

Ancestral bias in the *Hras1* gene and distal Chromosome 7 among inbred mice

Jennifer C. Drew · Andrew S. Kastenmeier · Norman R. Drinkwater

Received: 23 April 2007 / Accepted: 13 August 2007 / Published online: 29 September 2007
© Springer Science+Business Media, LLC 2007

Abstract Inbred strains of mice vary in their frequency of liver tumors initiated by a mutation in the *Hras1* (*H-ras*) proto-oncogene. We sequenced 4.5 kb of the *Hras1* gene on distal Chr 7 in a diverse set of 12 commonly used laboratory inbred strains of mice and detected no sequence variation to account for strain-specific differences in *Hras1* mutation prevalence. Furthermore, the *Hras1* sequence is essentially monoallelic for an ancestral gene derived from the *M. m. domesticus* species. To determine if the monoallelism and associated low rate of polymorphism are unique to *Hras1* or representative of the general chromosomal locale, we extended the sequence analysis to 12 genes in the final 8 Mb of distal Chr 7. A region of at least 2.5 Mb that encompasses

several genes, including *Hras1* and the *H19/Igf2* loci, demonstrates virtually no sequence variation. The 12 inbred strains share one dominant haplotype derived from the *M. m. domesticus* allele. Chromosomal regions flanking the monoallelic segment exhibit a significantly higher rate of variation and multiple haplotypes, a majority of which are attributed to *M. m. domesticus* or *M. m. musculus* ancestry.

Introduction

Several papers have recently described the first inquiries into the single nucleotide polymorphism (SNP) and haplotype structure of the inbred mouse genome on a global scale (Lindblad-Toh et al. 2000; Pletcher et al. 2004; Tsang et al. 2005; Wade and Daly 2005; Wade et al. 2002; Wiltshire et al. 2003). This research verifies and details a patchwork pattern of variation first posited in 1987, in which pairs of laboratory inbred strains share large blocks derived from one ancestral strain, primarily *M. m. domesticus* or *M. m. musculus* (Bonhomme et al. 1987). Within any pairwise inbred strain comparison, there are haplotype blocks defined by a low or high SNP rate, and the average length of these blocks is estimated to be 1.2–1.4 megabases (Mb) (Frazer et al. 2004; Wade et al. 2002; Zhang et al. 2005). Genome-wide studies of a few inbred strains estimate that the average rate of polymorphism between two such strains is 0.5 SNP/10 kb in the low-SNP-rate blocks and 35 SNP/10 kb in SNP-dense blocks (Frazer et al. 2004). Blocks of limited diversity are attributed to a recent coalescence in which inbred strains inherited the same ancestral allele, while SNP-dense blocks reflect inheritance of divergent ancestral alleles.

Electronic supplementary material The online version of this article (doi:10.1007/s00335-007-9061-1) contains supplementary material, which is available to authorized users.

Nucleotide sequence data reported here are available in NCBI dbSNP build 127 under accession numbers listed in Supplementary Table 2.

J. C. Drew · A. S. Kastenmeier · N. R. Drinkwater (✉)
McArdle Laboratory for Cancer Research, University of Wisconsin School of Medicine and Public Health,
1400 University Avenue, Madison, Wisconsin 53706, USA
e-mail: drinkwater@oncology.wisc.edu

Present Address:

J. C. Drew
Department of Microbiology and Cell Science,
University of Florida, Gainesville, FL 32611, USA

Present Address:

A. S. Kastenmeier
Department of Surgery, University of Utah Health Sciences Center, Salt Lake City, UT 84132, USA

Laboratory inbred strains of mice vary in their frequency of *Hras1* (*H-ras*) mutational activation during multistage hepatocarcinogenesis (Buchmann et al. 1991; J.C. Drew and N.R. Drinkwater, unpublished). Given its significant role in liver tumorigenesis and its strain-dependent frequency of activation, we tested the hypothesis that strain-specific polymorphisms in the *Hras1* gene could account for variable frequencies of *Hras1* initiation in liver tumors. We sequenced approximately 4.5 kb of the *Hras1* gene from 12 diverse, yet commonly used laboratory inbred mouse strains and two distantly related inbred strains. The set includes representatives from all six families of inbred strains as defined by genome-wide parsimony analysis of simple sequence length polymorphisms (Witmer et al. 2003).

In this article we report on the monoallelic inheritance of *Hras1* and its neighbors on distal Chr 7 among 12 classic inbred strains. Sequence analysis revealed a region of remarkably low diversity among the strains. No strain-specific SNPs account for differences in frequency of *Hras1* activation. Additional sequence analysis of surrounding genes in the final 8 Mb of distal Chr 7 exposed a unique 2.5-Mb block that is essentially devoid of any sequence variation among the 12 classic inbred strains. This block is flanked by regions with significantly greater diversity. Analysis of wild-derived inbred strains representing ancestral genomes of *M. m. domesticus* and *M. m. musculus* indicates that the 12 lab inbred strains have fixed alleles from the *M. m. domesticus* progenitor strain in the 2.5-Mb region on distal Chr 7 that includes the *Hras1* gene. Considering the well-established role of *Hras1* in cell signal transduction and of mutant *Hras1* in tumor development, these results have important implications for the study of *Hras1* in mouse models of neoplasia and contribute to the burgeoning understanding of laboratory inbred mouse genomes.

Materials and methods

Sequence analysis of *Hras1* and surrounding genes on Chr 7

The C57BL/6J (B6) *Hras1* and additional 5' and 3' nucleotide sequence was originally obtained from GenBank (accession No. z50013) and the Trace Archive database (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi>). Primer sets for PCR were designed with Primer3 software (http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) using the B6 sequence as a template to amplify a set of 400–500 bp of overlapping products. Genomic DNA from spleens was prepared according to a previously published protocol

(Bilger et al. 2004) from the following strains housed in our colony: B6, C57BR/cdJ, C3H/HeJ, CBA/J, and SM/J. Genomic DNA from 129X1/SvJ, 129P3/J, A/J, AKR/J, BALB/cByJ, DBA/2J, SWR/J, CAST/EiJ, SPRET/EiJ, WSB/EiJ, and CZECHII/Ei was obtained from The Jackson Laboratory (Bar Harbor, ME). Primers for 11 additional genes on distal Chr 7 were designed from the published B6 sequence (Mouse Genome Consortium 2002; http://www.ensembl.org/mus_musculus). For each of the 11 genes, we sequenced a 300–500-bp region. For 8 of 11 genes, the product spanned an intron/exon boundary. All primer sequences and accession numbers are listed in Supplementary Table 1.

PCR reactions included 10–20 ng of DNA, 62.5 μ M dNTPs (Amersham, Piscataway, NJ), 15 pmol of forward and reverse primers (Integrated DNA Technologies, Coralville, IA), 0.3 U Taq DNA polymerase, and 1 \times PCR buffer (Roche, Indianapolis, IN) in a total reaction volume of 20 μ l. The reactions were incubated in thermocyclers at 95°C for 3 min, followed by 40 cycles at 94°C for 15 sec, 60°C for 25 sec, 72°C for 90 sec, and a final step at 72°C for 7 min. Minor modifications to the standard protocol such as optimization of annealing temperatures and initial incubations at 95°C for 15 min with HotStar DNA Polymerase (Qiagen, Valencia, CA) were required for a small number of reactions. All PCR products were purified with a Qiagen PCR Purification Kit according to the recommended protocol, except for the final elution step. Instead of eluting DNA with the suggested buffer, we eluted twice with 50 μ l of ddH₂O. Approximately 35–50 ng of purified PCR products was used for sequencing in separate reactions with 1 pmol of forward and reverse primers using Big Dye terminator chemistry (Perkin Elmer, Boston, MA) and analyzed on an Applied Biosystems 373 DNA automated sequencer (Foster City, CA). Sequencing reactions were incubated at 94°C for 4 min, followed by 35 cycles at 94°C for 30 sec, 55°C for 30 sec, 60°C for 90 sec, and a final step at 60°C for 10 min. The sequences were processed at the sequencing facility at the McArdle Laboratory for Cancer Research.

The sequence of each mouse strain was analyzed and assembled using Phred/Phrap software (Ewing and Green 1998; Ewing et al. 1998) and strain sequences were aligned in ClustalX (Thompson et al. 1997).

SNP rate

The SNP rate, for a given region, was calculated by adding the total number of SNPs observed for all pairwise strain comparisons and dividing by the total number of compared nucleotides. The total number of pairwise comparisons for 12 strains is $(12 \times 11)/2 = 66$. For each region, SNP

rate = number of SNPs/(66 × length of high-quality sequence). The expected number of SNPs for any given region was calculated based on previously published estimates (Frazer et al. 2004; Wade et al. 2002). According to these data, low-SNP-rate regions define approximately two thirds of each inbred genome comparison with an overall low SNP rate of 0.5 SNP/10 kb, and high-SNP-rate regions comprise one third of the genome with an overall SNP rate of 35 SNP/10 kb. Therefore, to calculate the expected SNP rate within a region:

$$[0.67(0.5 \text{ SNP}/10 \text{ kb})] + [0.33(34 \text{ SNP}/10 \text{ kb})] = 11.6 \text{ SNP}/10 \text{ kb}$$

To determine the expected number of SNPs within a particular-size block:

$$(1.16 \text{ SNP}/\text{kb}) \times (66 \text{ comparisons}) \times (\text{kb of sequence})$$

Phylogenetic analysis

Sequence data from seven genes were concatenated for each of the 12 lab inbred strains. The genes included in the analysis were *Mgmt*, *Ebf3*, *Dpysl4*, *Hras1*, *Fgf3*, *Fgf4*, and *Ccnd1*. The linked strain sequences were aligned with ClustalX (Thompson et al. 1997). Loss-parsimony analysis according to the method of Fitch (1971) was executed with the DNAPARS program in the PHYLIP package. The output data were used to generate a phylogenetic tree with the DRAWGRAM program in the PHYLIP package (Felsenstein 1989).

Results

Sequence of the *Hras1* gene in 14 inbred strains

To identify any sequence variants in the *Hras1* gene that could account for strain differences in the frequency of *Hras1*-initiated liver tumors, we sequenced the gene from 14 inbred strains of mice: 12 classic laboratory inbred strains (A/J, AKR/J, C57BL/6J, C57BR/cdJ, BALB/cByJ, C3H/HeJ, CBA/J, DBA/2J, SM/J, SWR/J, 129X1/SvJ, and 129P3/J) and 2 distantly related inbred strains, *M. m. castaneus* (CAST/EiJ) and *M. m. spretus* (SPRET/EiJ). The 4.475-kb sequenced region of the *Hras1* gene spans all of the introns and exons and 0.9 kb of 5'- and 0.6 kb of 3'-flanking sequence based on the *Hras1* transcript expressed in the mouse liver (Counts et al. 1997). In this approximately 4.5 kb of sequence, 11 of the 12 classic inbred strains have exactly the same sequence. The only exception is the AKR/J (AKR) strain, which harbors two unique SNPs. Each of the SNPs is located in noncoding intronic sequence: one in intron 1 and one in the 3'-flanking region (Fig. 1).

When AKR is compared to each of the 11 other laboratory inbred strains, the two AKR SNPs aggregate to a total of 22 SNPs among all inbred pairwise comparisons, yielding an average of 0.3 SNPs between any 2 of the 12 inbred strains. This SNP rate is a much lower level of sequence variation than expected without any prior knowledge of the diversity of the region. Based on recent global estimates of sequence diversity, we expected to observe a total of 344 SNPs, or approximately 5 SNPs for

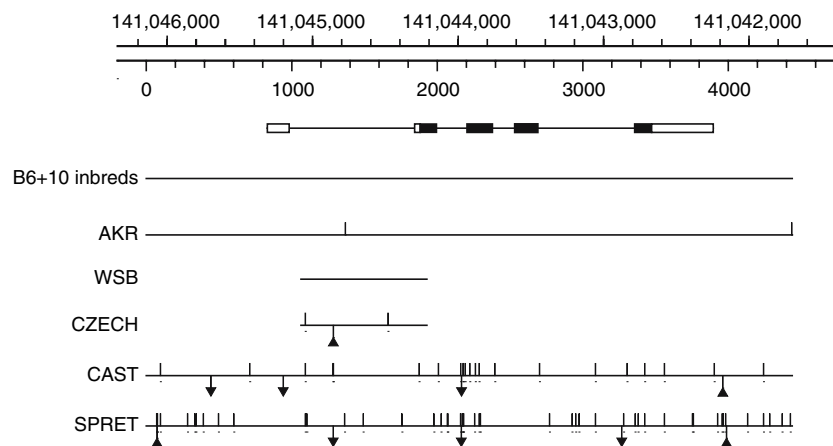


Fig. 1 Sequence variation of the *Hras1* gene across 14 inbred strains. To account for strain-dependent variation in mutant *Hras1* frequencies, we sequenced approximately 4.5 kb of the *Hras1* gene that spanned noncoding 5' and 3' regions from 14 strains of mice, including 12 common lab inbred strains and 2 distantly related strains (CAST and SPRET). The genomic position of *Hras1* along Chr 7, according to the NCBI m36 mouse assembly, is denoted on the top line of the figure with the sequence length demarcated on the

following line (www.ensembl.org). The genomic *Hras1* structure is shown with blocks indicating exons and the shaded portions representing the translated regions. This structure corresponds to an approximately 1-kb transcript expressed in the mouse liver (Counts et al. 1997). The horizontal lines next to the strain names indicate the genomic sequence coverage of that particular strain(s). Vertical hash marks represent SNPs, and arrows indicate insertions or deletions

any two-strain comparison (Frazer et al. 2004; Wade et al. 2002). The observed variation is a 17-fold reduction in the expected number of SNPs and is highly significant (the expected SNP rate is outside of the 99.99% confidence interval for the observed SNP rate).

As anticipated, CAST and SPRET genomes had a greater number of SNPs compared with the common laboratory inbred strains with 26 and 51 SNPs each, respectively (versus B6 and 10 other inbred strains); these results are consistent with previous estimations of 1 SNP/200 bp between CAST or SPRET and the laboratory inbred mouse (Frazer et al. 2004; Lindblad-Toh et al. 2000).

Ancestral origin of the common *Hras1* allele

To determine the origin of the inbred *Hras1* gene, we resequenced an 870-bp region that contains the AKR SNP located in intron 1 using genomic DNA from the WSB/EiJ (WSB) and CZECHII/EiJ (CZECH) strains. WSB and CZECH are wild-derived inbred strains representative of the ancestral species *M. m. domesticus* and *M. m. musculus*, respectively (Wade et al. 2002). Three SNPs were detected between the two archetypal strains, WSB and CZECH (SNP rate = 34/10 kb), and this rate is consistent with the upper level of diversity commonly observed between any two laboratory inbred strains (Frazer et al. 2004). All 12 of the classic laboratory inbred strains, including AKR, share the WSB nucleotides at this location (Fig. 1). The AKR SNP in intron 1 is unique because it is not observed in the WSB or CZECH sequence (Supplementary Table 2). Thus, the 12 laboratory inbred strains are effectively monoallelic for the *M. m. domesticus*-derived allele.

SNP frequencies of 12 genes on distal Chr 7

Hras1 resides in a gene-rich chromosomal region at the distal end of Chr 7. Because *Hras1* is monoallelic for the *M. m. domesticus*-derived allele and has a surprisingly low rate of variation, we sampled surrounding genes to determine the overall SNP frequency and haplotype structure of the distal-most 8-Mb chromosomal region. DNA from CAST, SPRET, WSB, CZECH, and the 12 laboratory inbred strains was used as a template to sequence 300–500-bp PCR products from 11 genes on distal Chr 7 (*Mgmt*, *Ebf3*, *Dpysl4*, *Echs1*, *Ptdss2*, *Rassf7*, *Dusp8*, *Igf2*, *Fgf3*, *Fgf4*, and *Ccnd1*). Supplementary Table 2 contains SNP data for each sampled gene. The average rate of sequence variation for all pairwise comparisons among the 12 laboratory inbred strains for each gene was calculated by the same method used in the *Hras1* analysis and is plotted against physical position in the lower panel of Fig. 2. The

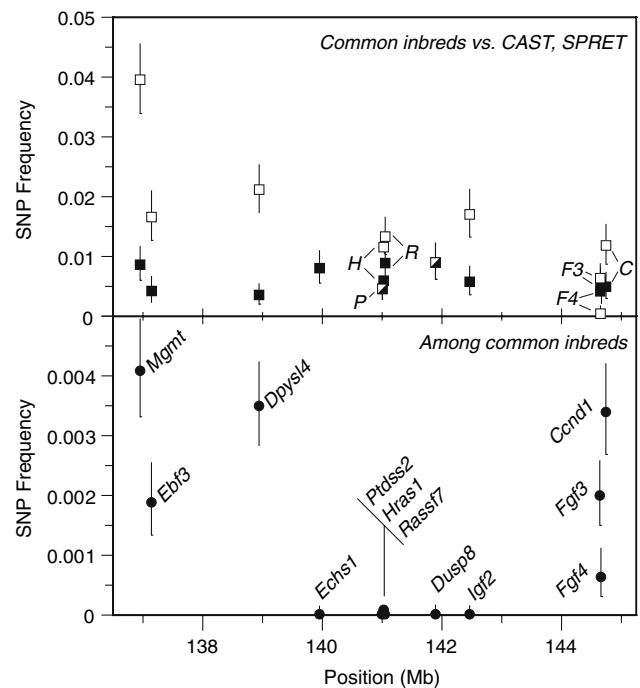


Fig. 2 Rates of variation of 12 genes in the distal-most 8 Mb of Chr 7. *Lower panel:* The frequency of SNPs among 12 lab inbred strains and their 95% confidence intervals for each gene are plotted against the chromosomal position. Each data point is labeled with its gene symbol. SNP frequency is defined as the total number of variants observed divided by the total number of bases compared for all 66 pairwise strain comparisons and is displayed as the number of SNPs per bp of compared sequence. Chromosomal position is assigned according to the NCBI m36 mouse assembly. *Upper panel:* The frequency of SNPs and their 95% confidence intervals for the lab inbred strains versus CAST (closed squares) or SPRET (open squares) are plotted against chromosomal position. The frequency of SNPs is displayed as the number of SNPs per bp of compared sequence. The data for each gene is directly above the corresponding information in the lower panel. For overlapping data points, a one- or two-character abbreviation is used to help distinguish the genes. SPRET sequence was not available for the *Echs1* gene, and the rates for the inbred strains versus CAST and SPRET were identical for the *Ptdss2* and *Dusp8* genes so only one symbol (half-filled) is shown

sequenced regions for five genes surrounding *Hras1* at positions 139.96–142.47 Mb were entirely devoid of any sequence variation as all 12 laboratory inbred strains were genetically identical at these locations. The combined analysis of the six genes, including *Hras1*, in this low-polymorphism region yields a SNP rate of 0.48 SNP/10 kb. For a region of this size in 12 laboratory inbred strains, the 2.5-Mb region has strikingly fewer SNPs than expected: 22 vs. 494.

Two regions of considerably greater diversity border the conserved block. The three sampled genes located in the 136.96–138.95-Mb interval, *Mgmt*, *Ebf3*, and *Dpysl4*, have SNP frequencies of 20–40 SNP/10 kb. The three sampled genes located in the final megabase of sequence on Chr 7, *Fgf3*, *Fgf4*, and *Ccnd1*, have similar, albeit slightly lower,

frequencies in the range of 5–30 SNP/10 kb for all pairwise comparisons (Fig. 2, lower panel). For 9 of the 12 genes, the sequenced fragment spans an intron-exon junction. For three genes, *Mgmt*, *Ebf3*, and *Igf2*, the sequenced region consists entirely of intronic sequence. For genes other than *Hras1*, all sequenced exon regions are coding, and there is approximately twice as much intron sequence as exon sequence (2741 bp vs. 1392). Supplementary Table 1 contains the exon and intron coverage per sequenced gene region. The average SNP rates of exonic and intronic sequence are 6 and 14 SNP/10 kb, respectively (data not shown), and this ratio is consistent with previous observations (Thomas et al. 2002).

The diversity between the laboratory inbred mouse and CAST or SPRET is significantly higher ($p < 0.05$, Fisher's exact test) than the level of variation among the laboratory inbred strains, and this SNP rate is constant throughout the 8-Mb region (Fig. 2, upper panel).

Haplotype and phylogenetic analysis of 12 genes in the final 8 Mb of Chr 7

All 12 laboratory inbred strains share WSB alleles in an interval that spans at least 2.5 Mb between 140.0–142.5 Mb, including *Hras1* and the five surrounding genes (Fig. 1; Supplementary Table 2). The interval size may extend to the outer borders of the two flanking regions of greater diversity (138.9–144.6 Mb), which means that the interval size lies between 2.5 and 5 Mb. Figure 3 depicts the overall haplotype structure of the entire 8-Mb region surveyed in this analysis. For the five genes with increased sequence diversity in the proximal and distal ends of the distal 8-Mb Chr 7 region, there are multiple haplotypes represented among the 12 laboratory inbred strains. For each gene fragment, multiple strains share identical haplotypes with WSB, CZECH, or, occasionally, unique haplotypes of unknown origin. Although with 12 different laboratory inbred strains, 12 distinct haplotypes are theoretically possible, no gene had more than three strain distribution patterns. Throughout the entire 9.1 kb of sequence for all of the 12 laboratory inbred strains, there is a total of three novel nucleotides (two in *Hras1* and one in *Fgf4*, each represented in a single strain). Most unexpectedly, several strains share identical haplotypes, without any SNPs, that span the entire 8 Mb.

The phylogeny of this Chr 7 region, as represented by the SNP data shown in Fig. 3, was analyzed using the PHYLIP loss-parsimony program (Felsenstein 1989). The resulting tree (Supplementary Fig. 1) reveals relatedness at distal Chr 7 that does not match the strains' overall relatedness based on genealogic, microsatellite, and SNP data (Beck et al. 2000; Petkov et al. 2004; Tsang et al. 2005;

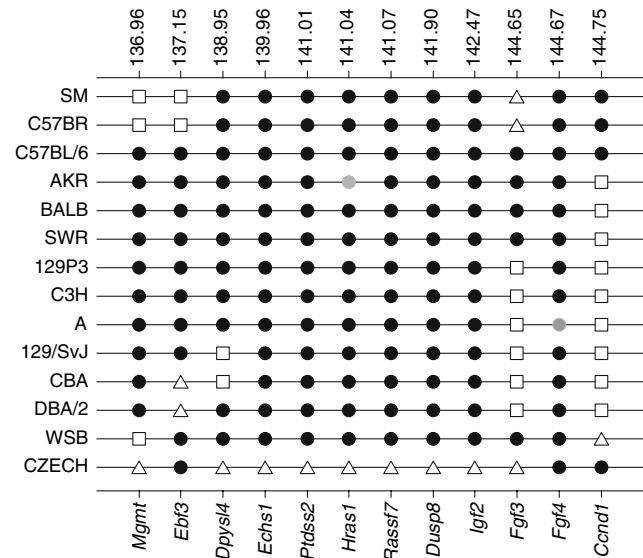


Fig. 3 Distal Chr 7 features a 2.5–5-Mb monoallelic block of *M. m. domesticus* origin. The 12 sampled loci are listed on the lower X axis with their corresponding position on the upper X axis. The strain names are listed on the left Y axis. Each strain at every position is indicated with a symbol and the distance between symbols along each row is fixed, not scaled, to actual positions. Each strain that shares the B6 haplotype at a particular locus has a black circle. The gray circles indicate strains with orphan SNPs (novel mutations) but an overall similarity to the B6 strain. For each gene, strains that have exactly the same haplotype have identical symbols. The strains are arranged by degree of similarity. Strains with similar haplotype patterns across the entire 8-Mb distal Chr 7 region are clustered together. WSB and CZECH haplotype information is denoted on the last two rows

Witmer et al. 2003). For instance, in genome-wide analyses, 129P3/J and C3H/HeJ align to two separate branches although they are identical at the distal Chr 7 region. C57BR/cdJ and C57BL/6J are closely related in genome-wide scans, but these two strains are distinct in the distal Chr 7 region and C57BR/cdJ and SM/J are entirely identical. These results are consistent with previous observations that local and global hereditary relationships often differ (Frazer et al. 2004; Park et al. 2002).

Validation of the haplotype structure

Previous estimates indicate that *M. m. domesticus* ancestry accounts for two thirds of the laboratory inbred mouse genome (Wade et al. 2002). Given this proportion, the chance of randomly encountering a locus in which 12 laboratory inbred strains have all fixed the *M. m. domesticus* allele, as for the monoallelic block on distal Chr 7, is $(0.67)^{12} = 0.008$ per block, or about 10 per genome. To determine the frequency of such blocks empirically, and to validate the presence of this large, monoallelic block on distal Chr 7, we queried the Mouse Phenome Database (<http://www.aretha.jax.org/pub-cgi/phenome/mpdcgi?rtm=>

[docs/home](#)). Although this database has information for all 12 lab inbred strains, the level of strain coverage varies. We restricted our analysis to the eight strains in our set for which at least 155,000 SNPs per genome had been analyzed: 129X1/SvJ, A/J, AKR/J, BALB/cByJ, C3H/HeJ, B6, DBA/2J, and SM/J. We discovered that for 354 nucleotide positions in a 2-Mb (141–143 Mb) window that lies within the monoallelic block on distal Chr 7, 100% were identical for all eight strains. We assessed the level of SNP variation throughout the entire genome to determine how often one would encounter a 2-Mb block with this level of sequence similarity at random among eight laboratory inbred strains. We analyzed all entered SNP information for the eight sampled strains in 2-Mb sliding windows. We identified five hundred 2-Mb intervals that contained information for at least 50 SNPs for the eight strains queried. In this analysis, 1% of these 2-Mb blocks are identical for all nucleotide positions in all eight laboratory strains. However, the block on distal Chr 7 had the highest level of significance as determined by comparing the lower bounds of the 95% confidence intervals (data not shown). Therefore, although the monoallelic block on distal Chr 7 is not singular, blocks of this size with similar sequence conservation (100% similarity among many strains) are rare in the mouse genome.

Discussion

Remarkably, a 2.5-Mb region on distal Chr. 7 that encompasses several genes, including *Hras1*, is essentially monoallelic among many laboratory inbred strains, all of which have fixed the *M. m. domesticus* allele. Although the AKR/J strain harbors two SNPs in the *Hras1* gene, they are both in noncoding regions. The AKR/J strain develops *Hras1* mutations in approximately 4% of its chemically induced liver tumors, which is very similar to the phenotype of the C57BL/6J and DBA/2J strains (J.C. Drew and N.R. Drinkwater, unpublished). These observations support the inference that the AKR-specific SNPs are phenotypically silent.

As a result of this high degree of haplotype conservation, we conclude that a 4.5-kb region carrying the *Hras1* gene is not responsible for strain-dependent variation in *Hras1* mutation rates in liver tumors. Although sequence analysis included the promoter regions and the 3' untranslated region, regulatory elements beyond the sequenced domain may affect expression and result in variable rates of *Hras1*-initiated liver tumors. We analyzed *Hras1* expression in the livers of five of the classic inbred strains (129X1/SvJ, C57BR/cdJ, C57BL/6J, C3H/HeJ, and SM/J) by microarray and did not observe any significant differences among the strains (A. Bilger, personal communication).

The extremely low rate of variation in the 2.5-5-Mb *Hras1* region, which contains up to 100 genes, is highly unusual. The overall SNP rate of the monoallelic region on distal Chr 7 is 0.48 SNPs/10 kb and is consistent with the observed rate in low-SNP regions for a single pair of strains. However, this low SNP rate is very rare for comparisons among many strains. Frazer et al. (2004) recently suggested that haplotype analysis with 12 inbred strains captures greater than 95% of the variation likely to be present among the classic laboratory inbred strains. The set of 12 lab inbred strains presented here are genetically diverse and include members from the six major families of inbred strains (Witmer et al. 2003). The monoallelic region on distal Chr 7 is the only block reported with such a low rate of variation over 66 pairwise strain comparisons. The two less-conserved regions flanking the monoallelic block are in agreement with the published average rate of 35 SNP/10 kb in a SNP-dense block, and our strain distribution of ancestral alleles is in accordance with the predicted proportion of two thirds with *M. m. domesticus* ancestry (Frazer et al. 2004; Wade et al. 2002).

The monoallelic interval, or SNP “desert,” on distal Chr 7 affects QTL analysis in two ways. First, this SNP desert is effectively noninformative and precludes linkage analysis in this region for pairs of the aforementioned, frequently used, 12 laboratory inbred strains. Second, the lack of genotypic differences among the 12 common lab strains examined in this study suggests that phenotypic differences are unlikely to be the direct result of genes located in this monoallelic 2-5-Mb interval, which includes *Hras1*, *H19*, *Igf2*, *Insulin 2 precursor (Ins2)*, and *Cyclin-dependent kinase inhibitor 1c (Cdkn1c)* (Mouse Genome Informatics, www.informatics.jax.org).

Functional selection may account for the shared *M. m. domesticus* ancestry in the distal Chr 7 monoallelic region. However, the *Hras1* gene is unlikely to be the driving force of such selection because *Hras1* is not an essential gene and it demonstrates functional redundancy with other RAS family members (Esteban et al. 2001; Ise et al. 2000; Johnson et al. 1997). The monoallelic block also includes the well-studied *H19/Igf2* imprinted region, which could drive selection for this haplotype. To our knowledge, this is the first detailed analysis of sequence variation among a large number of inbred strains in any imprinted region. Integrity of human imprinting clusters is essential and may be important for regulating expression. Since imprinted clusters are well conserved between human and mouse, murine imprinted regions may also require strict conservation (Casparly et al. 1998; Reik and Walter 2001). However, mice lacking functional *Igf2* are viable, which suggests that *Igf2* is not solely driving selection (Constania et al. 2000). A congenic mouse line carrying the *M. m. musculus* distal Chr 7 region on a *M. m. domesticus* or

standard inbred background would reveal whether the *M. m. musculus* Chr 7 allele results in a selective disadvantage.

Reduced variation is often observed at chromosomal regions linked to a beneficial mutation driving positive selection (Kaplan et al. 1989). For example, Schlenke and Begeen (2004) identified a chromosomal invariant region in a *Drosophila simulans* population that was best explained by recent fixation due to directional selection. Candidate gene evaluation of the region led to a locus containing a transposon that is associated with increased resistance to insecticides and thereby potentially improves fitness. As this example shows, monoallelic blocks can be signposts for biologically relevant loci and conserved regions such as that on distal mouse Chr 7 may be fruitful to explore.

Acknowledgments The authors are grateful to Kristin Liss, Kim Leutkehoelter, Andrew Schneider, and Matthew Gigot for expert technical assistance in various aspects of this work. They also thank Andrea Bilger for critical advice on the manuscript. This work was supported by grants CA22484 and CA009135 from the National Cancer Institute of the National Institutes of Health.

References

- Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, et al. (2000) Genealogies of mouse inbred strains. *Nat Genet* 24:23–25
- Bilger A, Bennett LM, Carabeo RA, Chiaverotti TA, Dvorak C, et al. (2004) A potent modifier of liver cancer risk on distal mouse chromosome 1: linkage analysis and characterization of congenic lines. *Genetics* 167:859–866
- Bonhomme F, Guenet JL, Dod B, Moriwaki K, Bulfield G (1987) The polyphyletic origin of laboratory inbred mice and their rate of evolution. *J Linn Soc* 30:51–58
- Buchmann A, Bauer-Hofmann R, Mahr J, Drinkwater NR, Luz A, et al. (1991) Mutational activation of the c-Ha-ras gene in liver tumors of different rodent strains: correlation with susceptibility to hepatocarcinogenesis. *Proc Natl Acad Sci U S A* 88:911–915
- Caspary T, Cleary MA, Baker CC, Guan XJ, Tilghman SM (1998) Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol Cell Biol* 18:3466–3474
- Constancia M, Dean W, Lopes S, Moore T, Kelsey G, et al. (2000) Deletion of a silencer element in *Igf2* results in loss of imprinting independent of *H19*. *Nat Genet* 26:203–206
- Counts JL, McClain RM, Goodman JI (1997) Comparison of effect of tumor promoter treatments on DNA methylation status and gene expression in B6C3F1 and C57BL/6 mouse liver and in B6C3F1 mouse liver tumors. *Mol Carcinog* 18:97–106
- Esteban LM, Vicario-Abejon C, Fernandez-Salguero P, Fernandez-Medarde A, Swaminathan N, et al. (2001) Targeted genomic disruption of *H-ras* and *N-ras*, individually or in combination, reveals the dispensability of both loci for mouse growth and development. *Mol Cell Biol* 21:1444–1452
- Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred. II. Error Probabilities. *Genome Res* 8:186–194
- Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using Phred. I. Accuracy Assessment. *Genome Res* 8:175–185
- Felsenstein J (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164–166
- Fitch WM (1971) Toward defining the course of evolution: minimum change for a specified tree topology. *Syst Zool* 20:406–416
- Frazer KA, Wade CM, Hinds DA, Patil N, Cox DR, et al. (2004) Segmental phylogenetic relationships of inbred mouse strains revealed by fine-scale analysis of sequence variation across 4.6 Mb of mouse genome. *Genome Res* 14:1493–1500
- Ise K, Nakamura K, Nakao K, Shimizu S, Harada H, et al. (2000) Targeted deletion of the *H-ras* gene decreases tumor formation in mouse skin carcinogenesis. *Oncogene* 19:2951–2956
- Johnson L, Greenbaum D, Cichowski K, Mercer K, Murphy E, et al. (1997) *K-ras* is an essential gene in the mouse with partial functional overlap with *N-ras*. *Genes Dev* 11:2468–2481
- Kaplan NL, Hudson RR, Langley CH (1989) The “hitchhiking effect” revisited. *Genetics* 123:887–899
- Lindblad-Toh K, Winchester E, Daly MJ, Wang DG, Hirschhorn JN, et al. (2000) Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nat Genet* 24:381–386
- Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562
- Park YG, Lukes L, Yang H, Debies MT, Samant RS, et al. (2002) Comparative sequence analysis in eight inbred strains of the metastasis modifier QTL candidate gene *Brms1*. *Mamm Genome* 13:289–292
- Petkov PM, Ding Y, Cassell MA, Zhang W, Wagner G, et al. (2004) An efficient SNP system for mouse genome scanning and elucidating strain relationships. *Genome Res* 14:1806–1811
- Pletcher MT, McClurg P, Batalov S, Su AI, Barnes SW, et al. (2004) Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* 2:e393
- Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2:21–32
- Schlenke TA, Begun DJ (2004) Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc Natl Acad Sci U S A* 101:1626–1631
- Thomas RS, Penn SG, Holden K, Bradfield CA, Rank DR, et al. (2002) Sequence variation and phylogenetic history of the mouse *Ahr* gene. *Pharmacogenetics* 12:151–163
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, et al. (1997) The CLUSTALX Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Tsang S, Sun Z, Luke B, Stewart C, Lum N, et al. (2005) A comprehensive SNP-based genetic analysis of inbred mouse strains. *Mamm Genome* 16:476–480
- Wade CM, Daly MJ (2005) Genetic variation in laboratory mice. *Nat Genet* 37:1175–1180
- Wade CM, Kulbokas EJ, Kirby AW, Zody MC, Mullikin JC, et al. (2002) The mosaic structure of variation in the laboratory mouse genome. *Nature* 420:574–578
- Wiltshire T, Pletcher MT, Batalov S, Barnes SW, Tarantino LM, et al. (2003) Genome-wide single-nucleotide polymorphism analysis defines haplotype patterns in mouse. *Proc Natl Acad Sci U S A* 100:3380–3385
- Witmer PD, Doheny KF, Adams MK, Boehm CD, Dizon JS, et al. (2003) The development of a highly informative mouse Simple Sequence Length Polymorphism (SSLP) marker set and construction of a mouse family tree using parsimony analysis. *Genome Res* 13:485–491
- Zhang J, Hunter KW, Gandolph M, Rowe WL, Finney RP, et al. (2005) A high-resolution multistrain haplotype analysis of laboratory mouse genome reveals three distinctive genetic variation patterns. *Genome Res* 15:241–249