HAL
open science

The porcine Major Histocompatibility Complex and related paralogous regions: a review

Patrick Chardon, Christine Renard, Claire Gaillard, Marcel Vaiman

To cite this version:

Patrick Chardon, Christine Renard, Claire Gaillard, Marcel Vaiman. The porcine Major Histocompatibility Complex and related paralogous regions: a review. Genetics Selection Evolution, 2000, 32 (2), pp.109-128. 10.1051/gse:2000101 . hal-00894302

HAL Id: hal-00894302

https://hal.science/hal-00894302v1

Submitted on 11 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Genet. Sel. Evol. 32 (2000) 109–128

© INRA, EDP Sciences

Review

The porcine Major Histocompatibility Complex and related paralogous regions: a review

Patrick CHARDON*, Christine RENARD, Claire ROGEL GAILLARD, Marcel VAIMAN

Laboratoire de radiobiologie et d'étude du génome,
Département de génétique animale,
Institut national de la recherche agronomique, Commissariat à l'énergie atomique,
78352, Jouy-en-Josas Cedex, France

(Received 18 November 1999; accepted 17 January 2000)

Abstract – The physical alignment of the entire region of the pig major histocompatibility complex (MHC) has been almost completed. In swine, the MHC is called the SLA (swine leukocyte antigen) and most of its class I region has been sequenced. Over one hundred genes have been characterised, including the classical class I and class I-related genes, as well as the class II gene families. These results in swine provide new evidence for the striking conservation during the evolution of a general MHC framework, and are consistent with the location of the class I genes on segments referred to as permissive places within the MHC class I region. Recent results confirm the involvement of the SLA region in numerous quantitative traits.

pig / major histocompatibility complex / physical map / performance / paralogous regions

Résumé – Le complexe majeur d'histocompatibilité du porc et les régions paralogues. La construction de la carte physique du Complexe Majeur d'Histocompatibilité du porc (SLA) est pratiquement achevée et la séquence nucléotidique d'une grande partie de la région SLA des gènes de classe I est d'ores et déjà disponible. Plus de 100 gènes de la région dont les différentes familles de gènes d'histocompatibilité ont été caractérisés. Les résultats obtenus chez le porc montrent l'existence d'une trame ancestrale de gènes, conservée durant l'évolution. Ils renforcent l'hypothèse de la multiplication et différenciation des gènes de classe I après spéciation et de leur confinement dans des segments particuliers, désignés comme permissifs, de la région de classe I. Des résultats récents confirment la participation de la région SLA dans de multiples caractères quantitatifs.

porc / histocompatibilité / carte physique / régions paralogues / caractères quantitatifs

* Correspondence and reprints
E-mail: chardon@biotec.jouy.inra.fr

1. INTRODUCTION

The Major Histocompatibility Complex (MHC), called the SLA (swine leukocyte antigen) in swine, is located on either side of the centromere of chromosome 7. It consists of three major gene clusters or regions which are schematically represented in Figure 1. As shown in the figure, the centromeric class III region and its contiguous telomeric class I region span about 1.5 megabases on the short arm [42]. The class II region spans about 0.5 megabases on the long arm, and the class II DRA gene and the RING1, KE4 and KE6 gene cluster are respectively located on its centromeric and telomeric ends [9].

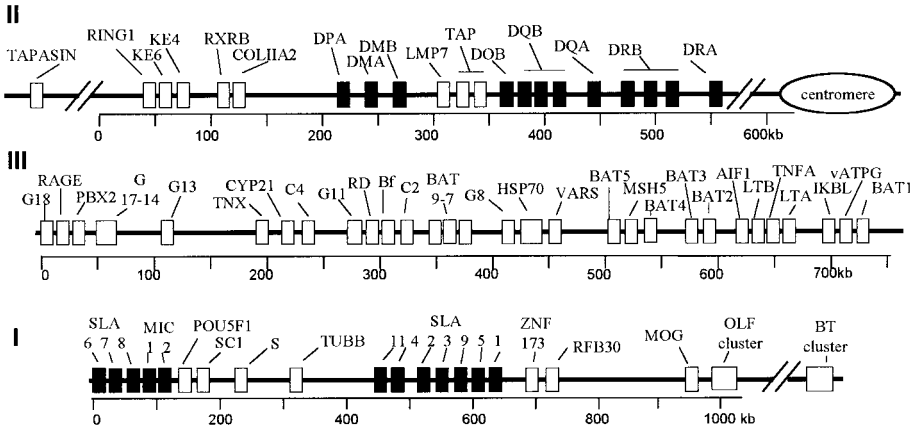

Figure 1. Physical map of the SLA complex. Black boxes: loci containing MHC-related sequences. White boxes: loci without MHC-related sequences. The name and function of the genes are specified in Tables III, IV and V. From the long arm to the short arm of the chromosome, the order of the regions is class II (II), class III (III) and class I (I).

Two-hundred and twenty-four genes have been located in the human MHC region [37]. Among them are 30 to 40 genes related to class I and class II genes, including many pseudogenes. The remaining genes comprise various oligo-copy gene families or single genes, including genes with proven functions, or genes to which functions inferred from related genes have been assigned, and also genes with no known function [39].

To date, about one-hundred genes and pseudogenes have been characterised in the SLA region. Numerous single-copy or anchor genes present throughout this region constitute a framework which has been well conserved during the evolution of the mammalian MHC region [1, 8, 42].

The main feature of the MHC is the incomparable polymorphism of the genes encoding its classical class I and class II membrane-anchored glycoproteins. These two classes of glycoproteins differ in their structure, the cells and tissues in which they are expressed, the origin of the peptides they present to the T cells and the subsets of T cells they activate. Thus, class I molecules primarily present peptides derived from nuclear and cytosolic proteins to the cytotoxic CD8 + T cells, whereas class II molecules mostly present peptides derived from exogenic molecules to helper CD4 + T cells.

2. THE SLA CLASS I REGION

Historically, the SLA region has been characterised using SLA class I serology. In fact, despite the present importance of molecular techniques in analysing the MHC region, SLA class I serology, using conventional allo-anti SLA reagents, still remains a powerful, quick and inexpensive tool for analysing large panels of individuals. On the basis of the segregation of SLA epitopes in more than 550 informative families, the existence of three class I series, A, B and C, has been postulated according to haplotype, with each haplotype corresponding to an allelic combination. At least 74 haplotypes have been characterised (Tab. I). Some of the haplotypes appear to be breed-specific, although the term of breed preference would be more consistent with the observations made. There are rare haplotypes which seemed to express more than three SLA class I series, while other haplotypes appear to express a single locus (Tab. I). The SLA class I region sequences for the haplotype H01 (Tab. I) recently obtained (Genbank, accession numbers AJ131112, AJ251829 and AJ251914) confirmed the presence of at least three potentially functional classical class I genes, but also indicated that one or two additional loci might be expressed. Evidence for the existence of strong linkage disequilibrium extending across the entire MHC region has been obtained in several haplotypes. The significance of this disequilibrium is not clear, and might simply result from a recent admixture of reproducers, or on the contrary from a selective advantage driving force.

2.1. Number of swine leucocyte antigen class I genes

Earlier biological molecular analysis provided evidence for the existence of about 10 SLA class I loci [34, 42]. This came as a surprise since the number of class I sequences found in humans and rodents is three to four times larger. However, among these numerous human and rodent class-I sequences, only two to three genes encode classical class I or Ia genes. The other genes encode class I-related gene families (Ib), the even more distant MIC gene family (Ic) and several pseudogenes. We recently sequenced 460 kb of the genomic SLA class I region, including two distinct segments, one comprising the Ia genes, (EMBL accession numbers AJ131112 and AJ251829) and the other, a tight cluster of Ib and Ic genes (EMBL accession number AJ251914) (Fig. I). Gene and exon predictions were carried out using the GENSCAN programme available on the web site: http://genomic.stanford.edu/GENSCANW.html. The BLAST and CLUSTALW programmes from the GCG package were used for sequence identification and multialignments.

2.2. The SLA Ia genes

From the most centromeric SLA-11 Ia gene, the order of the other Ia genes is SLA-4, -2, -3, -9, -5 and -1 (Fig. 1). The SLA-1, -2 and -3 sequences are functional genes [42], and respectively correspond to the serologically defined SLA C, B and A series. The SLA-4 and SLA-11 sequences are truncated, whereas in the third exon, the SLA-9 sequence displays a stop codon that causes premature termination of the transcription. The structure of the SLA

Table I. SLA haplotypes. The three A, B and C loci were ordered according to the physical map, where the B locus is closest to the centromere and the C locus to the telomere.

	locus C	locus A	locus B		locus C	locus A	locus B		locus C	locus A	locus B
H01	FJ1	W15	B1	H25	W2	FJ20		H49	FJ1	FJ20.W8	
H02	W14	W10		H26	SB19	FJ13	W6	H50	FJ1		
H03	W2	FJ20	FJ23	H27	W5		B11	H51	W5.W14		B11
H04	W9	FJ13		H28		FJ13	W6	H52	FJ1		B3
H05	W5		W4	H29		FJ13		H53	W9	FJ13	W6
H06	W5	FJ20	FJ4	H30	W9			H54		FJ20	
H07	W2	FJ20.W8	B11	H31	W16	W8	B11	H55	W9	FJ13	B3
H08		W17		H32			B1	H56	W5		W6
H09	W5		FJ4	H33	FJ1		B1	H57		FJ20.W7	FJ23
H10	W12	FJ33		H34	FJ1	W15	FJ23	H58	W14	FJ20	FJ23
H11	W2	FJ20	FJ23	H35	W9	W8		H59	fj14		
H12	FJ1	FJ13	B3	H36	W5			H60		FJ20.W8	
H13		FJ13	B3	H37		FJ27		H61	W14	FJ39	
H14	W16	FJ33	B11	H38		W33	FJ28	H62	W5	FJ13	B11
H15	W14		FJ25	H39	W16	W17		H63		W36	
H16	SB19	FJ13	FJ35	H40		SB		H64		BM37	
H17	SB19			H41			FJ23	H65	W2	BM38	
H18	W5			H42		FJ29	FJ4	H66	w2	fj20W13	B11
H19			W6	H43	FJ1	W15		H67		W8	FJ25
H20	W2	W13	B11	H44	W2	FJ20.W7	B11	H68	W36	FJ13	W18
H21	W2	FJ20	B11	H45		W10		H69			FJ11
H22	W14	W8		H46	W12	FJ20.W7		H70	L10	L9	FJ35FJ6
H23		W18		H47		FJ35		H71	FJ13	W9W5.	B11
H24	W5	FJ13	SB22	H48		W8	DC31	H72	L1	L10	FJ1

class Ia gene consists of a leader sequence, three exons encoding corresponding extracellular domains, a transmembrane exon and three intracytoplasmic exons (Fig. 2) [33]. All Ia genes code for a polymorphic heavy glycoprotein chain of about 45 kiloDaltons (kDa) which binds non-covalently to the monomorphic β2-microglobulin at the cell membrane [35].

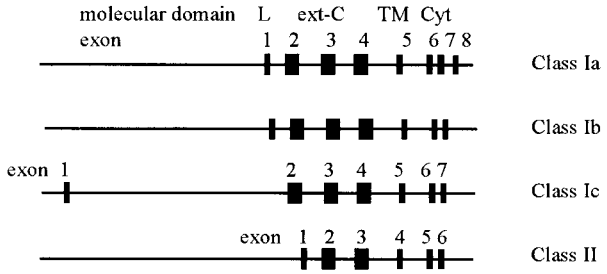

Figure 2. Organisation of different MHC genes. L: leader peptide; Ex-C: extracellular domain; TM: transmembrane domain; cyt: intracellular domain.

The polymorphism of swine SLA class Ia genes corresponds to amino acid substitutions in positions involved either in fixing the foreign peptides or in establishing direct contact with the T-cell-receptor [8]. This polymorphism is located in the extracellular α1 and α2 domains of SLA class I genes, and is almost perfectly superimposable on those of the human HLA class Ia sequences.

The SLA class I antigens are expressed constitutively on all nucleated cells, with however, great variations. The overall structural and regulatory organisation of the MHC class I genes among species, including the pig, is well conserved.

2.3. The SLA Ib genes

As mentioned above, the SLA-6 Ib gene is tightly linked to the SLA-7 and 8 genes, and to the Ic genes. The previously defined SLA-6 (PD6) gene exhibits an overall similarity of 55% with the classical swine Ia genes [13]. The position of the residues involved in glycosylation sites and disulfide bond formation are given in Table II. The size and eight-exon organisation of SLA-6 are similar to those of the class I genes. However, SLA-6 has a codon stop in exon 7 leading to a SLA-6 mature protein comprising some 270 amino acid residues. The SLA-6 gene was present in all the breeds tested, and displayed no polymorphism at the molecular level, thus presumably giving rise to a monomorphic protein. Research in humans and mice for a gene homologous to SLA-6 proved negative, as usually observed for non-classical class I genes.

The SLA-6 and the nearby SLA-7 genes are transcribed in the same direction, while further down the SLA-8 gene is coded for on the opposite strand. The alignments of all known SLA genomic sequences permitted the identification of exons 1 to 5 for SLA-7, and exons 1 to 6 for SLA-8. The putative additional exons of these genes could not be identified by comparison with the other class I gene sequences, either because of too much DNA sequence divergence, or simply because of their absence. The SLA-7 and SLA-8 genes

Table II. Potential glycosylation sites and cystein positions within mature SLA molecules

	Cystein position (amino acid number)				
	SLA2	SLA6	Sla7	Sla8	Mic2
Domain 1			60		
Domain 2	101, 164	101, 164	99, 101, 161	101, 164	95, 119, 126, 144, 162
Domain 3	203, 259	203, 259	203, 259	191, 203, 259	187, 199, 257
Domain 4		287	310		
	Potential glycosylation site (amino acid number)				
Domain 1	86 (1)	86 (2)	86 (1)	86 (1)	
Domain 3					195, 208, 235, 263 (3)

Glycosylation site corresponding to an NQS amino acid motif (1), NHS (2), and NLT, NIS, NGT, NHS (3).

are similar to class Ia genes with regards to the size and organisation of their characterised portions, and probably code for membrane-anchored molecules. The positions of the residues involved in glycosylation sites and disulfide bond formation are given in Table II. Note that at positions 86 to 88 the *N*-linked glycosylation consensus sequence NQS, found in all SLA class Ia sequences analysed so far, was also conserved in the SLA-7 and -8 genes, while for the SLA-6 gene this site comprises the amino acid residues NHS. The residues involved in the binding sites of human $CD8^+$ *T* cells, which have been localised on the $\alpha 2$ and $\alpha 3$ HLA domain, are well conserved in the porcine SLA Ia and Ib sequences.

The SLA-6 gene, previously referred to as PD6 gene, is transcribed in a tissue-specific manner [13]. The SLA-7 gene has been shown to give rise to a transcript in cultured swine fibroblasts [8], although neither its expression in these cells and other tissues nor its polymorphism has yet been studied. The transcription, expression and polymorphism of the SLA-8 gene must also be studied.

These swine class Ib genes appear to be closely related to the SLA Ia genes, and like the human and mouse Ib genes, have no counterparts in either of these species. The functions of the swine Ib genes are not known, but their tasks may include the control of NK cell activity or presentation of specific peptides like the mouse H2-M3 molecules [34, 27]. The functions of most class Ib genes in humans and mice are in fact still unknown, although a putative role in reproduction has been attributed to some of them.

2.4. MIC (Ic) genes

2.4.1. Human MIC genes

In humans, the MIC genes constitute a family of five related sequences, two of which are functional [18]. The overall predicted domain structures are similar to those of the class I genes. The MICA and MICB genes are closely related and exhibit a 91% similarity in their coding sequences, while the amino

acid similarity with the human or mouse class Ia or Ib genes ranges from 15 to 35%. Unlike the classical and other class I-related products, the MIC proteins do not associate with $\beta 2$ microglobulin. In humans, an unexpected polymorphism has been detected with at least 16 alleles for MICA, due to the occurrence of clustered amino acid substitutions in all three external domains. Curiously, the MIC polymorphic residues are located mainly at the periphery of the putative antigen-binding groove. The MIC proteins have a restricted pattern of expression with abundant expression in intestinal mucosa and also in a number of tumours [2]. In addition, they are expressed in the epithelial cells present in the subcapsular cortex of the thymus in freshly isolated keratinocytes, and in endothelial cells and monocytes [44]. Unlike the classical class I genes, the transcriptional control regions of the MICA and B genes lack an interferon-response element, but their promoter region contains a heat shock element, as observed in HSP70 genes. This finding is consistent with the large increase in the mRNA and protein expression of both MIC genes following heat shock induction. The MIC proteins are the ligands of the NKG2D receptor expressed by the $T\gamma\delta$ and NK cells [2]. NKG2D engagement causes the lysis of cells expressing MIC, the MIC proteins serving as an immune surveillance mechanism [18]. MIC genes have been identified in most mammalian species, but not in the mouse genome. In this species, however, the H2-T22^{b} and H-2T10^{b} genes might correspond to homologous evolutionary genes [17].

2.4.2. Swine MIC-like (Ic) genes

On the basis of the MIC sequences characterised, only the swine MIC2 gene is probably functional. Alignment with the human MIC genes permitted only the characterisation of the putative exons 2, 3 and 4. A putative exon 1 which might code for a peptide leader of 23 amino acids was localised 4648 bp upstream of exon 2 (Fig. 2). Upstream of exon 1, we found, as observed for the human MIC genes, a consensus heat shock element. In addition, the swine MIC2 sequence displays specific characteristics of the human MIC genes, such as four putative N-glycosylation sites, three of which are clearly counterparts of the human glycosylation sites. Similarly, several cystein residues of MIC2 domains 2 and 3 may participate in the building of disulfide bonds (Tab. II).

2.5. Relationships between the SLA Ia, Ib and Ic genes

The overall homology of the coding regions between the SLA Ia, Ib and Ic genes is low, ranging from 65 to 75%, compared to the homology of nearly 90% observed between the SLA class Ia genes. As shown in the phylogenic tree (Fig. 3), the SLA-7 and SLA-8 sequences are closer to each other and to the SLA-2 Ia genes than to the SLA-6 sequence. These findings suggest that although the three SLA Ib genes were all generated from a single SLA class Ia ancestor gene(s), it is likely that the SLA-6 gene diverged from a common ancestor before the generation of the SLA-7 and SLA-8 genes. Surprisingly, the swine Ib genes appear to be evolutionary more distant from the SLA class Ia genes than are the human Ib genes from the HLA Ia genes. The characterised domains of the swine MIC2 genes displayed a little more than 50% similarity to the human MIC sequences, but less than 25% to the SLA-2 Ia and SLA Ib sequences.

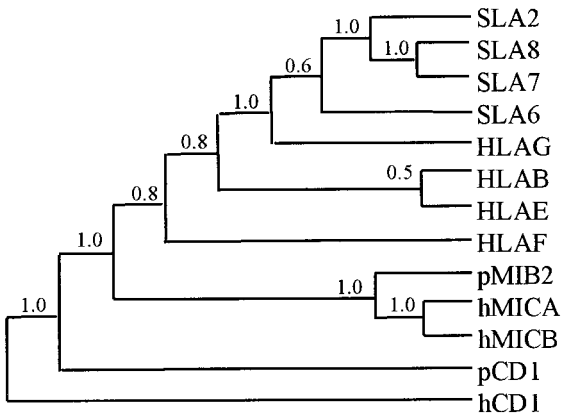

Figure 3. Relationship between constant domains of various class I molecules. Amino acid sequences were aligned using the Clustal *W* programme. The tree was constructed using the neighbour-joining method. The number at each fork indicates the % of time for which that node was supported in 1000 pseudo-replications. pMIC, pCD1: pig MIC or CD1 molecules; hMIC, hCD1: human MIC or CD1 molecules.

2.6. Non MHC class I sequences

The other loci of the SLA I-region, found in several species including swine, are listed in Table III, together with current information concerning their functions.

Table III. Genes of the class I region unrelated to the SLA class I gene family and their encoded protein putative functions.

SC1 (TCF19):	trans-activating factor with late growth-regulating activity.
POU5F1:	octamer transcription factor containing a POU domain.
p52:	protein subunit of the TFIIH transcription / DNA repair factor.
S:	keratin-like protein specifically expressed in the granular layer of the epidermis.
ZNF173:	acid finger protein.
RFB30:	ring finger B30 protein.
MOG:	myelin oligodendrocyte glycoprotein.
OLF42:	olfactory receptor-like genes.
OLF89:	
BUT:	butyrophylin (milk protein, related to RFB30).

3. THE SLA CLASS III REGION

The SLA class III region is centromeric and contiguous to the class-I region, with the most distal class III BAT-1 gene less than 20 kb from SLA-6. The class III region spans about 700 kb of DNA and contains over 35 characterised genes

[40]. Many of them are involved in important non-specific or innate defence mechanisms, such as the complement components, the TNF gene families, the Hsp70 gene family, the RAGE gene and the allograft inflammatory factor-1 (Tab. IV). Comparison with the human HLA class III region confirmed the good overall conservation of the class III gene cluster during evolution. Nevertheless, the segment containing CYP21, TNX and C4 which in humans and mice has been independently duplicated in tandem, has not been duplicated in pigs.

4. THE SLA CLASS II REGION

The SLA class II region spans about 500 kb and has been shown to harbour eighteen sequences including the classical class II DR and DQ genes, class II-related sequences, and also most of the other genes found in the HLA and H2 class II regions (Fig. 1) [9]. The swine DQ and DR genes display the overall organisation established for the class II genes in all mammalian species (Fig. 2) [29]. All functional class II molecules are transmembrane heterodimers which consist of an α-chain with a molecular weight of about 34 kDa non-covalently bound to a β-chain of about 29 kDa [19]. While the single swine DRA gene is monomorphic, the DQA gene and above all the DQB and DRB genes are polymorphic. This polymorphism, revealed by direct nucleotide sequencing, is essentially concentrated in the first domain (second exon) [8], in a region which forms the floor and wall of the peptide-binding groove. The serology of the SLA class II molecules has not been used routinely, mainly because of the existence of a wide range of class II cross-reactions among the classical reagents. In addition, there were too few monoclonal antibodies to permit class II typing.

The results of both Southern blot analyses and polymerase chain amplification indicated the existence of three DRB sequences per haplotype in pig breeds and wild boar. Only one DRB gene was however considered to be functional since other two displayed clear characteristics of pseudogenes [5]. Next to the DRB loci lies the DQA locus, and further telomeric are two DQB loci, one of which may correspond to a functional gene, and the other, to a pseudogene. Further telomeric, there is another class II-like sequence, provisionally called DOB since it might be the swine counterpart of the human DOB sequence. Another SLA class II sequence was designated as DPA, although no DPB locus has yet been identified. The LMP-7 locus has been tentatively located between the DMB and DOB loci close to the TAP1 and TAP2 genes, in accordance with the order of these genes in the class II HLA region. The telomeric end of the class II region comprises five tightly linked genes, including those coding for COL11A2, RXRB, KE6, KE4 and RING1. The putative functions of the other genes of the class II region are described in Table V.

SLA class II molecules are expressed in a tissue-restricted manner. They are found mainly on lymphoreticular *B* cells and macrophages, and are especially abundant on mature dendritic cells which are highly efficient initiators and regulators of immune responses. SLA class II antigens are also expressed on a significant proportion of circulating *T* cells, and on some parenchymatous cells such as kidney cells.

Table IV. Genes of the SLA class III region and their encoded protein putative functions.

RAGE:	receptor for advanced glycosylation end products of proteins. Major signal transduction receptor for members of a family of closely related polypeptides released from activated inflammatory cells.
PBX2:	homeodomain-containing protein.
G15:	lysophosphatidic acid acyltransferase.
CREB-RP (G13):	cAMP response element binding protein-related protein.
TNX:	tenascin *X* is a component of the extracellular matrix.
CYP21:	21-hydroxylase enzyme from the cortisol and aldosterone synthesis pathway.
C4:	complement component involved in the classical pathway cascade.
G11:	putative protein kinase activity.
BF:	factor *B*, a serine protease involved in the complement alternative pathway.
C2:	serine protease involved in the complement classical pathway.
HSP70:	cluster of three distinct heat shock proteins constitutively expressed or heat-inducible.
BAT7:	sialidase enzyme with optimal activity at acidic pH.
VARS2:	valyl tRNA synthetase.
MSH5:	MutS homologue 5 affects testicular size and ovarian structure.
G2 (BAT2):	prolin-rich protein with novel repeat elements.
AIF-1 (G1 ?):	allograft inflammatory factor-1 also involved in the control of insulin production.
LTB:	lymphotoxin *B* anchors LTA to cell membranes.
TNFA:	tumor necrosis factor A plays a major role in inflammation, immunomodulation and lipid metabolism.
LTA:	lymphotoxin A involved in lymphoid organ development and germinal center formation.
IKbL:	inhibitor of transcription factors.
V-ATPaseG:	vacuolar-ATPase subunit involved in a broad range of cellular functions, including glycosylation in the Golgi body and degradation of cellular debris in lysosomes
BAT1:	putative RNA helicase of the DEAD family.
G18, G17, G16, G14,G9a ,G7,G6, BAT9, BAT8, BAT5, BAT4, BAT3, RD:	genes with unknown functions

Table V. Genes of the class II region unrelated to the SLA class II gene family and their encoded protein putative functions

Tapasin:	TAP-associated protein is a transmembrane molecule of the immunoglobulin superfamily probably emerged from an ancestor of the MHC class I / II b-like genes. Plays a critical functional role in MHC class I-restricted antigen processing.
RING1:	RING finger protein.
KE6:	steroid and prostaglandin dehydrogenase-related protein.
KE4:	membrane hydrophobic protein with histidine and glycine-rich domains.
RXRB:	retinoid acid X receptor is a transcription factor for class I genes.
COL11A2:	collagen type 11 α2.
TAP1&2:	transporter associated with antigen processing 1 and 2 form a complex that translocates peptides generated in the cytoplasm into the endoplasmic reticulum vesicles.
LMP7:	Low molecular weight polypeptide catalytic subunit which translocates into the proteasome complex following activation by cytokines (notably γ-interferon).

5. THE MHC FRAMEWORK REGION AND ITS EVOLUTIONARY ASPECTS

5.1. Recombination within the SLA region

The recombination rate within the swine SLA complex is less than 1%. Fifteen of the 17 recombinations documented seem to have occurred within the class III region, although the precise crossover site is not known. The crossover site of the remaining two recombinants occurred between the *A* and *B* loci, that is at the centromeric end of the classical class I chromosomal segment (Fig. 1). The distance between the SLA-*A* and *B* loci is no more than 12 kb and might contain a recombinational hot spot.

5.2. The class I orthologous region, and permissive places

In species not closely related, such as humans, mice and swine, class I gene expansion seems to have resulted from duplication events which occurred after the divergence of these species. Nevertheless, although the orthologs of the class I genes between humans, mice and swine cannot be found, the precise physical organisation of the MHC class I regions of these species so far characterised shows that these regions are orthologous. They each comprise three segments, referred to as permissive places [1] in which all class I genes were located (Fig. 4). In humans and mice, the orthologous segment located between the TNF and POU5F1, single copy anchor sequences, harbours classical class Ia genes, whereas in swine, this orthologous segment contains the class I-related Ib genes. Conversely, whereas in man and the mouse the permissive place between the HSR1 and ZNF173 sequences contains only class Ib genes, in swine, this place contains all the classical class-Ia genes. The third permissive place, identified between the RFB30 and MOG sequences in humans and mice, is absent in pigs.

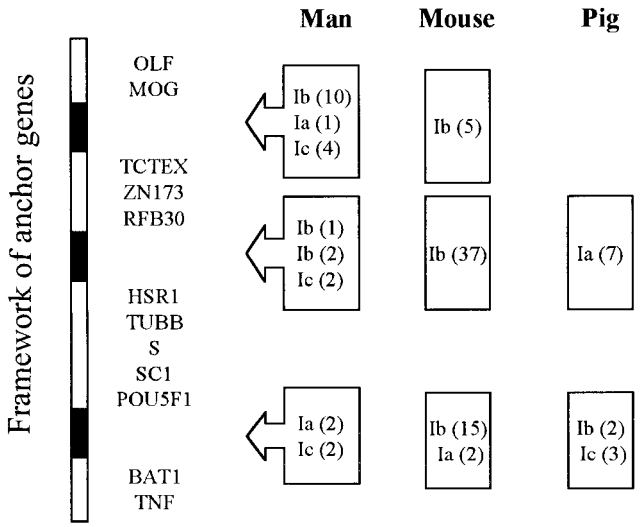

Figure 4. Location of class I-related sequences in human, mouse and pig MHC: Ia classical class, Ib non classical, Ic MIC sequence, in parenthesis: number of sequences. Black boxes: permissive regions for gene duplications.

In conclusion, studies in swine confirm that class I insertional events and their subsequent duplication and deletion occurred independently in most mammalian species. This finding constitutes new evidence for the remarkable conservation of the general MHC framework throughout long evolutionary periods. With regards to the origin of the MHC region, it has been shown that the MHC class I and class II genes tapasin, RAGE, BT and RFB30 have common ancestors and were derived from these ancestors by cis-duplication and shuffling. In addition, the discovery in *Caenorhabditis elegans* of a linkage group containing several MHC genes [38], mostly consisting of class III-like genes but also including a proteasome subunit, confirms that this syntenic group existed long before the emergence of prototype class I and class II genes.

5.3. Pig endogenous retroviral sequences

About 50 copies of the type *C* endogenous retrovirus were reported in the swine genome of which 10 to 20 copies might correspond to a full-length provirus. These copies are generally not arranged in tandem, although clusters of integration sites were observed in some areas, including the 7p1.1 band, where the SLA class I region is located [32]. Although these SLA – associated endoviral elements appear to be located outside the class I region itself, they might be partly responsible for the plasticity of the region.

5.4. Non-MHC-encoded class I-related genes

5.4.1. Related class I sequences

The results of both comparative gene mapping and chromosome painting point to conserved syntenic regions between the small arm of the human

6 chromosome which among other genes harbours the HLA complex, and the small arm of the swine 7 chromosome and part of its pericentric region [16]. One may therefore expect, that the genes located near the HLA complex, such as the class I-related HFE gene which has been mapped about 4Mb telomeric to the HLA-F gene, will also be genetically linked to SLA. Since HFE plays a major role in iron metabolism [30], its swine counterpart might be of interest, especially in piglets which usually display dramatic anaemia. In the mouse, the HFE gene maps on chromosome 13, in a region paralogous to the H-2 region located on chromosome 17.

Beside the MHC classical class-I genes and their variously related genes, the mammalian genome comprises a number of other class I genes located in other parts of the genome [20]. These include the CD1 family (cluster differentiation), the neonatal Fc receptor, the Zn-α2-glycoprotein, and the MR1 gene (for MHC class I-related). Except for the CD1 family, none of these genes has so far been identified in the pig.

5.4.2. The CD1 gene family

The swine CD1 genes constitute a family of at least four closely related genes, all clustered on pig chromosome 4 [15]. Their intron/exon molecular organisation is similar to that of MHC class I genes, although they display a significant homology to both class I and class II genes. The CD1 gene in fact emerged about 250-300 million years ago, from a primordial antigen-presenting ancestor gene which also engendered the precursors of todays MHC class I and class II genes. The CD1 proteins in the α2 and α3 domains display a homology of about 35% with the corresponding MHC class I domains. A direct homology was found between the CD1 genes in all mammalian species including man, despite great variations in the number of genes and loci expressed. The swine pCD1.1 gene, the best so far analysed, exhibited the greatest similarity to human CD1a, and another swine pCD1.1 gene seemed to be similar to the human CD1b gene [10].

The CD1 proteins have a characteristic folded structure that is similar to the MHC class I proteins, and possess a binding groove with specific features enabling the CD1 isoforms to bind lipid and glycolipid ligands [31]. These ligands are restricted to T cells comprising mainly circulating CD4-CD8-T cells or $\gamma\delta$ TCR + T cells, population types which are particularly abundant in swine. Note that the CD1 proteins also bind the β2 microglobulin noncovalently. As in humans, the swine pCD1.1 molecule was shown to be expressed by a subset of thymocytes, professional antigen-presenting cells such as dendritic cells, and macrophages and B lymphocytes. However, no CD1.1 expression has been found in endothelial or epithelial cells. The involvement of CD1 proteins in immune protection against microbial pathogens must now be evaluated.

5.5. The swine MHC paralogous regions

Paralogous regions are defined as chromosomal segments containing closely linked pairs of duplicated genes. In humans, certain regions of chromosomes 1, 9 and 19 display similarities to the MHC region of chromosome 6 in both their gene content and organisation [24]. Similarly, the swine MHC paralogous regions map to chromosomal segments 1q2.9-13, 4q1.5-1.6 and 2p14-17 (Tab. VI),

Table VI. Gene family members present in the 4 paralogous MHC regions in humans and the corresponding regions in pig.

Human chromosomes				Pig chromosomes			
6p21.3	19p13.1	9q33-34	1q21-25	7 p11-q11	2p14-17	1q28-13	4q1-5
BING1	HUNK1	ZNFX	BRDT				
RING3		ORFX		RING1			
COL11A2		COL5A2	COL11A1	COL11A2		COL5A1	
RXRB		RXRA	PBX1	PBX2			
PBX2	PBX4	PBX3	RXRG	RXRB			
NOTCH4	NOTCH3	NOTCH1	NOTCH2				
TNX		TNC	TNR	TNX		TNC	
C4	C3	C5		C4	C3		
CYP21				CYP21		CYP19[(1)]	
HSPA1		HSPD5	HSPA6,7			HSPD5	HSPA6
VARS1		VARS2		VARS			
BAT2		BAT2-1		BAT2			
BAT1	BAT1-1			BAT1			
OLF42, 89				OLF42, 89		OLF49[(2)]	
HLA class 1 2			CD1	SLA class 1 2		B2m[(3)]	CD1

[(1)] q14-17; [(2)] q2-12; [(3)] q-17.

which respectively have their corresponding segments in the human chromosomes 9, 1 and 19 [16].

6. MHC COMPLEX INVOLVEMENT IN PHYSIOLOGICAL AND PATHOLOGICAL SYNDROMES

Numerous HLA-associated diseases have been described, most of which are auto-immune syndromes [14]. Little is known about the mechanisms behind these diseases except for those caused by a deficiency of specific genes such as the C2 and 21-OH genes. Reports that class I or class II alleles exert direct protective effects against viral, bacterial or metazoan parasitic infections are few, both in humans and other vertebrates [21]. Nevertheless, there are consistent indications of human HLA complex involvement in different retroviral infections, such as HIV, where a significant association was observed between HLA class I homozygosity and rapid disease progression [7]. Similarly, HLA alleles affect the human *T*-lymphotropic virus-I (HTLV-I) proviral load and may increase the risk of HTLV-I-associated myelopathy. The outcome of hepatitis B and papilloma virus infections has also been associated with HLA alleles. The role of certain HLA antigens in protection against malaria caused by Plasmodium falciparum has been clearly shown, and recent results strongly indicate a host-parasite coevolution. HLA class II alleles have also been suggested to be involved in human onchocerciasis, a tissue nematode transmitted by insect vectors. The effect of the chicken MHC H-B complex on Mareck disesase has also been well documented [25], as has that of the BoLA complex in cattle in the bovine leukaemia virus infection (BLV). BLV is a member of the human HTLV subgroup of retroviruses which after infection cause persistent lymphocytosis whose development is controlled by the BolA-DR3 allele [28].

There is increasing evidence that genes of the MHC region other than those belonging to the class I or class II families may be directly involved in the outcome of infectious pathogenic syndromes. For instance, one nucleotide mutation in the TNF promoter region affecting the binding of the transcription factor OCT-1 is associated with a significantly increased risk to severe malaria [26]. TNF alleles and the TNF subregion have also been associated with various metabolic disorders such as insulin-dependent diabetes mellitus disease, severe debilitating inflammatory syndromes and cancer.

6.1. Involvement of the SLA region in pathological syndromes and the immune response

6.1.1. Cutaneous malignant melanoma

Segregation analyses of the occurrence of melanocytic lesions in the American miniature pig Sinclair line were consistent with a two-loci model involving an as yet undefined major initiator gene and a second locus belonging to the SLA region [4]. One particular SLA haplotype appeared to be necessary for the tumour initiator locus to be fully penetrant. In another herd of hybrid pigs comprising genes of five origins, including the Sinclair line, the spontaneous reappearance of the malignant melanoma also seemed to correlate with an ancestral SLA haplotype [22]. In contrast, segregation analysis of melanocytic

lesions in crosses with the Munich miniature swine Troll in Germany, showed no involvement of the SLA-complex. However, a linkage was found between TNFB-locus and a particular type of nevi [11].

6.1.2. The SLA complex and immune responsiveness

A favourite explanation for the class I and II genes being the most polymorphic genes known in mammals is the existence of an overdominant selection or heterozygote advantage effect. Thus, individuals heterozygous at the MHC-presenting molecules are more prone to present a larger set of peptidic epitopes than homozygotes, leading to a larger range of responses. However, at least in mice, the MHC polymorphism can also result from an MHC-incompatible mating preference in the wild, which indirectly leads to a similar advantage with regards to the immune response [43].

The role of the SLA complex in allograft tissues and organs has been fully assessed and therefore constitutes a well characterised model in biomedical research. Similarly a recently published review [40] was devoted to the responsability of the SLA complex for at least part of the genetic control of the immune response involving both antibodies and cellular responses to conventional antigens, pathogenic viruses and the nematode *Trichinella spiralis*.

No pathologies have been attributed to the SLA class III region so far. Like the human C4B component, the swine C4 molecule exhibits higher affinities to hydroxyl group-containing targets than to amino group-containing targets, although its average normal level is low. Nevertheless the SLA complex was shown to significantly affect complement haemolytic activity tests in a Large White herd [41].

6.2. The swine leukocyte antigen and production performances

A wide variety of productive and reproductive traits is affected by SLA specific haplotypes [40]. An extensive evaluation of these traits was recently conducted in a large herd of F2 pigs issued from initial matings of Meishan and French Large White grandparents. In all, eleven-hundred pigs were SLA typed and screened for more than 20 quantitative traits, from birth until they weighed 100 kg. Overall, the results of this evaluation confirmed that the SLA region indeed affects a number of traits, such as the level of androstenone, carcass composition, growth rate, the thickness of the dorsal fat layer, the intra-muscle fat contain, the malic enzyme activity level, and the development of reproductive organs in sires and females. In addition, parameters affected by the SLA region were discovered, including the basic glycaemia level at 120 days of life, and the hematocrit and haemoglobin levels [3].

7. CONCLUDING REMARKS

The SLA region is by far the best characterised region of the whole swine genome, and we are on the verge of obtaining its complete nucleotide sequence. The sequencing of the class I region was essential for the unravelling of the overall SLA class I organisation. In particular, it permitted individual characterisation of all class I genes, which because of their close homology,

had not been clearly assigned to a specific locus. SLA sequencing will also help to clarify the differences observed between the levels of expression of certain Ia alleles in miniature and French Large White pigs. We were also able to definitely demonstrate the existence in the class I region of three SLA class I-related genes which may in fact correspond to three distinct lineages. Two of these genes, namely SLA-6 and SLA-7 have been shown to be expressed. However, the tissular expression regulation and the role of the SLA Ib genes, as well as those of the MIC-like sequences must now be defined.

Analysis of the SLA structure has also confirmed the overall orthology between species for the entire MHC region, including the class I regions. The SLA class I region clearly comprises two permissive places in which the class I genes have evolved. However, in swine, the class Ia and Ib genes have opposite locations in these places compared to the locations of the Ia and Ib genes in men and mice (Fig. 4).

The involvement of the SLA region in various QTLs also needs to be defined. It is important to remember that many of the associations observed between the pig MHC region and its traits, were previously observed in the mouse H-2 region [23], and occasionally also in domestic species such as cattle and chickens. A number of these associations concerned traits related to reproductive functions, including embryo development, litter size, mating preference, and sometimes miscarriage [40]. The MHC class Ia molecules are probably not involved in any of these associations, since these molecules are not expressed at the surface of the embryos. However, in man the HLA-G Ib gene, which codes for an essentially monomorphic molecule, is primarily transcribed in cells of the trophoblastic lineage [6] starting from the preimplantation embryo stage. Similarly, in mice, class Ib but not Ia molecules have been shown to be expressed on the surface of mouse embryos before the implantation stages in particular strains. The function of these Ib molecules in reproduction is not known, although it has been postulated, at least for HLA-G, that they may protect the trophoblasts from NK cell-mediated lysis, because they behave like a self-element of the mother [6]. In mice, the Q9 Ib gene product controls the rate of cell cleavage and embryonic survival. It will therefore be of great interest to determine whether one of the SLA Ib genes is also preferentially expressed by the trophoblasts in swine. The presence of the β2-microglobulin molecule on the outer trophectoderm cell layer of the swine blastocysts before their implantation was readily observed. The association of this molecule with SLA classical antigens on blastocysts could not be established. Among the genes located in the MHC region which are not related to the class I genes, the presence of olfactory factors encoding genes is worth noting in connection with the importance for gilt puberty or boar taint. Similarly, the tissue level of androstenone, a compound which functions like a pheromone in the boar, is also controlled by the SLA region, although the gene involved has not yet been characterised [40]. Among the other peculiarities of swine reproduction, the segregation of the SLA H07 haplotype was observed to deviate significantly from the usual 1 :1 ratio in a number of back-cross families. It is not known whether this was due to the effect of a swine homologous region to the T/t complex, which in mice is located near the H-2 complex, or to genes of the SLA complex itself. Several genes within the human MHC in fact code for products which are specifically transcribed in the testis, including one involved

in gonadal maintenance [12]. Similarly, the mouse H-2 located Ring3 gene [36], has been suggested to play an important role in spermatogenesis.

REFERENCES

[1] Amadou C., Evolution of the Mhc class I region: the framework hypothesis, Immunogenetics 49 (1999) 362–367.

[2] Bauer S., Groh V., Wu J., Steinle A., Phillips J.H., Lanier L.L., Spies T., Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA, Science 285 (1999) 727–729.

[3] Bidanel J.P., Milan D., Iannuccelli N., Amigues Y., Boscher M.Y., Bourgeois F., Caritez J.C., Gruand J., Le Roy P., Lagant H., Bonneau M., Lefaucheur L., Mourot J., Prunier A., Désautés C., Mormède P., Renard C., Vaiman M., Robic A., Gellin J., Ollivier L., Chevalet C., Détection de locus à effets quantitatifs dans le croisement entre les races porcines Large White et Meishan. Résultats et perspectives, in: 32es Journées de la Recherche Porcine en France, 1-3 février 2000, Institut Technique du Porc, Paris, pp. 369–383.

[4] Blangero J., Tissot R.G., Beattie C.W., Amoss M.S., Genetic determinants of cutaneous malignant melanoma in Sinclair swine, Br. J. Cancer 73 (1996) 667–671.

[5] Brunsberg U., Edfors-Lilja I., Anderson L., Gustafsson K., Structure and organization of pig MHC class II DRB genes: evidence for genetic exchange between loci, Immunogenetics 44 (1996) 1–8.

[6] Carosella E.D., Rouas-Freiss N., Paul P., Dausset J., HLA-G: a tolerance molecule from the major histocompatibility complex, Immunol. Today 20 (1999) 60–62.

[7] Carrington M., Nelson G.W., Martin M.P., Kissner T., Vlahov D., Goedert J.J., Kaslow R., Buchbinder S., Hoots K., O'Brien S.J., HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage, Science 283 (1999) 1748–1752.

[8] Chardon P., Renard C., Vaiman M., The major histocompatibility complex in swine, Immunol. Rev. 167 (1999) 179–192.

[9] Chardon P., Rogel-Gaillard C., Peelman L., Yerle M., Velten F.W., Renard C., Vaiman M., Physical organization of the pig major histocompatibility complex class II region, Immunogenetics 50 (1999) 344–348.

[10] Chun T., Wang K., Zuckermann F.A., Gaskins H.R., Molecular cloning and characterization of a novel CD1 gene from the pig, J. Immunol. 162 (1999) 6562–6571.

[11] Distl O., Muller S., Wanke R., Hermanns W., Anderson L., Inheritance of melanocytic lesions and their association with the white color phenotype in miniature swine, in: Proceeding of the 6th world congress on genetics applied to livestock production, 26, January 11-16, 1998, Armidale, NSW, Australia, pp. 445–448.

[12] Edelman W., Cohen P.E., Kneitz B., Winand N., Lia M., Heyer J., Kolodner R., Pollard J.W., Kucherlapati R., Mammalian Muts homologue 5 is required for chromosome pairing in meiosis, Nat. Genet. 21 (1999) 123–127.

[13] Ehrlich R., Lifshitz R., Pescovitz M.D., Rudikoff S., Singer D.S., Tissue-specific expression and structure of a divergent member of a class I MHC gene family, J. Immunol. 139 (1987) 593–602.

[14] Foissac A., Crouau-Roy B., Fauré S., Thomsen M., Cambon-Thomsen A., Microsatellites in the HLA region: an overview, Tissue Antigens 49 (1997) 197–214.

[15] Garrido J.J., Yerle M., Chardon P., Barbancho M.J., Andrés-Cara D.F., Assignment of cluster of differentiation 1 locus (CD1) to pig chromosome bands 4q1.5→q1.6 by in situ hybridization, Cytogenet. Cell Genet. 83 (1998) 88–89.

[16] Goureau A., Yerle M., Schmitz A., Riquet J., Milan D., Pinton P., Frelat G., Gellin J., Human and porcine correspondence of chromosome segments using bidirectional chromosome painting, Genomics 36 (1996) 252–262.

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

[18] Groh V., Steinle A., Bauer S., Spies T., Recognition of stress-induced MHC molecules by intestinal epithelial $\gamma\delta$ T cells, Science 279 (1998) 1737–1740.

[19] Gustafsson K., LeGuern C., Hirsch F., Germana S., Pratt K., Sachs D.H., Class II genes of miniature swine. IV. Characterization and expression of two allelic class II DQB cDNA clones, J. Immunol. 145 (1990) 1946–1951.

[20] Hashimoto K., Hirai M., Kurosawa Y., A gene outside the human MHC related to classical HLA class I genes, Science 269 (1995) 693–695.

[21] Hill A.V.S., The immunogenetics of human infectious diseases, Annu. Rev. Immunol. 16 (1998) 593–617.

[22] Hruban V., Horâk V., Fortyn K., Presence of specific MHC haplotype in melanoblastoma-bearing minipigs, Anim. Genet. 25 (Suppl. 2) (1994) 20.

[23] Ivanyi P., Some aspects of the H-2 system, the major histocompatibility system in the mouse, Proc. R. Soc. Lond. B. 202 (1978) 117–158.

[24] Kasahara M., The chromosomal duplication model of the major histocompatibility complex, Immunol. Rev. 167 (1999) 17–32.

[25] Kaufman J., Salomonsen J., The 'minimal essential MHC' revisited: both peptide-binding and cell surface expression level of MHC molecules are polymorphisms selected by pathogens in chickens, Hereditas 127 (1997) 67–73.

[26] Knight J.C., Udalova I., Hill A.V.S., Greenwood B.M., Peshu N., Marsh K., Kwiatkowski D., A polymorphism that affects OCT-1 binding to the TNF promoter region is associated with severe malaria, Nat. Genet. 22 (1999) 145–150.

[27] Lenz L.L., Bevan M.J., CTL responses to H2-M3-restricted Listeria epitopes, Immunol. Rev. 158 (1997) 115–121.

[28] Lewin H.A., Russell G.C., Glass E.J., Comparative organization and function of the major histocompatibility complex of domesticated cattle, Immunol. Rev. 16 (1999) 145–158.

[29] Osborne B.A., Lunney J.K., Pennington L., Sachs D.H., Rudikoff S., Two-dimensional gel analysis of swine histocompatibility antigens, J. Immunol. 131 (1983) 2939–2944.

[30] Parkkila S., Waheed A., Britton R.S., Bacon B.R., Zhou X.Y., Tomatsu S., Fleming R.E., Sly W.S., Association of the transferrin receptor in human placenta with HFE, the protein defective in hereditary hemochromatosis, Proc. Natl. Acad. Sci. USA 94 (1997) 13198–13202.

[31] Porcelli S.A., Modlin R.L., The CD1 SYSTEM: Antigen-presenting molecules for T cell recognition of lipids and glycolipids, Annu. Rev. Immunol. 17 (1999) 297–329.

[32] Rogel-Gaillard C., Bourgeaux N., Billault A., Vaiman M., Chardon P., Construction of a swine BAC library: application to the characterization and mapping of porcine type C endoviral elements, Cytogenet. Cell Genet. 85 (1999) 205–211.

[33] Satz M.L., Wang L.-C., Singer D.S., Rudikoff S., Structure and expression of two porcine genomic clones encoding class I MHC antigens, J. Immunol. 135 (1985) 2167–2175.

[34] Shawar S.M., Cook R.G., Rodgers J.R., Rich R.R., Specialized functions of MHC class I molecules: I. An N-formyl peptide receptor is required for construction of the class I antigen Mta, J. Exp. Med. 171 (1990) 897–912.

[35] Singer D., Camerini-Otero R., Satz M., Osborne B., Sachs D., Rudikoff S., Characterization of a porcine genomic clone encoding a MHC antigen: expression in mouse L cells, Proc. Natl. Acad. Sci. USA 79 (1982) 1403–1407.

[36] Taniguchi Y., Matsuzaka Y., Fujimoto H., Miyado K., Kohda A., Okumura K., Kimura M., Inoko H., Nucleotide sequence of the *Ring3* gene in the class II region of the mouse MHC and its abundant expression in testicular germ cells, Genomics 51 (1998) 114-123.

[37] The MHC sequencing consortium. Complete sequence and gene map of the human major histocompatibility complex, Nature 401 (1999) 921–923.

[38] Trachtulec Z., Hamvas R.M.J., Forejt J., Lehrach H.R., Vincek V., Klein J., Linkage of TATA-binding protein and proteasome subunit C5 genes in mice and humans reveals synteny conserved between mammals and invertebrates, Genomics 44 (1997) 1–7.

[39] Trowsdale J., Campbell R.D., The 12th international MHC map, in: Charron D. (Ed.), Genetic diversity of HLA functional and medical implication, EDK, Paris, 1997, pp. 8–12.

[40] Vaiman M., Chardon P., Rothschild M.F., Porcine major histocompatibility complex, Rev. Sci. Tech. Off. Int. Epiz. 17 (1998) 95–107.

[41] Vaiman M., Hauptmann G., Mayer S., Influence of the major histocompatibility complex in the pig (SLA) on serum haemolytic complement levels, J. Immunogenet. 5 (1978) 59–65.

[42] Velten F.W., Renard C., Rogel-Gaillard C., Vaiman M., Schrezenmeir J., Chardon P., Spatial arrangement of pig MHC class I sequences, Immunogenetics 49 (1999) 919–930.

[43] Zavazava N., Eggert F., MHC and behavior, Immunol. Today 18 (1996) 8–10.

[44] Zwirner N. W., Fernandez-Vina M.A., Stastny P., MICA, a new polymorphic HLA-related antigen, is expressed mainly by keratinocytes, endothelial cells, and monocytes, Immunogenetics 47 (1998) 139–148.