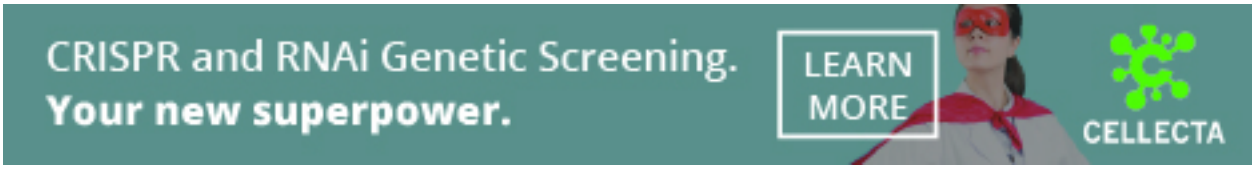
Phylogenetics of the laboratory rat Rattus norvegicus.

F Canzian

Genome Res. 1997 7: 262-267
Access the most recent version at doi:10.1101/gr.7.3.262



RESEARCH

Phylogenetics of the Laboratory Rat *Rattus norvegicus*

Federico Canzian[1]

Unit of Genetic Cancer Susceptibility, International Agency for Research on Cancer, F-69372 CEDEX 08, Lyon, France

A genealogic tree was constructed for inbred strains of the laboratory rat, including 63 strains and 214 of their substrains. Information on genetic and biochemical marker typings of these lines was collected from the literature and from the World Wide Web. Data on 995 polymorphisms were processed into a phylogenetic distance matrix, and a tree was obtained by the Fitch–Margoliash distance matrix method. The inbred strains of the laboratory rat showed an average polymorphism for pairwise comparison of 53%. Strain BN showed the highest genetic divergence from all the other ones. Comparison with the mouse phylogenetic tree indicated that laboratory rats possess a higher diversity than inbred strains of mice not derived from wild species. These results provide a phylogenetic basis in the choice of rat strains for genetic linkage experiments.

[Data described in this paper can be found at ftp://193.51.164.108/eps-xchange/Canzian/Rat/Phylogeny/Tree.jpg and ftp://193.51.164.108/eps-xchange/Canzian/Rat/Phylogeny/Matrix.jpg.]

After the mouse, the laboratory rat *Rattus norvegicus* is the most commonly used experimental organism. It is a model for a number of important human traits of biomedical importance, including susceptibility to cancer, hypertension, obesity, diabetes, and autoimmune diseases.

Although the mapping of the rat genome lags somewhat behind that of the mouse, in recent years there has been some effort made by several groups to increase the number of markers on the rat genetic map groups (Serikawa et al. 1992; Yamada et al. 1994; Jacob et al. 1995; Toyota et al. 1996). To date, this figure approaches 1000–1500 polymorphic markers.

The generation of new markers will be important, particularly in experiments of genetic mapping involving crosses between different inbred strains of rat. In many cases, the choice of strain will be directed by the maximal level of polymorphism achievable between the different rat strains.

A thorough understanding of the interstrain differences is therefore important, particularly at the phylogenetic level. There are hundreds of inbred strains of rats, and the history of their generation and evolution is not well known. A study has been published with a phylogenetic tree for 13 commonly used rat strains (Canzian et al. 1995a). Furthermore, the genotypes of many rat inbred strains have been characterized at a number of loci throughout the genome. This study incorporates most of the available information to produce a relatedness tree comprising 63 inbred strains and 214 substrains.

RESULTS AND DISCUSSION

In this study, most of the available information on genotypes in different inbred strains of rat was used for a phylogenetic analysis. The result is a relatedness tree for rat inbred strains, shown in Figure 1 and available on the World Wide Web at ftp://193.51.164.108/eps-xchange/Canzian/Rat/Phylogeny/Tree.jpg. The matrix showing pairwise differences among all the 63 strains is available at ftp://193.51.164.108/eps-xchange/Canzian/Rat/Phylogeny/Matrix.jpg.

The average polymorphism for pairwise comparisons of rat strains is 53%, higher than the 45% calculated from typings of biochemical and immunological loci in *Mus musculus domesticus*-derived inbred strains of mice (Fitch and Atchley 1985) and the 49% reported among non-wild-derived mice for microsatellites (Dietrich et al. 1994). Moreover, the average polymorphism of the combined Kyoto, MIT, Utrecht, NIH, and Oxford data sets (obtained only with microsatellites) is 64%, suggesting that the genetic patrimony of rat inbred strains is richer than that of non-wild-derived strains of mice.

The overall difference content of the Toku-

[1]E-MAIL canzian@iarc.fr; FAX 33-4-72738575.

7:262–267 ©1997 by Cold Spring Harbor Laboratory Press ISSN 1054-9803/97 $5.00

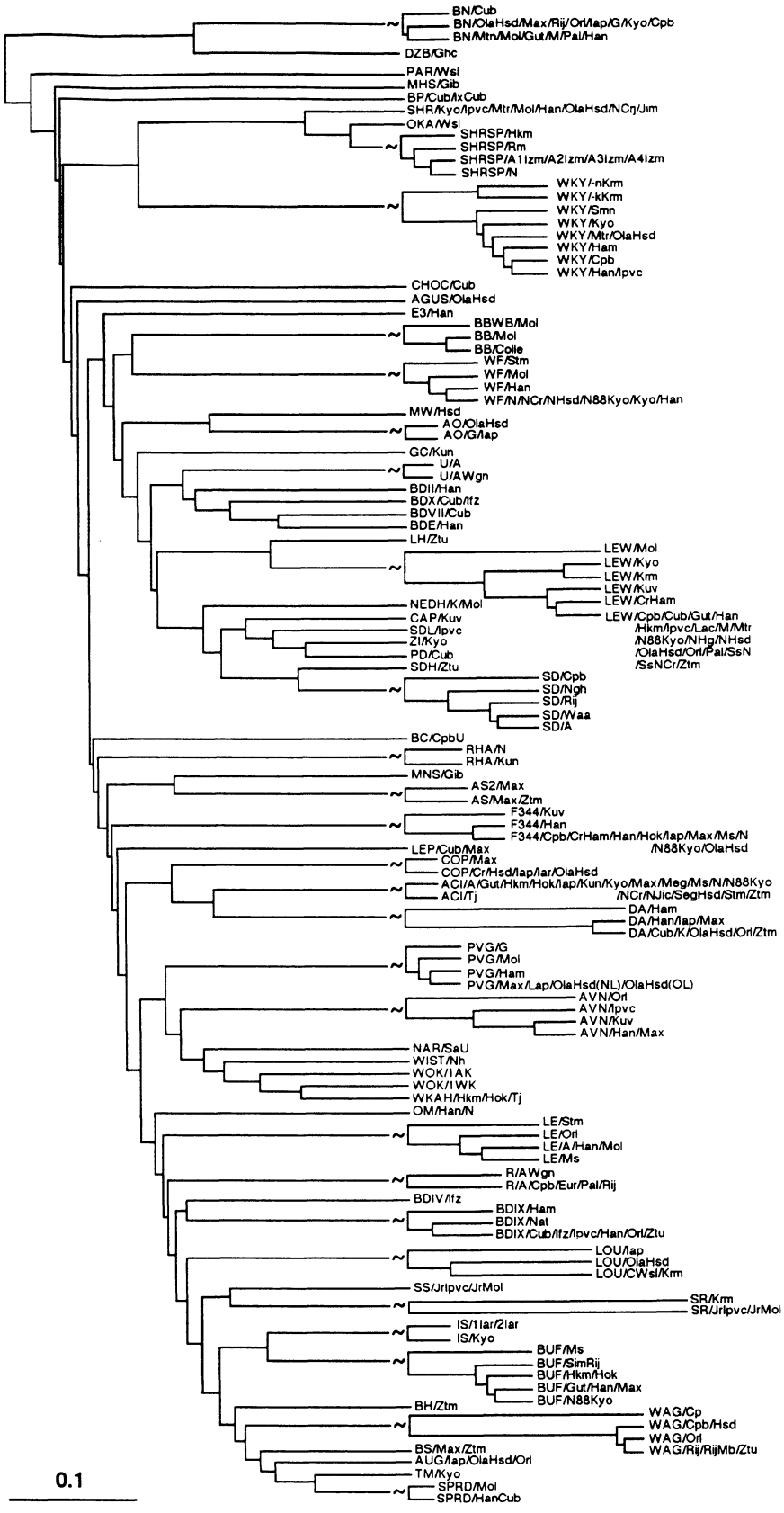


Figure 1 (*See following page for legend.*)

shima data set (based on biochemical polymorphism) is only 34%. This lower level of polymorphism may be attributable to (1) the fact that a single biochemical type (recorded as, e.g., enzyme activity) can arise from several distinct genetic polymorphisms, and/or (2) the fact that biochemical polymorphisms derive from coding sequences and are therefore subjected to a different selective pressure, especially in the case of the artificially directed evolution of inbred strains.

From the observation of the comprehensive relatedness tree (Fig.1) and comparison of the different trees from each data set (not shown), the following conclusions may be drawn:

1. Strain BN consistently possesses a genetic divergence most distant from all the other strains. This is in accordance with the tree reported previously (Canzian et al. 1995a) and is consistent with its reported origin from wild rats (Palm and Black 1971), in comparison to most other inbred strains, derived from collectionists and commercial breeders, which may have a heterogeneous genetic background prior to initiation of inbreeding. Therefore, the BN strain is likely in rat genetics to possess a role similar to *Mus spretus*, *Mus musculus castaneus*, and *Mus musculus molossinus*-derived strains for the mouse. The only strain closely related to BN is DZB, the origin of which is unknown.

2. In all single data set trees (SDTs) including them, strains PAR and CHOC also demonstrate divergence from the other strains, whereas not all SDTs are in agreement as to the position of BP, MHS, and AGUS.

3. Strains WKY, SHR, SHRSP, and OKA consistently belong to the same subtree, in accordance with reports about their origins as models for hypertension (Festing 1979). Moreover, in most SDTs, this group, after BN, is the most distant from all the others, in accordance with Canzian et al. (1995a). These strains derive from animals of outbred Wistar origin inbred at Kyoto or Hokkaido University (Festing 1979).

4. Strain BDE is reported to originate from a cross between E3 and BDVII rats (Festing 1979), and such information is supported in this work.

5. Strain AO is reported to be a subline of WAG (Festing 1979). The two strains are reported in the Utrecht and Tokushima SDTs; they are very close in the former, but distant in the Tokushima and the overall tree (Fig. 1).

6. F344, COP, AUG, and ACI were all bred by M.R. Curtis at Columbia University in 1918. Ancestors of these strains were purchased from local dealers, with the exception of the progenitors of COP, which came from Denmark (Palm and Black 1971). ACI is the product of a mating of AUG and COP sublines. DA was suggested to be related to ACI (Palm and Black 1971). These strains were in the same subtree in the tree proposed by Canzian et al. (1995a). In the present tree, ACI, COP, and DA are closely linked, whereas F344 and AUG appear to be more distant but still in the same supergroup, including strains from BC to SPRD in Fig. 1. From the matrix of pairwise differences (ftp://193.51.164.108/eps-xchange/Canzian/Rat/Phylogeny/Matrix.jpg) it is clear that ACI is closely related to its parent strains (ACI–COP, 37.2% difference; ACI–AUG 33.8%). However, the PHYLIP package is designed to infer phylogenetic relations that occur in nature, where two separate species do not mate and generate a third one. The fact that AUG and COP are phylogenetically distant (47.9% difference) may explain why AUG and ACI appear distant in the present tree. The relationship among AUG, COP, and ACI was more evident in the tree presented in Canzian et al. (1995a) because in that case the strains under investigation were much fewer.

7. Strain MNS was bred together with MHS and is its normal control in a model of hyperstension. In the Tokushima and Oxford SDTs these two strains are tightly related and close to the SHR group of strains (another model of hypertension, which could have been selected partially for the same genes). Nevertheless, in the Utrecht SDT, and in the general tree, the similarity between MNS and MHS is not apparent, although their overall difference is only 39.2%. Strains SS and SR, representing another model of hypertension, are always closely related, but remain consistently distant from the SHR and the MHS groups.

8. The assumption that the intrastrain genetic difference is lower than the interstrain difference (see Methods) holds in most cases, with the exceptions of substrains LEW/Mol (up to 50.0% difference with other LEW substrains), WKY/-nKrm and WKY/-kKrm (up to 46.7% difference), WAG/Cpb (45%), DA/Ham (43.5%), LE/Stm (40.0%), AVN/Orl (39.1%), and LOU/Iap (37.5%), which all show differences from other substrains of the same strain

Figure 1 Phylogenetic tree of inbred rat strains, obtained by the Fitch–Margoliash distance matrix method with evolutionary clock. The scale represents percentage of difference and is applicable only to the left part of the tree, including the consensus of the 63 strains. When a strain has substrains, the points of branching of the substrains from the main line are unknown and are therefore indicated as ~.

greater than the average interstrain difference in the whole data set. Users of these substrains for genetic studies should be careful, because the level of polymorphism with other strains could be different than expected. In contrast, other commonly used strains show remarkable homogeneity, that is, no polymorphism observed among substrains of AUG, BDX, BS, LEP, NEDH, OM, SHR, SS, and WKAH; up to 5.0% of polymorphisms among ACI substrains; 14.3% among BN substrains; and 16.7% among F344 substrains.

METHODS

Genotypes

Genotypes of inbred strains of rat, representing the whole genome were collected from the literature or from the World Wide Web, based on the work of the following groups:

1. The Massachusetts General Hospital/Massachusetts Institute of Technology group and associates listed 431 loci for 12 strains (Jacob et al. 1995; http://www-genome.wi.mit.edu/ftp/distribution/rat_sslp_releases/jan95). All of the loci are microsatellites, mostly anonymous markers, and a few intragenic polymorphisms.

2. The Kyoto University group published 126 polymorphic loci. All of the loci are microsatellite markers—90 intragenic ones, typed in 8 strains (Serikawa et al. 1992), some of which were subsequently used by other groups, 13 anonymous ones (Yokoi et al. 1996), and 13 obtained using primers for mouse microsatellites, typed in 11 strains (Kondo et al. 1993).

3. The Arthritis and Rheumatism Branch of The National Institutes of Health (Remmers et al. 1992, 1993a,b, 1995a,b; Goldmuntz et al. 1993a,b, 1995; Mathern et al. 1993, 1994; Zha et al. 1993, 1994; Du et al. 1994, 1995a,b; Ding et al. 1996) typed 130 loci on 16 strains. Most of the loci are microsatellites, with a few random amplified polymorphic DNA (RAPD) markers. Several markers are common to the MGH/MIT group and the Kyoto group.

4. The Wellcome Trust Centre for Human Genetics group at Oxford University (Kreutz et al. 1995; Gauguier et al. 1996; ftp://ftp.well.ox.ac.uk/pub/genetics/ratmap/) listed 275 loci for 11 strains, some with substrains, for a total of 24 lines. All of the loci are microsatellites, mostly developed at Oxford, with some common to the MGH/MIT group.

5. The Department of Laboratory Animal Science at Utrecht University typed 61 inbred strains and substrains at 37 MIT microsatellites (Otsen et al. 1995).

6. The Institute for Animal Experimentation at Tokushima University lists 58 markers for 141 strains (http://www.anex.med.tokushima-u.ac.jp/rat/w433-e.html). This data set, different from the others, is not based on genotypes, but on biochemical polymorphisms. Several substrains were also typed for most strains, so that the collection comprises a total of 433 lines of rats.

7. Other various groups reported genotypes obtained with different techniques from both anonymous and coding DNA (Mori et al. 1989; Canzian et al. 1995b; Deng and Rapp 1995; Deng et al. 1995; Kershaw et al. 1995).

Phylogenetic Analysis

The information from each source was compiled into a single file, listing all of the genotypes for the different strains. Of 145 strains, 63 were contained in at least two data sets and were typed at a sufficient number of loci; therefore, they were selected for the phylogenetic analysis. After pooling the data of markers typed by two or more groups, a total of 995 loci remained.

The Tokushima data set lists many strains with multiple sublines. These were not included in the comprehensive list, for two reasons: (1) Increasing the number of taxa in a tree is computationally difficult and reduces the probability of obtaining the correct tree; (2) most of the substrains are listed only in the Tokushima data set; therefore, insufficient data are available for meaningful comparisons. In the case of strains with multiple sublines, it was assumed that all substrains derive from the same inbred line and that the observed degree of differences among substrains is attributable either to contamination subsequent to the dissemination of the primary stock or fixation of residual heterozygosity. In such cases, a "consensus" type profile of the strain in question was compiled, choosing the genotype common to at least two-thirds of the substrains and leaving as unknown the genotypes of the loci for which no prevalent genotype was found.

For the comprehensive data set, a triangular matrix with the percentages of genotype differences was calculated by pairwise comparison of the strains, using a macro program written in Microsoft Excel Visual Basic (available from F.C.). The matrix was used as input for the program KITSCH with the software package PHYLIP (V. 3.572; J. Felsenstein, University of Washington, Seattle), which uses the Fitch–Margoliash distance matrix method (Fitch and Margoliash 1967). In the case of the 36 strains for which substrains are reported, trees of substrains were calculated separately with the method described above and data from the Tokushima data set; these substrain trees were then pasted onto the main tree.

Difference matrices and phylogenetic trees also were obtained separately for each data set (not shown).

ACKNOWLEDGMENTS

I thank Dr. Tommaso A. Dragani (Istituto Nazionale Tumori, Milan, Italy) for his encouragement and critical reading of the manuscript, Dr. Martina Creaven for correction of English, and Mrs. Mami Igarashi for help at all stages of this study.

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

Canzian, F., T. Ushijima, R. Pascale, T. Sugimura, T.A. Dragani, and M. Nagao. 1995a. Construction of a phylogenetic tree for inbred strains of rat by arbitrarily primed polymerase chain reaction (AP-PCR). *Mamm. Genome* **6:** 231–235.

Canzian, F., T. Ushijima, M. Nagao, I. Matera, G. Romeo, and I. Ceccherini. 1995b. Genetic mapping of the *RET* protooncogene on rat chromosome 4. *Mamm. Genome* **6:** 433–435.

Deng, A.Y. and J.P. Rapp. 1995. Linkage mapping of the endothelin-converting enzyme gene (*Ednce*) to rat chromosome 5. *Mamm. Genome* **6:** 759–760.

Deng, A.Y., J.P Rapp, H. Kato, and M.T. Bihoreau. 1995. Linkage mapping of the neuronal nitric oxide synthase gene (*Nos1*) to rat chromosome 12. *Mamm. Genome* **6:** 824.

Dietrich, W.F., J.C Miller, R.G. Steen, M. Merchant, D. Damron, R. Nahf, A. Gross, D.C. Joyce, M. Wessel, R.D. Dredge, A. Marquis, L.D. Stein, N. Goodman, D.C. Page, and E.S. Lander. 1994. A genetic map of the mouse with 4,006 simple sequence length polymorphisms. *Nature Genet.* **7:** 220–245.

Ding, Y.-P., E.F. Remmers, Y. Du, R.E. Longman, E.A. Goldmuntz, H. Zha, S. Kotake, G.W. Cannon, M.M. Griffiths, and R.L. Wilder. 1996. Genetic maps of polymorphic DNA loci on rat chromosome 1. *Genomics* **36:** 320–327.

Du, Y., E.F. Remmers, E.A. Goldmuntz, H. Zha, P. Mathern, L.J. Crofford, and R.L. Wilder. 1994. Simple sequence repeat length polymorphisms mapped to rat chromosome 11. *Cytogenet. Cell Genet.* **65:** 186–189.

Du, Y., E.F. Remmers, H. Zha, E.A. Goldmuntz, P. Mathern, L.J. Crofford, J. Szpirer, C. Szpirer, and R.L. Wilder. 1995a. Genetic map of eight microsatellite markers comprising two linkage groups on rat chromosome 6. *Cytogenet. Cell Genet.* **68:** 107–111.

Du, Y., E.F. Remmers, H. Zha, E.A. Goldmuntz, J. Szpirer, C. Szpirer, and R.L. Wilder. 1995b. A genetic map of microsatellite markers on rat chromosome 7. *Mamm. Genome* **6:** 295–298.

Festing, M.F.W. 1979. Inbred strains. In *The laboratory rat* (ed. H.J. Baker, J.R. Lindsey, and S.H. Weisbroth), pp. 55–73. Academic Press, New York, NY.

Fitch, W.M. and W.R. Atchley. 1985. Evolution in inbred strains of mice appears rapid. *Science* **228:** 1169–1175.

Fitch, W.M. and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* **155:** 279–284.

Gauguier, D., P. Froguel, V. Parent, C. Bernard, M.T. Bihoreau, B. Portha, M.R. James, L. Penicaud, M. Lathrop, and A. Ktorza. 1996. Chromosomal mapping of genetic loci associated with non-insulin dependent diabetes in the GK rat. *Nature Genet.* **12:** 38–43.

Goldmuntz, E.A., E.F. Remmers, H. Zha, J.M. Cash, P. Mathern, L.J. Crofford, and R.L. Wilder. 1993a. Genetic map of 12 polymorphic loci on rat chromosome 1. *Genomics* **16:** 761–764.

Goldmuntz, E.A., E.F. Remmers, H. Zha, P. Mathern, Y. Du, L.J. Crofford, and R.L. Wilder. 1993b. Genetic map of seven polymorphic markers comprising a single linkage group on rat chromosome 5. *Mamm. Genome* **4:** 670–675.

Goldmuntz, E.A., E.F. Remmers, Y. Du, H. Zha, P. Mathern, L.J. Crofford, and R.L. Wilder. 1995. Genetic map of 16 polymorphic markers forming three linkage groups assigned to rat chromosome 4. *Mamm. Genome* **6:** 459–463.

Jacob, H.J., D.M. Brown, R.K. Bunker, M.J. Daly, V.J. Dzau, A. Goodman, G. Koike, V. Kren, T. Kurtz, Å. Lernmark, G. Levan, Y.-P. Mao, A. Pettersson, M. Pravenec, J.S. Simon, C. Szpirer, J. Szpirer, M.R. Trolliet, E.S. Winer, and E.S. Lander. 1995. A genetic linkage map of the laboratory rat, *Rattus norvegicus*. *Nature Genet.* **9:** 63–69.

Kershaw, E.E., S.C. Chua, J.A. Williams, E.M. Murphy, and R.L. Leibel. 1995. Molecular mapping of SSRs for *Pgm1* and *C8b* in the vicinity of the rat *fatty* locus. *Genomics* **27:** 149–154.

Kondo, Y., M. Mori, T. Kuramoto, J. Yamada, J.S. Beckmann, D. Simon-Chazottes, X. Montagutelli, J.-L. Guenet, and T. Serikawa. 1993. DNA segments mapped by reciprocal use of microsatellite primers between mouse and rat. *Mamm. Genome* **4:** 571–576.

Kreutz, R., N. Hubner, M.R. James, M.T. Bihoreau, D. Gauguier, G.M. Lathrop, D. Ganten, and K. Lindpaintner. 1995. Dissection of a quantitative trait locus for genetic hypertension on rat chromosome 10. *Proc. Natl. Acad. Sci.* **92:** 8778–8782.

Mathern, P., E.A. Goldmuntz, H. Zha, Y. Du, L.J. Crofford, R.L. Wilder, and E.F. Remmers. 1993. Four polymorphic markers on rat chromosome 12 form a single linkage group. *Biochem. Genet.* **31:** 441–448.

Mathern, P., E.A. Goldmuntz, Y. Du, H. Zha, J.M. Cash, L.J. Crofford, R.L. Wilder, and E.F. Remmers. 1994. Nine polymorphic markers characterized by polymerase chain reaction techniques form two linkage groups on rat chromosome 8. *Cytogenet. Cell Genet.* **66:** 283–286.

Mori, M., K. Ishizaki, T. Yamada, H. Chen, T. Sugiyama, T. Serikawa, and J. Yamada. 1989. Restriction fragment length polymorphisms of the angiotensinogen gene in inbred rat strains and mapping of the gene on chromosome 19q. *Cytogenet. Cell Genet.* **50:** 42–45.

Otsen, M., M. Den Bieman, E.S. Winer, H.J. Jacob, J. Szpirer, C. Szpirer, K. Bender, and L.F. Van Zutphen. 1995. Use of simple sequence length polymorphisms for genetic characterization of rat inbred strains. *Mamm. Genome* **6:** 595–601.

Palm, J. and G. Black. 1971. Interrelationships of inbred rat strains with respect to Ag-B and non Ag-B antigens. *Transplantation* **11:** 184–189.

Remmers, E.F., E.A. Goldmuntz, J.M. Cash, L.J. Crofford, B. Misiewicz-Poltorak, H. Zha, and R.L. Wilder. 1992. Genetic map of nine polymorphic loci comprising a single linkage group on rat chromosome 10: Evidence for linkage conservation with human chromosome 17 and mouse chromosome 11. *Genomics* **14:** 618–623.

Remmers, E.F., E.A. Goldmuntz, H. Zha, L.J. Crofford, J.M. Cash, P. Mathern, Y. Du, and R.L. Wilder. 1993a. Linkage

map of seven polymorphic markers on rat chromosome 18. *Mamm. Genome* **4:** 265–270.

Remmers, E.F., E.A. Goldmuntz, H. Zha, P. Mathern, Y. Du, L.J. Crofford, and R.L. Wilder. 1993b. Linkage map of nine loci defined by polymorphic DNA markers assigned to rat chromosome 13. *Genomics* **18:** 277–282.

Remmers, E.F., Y. Du, H. Zha, E.A. Goldmuntz, and R.L. Wilder. 1995a. Ten polymorphic DNA loci, including five in the rat MHC (*RT1*) region, form a single linkage group on rat chromosome 20. *Immunogenetics* **41:** 316–319.

Remmers, E.F., E.A. Goldmuntz, J.M. Cash, H. Zha, L.J. Crofford, B. Misiewicz-Poltorak, P. Mathern, and R.L. Wilder. 1995b. Map of seven polymorphic markers on rat chromosome 14: Linkage conservation with human chromosome 4. *Mamm. Genome* **4:** 90–94.

Serikawa, T., T. Kuramoto, P. Hilbert, M. Mori, J. Yamada, C.J. Dubay, K. Lindpainter, D. Ganten, J.-L. Guenet, G.M. Lathrop, and J.S. Beckmann. 1992. Rat gene mapping using PCR-analyzed microsatellites. *Genetics* **131:** 701–721.

Toyota, M., F. Canzian, T. Ushijima, Y. Hosoya, T. Kuramoto, T. Serikawa, K. Imai, T. Sugimura, and M. Nagao. 1996. A rat genetic map constructed by Representational Difference Analysis markers with suitability for large scale typing. *Proc. Natl. Acad. Sci.* **93:** 3914–3919.

Yamada, J., T. Kuramoto, and T. Serikawa. 1994. A rat genetic linkage map and comparative maps for mouse or human homologous rat genes. *Mamm. Genome* **5:** 63–83.

Yokoi, N., K. Kitada, T. Kuramoto, T. Maihara, Y. Andoh, S. Takami, J. Higaki, T. Miki, T. Ogihara, G.M. Lathrop, and T. Serikawa. 1996. Mapping of 20 polymorphic DNA markers in the rat by somatic hybrid and linkage analysis. *Mamm. Genome* **7:** 71–73.

Zha, H., R.L. Wilder, E.A. Goldmuntz, J.M. Cash, L.J. Crofford, P. Mathern, and E.F. Remmers. 1993. Linkage map of 10 polymorphic markers on rat chromosome 2. *Cytogenet. Cell Genet.* **63:** 117–123.

Zha, H., E.F. Remmers, Y. Du, E.A. Goldmuntz, P. Mathern, H. Zhang, J.M. Cash, L.J. Crofford, and R.L. Wilder. 1994. A single linkage group comprising 11 polymorphic DNA markers on rat chromosome 3. *Mamm. Genome* **5:** 538–541.

Received October 28, 1996; accepted in revised form January 21, 1997.