Available online at www.sciencedirect.com
SCIENCE DIRECT
Genomics 88 (2006) 96–110
GENOMICS
www.elsevier.com/locate/ygeno

The genomic sequence and analysis of the swine major histocompatibility complex

C. Renard [a,1], E. Hart [b,1], H. Sehra [b], H. Beasley [b], P. Coggill [b], K. Howe [b], J. Harrow [b], J. Gilbert [b], S. Sims [b], J. Rogers [b], A. Ando [c], A. Shigenari [c], T. Shiina [c], H. Inoko [c], P. Chardon [a], S. Beck [b,*]

[a] *LREG INRA CEA, Jouy en Josas, France*
[b] *Wellcome Trust Sanger Institute, Genome Campus, Hinxton, Cambridge CB10 1SA, UK*
[c] *Department of Molecular Life Science, Tokai University School of Medicine, Bohseidai, Isehara, Kanagawa-Pref. 259-1193, Japan*

Received 7 December 2005; accepted 18 January 2006
Available online 2 March 2006

Abstract

We describe the generation and analysis of an integrated sequence map of a 2.4-Mb region of pig chromosome 7, comprising the classical class I region, the extended and classical class II regions, and the class III region of the major histocompatibility complex (MHC), also known as swine leukocyte antigen (SLA) complex. We have identified and manually annotated 151 loci, of which 121 are known genes (predicted to be functional), 18 are pseudogenes, 8 are novel CDS loci, 3 are novel transcripts, and 1 is a putative gene. Nearly all of these loci have homologues in other mammalian genomes but orthologues could be identified with confidence for only 123 genes. The 28 genes (including all the SLA class I genes) for which unambiguous orthology to genes within the human reference MHC could not be established are of particular interest with respect to porcine-specific MHC function and evolution. We have compared the porcine MHC to other mammalian MHC regions and identified the differences between them. In comparison to the human MHC, the main differences include the absence of *HLA-A* and other class I-like loci, the absence of *HLA-DP*-like loci, and the separation of the extended and classical class II regions from the rest of the MHC by insertion of the centromere. We show that the centromere insertion has occurred within a cluster of *BTNL* genes located at the boundary of the class II and III regions, which might have resulted in the loss of an orthologue to human *C6orf10* from this region.

Keywords: Adaptive immune system; Centromere repositioning; Comparative sequence analysis; Evolution; Swine leukocyte antigen (SLA) complex

The pig is an important model organism for both biomedical and agronomic research. The implications of swine leukocyte antigen (SLA) molecules in xenotransplantation and the association of the major histocompatibility complex (MHC) region in general with quantitative traits, such as growth rate and carcass fat accumulation, have led to a number of comprehensive studies into the porcine MHC [1–4]. In addition to its unique role in histocompatibility, the primary function of the porcine MHC is to provide protection against pathogens [5]. Consequently, a detailed analysis of the genes encoded within the MHC is essential to advance our understanding of the processes implicated in immune responses and the effects of intensive selection, e.g., via strict breeding programs, which have the potential to affect the haplotype structure and polymorphism of the MHC.

The porcine MHC or SLA complex is located on submetacentric chromosome 7 (SSC7p1.1–1q1.1). Physical mapping achieved contiguous BAC coverage of the entire region with the exception of the centromere, which separates the class II from the class III region [6]. The previously sequenced class I region [7–10,59] has been incorporated into the MHC reference sequence reported here, now comprising the entire class I, class II, and class III regions. The reference sequence has been subjected to comprehensive analysis and annotation, resulting in the first complete gene map of the SLA region. Differences between the orthologous maps in other mammals,

* Corresponding author. Fax: +44 1223 494919.
E-mail address: beck@sanger.ac.uk (S. Beck).
[1] These authors contributed equally to this work.

0888-7543/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2006.01.004

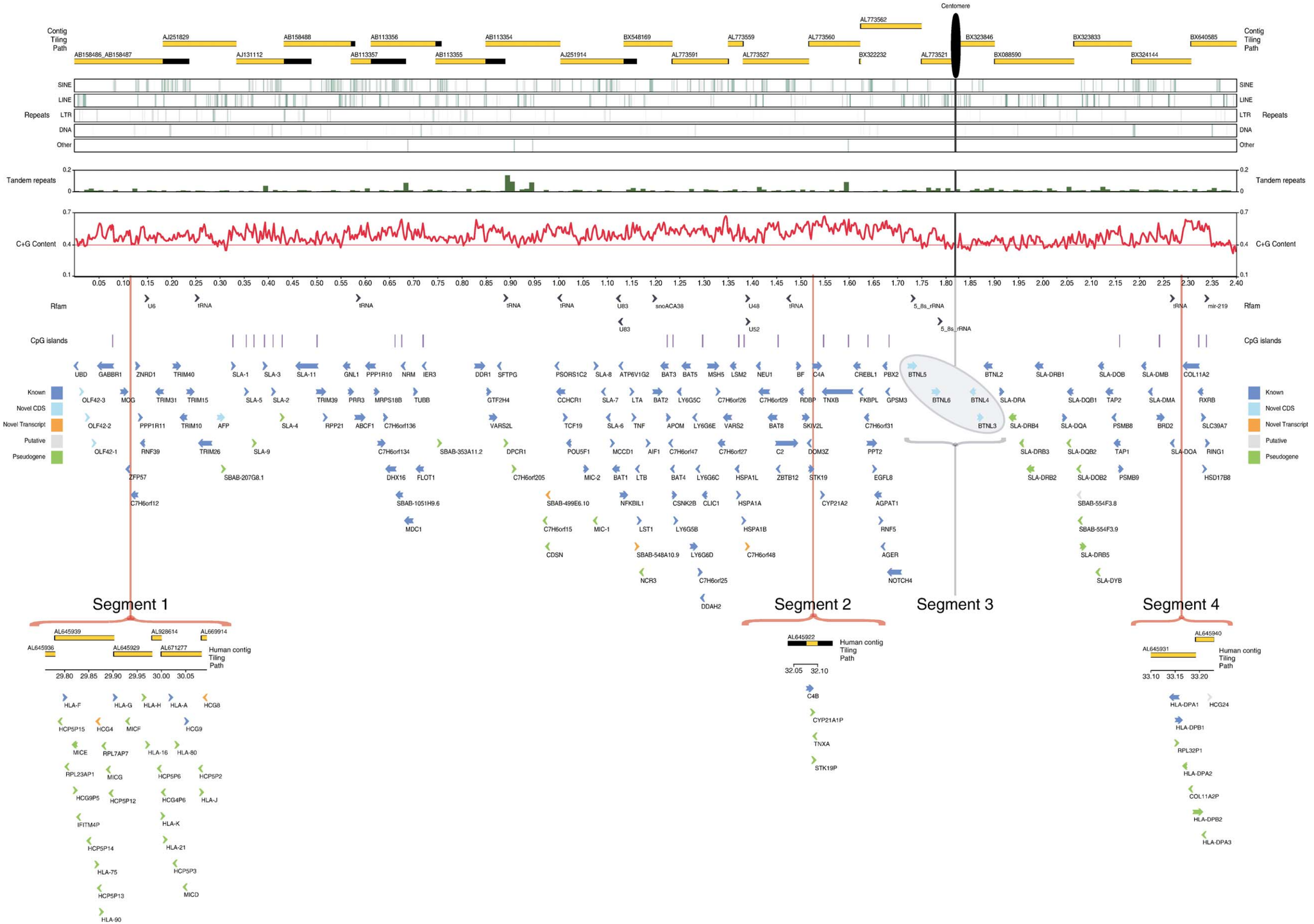

Fig. 1. Feature map of the SLA. Each locus is annotated according to type, orientation, and position within the SLA. The tiling path of the sequenced BACs is shown on the top with overlaps shown in black. Below this, the distribution of repeats and C + G content across the region is shown. RNA genes identified using Rfam and CpG islands are also depicted. Segment 1 illustrates the section of the HLA class I region that is absent in pig. Segment 2 illustrates the HLA class III RCCX module absent in pig. Segment 3 highlights the porcine-specific *BTNL* cluster surrounding the centromere. Segment 4 illustrates the section of the HLA class II region containing the *DP* loci that is absent from pig.

Table 1
Clone names and accession numbers of the pig BACs sequenced to provide coverage of the SLA region

BAC clone name	EMBL accession No.
SSC7p-telomeric	
Class I	
SBAB-649H10[a]	AB158486
SBAB-792A7[a]	AB158487
SBAB-207G8[b]	AJ251829
SBAB-490B10[b]	AJ131112
SBAB-771G4[a]	AB158488
SBAB-1111D10[a]	AB113357
SBAB-1051H9[a]	AB113356
SBAB-353A11[a]	AB113355
SBAB-499E6[a]	AB113354
SBAB-493A6[b]	AJ251914
Class III	
SBAB-548A10[c]	BX548169
SBAB-35B1[c]	AL773591
SBAB-711D2[c]	AL773559
SBAB-707F1[c]	AL773527
SBAB-514B12[c]	AL773560
SBAB- 446E4[c]	BX322232
SBAB-649D6[c]	AL773562
SBAB-339A5[c]	AL773521
Centromere	
Class II	
SBAB-43B6[c]	BX323846
SBAB-591C4[c]	BX088590
SBAB-554F3[c]	BX323833
SBAB-1044B7[c]	BX324144
SBAB-279G2[c]	BX640585
SSC7q-telomeric	

BACs were sequenced and submitted by [a]Tokai University School of Medicine (Japan); [b]INRA-CEA (France); or [c]Wellcome Trust Sanger Institute (UK).

particularly human, will be discussed in the context of MHC plasticity and evolution.

Results and discussion

SLA gene map

The 23 BAC clones that comprise the SLA span a region of 2.4 Mb and are listed in Table 1. They are derived from the SLA haplotype H01, which is prevalent in commercial breeds and particularly frequent in Yorkshire/Large White breeds. The entire SLA is represented by two BAC contigs interrupted by the centromere, as shown in Fig. 1. The first contig spans 1.8 Mb and consists of the contiguous class I and III regions, from the telomeric *UBD* gene in the extended class I region to 2 butyrophilin-like genes, *BTNL5* and *BTNL6,* at the centromeric end of the class III region. The second contig is 0.6 Mb in length and consists of the SLA class II region, extending from the centromeric *BTNL* gene cluster to the *RING1* gene at the telomeric end of the extended class II region. Within these two contigs we identified and annotated 151 gene loci; 121 of these are known genes that are predicted to be functional, 18 are pseudogenes, 8 loci are classified as novel CDS, 3 are novel transcripts, and 1 is putative. Because of the difficulty in unambiguously defining 1:1 orthologous relationships for loci in genomic regions that have undergone species-specific duplication, we were unable to assign orthology for 1 olfactory receptor locus, 4 butyrophilin-like loci, all 10 SLA class I loci, 7 class II pseudogene loci, 1 *TAP2*-like pseudogene locus, 2 novel transcript loci, and 1 putative gene locus. The most conserved part of the SLA complex is the class III region, which comprises 61 loci; of these 56 are known loci, 2 are novel CDS loci (*BTNL5, BTNL6*), 2 are novel transcripts (*SBAB-548A10.9, C7H6orf48*), and 1, *NCR3,* is a pseudogene. The functions of the 4 novel loci are currently unknown. Remarkably, all class III loci with the exception of *BTNL5, BTNL6,* and novel transcript *SBAB-548A10.9* have orthologues within the human and rodent class III regions. A summary of the gene annotation is listed in Table 2 and the full annotation is available online at the Vertebrate Genome Annotation (VEGA) database [11] (http://vega.sanger.ac.uk).

Compared with protein coding genes, non-protein-coding RNA genes are more difficult to predict. Transfer RNA (tRNA) genes are the functionally best understood class that can be predicted with high confidence. In the extended human MHC, for instance, over 150 tRNA genes have been identified [12]. Using the Rfam database [13] we were able to predict 15 RNA genes in the SLA region, comprising 6 tRNAs, 5 snoRNAs, 2 rRNAs, 1 miRNA, and 1 snRNA as shown in Table 3. The positions of two previously reported [14] snoRNAs conserved within introns of the *BAT1* gene in human, mouse, and pig are illustrated in Fig. 2. It is notable that, despite conservation of the snoRNAs, these noncoding regions are not consistently well conserved between species.

SLA plasticity

Overall, there is a high level of conserved synteny between the SLA and the HLA complex. There are, however, four segments of plasticity in which the two MHC regions differ significantly. These segments are shown in the lower part of Fig. 1 and will be discussed from left to right. Except for the insertion of the centromere around position 1.82 Mb, these segments constitute major deletions in the SLA or major insertions in the HLA complex.

Segment 1 maps between porcine genes *ZFP57* and *C7H6orf12,* defining the telomeric boundary of the human leukocyte antigen (HLA) class I region. It is about 300 kb in length and comprises 30 gene loci, including 1 classical (*HLA-A*) and 13 nonclassical class I loci. This apparent loss of class I loci is compensated for in the SLA by a cluster of 7 classical class I genes (*SLA-1, SLA-5, SLA-9, SLA-3, SLA-2, SLA-4,* and *SLA-11*) adjacent to the *TRIM* cluster and 3 nonclassical class I genes (*SLA-6, SLA-7, SLA-8*) between *POU5F1* and the start of the class III region. At this point it is unclear whether the 300-kb difference is due to a deletion in SLA or an insertion in HLA. Comparative analysis in cat [15] and horse [16] revealed an absence of MHC class I genes within the same region, suggesting an insertion event in human, although the expansion of class I genes in rodents [17,18] is more in line with a deletion event in pig, cat, and horse. The situation in cattle is still under

Table 2
List of annotated SLA gene loci, detailing gene name, locus type, locus description, and SLA coordinates for each locus

Gene name	Locus type	Locus description	Start	End
Extended Class I				
UBD	Known	Ubiquitin D	1038	3458
OLF42-3	Novel CDS	Novel olfactory receptor OLF42-3	16,138	17,076
OLF42-2	Novel CDS	Novel olfactory receptor OLF42-2	28,268	29,206
OLF42-1	Novel CDS	Novel olfactory receptor OLF42-1	41,896	42,834
GABBR1	Known	γ-Aminobutyric acid B receptor, 1	49,886	79,245
MOG	Known	Myelin oligodendrocyte glycoprotein	96,753	108,648
Class I				
ZFP57	Known	Zinc finger protein 57 homologue (mouse) (possible pseudogene)	109,436	111,924
C7H6orf12	Known	Likely orthologue of human chromosome 6 open reading frame 12	116,381	128,597
ZNRD1	Known	Zinc ribbon domain containing, 1	128,225	132,921
PPP1R11	Known	Protein phosphatase 1, regulatory (inhibitor) subunit 11	135,187	138,350
RNF39	Known	Ring finger protein 39 (HZFW1)	138,279	144,102
TRIM31	Known	Tripartite motif-containing 31	169,707	182,535
TRIM40	Known	Tripartite motif-containing 40 (orphan exon B30)	205,146	217,441
TRIM10	Known	Tripartite motif-containing 10 (ring finger protein B30)	220,499	231,022
TRIM15	Known	Tripartite motif-containing 15 (zinc finger protein B7)	233,573	245,288
TRIM26	Known	Tripartite motif-containing 26 (zinc finger protein 173)	258,677	282,575
AFP	Novel CDS	Putative acid finger protein (AFP) similar to TRIM26 (possible pseudogene)	296,686	308,097
SBAB-207G8.1	Pseudogene	RAP1A, member of RAS oncogene family (RAP1A) pseudogene	309,724	310,220
SLA-1	Known	Classical MHC class I antigen 1	327,007	330,592
SLA-5	Known	Classical MHC class I antigen 5 (possible pseudogene)	354,541	357,621
SLA-9	Pseudogene	Classical MHC class I antigen 9 (pseudogene)	370,802	373,883
SLA-3	Known	Classical MHC class I antigen 3	392,340	396,054
SLA-2	Known	Classical MHC class I antigen 2	410,204	413,623
SLA-4	Pseudogene	Classical MHC class I antigen 4 (pseudogene)	428,589	431,618
SLA-11	Known	Nonclassical/classical MHC class I antigen 11 (fossil gene)	459,588	501,608
TRIM39	Known	Tripartite motif-containing 39	501,141	517,176
RPP21	Known	Ribonuclease P 21-kDa subunit	518,569	520,496
GNL1	Known	Guanine nucleotide binding protein-like 1 (HSR1)	557,396	568,029
PRR3	Known	Proline-rich polypeptide 3	566,991	574,738
ABCF1	Known	ATP-binding cassette, subfamily F (GCN20) member 1	580,882	598,652
PPP1R10	Known	Protein phosphatase 1, regulatory subunit 10 (FB19)	603,491	620,345
MRPS18B	Known	Mitochondrial ribosomal protein S18-2	620,602	627,651
C7H6orf134	Known	Orthologous to human chromosome 6 open reading frame 134	628,082	640,120
C7H6orf136	Known	Orthologous to human chromosome 6 open reading frame 136	640,341	644,956
DHX16	Known	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16	645,290	662,750
SBAB-1051H9.6	Known	Likely orthologue of human KIAA1949	666,491	676,902
NRM	Known	Nurim (nuclear envelope membrane protein)	677,038	680,564
MDC1	Known	Mediator of DNA damage checkpoint 1	684,268	698,278
TUBB	Known	Tubulin, β-polypeptide (β5-tubulin)	702,512	707,024
FLOT1	Known	Flotillin 1	707,940	719,927
IER3	Known	Immediate early response 3	720,377	721,674
SBAB-353A11.2	Pseudogene	NADH dehydrogenase (ubiquitinone) 1β subcomplex 9 (NDUF9) pseudogene	755,706	756,498
DDR1	Known	Discoidin domain receptor family, member 1	828,411	846,175
GTF2H4	Known	General transcription factor IIH, polypeptide 4	853,062	858,810
VARS2L	Known	Valyl-tRNA synthetase 2 like	858,928	870,958
SFTPG	Known	Surfactant associated protein G	874,942	875,928
DPCR1	Pseudogene	Diffuse panbronchiolitis critical region 1	893,067	893,862
C7H6orf205	Pseudogene	Likely orthologue of human chromosome 6 open reading frame 206	907,332	910,776
C7H6orf15	Pseudogene	Likely orthologue of human chromosome 6 open reading frame 15	970,138	971,279
CDSN	Pseudogene	Corneodesmosin precursor (CDSN)	975,411	978,730
SBAB-499E6.10	Novel transcript	Novel transcript (overlaps *CDSN* pseudogene)	976,365	978,740
PSORS1C2	Known	Psoriasis susceptibility 1 candidate 2 (SPR1)	994,625	996,463
CCHCR1	Known	Coiled-coil α-helical rod protein 1	998,242	1,012,287
TCF19	Known	Transcription factor 19 (SC1)	1,012,640	1,016,353
POU5F1	Known	POU domain, class 5, transcription factor 1	1,018,942	1,024,930
MIC-2	Known	Similar to human MHC class I polypeptide-related sequence B	1,053,639	1,059,961
MIC-1	Pseudogene	Pseudogene similar to human MHC class I polypeptide-related sequence B	1,073,922	1,075,418
SLA-8	Known	Nonclassical MHC class I antigen 8	1,076,659	1,080,066
SLA-7	Known	Nonclassical MHC class I antigen 7	1,090,894	1,093,969
SLA-6	Known	Nonclassical MHC class I antigen 6	1,101,096	1,105,047

Table 2 (*continued*)

Gene name	Locus type	Locus description	Start	End
Class III				
MCCD1	Known	Mitochondrial coiled-coil domain 1	1,112,384	1,113,878
BAT1	Known	Orthologous to HLA-B associated transcript 1	1,113,865	1,124,847
ATP6V1G2	Known	ATPase, H^+ transporting, lysosomal 13-kDa, V1 subunit G isoform 2	1,126,768	1,129,100
NFKBIL1	Known	Nuclear factor of κ light polypeptide gene enhancer in B-cells inhibitor-like 1	1,128,720	1,140,899
LTA	Known	Lymphotoxin α (TNF superfamily, member 1)	1,152,039	1,154,186
TNF	Known	Tumor necrosis factor (TNF superfamily, member 2)	1,155,506	1,158,309
LTB	Known	Lymphotoxin β (TNF superfamily, member 3)	1,161,098	1,163,008
SBAB-548A10.9	Novel transcript	Novel transcript	1,162,065	1,164,394
LST1	Known	Leukocyte-specific transcript 1 (possible pseudogene)	1,167,601	1,169,317
NCR3	Pseudogene	Natural cytotoxicity triggering receptor 3	1,169,707	1,172,101
AIF1	Known	Allograft inflammatory factor 1	1,185,155	1,187,470
BAT2	Known	Orthologous to human HLA-B associated transcript 2	1,196,580	1,212,374
BAT3	Known	Orthologous to human HLA-B associated transcript 3	1,213,205	1,225,521
APOM	Known	Apolipoprotein M	1,224,732	1,229,637
C7H6orf47	Known	Orthologous to human chromosome 6 open reading frame 47	1,229,758	1,232,220
BAT4	Known	Orthologous to human HLA-B associated transcript 4	1,233,413	1,237,092
CSNK2B	Known	Casein kinase 2, β polypeptide	1,236,514	1,242,204
LY6G5B	Known	Lymphocyte antigen 6 complex, locus G5D	1,243,126	1,244,751
LY6G5C	Known	Lymphocyte antigen 6 complex, locus G5C	1,247,563	1,252,052
BAT5	Known	Orthologous to human HLA-B associated transcript 5	1,257,251	1,271,219
LY6G6D	Known	Lymphocyte antigen 6 complex, locus G6D	1,273,951	1,285,072
LY6G6E	Known	Lymphocyte antigen 6 complex, locus G6E	1,279,706	1,281,620
LY6G6C	Known	Lymphocyte antigen 6 complex, locus G6C	1,286,128	1,289,210
C7H6orf25	Known	Orthologous to human chromosome 6 open reading frame 25	1,291,342	1,295,001
DDAH2	Known	Dimethylarginine dimethylaminohydrolase 2	1,295,335	1,298,841
CLIC1	Known	Chloride intracellular channel 1	1,299,178	1,305,705
MSH5	Known	MutS homologue 5 (*Escherichia coli*)	1,309,334	1,329,480
C7H6orf26	Known	Orthologous to human chromosome 6 open reading frame 26	1,329,840	1,331,869
C7H6orf27	Known	Orthologous to human chromosome 6 open reading frame 27	1,331,949	1,342,553
VARS2	Known	Valyl-tRNA synthetase 2	1,342,823	1,356,236
LSM2	Known	LSM2 homologue U6 small nuclear RNA associated (*Saccharomyces cerevisiae*)	1,357,079	1,365,553
HSPA1L	Known	Heat shock 70-kDa protein 1-like	1,366,987	1,371,495
HSPA1A	Known	Heat shock 70-kDa protein 1A	1,372,025	1,374,461
HSPA1B	Known	Heat shock 70-kDa protein 1B	1,382,823	1,385,238
C7H6orf48	Novel transcript	Likely orthologue of human chromosome 6 open reading frame 48	1,390,475	1,392,838
NEU1	Known	Sialidase 1 (lysosomal sialidase)	1,410,790	1,415,736
C7H6orf29	Known	Orthologous to human chromosome 6 open reading frame 29	1,416,014	1,434,702
BAT8	Known	Orthologous to human HLA-B associated transcript 8	1,435,155	1,450,146
C2	Known	Complement component 2	1,450,246	1,492,584
ZBTB12	Known	Zinc finger and BTB domain containing 12	1,452,001	1,454,442
BF	Known	B-factor (properdin)	1,492,971	1,499,054
RDBP	Known	RD RNA binding protein	1,499,056	1,505,513
SKIV2L	Known	Superkiller viralicidic activity 2-like (*S. cerevisiae*)	1,505,557	1,516,164
DOM3Z	Known	Dom-3 homologue Z (*Caenorhabditis elegans*)	1,516,213	1,518,687
STK19	Known	Serine/threonine kinase 19	1,518,768	1,525,926
C4A	Known	Complement component 4A	1,526,547	1,541,623
CYP21A2	Known	Cytochrome P450, family 21, subfamily A, polypeptide 2	1,544,703	1,547,904
TNXB	Known	Tenascin XB	1,547,413	1,606,661
CREBL1	Known	cAMP-responsive element binding protein-like 1	1,611,898	1,622,257
FKBPL	Known	FK506 binding protein like	1,622,763	1,625,053
C7H6orf31	Known	Orthologous to human chromosome 6 open reading frame 31	1,634,093	1,639,848
PPT2	Known	Palmitoyl-protein thioesterase 2	1,638,961	1,655,520
EGFL8	Known	EGF-like-domain, multiple 8	1,652,209	1,655,524
AGPAT1	Known	1-Acylglycerol-3-phosphate *O*-acyltransferase 1 (acetoacetyl coenzyme A thiolase)	1,655,456	1,665,823
RNF5	Known	Ring finger protein 5	1,666,052	1,668,542
AGER	Known	Advanced glycosylation end product-specific receptor	1,668,679	1,671,726
PBX2	Known	Pre-B-cell leukemia transcription factor 2	1,672,129	1,677,516
GPSM3	Known	G-protein signaling modulator 3 (AGS3-like, *C. elegans*)	1,678,181	1,680,386
NOTCH4	Known	Notch homologue 4 (*Drosophila*)	1,681,906	1,706,942
BTNL5	Novel CDS	Novel protein similar to butyrophilin family proteins 5	1,723,457	1,738,340
BTNL6	Novel CDS	Novel protein similar to butyrophilin family proteins 6	1,773,093	1,786,375

(continued on next page)

Table 2 (*continued*)

Gene name	Locus type	Locus description	Start	End
Centromere				
Class II				
BTNL4	Novel CDS	Novel protein similar to butyrophilin family proteins 4	1,852,576	1,860,307
BTNL3	Novel CDS	Novel protein similar to butyrophilin family proteins 3	1,867,579	1,875,399
BTNL2	Known	Butyrophilin-like 2 MHC class II associated	1,881,380	1,894,763
SLA-DRA	Known	MHC class II, DR α	1,912,873	1,918,468
SLA-DRB4	Pseudogene	MHC class II, DR β-like 4 pseudogene	1,933,463	1,944,005
SLA-DRB3	Pseudogene	MHC class II, DR β-like 3 pseudogene	1,954,979	1,959,011
SLA-DRB2	Pseudogene	MHC class II, DR β-like 2 pseudogene	1,970,114	1,981,412
SLA-DRB1	Known	MHC class II, DR β1	1,987,113	1,999,786
SLA-DQA	Known	MHC class II, DQ α	2,038,519	2,044,340
SLA-DQB2	Pseudogene	MHC class II, DQ β-like 2 pseudogene	2,052,512	2,053,288
SLA-DQB1	Known	MHC class II, DQ β gene (locus 1)	2,053,707	2,062,115
SLA-DOB2	Pseudogene	MHC class II, DO β-like fragment	2,073,125	2,074,079
SBAB-554F3.8	Putative	Putative novel transcript (overlaps SLA-DOB2 and SBAB-554F3.9)	2,073,704	2,077,258
SBAB-554F3.9	Pseudogene	Pseudogene similar to part of transporter 2, ATP-binding cassette, subfamily B (MRD/TAP) (TAP2)	2,076,026	2,076,394
SLA-DRB5	Pseudogene	MHC class II, DR β-like 5 pseudogene	2,079,632	2,087,758
SLA-DYB	Pseudogene	MHC class II, DY/DQ β-like pseudogene	2,112,714	2,114,157
SLA-DOB	Known	MHC class II, DO β	2,117,722	2,125,314
TAP2	Known	Transporter 2, ATP-binding cassette, subfamily B (MDR/TAP)	2,132,977	2,144,285
PSMB8	Known	Proteasome (prosome, macropain) subunit, β type, 8	2,145,879	2,149,684
TAP1	Known	Transporter 1, ATP-binding cassette, subfamily B (MDR/TAP)	2,150,187	2,159,554
PSMB9	Known	Proteasome (prosome, macropain) subunit, β type, 9	2,158,749	2,164,942
SLA-DMB	Known	MHC class II, DM β	2,206,753	2,212,574
SLA-DMA	Known	MHC class II, DM α	2,220,965	2,225,342
BRD2	Known	Bromodomain containing 2	2,238,012	2,250,020
SLA-DOA	Known	MHC class II, DO α	2,266,309	2,270,108
Extended class II				
COL11A2	Known	Collagen, type X1, α2	2,292,167	2,322,532
RXRB	Known	Retinoid X receptor, β	2,323,649	2,330,469
SLC39A7	Known	Solute carrier family 39 (zinc transporter), member 7	2,330,633	2,335,036
HSD17B8	Known	Hydroxysteroid (17-β)dehydrogenase 8	2,335,249	2,337,415
RING1	Known	Ring finger protein 1	2,339,131	2,343,170

Loci for which unambiguous orthology could not be established to the corresponding region of human chromosome 6 are denoted in bold.

investigation [19,20]. A further difference worth noting is the absence of functional copies of the *PSORS1C1* and *CDSN* genes within the SLA class I region, both of which are implicated in human susceptibility to psoriasis [21,22]. A short fragment of sequence (between *CCHCR1* and *PSORS1C2*) that represents a possible remnant of porcine *PSORS1C1* has been previously described by Shigenari et al. [10] but falls below the annotation criteria used here and, therefore, has not been included in the gene list. The predicted CDS of porcine *CDSN* contains stop codons and frameshifts compared to other mammalian CDSN proteins and has been classified here as a pseudogene. However, there are two pig ESTs (Em:BG384333.1 and Em:BX918433.1)

Table 3
List of noncoding RNA genes within the SLA detected using RFAM analysis, detailing RNA type, RFAM accession, coordinates within the SLA, and orientation

Type/name	RFAM Accession No.	Start coordinate	End coordinate	Orientation	Score
U6 snRNA	RF00026	151,049	151,146	+	81.88
tRNA	RF00005	256,185	256,258	+	25.59
tRNA	RF00005	582,469	582,541	+	30.65
tRNA	RF00005	894,971	895,043	+	29.72
tRNA	RF00005	992,724	992,795	–	30.13
U83 snoRNA	RF00137	1,119,821	1,119,897	–	42.97
U83 snoRNA	RF00137	1,123,961	1,124,038	–	70.09
snoACA38	RF00428	1,198,925	1,199,055	+	90.25
U48 snoRNA	RF00282	1,390,871	1,390,933	+	26.86
U52 snoRNA	RF00276	1,391,535	1,391,601	+	37.49
tRNA	RF00005	1,469,339	1,469,410	–	26.22
5_8S_rRNA	RF00002	1,738,803	1,738,851	+	24.79
5_8S_rRNA	RF00002	1,786,838	1,786,886	+	25.50
tRNA	RF00005	2,271,734	2,271,805	+	30.58
mir-219	RF00251	2,338,481	2,338,552	+	46.71

The scores are bits (logs-odds) scores which represent the log(2) of the probability of the query given the model over the probability of random sequence given the model.

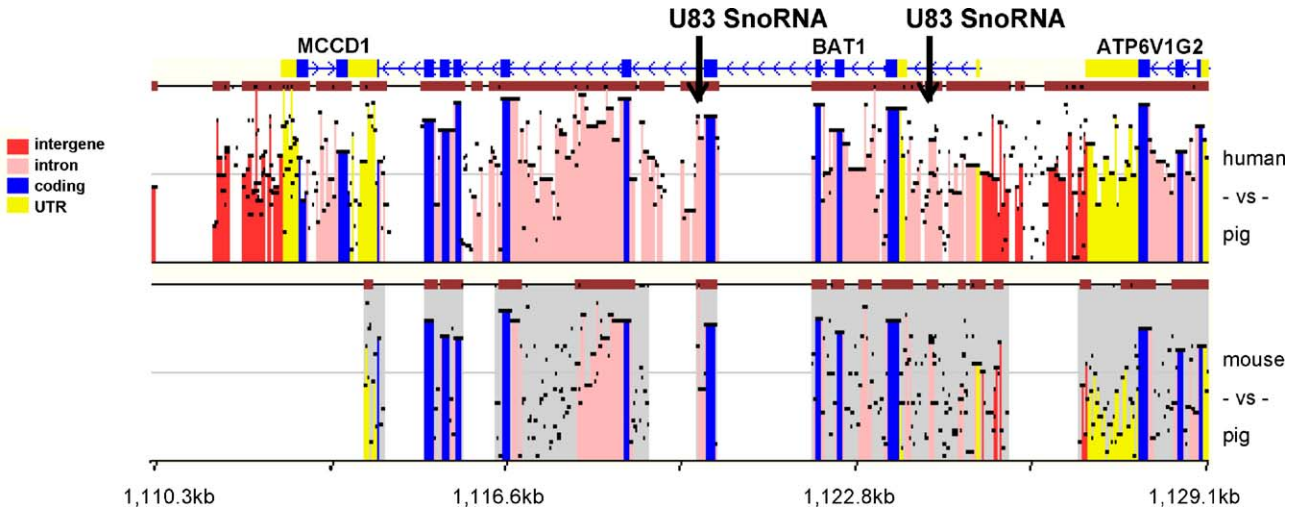

Fig. 2. Correlation of conserved noncoding sequences with RNA genes. Percentage identity plot performed using Z-PICTURE (see Materials and methods) illustrating the location of two U83 snoRNAs within conserved intron sequences of *BAT1*.

within the EMBL database that overlap pseudoexons 1 and 2 of *CDSN,* suggesting *CDSN* might be a transcribed pseudogene; a novel transcript (*SBAB-499E6.10*) has been included alongside *CDSN* to represent this.

Segment 2 maps to the central part of the class III region and involves four genes (*C4B, CYP21A2, TNXA,* and *STK19P*) that are collectively known as the RCCX module. While the copy number and gene status of this module vary depending upon haplotype in human [12] and in mice [23], no variation has been observed in a number of porcine haplotypes studied to date [24]. In rat, an additional module has translocated to the border of the class II region between *NOTCH4* and the *BTNL* cluster [25].

Segment 3 includes the centromere and will be discussed separately. Segment 4 maps to the telomeric end of the class II region and implicates eight genes, including all the *HLA-DP* loci. Functional loss, but not gene loss, of *HLA-DP* has also been observed in cat, which lacks *HLA-DQ* as well [26]. In both species, the loss of *HLA-DP* and *-DQ* (cat only) appears to be compensated for by an expansion of the *HLA-DR* gene family equivalent (Fig. 3). The SLA haplotype (H01) sequenced here contains one *SLA-DRA* gene and five *SLA-DRB* loci, although only *DRB1* is full length. *DRB4* has a deletion in exon 1. Exons 1 and 6 are missing in *DRB3,* while only exon 6 is missing in *DRB5* and *DRB2.* Four of five *DRB* loci are oriented and clustered in a pattern similar to that of other mammals; the remaining *SLA-DRB5* locus lies on the opposite strand within the *DQ–DO* interval. The *SLA-DRB5* gene is orthologous to *DLA-DRB2* and *FLA-DRB4* but no such *DRB* relic has been identified in the human or macaque *DQ* segment [27]. The *SLA-DQ* region comprises one *DQA* locus and two *DQB* loci of which only one is functional. We cannot exclude that the number of *DRB* and *DQB* copies could vary between different SLA haplotypes, as observed in the HLA [12]. The SLA *DQ–DO* interval also contains a putative locus (*SBAB-554F3.8*) and three pseudogenes (*SBAB-554F3.9, SLA-DOB2, SLA-DYB*) with similarities to *TAP2, DO,* and artiodactyl-specific *DYB,* respectively. In cattle and sheep, the class II *DQ–DO* interval is split into two subregions (separated by 17–30 cM) [28–30], giving rise to two loci—*DYA* and *DYB*—that are thought to have evolved from *DQ* [31]. Although there are some remnant matches on the DNA level, these are not sufficient to support the annotation of a porcine *DYA* locus according to the criteria used here. There is, however, supporting evidence for the presence of a *DYB* pseudogene, consisting of a two-exon fragment that shares similarity with predicted exons 2 and 3 of *Bos taurus DYB*. The SLA *DQ–DO* interval, with its heterogeneous set of pseudogenes, will be important in studies into the evolution of the artiodactyl MHC when the BoLA and OLA regions are fully sequenced. It is becoming increasingly clear that this subregion plays a significant role in the evolutionary divergence of mammalian lineages.

Centromere position within SLA

Among all the MHC regions studied to date [32], the porcine MHC is unique in that the class II region is separated from the class III and class I regions by the centromere [6]. Centromeres are specialized chromosomal regions of highly repetitive DNA that are defined cytogenetically as a dark-staining, heterochromatic structure within a chromosome. Centromere repositioning is a well-documented phenomenon in mammalian evolution [33]. The abundant repeat content of predominantly LINE elements between the class III and the class II regions in pig suggests the emergence of a neocentromere rather than translocation of an ancient centromere [34]. This hypothesis is supported by findings of the above study into the repositioning of centromeres in primates, which indicated that the position of a centromere could change radically over short periods of evolutionary time [33]. By sequencing the two BACs mapping closest to the centromere on both the short and the long arms of SSC7, we were able to determine the exact site of the centromere within the SLA.

The SSC7 centromere maps to a region that is extremely repeat-rich in human [35], is expanded in mouse [36], and spans

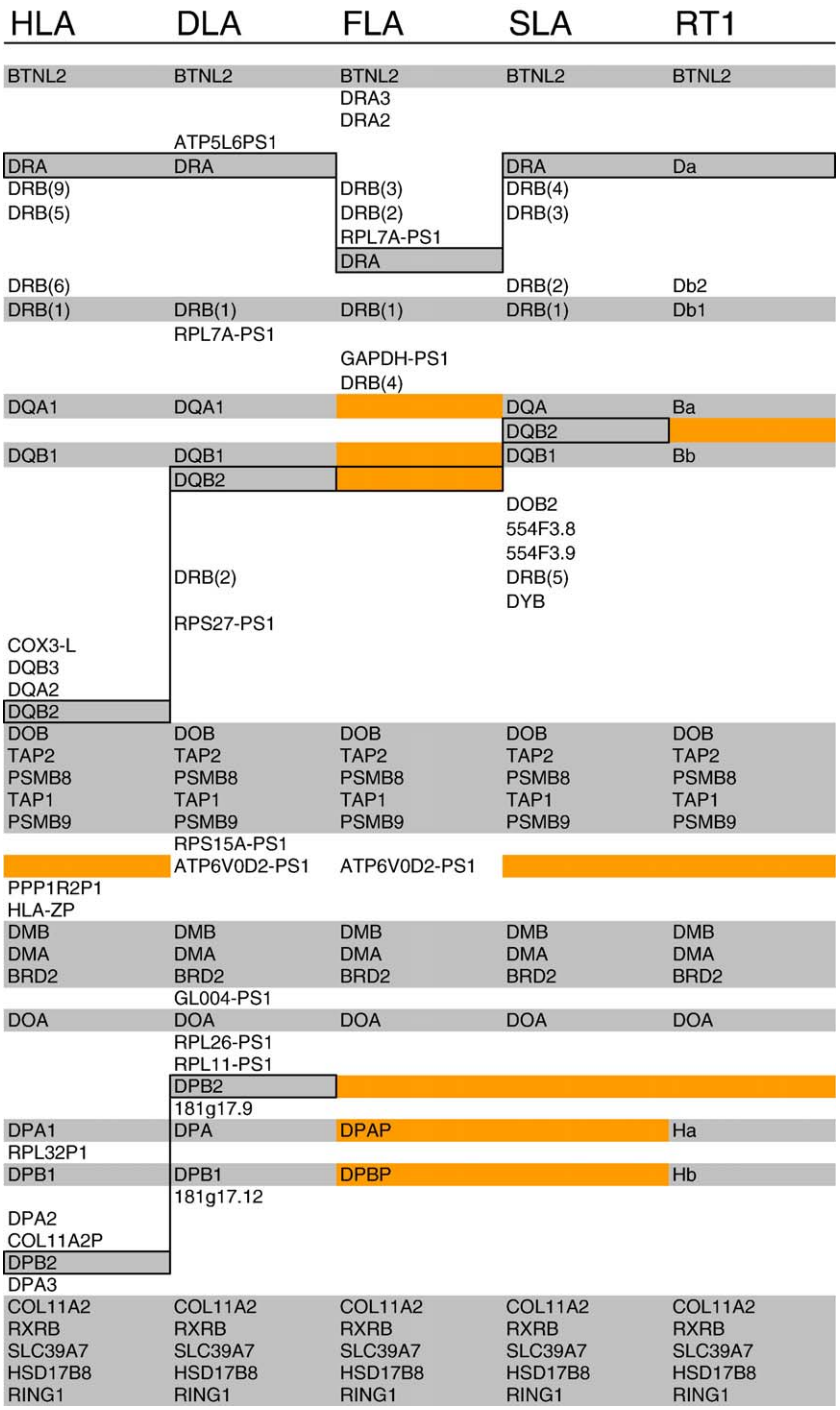

Fig. 3. Framework map of SLA class II region. The map shows a comparison of gene content within human (HLA), dog (DLA), cat (FLA), pig (SLA), and rat (RT1) class II regions. Orthologous framework loci conserved across two or more species are shaded in gray, whereas absent framework loci are highlighted in orange. The orange shading of *DPAP* and *DPBP* reflects functional rather than sequence loss in FLA. For the *DRB* loci, the numbers in parentheses indicate arbitrary copy numbers and do not reflect orthology. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a thus far uncloned gene desert of about 170 kb in horse [16]. Within the SLA, it is positioned between duplicated butyrophilin-like (*BTNL)* genes, in a G + C-poor region rich in tandem repeats (Fig. 1). In human, this corresponds to the region between *NOTCH4* and *BTNL2* that contains three loci (*C6orf10* (formerly *TSBP*), *HNRPA1P2,* and *HCG23*), all of which appear to be missing in SLA, possibly as a result of the centromere repositioning. Screening of the porcine BAC library with a *C6orf10*-specific probe failed to identify any positive clones. The position of the class II contig on the long arm of SSC7 has been confirmed by FISH and RH mapping [37].

Fig. 4 outlines the genomic organization of the SLA *BTNL* gene cluster flanking the centromere and compares it to the orthologous regions in human, dog, rat, and mouse. As illustrated in Fig. 4a, the five *BTNL* loci in the SLA differ somewhat in their number of exons and domains and, therefore,

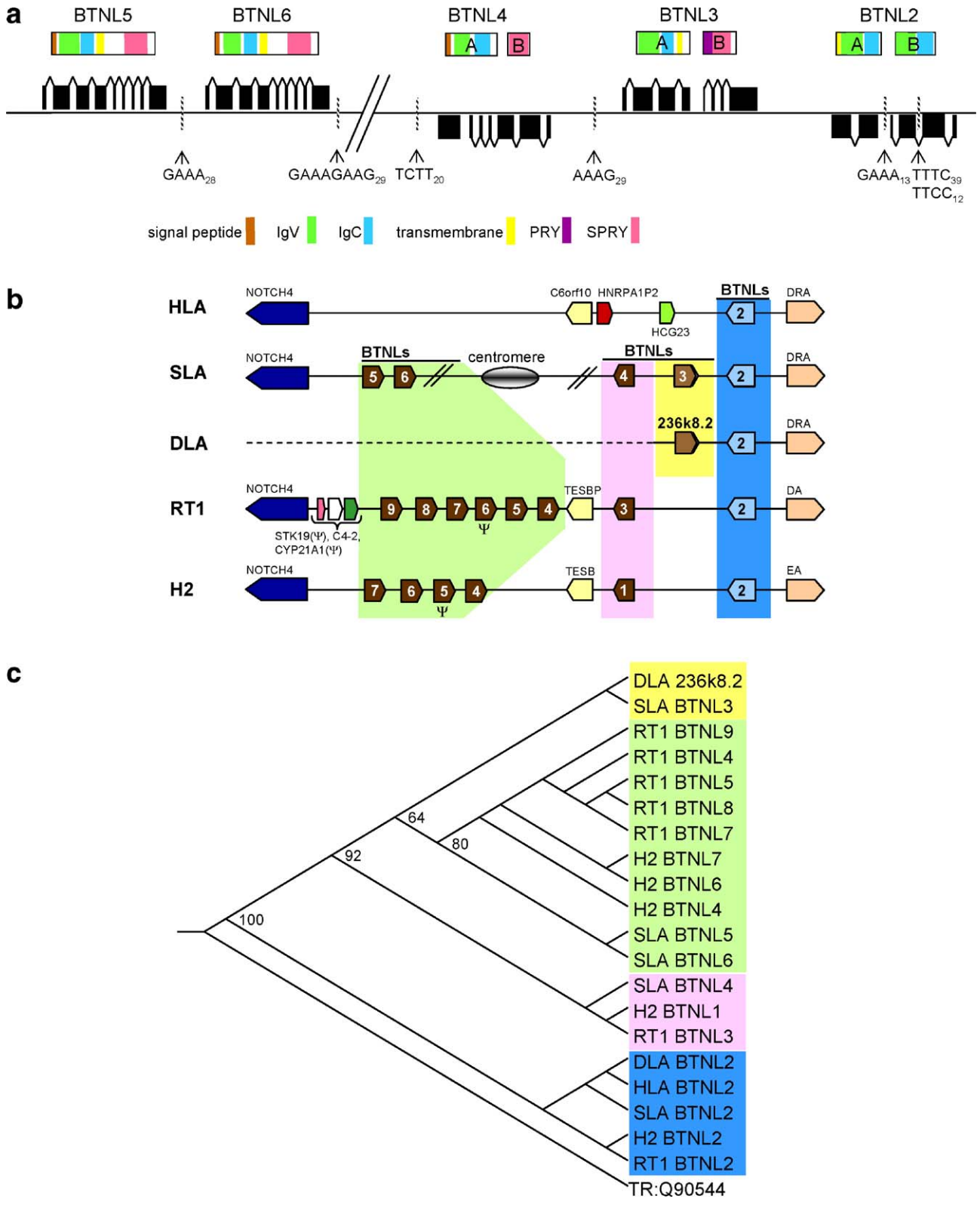

Fig. 4. Genomic organization and evolution of *BTNL* genes. (a) Schematic diagram illustrating the spatial organization, orientation, and exon structure of the *BTNL* loci within the SLA class II region (note that the introns are not shown to scale) and the domain architecture of BTNL proteins predicted by the translated sequence. (b) Schematic comparison of *BTNL* gene organization of human (HLA), pig (SLA), dog (DLA), rat (RT1), and mouse (H2). (c) Phylogenetic tree illustrating the relationships between class II BTNL proteins. Rat Btnl6 and mouse Btnl5 were excluded from the analysis because they are pseudogenes. The tree was rooted using TR: Q90544, a novel Ig domain-containing receptor from *Ginglymostoma cirratum* (nurse shark). Bootstrap values (% of 500 iterations) are shown for the nodes defining distinct BTNL lineages that have been coded with the same colors as in panels b and c. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are likely to have diverged since duplicating. Except for *BTNL2,* which is a known gene locus conserved in most mammals (the rhesus monkey [27] being a notable exception), the other four *BTNL* loci are each classified as novel CDS (see Materials and methods for details). Porcine *BTNL2, BTNL3,* and *BTNL4* have been annotated as two nonoverlapping fragments, A and B, as there is insufficient cDNA and protein evidence to define the exon/intron boundaries of the two fragments with confidence. All of the *BTNL* loci (including *BTNL2A* and *BTNL2B*) encode one set of IgV and IgC-like domains but differ in their other domains. For instance, *BTNL2* does not encode any SPRY domains that are common to the other *BTNL* loci and *BTNL3* encodes an additional PRY domain. With respect to the *BTNL* loci, the genomic organization of the SLA is more similar to that of rodents than to that of human (Fig. 4b).

The evolution of the class II and III *BTNL* loci is complex, as shown in the phylogenetic tree in Fig. 4c. Of the 22 *BTNL* loci shown in Fig. 4b, 20 loci were used for the phylogenetic analysis (rat *Btnl6* and mouse *Btnl5* were excluded because of their pseudogene status). According to this analysis, the *BTNL* loci can be grouped into four distinct lineages of which the *BTNL2* lineage (blue) is the most ancestral. *BTNL2* is also the only locus conserved in all the species studied here and therefore is likely to have orthologous function in these species. It is also the only *BTNL* locus in the corresponding human region [36] although further *BTNL* loci are present in the extended human class I region [38]. The second lineage (red) is defined by mouse *Btnl1,* rat *Btnl3,* and pig *BTNL4,* which also are likely to represent true orthologues. The third lineage (yellow), comprising pig *BTNL3* and dog *236k8.2,* appears to be specific to dogs and pigs only. Distinctive by their intracytoplasmic PRY and SPRY motifs, the predicted SLA BTNL3 and DLA 236k8.2 proteins are absent in rodents and humans. Thus, SLA *BTNL3* is predicted to be the orthologue of *236k8.2* in the DLA [39]. Finally, the most recent and largest lineage (green) consists of SLA *BTNL5* and *BTNL6* and 10 (8 plus 2 pseudogenes) paralogous copies of rodent *Btnl* genes. No orthologues could be assigned within this lineage. The expansion of *BTNL* genes observed in rodents, pigs, and possibly dogs is absent in the HLA in which deletion and/or translocation has occurred to give rise to another *BTNL* cluster in the extended class I region [38]. Despite their abundance, the function of BTNL family genes is not yet well established. The human *BTN1A1* locus has been implicated in milk droplet secretion and stability in human and mouse [40]. Recent data suggest that *BTNL2* plays a role as a costimulatory molecule involved in T cell activation, thus could be involved in immune response pathways [41].

Conclusion

The gene map and comparative analysis of the porcine MHC reported here can be expected to stimulate biomedical research into disease resistance and general well-being of farm animals and advance our understanding of the structure, function, and evolution of this complex region. Our analysis further confirms and extends previous observations that the MHC is a mosaic of highly conserved regions interspersed with highly plastic subregions that have undergone species-specific adaptation [42].

With respect to plasticity, our data confirm previous observations in rodents and primates that the class I region is the most dynamic region of the MHC. However, only one (albeit large) deletion and two moderate gene expansions could be identified in the SLA in comparison with the HLA class I region, suggesting limited evolution of the region compared to, e.g., rhesus monkey [27], mouse [17], and rat [25], in which multiple blocks of expansion have been observed. A similar trend was also observed in the class II region. Despite the additional expansion of pseudogenes within the *DO–DQ* interval, the pig class II region is shortened (compared to human) by the loss of *DP* loci. While the porcine *DR* region has undergone a limited expansion, giving rise to five *DRB* loci, only one of these is predicted to be functional. Compared to rodents, in which the scope for gene expansion is much larger, the generation of novel functional variants in pigs is reduced through effects of domestication and longer generation times. Polymorphism studies in pigs and other wild and domesticated artiodactyls will be necessary to advance this line of research.

The role of LINEs, SINEs, and other repeats in chromosome dynamics including centromere repositioning or the creation of neocentromeres is well documented [33,34,43,44]. Within the HLA, the highest density of such repeats is found in an area separating the class III and II regions [35,36] and, therefore, it is perhaps not surprising that the SSC7 centromere maps to the homologous region in pig. What is remarkable, however, is that the insertion of the (neo)centromere within the *BTNL* gene cluster does not seem to confound MHC function and, in fact, resulted in little disturbance of the molecular organization of the pig MHC. As far as we can tell from our analysis, only a few genes (*C6orf10* and perhaps some *BTNLs*) were possibly lost in the process. This is consistent with the observation of few disruptions resulting from centromere repositioning in other lineages [33]. In addition, it is known from studies in teleosts such zebrafish that the MHC can be fragmented without confounding function [45]. Despite the interruption by the centromere, the overall molecular organization of the pig MHC therefore remains similar to that of other mammals.

With the International Pig Genome Project (http://www.ncbi.nlm.nih.gov/projects/genome/guide/pig) gathering pace and recent publication of a shotgun survey of the pig genome [46], the high-quality finished sequence of the porcine MHC reported here represents a valuable reference sequence to guide future assemblies of this important genome.

Materials and methods

Mapping

The SBAB genomic pig BAC library was constructed from a Large White boar, SLA homozygous for the haplotype H01 [47]. A total of 158 BACs were

isolated and mapped as described previously [48], resulting in two contigs covering the SLA complex from which a minimum tile path of 23 clones was selected for sequencing. A breakdown of the individual BAC clones and their accession numbers is provided in Table 1. The overlap between BACs 207G8 and 490B10 and 499E6 and 493A6 was confirmed by PCR sequencing.

Sequencing and analysis

The SLA class I region was sequenced by INRA (France), Genoscope (France), and Tokai University (Japan) and the class II and III regions were sequenced by the Wellcome Trust Sanger Institute (United Kingdom), who also annotated the entire region. Subcloning and sequencing were performed using known procedures in operation at the time at each institute. For the analysis, we used a combination of BLAST [49] (http://www.ncbi.nlm.nih.gov/), DOTTER [50], PIPMAKER [51] (http://pipmaker.bx.psu.edu/pipmaker/), and Z-PICTURE [52] (http://zpicture.dcode.org/).

Sequence annotation

Manual annotation was uniformly performed on the entire SLA sequence by the Wellcome Trust Sanger Institute Havana team as follows: The finished porcine genomic sequence was analyzed using an automatic Ensembl pipeline [53] with modifications to aid the manual curation process. The G + C content of each clone sequence was analyzed and putative CpG islands were marked. Interspersed repeats were detected using RepeatMasker using the mammalian library along with porcine-specific repeats submitted to EMBL/NCBI/DDBJ and simple repeats using Tandem Repeats Finder [54]. The combination of the two repeat types was used to mask the sequence. This masked sequence was searched against vertebrate cDNAs and expressed sequence tags (ESTs) using WU-BLASTN and matches were cleaned up using EST2_GENOME. A protein database combining nonredundant data from SwissProt and TrEMBL was searched using WU-BLASTX. Ab initio gene structures were predicted using FGENESH and GENSCAN. The predicted gene structures were manually annotated according to the human annotation workshop guidelines (http://www.sanger.ac.uk/HGP/havana/hawk.shtml). The gene categories used were as described on the VEGA Web site [11] (http://vega.sanger.ac.uk/). *Known* genes are identical to known pig cDNAs or protein sequences or are orthologues of known human loci. *Novel CDS* loci have an open reading frame (ORF), are identical to spliced ESTs, or have some similarity to other genes or proteins. *Novel transcript* is similar to a novel CDS, but no ORF can be determined unambiguously. *Putative* genes are identical to spliced pig ESTs, but do not contain an ORF. *Pseudogenes* are nonfunctional copies of known or novel loci.

Phylogenetic analysis

Multiple sequence alignments of the butyrophilin IgV and IgC protein domains (220–226 amino acids) were constructed using CLUSTALW [55] (http://www.ebi.ac.uk/clustalw/). PHYLIP [56] was used to estimate protein distances (Jones–Tayor–Thornton model) and construct consensus trees using the neighbor-joining method [57] with 500 bootstrap replicates. MEGA version 3.1 [58] was used to present trees graphically.

Acknowledgments

The work carried out at the Sanger Institute was supported by the Wellcome Trust. We thank all members of the DNA Sequencing Division at the Sanger Institute and Jennifer Sambrook for assistance with phylogenetic trees. The work carried out at the INRA was supported by Genoscope–INRA funding. The work carried out at Tokai University was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan and the Animal Genome Research Project of the Ministry of Agriculture, Forestry, and Fisheries of Japan.

References

[1] J.P. Bidanel, et al., Detection of quantitative trait loci for growth and fatness in pigs, Genet. Sel. Evol. 33 (2001) 289–309.
[2] J.S. Logan, Prospects for xenotransplantation, Curr. Opin. Immunol. 12 (2000) 563–568.
[3] L. Wang, T.P. Yu, C.K. Tuggle, H.C. Liu, M.F. Rothschild, A directed search for quantitative trait loci on chromosomes 4 and 7 in pigs, J. Anim. Sci. 76 (1998) 2560–2567.
[4] Y.G. Yang, Application of xenogeneic stem cells for induction of transplantation tolerance: present state and future directions, Springer Semin. Immunopathol. 26 (2004) 187–200.
[5] M. Vaiman, P. Chardon, M.F. Rothschild, Porcine major histocompatibility complex, Rev. Sci. Tech. 17 (1998) 95–107.
[6] P. Chardon, C. Renard, M. Vaiman, The major histocompatibility complex in swine, Immunol. Rev. 167 (1999) 179–192.
[7] P. Chardon, et al., Sequence of the swine major histocompatibility complex region containing all non-classical class I genes, Tissue Antigens 57 (2001) 55–65.
[8] C. Renard, et al., Sequence of the pig major histocompatibility region containing the classical class I genes, Immunogenetics 53 (2001) 490–500.
[9] 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.
[10] A. Shigenari, et al., 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.
[11] J.L. Ashurst, et al., The Vertebrate Genome Annotation (VEGA) database, Nucleic Acids Res. 33 (2005) D459–D465.
[12] R. Horton, et al., Gene map of the extended human MHC, Nat. Genet. Rev. 5 (2004) 889–899.
[13] S. Griffiths-Jones, et al., Rfam: annotating non-coding RNAs in complete genomes, Nucleic Acids Res. 33 (2005) D121–D124.
[14] B.E. Jady, T. Kiss, Characterisation of the U83 and U84 small nucleolar RNAs: two novel 2′-O-ribose methylation guide RNAs that lack complementarities to ribosomal RNAs, Nucleic Acids Res. 28 (2000) 1348–1354.
[15] T.W. Beck, et al., The feline major histocompatibility complex is rearranged by an inversion with a breakpoint in the distal class I region, Immunogenetics 56 (2005) 702–709.
[16] A.L. Gustafson, et al., An ordered BAC contig map of the equine major histocompatibility complex, Cytogenet, Genome Res. 102 (2003) 189–195.
[17] J.K. Kulski, T. Shiina, T. Anzai, S. Kohara, H. Inoko, Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man, Immunol. Rev. 190 (2002) 95–122.
[18] A. Kumanovics, T. Takada, K.F. Lindahl, Genomic organization of the mammalian MHC, Annu. Rev. Immunol. 21 (2003) 629–657.
[19] F. Di Palma, S.D. Archibald, J.R. Young, S.A. Ellis, A BAC contig of approximately 400 kb contains the classical class I major histocompatibility complex (MHC) genes of cattle, Eur. J. Immunogenet. 29 (2002) 65–68.
[20] R.D. McShane, et al., Physical localization and order of genes in the class I region of the bovine MHC, Anim. Genet. 32 (2001) 235–239.
[21] A. Oka, et al., Association analysis using refined microsatellite markers localizes a susceptibility locus for psoriasis vulgaris within a 111 kb segment telomeric to the HLA-C gene, Hum. Mol. Genet. 8 (1999) 2165–2170.
[22] R. Tazi Ahnini, et al., Novel genetic association between the corneodesmosin (MHC S) gene and susceptibility to psoriasis, Hum. Mol. Genet. 8 (1999) 1135–1140.
[23] T. Xie, et al., Analysis of the gene-dense major histocompatibility complex class III region and its comparison to mouse, Genome Res. 13 (2003) 2621–2636.
[24] C. Geffrotin, C. Renard, P. Chardon, M. Vaiman, Marked genetic-polymorphism of the swine steroid 21-hydroxylase gene, and its location between the SLA class-I and class-II regions, Anim. Genet. 22 (1991) 311–322.
[25] P. Hurt, et al., The genomic sequence and comparative analysis of the

rat major histocompatibility complex, Genome Res. 14 (2004) 631–639.
[26] N. Yuhki, et al., Comparative genome organization of human, murine, and feline MHC class II region, Genome Res. 13 (2003) 1169–1179.
[27] R. Daza-Vamenta, G. Glusman, L. Rowen, B. Guthrie, D.E. Geraghty, Genetic divergence of the rhesus macaque major histocompatibility complex, Genome Res. 14 (2004) 1501–1515.
[28] M. Amills, V. Ramiya, J. Norimine, H.A. Lewin, The major histocompatibility complex of ruminants, Rev. Sci. Tech. 17 (1998) 108–120.
[29] V.L. Jarrell, H.A. Lewin, Y. Da, M.B. Wheeler, Gene-centromere mapping of bovine DYA, DRB3, and PRL using secondary oocytes and first polar bodies—Evidence for 4-strand double crossovers between DYA and DRB3, Genomics 27 (1995) 33–39.
[30] H. Wright, K.T. Ballingall, Mapping and characterization of the Dq subregion of the ovine MHC, Anim. Genet. 25 (1994) 243–249.
[31] 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.
[32] J. Kelley, L. Walter, J. Trowsdale, Comparative genomics of major histocompatibility complexes, Immunogenetics 56 (2005) 683–695.
[33] V. Eder, et al., Chromosome 6 phylogeny in primates and centromere repositioning, Mol. Biol. Evol. 20 (2003) 1506–1512.
[34] D.J. Amor, et al., Human centromere repositioning "in progress", Proc. Natl. Acad. Sci. USA 101 (2004) 6542–6547.
[35] The MHC Sequencing Consortium, Complete sequence and gene map of a human major histocompatibility complex, Nature 401 (1999) 921–923.
[36] M. Stammers, L. Rowen, D. Rhodes, J. Trowsdale, S. Beck, BTL-II: a polymorphic locus with homology to the butyrophilin gene family, located at the border of the major histocompatibility complex class II and class III regions in human and mouse, Immunogenetics 51 (2000) 373–382.
[37] P. Chardon, Physical organization of the pig major histocompatibility complex class II region, Immunogenetics 50 (1999) 344–348.
[38] D.A. Rhodes, M. Stammers, G. Malcherek, S. Beck, J. Trowsdale, The cluster of BTN genes in the extended major histocompatibility complex, Genomics 71 (2001) 351–362.
[39] S.L. Debenham, et al., Genomic sequence of the class II region of the canine MHC: comparison with the MHC of other mammalian species, Genomics 85 (2005) 48–59.
[40] J.L. McManaman, C.A. Palmer, S. Anderson, K. Schwertfeger, M.C. Neville, Regulation of milk lipid formation and secretion in the mouse mammary gland, Adv. Exp. Med. Biol. 554 (2004) 263–279.
[41] R. Valentonyte, et al., Sarcoidosis is associated with a truncating splice site mutation in BTNL2, Nat. Genet. 37 (2005) 357–364.
[42] C. Amadou, Evolution of the MHC class I region: the framework hypothesis, Immunogenetics 49 (1999) 362–367.
[43] E.E. Eichler, D. Sankoff, Structural dynamics of eukaryotic chromosome evolution, Science 301 (2003) 793–797.
[44] F.S. Kaplan, et al., The topographic organization of repetitive DNA in the human nucleolus, Genomics 15 (1993) 123–132.
[45] J.G. Sambrook, F. Figueroa, S. Beck, A genome-wide survey of major histocompatibility complex (MHC) genes and their paralogues in zebrafish, BMC Genom. 6 (2005) 152.
[46] R. Wernersson, Pigs in sequence space: a 0.66× coverage pig genome survey based on shotgun sequencing, BMC Genom. 6 (2005) 70.
[47] C. Rogel-Gaillard, N. Bourgeaux, A. Billault, M. Vaiman, P. Chardon, 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.
[48] F.W. Velten, et al., Spatial arrangement of pig MHC class I sequences, Immunogenetics 49 (1999) 919–930.
[49] S. Schwartz, et al., Human–mouse alignments with BLASTZ, Genome Res. 13 (2003) 103–107.
[50] E.L. Sonnhammer, R. Durbin, A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis, Gene 167 (1995) 1–10.
[51] S. Schwartz, et al., PipMaker—A Web server for aligning two genomic DNA sequences, Genome Res. 10 (2000) 577–586.
[52] I. Ovcharenko, G.G. Loots, R.C. Hardison, W. Miller, L. Stubbs, zPicture: dynamic alignment and visualization tool for analyzing conservation profiles, Genome Res. 14 (2004) 472–477.
[53] S.C. Potter, The Ensembl analysis pipeline, Genome Res. 14 (2004) 934–941.
[54] G. Benson, Tandem Repeats Finder: a program to analyze DNA sequences, Nucleic Acids Res. 27 (1999) 573–580.
[55] R. Chenna, Multiple sequence alignment with the Clustal series of programs, Nucleic Acids Res. 31 (2003) 3497–3500.
[56] J. Felsenstein, PHYLIP—Phylogeny Interference Package (version 3.2), Cladistics 5 (1989) 164–166.
[57] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.
[58] S. Kumar, K. Tamura, M. Nei, MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment, Brief Bioinform. 5 (2004) 150–163.
[59] A. Ando, et al., 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.