Immunogenetics (2005) 56: 773–777
DOI 10.1007/s00251-004-0744-4

BRIEF COMMUNICATION

Christian Roos · Lutz Walter

Considerable haplotypic diversity in the *RT1-CE* class I gene region of the rat major histocompatibility complex

Received: 21 August 2004 / Revised: 25 October 2004 / Published online: 1 December 2004

Abstract The major histocompatibility complex (MHC) class I region extending between the *Bat1* and *Pou5f1* genes shows considerable genomic plasticity in mouse and rhesus macaque but not in human haplotypes. In the rat, this region is known as the *RT1-CE* region. The recently published rat MHC sequence gave rise to a complete set of class I gene sequences in a single MHC haplotype, namely the $RT1^n$ haplotype of the widely used BN inbred strain. To study the degree of genetic diversity, we compared the *RT1-CE* region-derived class I genes of the $RT1^n$ haplotype with class I sequences of other rat haplotypes. By using phylogenetic tree analyses, we obtained evidence for extensive "presence and absence" polymorphisms of single loci and even small subfamilies of class I genes in the rat. Alleles of *RT1-CE* region class I genes could also be identified, but the rate of allelic nucleotide substitutions appeared rather low, indicating that the diversity in the *RT1-CE* region is mainly based on genomic plasticity.

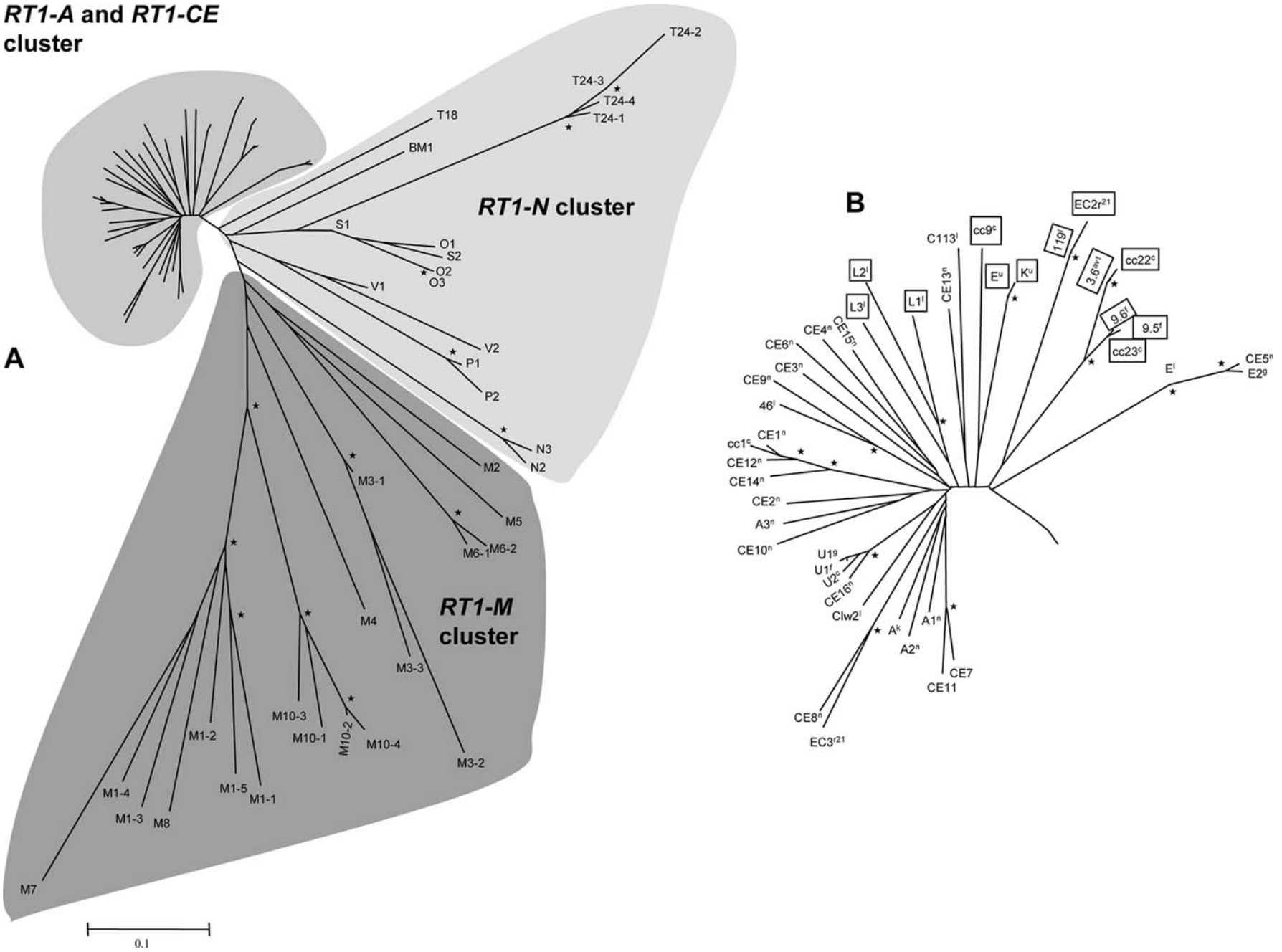
Keywords MHC · Rat class I genes · Gene tree analysis · Haplotypic diversity · Genomic plasticity

The laboratory rat (*Rattus norvegicus*) is a widely used experimental animal that provides important and well-established models for many human diseases, particularly infectious and autoimmune diseases (Günther 1999). Besides its central role in the control of immune responses against pathogens, the major histocompatibility complex (MHC) is also involved in the susceptibility of an individual to develop an autoimmune disease (Thorsby 1997). While many autoimmune diseases have been linked to MHC class II genes, e.g. the *HLA-DQB* gene in type I diabetes, associations are also found for class I genes, the most prominent being *HLA-B*2705* and ankylosing spondylitis (Hülsmeyer et al. 2004). The genetic basis of associations between the MHC and autoimmune diseases can be well demonstrated by using MHC congenic and intra-MHC recombinant congenic mouse or rat strains. Thus, it could be shown that in MOG-induced experimental autoimmune encephalomyelitis (EAE)—an animal model of multiple sclerosis —the rat MHC determines the degree of susceptibility and the clinical course of the disease in a haplotype-specific manner (Weissert et al. 1998). Recently, the rat MHC, the *RT1* complex, has been completely sequenced (Hurt et al. 2004). This sequence is based on a single *RT1* haplotype, namely the $RT1^n$ haplotype of the BN rat strain. To study the extent of genomic diversity in the *RT1-CE* class I region, which corresponds genomically to the *HLA-B* region in humans and the *H2-D/L/Q* region in mice, we aligned the *RT1-CE* region class I genes of the $RT1^n$ haplotype with class I sequences of other *RT1* haplotypes.

Rat MHC class I sequences were extracted from the DDBJ/GenBank/EMBL database and aligned using the Clustal X software (Thompson et al. 1997). Gene trees were constructed with the neighbor-joining algorithm (Saitou and Nei 1987) as implemented in PAUP, version 4.0b10 (Swofford 2002), and MEGA, version 2.1 (Kumar et al. 2001). Support of branching was assessed by bootstrapping based on 500 replications. MEGA was also used for the analysis of synonymous and nonsynonymous nucleotide substitutions.

The genomic sequence of the rat MHC (Hurt et al. 2004) revealed a complete set of class I gene sequences from a single haplotype, namely $RT1^n$. This prompted us to compare these class I genes with previously known rat class I sequences derived from other *RT1* haplotypes to analyze the extent of haplotypic diversity. Initially, a gene tree analysis was carried out showing that the $RT1^n$ class I sequences group roughly according to their genomic position, i.e., genes located in the *RT1-A* and *RT1-CE* region, in the *RT1-N* region, and in the *RT1-M* region cluster together (Fig. 1a). We were particularly interested in the *RT1-CE* region that contains 16 class I genes in the

C. Roos · L. Walter (✉)
Department of Primate Genetics, German Primate Center, Göttingen, Germany
e-mail: lwalter@gwdg.de
Tel.: +49-551-3851161
Fax: +49-551-3851228


RT1-A and RT1-CE cluster
A
T24-2
T24-3
T24-4
T24-1
T18
BM1
RT1-N cluster
S1
O1
S2
O2
O3
V1
V2
P1
P2
N3
N2
M2
M3-1
M5
M6-1
M6-2
RT1-M cluster
M4
M3-3
M3-2
M10-3
M10-1
M10-2
M10-4
M1-2
M1-4
M1-3
M8
M1-5
M1-1
M7
0.1
B
EC2r21
C113l
cc9c
119l
3.6int1
cc22c
L2l
Eu
Ku
L3l
L1l
CE13n
9.6l
9.5f
CE4n
CE15n
cc23c
CE6n
CE3n
El
CE5n
E2g
CE9n
46l
CE1n
cc1c
CE12n
CE14n
CE2n
A3n
CE10n
U1g
U1f
U2c
CE16n
Clw2l
Ak
A1n
A2n
CE7
CE11
CE8n
EC3r21

◀ **Fig. 1** Gene tree analysis of rat class I exon 2 to exon 8 sequences. Only bootstrap values exceeding 95% are shown and are indicated by an *asterisk*. **a** The database accession numbers of the class I genes can be found in a comprehensive review by Günther and Walter (2001) except for BX511170 (*RT1-A1*^{n} to *RT1-A3*^{n}, *RT1-CE1*^{n} to *RT1-CE16*^{n}, all *RT1-N* and *RT1-M* region class I loci), AF457139 (*RT1-L1*^{l}), AY397759 (*RT1-L2*^{l}), AY445668 (*RT1-L3*^{l}), and AF387339 (*RT1-46*^{l}). **b** shows an enlarged view of the *RT1-A* and *RT1-CE* region sequences. Loci of other *RT1* haplotypes that do not cluster with the *RT1*^{n} haplotype loci are indicated by a *box*

RT1^n haplotype (Hurt et al. 2004), which are all of the nonclassical type. The *RT1-CE* region is located between the *Bat1* and *Pou5f1* genes (Günther and Walter 2001), a genomic interval that exhibits considerable diversity in the mouse (Kumanovics et al. 2002) and the rhesus macaque (Daza-Vamenta et al. 2004). Therefore, we included class I sequences of other rat haplotypes in the phylogenetic analysis that are most likely derived from the *RT1-CE* region and do not represent class I genes of the *RT1-N* or *RT1-M* type. As expected, these additional class I sequences cluster with *RT1-A* and *RT1-CE*, and not with *RT1-N* or *RT1-M* region genes (Fig. 1a, b). Furthermore, some of these sequences cluster significantly with certain *RT1-CE* genes of the *RT1*^{n} haplotype, suggesting allelism of these loci. As can be seen in Fig. 1b, the *cc1* cDNA clone derived from the *RT1*^{c} haplotype (Leong et al. 1999) clusters with the *RT1-CE1*^{n} gene and, therefore, they most likely represent alleles. Likewise, *RT1-E2*^{g} (Lau et al. 2003) and *RT1-CE5*^{n} as well as *RT1–46*^{l} (Lambracht-Washington and Fischer Lindahl 2002) and *RT1-CE9*^{n} appear to be allelic (Fig. 1b). A significant clustering of *RT1-CE16*^{n} with *RT1-U1*^{f} and *RT1-U2*^{c} could also be observed, evolving the question whether *RT1-U1* and *RT1-U2* are indeed different loci or represent alleles of the same locus, namely *RT1-CE16*. With respect to the sometimes confusing nomenclature of *RT1* class I genes, it should be noted that a definitive nomenclature is not yet available and we follow our proposal published recently (Hurt et al. 2004), which is based on a systematic designation according to the physical mapping of class I genes.

In contrast to the mouse, where the evolution of the *H2-D/L/Q* region class I genes could be well deduced from the genomic sequence (Kumanovics et al. 2002), an evolutionary reconstruction was not possible for the rat *RT1-CE* region (Hurt et al. 2004), even by inclusion of certain repetitive sequences. However, inspection of the neighbor-joining tree shown in Fig. 1b provided some clues to the evolution of at least some of the *RT1-CE* region class I genes. The *RT1-CE7*^{n} and *RT1-CE11*^{n} genes and the *RT1-CE1*^{n}, *RT1-CE12*^{n}, and *RT1-CE14*^{n} genes cluster significantly (Fig. 1b), suggesting that they arose by recent duplications from respective ancestral genes. A clustering (bootstrap value 86%) of the *RT1-A3*^{n} and *RT1-CE10*^{n} genes was also found (Fig. 1b) and confirms the evolutionary relatedness of the centromeric and telomeric

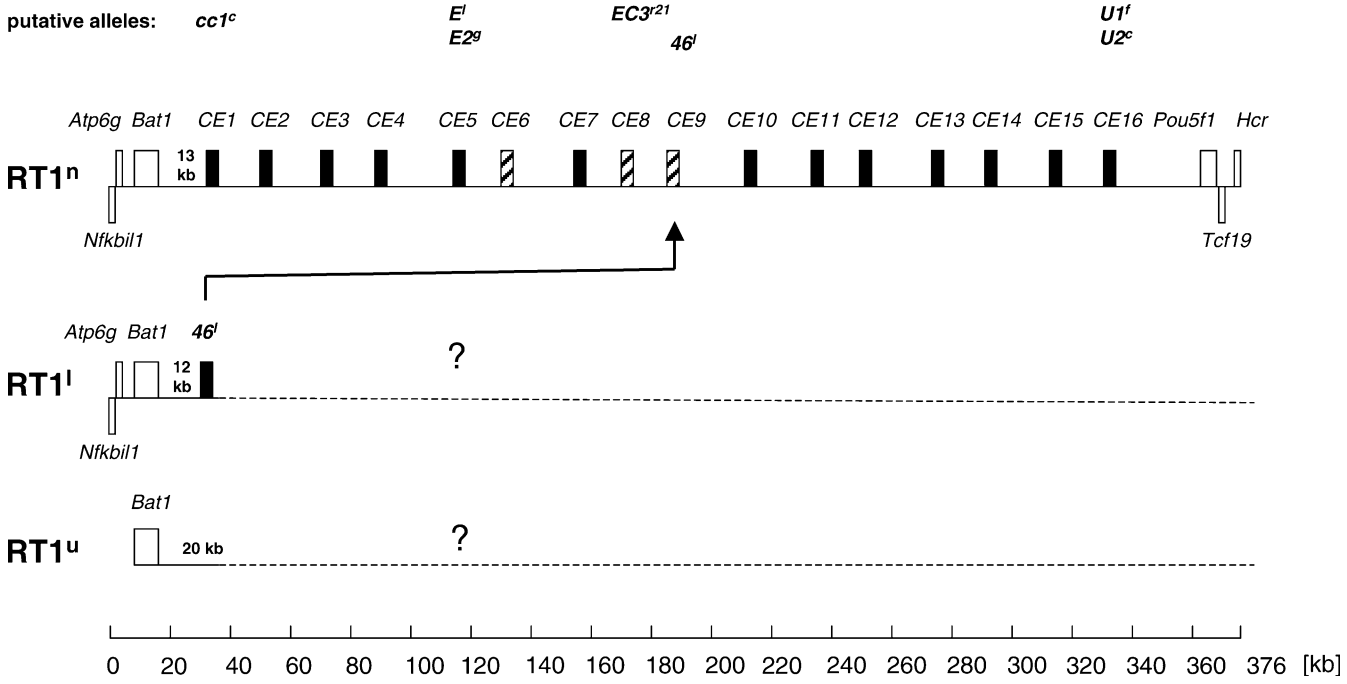


Fig. 2 Map of the *Bat1-Pou5f1* genomic interval that contains the *RT1-CE* region class I genes. The maps of the *RT1*^{n}, *RT1*^{l}, and *RT1*^{u} haplotypes are based on data reported by Hurt et al. (2004); Lambracht-Washington and Fischer Lindahl (2002) and Walter and Günther (1997), respectively. Functional and nonfunctional class I genes (Hurt et al. 2004) are indicated by *filled* and *hatched boxes*, respectively. The position of the box *above or below the line* represents the orientation of the gene. Putative alleles of the *RT1-CE* loci (according to the analysis shown in Fig. 1) are indicated. In the *RT1*^{u} haplotype no class I gene could be detected at 20-kb distance of the *Bat1* gene (Walter and Günther 1997). The *broken lines* and the *question marks* indicate that the *RT1*^{l} and *RT1*^{u} haplotypes are still continuing, but the genomic interval has not been characterized in detail by genomic sequencing or detailed physical mapping

class I regions *RT1-A* and *RT1-CE* (Lambracht-Washington et al. 2000; Hurt et al. 2004; Walter and Günther 2000).

Our finding that $RT1\text{-}CE9^n$ and $RT1\text{-}46^l$ are most likely alleles appears quite instructive. A cosmid sequence derived from the LEW rat ($RT1^l$ haplotype) has previously been reported by Lambracht-Washington and Fischer Lindahl (2002) showing that the $RT1\text{-}46^l$ maps adjacent to the *Bat1* gene at a distance of about 12 kb. Thus, we conclude that the $RT1^l$ haplotype lacks the *RT1-CE1* to *RT1-CE8* loci of the $RT1^n$ haplotype (Fig. 2). An alternative explanation might be that these loci were translocated to a different genomic region. In contrast, the loci $RT1\text{-}L1^l$, $RT1\text{-}L2^l$, and $RT1\text{-}L3^l$ which form a small subfamily of class I genes in the $RT1^l$ haplotype (Lambracht-Washington et al. 2004) do not cluster with any of the $RT1^n$-derived *RT1-CE* loci (Fig. 1b), indicating that the *RT1-L* subfamily either does not occur or got lost in the $RT1^n$ haplotype. Similarly, for the genes $RT1\text{-}3.6^{av1}$, $RT1\text{-}cc22^c$, $RT1\text{-}cc23^c$, $RT1\text{-}9.5^f$, $RT1\text{-}9.6^f$, $RT1\text{-}119^l$, $RT1\text{-}EC2^{r21}$, $RT1\text{-}E^u$, $RT1\text{-}K^u$, and $RT1\text{-}cc9^c$, no corresponding allele of the $RT1^n$ haplotype could be detected among the *RT1-CE* genes (Fig. 1b). It is not clear at the moment whether $RT1\text{-}C113^l$ and $RT1\text{-}Clw2^l$ represent alleles of $RT1\text{-}CE13^n$ and $RT1\text{-}CE16^n$, respectively, as bootstrap support for the clustering was low (48% and 56%, respectively). These data indicate that rat MHC haplotypes show considerable diversity with respect to absence and presence of certain class I loci in the class I region extending between *Bat1* and *Pou5f1*. Such genomic plasticity has also been reported for the *RT1-A* region that contains the classical class I genes of the rat (Walter and Günther 2000).

To study the level of allelic diversity, we compared the rates of synonymous (d_S) and nonsynonymous (d_N) substitutions in codons of the peptide-binding region (PBR) and non-PBR codons of exons 2 and 3 (Table 1). In contrast to the class Ia locus *RT1-A1*, which was included as a known example for positive selection, evidence for positive selection could neither be found for the expressed loci *RT1-CE1*, *RT1-CE5* and *RT1-CE16*, nor for *RT1-CE8* and *RT1-CE9*, which was expected as the latter two represent pseudogenes. These data further indicate that diversity of the *RT1-CE* region-derived class I genes is focused on presence and absence of certain loci. Interestingly, a similar mode of generating class I gene diversity in the corresponding genomic interval has recently been published for the mouse (Kumanovics et al. 2002) and the rhesus macaque (Daza-Vamenta et al. 2004), adding further impact on the "birth and death" model of MHC evolution (Klein et al. 1998). This diversity has to be taken into account when these species are used as animal models for human MHC-associated diseases, as most of these class I loci are expressed, or at least appear expressible from inspection of their genomic sequences.

Table 1 Mean number of synonymous [(d_S)±SE] and nonsynonymous [(d_N)±SE] substitutions per 100 sites in codons of the peptide-binding region (*PBR*) and remaining codons (*non-PBR*) of exon 2 and exon 3

Comparison	d_S PBR[a]	d_N PBR	d_S Non-PBR	d_N Non-PBR
$CE1^n+cc1^c$	0.0±0.0	0.8±0.8	0.0±0.0	1.1±0.6
$CE5^n+E^l$	0.0±0.0	0.8±0.7	4.6±2.3	1.5±0.7
$CE5^n+E2^g$	2.6±2.6	0.0±0.0	0.0±0.0	0.7±0.5
$CE8^n+EC3^{r21}$	10.5±5.1	3.1±1.4	12.4±4.2	4.4±1.2
$CE9^n+46^l$	5.1±3.8	0.8±0.7	10.5±3.7	1.8±0.8
$CE16^n+U1^f$	5.0±3.1	7.3±2.9	1.1±1.0	1.5±0.7
$CE16^n+U1^g$	5.0±3.1	9.0±3.3	3.3±1.8	1.5±0.7
$CE16^n+U2^c$	5.0±3.0	8.2±3.0	2.2±1.6	2.2±0.9
$A1^n+A1^c$	11.3±5.3	29.5±6.5*	5.6±2.6	5.3±1.5

[a] Tests of the hypothesis that $d_S=d_N$, *$P<0.005$

In summary, we have analyzed the genetic diversity of rat class I genes that map between the *Bat1* and *Pou5f1* genomic interval and have obtained evidence for extensive polymorphism, which is mainly manifested in presence and absence of loci, and not in allelic substitutions.

Acknowledgements This paper is dedicated to Prof. Eberhard Günther, who died on February 20, 2004.

References

Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE (2004) Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Res 14:1501–1515

Günther E (1999) Rat models of disease. Rat Genome 5:75–96

Günther E, Walter L (2001) The major histocompatibility complex of the rat (*Rattus norvegicus*). Immunogenetics 53:520–542

Hülsmeyer M, Fiorillo MT, Bettosini F, Sorrentino R, Saenger W, Ziegler A, Uchanska-Ziegler B (2004) Dual, HLA-B27 subtype-dependent conformation of a self-peptide. J Exp Med 199:271–281

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

Klein J, Sato A, O'hUigin C (1998) Evolution by gene duplication in the major histocompatibility complex. Cytogenet Cell Genet 80:123–127

Kumanovics A, Madan A, Qin S, Rowen L, Hood L, Fischer Lindahl K (2002) QUOD ERAT FACIENDUM: sequence analysis of the H2-D and H2-Q regions of 129/SvJ mice. Immunogenetics 54:479–489

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Lambracht-Washington D, Düvel H, Hänisch L, Dinkel A, Wonigeit K (2004) RT1.L: a family of MHC class Ib genes of the rat major histocompatibility complex with a distinct promoter structure. Immunogenetics 56(1):28–37

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

Lambracht-Washington D, Fischer Lindahl K, Wonigeit K (2000) Promoter structures suggest independent translocations of ancestral rat RT1.A and mouse H2-K class I genes. Immunogenetics 51:873–877

Lau P, Amadou C, Brun H, Rouillon V, McLaren F, Le Rolle AF, Graham M, Butcher GW, Joly E (2003) Characterisation of RT1-E2, a multigenic family of highly conserved rat non-classical MHC class I molecules initially identified in cells from immunoprivileged sites. BMC Immunol 4:7

Leong LY, Le Rolle AF, Deverson EV, Powis SJ, Larkins AP, Vaage JT, Stokland A, Lambracht-Washington D, Rolstad B, Joly E, Butcher GW (1999) RT1-U: identification of a novel, active, class Ib alloantigen of the rat MHC. J Immunol 162:743–752

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.0, Sinauer Asscociates, Sunderland

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acid Res 24:4876–4882

Thorsby E (1997) Invited anniversary review: HLA associated diseases. Hum Immunol 53:1–11

Walter L, Günther E (1997) Clonal contigs of the class I regions of the rat major histocompatibility complex. Transplant Proc 29:1660

Walter L, Günther E (2000) Physical mapping and evolution of the centromeric class I gene-containing region of the rat MHC. Immunogenetics 51:829–837

Weissert R, Wallström E, Storch MK, Stefferl A, Lorentzen J, Lassmann H, Linington C, Olsson T (1998) MHC haplotype-dependent regulation of MOG-induced EAE in rats. J Clin Invest 102:1265–1273