Difference in number of loci of swine leukocyte antigen classical class I genes among haplotypes ☆

Maiko Tanaka-Matsuda [a,b], Asako Ando [c], Claire Rogel-Gaillard [d], Patrick Chardon [d], Hirohide Uenishi [b,e,*]

[a] Second Research Division, Institute of Society for Techno-innovation of Agriculture, Forestry and Fisheries (STAFF), 446–1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305–0854, Japan
[b] Animal Genome Research Program (NIAS/STAFF), 446–1 Ippaizuka, Kamiyokoba, Tsukuba, Ibaraki 305–0854, Japan
[c] Department of Molecular Life Science, Division of Basic Molecular Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259–1193, Japan
[d] INRA, Institut National de la Recherche Agronomique; CEA, Commissariat à l'Energie Atomique, Laboratoire Mixte de Radiobiologie et d'Etude du Génome, Domaine de Vilvert, 78352 Jouy-en-Josas, France
[e] Division of Animal Sciences, National Institute of Agrobiological Sciences (NIAS), 2–1–2 Kannondai, Tsukuba, Ibaraki 305–8602, Japan

ARTICLE INFO

ABSTRACT

The structure of the entire genomic region of swine leukocyte antigen (SLA)—the porcine major histocompatibility complex—was recently elucidated in a particular haplotype named Hp-1.0 (H01). However, it has been suggested that there are differences in the number of loci of *SLA* genes, particularly classical class I genes, among haplotypes. To clarify the between-haplotype copy number variance in genes of the *SLA* region, we sequenced the genomic region carrying *SLA* classical class I genes on two different haplotypes, revealing increments of up to six in the number of classical class I genes in a single haplotype. All of the *SLA-1*(-like) (*SLA-1* and newly designated *SLA-12*) and *SLA-3* genes detected in the haplotypes thus analyzed were transcribed in the individual. The process by which duplication of *SLA* classical class I genes was likely to have occurred was interpreted from an analysis of repetitive sequences adjacent to the duplicated class I genes.

© 2008 Elsevier Inc. All rights reserved.

Introduction

The major histocompatibility complex (MHC) is one of the most complicated regions in the whole mammalian genome [1]. It encodes many proteins that are closely involved in the immune response through the presentation of peptidic antigens to T lymphocytes. In the MHC region, many paralogous genes have been generated in the process of evolution from particular ancestral genes. Among the MHC class II genes of mammals, many orthologs, such as *DR*, *DQ*, *DP*, *DM*, and *DO*, are shared among species; the relationships between orthologs can be easily identified [1,2], although some species such as cats and cattle possess different member genes at loci in the shared class II genes [3,4]. In contrast, the process of evolution of MHC class I genes in mammals is extremely complicated, and there are many types of species-specific expansions of class I genes [5]. Consequently, the structure of the gene family of MHC class I varies among species, and it is often difficult to find counterparts of a particular class I gene in close-relative species [6,7].

The entire sequence of the porcine MHC (swine leukocyte antigen; SLA) region was completely sequenced recently by using a particular haplotype named Hp-1.1 (conventionally called H01) found in a Large White individual [8–12]. Examination of the sequence demonstrated that representative class II genes such as the α and β chain genes of *DR*, *DQ*, *DM*, and *DO* are well conserved in terms of structure and location, as is the case in other species, but functional *DPA* and *DPB* genes could not be identified. The porcine class I genes are divided into two groups— classical and nonclassical—on the basis of their locations on the genome and their similarity to human or other mammalian class I genes [13]. There are six classical class I genes: *SLA-1*, *SLA-2*, *SLA-3*, *SLA-4*, *SLA-5*, and *SLA-9*. Among the proteins encoded by these loci, SLA-1, SLA-2, and SLA-3 are functional and play a role in the presentation of intracellular peptidic antigens to cytotoxic T cells. *SLA-4* and *SLA-9* are pseudogenes, and there is no evidence for the expression of *SLA-5*. An additional class I pseudogene, *SLA-11*, which has characteristics of both the classical and nonclassical class I genes, is located adjacent to the region containing the classical class I genes [14]. Sequencing of this region has demonstrated that these classical class I genes are arranged in a single contiguous region, and are ordered as *SLA-1*, *SLA-5*, *SLA-9*, *SLA-3*, *SLA-2*, *SLA-4*, and *SLA-11* from the p-terminus of the chromosome [11]. On the other hand, there are three nonclassical class I genes, *SLA-6*, *SLA-7*, and *SLA-8*, in a cluster [9].

Studies to date have indicated that there is copy number variance of classical class I genes in the porcine MHC region. In particular,

several studies have implied that the typical classical class I gene, named *SLA-1*, which is clearly defined in the Hp-1.0 haplotype (the portion of the region containing class I genes of the Hp-1.1 haplotype), has duplication or deletion on the chromosome. On the other hand, another type of classical class I gene, named *SLA-3*, has deletion in particular haplotypes [14]. However, genotyping of *SLA-1* and *SLA-3* in individuals whose haplotypes are not already known has been performed often by using cDNA, because the complexity and polymorphism of the sequences flanking the classical class I genes hinder the design of appropriate primers for polymerase chain reaction (PCR) amplification in genotyping. Therefore, the actual number of *SLA-1* or *SLA-3* loci on the genome has not been clearly demonstrated. We previously designed 40 microsatellite (MS) markers for the genotyping of haplotypes of *SLA*. The MS markers within the genomic region carrying MHC classical class I genes frequently produced more than two fragments from diploid genomic DNA [15]. This implies that in these individuals there was duplication of the genomic region containing the classical class I genes, unlike in the Hp-1.0 haplotype. Previous studies have also implied that such copy number variation between haplotypes exists in MHC class I genes of other artiodactyls, i.e., cattle and sheep. In these animals, as in pigs, different haplotypes express varying numbers of class I genes. The variation in the number of class I genes expressed in cattle and sheep makes it difficult to designate locus names [16,17].

The pig is considered a useful model animal for biomedical research—especially for transplantation research—because of the similarity of its organ size and cardiovascular system structure to those in humans [18,19]. In such studies or transplantation applications using pigs, precise knowledge of the expression of antigen-presenting SLA molecules is indispensable [20]. Therefore, we have to know the structure of the loci of the MHC genes, and their expression, in the tissues of pigs with various haplotypes. Here, we demonstrate differences among haplotypes in the structure of the genomic region containing *SLA* classical class I genes. We also demonstrate the expression of multiple loci similar to *SLA-1* on particular chromosomes.

Results

Variance in number of loci of SLA classical class I genes on different haplotypes

We had formerly designed MS markers distributed throughout the entire *SLA* region, and one MS marker named *SLAMS034* was located adjacent to a typical classical class I gene, *SLA-1* [15]. Interestingly, we found, as in our previous report, that PCR performed with *SLAMS034* in several individuals with heterozygous haplotypes generated more than two fragments, although the genomic DNA derived from an individual with the homozygous haplotype Hp-1.0, which has been entirely sequenced and extensively analyzed for its genic loci [9,11,12,21], produced only one fragment by PCR with *SLAMS034* [15]. We assumed that this increment in the number of fragments was caused by duplication of the genomic region carrying the *SLA* classical class I genes.

We had previously constructed a BAC library by using an individual of Landrace breed (L14–216) [22] that possessed two different haplotypes for the genomic region containing the *SLA* classical class I genes, inferred by genotyping with MS markers in the *SLA* region designed previously [15]. Both of the haplotypes of L14–216 were different from haplotype Hp-1.0. Furthermore, PCR with *SLAMS034* using the genomic DNA of L14–216 generated three fragments, the sizes of which were 309, 313, and 334 bp. We isolated BAC clones carrying the *SLA* classical class I genes derived from the respective haplotypes in the library, and we constructed two BAC contigs, corresponding to each haplotype (Fig. 1).

We determined the whole sequences of the BAC clones contained in the two contigs. The contig corresponding to the first haplotype,

which was designated Hp-28.0, was 541,313 bp long and ranged from the genomic region carrying *TRIM26* and other members of the *TRIM* gene family (this region is located on the side of the p-terminal flanked by *SLA-1* on swine chromosome [SSC] 7 of the Hp-1.0 sequence), to *SLA-11*, which is the class I gene distal-most from the p-terminal of SSC7 in the region containing classical class I genes. The contig was extended to the middle of *TRIM39*, adjacent to *SLA-11*. The second contig, which was derived from the other haplotype, which was designated Hp-62.0, was 254,346 bp long. It ranged from *TRIM10* to *SLA-4*. We could not isolate the BAC clones containing *SLA-11* of Hp-62.0 from the library (Fig. 1). However, both of the contigs contained the genomic region containing loci corresponding to the typical *SLA* classical class I genes *SLA-1*, *SLA-2*, and *SLA-3*. Furthermore, the structure of both ends of the determined sequence was markedly well conserved, as indicated by dot-plot analysis (Supplementary Fig. 1). A tandem highly duplicated structure was observed in the region between downstream of *SBAB-207G8.3* and upstream of *SLA-3* or its corresponding gene. The region with tandem duplication was also covered by the contig derived from the Hp-62.0 haplotype (Supplementary Fig. 1B).

The sequences derived from the Hp-28.0 and Hp-62.0 haplotypes contained 13 and 8 *SLA* classical class I genes, respectively (Fig. 1). This was strikingly different from the case in the Hp-1.0 haplotype, which possesses seven *SLA* classical class I genes. We performed sequence comparison of the observed loci with those on the Hp-1.0 haplotype (Fig. 2A). We detected five loci homologous to *SLA-1/SLA-3* on the Hp-28.0 haplotype (*A01*, *A03*, *A05*, *A08*, and *A10*) (Fig. 1, 2A, and Table 1). One of the loci corresponding to *SLA-1/SLA-3* had a 2-bp deletion on exon 5 that generated a stop codon in exon 6 (*A05*). Another locus corresponding to *SLA-1/SLA-3* had a 4-bp elongation in exon 6, resulting in a stop codon in exon 7 (*A03*). On the other hand, the Hp-62.0 haplotype had three loci (*B01*, *B04*, and *B06*) homologous to *SLA-1/SLA-3*, one of which possessed an elongation similar to *A03*, resulting in a stop codon in exon 7 (*B04*) (Figs. 1, 2A, and Table 1). Phylogenetic analysis using entire coding sequences (data not shown) or using sequences of exons 2 and 3, as in a previous report of *SLA* class I nomenclature [14] and the database for MHC polymorphisms, IPD-MHC [23], demonstrated that *A10* and *B06* were included in the clade of known *SLA-3* sequences (Figs. 2A and B); we therefore classified these loci into *SLA-3* (*A10*, *SLA-3*070102*; *B06*, *SLA-3*hb06*). Among the remaining loci similar to *SLA-1/SLA-3*, *A01*, *A08* and *B01* were grouped into a large clade containing many *SLA-1* alleles (Fig. 2B). In this study, we designated *SLA-1* for these loci similar to authentic *SLA-1* (*A01* [*SLA-1b*], *SLA-1*1501*; *A08* [*SLA-1a*], *SLA-1*0901*; *B01*, *SLA-1*1401*). On the other hand, *A03*, *A05*, and *B04* were grouped by themselves, away from the other *SLA-1* and *SLA-3* loci, although these loci shared two characteristic bases—C at position 51 and G at position 118 in the coding sequence (CDS)—that clearly discriminate known *SLA-1* alleles from *SLA-3*. These loci were designated *SLA-12* (*A03* [*SLA-12b*], *SLA-12*ha03Q*; *A05* [*SLA-12a*], *SLA-12*ha05Q*; *B04*, *SLA-12*hb04Q*). We noted that they could be clearly discriminated from the *SLA-1* loci such as *A01*, *A08*, and *B01*. For example, *A03*, *A05*, and *B04* shared TCGGGCA at positions 57 to 63 in their CDS, whereas the *SLA-1* sequences possessed CCAGGCG (Fig. 3). Among the known alleles assigned as *SLA-1*, *SLA-1*es11* (**EU440342**), *SLA-1*w10sm21* (**AY135589**) and *SLA-1*w11yn01* (**AY102469**), as well as *B01* (*SLA-1*1401*, identical to *SLA-1*es12* [**EU440343**]), were not grouped with other *SLA-1* sequences, like *A03*, *A05*, and *B04* (Fig. 2B). However, these loci, unlike the majority of the *SLA-1* alleles in the phylogenetic tree, did not have the 7-bp TCGGGCA sequence characteristic of *SLA-12* (Fig. 3). Several sequences with the same characteristics as newly found *SLA-12* sequences have been found in the public database for expressed sequence tags (ESTs), dbEST, in GenBank [24]. These sequences belonged to the clade for *SLA-12* loci in the phylogenetic analysis (Fig. 2B). The TCGGGCA sequences were quite rare in the *SLA-1*-like sequences observed in the public databases. Nevertheless, two
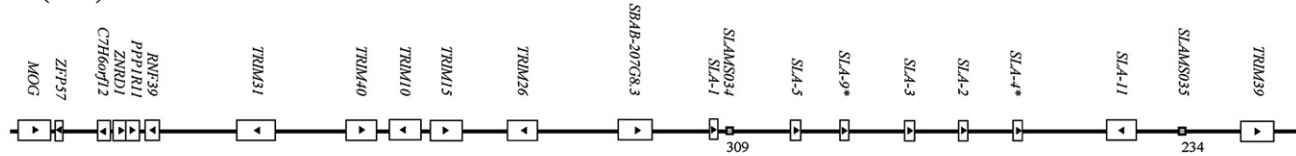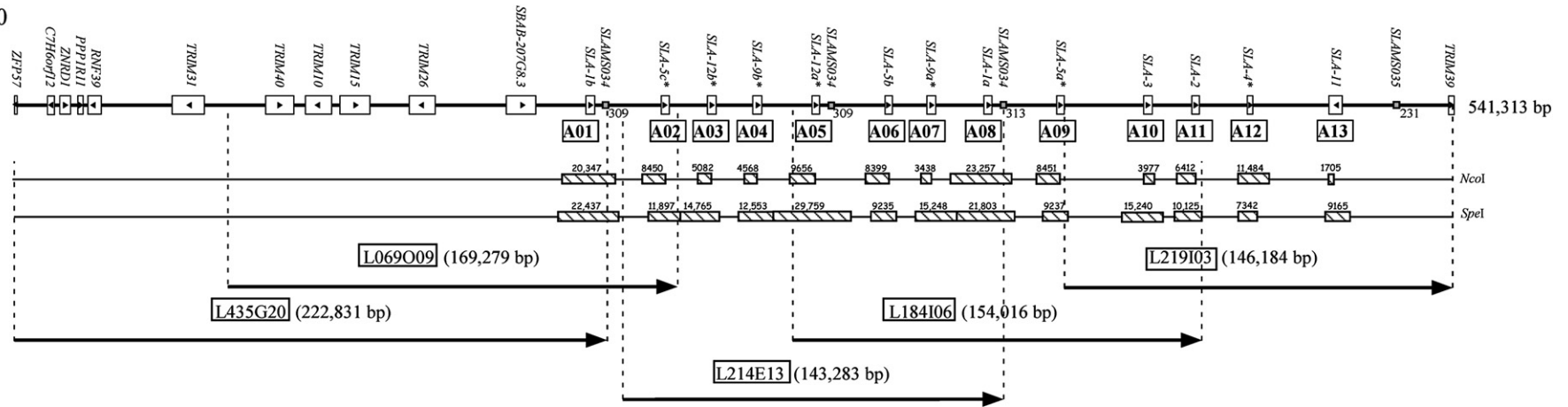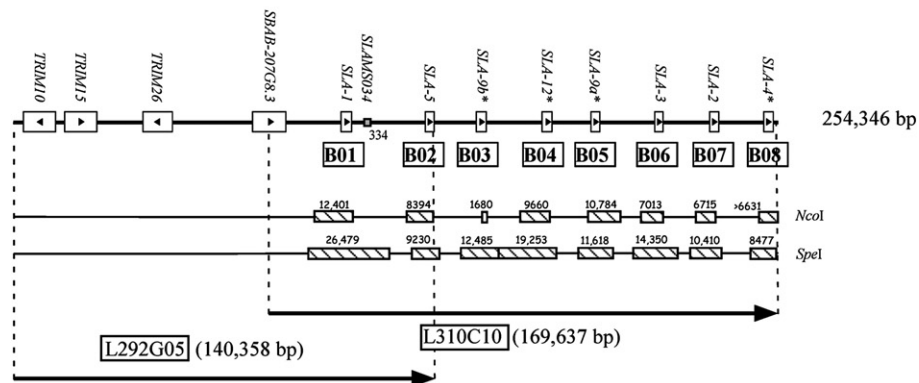
Fig. 1. BAC contigs derived from two haplotypes possessed by a Landrace individual. The haplotypes of the individual differed from haplotype Hp-1.0, which had been sequenced completely in previous studies [9–12]. Genes on the genomic sequences are indicated by rectangles with arrows showing their orientation, according to a previous study [10]. Loci with stop codon(s) in the region corresponding to CDSs or loci regarded as pseudogenes are asterisked. Detected genes corresponding to *SLA* classical class I genes are demonstrated as *A01* to *A13* on the Hp-28.0 haplotype and *B01* to *B08* on the Hp-62.0 haplotype, with the corresponding class I genes named on the Hp-1.0 haplotype. *SLA-1*-like loci with different characteristic bases, newly found in this study, were designated as *SLA-12*. The duplicated loci were followed by sequential alphabets (a, b and c) from the most centromeric locus. MS markers (*SLAMS034* and *SLAMS035*) from previous reports [15] are indicated by gray rectangles on the genomic sequences, along with the lengths of the PCR amplicons for the respective loci. BAC clones consisting of the contigs for the respective haplotypes are indicated by the large arrows, which are directed as registered in the DDBJ/EMBL/GenBank databases (L435G20: **AP009553**; L069O09: **AP009554**; L214E13: **AP009555**; L184I06: **AP009556**; L219I03: **AP009557**; L292G05: **AP009558**; and L310C10: **AP009559**) with lengths in base pairs in parentheses. Locations of the fragments detected by Southern hybridization with the universal *SLA* classical class I probe (Supplementary Fig. 2 and Supplementary Table 1) are shown under the respective genomic sequences by hatched rectangles, with lengths in base pairs.

sequences with high similarity with those of *SLA-12* in the ESTs derived from backfat tissue of commercial crossbred pigs (Yorkshire× Landrace) (**DT320911** and **DT324475**) possessed TCGGGCA at positions 57 to 63 in their putative CDSs. The EST **DT324475** was exactly matched to the first 728 bases from the putative start codon of *A05*. Both **DT320911** and **DT324475** had sequences in exons 2 and 3 identical with those of *A05*. Moreover, we found additional sequences possessing TCGGGCA in our cDNA libraries derived from the intestine (**BW976872**) and uterus (**BP169723**) of crossbred pigs ((Landrace× Large White)×Duroc) [25]. These four sequences shared other characteristic bases such as CAC at positions 549 to 551 in the putative CDSs with *A03*, *A05*, and *B04*, although several *SLA-1* alleles (*SLA-1*es11*, *SLA-1*es12* and *SLA-1*st11*) possess these bases (Fig. 3). These observations demonstrated that the *SLA-12* loci are possessed by particular haplotypes and are transcribed, although the transcription level may be low. In addition, *SLA-12* had a premature stop codon in the region encoding their cytoplasmic tail; therefore, these loci might be nonfunctional.

The loci corresponding to *SLA-2* possess high similarity to *SLA-1* and *SLA-3* in terms of their exon 2 and 3 sequences, but *SLA-1* and *SLA-3* can be obviously discriminated from *SLA-2* because *SLA-2* sequences, unlike *SLA-1* and *SLA-3* ones, commonly have a nine-base elongation at the head of the CDS [14]. One locus corresponding to *SLA-2* was found in each haplotype in this study (*A11* in Hp-28.0, *SLA-2*0503*; *B07* in Hp-62.0, *SLA-2*0602*), and the location of the locus was conserved among haplotypes, including Hp-1.0 (Fig. 1).

The loci corresponding to *SLA-5*, which possess high similarity in their sequence and structure to *SLA-1* and *SLA-3* but were clearly

separated from other *SLA* classical class I genes, including *SLA-1* and *SLA-3*, in the phylogenetic tree (Figs. 2A and B), were also multiplied as three loci (*A02* [*SLA-5*ha02Q*], *A06* [*SLA-5*ha06Q*], and *A09* [*SLA-5*ha09Q*]) in the Hp-28.0 haplotype (Fig. 1). On the other hand, there was only one locus (*B02* [*SLA-5*0101Q*]) for *SLA-5* in the Hp-62.0 haplotype; similarly, there is only one locus in Hp-1.0. Among the *SLA-5* loci, *A06* and *B02* were identical and almost identical (1-base substitution), respectively, to *SLA-5* in the Hp-1.0 haplotype. The exon 2 and 3 sequences of *B02* exactly matched those of *SLA-5* in Hp-1.0 (Fig. 2B).

Two loci corresponding to *SLA-9* were observed in the respective haplotypes (*A04* and *A07* for Hp-28.0; *B03* and *B05* for Hp-62.0) (Fig. 1). The loci proximal to the centromere, *A07* and *B05*, were more similar to *SLA-9* in Hp-1.0 than were *A04* and *B03* (Fig. 2C). However, *B03* and *B05* shared a 4-bp deletion in exon 2, which was not found in *A04*, *A07*, or *SLA-9* in Hp-1.0 (data not shown). Among the loci similar to *SLA-9*, only *A07* did not possess any deletions in comparison with *SLA-9* in Hp-1.0.

The loci corresponding to *SLA-4* (*A12* and *B08*), which were clearly demonstrated by phylogenetic analysis (Fig. 2A), were conserved in terms of position and number among haplotypes such as *SLA-2* and *SLA-3* (Fig. 1).

A locus corresponding to *SLA-11* was observed only in the contig of the Hp-28.0 haplotype (*A13*) (Fig. 1). Southern blotting with a probe that universally recognized *SLA* classical class I genes demonstrated that the restriction fragments generated from the genome of individual L14–216 corresponded to the 21 classical class I genes observed in the two contigs. However, one additional locus was
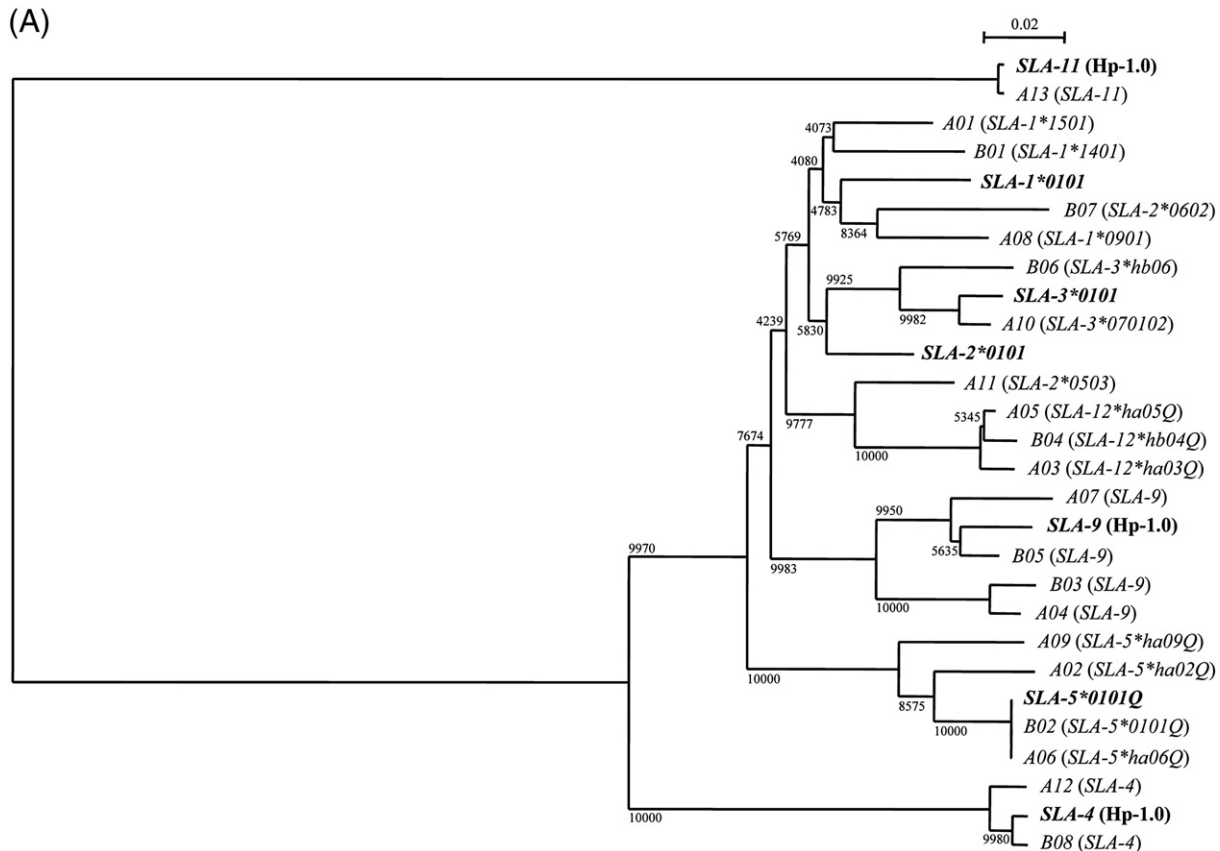
(A)

Fig. 2. (A) Phylogenetic tree of the *SLA* classical class I genes in the Hp-1.0, Hp-28.0, and Hp-62.0 haplotypes. The tree was constructed with sequences corresponding to exons 2 and 3 of the genic sequences. The loci found in Hp-1.0 are shown in bold letters. *SLA-11* of Hp-1.0 and the loci corresponding to *SLA-11* in Hp-28.0 (*A13*) were used as outgroups. (B, C) Phylogenetic trees of *SLA-1*, *SLA-3*, and *SLA-5* (B) and *SLA-9* (C) with their corresponding loci observed in the Hp-28.0 and Hp-62.0 haplotypes. The alleles of *SLA-1*, *SLA-3*, *SLA-5* and *SLA-9* were retrieved from IPD-MHC for *Sus scrofa* (http://www.ebi.ac.uk/ipd/mhc/download.html) [23]. The trees were constructed with putative exons 2 and 3 estimated by sequence similarity to corresponding loci. The putative exon 2 and 3 sequences of the four ESTs similar to *A03*, *A05*, and *B04* (*SLA-12*) are shown in square brackets. The exon 2 and 3 sequence of human *HLA-A*02010101* (**EU445471**) was used as an outgroup. Bootstrap values for 10,000 replicates are indicated beside the branches. The clades consisting of sequences clearly discriminated from *SLA-1* sequences in (B) are indicated as *SLA-3*, *SLA-5* and *SLA-12*.
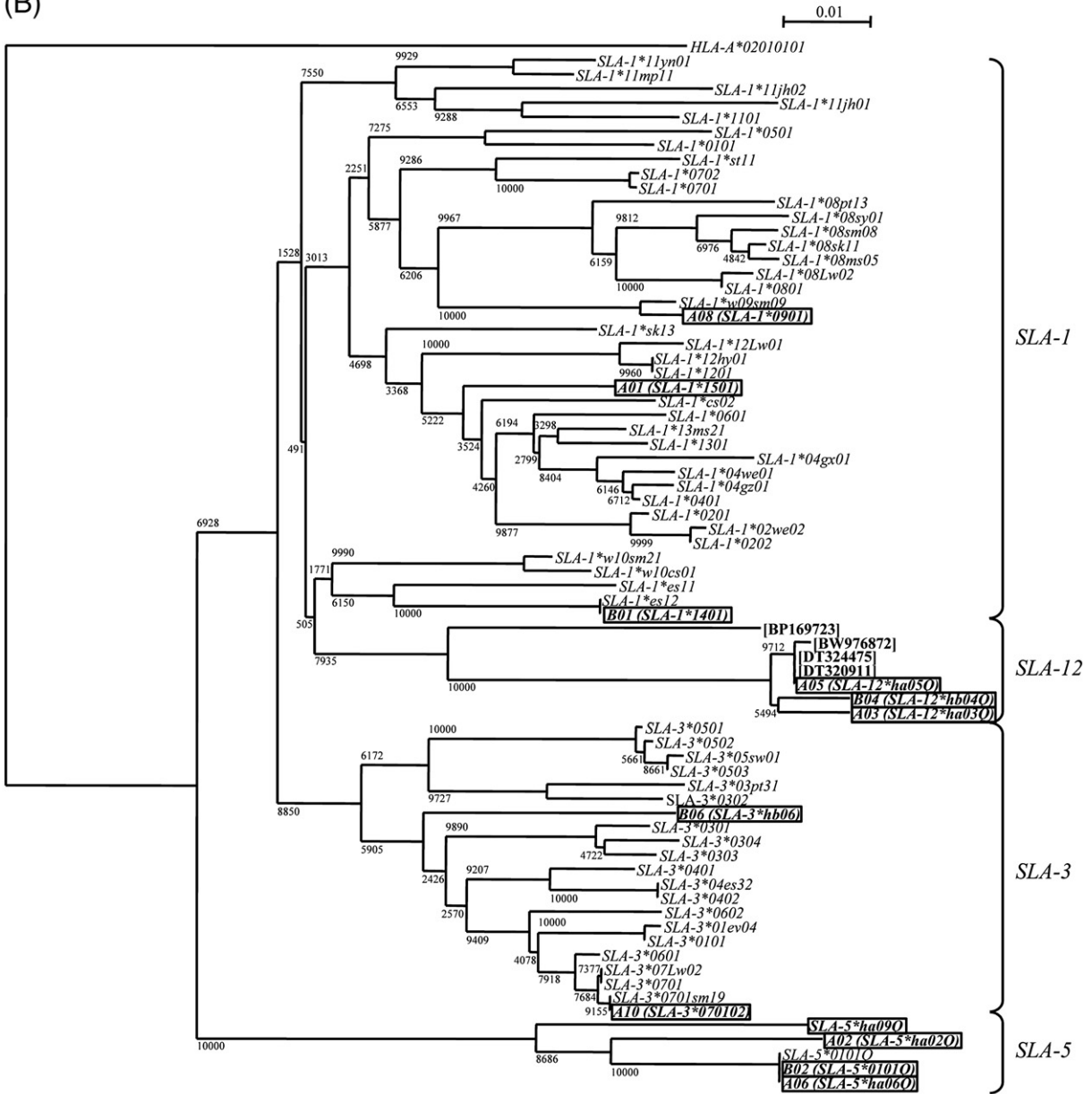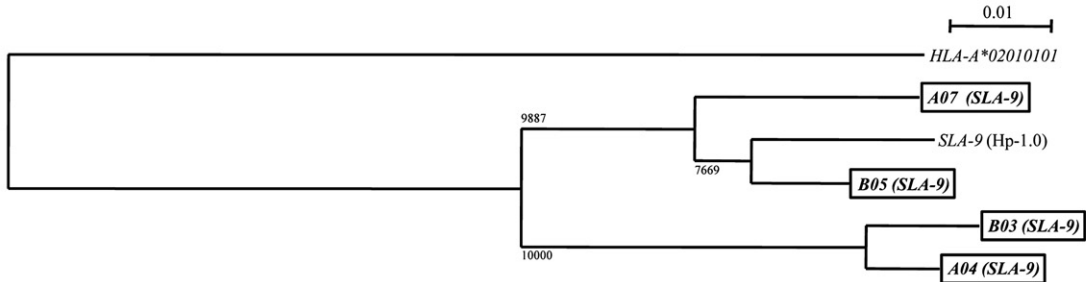
Fig. 2 (*continued*).

inferred by the Southern blot; it might have corresponded to *SLA-11* on the Hp-62.0 haplotype, which had not been cloned (Supplementary Fig. 2). The sizes of the fragments with the regions containing the classical class I genes generated from the BAC clones in the contigs were inferred from the sequences obtained (Supplementary Table 1) and were compared with the observed fragments in the Southern blotting (Supplementary Fig. 2). The results demonstrated that the sequences of the contigs were correctly assembled. In total, 25 genes were observed in the Hp-28.0 haplotype and 12 in the Hp-62.0 haplotype (Fig. 1 and Table 1).

Table 1
Genes detected in the genomic regions carrying *SLA* classical class I genes in the respective haplotypes

Haplotype[a]	Locus	Allele designation[a]	Description[b]	Orientation	Pseudogenes or genes with stop codon(s)[c]	Start	End	Length of putative CDS (bp)[d]
Hp-28.0	ZFP57		Zinc finger protein 57 homolog (mouse)	−		239	166	>74
	C7H6orf12		Chromosome 7 open reading frame, human c6orf12	−		14,543	11,880	1065
	ZNRD1		Zinc ribbon domain-containing 1	+		16,836	21,052	372
	PPP1R11		Protein phosphatase 1, regulatory (inhibitor) subunit 11	+		23,795	25,656	381
	RNF39		Ring finger protein 39	−		32,155	27,455	1059
	TRIM31		Tripartite motif-containing 31	−		71,438	58,801	1731
	TRIM40		Tripartite motif-containing 40	+		94,376	104,704	936
	TRIM10		Tripartite motif-containing 10	−		119,289	109,018	1449
	TRIM15		Tripartite motif-containing 15	+		122,319	133,234	1386
	TRIM26		Tripartite motif-containing 26	−		158,260	148,402	1638
	SBAB-207G8.3		Novel protein similar to tripartite motif-containing 26 (TRIM26)	+		184,895	195,968	1632
	SLA-1b (A01)	SLA-1*1501	SLA-1 histocompatibility antigen, class I (duplicated)	+		214,954	218,031	1086
	SLA-5c (A02)	SLA-5*ha02Q	Similar to SLA-5 histocompatibility antigen, class I	+	*	243,006	246,093	1085
	SLA-12b (A03)	SLA-12*ha03Q	SLA-12 histocompatibility antigen, class I (newly designated, duplicated)	+	*	260,451	263,315	1041
	SLA-9b (A04)		Similar to SLA-9 histocompatibility antigen, class I	+	*	277,640	280,722	1089
	SLA-12a (A05)	SLA-12*ha05Q	SLA-12 histocompatibility antigen, class I (newly designated)	+	*	299,642	302,720	999
	SLA-5b (A06)	SLA-5*ha06Q	SLA-5 histocompatibility antigen, class I	+		327,044	330,124	1086
	SLA-9a (A07)		SLA-9 histocompatibility antigen, class I	+	*	343,337	346,404	1086
	SLA-1a (A08)	SLA-1*0901	SLA-1 histocompatibility antigen, class I	+		364,201	367,283	1086
	SLA-5a (A09)	SLA-5*ha09Q	Similar to SLA-5 histocompatibility antigen, class I	+	*	391,657	394,733	1082
	SLA-3 (A10)	SLA-3*070102	SLA-3 histocompatibility antigen, class I	+		424,722	427,790	1095
	SLA-2 (A11)	SLA-2*0503	SLA-2 histocompatibility antigen, class I	+		442,599	445,520	1095
	SLA-4 (A12)		SLA-4 histocompatibility antigen, class I	+	*	<463,262	>465,290	>842
	SLA-11 (A13)		SLA-11 histocompatibility antigen, class I	−		>499,728	494,380	>1049
	TRIM39		Tripartite motif-containing 39	+		539,128	>540,961	>1009
Hp-62.0	TRIM10		Tripartite motif-containing 10	−		13,823	3517	1449
	TRIM15		Tripartite motif-containing 15	+		16,853	27,770	1386
	TRIM26		Tripartite motif-containing 26	−		52,821	42,941	1638
	SBAB-207G8.3		Novel protein similar to tripartite motif-containing 26 (TRIM26)	+		79,450	90,583	1632
	SLA-1 (B01)	SLA-1*1401	SLA-1 histocompatibility antigen, class I	+		109,426	112,523	1086
	SLA-5 (B02)	SLA-5*0101Q	SLA-5 histocompatibility antigen, class I	+		136,886	139,966	1086
	SLA-9b (B03)		Similar to SLA-9 histocompatibility antigen, class I	+	*	154,004	157,083	1081
	SLA-12 (B04)	SLA-12*hb04Q	SLA-12 histocompatibility antigen, class I (newly designated)	+	*	175,907	178,976	1090
	SLA-9a (B05)		Similar to SLA-9 histocompatibility antigen, class I	+	*	192,110	195,175	1082
	SLA-3 (B06)	SLA-3*hb06	SLA-3 histocompatibility antigen, class I	+		213,380	216,439	1086
	SLA-2 (B07)	SLA-2*0602	SLA-2 histocompatibility antigen, class I	+		231,386	234,449	1095
	SLA-4 (B08)		SLA-4 histocompatibility antigen, class I	+	*	<249,807	>251,934	>577

Locations of the loci are demonstrated by the distance from the p-terminus of the respective contigs.

[a] Haplotype and allele names were designated by the ISAG SLA Nomenclature Committee [23].

[b] Descriptions of the respective loci are in accordance with the annotation of the sequence of the Hp-1.0 haplotype (**AJ251829** and **AJ131112**), and Vertebrate Genome Annotation (VEGA) database for *Sus scrofa* (http://vega.sanger.ac.uk/Sus_scrofa/) [41], considering the AG–GT rule at splicing.

[c] Loci without any obvious start codons and loci possessing stop codon(s) within their CDSs, as estimated by comparison with the genic sequences on the Hp-1.0 haplotype, are indicated by asterisks.

[d] Length of putative CDSs, or sequences corresponding to CDSs of paralogous genes are shown.

Characteristic sequences observed upstream of SLA classical class I genes

We examined characteristic repetitive sequences such as DNA transposons and MS repeats within the flanking sequences of *SLA* classical class I loci in the Hp-1.0, Hp-28.0 and Hp-62.0 haplotypes (Fig. 4). In the Hp-28.0 haplotype, all four *SLA-1*-like (*SLA-1* and *SLA-12*) loci were flanked by Tigger1a [26] and Charlie1a [27] on their p-terminal sides; these are fossils of a DNA transposon observed in the upstream of the *SLA-1* and *SLA-3* loci in the Hp-1.0 haplotype [21]. The Tigger1a sequence upstream of the *A10* (*SLA-3*) locus might have collapsed during the process of establishment of this genomic region. In the Hp-62.0 haplotype, both the *SLA-1*-like (*B01* [*SLA-1*] and *B04* [*SLA-12*]) and *SLA-3* (*B06*) loci had Tigger1a on their p-terminal flanking sequences. A 5S rRNA pseudogene was observed upstream of all *SLA-1*, *SLA-3* and *SLA-12* loci in the respective haplotypes (Fig. 4). Like the *SLA-1*, *SLA-3*,

and *SLA-12* loci, all *SLA-5* loci possessed Charlie1a on their p-terminal side. Charlie1a was also observed upstream of the *SLA-2* loci on both haplotypes that we studied (data not shown). In each region between the *SLA-1*/*SLA-12* and *SLA-5* loci, MS repeats (TAAAA)n and (CA)n were observed. These (CA)n repeats correspond to the *SLAMS034* marker, which is found adjacent to *SLA-1* in the Hp-1.0 haplotype [15]. There were three sequences corresponding to *SLAMS034* in the Hp-28.0 haplotype and one corresponding sequence in the Hp-62.0 haplotype (Fig. 4). The flanking sequences, corresponding to the primers for *SLAMS034*, of the respective repeats were well conserved and completely identical, with the exception of one sequence found in the region adjacent to *A01* (Supplementary Fig. 3). Estimated amplicons derived from the three sequences encompassing the (CA)n repeat with exactly matched sites for primer annealing were consistent with the observed amplicons, which were 309, 313, and 334 bp.

Fig. 3. Comparison of the alleles of *SLA-1* (A) and *SLA-3* (B) sequences that have been so far determined (cited from IPD-MHC; http://www.ebi.ac.uk/ipd/mhc/download.html) [23] with the loci assumed to be *SLA-1* and *SLA-12* (A) and *SLA-3* loci (B) in the haplotypes we sequenced. Porcine ESTs (**BP169723**, **BW976872**, **DT320911**, and **DT324475**) that share sequence characteristics with *A03*, *A05*, and *B04* (*SLA-12*) are incorporated into the figure for *SLA-1*/*SLA-12* (A). The entire sequences of exon 1 and partial sequences of exons 2 and 3 are shown; the characteristic bases of *SLA-1* and *SLA-3* demonstrated in a former study (positions 9, 10, 51, 118, 559, and 560 in the coding sequence) [14] are indicated by shading.

(A) SLA-1/SLA-12

exon 1   exon 2 (partial)   exon 3 (partial)

Row labels:

- SLA-1*0101
- Hp-28.0 A01 (SLA-1*1501)
- A03 (SLA-12*ha03Q)
- A05 (SLA-12*ha05Q)
- A08 (SLA-1*0901)
- Hp-62.0 B01 (SLA-1*1401)
- B04 (SLA-12*hb04Q)
- SLA-1*0201
- SLA-1*0202
- SLA-1*02we02
- SLA-1*0401
- SLA-1*04gx01
- SLA-1*04gz01
- SLA-1*04we01
- SLA-1*0601
- SLA-1*0701
- SLA-1*0702
- SLA-1*0801
- SLA-1*08Lw02
- SLA-1*08ms05
- SLA-1*08pt13
- SLA-1*08sk11
- SLA-1*08sm08
- SLA-1*08sy01
- SLA-1*1101
- SLA-1*11jh01
- SLA-1*11jh02
- SLA-1*11mp11
- SLA-1*11yn01
- SLA-1*1201
- SLA-1*12Lw01
- SLA-1*12hy01
- SLA-1*1301
- SLA-1*13ms21
- SLA-1*cs02
- SLA-1*es11
- SLA-1*es12
- SLA-1*sk13
- SLA-1*st11
- SLA-1*w09sm09
- SLA-1*w10cs01
- BP169723
- BW97872
- DT320911
- DT324475

(B) SLA-3

exon 1   exon 2 (partial)   exon 3 (partial)

Row labels:

- SLA-3*0101
- A10
- B06
- SLA-3*01ev04
- SLA-3*0301
- SLA-3*0302
- SLA-3*0303
- SLA-3*0304
- SLA-3*03pt31
- SLA-3*0401
- SLA-3*0402
- SLA-3*04es32
- SLA-3*0501
- SLA-3*0502
- SLA-3*0503
- SLA-3*05sw01
- SLA-3*0601
- SLA-3*0602
- SLA-3*0701
- SLA-3*0701sm19
- SLA-3*07Lw02

SLA-9 loci were flanked by characteristic MS repeats $(GGGGA)_n$ and $(TCCCC)_n$, and long interspersed nucleotide element (LINE)/L1 sequences, L1M4 and L1MC3, on the p-terminal side. Only *A04* had no traces of the MS repeat $(GGGGA)_n$, probably because the MS repeat was collapsed by deletion of the genomic region encompassing a locus corresponding to *SLA-5* (Fig. 4). Apart from the repetitive sequences generally found in other regions on the pig genome, a 0.9- to 1.4-kb sequence is commonly observed 4 to 11 kb upstream of the *SLA* classical class I genes; it has been designated *SLA* classical class I-associated repeat element (SCRE) [21]. The SCRE was also commonly observed at the *SLA* classical class I loci in the haplotypes sequenced here, in their upstreams. The most conserved region of approximately 240 bp (designated the SCRE core sequence here), which was located 6 to 7 kb upstream of *SLA-1* and *SLA-3* and 12 kb upstream of *SLA-5*, was detected in all *SLA-1*, *SLA-3*, *SLA-5*, and *SLA-12* loci in the Hp-1.0, Hp-28.0, and Hp-62.0 haplotypes, with the exception of the *A03* locus (*SLA-12*) (Fig. 4). At 9 kb upstream of each SCRE of the *SLA-5* locus, we detected an additional DNA transposon named MER5A [28] (Fig. 4).

Duplication of SLA-1-like loci and flanking regions containing SLA classical class I genes

Two (*A01* and *A08*) and one (*B04*) loci designated *SLA-1* were observed in haplotypes Hp-28.0 and Hp-62.0, respectively. Furthermore, there were additional two (*A03* and *A05*) and one (*B04*) loci sharing the characteristics of *SLA-1* in the respective haplotypes, which were designated *SLA-12*. In the Hp-1.0 haplotype, there is only one locus for *SLA-1*. Similar amplification of loci was observed in *SLA-5* and *SLA-9*. These *SLA-1*, *SLA-5*, *SLA-9* and *SLA-12* loci observed in the Hp-28.0 and Hp-62.0 haplotypes may have been generated from the respective common ancestral loci. Comparison of structural similarity among the genomic regions containing the *SLA* classical class I loci in the two haplotypes studied here and in the Hp-1.0 haplotype suggests that more than one duplication event in the Hp-28.0 haplotype, and at least one such event in Hp-62.0, have occurred, if we assume that Hp-1.0 was the ancestral haplotype (Supplementary Fig. 1).

The genic sequences of *SLA-1*-like (*SLA-1* and *SLA-12*) and *SLA-3* were highly similar to each other. However, the phylogenetic analysis demonstrated that *A10* in Hp-28.0 and *B06* in Hp-62.0, the loci similar to *SLA-3* and most distal from the p-terminal in their respective haplotypes, were clearly separated from the *SLA-1*-like loci (*SLA-1* and *SLA-12*) (Figs. 2A and B). Therefore, the *SLA-1* and *SLA-12* loci in the Hp-28.0 or Hp-62.0 haplotype that are not observed in the Hp-1.0 haplotype might have been generated from the ancestral *SLA-1* locus. The loci that we regarded as having been generated by duplication of *SLA-1* on the Hp-28.0 and Hp-62.0 haplotypes were obviously discriminated from *SLA-3* with nucleotide positions 51 and 118, as described above (Fig. 3).

Consideration of the characteristic sequences flanking the *SLA* classical class I genes revealed several genome blocks conserved among the haplotypes. Comparison of the sequence of the Hp-1.0 haplotype with those of Hp-28.0 and Hp-62.0 demonstrated that the approximately 65-kb unit in the Hp-1.0 haplotype containing *SLA-1*, *SLA-5*, and *SLA-9* appeared four times in Hp-28.0 and twice in Hp-62.0. Furthermore, it suggested that several deletions of genomic regions have occurred along with these duplications (Fig. 4).

We assigned the blocks to make the duplication and deletion process clear. The region encompassing two characteristic repetitive sequences, 5S rRNA and Tigger1a, was designated block A. The regions encompassing *SLA-1*-like (*SLA-1* and *SLA-12*), *SLA-5*, *SLA-9*, and *SLA-3*

loci and their flanking characteristic sequences were designated B, C, D, and E, respectively. There were four repetitive units in the Hp-28.0 haplotype for the sequence corresponding to blocks B, C, D, and A in the Hp-1.0 haplotype (Fig. 4), although several regions were deleted in the repetitive units. The deletion of block D in the first repetitive unit (designated Hp-28.0-I in Fig. 4) corresponded to the region carrying an *SLA-9* locus and its upstream sequence. The second repetitive unit (Hp-28.0-II) lacked a large sequence, block C and the p-terminal side of block D, ranging from the downstream of *A03* (*SLA-12*) to the $(GGGGA)_n$ MS repeat located downstream of the absent *SLA-5* locus. The second unit also lacked the SCRE upstream of *A03* in block B. The most proximal unit to the centromere (Hp-28.0-IV) had a small deletion including a locus corresponding to *SLA-9* (block D). In the Hp-62.0 haplotype, a large duplication of the sequence covering blocks B to A was observed; it also lacked block C, including a locus corresponding to *SLA-5* and its upstream region in the unit proximal to the centromere (Hp-62.0-II). These observations show that active duplication and deletion processes have occurred in the region upstream of *SLA-3* in the respective *SLA* haplotypes.

Phylogenetic analysis using the characteristic sequences observed around the assumed duplicated units also demonstrated active duplication and deletion processes. Examination of the phylogenetic tree constructed with the SCRE core sequences flanked by *SLA-1*, *SLA-3*, *SLA-5*, and *SLA-12* loci indicated that the SCRE sequences adjacent to *SLA-3* were distant from the SCRE sequences near *SLA-1*, *SLA-5*, and *SLA-12* loci (Fig. 5A). The Tigger1a and 5S rRNA sequences adjacent to *SLA-1* loci proximal to the p-terminus were divergent from the others located in the remaining region covering *SLA-1*, *SLA-3*, and *SLA-12* loci (Figs. 5B and C). These findings implied that the boundaries of the duplication unit are located upstream of SCRE of *SLA-1*/*SLA-12* and *SLA-3*, and that the unit corresponded to B–C–D–A in the Hp-1.0 haplotype (Fig. 4).

Transcription of SLA classical class I genes within different haplotypes

The precise genomic structures of the various *SLA* haplotypes have not been elucidated. Therefore, genotyping of the *SLA* classical class I genes has often been performed by using cDNA because we do not have appropriate genotyping primers for various alleles of the *SLA* classical class I genes for genomic DNA. In genotyping using cDNA, more than one locus for *SLA-1* is often observed in particular haplotypes, although in several haplotypes no *SLA-1* or *SLA-3* loci are found [14]. The sequences of the two haplotypes examined here possessed four and two *SLA-1*-like (*SLA-1*/*SLA-12*) loci, respectively, on the genome. We designed PCR primers for amplicons spanning intervening exon boundaries with cDNA and thus examined the transcription of *SLA* classical class I loci on the genomes of the Hp-28.0 and Hp-62.0 haplotypes (Table 2). In *B05*, corresponding to *SLA-9*, primers generating amplicons that could be discriminated from other *SLA-9* loci could not be designed. All of the amplicons generated by the primers were subjected to sequencing, and the amplicons were confirmed to be correct for the targeted loci.

The results of reverse transcription-polymerase chain reaction (RT-PCR) analysis are shown in Table 2. In the Hp-28.0 haplotype all four *SLA-1*(-like) loci (*A01*, *A03*, *A05*, and *A08*) were transcribed. *SLA-3* (*A10*) was also transcribed in the Hp-28.0 haplotype. In the Hp-62.0 haplotype, both *SLA-1*(-like) loci (*B01* and *B04*) and a locus for *SLA-3* (*B06*) were transcribed. We also confirmed the transcription of *SLA-2* (*A11* and *B07*) on both haplotypes (Table 2). Interestingly, transcription from all of the tested *SLA-9* loci in both haplotypes (*A04*, *A07*, and *B03*) was detected, although these loci had stop codon(s). The

---

Fig. 4. Structural comparison around the duplicated regions containing loci corresponding to *SLA-1*, *SLA-5*, and *SLA-9* among haplotypes. *SLA* classical class I loci, as well as characteristic repetitive sequences and MS repeats adjacent to the class I loci, are indicated in the three haplotypes. The regions encompassing two characteristic repetitive sequences, 5S rRNA and Tigger1a, were designated block A. The other regions encompassing loci and their characteristic flanking sequences were designated as follows: B (*SLA-1*/*SLA-12*), C (*SLA-5*), D (*SLA-9*), and E (*SLA-3*). Loci that might have been deleted from the putative ancestral genomic sequence are asterisked. Repetitive units in the Hp-28.0 and Hp-62.0 haplotypes are designated Hp-28.0-I to -IV and Hp-62.0-I to -II, respectively, at the right side of the figure.
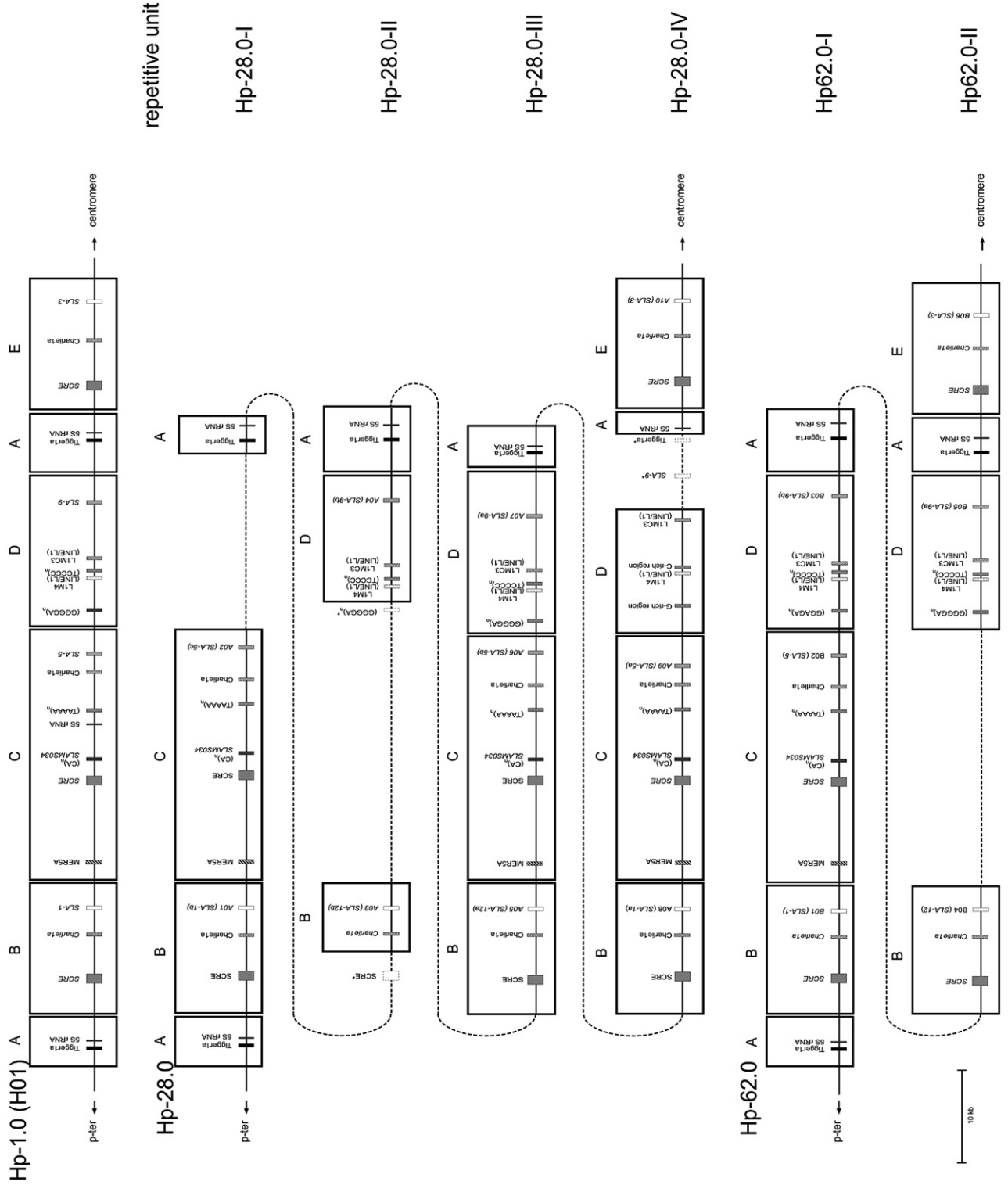
transcription levels of the respective loci could not be determined by RT-PCR, because the extreme similarity of the sequences of *SLA-1*(-like) and *SLA-9* loci hindered the specificity of amplification of correct fragments corresponding to the respective loci. In particular, we did

not succeed in amplifying *A03*, *A05*, and *A08* as single fragments. Because the fragments obtained were mingled with each other, we have to confirm the amplicons by cloning of the fragments (data not shown). This implies that there may be a relatively small amount of
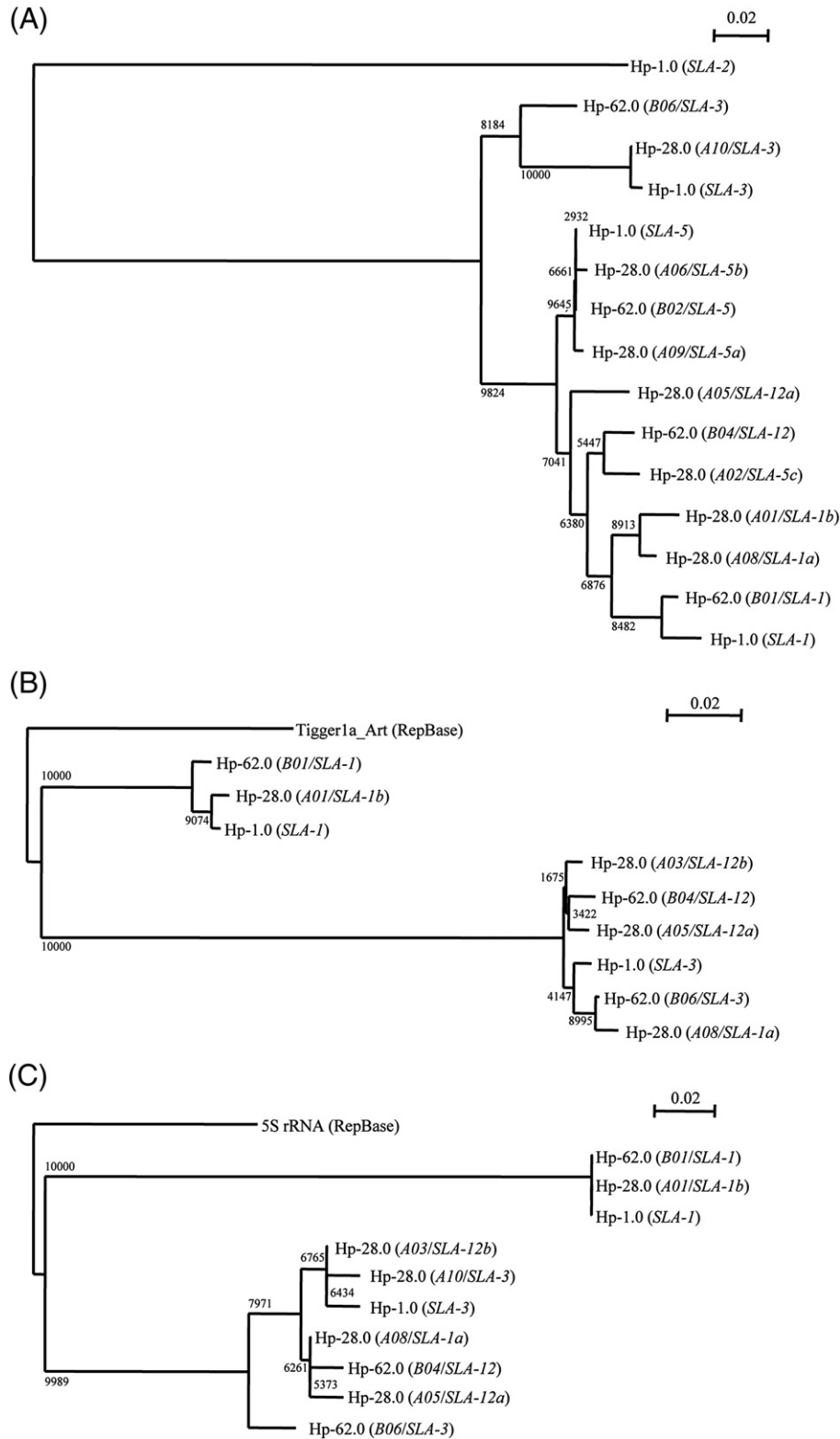

Fig. 5. Phylogenetic trees of characteristic sequences adjacent to the *SLA* classical class I genes. Bootstrap values for 10,000 replicates are indicated beside the branches. (A) Tree of SCRE core sequences located in the upstream regions of the *SLA-1*, *SLA-5*, and *SLA-3* loci in the Hp-1.0 haplotype and the loci corresponding to these *SLA* genes in the Hp-28.0 and Hp-62.0 haplotypes. The sequence corresponding to the SCRE core sequence in the upstream region of *SLA-2* in the Hp-1.0 haplotype (67,672 to 67,910 bp in **AJ131112**) was used as an outgroup. (B and C) Trees of Tigger1a (B) and 5S rRNA (C) sequences located in the upstream regions of *SLA-1* and *SLA-3* in Hp-1.0 and their corresponding loci in Hp-28.0 and Hp-62.0. The corresponding sequences observed in the RepBase database [27] were used as outgroups.

Table 2
Expression of *SLA-1*, *SLA-2*, *SLA-3*, *SLA-9*, and *SLA-12* loci located on the genomic sequences of the Hp-28.0 and Hp-62.0 haplotypes, as determined by RT-PCR

Haplotype	Locus	Forward primer (5′→3′)	Reverse primer (5′→3′)	Boundary of intervening exons	Detected amplicon			
					Genomic DNA	Size (bp)	cDNA	Size (bp)
Hp-28.0	A01 (SLA-1b)	AAAGTCAAGGACAACGCACA	GTCCTGTCTGTACCCGTGGA	exon 2/3	+	378	+	153
	A03 (SLA-12b)	GTCCCCACTCCCTGAGCTATTTCTA	TCCCTCCGCCTCATTGGCA	exon 2/3	+	690	+	464
	A04 (SLA-9b)	GTCAAGGAAGAGGCACAGATTTAC	CTTCCCCATCTCCAGGTATTTC	exon 2/3	ND[c)]	521	+	330
	A05 (SLA-12a)	GGTCTCACACCTACCAGTGGATGTAC	ATGTCCTGGCTCTGGTCCTGGCCCTCG	exon 3/4	+	1020	+	412
	A07 (SLA-9a)	GTCGGCTACGTGAACGACA	TAGCTCCTCTCCTGCTCTGC	exon 2/3	+	594	+	404
	A08 (SLA-1a)	TTCGACAGCGACGCCCCGAATCCGA	CCTTCCCCATCTCCAGGTATTC	exon 2/3	+	649	+	424
	A10 (SLA-3)	GTTTCATCTCCGTCGGCTAC	AGCCGTGTTATTCCCCATCTG	exon 2/3	+	709	+	484
	A11 (SLA-2)	GTGGATAGAGAAGGAGGGGC	TCCAGGTATTTCTGGAGCGA	exon 2/3	+	595	+	369
Hp-62.0	B01 (SLA-1)	GAATGTCATGGACAACGCAC	GTAGGCGTCCTGATCGTACC	exon 2/3	ND[c)]	387	+	160
	B03 (SLA-9b)	GCGCTCTCGGCTACTACAAT	GGTAACTTCTCATGCGCTCC	exon 3/4	ND[c)]	432	+	240
	B04 (SLA-12)	CGATTACGTCGCCCTGAACCAGGACCA	TCCATGTCCTGGCTCTGGTCCTGC	exon 3/4	ND[c)]	918	+	324
	B05 (SLA-9a)			ND[a)]	ND[c)]		ND[b)]	
	B06 (SLA-3)	ATACAGCAGAAGGGGCAGAA	GAGGAAGAATCCGTCTGGC	exon 2/3	ND[c)]	403	+	177
	B07 (SLA-2)	AACAGGAGGGGCAGGACTAT	AGGCGAACTGTATGTACCCG	exon 2/3	ND[c)]	422	+	195

Each locus is demonstrated with the corresponding locus on the Hp-1.0 haplotype in parentheses. Primers were designed to amplify fragments in which exon boundaries intervened, and the same primers were used for amplification with genomic DNA to confirm whether they worked correctly.
[a)] Primers could not be designed because the sequence of *B05* loci was too similar to those of the other *SLA-9* loci to discriminate.
[b)] Not done because the corresponding primers could not be prepared.
[c)] The detected fragments consisted of plural amplicons, and it could not be determined whether the correct amplification had been performed from cloning and sequencing of the amplicons.

transcription of *A03*, *A05*, and *A08* in comparison with that of *A01* or *B01*.

We have not succeeded in amplifying PCR fragments derived from cDNA for *SLA-5* loci of either haplotype by any primers thus designed so far (data not shown). Therefore, we concluded that no loci corresponding to *SLA-5* were transcribed on either haplotype.

Discussion

We demonstrated that the structure of the genomic region containing the *SLA* classical class I loci varied according to haplotypes. Furthermore, in particular, many of the loci similar to *SLA-1* (*SLA-1* and *SLA-12*), the number of which differed among the haplotypes analyzed, could be transcribed. This was consistent with the inference from RT-PCR-based genotyping that some haplotypes possess plural *SLA-1* loci or lack *SLA-1* or *SLA-3* loci [14,29]. Variance of genomic structure with haplotypes is implied by the diverse numbers of fragments generated by particular MS markers located in the *SLA* region, especially in the genomic region containing the classical class I genes, in different individuals [15]. However, the structures of the genomic regions containing *SLA* classical class I genes were strikingly different among haplotypes. In the sequence of the Hp-28.0 haplotype alone, two *SLA-1*, three *SLA-5*, and two *SLA-9* loci were observed. In addition, two *SLA-12*, similar to *SLA-1*, were observed in this haplotype. In contrast, the Hp-62.0 haplotype possessed one *SLA-1*, two *SLA-9*, and one *SLA-12* loci. In total, 13 and 8 (and perhaps one additional) class I genes were observed in the region between *SBAB-207G8.3* and *TRIM39* of the Hp-28.0 and Hp-62.0 haplotypes, respectively, whereas the Hp-1.0 haplotype possesses only seven class I genes in the corresponding region. The complexity of the structure of the *SLA* classical class I region is an obstacle to the genotyping of *SLA* haplotypes by using the genic sequences of *SLA per se* with genomic DNA. Therefore, precise determination of *SLA* haplotypes requires the compilation of sufficient knowledge of the structure of the existing *SLA* haplotypes, particularly in highly duplicated genomic regions such as that carrying the classical class I genes. Because currently the structures of only a few haplotypes, including the two in this study, have been elucidated, genotyping of *SLA* haplotypes using combinations of polymorphic markers such as MS markers is a realistic alternative method of classifying *SLA* haplotypes [30].

Copy number variance of loci for MHC class I genes has been also observed in other mammals. For example, previous studies have

demonstrated that the numbers of loci encoding murine H2-D and H2-Q vary among the haplotypes [31–33]. Structural and functional studies of antigen presentation by MHC molecules have been mostly limited to human and rodents [1], and the roles allotted to the respective molecules encoded by *SLA* class I genes remain unclear. However, our results, taken together with the results of genotyping by RT-PCR-based methods [14], indicate that there is at least one *SLA-1*(-like) or *SLA-3* locus, which retains an extremely high level of polymorphism and probably plays an important role in antigen presentation to T lymphocytes, in the majority of the *SLA* haplotypes, although the number of loci may vary. In some haplotypes, RT-PCR-based genotyping has failed to indicate the existence of *SLA-1* [14]. However, other *SLA* classical class I genes such as *SLA-3* may complement the function of *SLA-1*. The existence of *SLA-1*(-like) loci on the genomes of such haplotypes remains to be elucidated.

In the case of several loci for duplicated *SLA* classical class I genes, we observed transcription of the loci that had stop codons in their regions corresponding to CDSs. We confirmed that one *SLA-1* with a stop codon was transcribed on each haplotype in the individual investigated. On the other hand, three of the four *SLA-9* loci on the two haplotypes were transcribed, although all of these loci possessed stop codons and seemed nonfunctional for antigen presentation to T cell receptors. The significance of transcription of the *SLA-9* loci is unclear, as many MHC nonclassical class I genes with premature stop codons are transcribed [34]. However, it is plausible that the *SLA* classical class I genes, including pseudogenes, play a role in supplying sequences of exons and promoters to other functional loci of class I genes by a gene-conversion-like process through evolution. This possibility is implied in the case of MHC class II genes in primates [35].

Unfortunately, we were not able to isolate the genomic region containing *SLA-11*, the *SLA* class I gene most proximal to the centromere in the region carrying *SLA* classical class I genes, on the Hp-62.0 haplotype. However, comparison of the structures of the three haplotypes that have been sequenced to date revealed that extensive duplication or deletion of the genomic region occurred within the areas of *SLA-1*(-like), *SLA-5*, and *SLA-9* loci. The results of genomic Southern hybridization implied that only one locus, probably corresponding to *SLA-11* on the Hp-62.0 haplotype, exists besides the loci contained in the region that we sequenced. This implies that the members of loci for the *SLA* classical class I genes are generally conserved among haplotypes, but copy number variance of the loci corresponding to *SLA-1*(-like), *SLA-5*, and *SLA-9* occurs among the haplotypes.

Genotyping using MS markers has also implied that structural variance among *SLA* haplotypes exists in the regions different from those carrying the classical class I genes [15]. Genotyping using the genomic DNA of individuals of the Berkshire breed generated more than two fragments by MS markers located within class II regions [15]. Copy number variance of *HLA-DRB* is observed in the MHC class II region in humans [36]. Examination of the recently released Hp-1.0 *SLA* genomic sequence revealed that porcine *DRB* is duplicated and possesses multiple loci [10]. Variance in the number of fragments generated by MS markers adjacent to *DRB* may be reflected in the copy number variance of *DRB* in the *SLA* class II region; this variance will be elucidated by precise sequencing of the region in several haplotypes.

It is interesting to estimate the period of establishment of the structure covering the duplicated classical class I loci in the haplotypes sequenced here. By using SCRE core sequences and assuming an average substitution rate of 4.6×10^{-9} per site and per year of nucleotides in pseudogenes [37], the period of divergence between *SLA-1*(-like) (*SLA-1* and *SLA-12*) and *SLA-3* loci was estimated to be 15 and 11 million years ago (Mya) in the Hp-28.0 and Hp-62.0 haplotypes, respectively. On the other hand, phylogenetic analysis indicated that the divergence period of Tigger1a sequences flanked by the *SLA-1*(-like) locus at the end most proximal to the p-terminal was 20 to 26 Mya. This was quite similar to the result for the Hp-1.0 haplotype—approximately 15 Mya—in an earlier study [21]. The SCRE sequences flanking *SLA-1*(-like) and *SLA-5* loci possessed higher similarity to each other than to that flanking *SLA-3*, and the divergence was estimated to occur between 0.9 and 7.2 Mya. Meanwhile, the Tigger1a sequences except that flanking the *SLA-1*(-like) locus at the end most proximal to the p-terminal diverged from each other relatively recently (0.7 to 3 Mya). Taken together, these estimations suggest that the copy number variance in *SLA* classical class I loci occurred recently over several million years. However, analysis of additional genomic sequences encompassing the *SLA* classical class I loci in different haplotypes is required for precise estimation of the process of establishment of the structure of this region. Further investigation of the sequences in other *SLA* haplotypes will be needed to clarify when *SLA-12* diverged from *SLA-1*.

The active duplication and deletion process of the region containing the classical class I genes may contribute to an increase in the variety of functional alleles for antigen presentation, thus augmenting disease resistance against more diverse pathogens. However, the existence of an excessive number of such alleles could limit the valid repertoire of T cell receptors for antigens, because the presence of too many MHC class I alleles could immoderately eliminate the TCR repertoire by a negative selection process in the thymus. Therefore, the number of loci encoding functional class I molecules is restricted, although active duplications and deletions have occurred in the regions carrying these genes [38,39]. In humans, three functional classical class I loci, *HLA-A*, *HLA-B*, and *HLA-C*, are expressed in a single haplotype. The Hp-1.0 haplotype also possesses three functional classical class I genes. However, four functional classical class I genes are expressed in some haplotypes in pigs and other artiodactyls [14,16]. In the Hp-28.0 haplotype in this study, six classical class I loci were expressed, four of which possessed intact CDSs. *SLA-12* loci in the Hp-28.0 haplotype had a premature stop codon in its cytoplasmic tail; the function of their antigen presentation may be hampered. It would be interesting to determine whether the *SLA-1*-like locus with an intact CDS encoding molecules with the ability to present antigens. Moreover, the relatively large numbers of MHC classical class I loci in several haplotypes of the artiodactyls in comparison with those in primates or rodents may be correlated with a unique system for generating TCR diversity in artiodactyls. Otherwise, the number of functional classical class I molecules may be limited by generation of premature stop codons in the "excessive" *SLA-1*-like (*SLA-12*) loci.

Sequencing of the whole pig genome by an international consortium is now progressing [40], and recent progress in sequen-

cing technology has made it possible to sequence the genomic regions of different *SLA* haplotypes more efficiently. These developments may enable us to obtain structural information on many different haplotypes of *SLA*, and they may also enable us to develop a more precise and efficient method of determining *SLA* haplotypes by using genomic DNA. This will contribute to the analysis of how *SLA* haplotypes affect porcine immunological responses and disease resistance, and it will also help in the experimental design of transplantation research in pigs.

Materials and methods

Pig genomic DNA library and clone screening

A bacterial artificial chromosome (BAC) library that used genomic DNA from a male of Landrace breed, named L14–216, maintained by the National Institute of Livestock and Grassland Science (Tsukuba, Japan) was constructed as previously described [22]. BAC clones with the genomic region containing the *SLA* classical class I genes were screened with PCR primers designed within the sequences of the flanking genes *TRIM26* (forward: 5′-GGTGCTGTGCTCTATCTGTCTT-3′; reverse: 5′-TATC-TGCTGGTCGTCCTTGC-3′) and *TRIM39* (forward: 5′-AGGCAAGTTGTTC-TGTGTGTGC-3′; reverse: 5′-CCTCCTGGTCCTCATAGCAG-3′) by a multi-step PCR screening system constructed for the library, as previously described [41]. Overlaps of the BAC clones were confirmed by direct sequencing of the clones, using primers designed within the end sequences of those clone inserts estimated to overlap. Gaps in contiguous BAC clones were filled by the isolation of bridging clones, using primers designed within the end sequences of clone inserts on both sides of the gaps. Incorporation of the *SLA* classical class I genes in the isolated BAC clones was confirmed by rough shotgun sequencing before the complete sequencing, as described below.

DNA sequencing, assembly, and computational analysis

Sequencing of the isolated BAC clones was performed by a shotgun sequencing method, as previously described [42]. The assembled sequences of the BAC clones obtained were confirmed by digestion of the DNA of the clones with restriction enzymes and pulsed-field gel electrophoresis.

The sequences of genes located on the genomic sequences cloned to the BAC clones thus sequenced were estimated by comparison with the sequences of all the alleles reported so far for *SLA* classical class I genes [14], using BLAST similarity analysis of each two sequences (bl2seq) [43] and considering the AG–GT rule at splicing. Identity of the sequences between the overlapping BAC clones belonging to each haplotype was confirmed also by using bl2seq. Phylogenetic analysis was performed by the neighbor-joining method [44] using ClustalX [45].

Detection of transcription of putative genes by RT-PCR

Primers to detect expressed *SLA* classical class I genes were designed by manual inspection, to avoid amplifying other loci or alleles possessed by the individual thus used, for amplification of the classical class I genes within the BAC clones thus determined. The forward and reverse primers were designed, respectively, in exons 2 and 3, or exons 3 and 4, in particular loci, including the intermediate intron within the amplicon. The amplicons derived from RT-PCR were easily discriminated by their length from those generated from contaminated genomic DNA. The primers designed are shown in Table 2.

Reverse transcription was performed with ReverTra Dash (Toyobo, Osaka, Japan) in accordance with the manufacturer's instructions, using total RNA derived from a kidney of the Landrace individual; this individual was identical to that used for construction of the BAC library. PCR with the single strand cDNA thus obtained was performed with AmpliTaq Gold (Applied Biosystems). All of the PCR fragments

amplified with the primers thus designed were sequenced directly, or were cloned into plasmids by using a TOPO TA Cloning Kit (Invitrogen, Carlsbad, CA, USA) and sequenced to confirm the correct amplification of the targeted alleles.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2008.10.004.

References

[1] A. Kumánovics, T. Takada, K.F. Lindahl, Genomic organization of the mammalian MHC, Annu. Rev. Immunol. 21 (2003) 629–657.
[2] A.L. Hughes, M. Nei, Evolutionary relationships of class II major-histocompatibility-complex genes in mammals, Mol. Biol. Evol. 7 (1990) 491–514.
[3] K.T. Ballingall, S.A. Ellis, N.D. MacHugh, S.D. Archibald, D.J. McKeever, The *DY* genes of the cattle MHC: expression and comparative analysis of an unusual class II MHC gene pair, Immunogenetics 55 (2004) 748–755.
[4] N. Yuhki, T. Beck, R.M. Stephens, Y. Nishigaki, K. Newmann, S.J. O'Brien, Comparative genome organization of human, murine, and feline MHC class II region, Genome Res. 13 (2003) 1169–1179.
[5] A.L. Hughes, M. Yeager, A.E. Ten Elshof, M.J. Chorney, A new taxonomy of mammalian MHC class I molecules, Immunol. Today 20 (1999) 22–26.
[6] E.J. Adams, P. Parham, Species-specific evolution of MHC class I genes in the higher primates, Immunol. Rev. 183 (2001) 41–64.
[7] A.L. Hughes, Independent gene duplications, not concerted evolution, explain relationships among class I MHC genes of murine rodents, Immunogenetics 33 (1991) 367–373.
[8] A. Ando, A. Shigenari, J.K. Kulski, C. Renard, P. Chardon, T. Shiina, H. Inoko, 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 (2005) 864–873.
[9] P. Chardon, C. Rogel-Gaillard, L. Cattolico, S. Duprat, M. Vaiman, C. Renard, Sequence of the swine major histocompatibility complex region containing all non-classical class I genes, Tissue Antigens 57 (2001) 55–65.
[10] C. Renard, E. Hart, H. Sehra, H. Beasley, P. Coggill, K. Howe, J. Harrow, J. Gilbert, S. Sims, J. Rogers, A. Ando, A. Shigenari, T. Shiina, H. Inoko, P. Chardon, S. Beck, The genomic sequence and analysis of the swine major histocompatibility complex, Genomics 88 (2006) 96–110.
[11] C. Renard, M. Vaiman, N. Chiannilkulchai, L. Cattolico, C. Robert, P. Chardon, Sequence of the pig major histocompatibility region containing the classical class I genes, Immunogenetics 53 (2001) 490–500.
[12] A. Shigenari, A. Ando, C. Renard, P. Chardon, T. Shiina, J.K. Kulski, H. Yasue, H. Inoko, Nucleotide sequencing analysis of the swine 433-kb genomic segment located between the non-classical and classical SLA class I gene clusters, Immunogenetics 55 (2004) 695–705.
[13] P. Chardon, C. Renard, M. Vaiman, The major histocompatibility complex in swine, Immunol. Rev. 167 (1999) 179–192.
[14] D.M. Smith, J.K. Lunney, G.W. Martens, A. Ando, J.H. Lee, C.S. Ho, L. Schook, C. Renard, P. Chardon, Nomenclature for factors of the SLA class-I system 2004, Tissue Antigens 65 (2005) 136–149.
[15] M. Tanaka, A. Ando, C. Renard, P. Chardon, M. Domukai, N. Okumura, T. Awata, H. Uenishi, Development of dense microsatellite markers in the entire SLA region and evaluation of their polymorphisms in porcine breeds, Immunogenetics 57 (2005) 690–696.
[16] K.T. Ballingall, D. Miltiadou, Z.W. Chai, K. McLean, M. Rocchi, R. Yaga, D.J. McKeever, Genetic and proteomic analysis of the MHC class I repertoire from four ovine haplotypes, Immunogenetics 60 (2008) 177–184.
[17] S.A. Ellis, E.C. Holmes, K.A. Staines, K.B. Smith, M.J. Stear, D.J. McKeever, N.D. MacHugh, W.I. Morrison, Variation in the number of expressed MHC genes in different cattle class I haplotypes, Immunogenetics 50 (1999) 319–328.
[18] D.K. Cooper, B. Gollackner, D.H. Sachs, Will the pig solve the transplantation backlog? Annu. Rev. Med. 53 (2002) 133–147.
[19] J.S. Logan, Prospects for xenotransplantation, Curr. Opin. Immunol. 12 (2000) 563–568.
[20] D.K. Cooper, Xenoantigens and xenoantibodies, Xenotransplantation 5 (1998) 6–17.
[21] C. Renard, P. Chardon, M. Vaiman, The phylogenetic history of the MHC class I gene families in pig, including a fossil gene predating mammalian radiation, J. Mol. Evol. 57 (2003) 420–434.
[22] M. Tanaka, K. Suzuki, T. Morozumi, E. Kobayashi, T. Matsumoto, M. Domukai, T. Eguchi-Ogawa, H. Shinkai, T. Awata, H. Uenishi, Genomic structure and gene order of swine chromosome 7q1.1→q1.2, Anim. Genet. 37 (2006) 10–16.
[23] S.A. Ellis, R.E. Bontrop, D.F. Antczak, K. Ballingall, C.J. Davies, J. Kaufman, L.J. Kennedy, J. Robinson, D.M. Smith, M.J. Stear, R.J. Stet, M.J. Waller, L. Walter, S.G. Marsh, ISAG/IUIS-VIC Comparative MHC Nomenclature Committee report, 2005, Immunogenetics 57 (2006) 953–958.
[24] M.S. Boguski, T.M. Lowe, C.M. Tolstoshev, dbEST — database for "expressed sequence tags", Nat. Genet. 4 (1993) 332–333.
[25] H. Uenishi, T. Eguchi-Ogawa, H. Shinkai, N. Okumura, K. Suzuki, D. Toki, N. Hamasima, T. Awata, PEDE (Pig EST Data Explorer) has been expanded into Pig Expression Data Explorer, including 10 147 porcine full-length cDNA sequences, Nucleic Acids Res. 35 (2007) D650–D653.
[26] A.F. Smit, Interspersed repeats and other mementos of transposable elements in mammalian genomes, Curr. Opin. Genet. Dev. 9 (1999) 657–663.
[27] J. Jurka, Repbase update: a database and an electronic journal of repetitive elements, Trends Genet. 16 (2000) 418–420.
[28] A.F. Smit, A.D. Riggs, Tiggers and DNA transposon fossils in the human genome, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 1443–1448.
[29] A. Ando, H. Kawata, A. Shigenari, T. Anzai, M. Ota, Y. Katsuyama, M. Sada, R. Goto, S. N. Takeshima, Y. Aida, T. Iwanaga, N. Fujimura, Y. Suzuki, T. Gojobori, H. Inoko, Genetic polymorphism of the swine major histocompatibility complex (SLA) class I genes, SLA-1, -2 and -3, Immunogenetics 55 (2003) 583–593.
[30] A. Ando, H. Uenishi, H. Kawata, M. Tanaka-Matsuda, A. Shigenari, L. Flori, P. Chardon, J.K. Lunney, J.K. Kulski, H. Inoko, Microsatellite diversity and crossover regions within homozygous and heterozygous SLA haplotypes of different pig breeds, Immunogenetics 60 (2008) 399–407.
[31] A. Kumánovics, A. Madan, S. Qin, L. Rowen, L. Hood, K. Fischer Lindahl, Quod erat faciendum: sequence analysis of the H2-D and H2-Q regions of 129/SvJ mice, Immunogenetics 54 (2002) 479–489.
[32] E.H. Weiss, D. Bevec, G. Messer, S. Schwemmle, C. Grosshaus, M. Steinmetz, W. Schmidt, Organization of the AKR Qa region: structure of a divergent class I sequence, Q5^k, J. Immunogenet. 16 (1989) 283–290.
[33] E.H. Weiss, L. Golden, K. Fahrner, A.L. Mellor, J.J. Devlin, H. Bullman, H. Tiddens, H. Bud, R.A. Flavell, Organization and evolution of the class I gene family in the major histocompatibility complex of the C57BL/10 mouse, Nature 310 (1984) 650–655.
[34] J. Trowsdale, "Both man and bird and beast": comparative organization of MHC genes, Immunogenetics 41 (1995) 1–17.
[35] G.G. Doxiadis, M.K. van der Wiel, H.P. Brok, N.G. de Groot, B.A. 't Hart, J.J. van Rood, R.E. Bontrop, Reactivation by exon shuffling of a conserved HLA-DR3-like pseudogene segment in a New World primate species, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 5864–5868.
[36] R.E. Bontrop, N. Otting, N.G. de Groot, G.G. Doxiadis, Major histocompatibility complex class II polymorphisms in primates, Immunol. Rev. 167 (1999) 339–350.
[37] W.H. Li, T. Gojobori, M. Nei, Pseudogenes as a paradigm of neutral evolution, Nature 292 (1981) 237–239.
[38] J.W. Kappler, N. Roehm, P. Marrack, T cell tolerance by clonal elimination in the thymus, Cell 49 (1987) 273–280.
[39] K.M. Wegner, M. Kalbe, J. Kurtz, T.B. Reusch, M. Milinski, Parasite selection for immunogenetic optimality, Science 301 (2003) 1343.
[40] S.J. Humphray, C.E. Scott, R. Clark, B. Marron, C. Bender, N. Camm, J. Davis, A. Jenks, A. Noon, M. Patel, H. Sehra, F. Yang, M.B. Rogatcheva, D. Milan, P. Chardon, G. Rohrer, D. Nonneman, P. de Jong, S.N. Meyers, A. Archibald, J.E. Beever, L.B. Schook, J. Rogers, A high utility integrated map of the pig genome, Genome Biol. 8 (2007) R139.
[41] K. Suzuki, S. Asakawa, M. Iida, S. Shimanuki, N. Fujishima, H. Hiraiwa, Y. Murakami, N. Shimizu, H. Yasue, Construction and evaluation of a porcine bacterial artificial chromosome library, Anim. Genet. 31 (2000) 8–12.
[42] H. Shinkai, T. Morozumi, D. Toki, T. Eguchi-Ogawa, Y. Muneta, T. Awata, H. Uenishi, Genomic structure of eight porcine chemokine receptors and intergene sharing of an exon between CCR1 and XCR1, Gene 349 (2005) 55–66.
[43] T.A. Tatusova, T.L. Madden, BLAST 2 sequences, a new tool for comparing protein and nucleotide sequences, FEMS Microbiol. Lett. 174 (1999) 247–250.
[44] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[45] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, Nucleic Acids Res. 25 (1997) 4876–4882.