Selection on *cis*-Regulatory Variation at *B4galnt2* and Its Influence on von Willebrand Factor in House Mice

Jill M. Johnsen,[1] *Meike Teschke,† Pavlos Pavlidis,‡ Beth M. McGee,§|| Diethard Tautz,† David Ginsburg,*§|| and John F. Baines‡*

*Department of Internal Medicine, University of Michigan; †Max-Planck Institute for Evolutionary Biology, Plön, Germany; ‡Department Biology II, Section of Evolutionary Biology, University of Munich (LMU), Planegg-Martinsried, Germany; §Department of Human Genetics, Howard Hughes Medical Institute, University of Michigan; and ||The Life Sciences Institute, University of Michigan

The RIIIS/J inbred mouse strain is a model for type 1 von Willebrand disease (VWD), a common human bleeding disorder. Low von Willebrand factor (VWF) levels in RIIIS/J are due to a regulatory mutation, *Mvwf1*, which directs a tissue-specific switch in expression of a glycosyltransferase, B4GALNT2, from intestine to blood vessel. We recently found that *Mvwf1* lies on a founder allele common among laboratory mouse strains. To investigate the evolutionary forces operating at *B4galnt2*, we conducted a survey of DNA sequence polymorphism and microsatellite variation spanning the *B4galnt2* gene region in natural *Mus musculus domesticus* populations. Two divergent haplotypes segregate in these natural populations, one of which corresponds to the RIIIS/J sequence. Different local populations display dramatic differences in the frequency of these haplotypes, and reduced microsatellite variability near *B4galnt2* within the RIIIS/J haplotype is consistent with the recent action of natural selection. The level and pattern of DNA sequence polymorphism in the 5′ flanking region of the gene significantly deviates from the neutral expectation and suggests that variation in *B4galnt2* expression may be under balancing selection and/or arose from a recently introgressed allele that subsequently increased in frequency due to natural selection. However, coalescent simulations indicate that the heterogeneity in divergence between haplotypes is greater than expected under an introgression model. Analysis of a population where the RIIIS/J haplotype is in high frequency reveals an association between this haplotype, the *B4galnt2* tissue-specific switch, and a significant decrease in plasma VWF levels. Given these observations, we propose that low VWF levels may represent a fitness cost that is offset by a yet unknown benefit of the *B4galnt2* tissue-specific switch. Similar mechanisms may account for the variability in VWF levels and high prevalence of VWD in other mammals, including humans.

Introduction

von Willebrand factor (VWF) is a central component of hemostasis, and defects in VWF result in von Willebrand disease (VWD), the most common inherited bleeding disorder in humans (Johnsen and Ginsburg 2006). Type 1 VWD accounts for 70% of all of VWD cases and is characterized by a quantitative reduction in VWF. The RIIIS/J inbred mouse strain has a phenotype similar to Type 1 VWD in humans (Sweeney et al. 1990). The low VWF levels in RIIIS/J are due to a mutation previously termed *Mvwf1*, for "Modifier of" *VWF 1* (Mohlke et al. 1996). *Mvwf1* is a naturally occurring, *cis*-regulatory mutation that results in a unique, tissue-specific switch in the expression of a glycosyltransferase, B4GALNT2, from the intestinal epithelial pattern seen in most mice (and humans; Montiel et al. 2003) to a vascular endothelial cell-specific pattern. Endothelial cell expression of B4GALNT2 leads to aberrant posttranslational modification of VWF, resulting in accelerated clearance from circulation and a reduced plasma VWF phenotype (Mohlke et al. 1999).

Further characterization of the *B4galnt2* locus in RIIIS/J identified a ~30-kb genomic region of high sequence divergence (>2% relative to C57BL6/J) flanking *B4galnt2* exon 1 (Johnsen et al. 2008). Divergence within this region is heterogeneous, with a peak of nucleotide divergence (up to 8%) in the immediate 5′ flanking region of the gene that is several fold higher than the expected difference between inbred mouse strains (Wade et al. 2002). Transgenic experiments confirmed that the mutation causing the tissue-specific switch in *B4galnt2* gene expression lies within this region (Johnsen et al. 2008). A survey of 65 laboratory and wild-derived inbred mouse strains found a haplotype block spanning ~97 kb of nearly identical sequence to RIIIS/J in 12 other inbred strains that also exhibited the *Mvwf1* tissue-specific switch in *B4galnt2* gene expression. Five of these strains were independent, wild-derived strains of *Mus musculus domesticus* origin.

These observations raise important evolutionary questions regarding the origin and maintenance of these divergent haplotypes. However, until now, only inbred strains have been analyzed. In this study, we survey both DNA sequence polymorphism and microsatellite variability surrounding the *B4galnt2* gene region in natural populations of *M. m. domesticus* and determine the B4GALNT2 expression pattern and VWF levels in wild-caught mice. We find that the pattern of haplotypic variation previously observed in laboratory and wild-derived inbred strains is also present in natural populations, that there is evidence of a recent shift in frequency of one of these haplotypes due to natural selection, and that the level and pattern of polymorphism in the 5′ flanking region significantly deviates from the neutral expectation. Coalescent simulations performed to evaluate the role of introgression indicate that the heterogeneity in divergence between haplotypes is highly unlikely under an introgression event alone. Furthermore, we demonstrate an association between genotype at *B4galnt2*, B4GALNT2 expression pattern, and VWF levels in a natural population.

[1] Now at the Research Division, Puget Sound Blood Center and Department of Internal Medicine, University of Washington

Key words: balancing selection, introgression *cis*-regulatory variation, house mouse, *Mus musculus*, von Willebrand disease (VWD), von Willebrand factor (VWF).

E-mail: ginsburg@umich.edu.

Mol. Biol. Evol. 26(3):567–578. 2009
doi:10.1093/molbev/msn284
Advance Access publication December 16, 2008

Materials and Methods

Animal Material

Population samples of *M. m. domesticus* from France, Germany, and Cameroon are comprised of wild-caught individuals as previously described (Ihle et al. 2006; Baines and Harr 2007). The population sample of *M. m. domesticus* from Chicago was kindly provided by Bettina Harr. DNA from a single *M. famulus* individual (kindly provided by F. Bonhomme) was sequenced as an outgroup. Individuals screened for their VWF levels and *B4galnt2* expression pattern were caught in the summer of 2005 at the same sampling localities in the Massif Central of France as described in Ihle et al. (2006). Mice were acclimated to laboratory conditions for 2 months prior to analysis. Animal maintenance and handling was according to the required legal standards and institutionally approved by the Bezirksregierung Köln.

Preparation of Biological Samples

DNA was prepared from either ethanol-preserved or frozen tissues as previously described (Nichols et al. 1994). Platelet-poor plasma was collected as previously described (Nichols et al. 1994) and stored at −80 °C until analysis. Tissues were formalin-fixed and paraffin embedded by the Tissue Core of the University of Michigan Comprehensive Cancer Center.

Analysis of *B4galnt2* Expression Patterns and Plasma VWF Levels

Dolichos biflorus agglutinin (DBA) lectin staining was performed on formalin-fixed, paraffin-embedded tissues as previously described (Mohlke et al. 1999) using horseradish peroxidase conjugated DBA lectin (H-Y Labs, Inc.). Small intestine from C57BL/6J and RIIIS/J were simultaneously stained as controls.

Plasma VWF levels were determined by an enzyme linked immunosorbent assay in duplicate, as previously described (Mohlke et al. 1996) using aliquots of pooled C57BL6/J plasma to generate a standard curve. Statistical significance in VWF levels between the two groups (with or without RIIIS/J nucleotide polymorphisms at fragment 6) was determined using Student's unpaired two-tailed *t*-test.

Polymerase Chain Reaction (PCR) and DNA Sequencing

Genomic sequencing was performed at the University of Michigan, University of Cologne, and University of Munich. Briefly, nine primer pairs that span the RIIIS/J haplotype block were designed for PCR using RIIIS/J genomic sequence (GenBank EF372924). A list of the PCR primers, cycling conditions, and sequencing primers for each amplicon is provided in supplemental table S1, Supplementary Material online.

Genotyping at *B4galnt2*

A PCR fragment diagnostic of the RIIIS/J and C57BL6/J haplotypes was sequenced to screen natural populations and assign genotypes to wild-caught individuals evaluated for B4GALNT expression and VWF levels. This fragment (fragment 6) is also included in the more extensive survey conducted in the German and French populations described below (see supplemental table S1, Supplementary Material online, for primers). This fragment was chosen because it is located near the center of the previously identified haplotype block and peak of divergence in the 5′ flanking region of *B4galnt2* (fig. 1).

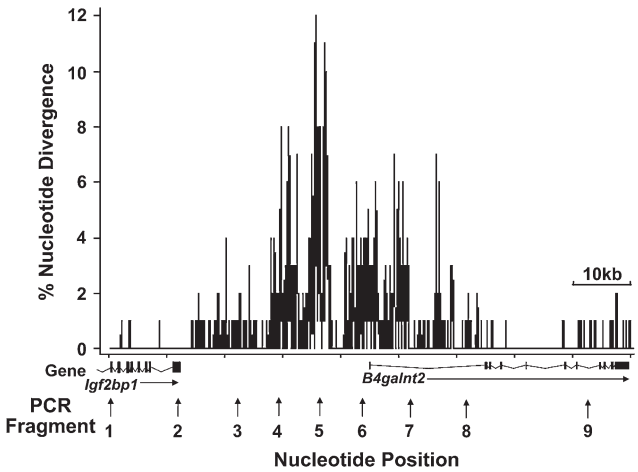

FIG. 1.—Location of fragments 1–9 relative to the region of high sequence divergence. Sliding window plot of single nucleotide divergence between C57BL6/J and RIIIS/J at the *B4galnt2* locus adapted from Johnsen et al. (2008) to show the positions of the PCR fragments (1–9) analyzed in this study (indicated by arrows in the second row). The gene structure of *B4galnt2* and the 3′ end of the upstream gene, *Igf2bp1*, are depicted in the first row under the *x*-axis.

Microsatellite Genotyping

Microsatellite typing was performed at the University of Cologne. Forward primers were labeled with 6-carboxyfluorescein or 6-carboxy-2′,4,4′,5′,7,7′-hexachlorofluorescein dye at the 5′ end. PCR was performed using a Multiplex PCR kit (Qiagen, Hilden, Germany). Dinucleotide repeat loci were chosen from either the mouse genetic mapping panel contained in the Whitehead Institute database (Dietrich et al. 1996) or identified within the mouse genome sequence (NCBI build 36). A list of primers and their genomic location is provided in supplemental table S2, Supplementary Material online.

Data Analysis

Sequence chromatograms were aligned to the homologous sequence from RIIIS/J (GenBank EF372924) and C57BL6/J (NCBI build 36) and edited using DNAstar (Madison, WI) software. For basic polymorphism statistics, heterozygous positions were randomly assigned to one of two alleles for each individual (i.e., "pseudohaplotypes" were produced), and analysis was performed using DnaSP 4 (Rozas et al. 2003). The multilocus Hudson-Kreitman-Aguasw (HKA) test was performed using the program HKA provided by J. Hey (http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm). For haplotype analysis,

we reconstructed the phase of diploid individuals using the program PHASE version 2.1 (Stephens et al. 2001; Stephens and Donnelly 2003). The algorithm was run five times automatically using the "-x" option and the output providing the best "goodness of fit" was used. Although the phase could be reconstructed with a certainty of $\geq$90% for a majority of individuals and heterozygous positions, we chose the composite genotypic disequilibrium, Δ, for analysis of linkage disequilibrium (LD), as it does not require a known phase of alleles (Weir 1996). In addition, this measure of LD does not require an assumption of random mating, which may be more appropriate for wild mouse populations where deviations from Hardy–Weinberg proportions are common (Laurie et al. 2007). The composite genotypic disequilibrium was computed using the "composite_LD" function submitted to the Bioperl project (Stajich et al. 2002), and LD is reported as the composite genotypic r^2, the squared correlation of genotypic indicators at two loci in diploid individuals (Weir et al. 2004; Laurie et al. 2007).

Microsatellite gene diversity estimates (expected heterozygosity, corrected for sample size) were calculated using the program MS Analyzer 3.15 (Dieringer and Schlotterer 2003). The ln RH statistic (Kauer et al. 2003) was applied to compare variability levels between populations. Significance was estimated by comparing the ln RH values of the typed microsatellites against a reference data set of microsatellite loci distributed throughout the genome (Ihle et al. 2006). To obtain estimates of within-haplotype class heterozygosity at microsatellites, the phase of the 12 microsatellite loci and two haplotype classes (as determined by diagnostic PCR fragment 6) was reconstructed using PHASE (Stephens et al. 2001; Stephens and Donnelly 2003) as described above. Only the genotype data for which an estimate of the phase with respect to the alternative *B4galnt2* haplotypes had a certainty $\geq$90% were included in within-haplotype heterozygosity estimates and analyses.

Allele Age Estimates

Two separate methods were used to estimate the age of the RIIIS/J haplotype, one based on the accumulation of new mutations since the common ancestor of RIIIS/J-bearing chromosomes (Rozas et al. 2001) and the other based on allele frequency (Kimura and Ota 1973). The age based on the accumulation of mutations among RIIIS/J chromosomes was estimated assuming a star phylogeny, with the number of mutations following a Poisson distribution. The presence of at most two nucleotide mutations among a sample of eight RIIIS/J chromosomes sequenced at the nucleotide level (one singleton within an RIIIS/J homozygote and one singleton of uncertain phase occurring in a heterozygote within sequence fragments 2 through 8) and two microsatellite loci monomorphic among 23 chromosomes (the minimum number for which reliable phase estimates were obtained) were used for this calculation. The nucleotide mutation rate was assumed to be 2×10^{-9} (range 1.67×10^{-9}–2.98×10^{-9}; Eyre-Walker et al. 2002), whereas microsatellite mutation rates were estimated by the method of Xu and Fu (2004) in the German population, which is less likely to be influenced by the effects of selection, and assuming an effective population size of 58,000 (Salcedo et al. 2007). The 95% confidence interval (CI), $[t_{\text{lower}} - t_{\text{upper}}]$, was estimated as follows: t_{lower} is the value of t for which the probability of two nucleotide and zero microsatellite mutations plus the probabilities of all outcomes with equally many or more mutations (i.e., two, three, four, etc., mutations at microsatellite and nucleotide loci combined) is equal to 0.025. t_{upper} is the value of t for which the probability of two nucleotide and zero microsatellite mutations plus the probabilities of all equally or more extreme outcomes (e.g., no mutations at microsatellite or nucleotide loci) is equal to 0.025. The 95% CI for the expected age based on allele frequency (Kimura and Ota 1973) was obtained using the approximation given by Slatkin and Rannala (2000). Estimates are given in units of years, assuming two generations per year.

Coalescent Simulations

Coalescent simulations were performed in order to carry out an extension of the haplotype test of Hudson et al. (1994) and to simulate the effect of an introgression event on the pattern of nucleotide variability at *B4galnt2*. The test of Hudson et al. (1994) gives the probability of observing a subset of i alleles with j or fewer segregating sites given an overall sample of n alleles with S segregating sites. To generate genealogies containing mutation events both at nucleotide and microsatellite loci, the program simcoal2 (Laval and Excoffier 2004) was used. Given the larger sample size of individuals typed for microsatellite loci (31–46) than for nucleotide polymorphism (six to eight), the number of segregating sites expected within the RIIIS/J haplotype in the total larger sample was estimated using Watterson's (1975) estimator of θ.

To evaluate an introgression model, coalescent simulations similar to those described by Evans et al. (2006) were performed using the software ms (Hudson 2002). Random genealogies for a sample of 16 chromosomes were simulated for a ~60-kb chromosomal region. The number of segregating sites was fixed based on the number expected over the entire ~60 kb given the number observed in fragments 2–8. We used the following scenario to reproduce the effects of introgression: Eight chromosomes were obtained from population A (recipient population) and eight from population B (donor population), representing the same frequency of RIIIS/J haplotypes present in the sequenced sample. Two demographic models were considered for population A, a standard constant size model assuming an effective population size of 58,000 (Salcedo et al. 2007), and the most likely demographic scenario (including a population size bottleneck and migration within and between species) estimated for the French population by Baines and Harr (2007). In both cases, population A was assumed to have split from population B at 500,000 years (1,000,000 generations) in the past. Given the low level of variation observed within RIIIS/J haplotypes, population B was subjected to a severe population bottleneck in the recent past ($N_b/N_0 = 0.000001$, $t_b = 0.000001$, time measured in $4N_e$ generations). This simulates the event that only a single haplotype introgressed into population A and

subsequently increased in frequency. For each model, simulations were performed with both zero recombination and using the rate estimated for this chromosomal region (Jensen-Seaman et al. 2004). For each simulated genealogy, seven representative regions of the same length and relative position of fragments 2–8 were taken to compare the variance in both nucleotide diversity, π, and the average pairwise divergence between haplotype classes, $\pi_{between}$, with the observed values. *P* values are given as the number of simulations out of 20,000 for which the observed variance in either π or $\pi_{between}$ were equal to or greater than the observed value.

Results

Screen of the RIIIS/J Haplotype in Natural Populations

Our previous survey of laboratory and wild-derived inbred strains (Johnsen et al. 2008) suggests that the haplotype associated with the RIIIS/J strain segregates in natural populations. This was based on extensive sequence characterization of six common laboratory strains and 23 wild-derived strains, including representatives of the *M. musculus* subspecies *musculus*, *domesticus*, *castaneus*, and *molossinus*, as well as the closely related *Mus spretus* and *Mus spicilegus*. The RIIIS/J haplotype appeared to be confined to laboratory strains and wild-derived strains of *M. m. domesticus* origin.

To screen for the distribution and frequency of the RIIIS/J haplotype in natural populations of *M. m. domesticus*, a diagnostic PCR fragment (fragment 6; see Materials and Methods and fig. 1) was typed in 126 individuals from four separate populations. There are 11 single nucleotide polymorphism in this fragment between the RIIIS/J and C57BL6/J haplotype classes, and assignment of haplotype class was confirmed by the generation of Neighbor-Joining (Saitou and Nei 1987) trees for each population (supplementary figure S1, Supplementary Material online). The frequency of the RIIIS/J haplotype was highly variable (table 1), with the two *M. m. domesticus* populations in Europe displaying the most striking difference in haplotype frequency (0% vs. 39%). No significant deviation from Hardy–Weinberg equilibrium was observed in Cameroon ($P = 0.56$) or Chicago ($P = 0.53$), whereas an excess of homozygotes was observed in France ($P = 0.015$). Previous studies of both this French population and other natural populations of *M. m. domesticus* indicate that an excess of homozygotes is a common observation at loci throughout the genome, however, likely due to inbreeding effects associated with communal nesting in wild mouse populations (Ihle et al. 2006; Laurie et al. 2007).

Table 1
C57 and RIIIS/J Haplotype Frequencies in Populations of *Mus musculus domesticus*

M. m. domesticus Population	C57 Haplotype	RIIIS/J Haplotype
France ($n = 51$)	60.8% (62/102)	39.2% (40/102)
Germany ($n = 36$)	100% (72/72)	0% (0/72)
Cameroon ($n = 29$)	77.6% (45/58)	22.4% (13/58)
Chicago ($n = 10$)	75% (15/20)	25% (5/20)

DNA Sequence Variation in European *M. m. domesticus*

For a more detailed analysis of DNA sequence variation at *B4galnt2*, we chose to focus on the *M. m. domesticus* populations from Germany and France for several reasons. Although populations from Chicago and Cameroon both display frequencies of the RIIIS/J haplotype of ~20% to 25%, these represent much more recent colonizations and are thus more likely to be confounded by demography. Additionally, the German and French populations have been extensively characterized at both the microsatellite (Ihle et al. 2006) and DNA sequence level (Baines and Harr 2007), which provide powerful points of comparison, and phenotypic screening of wild-caught mice from the same locations sampled by Ihle et al. (2006) was possible.

Nine sequence fragments spanning approximately 100 kb of the *B4galnt2* structural gene and 5′ flanking region were sequenced in six to eight individuals each in the German and French populations (fig. 1; table 2), for which data from seven unlinked autosomal reference loci also exist (Baines and Harr 2007). These fragments were chosen to encompass the region previously identified as highly divergent between the C57BL6/J and RIIIS/J strains (Johnsen et al. 2008), in addition to sequence flanking this region. An outgroup sequence from *Mus famulus* was obtainable for six of these nine segments. Figure 2 displays a plot of nucleotide diversity across the *B4galnt2* region relative to the genome average as determined from seven unlinked autosomal reference loci (Baines and Harr 2007). A clear peak of polymorphism in the immediate 5′ flanking region of the gene is present in the French population, which decreases on either side. Polymorphism levels at this peak are nearly 10-fold higher than the genome average ($\pi = 2.43\%$ vs. $\pi_{average} = 0.26\%$) for this population. In contrast, the German population displays consistently low polymorphism across the region comparable to levels seen at the reference loci.

Haplotype Structure and LD

To analyze the pattern of haplotypic variation at *B4galnt2*, the phase of the diploid sequence data was reconstructed using the program PHASE (Stephens et al. 2001; Stephens and Donnelly 2003). This could be estimated with a certainty of >90% for a majority of polymorphic sites and individuals (supplementary figure S2, Supplementary Material online). Inspection of these results reveals the presence of two major haplotype classes spanning sequence fragments 2–8, which are very similar to those present in the C57BL6/J and RIIIS/J inbred strains. Although the C57BL6/J haplotype class appears fixed in the German population, both haplotype classes segregate at intermediate frequency in the French population.

Although the haplotype reconstruction was reliable over most of the sampled region, there was less certainty at the 5′ and 3′ extremes (i.e., fragments 1 and 9). Thus, to quantify LD across the entire region, we report the composite genotypic r^2, as it does not require a known phase of alleles (see Materials and Methods). LD between sites in the

Table 2
Summary of Polymorphism and Divergence across *B4galnt2* Region

Fragment (Position[a])	Population	Sample Size	Length (bp)	S^b	π^c (%)	θ^d (%)	K^e (%)	Tajima's *D*
	France	14	749	1	0.07	0.04	—	1.434
1 (−55.3 kb)	Germany	12	736	1	0.02	0.05	—	−1.140
	France	16	427	3	0.31	0.21	4.85	1.376
2 (−38.3 kb)	Germany	12	427	1	0.04	0.08	4.63	−1.140
	France	16	353	2	0.30	0.17	2.79	1.970
3 (−29.2 kb)	Germany	12	353	2	0.19	0.19	2.70	−0.047
	France	16	638	4	0.33	0.19	2.88	1.141*
4 (−18.7 kb)	Germany	12	638	0	0	0	3.04	—
	France	16	395	18	2.43	1.37	3.04	3.069***
5 (−10.4 kb)	Germany	12	395	0	0	0	4.81	—
	France	14	476	13	1.30	0.86	4.53	2.084*
6 (−1.1 kb)	Germany	14	478	0	0	0	4.47	—
	France	14	234	3	0.68	0.40	—	2.074*
7 (+4.2 kb)	Germany	14	382	1	0.04	0.08	—	−1.155
	France	16	739	6	0.39	0.25	4.63	2.004*
8 (+20.2 kb)	Germany	14	739	2	0.06	0.09	4.56	−0.959
	France	16	668	10	0.65	0.45	—	1.623
9 (+41.4 kb)	Germany	12	668	3	0.18	0.15	—	0.672

$^*P < 0.05$; $^{***}P < 0.001$.
[a] Position relative to *B4galnt2* start codon.
[b] S = the number of segregating sites.
[c] π = nucleotide diversity (Tajima 1983).
[d] θ = nucleotide diversity (Watterson 1975).
[e] K = divergence.

French population is high, with an average r^2 over the entire sequenced region of 0.686. In particular, the association between sites spanning the ~60-kb interval between fragments 2–8 reaches the maximum ($r^2 = 1.00$; $P = 0.0047$). This value far exceeds the mean LD between sites at similar distance in a large-scale study of a *M. m. domesticus* population from Arizona ($r^2 \approx 0.15$; Laurie et al. 2007). However, due to the large number of pairwise comparisons (1,771) associated with the many divergent sites between haplotypes, no pairwise comparison remains significant after Bonferroni correction. LD in the German population, where only the C57BL6/J haplotype class is present, is comparatively lower (average $r^2 = 0.201$) and also displays no significant association between sites.

The region spanning fragments 4–7 is particularly striking. These fragments total ~2 kb and span a ~23-kb region (of which ~19 kb lies upstream of the start codon). In this region, only three polymorphic sites were identified "within" haplotype classes (one within RIIIS/J, one within C57BL6/J in the German population, and one singleton of uncertain haplotypic phase in a heterozygous individual), whereas 36 polymorphic sites are found "between" them.

Tests of Neutrality

To test if the pattern of polymorphism at *B4galnt2* is consistent with a neutral model of molecular evolution, we first performed a test based on the distribution of allele frequencies. Tajima's *D* (Tajima 1989) is consistently positive across the *B4galnt2* region in the French population, indicating an excess of intermediate frequency variants (table 2). Sequence fragments 4–7, which span the peak of polymorphism (in fig. 2), all display significantly positive values. In contrast, with the exception of the most 3′ locus (fragment 9), the German population displays consistently negative values of Tajima's *D*, indicating an excess of low-frequency variants. We next compared the ratio of polymorphism with divergence across the six fragments for which outgroup data were attainable and seven unlinked autosomal reference loci (Baines and Harr 2007) using a multilocus version of the HKA test (Hudson et al. 1987). The deviation from the neutral expectation for both polymorphism and divergence is plotted in figure 3. The neutral model could be rejected for both populations ($P < 10^{-5}$

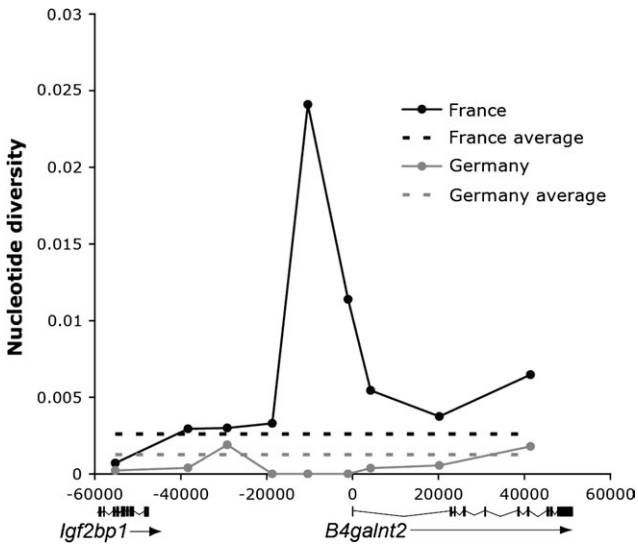

FIG. 2.—Nucleotide diversity, π (Nei 1987), for *Mus musculus domesticus* populations across the *B4galnt2* gene region. Diversity in France (black line) and Germany (gray line) are strikingly different in the region near *B4galnt2* exon 1. Dashed lines represent the genome average at seven unlinked autosomal reference loci (Baines and Harr 2007).

and $P = 0.002$ for France and Germany, respectively). However, the loci contributing most to the overall chi-square statistic differ between populations. In particular, significance in the French populations is due to fragments 5 and 6, which display both an excess of polymorphism and a deficiency of divergence. In contrast, sequence fragments spanning *B4galnt2* appear to be evolving similarly to those throughout the genome in the German population, although the test remains significant due to elevated polymorphism at a single unlinked reference locus in this population. It should be noted, however, that this reference locus displays no evidence of a deviation from neutrality by other measures (e.g., Tajima's *D*), nor does it display evidence of selection in other *M. m. domesticus* populations, including the French population (Baines and Harr 2007). Furthermore, removal of this locus does not change the significance of the results in the French population due to *B4galnt2* fragments 5 and 6.

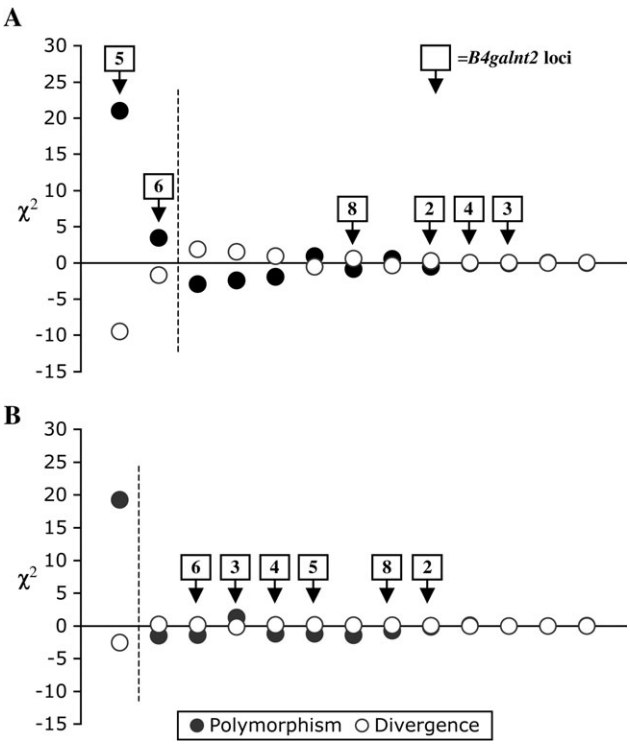

FIG. 3.—Contribution of *B4galnt2* and unlinked autosomal reference loci (Baines and Harr 2007) to the multilocus HKA statistic. *B4galnt2* loci (with sequence fragment numbers from table 2) are indicated by arrowed boxes and are shown for both the French population (*a*) and German population (*b*). The contribution to the overall test statistic is shown for both polymorphism (black circles) and divergence (open circles) for each locus. Values above the *x*-axis indicate a positive deviation from the neutral expectation, whereas those below the *x*-axis indicate a negative deviation. The loci are ordered according to their total contribution from left to right. The dotted vertical lines represent the cutoff, after which removal of the loci to the left of the line from the analysis results in no significant deviation from the neutral expectation.

Microsatellite Variability

The observation of high divergence between haplotypes, low variation within haplotypes, and large differences in haplotype frequency between populations suggest complex dynamics at *B4galnt2*. To further explore the nature of the selective forces operating, we genotyped 12 microsatellites across this region in 31–46 individuals per population (supplemental tables S2 and S3, Supplementary Material online) and compared them with the pattern seen at 118 microsatellites distributed randomly throughout the genome in the same German and French population samples (Ihle et al. 2006). Due to their high polymorphism levels and high mutation rate, these microsatellite markers serve as sensitive indicators of selection at linked sites and have the ability to capture more recent events.

The relative amount of heterozygosity between the two populations at individual loci may be described by the ln RH statistic (Kauer et al. 2003). After standardizing the values of the 12 loci using the mean and standard deviation of ln RH from the random data set, we detected no significant outlier after Bonferroni correction. However, when we analyzed heterozygosity with respect to the RIIIS/J and C57BL6/J haplotypes classes individually, a striking pattern emerged (fig. 4). At six loci within a ~160-kb window surrounding *B4galnt2*, the average heterozygosity within the RIIIS/J haplotype class is nearly half that of the C57BL6/J class (0.446 vs. 0.807), whereas the average values between the two classes at six more distal flanking loci (1–2.7 Mb from the gene) display nearly identical values (0.732 vs. 0.733). Even more striking, the two microsatellite loci located within the haplotype structure associated with the RIIIS/J and C57BL6/J classes are monomorphic (i.e., contain no variation) within the RIIIS/J class. Specifically, a $(GT)_{6–27}$ repeat located ~30 kb upstream of the *B4galnt2* start codon is polymorphic within the C57BL6/J haplotype, but harbors only a $(GT)_{23}$ allele within the RIIIS/J class, and a $(GA)_{10–37}$ locus ~1 kb downstream of the *B4galnt2* start codon is polymorphic within the C57BL6/J class, but harbors a single $(GA)_{10}$ allele within the RIIIS/J class.

The observation of a single allele at two microsatellite loci within a haplotype of intermediate frequency is unlikely under a neutral model and suggests a young age of the RIIIS/J haplotype. To assess the probability of observing a single allele at two microsatellite loci within a haplotype of given frequency, we performed an extension of the haplotype test of Hudson et al. (1994) including microsatellites (see Materials and Methods). Coalescent simulations were performed under both a standard constant size model and under the most likely demographic model for the French population (including a population size bottleneck) estimated by Baines and Harr (2007) under the conservative assumption of no recombination. In both cases, the observed level of nucleotide and microsatellite variation within the RIIIS/J haplotype significantly deviated from the neutral expectation ($P < 0.0002$ and $P = 0.0006$ for constant size and demographic models, respectively).

In addition, we compared the age of the RIIIS/J haplotype estimated from within-haplotype nucleotide and microsatellite polymorphism with that predicted from its frequency. Assuming a star phylogeny and that the mutations among RIIIS/J chromosomes follow a Poisson distribution, an estimate of the age of the RIIIS/J haplotype can

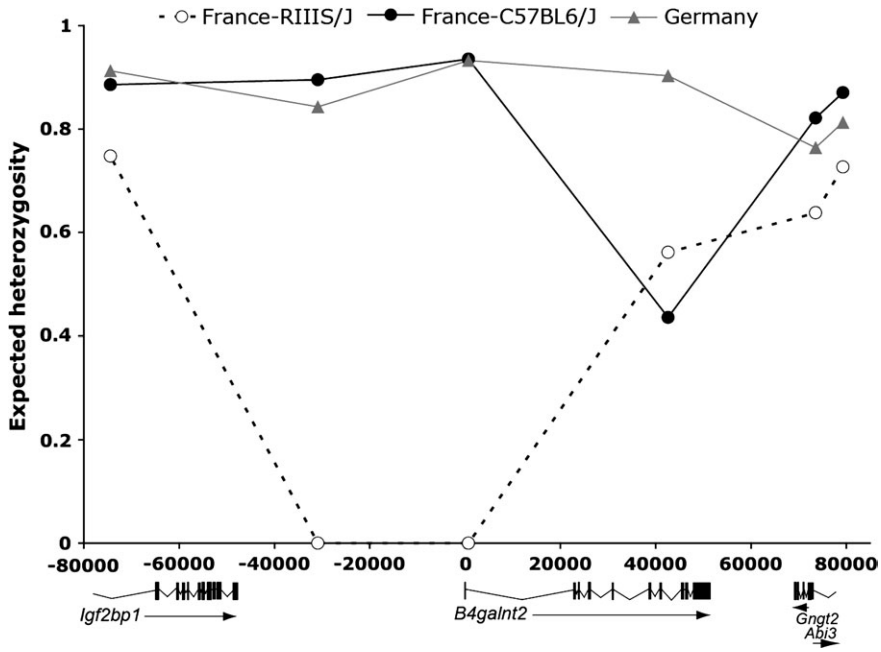

FIG. 4.—Heterozygosity at microsatellites spanning a ~160-kb region containing *B4galnt2*. Values of the German population are displayed by triangles (gray) and those of the French population with circles. The C57BL6/J haplotype class within France is displayed by a solid black line (closed circles), the RIIIS/J class with a dotted black line (open circles). The solid gray bar within the 5′ half of the *B4galnt2* gene depicts the region of haplotype structure associated with the RIIIS/J and C57BL6/J classes. The values of six other flanking loci are not shown (supplementary table S3, Supplementary Material online).

be made given the observation of at most two nucleotide mutations (one singleton within an RIIIS/J homozygote and one singleton of uncertain phase occurring in a heterozygote within fragments 2–8) and assuming that no mutations have occurred at the two monomorphic microsatellite loci, giving a maximum likelihood estimate of 97 years (95% CI = 12–350 years). In contrast, the age predicted by the frequency of the RIIIS/J haplotype is 70,040 years (95% CI = 12,650–109,246 years) (Kimura and Ota 1973). Taken together, these observations suggest the RIIIS/J haplotype has risen to intermediate frequency very recently in the French population due to the action of natural selection.

Introgression Simulations

The high sequence divergence, LD, and low variation within the RIIIS/J haplotype suggest introgression from a closely related species as a viable explanation for the origin of this allele in *M. m. domesticus*, given that genetic exchange with even more distant, noncommensal species such as *M. spretus* (average nucleotide divergence ~2%; Galtier et al. 2004) is known to occur (Greene-Till et al. 2000; Orth et al. 2002; Bonhomme et al. 2007). However, although the region of LD spans a large distance (~60 kb), as might be expected if the RIIIS/J haplotype recently introgressed, the level of sequence divergence between haplotypes is highly heterogeneous across this region. For example, the average pairwise divergence, $\pi_{between}$, between haplotypes at fragment 5 (4.6%), near the center of the region of LD, is roughly 10-fold higher than at fragments 2 and 8 (0.47% and 0.54%, respectively). In the absence of comparable heterogeneity in the mutation rate across the region, this pattern would be unexpected under a simple introgression model, and the range of divergence to the more distantly related *M. famulus* across sequence fragments (2.7–4.9%) varies less than 2-fold (table 2).

To test if the heterogeneity in divergence between haplotypes is compatible with a simple introgression model, we performed coalescent simulations allowing for the outcome of sampling alleles from two divergent populations to be evaluated (see Materials and Methods). Genealogies of a ~60-kb chromosomal region (corresponding to the region of LD associated with fragments 2–8) were simulated under a standard constant size model and a demographic model suited to the history of house mice as above (Baines and Harr 2007), both with and without recombination. The observed variances of both π and $\pi_{between}$ across the region are significantly greater than expected under all scenarios ($P < 5 \times 10^{-5}$). Thus, this heterogeneity, largely resulting from the extreme peak of divergence between haplotypes, appears highly unlikely under a simple introgression model.

Mvwf1 Phenotype in a Natural *M. m. domesticus* Population

To determine if the presence of the RIIIS/J haplotype in the 5′ flanking region of *B4galnt2* in wild mice is associated with the same tissue-specific switch in expression conferred by the *Mvwf1* mutation in inbred strains, the same diagnostic fragment (fragment 6) as above was typed in 24 wild-caught *M. m. domesticus* individuals from the Massif Central region of France and small intestine was stained with DBA lectin (which is specific for B4GALNT2 glycosylation products; Mohlke et al. 1999). Individuals

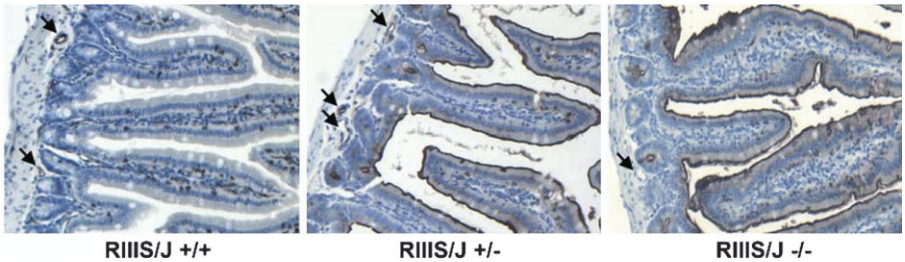

FIG. 5.—DBA lectin staining of small bowel. DBA lectin staining (brown) is specific for B4GALNT2 glycosylation (Mohlke et al. 1999) and is seen in serosal blood vessels (indicated by arrows in the left and center panels) in individuals homozygous (+/+) and heterozygous (+/−) for the RIIIS/J haplotype class, but is absent in individuals lacking RIIIS/J polymorphisms (arrow in the right panel). Intestinal epithelium is positive in animals heterozygous or homozygous for the C57BL6/J haplotype class (+/− and −/−), whereas intestinal epithelium is negative for DBA lectin staining in animals homozygous for the RIIIS/J haplotype class (+/+).

homozygous for the RIIIS/J haplotype uniformly exhibited DBA–lectin staining of blood vessels but demonstrated no intestinal epithelial DBA lectin staining (fig. 5), consistent with the tissue-specific switch in *B4galnt2* expression first characterized in the RIIIS/J inbred mouse strain. All individuals lacking the RIIIS/J haplotype demonstrated intestinal epithelial DBA lectin staining and never exhibited vascular DBA lectin staining (fig. 5). Individuals heterozygous for RIIIS/J polymorphisms exhibited both intestinal epithelial and vascular endothelial DBA lectin staining, consistent with codominant expression of *B4galnt2* alleles in these tissues (fig. 5). In addition, these data indicate that fragment 6 is a reliable diagnostic marker of (i.e., is tightly linked to) the mutation conferring the regulatory switch.

Plasma VWF levels were determined in all 24 wild-caught individuals and correlated to genotype at fragment 6. The low VWF phenotype conferred by the RIIIS/J allele has been extensively characterized in laboratory mice and is inherited in an autosomal dominant manner (i.e., mice both heterozygous and homozygous for the RIIIS/J haplotype exhibit similarly low VWF levels) (Nichols et al. 1994; Mohlke et al. 1996, 1999). Individuals either heterozygous or homozygous for the RIIIS/J haplotype displayed less than half the VWF levels of individuals that did not have the RIIIS/J haplotype (5.5 ± 2.0 vs. 13.3 ± 2.3, $P < 1 \times 10^{-7}$ using Student's unpaired two-tailed *t*-test; displayed in fig. 6), consistent with vascular endothelial expression of *B4galnt2* leading to altered glycosylation and subsequent accelerated clearance of VWF from circulation. Thus, the phenotypic effect of the RIIIS/J haplotype is confirmed in this natural population of *M. m. domesticus*, where it clearly confers the *Mvwf1* tissue-specific switch and causes a significant reduction in plasma VWF.

Discussion

The data reported here indicate that the pattern of haplotypic variation at *B4galnt2* in natural populations is similar to that observed in laboratory mouse strains (i.e., two main divergent haplotype classes similar to either C57BL6/J or RIIIS/J). The level and pattern of polymorphism and divergence in populations in which these haplotypes segregate are extreme, suggesting that the RIIIS/J haplotype may have been introduced into *M. m. domesticus* by introgression and/or is maintained by balancing selection, whereas analysis of microsatellite variation within this region suggests the action of very recent natural selection, which may account for the differences in haplotype frequency observed between different local populations. Furthermore, wild-caught individuals display a clear association between haplotype class and variation in the tissue-specific expression pattern of B4GALNT2, and individuals homo- or heterozygous for the RIIIS/J haplotype class display significant reductions in VWF level.

Origin and Maintenance of Divergent *B4galnt2* Alleles

Compared with the genomic regions between laboratory mouse strains deemed as "ancestrally divergent" (i.e.,

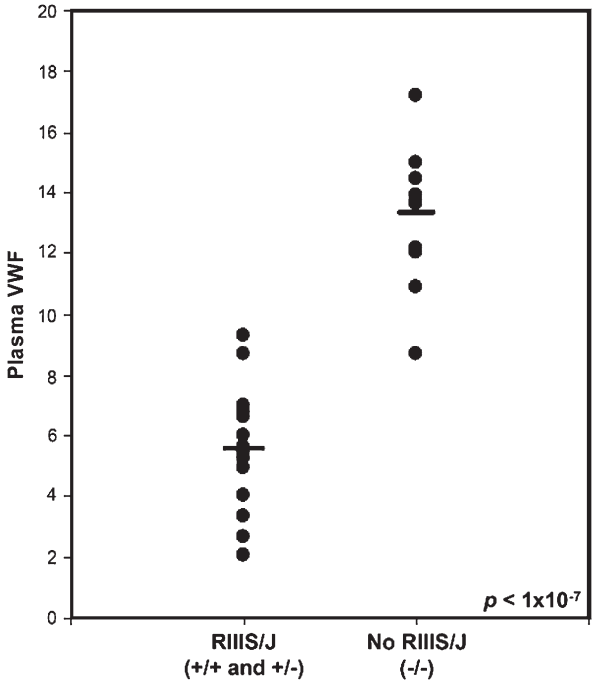

FIG. 6.—VWF levels in wild-caught mice. Plasma VWF levels (in arbitrary units) for wild-caught individual French mice are grouped by the presence or absence of the RIIIS/J haplotype class at fragment 6. The mean for each group is indicated by the solid bars. The presence of RIIIS/J-like sequence correlates with a highly significant decrease in plasma VWF levels ($P < 1 \times 10^{-7}$).

originating from different *M. musculus* subspecies), where a nucleotide difference occurs on average every ~200 bp (~0.5% divergence) (Wade et al. 2002), the level of sequence divergence between the RIIIS/J and C57BL6/J inbred strains and the haplotypes that largely correspond to these sequences in natural populations is highly unusual. The overall high sequence divergence, LD, and low variation within a haplotype at intermediate frequency suggest that selection may have operated on an allele derived from a rare interbreeding event between partially isolated populations to give rise to the pattern of variation observed at *B4galnt2*. However, the sliding window analysis of the RIIIS/J sequence (shown adapted from Johnsen et al. 2008, in fig. 1) demonstrates that this level of divergence is not uniform, displaying a peak of ~8% divergence in the immediate 5′ flanking region of the gene, which decreases on either side. Under a simple neutral introgression model, divergence between haplotypes would be largely governed by the mutation rate, and thus be expected to be more uniform.

Although introgression may be a contributing factor to the history of these two divergent allele classes at *B4galnt2*, the peak of divergence between them is also consistent with a prediction of long-term balancing selection: linked neutral mutations are expected to accumulate between allele classes with increasing proximity to the selected site(s), where the opportunity to "migrate" (i.e., to recombine) between allele classes becomes increasingly smaller (Hudson and Kaplan 1988; Charlesworth et al. 1997; Takahata and Satta 1998; Charlesworth 2006). However, simple long-term maintenance of alleles within a single population would also predict an accumulation of within-allele-class variants, which is not observed in our survey. Together with the observation of large differences in frequency between local populations, this suggests that the selective force(s) maintaining this variation may be heterogeneous over space and/or time. Stahl et al. (1999) proposed a model termed "trench warfare" to explain the heterogeneous distribution of alternative wild-type and disease-resistant alleles segregating at the *Rpm1* locus of *Arabidopsis thaliana* across its geographic range. Under this model, if a resistance allele carries a cost, it will be disfavored in the absence of the pathogen against which it confers resistance, and recurrent advances and retreats of the resistance allele/pathogen will act to maintain variation. It is possible that the German and French populations, which harbor frequencies of the RIIIS/J haplotype of 0% and 39%, represent different stages of such a fluctuating model, with the RIIIS/J haplotype having very recently increased in frequency in the French population.

Another aspect of the data inconsistent with simple long-term maintenance of alleles is the large size of the region of LD, which would not be maintained over long periods unless recombination is suppressed across this region. Although the chromosomal region containing *B4galnt2* experiences greater than average levels of recombination for the mouse genome (1.3 cM/Mb; Jensen-Seaman et al. 2004), it is possible that the many nucleotide differences between C57BL6/J and RIIIS/J, in conjunction with several larger insertions/deletions (Johnsen et al. 2008), are acting to suppress recombination between these alleles. The large region of LD would, however, be consistent with the recent increase in frequency of an allele due to natural selection.

VWF Variability in Mammals

In addition to identifying signatures of selection at the *B4galnt2* locus in these natural populations, we tested wild-caught individuals from a French house mouse population for the *Mvwf1* phenotype initially described in the RIIIS/J inbred mouse strain. RIIIS/J exhibits a low VWF phenotype similar to human type 1 VWD (Sweeney et al. 1990), the most common inherited bleeding disorder in humans. In addition to the mouse, VWD has been reported in other mammals, including pigs, calves, dogs, cats, horses, and rabbits (Mohlke and Ginsburg 1997). VWD in humans exhibits variable expressivity and incomplete penetrance, and this observation, in conjunction with the wide range of VWF levels found in normal subjects (from 50% to 150%), implies the influence of modifier genes outside the *VWF* gene itself (Levy and Ginsburg 2001). Although one major modifier, "ABO" blood group (Orstavik et al. 1985), and other minor modifiers of VWF have been described (O'Donnell et al. 2002), the cause of the majority of genetic variability in VWF levels in humans remains unexplained. The reason such wide variability in a critical coagulation factor is tolerated within human populations is unknown, and recent work has uncovered similar VWF variability in mice (Lemmerhirt et al. 2006, 2007).

In the current study, we demonstrate that the RIIIS/J haplotype is associated with a remarkably significant decrease in VWF in a wild-caught mouse population despite numerous uncontrolled acquired, environmental, and genetic factors that would be expected to influence VWF levels. Though we cannot exclude an offsetting potential benefit from low VWF levels, we propose another, as yet unknown, benefit of carrying the RIIIS/J haplotype, resulting in balancing selection at the *B4galnt2* locus, with low VWF levels representing a pleiotropic byproduct.

Possible Selective Forces Operating on Variation in *B4galnt2* Expression

The endothelium plays an active role in thrombosis, inflammation, and vascular wall remodeling (Michiels 2003) in response to injury. Consequently, a number of cell adhesion molecules and/or soluble endothelial cell-derived products, such as those critical for the host inflammatory response (Reinhart et al. 2002), are potential candidates for glycosylation modification by endothelial expressed B4GALNT2. The resulting altered function could lead to an advantageous phenotype under specific circumstances by modifying the host response to inflammation, injury, or a pathogen (Laschke et al. 2005; Doulet et al. 2006; Biswas et al. 2007). Such a scenario could favor the RIIIS/J haplotype under certain conditions, such as an infectious outbreak.

Although there is a higher frequency of individuals gaining expression of *B4galnt2* in blood vessels than losing expression in intestine in this wild mouse population, the loss of bowel *B4galnt2* expression in animals homozygous

for the RIIIS/J haplotype is also a candidate to affect fitness in these populations. Glycosylation of intestinal mucins has been shown to be critical for colonization, adhesion, and infection of enteric pathogens (Karlsson et al. 2000; Petri et al. 2002; Robbe et al. 2003; Kawakubo et al. 2004), and loss of carbohydrate ligands for an invading pathogen could result in a fitness advantage for mice homozygous for RIIIS/J-like sequences. However, expression of *B4galnt2* orthologs are conserved in intestine in mice and humans (Capon et al. 2001; Montiel et al. 2003) and likely other mammals (Chae 1997; Karlsson et al. 1997), suggesting that intestinal B4GALNT2 may serve a function that would be lost in mice homozygous for the RIIIS/J haplotype.

Conclusion

Differences in transcriptional regulation appear to account for a significant proportion of variation subject to selection. Previous application of population genetics to the study of variation in *cis*-regulatory DNA has identified several examples of balancing selection between disease and protective states (reviewed by Hahn 2007). The *Mvwf1* regulatory switch in gene expression from one tissue-restricted pattern to another is an unusual molecular mechanism with few reported precedents (Mohlke et al. 1999). The multifaceted phenotype resulting from the tissue-specific switch in *B4galnt2* expression pattern likely underlies complex selective forces at work at the *B4galnt2* locus. Whether a fitness advantage is conferred by the gain of *B4galnt2* vascular expression or the loss of *B4galnt2* intestinal expression, a specific advantageous phenotype conferred by the RIIIS/J haplotype class in natural populations may be difficult to identify. The molecular evidence at the *B4galnt2* locus in this natural mouse population is consistent with tolerance of the presumably detrimental low VWF phenotype in exchange for an as yet unknown benefit. We speculate that similar selection may account for the tolerance of variability in VWF and high prevalence of VWD in other mammalian species, including humans.

Supplementary Materials

Supplementary tables S1, S2, and S3 and figures S1 and S2 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We wish to thank Bettina Harr, John Parsch, and Laura Rose for helpful comments and discussion, Lukasz Grzeskowiak, Stephan Hutter, and Pleuni Pennings for assistance in data analysis and Tobias Heinen for assistance in mouse handling. We are especially grateful to Barbara Dod for trapping wild mice in the Massif Central. This work was supported by the Tissue Core of the University of Michigan Comprehensive Cancer Center, with special thanks to Alan Burgess. Tissue Core support was provided by The University of Michigan Cancer Center (NIH grant number CA46592). We also thank the University of Michigan DNA Sequencing Core for sequencing and genotyping services. This work was supported by an American Heart Association Fellow-to-Faculty Award #0575033N (J.J), Volkswagen Stiftung grants VW (I/824234) (P.P.), and VW (1/81 911) (D.T.), the International Graduate School in Genetics and Functional Genomics—Cologne (D.T. and J.F.B.), NIH grants PO1 HL 057346 (D.G.) and 4 R37-HL 036963 (D.G.), and the University of Munich (J.F.B.). D.G. is an investigator of the Howard Hughes Medical Institute.

Literature Cited

Baines JF, Harr B. 2007. Reduced X-linked diversity in derived populations of house mice. Genetics. 175:1911–1921.

Biswas AK, Hafiz A, Banerjee B, Kim KS, Datta K, Chitnis CE. 2007. *Plasmodium falciparum* uses gC1qR/HABP1/p32 as a receptor to bind to vascular endothelium and for platelet-mediated clumping. PLoS Pathog. 3:1271–1280.

Bonhomme F, Rivals E, Orth A, Grant GR, Jeffreys AJ, Bois PR. 2007. Species-wide distribution of highly polymorphic minisatellite markers suggests past and present genetic exchanges among house mouse subspecies. Genome Biol. 8:R80.

Capon C, Maes E, Michalski JC, Leffler H, Kim YS. 2001. Sd(a)-antigen-like structures carried on core 3 are prominent features of glycans from the mucin of normal human descending colon. Biochem J. 358:657–664.

Chae C. 1997. Lectin histochemical characteristics of the epithelial surface of ileal Peyer's patches in 3-week-old pigs. J Vet Med Sci. 59:931–934.

Charlesworth B, Nordborg M, Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 70:155–174.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2:e64.

Dieringer D, Schlotterer C. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Mol Ecol Notes. 3:167–169.

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ. 1996. A comprehensive genetic map of the mouse genome. Nature. 380:149–152.

Doulet N, Donnadieu E, Laran-Chich MP, Niedergang F, Nassif X, Couraud PO, Bourdoulous S. 2006. Neisseria meningitidis infection of human endothelial cells interferes with leukocyte transmigration by preventing the formation of endothelial docking structures. J Cell Biol. 173:627–637.

Evans PD, Mekel-Bobrov N, Vallender EJ, Hudson RR, Lahn BT. 2006. Evidence that the adaptive allele of the brain size gene microcephalin introgressed into *Homo sapiens* from an archaic Homo lineage. Proc Natl Acad Sci USA. 103:18178–18183.

Eyre-Walker A, Keightley PD, Smith NG, Gaffney D. 2002. Quantifying the slightly deleterious mutation model of molecular evolution. Mol Biol Evol. 19:2142–2149.

Galtier N, Bonhomme F, Moulia C, Belkhir K, Caminade P, Desmarais E, Duquesne JJ, Orth A, Dod B, Boursot P. 2004. Mouse biodiversity in the genomic era. Cytogenet Genome Res. 105:385–394.

Greene-Till R, Zhao Y, Hardies SC. 2000. Gene flow of unique sequences between *Mus musculus domesticus* and *Mus spretus*. Mamm Genome. 11:225–230.

Hahn MW. 2007. Detecting natural selection on *cis*-regulatory DNA. Genetica. 129:7–18.

Hudson RR. 2002. Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics. 18: 337–338.

Hudson RR, Bailey K, Skarecky D, Kwiatowski J, Ayala FJ. 1994. Evidence for positive selection in the superoxide dismutase (Sod) region of *Drosophila melanogaster*. Genetics. 136:1329–1340.

Hudson RR, Kaplan NL. 1988. The coalescent process in models with selection and recombination. Genetics. 120:831–840.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics. 116:153–159.

Ihle S, Ravaoarimanana I, Thomas M, Tautz D. 2006. An analysis of signatures of selective sweeps in natural populations of the house mouse. Mol Biol Evol. 23:790–797.

Jensen-Seaman MI, Furey TS, Payseur BA, Lu Y, Roskin KM, Chen CF, Thomas MA, Haussler D, Jacob HJ. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 14:528–538.

Johnsen JM, Ginsburg D. 2006. von Willebrand disease. In: Lichtman MA, Beutler E, Kaushansky K, Kipps TJ, Seligsohn U, Prchal J, editors. In: Williams' hematology. New York: McGraw-Hill, Inc. p. 1929–1945.

Johnsen JM, Levy GG, Westrick RJ, Tucker PK, Ginsburg D. 2008. The endothelial-specific regulatory mutation, *Mvwf1*, is a common mouse founder allele. Mamm Genome. 19:32–40.

Karlsson NG, Herrmann A, Karlsson H, Johansson ME, Carlstedt I, Hansson GC. 1997. The glycosylation of rat intestinal Muc2 mucin varies between rat strains and the small and large intestine. A study of O-linked oligosaccharides by a mass spectrometric approach. J Biol Chem. 272:27025–27034.

Karlsson NG, Olson FJ, Jovall PA, Andersch Y, Enerback L, Hansson GC. 2000. Identification of transient glycosylation alterations of sialylated mucin oligosaccharides during infection by the rat intestinal parasite *Nippostrongylus brasiliensis*. Biochem J. 350:805–814.

Kauer MO, Dieringer D, Schlotterer C. 2003. A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. Genetics. 165:1137–1148.

Kawakubo M, Ito Y, Okimura Y, Kobayashi M, Sakura K, Kasama S, Fukuda MN, Fukuda M, Katsuyama T, Nakayama J. 2004. Natural antibiotic function of a human gastric mucin against Helicobacter pylori infection. Science. 305:1003–1006.

Kimura M, Ota T. 1973. The age of a neutral mutant persisting in a finite population. Genetics. 75:199–212.

Laschke MW, Kerdudou S, Herrmann M, Menger MD. 2005. Intravital fluorescence microscopy: a novel tool for the study of the interaction of *Staphylococcus aureus* with the microvascular endothelium in vivo. J Infect Dis. 191:435–443.

Laurie CC, Nickerson DA, Anderson AD, Weir BS, Livingston RJ, Dean MD, Smith KL, Schadt EE, Nachman MW. 2007. Linkage disequilibrium in wild mice. PLoS Genet. 3:e144.

Laval G, Excoffier L. 2004. SIMCOAL 2.0: a program to simulate genomic diversity over large recombining regions in a subdivided population with a complex history. Bioinformatics. 20:2485–2487.

Lemmerhirt HL, Broman KW, Shavit JA, Ginsburg D. 2007. Genetic regulation of plasma von Willebrand factor levels: quantitative trait loci analysis in a mouse model. J Thromb Haemost. 5:329–335.

Lemmerhirt HL, Shavit JA, Levy GG, Cole SM, Long JC, Ginsburg D. 2006. Enhanced VWF biosynthesis and elevated plasma VWF due to a natural variant in the murine *Vwf* gene. Blood. 108:3061–3067.

Levy G, Ginsburg D. 2001. Getting at the variable expressivity of von Willebrand disease. Thromb Haemost. 86:144–148.

Michiels C. 2003. Endothelial cell functions. J Cell Physiol. 196:430–443.

Mohlke KL, Ginsburg D. 1997. von Willebrand disease and quantitative variation in von Willebrand factor. J Lab Clin Med. 130:252–261.

Mohlke KL, Nichols WC, Westrick RJ, Novak EK, Cooney KA, Swank RT, Ginsburg D. 1996. A novel modifier gene for plasma von Willebrand factor level maps to distal mouse chromosome 11. Proc Natl Acad Sci USA. 93:15352–15357.

Mohlke KL, Purkayastha AA, Westrick RJ, Smith PL, Petryniak B, Lowe JB, Ginsburg D. 1999. Mvwf, a dominant modifier of murine von Willebrand factor, results from altered lineage-specific expression of a glycosyltransferase. Cell. 96:111–120.

Montiel MD, Krzewinski-Recchi MA, Delannoy P, Harduin-Lepers A. 2003. Molecular cloning, gene organization and expression of the human UDP-GalNAc:Neu5Acalpha2–3Gal-beta-R beta1,4-*N*-acetylgalactosaminyltransferase responsible for the biosynthesis of the blood group Sda/Cad antigen: evidence for an unusual extended cytoplasmic domain. Biochem J. 373:369–379.

Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.

Nichols WC, Cooney KA, Mohlke KL, et al. 1994. von Willebrand disease in the RIIIS/J mouse is caused by a defect outside of the von Willebrand factor gene. Blood. 83:3225–3231.

O'Donnell J, Boulton FE, Manning RA, Laffan MA. 2002. Genotype at the secretor blood group locus is a determinant of plasma von Willebrand factor level. Br J Haematol. 116:350–356.

Orstavik KH, Magnus P, Reisner H, Berg K, Graham JB, Nance W. 1985. Factor VIII and factor IX in a twin population. Evidence for a major effect of ABO locus on factor VIII level. Am J Hum Genet. 37:89–101.

Orth A, Belkhir K, Britton-Davidian J, Boursot P, Benazzou T, Bonhomme F. 2002. Natural hybridization between 2 sympatric species of mice, *Mus musculus domesticus* L. and *Mus spretus Lataste*. C R Biol. 325:89–97.

Petri WA Jr., Haque R, Mann BJ. 2002. The bittersweet interface of parasite and host: lectin-carbohydrate interactions during human invasion by the parasite *Entamoeba histolytica*. Annu Rev Microbiol. 56:39–64.

Reinhart K, Bayer O, Brunkhorst F, Meisner M. 2002. Markers of endothelial damage in organ dysfunction and sepsis. Crit Care Med. 30:S302–S312.

Robbe C, Capon C, Maes E, Rousset M, Zweibaum A, Zanetta JP, Michalski JC. 2003. Evidence of regio-specific glycosylation in human intestinal mucins: presence of an acidic gradient along the intestinal tract. J Biol Chem. 278:46337–46348.

Rozas J, Gullaud M, Blandin G, Aguade M. 2001. DNA variation at the rp49 gene region of *Drosophila simulans*: evolutionary inferences from an unusual haplotype structure. Genetics. 158:1147–1155.

Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics. 19:2496–2497.

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

Salcedo T, Geraldes A, Nachman MW. 2007. Nucleotide variation in wild and inbred mice. Genetics. 177:2277–2291.

Slatkin M, Rannala B. 2000. Estimating allele age. Annu Rev Genomics Hum Genet. 1:225–249.
Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999. Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature. 400:667–671.
Stajich JE, Block D, Boulez K, et al. (21 co-authors). 2002. The Bioperl toolkit: perl modules for the life sciences. Genome Res. 12:1611–1618.
Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet. 73:1162–1169.
Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet. 68:978–989.
Sweeney JD, Novak EK, Reddington M, Takeuchi KH, Swank RT. 1990. The RIIIS/J inbred mouse strain as a model for von Willebrand disease. Blood. 76:2258–2265.
Tajima F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics. 105:437–460.
Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 123:585–595.
Takahata N, Satta Y. 1998. Footprints of intragenic recombination at HLA loci. Immunogenetics. 47:430–441.
Wade CM, Kulbokas EJ III, Kirby AW, Zody MC, Mullikin JC, Lander ES, Lindblad-Toh K, Daly MJ. 2002. The mosaic structure of variation in the laboratory mouse genome. Nature. 420:574–578.
Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. Theor Popul Biol. 7:256–276.
Weir BS. 1996. Genetic data analysis II. Methods for discrete population genetic data. Sunderland (MA): Sinauer Associates, Inc.
Weir BS, Hill WG, Cardon LR. 2004. Allelic association patterns for a dense SNP map. Genet Epidemiol. 27: 442–450.
Xu H, Fu YX. 2004. Estimating effective population size or mutation rate with microsatellites. Genetics. 166:555–563.

Michael Nachman, Associate Editor

Accepted November 17, 2008