

# The Genomic Sequence and Comparative Analysis of the Rat Major Histocompatibility Complex

Peter Hurt,<sup>1</sup> Lutz Walter,<sup>2</sup> Ralf Sudbrak,<sup>1</sup> Sven Klages,<sup>1</sup> Ines Müller,<sup>1,3</sup> Takashi Shiina,<sup>4</sup> Hidetoshi Inoko,<sup>4</sup> Hans Lehrach,<sup>1</sup> Eberhard Günther,<sup>2,5</sup> Richard Reinhardt,<sup>1,5</sup> and Heinz Himmelbauer<sup>1,6</sup>

<sup>1</sup>Max-Planck-Institut für molekulare Genetik, D-14195 Berlin-Dahlem, Germany; <sup>2</sup>Abteilung Immungenetik, Universität Göttingen, D-37073 Göttingen, Germany; <sup>3</sup>RZPD Deutsches Ressourcenzentrum für Genomforschung GmbH, D-14059 Berlin, Germany;

<sup>4</sup>Tokai University School of Medicine, Kanagawa 259-1193, Japan

We have determined the sequence of a 4-Mb interval on rat chromosome 20p12 that encompasses the rat major histocompatibility complex (MHC). This is the first report of a finished sequence for a segment of the rat genome and constitutes one of the largest contiguous sequences thus far for rodent genomes in general. The rat MHC is, next to the human MHC, the second mammalian MHC sequenced to completion. Our analysis has resulted in the identification of at least 220 genes located within the sequenced interval. Although gene content and order are well conserved in the class II and class III gene intervals as well as the framework gene regions, profound rat-specific features were encountered within the class I gene regions, in comparison to human and mouse. Class I region-associated differences were found both at the structural level, the number, and organization of class I genes and gene families, and, in a more global context, in the way that evolution worked to shape the present-day rat MHC.

[Supplemental material is available online at [www.genome.org](http://www.genome.org). The nucleotide sequences reported in this work have been submitted to GenBank under accession nos. BXIII70 and AL603724.]

The major histocompatibility complex (MHC) was initially described in the context of tumor transplantation studies in mouse (Gorer 1937). It plays a central role in controlling immune responsiveness and susceptibility to a great variety of mainly autoimmune and infectious diseases. Homologous regions were subsequently discovered in all vertebrate species ranging from shark to human (Trowsdale 1995; Flajnik and Kasahara 2001; Kulski et al. 2002). The mammalian MHC is usually divided into three regions, the class I, II, and III (also called central) gene regions (Fig. 1), based on the genes contained therein. Class I regions encode the classical I (class Ia) and nonclassical I (class Ib) molecules. The widely expressed class Ia molecules perform the well described antigen-presenting function to CD8 T-lymphocytes, whereas the function of class Ib genes has largely remained unknown. The class II region encodes the class II molecules that present antigen to CD4 T-lymphocytes. Both class I and class II genes are embedded between well conserved framework genes, some of which are essential for adaptive immunity (e.g., genes involved in antigen-processing and peptide transport). The class III region harbors a diverse array of structurally unrelated genes, among them several genes involved in innate immunity (e.g., complement components and cytokines). The human MHC is currently the only mammalian MHC for which the complete sequence has been determined (MHC Sequencing Consortium 1999; Shiina et al. 1999). Other mammals for which partial MHC sequences are available are the chimpanzee (class I region, complete), pig (class I region, partial), cat (class II region), and mouse (class II, III, and incomplete class I region). The comparison of MHC sequences from human and these mammals has

revealed a high degree of plasticity that is encountered within class I regions and to a lesser extent in the class II region (Chardon et al. 2001; Renard et al. 2001; Kumánovics et al. 2002; Anzai et al. 2003; Takada et al. 2003; Yuhki et al. 2003). After having established a BAC/PAC contig-based map of the rat MHC (Walter and Günther 2000; Ioannidu et al. 2001; Walter et al. 2002), we determined the sequence of a genomic segment of 3921 kb on the short arm of rat chromosome 20 that encompasses the MHC of the rat, the RT1 complex. This sequence will shed further light onto evolution and biology of the MHC, in particular as the rat represents the most important model for complex diseases such as rheumatoid arthritis, multiple sclerosis, and type I diabetes mellitus, and their control by the MHC.

## RESULTS AND DISCUSSION

### Molecular Architecture of the Rat MHC and Comparative Analysis

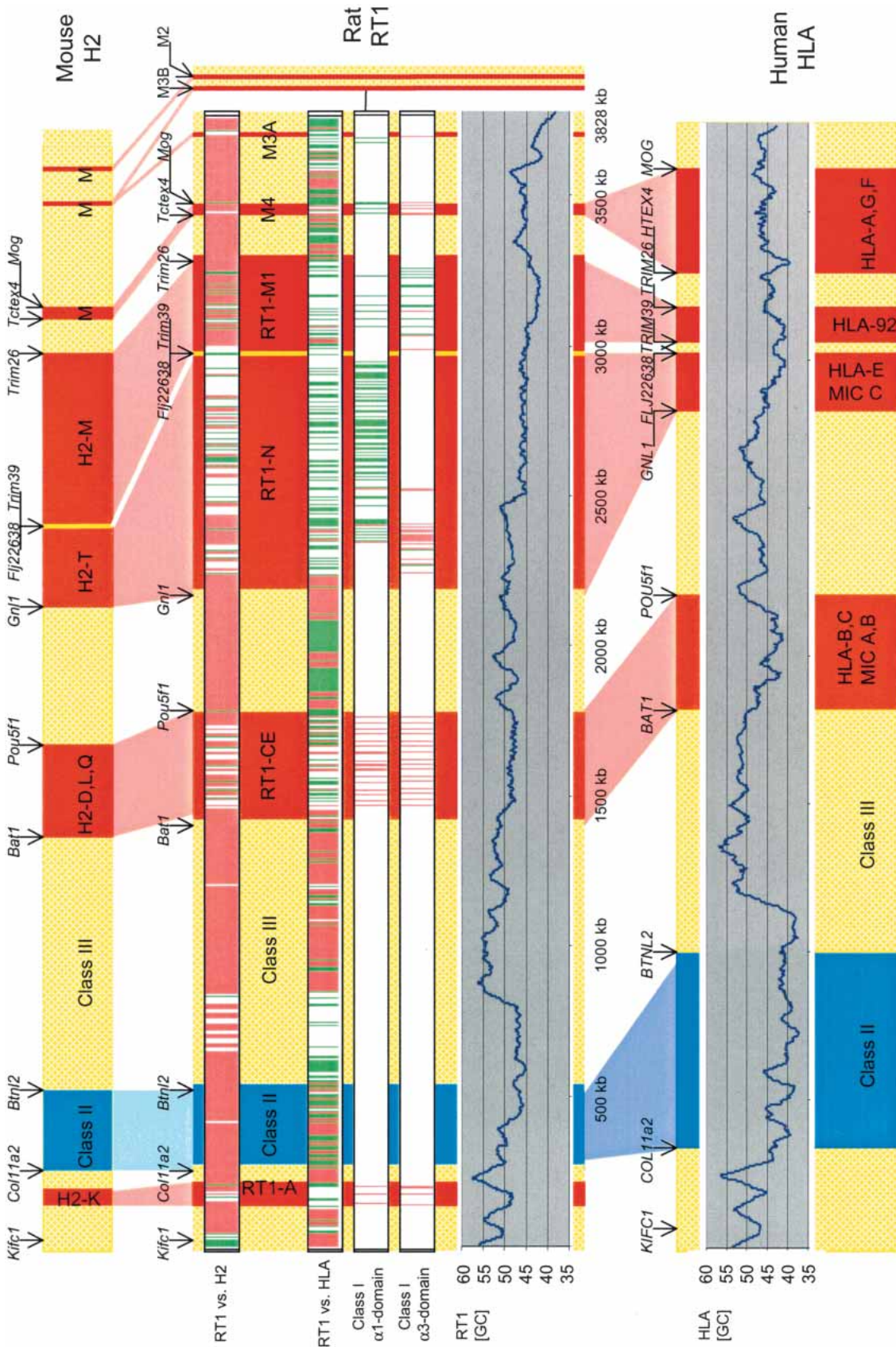
Our sequences were produced from the BN/ssNHsd rat strain, representing the rat MHC haplotype RT1<sup>a</sup> and are contained within two contigs, one major contig of 3.8 Mb and, located distally at a distance of 0.2 Mb, a minor contig of 0.09 Mb (Figs. 1, 2). The first gene on the centromeric side of the 3.8-Mb contig is the zinc finger gene *3930402F13Rik*, located 40 kb proximal to the *Kifc1* gene. *Kifc1* is a better marker for the proximal border of the MHC, because in comparison between mouse and human MHCs, colinearity disappears proximal to *Kifc1* (Fig. 1; Supplemental Fig. 6 available online at [www.genome.org](http://www.genome.org)). However, in the rat colinearity to the human HLA extends further towards the centromere, indicating the presence of an evolutionary breakpoint within this interval that disrupts syntenic specifically in mouse, but not generally in rodents. Indeed, in mouse, genes *Phf1* (*Tctex3*) and *Syngap* have been transposed to a position located 2 cM proximal to the MHC on mouse chromosome 17.

<sup>5</sup>These authors share second-to-last senior coauthorship.

<sup>6</sup>Corresponding author.

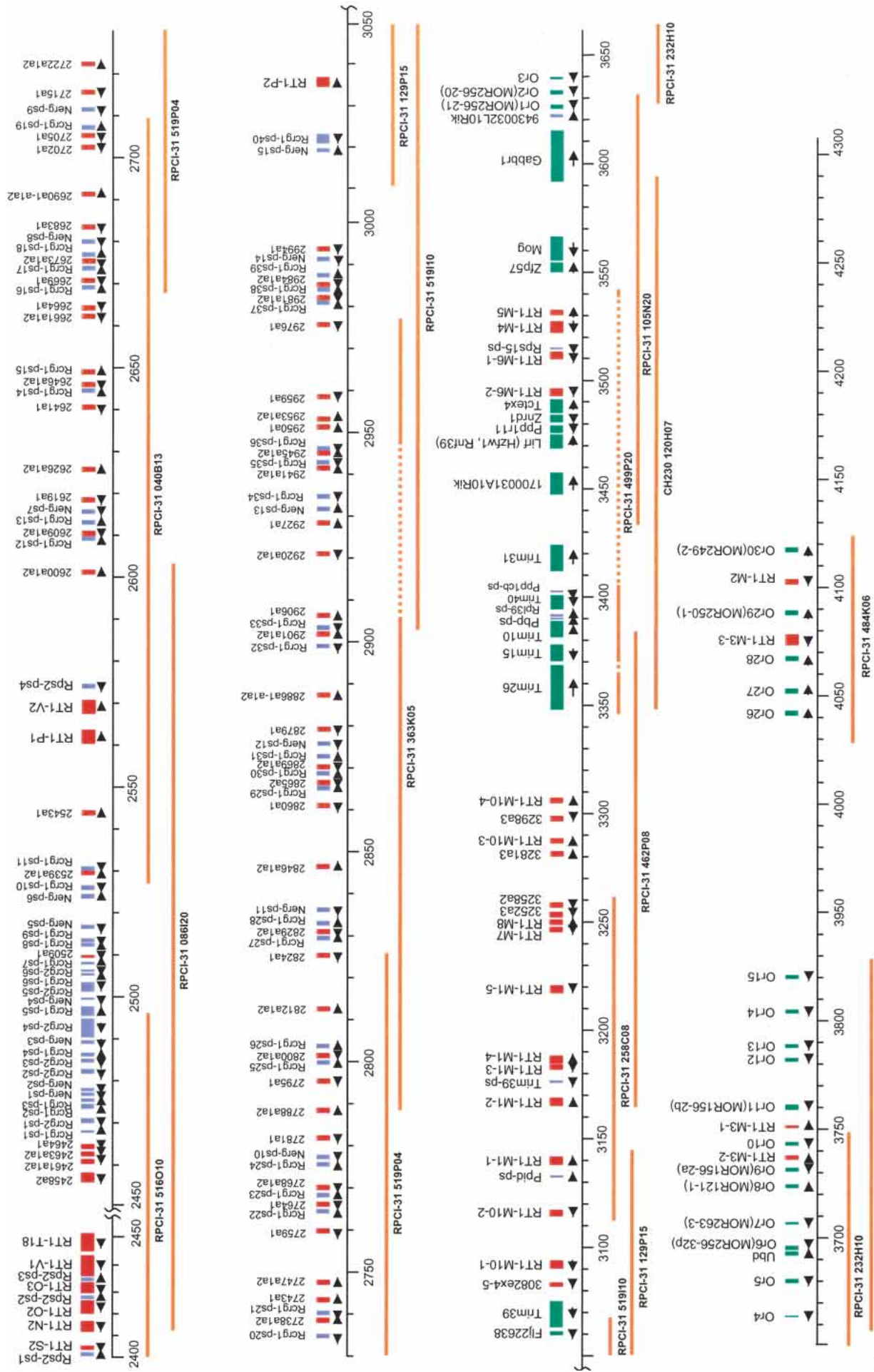
E-MAIL [himmelbauer@molgen.mpg.de](mailto:himmelbauer@molgen.mpg.de); FAX 49 30-8413-1380.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1987704>.

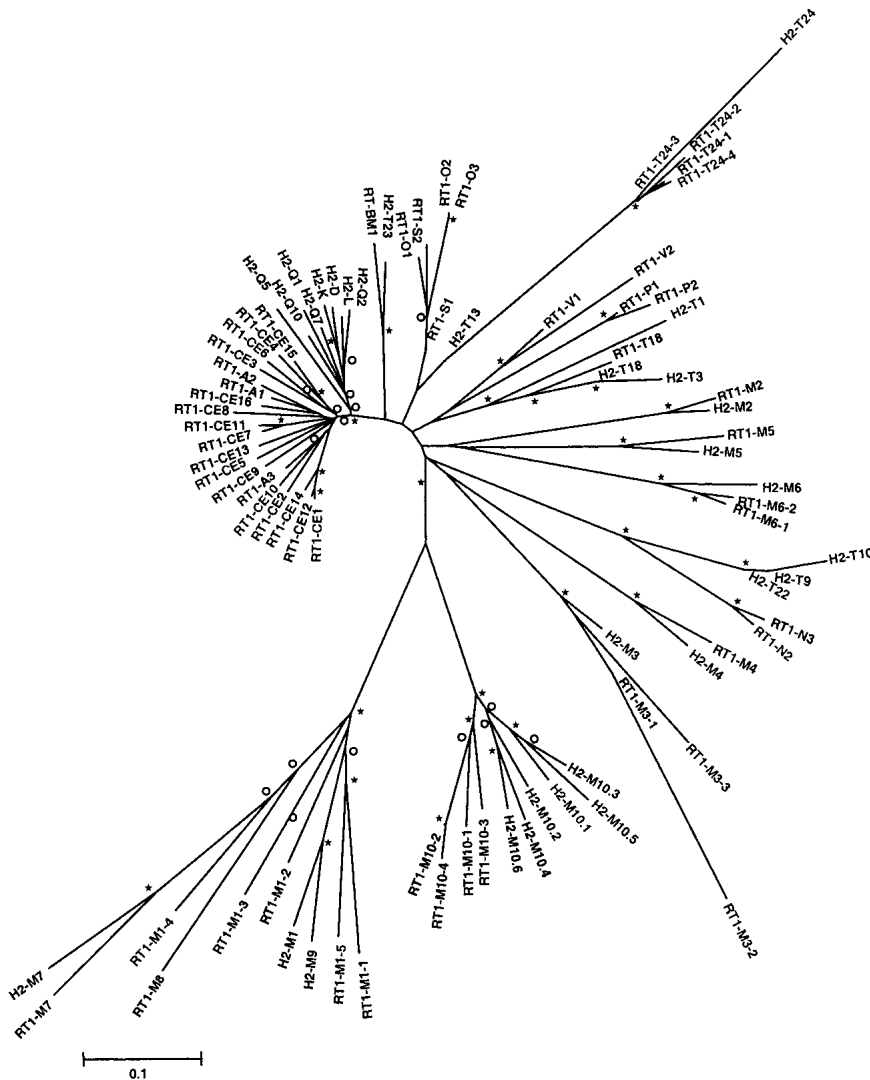


**Figure 1** Comparative analysis of structure and sequence of the MHC region in mouse H2 (top), rat RT1 (center), and human HLA (bottom). We detected considerable differences in the size of orthologous regions in rat, human, and mouse. Regions are color-coded, based on their gene content. Red, class I gene regions; blue, class II gene regions; yellow, framework gene regions including class III regions. Positions of selected framework genes, indicative of borders between class I, class II, and class III intervals and framework regions are shown. Four horizontal “heat map” bars embedded within the diagram that represent RT1 vs. H2, comparison of 3.8-Mb RT1 sequence to H2. Finished sequence of the complete mouse MHC from a single haplotype is not yet available. We have therefore merged 1.7-Mb finished sequence from proximal H2 (haplotype bc; NT\_002588) with whole-genome shotgun (WGS) draft sequence (haplotype b; NT\_039650). RT1 vs. HLA, comparison of 3.8-Mb RT1 sequence to HLA (NT007592; segment 20–25 Mb). Class I α1- and α3-domain homologies were identified by rat RT1-A gene exons 2 and 4, respectively (from U38972). Analysis was performed using PipMaker. Colors within “heat map” bars indicate the degree of conservation. Red, homology exceeds 70% in 100-bp interval. Green, homology between 50% and 70% within 100 bp. GC content for RT1 and HLA was calculated with a 50-kb window and a shift of 0.2 kb. The rat M3B/M2 region is unconnected to our major contig (indicated by thin line).





**B** Figure 2 (Legend on next page)

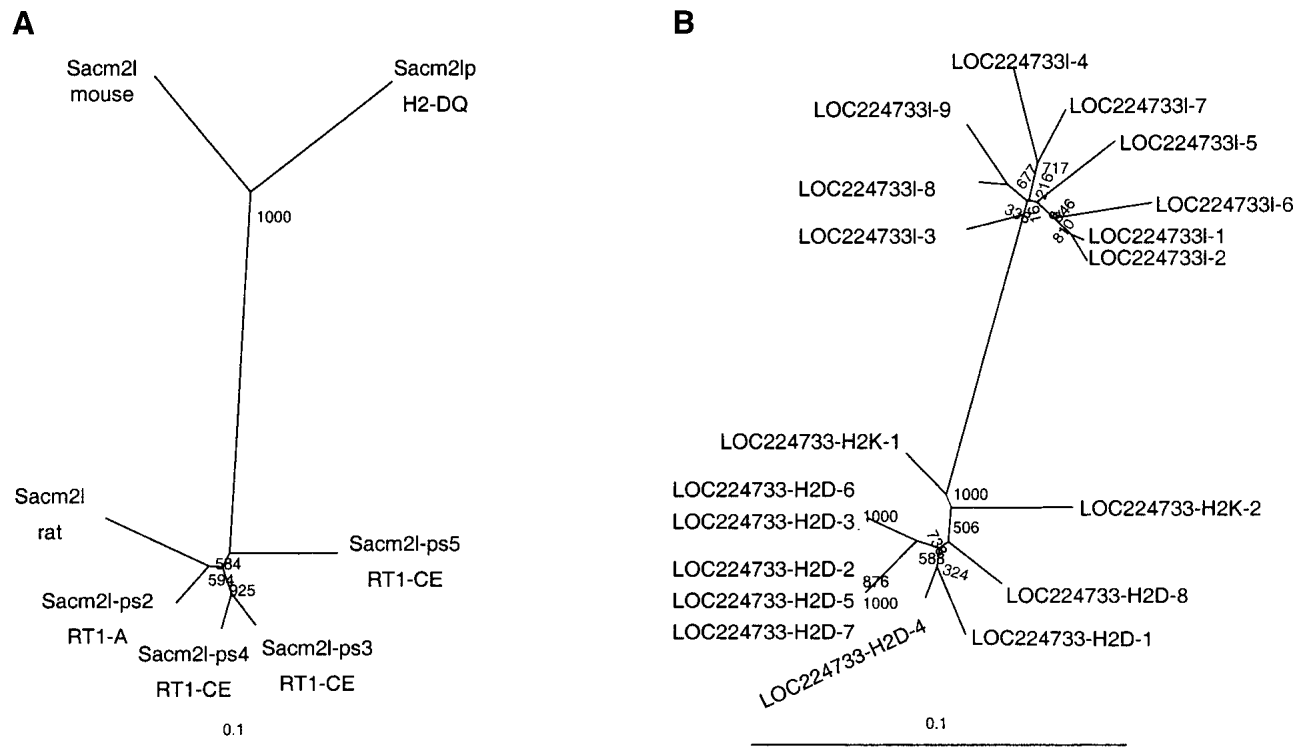


**Figure 3** Phylogenetic tree analysis of rat class I genes identified in the RT1<sup>N</sup> genomic sequence determined by us, combined with mouse class I gene sequences (rat gene prefix RT1; mouse gene prefix H2). Noteworthy are the species-specific amplification of *A* and *CE* genes in rat and *H2-K*, *-D*, *-L* genes in mouse. The class I genes in the rat *N*, mouse *T*, and in the *M*-gene regions of both species show clear orthology, with radiation in some cases that took place after species divergence, for example, the *RT1-T24* and *RT1-M10* families. A minimum evolution tree was generated by calculating Jukes-Cantor distance based on class I gene exons 2–8. Asterisks and open circles indicate bootstrap levels (500 replicates) of  $\geq 95\%$  and  $\geq 70\%$ , respectively.

**Figure 2** Physical map of the RT1 region including the positions of genes, pseudogenes, and gene fragments as detected by our sequence analysis. Nomenclature of non-class I and non-class II genes generally follows LocusLink (<http://ncbi.nlm.nih.gov/LocusLink>). Where no gene designation in mouse or human exists to date, accession numbers, Riken cDNA clone numbers, or EST designations are used. For brevity, Riken clone 5830469G19-related pseudogenes are designated as *Rcrg1-ps*, Riken clone 2210412D01-related pseudogenes as *Rcrg2-ps*, and EST 221625 (A1177969)-related pseudogenes as *Nerg-ps*. Class I gene fragments of the RT1-*N* and RT1-*M1* regions are designated according to their position (kb) and the domains (a) or exons (ex) present. No unifying nomenclature exists for olfactory receptor genes. Therefore, the designation *Or* with consecutive numbers was used, leaving a gap for those *Or* genes that map between *Or15* and *Or26* according to the Baylor rat genome sequence. The mouse gene (MOR designation) was included in brackets where orthology was evident (see Fig. 5B). Red, class I genes and class I-derived gene fragments; blue, class II genes; purple, non-class I pseudogenes and non-class I gene fragments; green, framework genes (functional genes with exception of class I and class II genes). Map positions are indicated in kilobasepairs. For sake of clarity, the segment between 2450 and 3050 kb is shown at higher resolution (breaks indicated by pair of wavy lines). Sequenced clones are indicated with their library origin and plate coordinates. We used the RPCI-31 and RPCI-32 library resources (Woon et al. 1998). A minimal tiling path consisting of 38 BAC and PAC clones was derived from our previously constructed RT1 contig maps (Walter and Günther 2000; Ioannidu et al. 2001; Walter et al. 2002). Deletions within clones are displayed as dotted lines. We encountered RT1 intervals that were prone to deletions in a number of our clones, that is, between *RT1-DMa* and *RT1-Db1* in the class II region, in interval *G6f-Bat5* and proximal to *Mog*. For these regions, clones sequenced at low redundancy in the rat genome project at Baylor College were identified using BLAST and sequenced to high coverage for gap closure (BAC clones CH230–103E15, CH230–003K4, and CH230–120H07). The three library resources used each represent the rat RT1<sup>N</sup> haplotype. Clone RPCI31–484K06 containing class I genes *RT1-M3-3* and *RT1-M2* is connected to the major contig of 3.8 Mb by draft-quality sequence generated at Baylor College.

The rat MHC contains eight class I gene clusters or gene blocks (RT1-A, -CE, -N, -M1, -M4, -M3A, -M3B, and -M2; suggested designation according to principle genes). In comparison to mouse H2 and human HLA, similarities but also discrepancies are apparent. A structural feature that distinguishes the rodent MHCs from the HLA is the presence of a class I gene cluster proximal to class II and class III gene regions, called RT1-A in rat and H2-K in mouse (Fig. 1). RT1-A locates between positions 139 and 227 kb on our sequence. RT1-A and H2-K are orthologous regions with respect to their position on the chromosome, indicated by the same pair of flanking framework genes *Sacm2l* (proximal) and *Ring1* (distal) in the two rodent species (Walter and Günther 2000). RT1-A and H2-K are not orthologs with respect to their gene content and the genomic context within which the encoded class I genes reside (Figs. 3, 4).

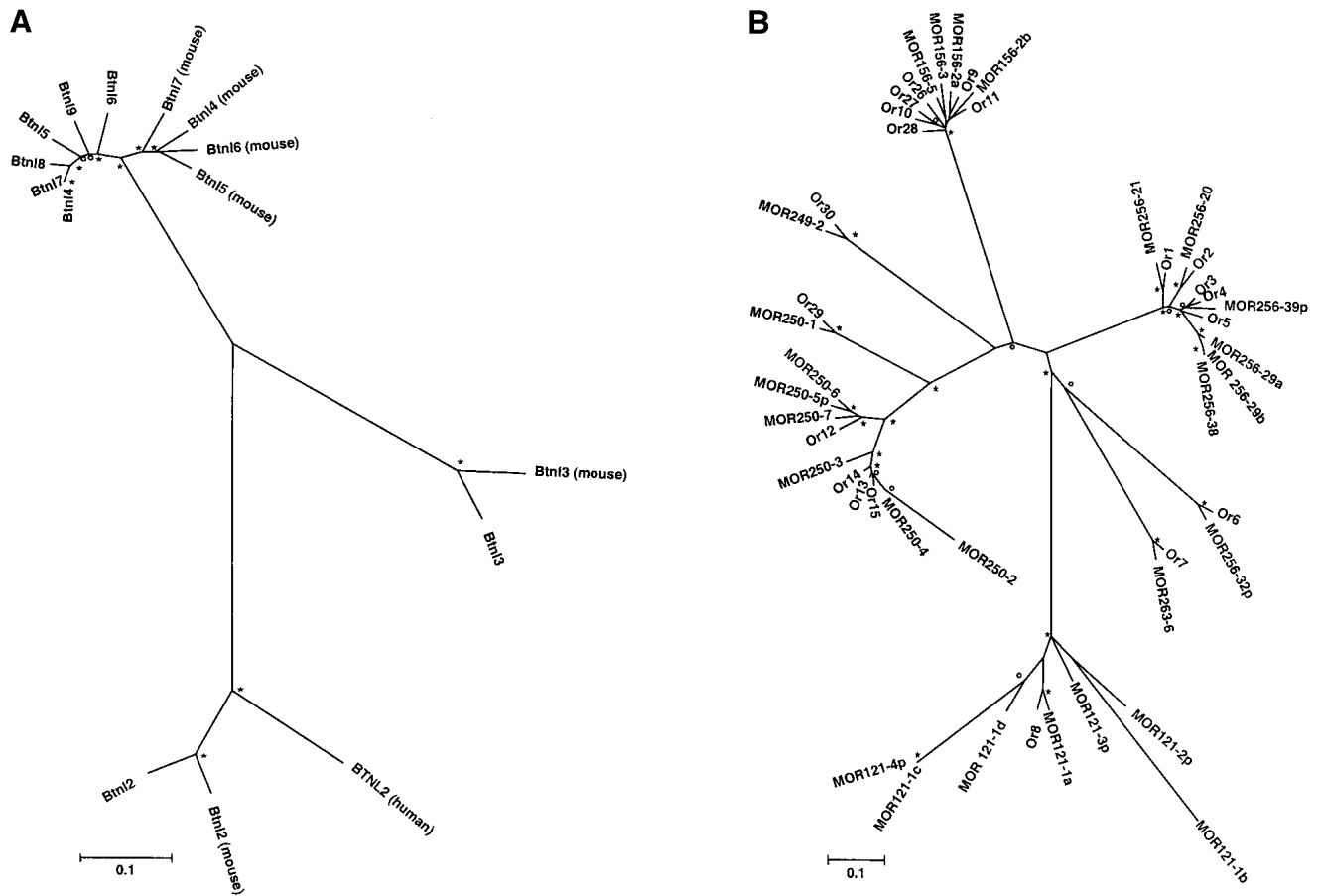
The rat centromeric class I region and the first telomeric class I gene cluster RT1-CE are separated by the class II and class III regions (positions 281–1485 kb). The overall organization of class II and class III regions in rat is much like in mouse and human, with a few noteworthy exceptions. Firstly, at the distal border of the class II region telomeric to *Tesb* (position 652–823 kb), a rodent-specific expansion of butyrophilin-like (*Btnl*) genes is observed. The function of *Btnl* proteins is unclear (Stammers et al. 2000). Interestingly, the expansion of mouse and rat *Btnl* genes occurred independently (Fig. 5A), implying rapid evolution of these genes. Secondly, we previously noted a rat-specific duplication of a module encompassing the complete *C4* complement gene and parts of the flanking loci *Cyp21a1* and *Stk19* between the *Btnl* gene cluster and *Notch4* (Walter et al. 2002). A modular duplication of *C4* and flanking genes has been found for human and mouse. In contrast to mouse and human, the duplicated modules in rat are not arranged in tandem, but are separated by 167 kb.



**Figure 4** Independent evolution of proximal and first distal class I gene cluster in rat and mouse. (A) We detected a total of seven partial homologies to the *Sacm2l* gene in class I gene regions RT1-A and -CE, four of which match the 3'-end of *Sacm2l* (position 138894–139294). A similar transposition occurred in mouse leading to the presence of a *Sacm2l*-gene fragment in the H2-D/Q region (position 1554403–1554806 in NM\_002588). Our phylogenetic analysis indicates that duplication and transposition took place after the speciation of mouse and rat. (B) Phylogenetic relationship between pseudogenes with homology to mouse gene LOC224733 in the MHC of rat and mouse. LOC224733-like sequences 1 and 2 are located in the rat A region; LOC224733-like sequences 3 to 9 map to rat CE (Fig. 2). The localization of 10 LOC224733-derived copies in mouse H2-K and H2-D is indicated. Species-specific clustering supports independent transposition and amplification of LOC224733 elements in rat and mouse. We calculated neighbor-joining trees as implemented in CLUSTALX. Bootstrap values (1000 replicates) are indicated.

Telomeric to class III, our RT1 sequence identifies seven clusters of class I genes. The cluster architecture that we found in the rat exhibits major differences compared to both mouse and human, even though individual clusters are flanked by precisely the same framework genes in either species, which is in line with the “framework hypothesis” brought forward previously (Amadou 1999). In the HLA region, human class I genes are distributed over only four clusters, with Ia genes located in clusters 1 (*HLA-B*, -C) and 4 (*HLA-A*) and Ib genes localized to clusters 2 (*HLA-E*) and 4 (*HLA-G*, -F). Cluster 3 in the HLA region contains only a class I pseudogene (*HLA-92*). Clusters 1, 2, and 4 contain the *MIC* genes *MICA* to *MICG*. We did not find any *MIC*-related genes in the rat. *MIC* genes are therefore not only absent from the mouse MHC, but are also missing in the rat. Despite the fragmented nature of class I gene-containing regions in human, evolution by duplication has been proposed such that one ancestral module containing an HLA class I gene has given rise to the current array of human class I genes and pseudogenes as well as *MIC* genes (Shiina et al. 1999). It is not possible to devise a similarly simple model for the evolution of the rat MHC. Structural similarities between clusters, aside from the presence of class I genes, are restricted to the centromeric and the first telomeric class I gene cluster in rat. On our sequence, the first telomeric cluster RT1-CE spans positions 1485–1828 kb and contains 16 class I genes in the RT1<sup>D</sup> haplotype that we studied. Of these, 13 genes could be functional based on their coding potential. It was shown previously by phylogenetic analysis of mouse and rat class I gene coding and intronic sequences that rat *RT1-A* and *CE* genes are more closely related to each other than to the centromeric mouse

*H2-K* genes and the *H2-D*, -L and -Q class I genes in the first telomeric cluster (Lambrecht-Washington and Fischer Lindahl 2002). Our findings that the RT1-A and -CE regions have several rat-specific genomic features in common are of interest in this context, because they strongly support independent evolution following the speciation of mouse and rat (Fig. 4). Firstly, genomic *Sacm2l* gene fragments in RT1-A and -CE are more closely related to the rat *Sacm2l* than to the mouse *Sacm2l* gene and a *Sacm2l* gene fragment in the H2-D/Q region. Secondly, multiple copies of a pseudogene with homology to mouse LOC224733 can be found in the RT1-A, -CE and H2-K and D/Q intervals, respectively. Again, amplified copies cluster in a species-specific manner, indicating that transposition and amplification have occurred independently in mouse and rat. Thirdly, eight copies of a pseudogene homologous to Cdk2-associated protein 1 (*Cdk2ap1*) can be found in the rat RT1-A and -CE region. *Cdk2ap1*-related sequences are absent from the mouse H2-K and -D/Q clusters. It is noteworthy that class I genes of the RT1-A and CE regions encode inhibitory and activating ligands of natural killer (NK) cells, which are an important part of the innate immune system (Naper et al. 1999). The large number of class I genes in the RT1 complex might play a particular role in the coevolution of class I and NK cell receptor genes, their composition, and expression (Naper et al. 1999, 2002). Interestingly, NK cell receptors of the rat Ly49 killer cell lectin-like receptor (KLR) type predominate over killer cell Ig-like receptor (KIR) genes (Naper et al. 1999; Hoelsbrekken et al. 2003), similar to the mouse, but at variance with human.



**Figure 5** (A) Phylogenetic relationship of butyrophilin-like genes located in the proximal part of the class III region (position 558–792 kb). Rat genes *Btn14* to *Btn19* and mouse genes *Btn14* to *Btn17* duplicated and diverged after the split of the rat and mouse lineages (mouse *Btn17* corresponds to accession number AF030001). (B) Phylogenetic analysis of rat (Or) and mouse olfactory receptor (MOR designation) genes in the distal part of the MHC. Considerable orthology is detected among rat and mouse olfactory receptors. Trees were generated using UPGMA (A) and minimum evolution (B). Asterisks and open circles indicate bootstrap levels (500 replicates) of  $\geq 95\%$  and  $\geq 70\%$ , respectively.

Our sequence of the second telomeric class I gene cluster RT1-N (position 2273–3059 kb) has revealed 16 class Ib genes, most of them potentially expressed and concentrated at the proximal end of the cluster. Remarkably, the prototype gene to the class I *RT1-T24* gene family was originally identified as a pseudogene in the haplotype r21 (Salgar et al. 1995). In RT1<sup>n</sup>, all four *RT1-T24* genes are presumably functional, based on their coding potential. We have found that the RT1-N region is extremely rich in non-class I pseudogenes and fragments derived from class I genes. We detected 40 copies of a gene homologous to mouse RIKEN cDNA 5830469G19 (*Rcrg1-ps1-40*, Fig. 2), 15 copies related to rat EST 22165 (acc. number AI177969; *Nerg-ps1-15*, Fig. 2), and six copies of mouse RIKEN gene 2210412D01 (*Rcrg2-ps1-6*, Fig. 2). By BLAST analysis, 52 genomic fragments with class I exon/intron homology were identified within a 600-kb interval up to *RT1-P2*. In 27 cases, class I exons 2 and 3 encoding  $\alpha 1$  and  $\alpha 2$  domains were found to colocalize, and in 25 cases matches displayed class I exon 2 homology only. However, none of these fragments contain class I gene exon 4 homologies. Because exon 4 encodes the  $\alpha 3$ -domain that is particularly well conserved between different class I genes, we assumed that none of the gene fragments detected are part of a functional class I locus. According to phylogenetic analysis, the fragments cannot be readily assigned to a distinct RT1 class I gene, but show a modular composition repeated 12 times (Suppl. Fig. 7).

Even though sequencing of the homologous H2-T region in mouse is still incomplete, several points can be made. The rat N-region is almost fourfold larger than mouse H2-T and appears to have acquired rat-specific characteristics during evolution. Class I genes located within the N-region display orthology to mouse T-region genes, but have duplicated since the divergence of rat and mouse, such that several rat gene paralogs match one mouse gene, for example, the *RT1-T24* genes (Fig. 3). Subfamily structure in the N-region is observed and reflects evolutionarily young duplication events. This applies to units encompassing functional class I loci (*RT1-N2/RT1-N3*; *RT1-O2/RT1-O3*; *RT1-T24* genes) and to serially amplified regions containing pseudogenes and class I gene fragments. Thus, a module of *RT1-N*, *RT1-O*, *Rps2*, and *RT1-S* genes is duplicated in this region and occurs as triplicate in the r21 haplotype (Salgar et al. 1998).

The telomeric 1000 kb of the MHC complex in the rat is characterized by the presence of the M-like class I gene clusters M1, M4, M3A, M3B, and M2. Clusters M1 and M4 match the positions of human class I gene blocks HLA-92 and HLA-A/G/F, without any further structural and functional resemblance inside the class I regions. In fact, M-like class I genes are absent from the HLA complex. Between mouse and rat, an orthologous relationship between most M-region class I genes is evident, with some exceptions, for example, the species-specific expansion of the subfamily of *M10* class I genes (Fig. 3), recently shown to be

involved in pheromone reception in mouse (Loconto et al. 2003). The telomeric border of the human MHC is often marked by the framework gene *MOG*. In contrast, the rat MHC extends further distal, with three different *M3*-class I genes and the *M2*-class I gene embedded in a framework of olfactory receptor (OR) genes. These OR genes to a large extent have orthologs in the mouse (Fig. 5B). This is remarkable, because rapid evolution of OR genes has been proposed on the basis of a comparison of mouse and human OR genes in the MHC region (Younger et al. 2001). The conservation of the M region with its involvement in pheromone reception and olfaction could reflect the importance of these functions for mouse and rat relative to human.

### Evolution of Class I Gene Regions

Taken together, our results demonstrate that also moderately diverged mammals display very pronounced differences in their class I gene regions, as the split between the rat and mouse lineage was dated to 16–23 million years ago (Springer et al. 2003). The differences relate to class I gene numbers, their organization into class I gene subfamilies, and also to the genomic context encountered within class I blocks. In combination with our phylogenetic data (Figs. 3, 4; Suppl. Fig. 7), this strongly testifies to the rapid species-specific evolution of class I genes and gene clusters, following speciation. In part, this variability will reflect the different parasite load that challenges the immune systems of rat, mouse, and human. Our data also highlight that different mechanisms exist to shape the MHCs of extant mammals. In the primate species investigated thus far (human and chimpanzee), class I genes are embedded within modules that share homology between different clusters. This is a clear indication of coevolution between the different clusters by amplification, transposition, and diversification. In contrast, sequence exchange between clusters in the rat is only recognizable for the RT1-A and RT1-CE clusters. Modular architecture that is observed in the RT1-N cluster is the result of local duplications only, and modules, except for the presence of class I genes, have no homology to sequences located in other clusters, supporting independent evolution. In contrast to rodents and primates, the pig class I region, which has been analyzed as a representative of artiodactyls, is much simpler in its organization. In pig, class I genes are concentrated within only two clusters (Chardon et al. 2001; Renard et al. 2001), at positions equivalent to the rat RT1-CE and RT1-N clusters.

### Rat MHC Polymorphisms

Few sequence data are available for an analysis of MHC polymorphism in the rat. The polymorphism of the class Ia and class II genes can be examined on the basis of cDNA sequence data from the available defined independent RT1 haplotypes of inbred rat strains. Variability is found in the range to be expected from the mouse, highest variability being observed in the peptide-binding domain-encoding exons (Suppl. Fig. 9). Furthermore, presence/absence polymorphisms are evident in the class Ia-encoding RT1-A region. The different RT1 haplotypes carry one, two, or three of the A1/A2/A3 set of class I genes (Günther and Walter 2001; Walter and Günther 2000). Variability of class I gene content is also evident in the RT1-CE class I cluster between *Bat1* and *Pou5f1*. Comparison of the sequence reported here and that of a *Bat1*-encompassing cosmid clone from the RT1<sup>1</sup> haplotype of the LEW strain (GenBank acc. no. AF387339, Lambrecht-Washington and Fischer Lindahl 2002) reveals that not only the distance between *Bat1* and the first telomeric class I gene differs between LEW and BN (13 kb vs. 12 kb), but also that these class I genes are not orthologs. Rather, the first class I gene in the RT1-CE cluster of the RT1<sup>1</sup> haplotype shows the highest degree of sequence similarity to the CE9 gene of the RT1<sup>n</sup> haplotype (data not shown).

### Rat MHC and Disease Association

The rat, in particular the BN strain from which the BAC and PAC libraries are derived, is an important model system for studying complex genetic diseases, and for organ transplantation. The BN strain is known for its predisposition to develop Th2-polarized immune responses that in part might be MHC-controlled (Fournié et al. 2001). Furthermore, after administration of mercury or gold salt, BN rats develop autoimmune traits that are under MHC control (Druet et al. 1977; Kermarrec et al. 1996). In myelin oligodendrocyte glycoprotein (MOG)-induced experimental autoimmune encephalomyelitis (EAE), which represents a model of multiple sclerosis, the RT1<sup>n</sup> haplotype studied here confers an early-onset, acute type, whereas other RT1 haplotypes are associated with relapsing/remitting, or mild forms. Susceptibility has been attributed to the class II region and to the telomeric class I region as well (Weissert et al. 1998; Stefferl et al. 1999). Interestingly, *Mog* maps to the telomeric end of the MHC, and the highly susceptible BN rat exhibits a stronger expression of the MOG protein compared to less susceptible strains (Pagany et al. 2003). The rat RT1<sup>n</sup> haplotype sequence generated by us can now be used to study the molecular basis of MHC-controlled disease traits and serve as a reference sequence. Sequencing of further RT1 haplotypes will delineate the features that distinguish the different haplotypes, causing animals to be either disease-prone or -resistant.

## METHODS

### Clone Resources, Sequencing, and Sequence Annotation

Rat RPCI-31 PAC and RPCI-32 BAC clones were obtained from the Resource Centre of the German Human Genome Project (<http://www.rzpd.de>). These libraries, and the CH230 rat BAC library as well, were prepared from the BN rat strain. The sequence therefore represents a single rat MHC haplotype (RT1<sup>n</sup>). BAC/PAC-DNA was purified by CsCl-density centrifugation according to standard protocols and sheared by sonification. Size-selected fragments (1500 bp or 3500 bp) were cloned into pUC18. Sequencing was carried out on PCR-amplified pUC18 inserts using Big Dye-terminator chemistry on ABI3700 sequencers. All raw sequences were processed by Phred and assembled by Phrap (Ewing and Green 1998; Staden et al. 2000). Clones were initially sequenced to 10-fold coverage to resolve reads into ordered contigs. Gaps were subsequently closed by directed sequencing. Analyzed regions were manually edited in GAP4 (<http://www.sanger.ac.uk/Software/sequencing/docs/phrap2gap>). The quality of these sequence data was finished to justify the “Bermuda Rules” (<http://www.ornl.gov/hgmis/research/bermuda.html#1>). Repetitive elements were identified using RepeatMasker (A. Smit and P. Green, unpubl.). For annotation, only homologies to experimentally validated genes and spliced ESTs were taken into account. Gene structures were generated using EST\_GENOME, which produces optimal alignment of spliced to unspliced DNA (Mott 1997), followed by manual refinement. Cross-species homologies were identified using PipMaker (Schwartz et al. 2000), and phylogenetic analysis was carried out using CLUSTALX and MEGA 2.0 (Thompson et al. 1994; Kumar et al. 2001).

## ACKNOWLEDGMENTS

We thank K. Borzym, S. Gelling, V. Gimmel, K. Heitmann, S. Lehrack, and M. Sontag for template preparation, sequencing, and sequence assembly, A. Kosiura for shotgun library preparation, and S. Palczewski for BAC clone confirmation. D. Groth, S. Hennig, and A. Beck provided useful scripts for data analysis and data submission. We thank S. Scherer for CH230 rat BAC clones. This work was supported by the German National Genome Research Network (NGFN), the Max-Planck Society (MPG), and by Grant-in-Aid for Scientific Research on Priority Areas (C) “Genome Science” from the Ministry of Education, Culture, Sports, Science and Technology of Japan.



The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

- Amadou, C. 1999. Evolution of the MHC class I region: The framework hypothesis. *Immunogenetics* **49**: 362–367.
- Anzai, T., Shiina, T., Kimura, N., Yanagiya, K., Kohara, S., Shigenari, A., Yamagata, T., Kulski, J.K., Naruse, T.K., Fujimori, Y., et al. 2003. Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. *Proc. Natl. Acad. Sci.* **100**: 7708–7713.
- Chardon, P., Renard, C., Rogel-Gaillard, C., Cattolico, L., Duprat, S., Vaiman, M., and Renard, C. 2001. Sequence of the swine major histocompatibility complex region containing all nonclassical class I genes. *Tissue Antigens* **57**: 55–65.
- Druet, E., Sapin, C., Günther, E., Feingold, N., and Druet, P. 1977. Mercuric chloride-induced antglomerular basement membrane antibodies in the rat: Genetic control. *Eur. J. Immunol.* **7**: 348–351.
- Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* **8**: 186–194.
- Flajnik, M.F. and Kasahara, M. 2001. Comparative genomics of the MHC: Glimpses into the evolution of the adaptive immune system. *Immunity* **15**: 351–362.
- Fournié, G.J., Cautain, B., Xystrakis, E., Damoiseaux, J., Mas, M., Lagrange, D., Bernard, I., Subra, J.F., Pelletier, L., Druet, P., et al. 2001. Cellular and genetic factors involved in the difference between Brown Norway and Lewis rats to develop respectively type-2 and type-1 immune-mediated diseases. *Immunol. Rev.* **184**: 145–160.
- Gorer, P.A. 1937. The genetic and antigenic basis of tumour transplantation. *J. Pathol. Bacteriol.* **44**: 691–697.
- Günther, E. and Walter, L. 2001. The major histocompatibility complex of the rat (*Rattus norvegicus*). *Immunogenetics* **53**: 520–524.
- Hoelsbrekken, S.E., Nylenna, Ø., Saether, P.C., Slettedal, I. Ø., Ryan, J.C., Fossum, S., and Dissen, E. 2003. Cutting edge: Molecular cloning of a killer cell Ig-like receptor in the mouse and rat. *J. Immunol.* **170**: 2259–2263.
- Ioannidu, S., Walter, L., Dressel, R., and Günther, E. 2001. Physical map and expression profile of genes of the telomeric class I gene region of the rat MHC. *J. Immunol.* **166**: 3957–3965.
- Kermarrec, N., Dubay, C., De Gouyon, B., Blanpied, C., Gauguier, D., Gillespie, K., Mathieson, P.W., Druet, P., Lathrop, M., and Hirsch, F. 1996. Serum IgE concentration and other immune manifestations of treatment with gold salts are linked to the MHC and IL4 regions in the rat. *Genomics* **31**: 111–114.
- Kulski, J.K., Shiina, T., Anzai, T., Kohara, S., and Inoko, H. 2002. Comparative genomic analysis of the MHC: The evolution of class I duplication blocks, diversity and complexity from shark to man. *Immunol. Rev.* **190**: 95–122.
- Kumánovics, A., Madan, A., Qin, S., Rowen, L., Hood, L., and 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, I.B., and Nei M. 2001. MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* **17**: 1244–1245.
- Lambracht-Washington, D. and Fischer Lindahl, K. 2002. Does the rat have an H2-D orthologue next to Bat1? *Immunogenetics* **53**: 1039–1046.
- Loconto, J., Papes, F., Chang, E., Stowers, L., Jones, E.P., Takada, T., Kumánovics, A., Fischer Lindahl, K., and Dulac, C. 2003. Functional expression of murine V2R pheromone receptors involves selective association with the M10 and M1 families of MHC class Ib molecules. *Cell* **112**: 607–618.
- The MHC Sequencing Consortium 1999. Complete sequence and gene map of a human major histocompatibility complex. *Nature* **401**: 921–923.
- Mott, R. 1997. EST\_GENOME: A program to align spliced DNA sequences to unspliced genomic DNA. *Comput. Appl. Biosci.* **13**: 477–478.
- Naper, C., Ryan, J.C., Kirsch, R., Butcher, G.W., Rostad, B., and Vaage, J.T. 1999. Genes in two major histocompatibility complex class I regions control selection, phenotype, and function of a rat Ly-49 natural killer cell subset. *Eur. J. Immunol.* **29**: 2046–2053.
- Naper, C., Hayashi, S., Løvik, G., Kveberg, L., Niemi, E.C., Rolstadt, B., Dissen, E., Ryan, J.D., and Vaage, J.T. 2002. Characterization of a novel killer cell lectin-like receptor (KLRH1) expressed by alloreactive rat NK cells. *J. Immunol.* **168**: 5147–5154.
- Pagany, M., Jagodic, M., Bourquin, C., Olsson, T., and Linington, C. 2003. Genetic variation in myelin oligodendrocyte glycoprotein expression and susceptibility to experimental autoimmune encephalomyelitis. *J. Neuroimmunol.* **139**: 1–8.
- Renard, C., Vaiman, M., Chiannikulchai, N., Cattolico, L., Robert, C., and Chardon, P. 2001. Sequence of the pig major histocompatibility region containing the classical class I genes. *Immunogenetics* **53**: 490–500.
- Schwartz, S., Zhang, Z., Frazer, K.A., Smit, A., Riemer, C., Bouck, J., Gibbs, R., Hardison, R., and Miller, W. 2000. PipMaker—A web server for aligning two genomic DNA sequences. *Genome Res.* **10**: 577–586.
- Salgar, S.K., Kunz, H.W., and Gill III, T.J. 1995. Nucleotide sequence and structural analysis of the rat RT1.Eu and RT1.Aw3l genes, and of genes related to RT1.O and RT1.C. *Immunogenetics* **42**: 244–253.
- Salgar, S.K., Kunz, H.W., and Gill III, T.J. 1998. Structural organization, sequence analysis, and physical mapping of the Grc-linked class Ib gene RT1.S3 in the rat. *Immunogenetics* **48**: 76–81.
- Shiina, T., Tamiya, G., Oka, A., Takishima, N., Yamagata, T., Kikkawa, E., Iwata, K., Tomizawa, M., Okuaki, N., Kuwano, Y., et al. 1999. Molecular dynamics of MHC genesis unravelled by sequence analysis of the 1,796,938-bp HLA class I region. *Proc. Natl. Acad. Sci.* **96**: 13282–13287.
- Springer, M.S., Murphy, W.J., Eizirik, E., and O'Brien, S.J. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. *Proc. Natl. Acad. Sci.* **100**: 1056–1061.
- Staden, R., Beal, K.F., and Bonfield, J.K. 2000. The Staden package, 1998. *Methods Mol. Biol.* **132**: 115–130.
- Stammers, M., Rowen, L., Rhodes, D., Trowsdale, J., and Beck, S. 2000. 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**: 373–382.
- Steffel, A., Brehm, U., Storch, M., Lambracht-Washington, D., Bourquin, C., Wonigeit, K., Lassmann, H., and Linington, C. 1999. Myelin oligodendrocyte glycoprotein induces experimental autoimmune encephalomyelitis in the "resistant" Brown Norway rat: Disease susceptibility is determined by MHC and MHC-linked effects on the B cell response. *J. Immunol.* **163**: 40–49.
- Takada, T., Kumanovics, A., Amadou, C., Yoshino, M., Jones, E.P., Athanasiou, M., Evans, G.A., and Fischer Lindahl, K. 2003. Species-specific class I gene expansions formed the telomeric 1 Mb of the mouse major histocompatibility complex. *Genome Res.* **13**: 589–600.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Trowsdale, J. 1995. "Both man and bird and beast": Comparative organization of MHC genes. *Immunogenetics* **41**: 1–17.
- Walter, L. and Günther, E. 2000. Physical mapping and evolution of the centromeric class I gene-containing region of the rat MHC. *Immunogenetics* **51**: 829–837.
- Walter, L., Hurt, P., Himmelbauer, H., Sudbrak, R., and Günther, E. 2002. Physical mapping of the major histocompatibility complex class II and class III regions of the rat. *Immunogenetics* **54**: 268–275.
- Weissert, R., Wallstrom, E., Storch, M.K., Steffel, A., Lorentzen, J., Lassmann, H., Linington, C., and Olsson, T. 1998. MHC haplotype-dependent regulation of MOG-induced EAE in rats. *J. Clin. Invest.* **102**: 1265–1273.
- Woon, P.Y., Osoegawa, K., Kaisaki, P.J., Zhao, B., Catanese, J.J., Gauguier, D., Cox, R., Levy, E.R., Lathrop, G.M., Monaco, A.P., et al. 1998. Construction and characterization of a 10-fold genome equivalent rat P1-derived artificial chromosome library. *Genomics* **15**: 306–316.
- Younger, R.M., Amadou, C., Bethel, G., Ehlers, A., Fischer Lindahl, K., Forbes, S., Horton, R., Milne, S., Mungall, A.J., Trowsdale, J., et al. 2001. Characterization of clustered MHC-linked olfactory receptor genes in human and mouse. *Genome Res.* **11**: 519–530.
- Yuhki, N., Beck, T., Stephens, R.M., Nishigaki, Y., Newmann, K., and O'Brien, S.J. 2003. Comparative genome organization of human, murine, and feline MHC class II region. *Genome Res.* **13**: 1169–1179.

## WEB SITE REFERENCES

- <http://www.rzpd.de>; German Resource Center.
- <http://www.sanger.ac.uk/Software/sequencing/docs/phrap2gap>; Sequence assembly software.
- <http://www.ornl.gov/hgmis/research/bermuda.html#1>; Policies on release of genomic sequence data.
- <http://ncbi.nlm.nih.gov/LocusLink>; LocusLink database. Provides a single query interface to curated sequence and descriptive information about genetic loci.

Received September 18, 2003; accepted in revised form December 8, 2003.