGCAGAGGCCGGATGGGGTCTGCTTTTTCTAGAACCCACCGTAGTCACTGCTGG
GGTCCTCTTCATGGTGCACTCTAGTCCATGCCATGGGGATGATAAATAACCTTGGTGGATGGATTTAACAGTGCCAGGAAA
CACTCTGCCTAGGGAGGAAGAAAATGGTTTGGTTTTAAGACTCCGTTCCTCTCTTAACCCTGGTGGTTTGTAATCTATCACT
TCTGCCCAGTCCCGTTTCCCTGGCTCTGTGAGTCCTCTAGAGGTCACTGTTTTCTTCAGATCCTGACACTGACTACTAGACA
TC

19. 123H7 T729

CGTACTTCATAGGGAATTCCTTTTGGAGGATGACAGTGGATTCATTGCCCTCGGGGGTTCAAACTAGTTATGAGTGAGGGA
TTGGCCAGAAGATCGGGGCGCAGGCCAGCAGGAATGCTCTGTTAGGATAAGCAAGTTTGATAGGAAGAAGCTGCTCTTCT
CCAAATTACAGAGAGGTGATGTGTTCGTATTGCACGTAGACTGTTTGTGTAACAAGACCTCCCCGCCCTGCCACCCCGACA
CACAAACGAGCTGCCTAAAATATCCTTGCCTTGCAGATTGGAGGTTCTCCAAATATTTTGTGACCTGAGGATCCAGCTCAAG
TGAGGTGCCATAGGATGCGTTCCTGAGTTTGCATTGTATGGAGACTTTTCTGGAATCTCTCAGTAGCAAGTCAGCTTCACAA
CAAATTTTGCATTGAGTCCTAGACTGCTTGCAGTCTGAATTTGGACTACTTAGGTAGTGTGCTCAAAGTTGAACCTGGACAC
AGCACAACTCAAGTTTGCATCAGACTGGGAAGCAAACTT

20. 169G2 TCF19

TGGTCTAGGGGAGAGCTCCAGGCTGGGGCTGGGGCTGGCTTCCGCCCTATGCTGCCCTCCCTCGGGGCTCCACAGCGGC
CTCTTAGCACCTTGTCCTCTACCCCCAAGGCCACACTGATCCTGAACTCCATTGGCAGCCTCAGCAAGCTCCGGCCCCAGC
CCCTCACCTTCTCCCCTAGGGGGGGTAGACCAAAGATCTTGCCTGTTTCCGCCCCACCTGGGGAAGTGGGTATCATACCTT
CTGCTCCACCCCCACGCAATCGGAGGAAATCTGCTCACCAAGTGTTAGCAGAACTGGATGATGAGAGTGAGCAGATTTCCT
CGATTGTCGGGGGGTGGAACAAAGGTTATGTACTATCCMGGGTGTGGCGAACAGGCAGATCTTGTGTTCTCCCCTAGGGG
GGAGAAGGGGGAGGGGGCGGGCGATTTAAGCCACTGAATTTGGAAATTTGGCTGGGGTAAAAGAAGGTGTAACCGTGCA
CAAGGAGGGCCTAAGGGGACCACCCCCCCGGCTTCCCCAACCGGGGGAGGATGCGCAGTGAGCTCGATAGCAGCATCAG
CGACTAT

21. 99L19 TCF19

GGCCCTAGGGGAGGAGTTCCAGGCTGGGCTGGGGCTGGCTTCCGCCCTTGCTGCCCTCCCTCGGCTCCACACGGCCTCT
TAACACCTTGTCCTCTACCCCCAAGGCCACACTGATCCTGAACTCCATTGGCAGCCTCACAAGCTCCGGCCCCAGCCCCTC
ACCTTCTCCCCTAGGGGGGGTAACCAAAATCTTGCCTGTTTCCGCCCCACCTGGGGAAGTGGRTATCATACCTTCTGCTCC
CCCCAGCAATCGAGAAATCTGCTCACCAAGTGTTAGCAAAGCTGGATGATGAAAGTGACATTCCCCCGGGGGGGGGGGTA
AAAAAAACAAACTCCCGCGGAAAAAAAAAATTTTTTATTAAGGGGAGGGGGGGGGT

TS1. 161C12-1

AGCCTGCAGGTCATAATAGACCAATAGAAGGAGTGGGCATCGTGGTTCACACTTGTAATCCCTACATTTTGGGAAGCTGAG
GTGGGAGGATCATTTCAGTTCAGGAGTTTGAGACCAGCCTGATAGTGAGACCTCATCTCTACCAAGTAAAAATAATTTTAAA
AGGAAAAAGATCAACAGACAGGAAAGGAAGAAATATGTCTTTGTCACCAGCTTCATTGCACATGCAGAAAACACTAGGAAAT
TCTGAAAAAGTCTCTGGAATTAATTATTGAATTTACAAATAGTTCAGAAAATATATGTGATAATCCACTTAGATGGACATGAAA
AGATAGCAAACAATACTGGTCATCAAAGAAGTGAAAGTAAAAACCACAATCAGGAACCACCACACATCACCTGGAACAGGTA
AAGTCAAAAAGACATATGATAATCTAAATACTGGCAAGAATATGGAACAAATACCAATGTCTGTATTGCTGGTTTGAATGCAA
AAAATGTTGCAGCCAATTTGTAAAGTGGTTTAGCAATTTCTTATACAGCTGCCCATCCACTACCACAGGACCCAGGAATCCC
ACATATATGTATTTATCCAAAAGAAATAAAAGTGTAAGGCCACACTTGTAAGCAGTTACTTGTAGTGGCTTCATTAATAACAC
ACCCTAACTGGAGTGATCCACATGTCTATCAACTGGAGAACAGAAACCGATGAGTAAACTGGGATTCCAGCAATACTCAGC
AGCTACTCAGCAACAAAAATCAATGAATGGCATCATCTCAAACATTGTTACGCTGCGTGAGAGATCAAACGAAGGACTACAT
CACATATAATTGCACGTCCATGAAATTCTAGAAATTTCACTAATGCAGTGACAGAAAGCAGAGCAGTAGCTGAGTGAAGGGA
AGGATTGGGGGTGGAAAGCTAGAATTAAATAGAAAGGGGCATAGAGAAAGTTTTTAGGGAAAAGAAACTGTTCTCTACAGT
GTCACAATTCTGGAGTGACTGGGCAGTGGGAACTAAGGAGAAAGGAGGGAGGAGAGGGGCAGAGAGAAGGTCTGGGGAA
GCAGGAGGTAAGGAGAAGGAGCAGGGAAAAGGACTCTAAAGCAGTGAAGGAGCCTAGCAGGGGGTTCTTTGCATGCAGA
GTTTCTCTACTGACCAGTGTGGTAGTTACACGACTACAAATAATATCAATATGCACCAAAAAGCGCAGCTAAACCTGGTAAA
TTTTATTATGCATAAATGCCCTAAGCCAAAAAAAAAAAAAAAAAGGAGGCAGAGGGGAAGAAGGCAAATAAACACATTTTTAG
ATCATCAAAACCATAAAATTTATTTCCTAGGGGTCCTGTACTACATGTAATTTTAAAGGAAGTTATTAACAACAGAAATCATGC
ACACACACATATCACATACACACTCATTTTCTTAAGTTCCTGAAGGCATATGGCTCCTTTTCTAAAACAAAAACGATTACACT
GTATTGTTGAGTTTGTATAGTATATTAGTGTAATATATACAACAATAACAGCACAATAGTAGGTTAAATGAAACTCCACTGTTA
CAAGTGTCCTTTATTTTGCTGGATGTAGTTTAATATTACCTGAACTTCACTGTGAAAAGTCAAGGAATCGGGTTTCCATTCTT
ACAACAATAAAACATGAGTGTAAAGAAATATAGCTAAAAGCCACTAGGAGTATTAAAACCATAACCTAAAAAATGTGTACTTG
ACACATAAGAAACTAGCAAAGAAGGAACAGAAACATAAAGATATGAGACAAATTGAAAGCATATAGAAAATGGTAGACTTTTT
CTTTTTTTTTTTTTTTTTTGGCACCAAATTTCTTTATTTGAAGGAATGGTACAAAGAACTTAAGTGGATGTTTTGGTACAACTTAT
AGAAAGGGTAAAGGGAACCCCAATATGCATGCAACTGTCTTGGTGACCAGGGAAGTCACCCCTTGGTTATGGGGAAATTAG
CCTAAGTCTTAGCTTTTGCTATCACTGTCTTCCAGGGTGTGCTTGTCAAAGAGATATTCTTCCAAGCCAGATTCAGTTGCTCC

CAACTTGAGCAAGTTGGTCACGTGGCCACCCAATTCTTTGATGGATTTCACCTGCTTATTCAGGTAATGTGTCTCAATGAAG
TTTCACAAATGGGGGTCATTTTTGTCAGTGGCCAGTTTGTGCAGTTCCAGTAGTGACTGATTCACATCTTTTTCCAAATGTAA
TGCACACTCCATCGCATTCAGCCCGCTCTTTCAGTCATCATGGTCTGGTTTCTTGACATCCTGAAGGAAGATGAGGCCACCT
CGTTGGTTCTGCAGCTTCATCAGTTTCTCGGCATGTTCCCTCTCCTCATGAGATTGGTGAAGAAAGTACTTGGCAAAGTTCT
TCAAAGCCACGTCATCACGGTCAAAGTAGTAAGACATGGACAGGTAAACATAGGAGGCGTAGAGCTCCGGGTTGATCTGG
CGGTTGATGGCGGCCTCTGAATCCTGGTGGTAGTTCTGGCGCACCTGCGAAGGTGATGCGGTCGTCACCGCGGCTGCTAA
GGAGAGGCGGCGGCGGCGGCGGTGGCTGCGCGGCGCTGGAGAGGCGGCGGGGCCTTGGAGAGGTCGGAGGGCGCGG
TGTGGAGGTGAGGGAGGGCTGGCTGTGGGCGGCCGGCCGGGGTGGGGGACGAGCGCCGGGTTCCGTCCAAGCCCTGT
TGAAGCAGGAAACCGCGGCGACTCTCGGCGGAGAACGTCTTGACTTTTTCACTTACATTGGTAGAATCCAACCATTTTAAGT
GAAGAAATGACACGACCTGAGTCACATTAGCAGGATTGCTGAGCACTGTGGGGAGAACAGACATGGGCAGGAAGTGAGGG
ACGGTGTTAGTGCCGCAATTCAGGAGTGACAGGGTGGCAGGGACTAAGGGGAGGGGAGGATGTGAGGGATGAGAGGGG
CAAAGAGAAGGGATGGAGAAGCAGGAGGTGAGGAAAAGGGGCAGGGGAAAGAATTC

TS2. 161C12-2

GAATTCTTGTCCTGACTGTTCTGAGTAGTTCTGTAACTTGAGGCAATCACTTTTGTTATAAACCTCAGTGTTTGGTGGGGCG
ACATAAAATTTTTAAATTAGGTGACCTCTAAAAGTTTATTGTAGTTTTGAAACATGCTAAAAATTGGTTTAGCTCAAACTGAGA
GTGGGAACCCTGGGGGGATTGGACTCTTTCCTTCCTTGTTTTTGAGACTCTTTGCTCTTGGCTTGGCAGGTTGCTGTGTTTT
TTGGTGGTCTGTCTATCAAGAAGGATGAAGAGGTGCTGAAGAAGAACTGCCCGCATATCGTCGTGGGGACTCCAGGCCGT
ATCCTAGCCCTGGCTCGAAATAAGAGCCTCAACCTCAAACACATTAAACACTTTATTTTGGATGAATGCGATAAAATGCTTGA
ACAGCTCGGTGAGTGGCAGTGCTGGGGCTTGGCTAATGCTGGGGAGTTGTTCTTTGGAGCCAAATGATGTTTATTTGAAAC
AGGAGCACCTCAGTGTAAGGACGACTCTTATCTATCACCCATGACTGATGGCTCTGGGTTCCCTGGTTGGTTTTTATTCTTT
TAAGCACAGTAAGGGGGTCATCTATCATCTTTCTATGATTTCTTTTTAACCTTTGAGAATAGGGGACTTTGATAATTTAGGCA
TAAGTCGTCACCACCACTACCCTTTCCATTAAATATTTCTGTACTGGGAGTCATAGGGGATATTCTAAACTGAAGGAGAAGA
CTGTCTACCCTTATGGCATCTCCAGCACAACATACACAGTCAGAATTTATACCTGAATCAGTTTCTAGACTCGGGTCTGAAAT
TAATGTAGAGAGTGTTTGTAGGAGTTCATGTGGTGTGGTATTCTAGTGCTAAGTGGGGTTAATGAAGAAAGATTTTTCTGTC
AGAATTGAATCTTGGTAAATGAGAGTGGGGATGGACATAGGCTCCAGAGAAAAGTTATTACAAATGATCTTTGGCAATTCAA
TATGGCGAGCTATAAAGGTGGGCTCCAGGTAGAGATGTCTTATTTGTCCAACTGGATAGAAAAGTGATCCAGAGGGTCATA
GATGAACTTTGATAGCTGGAATGTAATATGTGCTTGATATTTGTAGTCTGCTGAAGGCTGGCCTGGGCTTGGGCAGGAAAG
AGTTGGGAGAATGGCCTGTAAAGCATGTTTTGAAGGCCTTATGCCTGAGAGTATCTGCACAGGGCTTTGTCCCCATTTTATA
GTTGTGAACTTCGGGGTTTTTATCTTTCTTGCTTAAACAGTTGGAATATAATTCCAATTCTTGATTCCAAGTTCATTGCAGTCT
TCACCTCACTGGGAGGTCTACCTTTAATCTTTACTATACCTGCTTATTTTCTCCTACAATTCTTTTCTTGTTCTCATAGACATG
CGTCGGGATGTCCAGGAAATTTTTCGCATGACCCCCCATGAGAAGCAGGTCATGATGTTCAGTGCTACCTTGAGCAAAGAG
ATCCGTCCAGTCTGCCGCAAGTTCATGCAAGATGTAAATACCCTTCTACCTTCTCTCCCTCCACTCCCCGCCCGCTGCCTC
CTCCCCTTCCTCGCCCTCTTCCTCCAGACTCCCTTGTCATTCAAGTGCCAAGAAGTCGGCTTATGCCCAGCTGGGAGTAAC
GACTCCTTGAAGAGACATACAGAAGCAGAGACAGCTAGTGTTAGGGTCTGCGCGGGTGCCAGGGAAACTCCGGAAGACTT
GGTCGGGTTAATGTGAGAGCGGGTAGTGTTCGACTTTTTCTCAAAAATCGCAACATTTTTGAACCTCTTCTCCTTTTGGGGG
AGGGCAGGATTTTTCTGCCCTACCACCCACCCATCGTGTCTTACGTGCCCCCTACAGCCACGCATCCTCAAGGTGGCATCG
AGCATACAGCTGGAGCCTTCTGCTCACCAAAACTCCTACTTCCCAGTGGCAGGAGAGCAAGAGAGGGACAGACAGATGGC
AGGGCATGTCCAAAAGAAGAGCATCAGCACAGATGAATCCTCCCCTTCCCCACCTCCAGGGGTGGGGGCCTTTGGCACCT
CAATCCCCGATACCCTACTCCTTCCCACCCACATCTCCTTGCACCCATCCGGAACCTCGGTTGATGTGAGCCGGCAACAGA
GAAGCACCGTGGCGCGGCGAGGGAATGCGGACGGCACCCAGCGGTGGATGGCGGCAGCGGAGGCCGCGGGGAAACCT
GACCAGGAAGCTGAGGACCAAACCAGCCTCTTTTTCCGTTCCCGGTTTTTTTCCTGAACCCAACGCGTGCCGTGCCCCGTT
TCCCCCCATATGTGTTGGGGAGGGGTGTCCTGAATGGGGTGGTAGATTTTTTTTTTATTTTTTTTTTTTTTTTCTTAAAAAAAATT
TTTTTTATTACTCAGAGGAGAGGGACATAGGAAAGGTAAAGTAGATGTAATCGGGTGAGTGTTAGGGTTTGGGGCTAGGTG
GGTCCGATTGCATAAGCAGTGGAGTGTGTTCTTCCCCTCCCTGCAGTGTTCCTTCCTGTGGGATAACTCTTTATCTGTATTT
GGGGCTAGAATGAAATTTCAAGGAGGCCATGGAACTTGTCTTTAAAAAGCCTGCCTTGGCTGAGCATGGTGGCTCACCTCT
AATCCCAGTGCTTTGGGAGGCCAAGGTGGGAGGATTGCTTGAGCCCATTGAGACTAGCCTAGGCAATGTAGTGAGACCTT
GTCTCTACCAAAAAAGCTGGGTGTGGTGGCACATGCCTGTGGTCCCAGCTATCTGGGAGGCTAAGGCAGGAGGGTTTGCC
TGAGCCCAGGATGTGGAGGCTGCAGTAAGCTGTAATTGTGCCACTGTACTCCAGCCTGGGTGATAGAGGCCCTGTATCAA
AAAAAGAAAAGAAAACCTGCCTTGTTTTTTTTTTCCTATGGTACATGCATCATTTCCTATTTGGGCCGCGGGGTCTTCATGTT
AACTATCTTTCCTAGGATAGTATAAAAAGTGTTAGGAAAATTGGGTTTTGGAGTTGTGCTAATTTC

22. 261L9 HLA-B 29

GGACATCGAGGGCTGCTTTTTTGCTTTTCTACCCAGGGCTGCTAGTGTTTTTYTCCCTCCCTCATCCCCTCCTCCCACACCA
ACTCCCAACCCACACACACCCCTGCACACTGCAGCACACAATCACGGTTCTCTCTTCAGGAGAGAATAGTCCTTGATGATG
GGTCCAATTTCACAAACAAATGTAAGTCTAAATTACACTCTGCT

23. 271C5 T7.29

GTGGGAACCTAATAACTTTTTATAGCATACATTACACGAAGTTGTATTAATGTGGTTCCGGATCAATTCTCATGTTTGACAGC
TTATCATCGATAAGCTTTAATGCGGAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAAGCTAT
CGC

24. 271C5 32A

TCGGCCCCACCACCACCACCCCAGTCAGCAAATGCAGCACTTAATAAGGTTTCTCTCTTCAGAACATTTTTGACTATGGACC
CATTTTACAACAACTGACGACTGAAATTAGACTGCTTTATAGATTCATGAGTTGGGATTGGAGCCAGCACCAAGATCACTGG
AACCAGGGCAGGGAGAGAGGGCAGGAGAGCAGAGCAGAAGAGGAGCCCTAGAAGGAGGCCAGGAGCTGAATGGGTCTG
AAAATGTCGTTCCAGACAGAGACCATCACCTTGCTGATCGTAGTCGTAGACAGATGCACGTGAGCTACTACGTACGTACGA
TGCCATGTAAGTATACGATCTGAGACATGATCTATATCTACGCTAACTAGTACGTCAGTAAGTCGATCATACTGTACTAGCAT
ACTGTAGTACAGAGATACGTACATCTTCATAACATAGACGACTCAGTAGTGTGCGTAGCAGCCACTTTAGCATCACTGCTAT
GATGACTATACAGACTACGATCAGTACGACGATACGAACATGACTGCTGATCTCGTATGCATAATGATGCGCGCTGACTAC
GTCTAC

25. 271C5 T7.29

GTGGGAACCTAATAACTTTTTATAGCATACATTACACGAAGTTGTATTAATGTGGTTCCGGATCAATTCTCATGTTTGACAGC
TTATCATCGATAAGCTTTAATGCGGYAGTTTATCACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAAGCTA
TCGC

26. 127E13 HLA-B29-

TTTTGAAGTGAAGACTCACCCTCTGCTTTTCTGAGCAAGACGCTGCTTGTGTCCTCCCTCTGGCATCCCTCCTCTCACACCA
ACCCCCCTGCACGCGCGTGCACACACACACACACACATTGCAGCACACAATCAGGTTTCTCTCTTCAGAAAAGAACAGTCC
TTGATGATGGGTCCAATTTCACAAATAAATGTAAGTCTAAATTGGACTCTGCTTTATAGATTCATTATTTGGGGGTGGGGCCG
ACACCAAGGTCACTAGGACCCGGGCACATAAACA

27. 127E13 HLB-29B

CCCCGGTTACGAAAAGGTGGGGCTGGCTGTCTCCATCTCTGTCTCAACTTCATGGTGCACTGAGCTACAACCTCTTGCTTC
CTTACTAAAAATAACAATCTGAATATAAATGTGCTTCCTCAAATATTTGCCATGAGAGGTTGATGAGTTAATTAAATCCCCACT
TCCCCACCCC

28. 282L12 ptarbac3

GAGTACTTATAGGGAATTCCTGCTCACAGCGGCGGGTGTGAGCGCCCCAAAGCTGGAAATCAGAACCCTAAAGGCAGCTC
GGCCCTCCCCAGCCCACAGCGCCGTTATTCCGTTTCTATATCGGTAAACACGCTTCATTTTCCATAGCCCAGGGCGGGGTG
ACGGGTGATCCCAGTCTTCGCAGTGAACTGCGGGCAGTAGAATTCAAAAAGCTTGCGGGCAAAGCCGGGCGCGGTGGCC
CACGCCTGTAATGCCGTACGTTTCTTTGTATA

29. 245C6 T729

AATTCGGGGGGACATTCATAGGGAATTCAGCCCCTCTGTCCACCCATTCTTCTCCCCACAGGATCGTCTGAGGAGTCGATT
GGAAGCCCTGAGCATGGAGAGAGATGAGATTGAGGATGTAAAGTGTGGAGAAGATCAGAAGCTTCAAGTGCTGCTGGTAC
AGGCCACGTCACTGTCTACCTTTTCCTTTGAAGGTTTTTCTTAACAGACTCTGGGGAAAACTGTTGGCTGGTATCTGTCGCC
TGGCTGAAAAGAACTGGCACACTGTTCTCATTCACCT

30. 463N5 HLA-B

AACACCAACACGACTCCGGGGCCTGCTGGTGCTTTTCTACGAGGGCTGCTAGTGTGTCCTCCCTCCCTCATCCCTCCTCCC
ACACCAACTCCCAACCAACACACACCGCTGCACACTGCAGCACACAATCAGTCTTCTTTCTTCAGGAAAGAATAGTCCTTGA
TGATGGGTCCAATTTCACAAACAAATGTCAGTCTAAATTACACTCTGCTTTATAGATTCATGAGTTGGGATTGGRGGCAGCA

CCAAGATCACTAGAACCAGGGCAGGGAGAGAGGGCAGGACAGCAGAGCAGAAGAGGAGCCCTAGTAGCATGGCAGGAG
GTGAATGGC

30. 463N5 ptarbac3

TGACATCTAGAGGGCGCATTCGTTTCTCTACGCAGTTACCCAATAGTGGTGATCTACTTTTGGTTACTTAAGGATCTCCAAA
CTGTTTTCCATAGTTGTTGTACTAGTTTGCATTCCCACCAGCGTGTAAAAGTGTTCCCTTTCACTGCATCCACGCCAACATCT
ATTGTGCTTTGATATTTTGACTCTGGCCATTCTTGCAGGAGTGAGTTGCTATTGCACTGTGGTTTTGATTTGCATTTCCCCGC
TCATTAGTGAGGTTGGCCATTTTTTCATATGTTTCTTGACCATTTGCATATTTCTTTCGAGAATTTTCTATTTATGCCCTCAGC
ACATTTTTTGATGGGATTGTTTGTTCCTGTCCC

32. 436B4 T7.29

CGCCGGCACCGGGGGCTGTTTTTTTGGCTTTTCTCCCCCAGGGCTGCTAGTGTTTTTCTCTCCCTCCCTCATCCCCTCCTC
CCACACCAACTCCCAACCCACACACACCCCTGCACACTGAAGCACACAATCACGGTTCTCTCTTCAGGAGAGAATAGTCCT
TGATGATGGGTCCAATTTCACAAACAAATGTAAGTCTAAATTAGACTCTGCTTTATAGATTCATGAGTTGGGATTGGACCCAA
CATTCTGAAAATTTTTCTCAGACTGCTG

33. 325P21 T7.29

GACTTCATAGTGGAATTCTGCACAATGGTCTCTCCTTTTTGAGGACTTCTCATTCTCTCCCATTCTGACATAGGCTGCTTACC
TGGCTCATAGTCCAACTCAAAGCATAGTTTTTCTGTAAAGAAGAAATAAAACAGGATGATATTTTATTAGTCTTACAAAACCAT
CAGACACTTAACAATGAGAAAACTGAGGCCCAGAGGAGGGAAGGGACAAGAAGAATGAAGCTGGAATCCTCTAGACCAGT
TTCCAAGCTTGCATCAAATCACCTGGAGCACTTGCGAAAACACACCTGTTTCAATTCAGATGCTCTGGAGTGGGGA

34. 325P21 ptarbac3

AAGACTTCTAGGGGAATTCTGGGTTTCTAACTCCTGTACAGAGACTGATTTTTGAGGGCCCGCCCCGCTCTCTGGGACAAT
TAAGGGATGAAGTCTTTGAGGGAGTTGTGGGGAATTACAATCCAGGAATACTGATCACCTGTCCCTTTTGACCCCTGCACC
ATATCATGG

35. 279K21 ptarbac3

GGTCTTCTAGGGAATTTCTGGCCTCAGCGATCAGCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTCTGAGCCACCA
AGGCTCTTTCTTCCTGTCTCATGTAAGCACTTAAAGCTTCATGTTTCCCTCAGAGCAATGCCTGAGCTGCATTCCACAAACTT
GGATATGTTGTTTTGTGACCATGCAATTCAAAATATTTTCTTACTTTCCTTGTGATTTTGTCTTTGGCCCTTAATTTACATAGAT
AAATGTGGTTTCATTCCATACAGTTGATGTTTCCAAAATATCTTCTTTGTGTTAATTTCTAATTTAGTATTATCATAGAGAATAC
ACTCTGGTGACTTTTACACTGTTGAATTAGTTGAGGCCCATTTTTCAGCTCAGCATATTGTCCTGGAATTGCACAGTGAATCT
ACCGTATGTAATTGGTCACCTGTTTACCTGCTTTATATTTTCTGAGGGTTTGTGAGGTTTTCTCC

36. 224H20 T7.29

CTCGAGAGAGGGAGTGCTTATTGTAGCGGCTTCCGAGATTTTAGCTGAGCTAAAACCCGACATTTGCTTCCCTATCTCATTT
CTTGGCCTCATTGGTAATCTTTTAATGAGGAAGACATTTCCCCACGCGTGGAGCCGTGGCCCTTGACCTCTCACTCA

37. 348C21 T7.29

CGTCTTCATAGGGAATTATCCCGCATGCTTGCTTGTCTATTTCTTCCACATGAAAAAGGCTTTCAGCCCACCTGGCTGAGCA
GAAAGAATCTGAATGTTAGAGTTAGGTAGCCTGGGTTTGAATTCCATCTCGGGTACCACACTATGGCAAAATTCTTAGATTC
TCCAAGCCTCAGTTGCCTTGTCTGCCAAATAGAGAACACATCCTCCCTCCTAAATTGTAAAAAGGATTAAAGTCATGTAAAGT
GCATACTACAAATCCAGTCACAAAGTAGCTAGCTACTCACTAAGTGTTCAGCTCCTCCTTCCTCATTCAGATGGGAAGTGGC
TTTAGATCAACAAAGTGGCAACCCAGTGGGCTTGAGCAGCACTGTGAACTTAGAATCCAAGAAAACTTGCAAAGAGCAGCT
GGGAGGTAGTAGATCCTTGTGTTGCTTTTGGAAATTGTTCCATTTCTTTATTTCCCTATTGAATCTAGGCTCTAG

38. 171K8 T7.29

TGACTTCATAGGGAATTCATCCCAGATGCTTGCTTGTCTATTTCTTCACATGAAAAAGGCTTTCAGCCCACCTGGCTGAGCA
GAAAGAATCTGAATGTTAGAGTTAGGTAGCCTGGGTTTGAATTCCATCTCGGGTACCACACTATGGCAAAATTCTTAGATTC
TCCAAGCCTCAGTTGCCTTGTCTGCCAAATAGAGAACACATCCTCCCTCCTAAATTGTAAAAAGGATTAAAGTCATGTAAAGT
GCATACTACAAATCCAGTCACAAAGTAGCTAGCTACTCACTAAGTGTTCAGCTCCTCCTTCCTCATTCAGATGGGAAGTGGC

TTTAGATCAACAAAGTGGCAACCCAGTGGGCTTGAGCAGCACTGTGAACTTAGAATCCAAGAAAACTTGCAAAGAGCAGCT
GGGAGGTAGTAGATCCTTGTGTTGCTTTTGGAAATTGTTCACATTCTTTATTCACCTATTGAATCTAGACTCTAGCTAGAGAA
AGAGCGGCAACCCTTGGCTTCAAATCCAGTGCTCTTCTCACACTCTTTAAGGTTGCTTTCCCGCTGCAGCCGCGTAACCTG
GCAAAATCGGTTACGGGTTGAGTAATAAATGGATGCCCTGCGTAAGAAAACGCGCTCCCAAAT

39. 171K8 ptarbac2

TTCTAACCGTAACCGATTTTGCCGGTTACGCGGCTGCAGCGGGAAAGCAACCTTAAAGAGTGTGAGAAGAGCACTGGATTT
GAAGCCAAGGGTTGCCGCTCTTTCTCTAGCTAGAGTCTAGATTCAATAGGTGAATAAAGAATGTGAACAATTTCCAAAAGCA
ACACAAGGATCTACTACCTCCCAGCTGCTCTTTGCAAGTTTTCTTGGATTCTAAGTTCACAGTGCTGCTCACAGCCCACTGG
GTTGCCACTTTGTTGATCTAAAGCCACTTCCCATCTGAATGAGGAAGGAGGAGCTGAACACTTAGTGAGTAGCTAGCTACTT
TGTGACTGGATTTGTAGTATGCACTTTACATGACTTTAATCCTTTTTACAATTTAGGAGGGAGGATGTGTTCTCTATTTGGCA
GACAAGGCAACTGAGGCTTGRAGAATCTAAGAATTTTGCCATAGTGTGGTACCCGAAATGGAATTCAAACCCAGGCTACCT
AACTCTAACATTCAGATTCTTTCTGCTCAGCCGGGTGGGCTGAAAGCCTTTTTGATGTGAAAAAAATAAAACCCGC

40. 171K8 Caja G

CGCCTCGTGGCTGCCTCCCACTTGCGCTGGGTGATCTCGAGCCGCCAAGTCCGCGGCCGTCCCGGAGCGCAGGTCCTCG
TTCAAGGCGATGTAGTCCTTGCCATCATAGGCGTGCTGGTTATACCCGCGGAGGAGGCGTCAGGCAGGCCCCAGGTCGCA
TCCATACATCCTCTGGTAGGTGTGAGACCCTGGCCCCGCCCCCGAGGTCGGCCCCGCCCCAGCCCAACTTGCGGGGATT
TTGGACCCAACTGAAAATGAAACCGGGTAAAGGCGCCTGGGGCTCTCCTGGATTGCGGGTTTGGGCGGGTGCCGCAGCC
TCACGGTGAATCTTGGACCTGGAGACTCAGGGCGACCTGGGCCCT

41. 239N18 T7.29

CTCTTCATAGGGAATTCATCCCAGCATGCTTGCTTGTCTATTTCTTCACATGAAAAAGGCTTTCAGCCCACCTGGCTGAGCA
GAAAGAATCTGAATGTTAGAGTTAGGTAGCCTGGGTTTGAATTCCATCTCGGGTACCACACTATGGCAAAATTCTTAGATTC
TCCAAGCCTCAGTTGCCTTGCTCTGCCAAATAGAGAACACATCCTCCCTCCTAAATTGTAAAAAGGATTAAAGTCATGTAAA
GTGCATACTACAAATCCAGCCACACAAGTAGCTAGCTACTTAGTAAGTGTTCAGCTTCCTCCTTCCCTCATCC

42. 120D3 32b

CCCCGGTTACGAAAAGGTGGGGCTGGCTGTCTCCATCTCTGTCTCAACTTCATGGTGCACTGAGCTACAACCTCTTGCTTC
CTTACTAAAAATAACAATCTGAATATAAATGTGCTTCCTCAAATATTTGCCATGAGAGGTTGATGAGTTAATTAAATCCCCACT
TCCCCACCCC

43. 120D3 HLAB 32B-

TTTCAGCAGAAGTGGGGCTGGGTGCCTCCATCTCAGTCTCAACTTCATGGTGCACTGAGCTGCAACCTCTTACTTCCCTAC
TGAAAATGAGAATCTGAAGATAAATGTGTTTTCTCAAATATTTGCCATGAGAAGTTGATGGGTTAATTAGTCAATTCCTAAAAT
TTGAGAGAGGAAATAAAGACCTGAGAACCTTCCAGAATCTGCATGTTCACTGGCTGAGTCTGTTGCAGGTGGGGATGGA

44. 120D3 HLAB -29-

TTGAATCGAGGCCCCTCTGCTTTTCTCAGCGGTGCTGCTTGTGTCCTCCCTCTGGCATCCCTCCTCTCACACCAACCCCCC
TGCACGCGCGTGCACACACACACACACACATTGCAGCACACAATCAGGTTTCTCTCTTCAGAAAAGAACAGTCCTTGATGAT
GGGTCCAATTTCACAAATAAATGTAAGTCTAAATTAACTCTGCTTTATAGATTTATAGTGGGGTGGGCAACA

45. 120D3-Caja-G

TCAACACTGCTACGTCCCGGCCAGGCTGCGGGGAGCCCCGGTACTTCGAAGTCGGCTACGTGGACAACACGCAGTTCGT
GCGGTTTGACAGCGACGCCGCGAGTCCGAGGATGGAGCCGCTGGCGCCGTGGGTGGAGCAGGAGGGCCGGAGTACTGG
GACCCGGAGGCGCAGAAAGCCAAGGCCGCCGCACAGAATATCCGAGGGAACCGGCGGACCCTGCGCGGCCACTACAAC
TAGAGCGGGCCGGTGAGTGACCTCGGCCCTGAGCGCAGGTCACGACGCCTCCCCATCCCCCACGGACAGCCCAGGTCCC
CGTGAGTCTCCGGTTCCGAGATCATCCACCCCGAGGCTGCGAGACCCGCCCAGACCCTCGACCCGGGAGAGGCCCAGGC
GCCTTTACCCGGTTTCATTTTCAGTTTAGGCCAAAATCCCCACGGATTGGTCCGGGCGGGGGCGGGGCTCGGTTGGTAGG
GCTTACCTGGGGTCTGACCAGCGCCTCCTCTGCGGATATTACCAGTTAGCCTAAGACCGCAAGGACTACCTCTCTTTGAAC
GAGGACTTGCGCTCCTGGACAGCCTCCTGCACAGCAGCTCAGATCACCCAGCCCAAGTGGGAGGCGGCCAGTGTGGCGG
GAGCAGTGAAGAGCCTACCCTGGACGACACGTG

46. 120D3 Caja-E-

GCTTTCCTGCTCCGCCACACTGGCCGCCTCCCACTTGGGCTGGGTGATCTGAGCTGCTGTGCAGGAGGCTGTCCAGGAG
CGCAAGTCCTCGTTCAAAGAGAGGTAGTCCTTGCGGTCTTAGGCTAACTGGTAATATCCGCAGAGGAGGCGCTGGTCAGA
CCCCAGGTAAGCCCTACCAACCGAGCCCCGCCCCCGCCCGGACCAATCCGTGGGGATTTTGGCCTAAACTGAAAATGAAA
CCGGGTAAAGGCGCCTGGGCCTCTCCCGGGTCGAGGGTCTGGGCGGGTCTCGCAGCCTCGGGGTGGATGATCTCGGAA
CCGGAGACTCACGGGGACCTGGGCTGTCCGTGGGGGATGGGGAGGC

47. 120D3 MIC

AAGTCCTTGGGTACATCTGGTCCTAATTAAGCTTGGAAACCTTGAGTTTGGTCCTCCCCATCCTGCTGCTAAGTCAATGTGA
TCTTTGCCGAGTAGAAGCCCAGGGCCCAGCACCTCAGCCTCATGGTCAGAGATGGGGTGG

48. 120D3 Klex2-

AACACACTGCTACGTCCCGGCCGGCTGCGGGGAGCCCGGTACTTCGAAGTCGGCTACGTGGACTTTCACGCAGTCGGCA
ACCTT

49. 193P12 T7.29

GTCTTCATATTGGAATTCTGGCAATCTCTCTCTTTTTTTTGCAATAAAAGCCTTATAAGAAACTTATTGAACGGCATAGGAATC
AGTGCCGTTCAACTTGATTCAGAACGGTCAGTACATTATAATCTTGCCAAGGTACCTACTTATTAATTCCATGTCTTAATTTTA
TGAAGTATTACAGGCTTTTAATACAGATAATTGGATATTTGCCACATATTTAAAAAAAATGTTATAAAATCTCTGTGTACTTGA
TATCTAAGGTTTATTTTGGACTAAAACCCTATTACAGTATTCCTCAAAAGTGAAAATAAAATTTCCCTAACAAATAAAATAAAA
AAGATACATAGATTGAATTGAAATAGGAAATTTCTAAATCCATGAACTGCAATTCAAAACTGGTTGTTTTCAGCCTTGCCCTA
TTTCTCCTGATCTAGCTATCATCTTGACCTATAATGGTCCTGCTATGCCATAATATGAGGTTGCATCCTGAGAAGATATATTT
GATTCTTATAAGACTCAGAGATACCGTAAGGGAAAATAGTGAATTCAAATAACCATGTGAAAGAACTATAAGCCCTCAAAATA
TTGTGCTTAACCATAACGCCTCAATTGTACCAAAATTGTTGAACTATGTTTATTACTGACCTCCTTTAATTCAAAACCAACAAG
ACCAACAAAGTTGATTCTTTGGCA

50.18G14 ptarbac2

TTTTCGAGGCCTAATAACAAACTAGAGCGACAACATACGAAGCCCGTATTATGTGTGGCGGATAAAATCTCATGTTTGACAG
CTTATCATCGATAAGCTTTAATGCGGTAGTTTATTACAGTTAAATTGCTAACGCAGTCCGGCACCGTGTATGAAATCTAACAA
TGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTAGGCATAGGCTTGGTCATGCCGGTACTGCCGGGCCTCT
TGCGGGATATCTGCCATTCCGACAGCATGCCAGT

51. 19I16 T729

GTCTTCATAGTGGAATTCCATTTCTACTCTGCCTCACCCAGTAAATTAATACAAATGCTTCAGAAGTCAGTTTAAGGGATTGT
TCCATGACCCTACAGATCATAGTTTCTTCAGGTCAGGGTTTCTCAACAATGGCCCTATGGCATTTTGGACCAGGGAGTTGTT
TGTTGTGGATAGCTGTCGTGTGCATTGTAGGATGTTTAATGGATGCCCTGCGTAAG

52. 174N1 T7-29

AGACTTTCATATGGGATTCATCGCATATGCCTCAGTGTTTATGTGATCCAGAGCCTCACCACTGCCTATACAGACATGGGCA
AACAGAGGAGAGTCTGGAAAATCATGTTGAACCAGCAAAAGTCATGCTTAGGGAATAGCCCTCTATGCAGTGGCACCCTCT
TTTGCAGCCTCTTCTGGATTGGCTGAAGGAGGCGCATGA

53. 169G2 T729

TGTACTTCATAGTGGAATTCCATCAGTGCAGACTCAGGGACCACTGAGAACTACGTCACTGCAGCCGCGTAACCTGGCAAA
ATCGGTTACGGTTGAGTAATAAATGGATGCCCTGCGTAAG

Zusammenfassung

Der Weissbüschelaffe *Callithrix jacchus* dient als Tiermodell zur Erforschung von Infektionen, Arzneimittelwirksamkeiten, der Parkinson-Krankheit sowie Organtransplantationen. Ferner wurde an ihm die Durchführbarkeit von Zell- und Gentherapie bei immunologischen und hämatologischen Störungen getestet. Dieser Neuweltaffe ist von besonderer Bedeutung wegen seiner Ähnlichkeit zum Menschen.
Ziel dieser Arbeit war es die vier Klasse-I-Intervalle (Contig 1-4), einschließlich der flankierenden *Framework*-Gene, der MHC-Klasse-I-Region des Weissbüschelaffen *Callithrix jacchus* auf genomischer Ebene zu analysieren.
Hierzu wurde eine BAC-Bank (CHORI 259) von *Callithrix jacchus* mit Sonden von zwei MHC-Klasse-I-Genen und mehreren *Framework*-Genen gescreent. Southern-Blot- und Sequenzanalyse der MHC-Klasse-I-positiven BAC-Klone ermöglichte die Erstellung einer Karte über die Anordnung der MHC-Klasse-I-Gene im Weissbüschelaffen.
Die physikalische Kartierung der MHC-Klasse-I-Region von *Callithrix jacchus* liefert neue Grundlagen zur weiteren Analyse der MHC Evolution und zur Erforschung von verschiedenen Krankheiten, die mit MHC assoziert sind. Außerdem dient die Karte als Grundlage für die komplette Sequenzierung der MHC Region.

ACKNOWLEDGEMENTS

The work described in this thesis has been carried out experimentally at the Deutsches Primatenzentrum, Abteilung Primatengenetik, under the supervision of PD Dr. Lutz Walter. The evaluation of the *Callithrix* sequences obtained was done at the Max-Planck-Institut für experimentelle Medizin with guidance by Prof. Dr. Johannnes H. Matthaei, to whome my special and sincere thanks and gratitude are due for his motivation, suggestions and discussions. Prof. Dr. Jürgen Wienands as my advisor has kindly contributed his counsel.

I like to thank Prof. Dr. Jürgen Wienands (Referent) and Prof. Dr. Ernst A. Wimmer (Korreferent) of the Universität Göttingen for their support. Sincere gratitude is felt towards my helpful advisors of the Institut für Genetik der Universität zu Köln, Prof. Dr. Thomas Wiehe and Prof. Dr. Jonathan Howard. Of the Göttingen Faculty of Biology, I wish to thank Prof. Dr. Hans Joachim Fritz, Prof Dr. Botho Bowien, Prof. Dr. Dieter Heineke and at Faculty of Medicine Prof. Dr. Nicolai Miosge. Internationally, Claire Amadou (France), Prof. Dr. Carlos Burgos (Colombia), Prof. Dr. Luis Fernando Cadavid (Colombia), Prof. Dr. Helena del Coral (Colombia) and Dr. Takashi Shiina (Japan), who kindly provided the sequences called TS1 and TS2. With gratitude I appreciate the invitation and support by Prof. Dr. Walter Stühmer, Director of the MPI für experimentelle Medizin.

I also thank all colleagues of our research group at the DPZ for the good and motivating working atmosphere, especially Mrs. Christiane Schwarz for her friendly help with lab work. I would like to say thanks a lot to, Anne (Anecita) Maicito (Conny), Jenicita (Jenny), Ulrikita (Ulrike) and Seray for their friendliness.

The Ph.D. scholarship granted by Deutscher Akademischer Austauschdienst (DAAD) is gratefully acknowledged.

Finally, I like to express my special thanks, to my family and to all the understanding and helpful people, of whom I mention Martin Mesa, Dr. Glenda Rodriguez, Dr. Maria Laura Ponce, Dr. Leslie Pineda, Maria Isabel Gomez, Omar Diaz, Dr. Jorge Alonso Leiva, Anne Averdam, Ute Rust, Svea Viola Dettmer, Dr. Andrea Gomez, Cyrilla Maelicke, Ralf Bosselmann and Sridhar Boppana, thanks for their help and friendliness.

Lebenslauf

Natalia Regina Mesa Herrera

Geboren am	01.10.1970 in Medellín, Colombia
Eltern:	Luis Hernan Mesa Santamaria
	Maria Leonisa Herrera de Mesa
Familienstand:	ledig
Staatsangehörigkeit:	Kolumbianisch

Schulische Ausbildung:

1975–1981	Palermo de San Jose, Medellín, Colombia
1982–1989	Palermo de San Jose, Medellín, Colombia

Universitäre Ausbildung:

1990–1995	Bacteriologie und Klinischen Labor im Studiengang Mikrobilogie an Colegio Mayor de Antioquia, Medellín, Colombia
1996–1999	Anfertigung der Magister Scientiae-Arbeit an der Universidad de Antioquia, Medellín, Colombia
Juni 1999	Magister Scientiae-Examen
Januar 04–August 07	Anfertigung und Anleitung der Dissertation am Deutschen Primatenzentrum, Göttingen von Prof. PD. Lutz Walter (DPZ, Laborarbeit) und am Max-Plank-Institut für Experimentelle Medizin, Göttingen von Prof. Johannes H. Matthaei (MPI, Auswertung und Textfassung).