Biological Journal of the Linnean Society, 2005, **84**, 523–534. With 3 figures

The genus Mus as a model for evolutionary studies
Edited by J. Britton-Davidian and J. B. Searle

The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*

BRET A. PAYSEUR and MICHAEL W. NACHMAN*

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ 85721, USA

Received 11 November 2003; accepted for publication 7 October 2004

Understanding the genetic details of reproductive isolation is a key goal in the study of speciation. Hybrid zones, geographical regions where two species meet and exchange genes, can provide insight into the genetic basis of reproductive isolation. This is especially true in species with mapped molecular markers because patterns of gene flow can be compared among different genomic regions. Even greater insight can be obtained in species with complete genome sequences because gene identity, gene number and other features of interest can be assessed for genomic regions with different patterns of introgression. Here, we review recent studies on the well-characterized hybrid zone between *Mus domesticus* and *M. musculus*, including a detailed survey of patterns of introgression for 13 markers on the X chromosome. We then compare levels of introgression for these 13 regions to a number of genomic attributes inferred from the complete sequence of the X chromosome, with two purposes. First, we identify candidate genes for reproductive isolation by finding genes that map to an X-linked region of reduced introgression and that are only expressed in the male germ line or that show high rates of protein evolution in comparison with rat. Second, we ask whether patterns of gene flow are correlated with recombination rate, gene density, base composition, CpG island density, mutation rate and the rate of protein evolution, as might be expected if many genes contribute to reproductive isolation. We identify seven candidate genes for reproductive isolation between *M. domesticus* and *M. musculus*, and our analyses reveal no general correlations between levels of introgression and other measured sequence characteristics. These results underline the utility of the house mouse as a model system for the study of speciation.

ADDITIONAL KEYWORDS: house mouse – natural selection – reproductive isolation.

INTRODUCTION

Identification of the genetic differences that cause reproductive isolation is a key goal in the study of speciation. Detailed knowledge about the genetic basis of reproductive isolation in animals comes primarily from laboratory crosses between species pairs (particularly in the fruit fly genus *Drosophila*), in which genetic markers are associated with phenotypes related to reproductive isolation to identify the genomic regions involved (Dobzhansky, 1936). Two major generalities have emerged from empirical and theoretical research on the genetics of reproductive isolation. First, reproductive isolation is typically caused by substitutions at different, interacting loci (Bateson, 1909; Dobzhansky, 1936; Muller, 1940; Orr, 1996, 1997). Second, the X chromosome has a disproportionate effect on reproductive isolation (Coyne & Orr, 1989).

In nature, geographical areas of overlap between species, where hybridization occurs (hybrid zones), can also be used to identify genomic regions underlying reproductive isolation. Under this approach, variation in the degree of gene flow between species is measured using molecular markers, with loci showing unusually low levels of admixture pointing towards

*Corresponding author. E-mail: nachman@u.arizona.edu

genomic regions of interest. Rieseberg, Whitton & Gardner (1999) successfully applied this strategy to locate 26 genomic regions contributing to reproductive isolation between two sunflower species, *Helianthus petiolaris* and *H. annuus*. This approach is especially powerful in species with complete genome sequences because genomic correlates of levels of introgression can be identified and particular genes associated with reduced introgression can be found.

In this regard, house mice offer exceptional opportunities to elucidate the genetic basis of reproductive isolation for at least four reasons. First, *Mus domesticus* and *M. musculus*, two species that diverged from each other approximately 350 000 years ago (She *et al*., 1990), form a narrow hybrid zone that stretches across Europe, from the Jutland peninsula to the Bulgarian coast of the Black Sea (Boursot *et al*., 1993; Sage, Atchley & Capanna, 1993). These taxa are referred to as species by some authors (e.g. Tucker *et al*., 1992) and as subspecies of *M. musculus* by other authors (e.g. Dod *et al*., 1993). Second, *M. domesticus* and *M. musculus* appear to be partially reproductively isolated. Hybrid males produced by crossing wild *M. musculus* with some laboratory strains (which are of primarily *M. domesticus* origin) are sterile (Forejt & Ivanyi, 1975; Forejt *et al*., 1991). Additionally, crosses between wild *M. musculus* and wild *M. domesticus* sometimes yield sterile hybrid males (J. Piálek, pers. comm.; Britton-Davidian *et al.*, 2005, this issue). Natural hybrids also harbour more parasites than do pure-species individuals (Sage *et al*., 1986; Moulia *et al*., 1993, 1995). Moreover, allele frequencies at diagnostic molecular markers change rapidly across the European hybrid zone (Hunt & Selander, 1973; Vanlerberghe *et al*., 1986; Tucker *et al*., 1992 and Božíková *et al*., 2005; Dod *et al.*, 2005; Raufaste *et al*., 2005, all this issue), indicating that the zone is probably maintained by a balance between selection against hybrids and dispersal (Barton & Hewitt, 1985). Third, the complete genome sequence of the C57BL/6J inbred strain was recently described (Mouse Genome Sequencing Consortium, 2002). Finally, a wide range of genetic techniques, including mapping, gene knock-outs and transgenic methods, are available in mice, allowing experimental tests of the significance of particular genomic regions and genes for reproductive isolation.

THE *M. DOMESTICUS*–*M. MUSCULUS* HYBRID ZONE

The European contact zone between *M. domesticus* and *M. musculus* has been intensively studied for decades and is one of the best understood hybrid zones in the world. Studies of five different transects, in Denmark, in northern Germany, in the Czech Republic, in southern Germany and in Bulgaria (Fig. 1), have offered rich

Figure 1. Map of the hybrid zone between *Mus domesticus* and *M. musculus* (modified from Sage *et al*., 1993). The solid line indicates the approximate position of the hybrid zone and the black boxes represent transects that have been studied. The species ranges of *M. domesticus* and *M. musculus* are to the west and east, respectively, of the solid line.

opportunities for collaboration among researchers interested in the biology of the house mouse. In particular, genetic surveys in the hybrid zone have uncovered several patterns that provide important insights into the genetic basis of reproductive isolation between *M. domesticus* and *M. musculus*.

First, allele frequency changes at molecular markers usually occur over short distances (relative to the large geographical ranges of both species; Boursot *et al*., 1993; Sage *et al*., 1993) and in broadly similar places for different loci, suggesting that the hybrid zone is maintained by a balance between selection against hybrids and dispersal (Barton & Hewitt, 1985). Second, allele frequency clines are asymmetrical: loci with greater degrees of introgression tend to have cline positions shifted into *M. musculus* territory (Hunt & Selander, 1973; Tucker *et al*., 1992; Dod *et al*., 1993; Sage *et al*., 1993; Munclinger *et al*., 2002). Third, there is some variation in levels of introgression for the same genomic region in different transects [e.g. *Mpi-1* shows a greater cline width in Denmark (Dod *et al*., 1993) than in Southern Germany (Tucker *et al*., 1992)]. Finally, interlocus comparisons reveal clear heterogeneity in patterns of introgression between different genomic regions in the same transect. Mitochondrial DNA markers introgress to a greater degree than autosomal loci, and sex-chromosomal markers usually display reduced gene flow (Vanlerberghe *et al*., 1986, 1988; Tucker *et al*., 1992; Dod *et al*., 1993; Sage *et al*., 1993). In agreement with mapping studies in *Drosophila* (Coyne & Orr, 1989) and hybrid zone research in other species (Hagen, 1990; Sperling & Spence, 1991; Porter *et al*., 1997), X-linked loci show reduced introgression across three transects of the hybrid zone between *M. domesticus* and *M. musculus* (Tucker *et al*., 1992; Dod *et al*., 1993; Munclinger *et al*., 2002), suggesting that the X chromosome may contribute differentially to reproductive isolation.

This idea motivated Payseur, Krenz & Nachman (2004) to conduct a targeted survey of introgression in the hybrid zone at 13 loci with known positions on the X chromosome. These markers were single nucleotide differences between *M. domesticus* and *M. musculus* that were fixed in a panel including ten individuals of each species from outside the hybrid zone. Payseur *et al.* (2004) identified a clear valley of reduced introgression, as measured by cline width, centred on a marker placed at 78.4 Mb (27.3 cM) in the sequence (Fig. 2). The lack of introgression indicates that this genomic region may contain genes underlying reproductive isolation between *M. domesticus* and *M. musculus*. X-linked loci with reduced gene flow also exhibited shifts in cline position toward the *M. domesticus* side of the hybrid zone, suggesting that the crucial incompatibilities derive from disrupted interactions between the *M. domesticus* X chromosome and other parts of the *M. musculus* genome. The chromosomal arrangement of these patterns of introgression showed evidence of linkage, with adjacent loci displaying similar trends (Fig. 2) and a relatively smooth change in cline width and cline centre at markers situated at increasing distances from the region of reduced introgression.

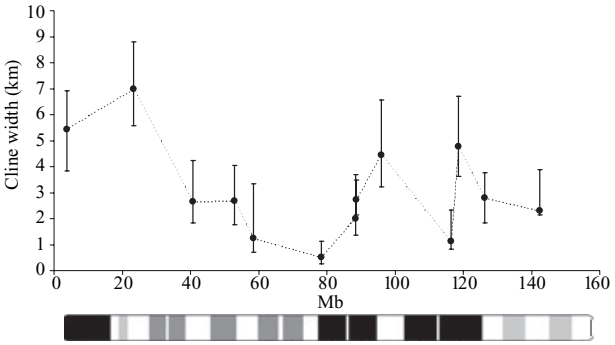

Figure 2. Scatterplot of cline width vs. sequence position for 13 markers on the X chromosome (from Payseur *et al.*, 2004). Error bars display two-unit support limits from likelihood analyses. An idiogram of the X chromosome is also shown.

The study by Payseur *et al.* (2004) constitutes the first survey of gene flow across a hybrid zone using molecular markers with known positions in a complete genome sequence. The availability of the complete genome sequence in *Mus* allows us to address several important and previously intractable questions. First, which genes are located in the X-linked region of reduced introgression and which of these show characteristics that we might expect for genes underlying reproductive isolation? Although many mouse genes have not yet been functionally characterized, certain categories of genes are predicted to play a role in reproductive isolation between *M. domesticus* and *M. musculus*. The observation that these two species may be isolated by hybrid male sterility indicates that those genes only expressed in the male germ line represent reasonable candidates. Work in *Drosophila melanogaster* suggests that proteins that function in the seminal fluid are associated with variation in male mating success via competition between different males or interactions between male proteins and the female reproductive tract, and thus may play a role in prezygotic isolation between species (Clark *et al*., 1995). Furthermore, male reproductive proteins frequently show elevated rates of divergence (Coulthart & Singh, 1988; Aguade, Miyashita & Langley, 1992; Wyckoff, Wang & Wu, 2000; Swanson, Aquadro & Vacquier, 2001; Torgerson, Kulathinal & Singh, 2002) and this rapid evolution may contribute to the formation of reproductive barriers (Swanson & Vacquier, 2002). In

addition to genes associated with male reproduction, we might also expect rapidly evolving genes in general to be differentially involved in reproductive isolation because divergence increases the likelihood of dysfunctional interactions between genes in hybrid individuals. For example, genes responsible for hybrid male inviability (*Nup96*; Presgraves *et al*., 2003), hybrid male sterility (*Odysseus*; Ting *et al*., 1998) and hybrid female sterility (*Hmr*; Barbash *et al*., 2003) in *Drosophila* show high divergence rates among closely related species.

A second question of interest is whether large-scale genomic features correlate with patterns of introgression. Such correlations might be expected for several reasons, but most require that reproductive isolation has a polygenic basis. For example, if many genes contribute to isolation and these genes are randomly distributed along the X chromosome, we might expect markers mapping to regions of low recombination to show reduced introgression simply because they are more likely to be linked to targets of selection. Recombination may also be important in the development of reproductive isolation. A recent model inspired by data from *Drosophila* and *Helianthus* suggests that chromosomal rearrangements can reduce gene flow by suppressing recombination in heterozygotes (rather than by simple underdominance; Noor *et al*., 2001; Rieseberg, 2001; Navarro & Barton, 2003a). As might be predicted by this model, humans and chimpanzees show greater rates of amino acid divergence at genes on rearranged chromosomes than at genes on chromosomes without rearrangement (Navarro & Barton, 2003b). No direct comparison between absolute rates of recombination and levels of introgression in a hybrid zone has been reported; such a comparison requires integrated genetic and physical maps for species that form hybrid zones.

If reproductive isolation is caused by many genes randomly placed throughout the genome, chromosomal regions with high gene density or high CpG island density might also be predicted to introgress less because markers in these regions will be linked to targets of selection with a higher probability. Furthermore, reproductive isolation might be expected to increase as a function of divergence in coding sequence, suggesting that markers situated near rapidly evolving genes may show reduced introgression. In general, correlations between patterns of introgression and genomic variables are not expected if only a few genes of major effect underlie reproductive isolation.

These considerations motivated us to compare patterns of introgression of X chromosomal regions across the hybrid zone between *M. domesticus* and *M. musculus* to several genomic attributes, with the goals of (i) identifying candidate genes for reproductive isolation and (ii) evaluating potential genomic correlates of introgression.

MATERIAL AND METHODS

Differential patterns of introgression

Payseur *et al.* (2004) genotyped mice (collected by R. Sage) from a 170-km transect of the hybrid zone in southern Germany and western Austria at 13 species-specific X-linked markers. Introgression was measured by fitting allele frequencies to a model derived from cline theory (Szymura & Barton, 1986; Porter *et al*., 1997). Here, we use two of these parameters to characterize patterns of gene flow: cline width, the geographical distance over which the most extreme change in allele frequency occurs; and cline centre, the geographical location where the expected allele frequency is 0.5. For more details of model-fitting, see Payseur *et al.* (2004).

Candidate genes for reproductive isolation

To find genes potentially involved in hybrid male sterility between *M. domesticus* and *M. musculus*, we identified X-linked genes with expression restricted to mouse male germ cells from the literature. Most information came from three large-scale studies, one using cDNA subtractive hybridization (Wang *et al*., 2001) and two applying microarray expression techniques (Su *et al*., 2002; Schultz, Hamra & Garbers, 2003). Sequence positions for these genes were assembled using a variety of on-line databases, including UniGene (www.ncbi.nlm.nih.gov/UniGene), LocusLink (www.ncbi.nlm.nih.gov/LocusLink), Mouse Genome Informatics (www.informatics.jax.org) and the Mouse Genome Browser (www.genome.ucsc.edu). We also analysed rates of protein evolution at 500 genes along the X chromosome (in comparison with rat coding sequences; see below). Those genes that were located in the X-linked region of reduced introgression and showed either expression restricted to the male germ line or high rates of protein evolution (or both) were selected as candidate genes for reproductive isolation between *M. domesticus* and *M. musculus*.

Genomic correlates of patterns of introgression

Genomic variables were estimated using data from on-line databases. All analyses used the most recent annotation of the mouse genome sequence (MGSCv3, update of the sequence produced by the Mouse Genome Sequencing Consortium, 2002; NCBI Build 30). All analyses, except those involving recombination rate, were completed for chromosomal windows including 500 kb, 1 Mb and 2 Mb on either side of each

molecular marker for which measures of introgression were estimated.

Recombination rates were calculated by comparing the genetic positions (from the Whitehead MIT genetic map; Dietrich *et al*., 1996) and physical positions (from the genome sequence) of X-linked microsatellites (downloaded from www.genome.ucsc.edu). Recombination rate was estimated as the slope of a linear regression of genetic position vs. physical position for microsatellites within 5 Mb on either side of each marker of interest.

Gene density was estimated in three ways, using numbers of known genes (named genes for which some functional information is available), RefSeq genes (genes for which mRNA expression has been verified) and GenScan genes (genes predicted based on DNA sequence alone). Lists of genes and their positions in the sequence of the X chromosome were downloaded from the Mouse Genome Browser at www.genome.ucsc.edu. Lists of CpG islands and their positions were also obtained at www.genome.ucsc.edu. G+C content was estimated from downloaded genomic contigs with sequence flanking the markers of interest (www.genome.ucsc.edu). Ambiguous nucleotides (Ns) were ignored in the G+C content estimation procedure.

Because coding sequences were not yet available for wild *M. domesticus* and *M. musculus*, rates of synonymous (k_s) and non-synonymous (k_a) substitution per site were estimated by comparison against putative orthologous genes from rat. All X-linked coding sequences from mouse with putative homologues in rat and all X-linked coding sequences from rat with putative homologues in mouse were downloaded from www.ensembl.org. The rat sequences were used to assemble a BLAST (Altschul *et al*., 1990) database. Each mouse sequence was blasted against the rat database and the gene with the highest bitscore was selected as a putative orthologue. All genes that returned multiple matches with high similarity were discarded as potential paralogues. Orthologous gene pairs were aligned using the GAP routine from Wisconsin Package Version 10.2 [Genetics Computer Group (GCG)] software. The parameters k_s and k_a were estimated with the Diverge routine in GCG, which uses a modified version of Li's (1993) method. This approach utilizes a two-parameter model (Kimura, 1980) to correct for differences in the rates of transition and transversion and to account for multiple hits. Rates of substitution for a given window were estimated as averages of rates for all genes included in the window.

Genomic variables were compared to cline width and cline centre visually, and using Spearman's non-parametric rank correlation tests.

RESULTS

Candidate genes for reproductive isolation

Genes that confer sterility in hybrid males may be differentially expressed in the male germ line. Using this rationale, we identified three genes located in the X-linked region of reduced introgression that are expressed only in male germ cells (Table 1). *Tktl1* produces an enzyme involved in carbohydrate transport and metabolism (transketolase activity) and *Halapx* encodes an acidic protein that is rich in alanine.

Rapidly evolving genes may also contribute disproportionately to the formation of reproductive barriers between closely related species. Two of the loci expressed solely in the male germ line, *Tktl1* and *Tex11*, exhibit higher than average k_a/k_s values in comparison with other X-linked loci (Table 1; average k_a/k_s for 500 genes = 0.274; median = 0.152). Four additional genes show relatively high rates of protein evolution ($k_a/k_s > 0.5$) and map to the X-linked region of reduced introgression (Table 2). *Pbsn* functions in odorant binding, and reference cDNA sequences are derived from libraries constructed from urinary bladder, vesicular gland and bone. *Pet2* causes plasmacytoma (when mutated) and cDNA reference sequences come from testis.

Additional functional information about these genes and the other genes in Tables 1, 2 is not currently available, but based on their locations, expression patterns and rates of evolution, they are candidate genes

Table 1. Candidate genes for reproductive isolation between *Mus domesticus* and *M. musculus* based on expression in the male germ line

Gene	Description	k_a [a]	k_s [a]	k_a/k_s [a]	Position[b]
Tktl1	Transketolase activity; carbohydrate metabolism	0.038	0.109	0.352	58307522
Halapx	Haploid specific alanine-rich protein	[c]	[c]	[c]	68169064
Tex11	[c]	0.104	0.094	1.108	85652084

[a]Divergence rates are calculated by comparison with orthologous sequences from rat.

[b]Position represents the sequence position of the translation start site.

[c]No information available. No orthologous gene sequence for *Halapx* could be located in the rat genome.

Table 2. Candidate genes for reproductive isolation between *Mus domesticus* and *M. musculus* based on high k_a/k_s values

Gene[a]	Description	k_a [b]	k_s [b]	k_a/k_s [b]	Position[c]
Mm.21705	[d]	0.106	0.109	0.972	58481953
Pbsn	Odorant-binding function	0.133	0.148	0.899	62321536
Pet2	Plasmacytoma	0.156	0.115	1.357	74093574
ENSMUSG00000050332.1	[d]	0.064	0.109	0.587	80204267

[a]Gene name is given for known genes. For predicted genes, UniGene or Ensembl identifiers are provided.
[b]Divergence rates are calculated by comparison with orthologous sequences from rat.
[c]Position represents the sequence position of the translation start site.
[d]No information available.

for reproductive isolation between *M. domesticus* and *M. musculus*.

Genomic correlates of patterns of introgression

Cline width and genomic attributes of the X-linked regions investigated by Payseur *et al.* (2004) are plotted as a function of sequence position in Figure 3, and data are provided in Table 3. A few suggestive trends can be seen in Figure 3, including an apparent spike in gene density and CpG island density corresponding to the region of low cline width. Additionally, most variables show very weak trends in predicted directions. For example, cline width is positively related to recombination rate, and negatively related to gene density, CpG island density and k_a; however, none of these relationships approaches statistical significance in correlation analyses ($P > 0.05$) with total window sizes of 1, 2 or 4 Mb. Additionally, comparisons between loci in regions of reduced introgression and all other loci yield no significant differences (Mann–Whitney U; $P > 0.05$ in all tests). Relationships between cline centre and genomic variables are also not statistically significant.

DISCUSSION

Candidate genes for reproductive isolation

Perhaps the clearest prediction about sequence characteristics and their associations with patterns of introgression is that genes with functions related to the observed phenotype of reproductive isolation should be located near markers exhibiting reduced introgression. The form and extent of reproductive isolation between natural populations of *M. domesticus* and *M. musculus* is currently not well understood. However, if these species are isolated by hybrid male sterility, we might expect markers showing reduced introgression to be linked to genes expressed in the male germ line. This prediction is further motivated by the observation that genes associated with sperm production are among the most rapidly evolving genes (at the protein level) in mice (Mouse Genome Sequencing Consortium, 2002; Torgerson *et al.*, 2002). Furthermore, sperm-related genes are over-represented on the X chromosome (Wang *et al.*, 2001) and evolve faster than their autosomal counterparts (Torgerson & Singh, 2003).

Although we uncovered no general correspondence between the number of genes expressed solely in male germ cells and levels of introgression across the X chromosome, three candidate genes for reproductive isolation, *Tktl1*, *Halapx* and *Tex11*, were identified on the basis of their expression and introgression profiles. Two of these genes also show evidence of rapid divergence in comparisons with rat.

Under the assumption that rapidly evolving genes might be differentially involved in reproductive isolation, we identified four additional genes with high k_a/k_s values that map to the region of reduced introgression. One of these genes, *Pbsn*, has odorant binding capacity. The combination of this function with a pattern of expression that includes the urinary bladder and vesicular gland is notable given the role of pheromones in mouse mating behaviour (Bronson, 1979; Ganem *et al.*, 2005, this issue). Another rapidly evolving gene, *Pet2*, is located in the region of reduced introgression and is expressed in testis, although the extent of its expression in other tissues is unclear. Our approach assumes that genes that have evolved rapidly between mouse and rat have also evolved rapidly between *M. domesticus* and *M. musculus*. The difference in divergence time in these two comparisons (mouse–rat: ~12–24 Mya, Adkins *et al.*, 2001; *M. domesticus–M. musculus*: ~350 000 years ago, She *et al.*, 1990) implies that we may have identified genes that have not actually diverged rapidly between *M. domesticus* and *M. musculus*, and we also may have failed to detect genes with accelerated rates of evolution between these two species.

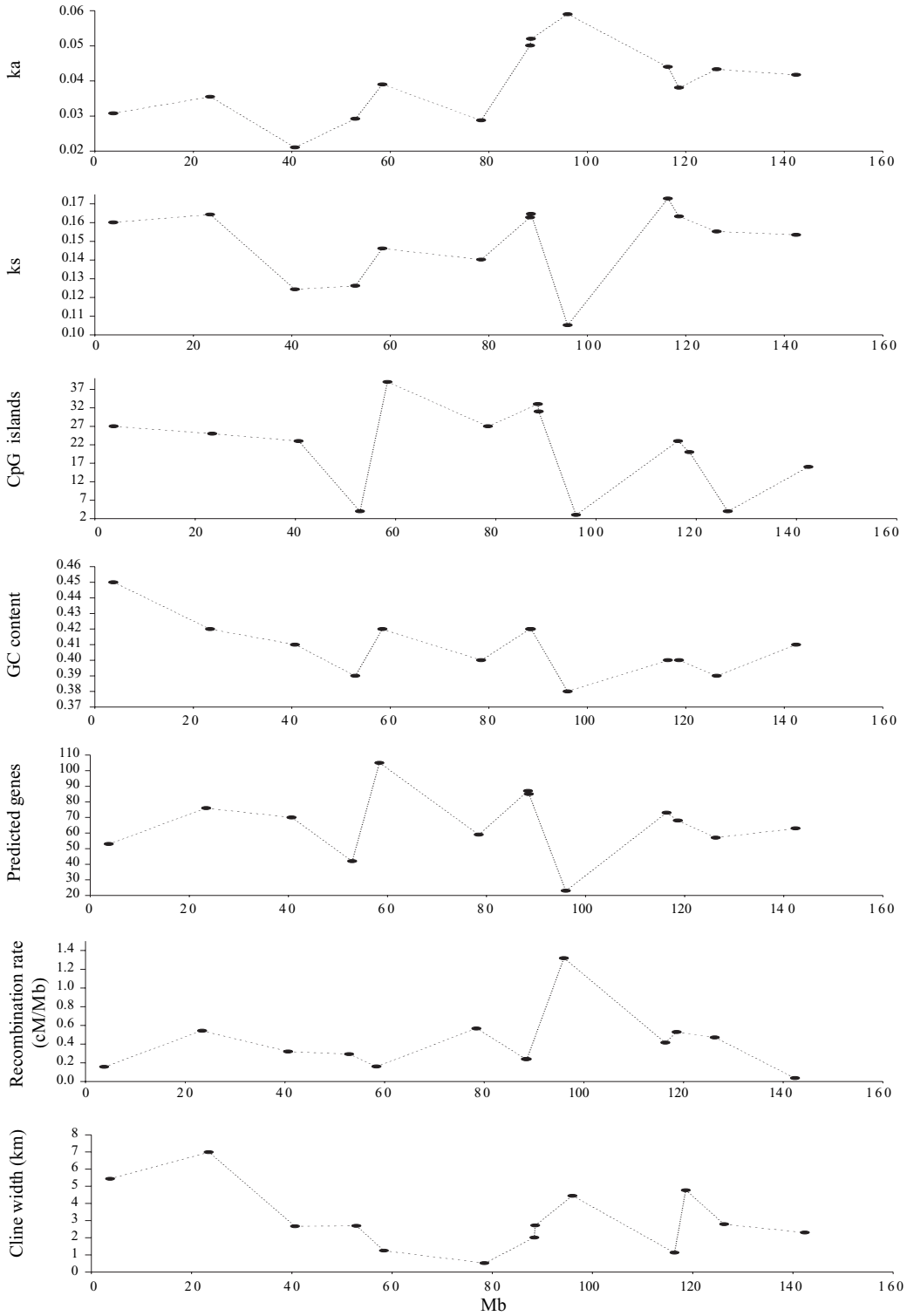

Figure 3. Cline width and genomic attributes plotted against sequence position for 13 X-linked loci. Results are shown for 10-Mb windows for recombination rate and 4-Mb windows for all other genomic characteristics.

Table 3. Genomic characteristics for 4Mb windows centred on markers for which introgression across the hybrid zone was assessed

Locus	Sequence position (bp)	Genetic position (cM)[a]	Cline width (km)[b]	Cline centre (km)[b]	Recomb. rate (cM/Mb)	GeneScan genes	RefSeq genes	Known genes	CpG islands	GC content	k_s[c]	k_a[c]	k_a/k_s[c]
Foxp3	3748097	1.1	5.43	56.27	0.16	53	46	44	27	0.45	0.155	0.026	0.192
Sepb6	23407551	15.3	6.99	55.94	0.54	76	29	24	25	0.42	0.159	0.030	0.181
Nt	40644378	20.8	2.67	54.40	0.32	70	25	12	23	0.41	0.119	0.016	0.130
Fmr1	52891104	24.0	2.69	54.57	0.29	42	6	7	4	0.39	0.121	0.024	0.174
Emd	58386749	24.0	1.25	54.15	0.16	105	69	69	39	0.42	0.141	0.034	0.211
Pola1	78445479	27.3	0.52	53.77	0.57	59	19	14	27	0.40	0.135	0.024	0.141
Xist	88396856	29.5	2.01	54.15	0.24	87	19	19	33	0.42	0.158	0.045	0.293
DXMit 18.2	88564060	29.5	2.72	54.54	0.24	85	18	18	31	0.42	0.160	0.047	0.306
Pou3f4	96002970	40.4	4.44	55.69	1.32	23	4	3	3	0.38	0.100	0.054	0.479
Btk2	116363735	43.7	1.13	53.80	0.42	73	37	36	23	0.40	0.168	0.039	0.193
Plp	118630362	45.9	4.77	56.13	0.53	68	27	29	20	0.40	0.158	0.033	0.191
Trrp5	126291138	50.3	2.79	54.68	0.47	57	11	8	4	0.39	0.150	0.038	0.294
Glra2	142398736	57.9	2.30	55.34	0.04	63	20	21	16	0.41	0.148	0.037	0.309

[a]Position on the Whitehead MIT Genetic Map (Dietrich *et al*., 1996) of the locus closest in the sequence.
[b]From Payseur *et al*. (2004). Differences between these estimates and those presented in Payseur *et al*. (2004) reflect errors in locality assignment for a few animals in that paper, subsequently corrected in an erratum.
[c]Average of values for all mouse–rat orthologous gene pairs.

In addition to identifying candidate genes for reproductive isolation, genomic data can also be used to find genes associated with positive selection in hybrid zones. Payseur *et al*. (2004) noted a pattern consistent with adaptive introgression at the *Xist* marker: although cline width was close to the average for the X chromosome, many populations on the *M. musculus* side of the hybrid zone contained *M. domesticus* alleles at this locus, suggesting that positive selection might drive the spread of alleles on to the heterospecific genomic background. We found four genes [Mm.46128 (UniGene identifier), Mm.54292, *Tsx* and *Cnbp2*] that are exclusively expressed in male germ cells within about 500 kb proximal to the *Xist* locus. Given that genes expressed in the male germ line may frequently be targeted by positive selection in mice (Mouse Genome Sequencing Consortium, 2002; Torgerson *et al*., 2002), the inspection of patterns of introgression at these genes would provide useful information about the mode and target(s) of selection in this interesting genomic region.

In combination with patterns of introgression across the hybrid zone, we used two criteria to identify the seven candidate genes for reproductive isolation listed in Tables 1, 2: expression in the male germ line and high rates of protein evolution. Although these criteria seem reasonable on biological grounds, additional types of genes may be involved in reproductive isolation between *M. domesticus* and *M. musculus*. Genes with reproductive functions in females, including loci affecting mating behaviour, may be important in isolation. Additionally, the observed increase in parasite loads in natural hybrids (Sage *et al*., 1986; Moulia *et al*., 1993, 1995) suggests that genes with functions related to immunity may contribute to reproductive barriers. Finally, many genes responsible for reproductive isolation may perform functions that are unpredictable from the nature of the isolation. For example, a gene that causes hybrid inviability between *D. melanogaster* and *D. simulans* (*Nup96*) encodes a nuclear pore protein, whose function and sequence is strongly conserved across deep phylogenetic boundaries (Presgraves *et al*., 2003). More genes underlying reproductive isolation need to be characterized before generalizations about functional biases can be drawn.

Genomic correlates of patterns of introgression

In addition to the identification of candidate genes for reproductive isolation, the study by Payseur *et al*. (2004) and the availability of the complete genome sequence of the mouse allows comparisons between patterns of introgression and general genomic attributes for the first time. Our analyses reveal no significant correlations between patterns of gene flow and measured genomic characteristics. To guide the interpretation of these results, we review reasons why patterns of introgression might be predicted to be correlated with a number of genomic attributes.

The rate of recombination may affect patterns of introgression in multiple ways. The extent of association between alleles at different loci in a hybrid zone is predicted to decay as a function of recombination rate. Rieseberg *et al*. (1999) provided empirical evidence supporting this prediction: closely linked markers show more linkage disequilibrium than unlinked markers in a sunflower hybrid zone. Signs of recombination are also visible in the results of Payseur *et al*. (2004), with neighbouring loci showing correlated cline widths.

Because recombination controls the rate at which loci approach linkage equilibrium, the power to find genes underlying reproductive isolation in hybrid zones using markers depends on the local rate of recombination. If reproductive isolation is caused by many loci randomly placed throughout the genome, markers situated in regions experiencing little recombination are more likely to be linked to these genes, and thus less likely to flow between species. Therefore, recombination rate may be positively correlated with levels of introgression if reproductive isolation has this type of genetic architecture. If the incompatibilities underlying reproductive isolation usually involve coding regions, a similar line of reasoning predicts that introgression should be negatively correlated with gene density.

Another potential genomic correlate of introgression is the number of CpG islands. In mammals, the cytosine residue in CpG dinucleotides is usually methylated, and deamination of this cytosine causes a genome-wide reduction of CpG sites across the human genome (International Human Genome Sequencing Consortium, 2001). In some genomic regions (CpG islands), CpG dinucleotides are not methylated, and may have regulatory functions. Hence, regional differences in CpG island density between *M. domesticus* and *M. musculus* could disrupt regulatory interactions in hybrids.

Rates of evolutionary divergence might also be predicted to be associated with levels of gene flow. If the rate of appearance of incompatibilities is a function of the neutral mutation rate, we might expect more problems to arise in genomic regions with higher neutral divergence rates (as measured by k_s), leading to the prediction of reduced introgression in such regions. Additionally, there are two reasons why patterns of introgression might be associated with rates of protein evolution. First, if functional divergence at genes involved in reproductive isolation is frequently driven by positive selection, a negative correlation between

rates of protein evolution and the degree of introgression might be expected. Alternatively, genes with more functional constraint at the protein level (within species) may be more easily perturbed in hybrids, suggesting that rates of protein evolution may be positively correlated with levels of gene flow.

Despite these predictions, we did not find significant correlations between recombination rate, gene density, CpG island density, k_s or k_a, and either cline width (Fig. 3) or cline centre. Our failure to identify associations between patterns of introgression and these genomic attributes may be caused by the limited number of genomic regions we have surveyed. Relationships of weak or moderate strength are probably not detectable and the effects of covariation between genomic variables cannot be accounted for with a sample size of only 13. Additionally, the scale over which relationships between introgression and genomic variables should exist is unclear, making the determination of statistically independent chromosomal regions difficult. Nevertheless, scaling-up analyses of the kind presented here to large numbers of loci from throughout the genome should provide useful information about the genetic basis of reproductive isolation between *M. domesticus* and *M. musculus*. If analyses based on larger numbers of markers also fail to find associations between levels of introgression and genome-scale characteristics, this would suggest that the genetic basis of reproductive isolation in house mice may be simple, involving a small number of loci with major effects.

ACKNOWLEDGEMENTS

We thank J. Piálek for permission to cite his unpublished work and we thank the reviewers for comments on the manuscript. We also thank Jeremy Searle and Janice Britton-Davidian for inviting us to contribute to this volume.

REFERENCES

Adkins RM, Gelke EL, Rowe D, Honeycutt RL. 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. *Molecular Biology and Evolution* **18:** 777–791.

Aguade M, Miyashita N, Langley CH. 1992. Polymorphism and divergence in the *Mst26A* male accessory gland gene region in *Drosophila*. *Genetics* **132:** 755–770.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403–410.

Barbash DA, Siino D, Tarone AM, Roote J. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. *Proceedings of the National Academy of Sciences, USA* **100:** 5302–5307.

Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annual Review of Ecology and Systematics* **16:** 113–148.

Bateson W. 1909. Heredity and variation in modern lights. In: Seward AC, ed. *Darwin and modern science*. Cambridge: Cambridge University Press, 85–101.

Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F. 1993. The evolution of house mice. *Annual Review of Ecology and Systematics* **24:** 119–152.

Božíková E, Munclinger P, Teeter KC, Tucker PK, Macholán M, Piálek J. 2005. Mitochondrial DNA in the hybrid zone between *Mus musculus musculus* and *Mus musculus domesticus*: a comparison of two transects. *Biological Journal of the Linnean Society* **84:** 363–378.

Britton-Davidian J, Fel-Clair F, Lopez J, Alibert P, Boursot P. 2005. Postzygotic isolation between the two European subspecies of the house mouse: estimates from fertility patterns in wild and laboratory-bred hybrids. *Biological Journal of the Linnean Society* **84:** 379–393.

Bronson FH. 1979. The reproductive ecology of the house mouse. *Quarterly Review of Biology* **54:** 265–299.

Clark AG, Aguade M, Prout T, Harshman LG, Langley CH. 1995. Variation in sperm displacement and its association with accessory gland protein loci in *Drosophila melanogaster*. *Genetics* **139:** 189–201.

Coulthart MB, Singh RS. 1988. High level of divergence of male-reproductive-tract proteins, between *Drosophila melanogaster* and its sibling species, *D. simulans*. *Molecular Biology and Evolution* **5:** 182–191.

Coyne JA, Orr HA. 1989. Two rules of speciation. In: Otte D, Endler J, eds. *Speciation and its consequences*. Sunderland, MA: Sinauer Associates, 180–207.

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ, Evans CA, DeAngelis MM, Levinson DM, Kruglyak L, Goodman N, Copeland NG, Jenkins NA, Hawkins TL, Stein L, Page DC, Lander ES. 1996. A comprehensive genetic map of the mouse genome. *Nature* **380:** 149–152.

Dobzhanksy T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in *Drosophila pseudoobscura* hybrids. *Genetics* **21:** 113–135.

Dod B, Jermiin LS, Boursot P, Chapman VH, Tonnes-Nielsen J, Bonhomme F. 1993. Counterselection on sex chromosomes in the *Mus musculus* European hybrid zone. *Journal of Evolutionary Biology* **6:** 529–546.

Dod B, Smadja C, Karn RC, Boursot P. 2005. Testing for selection on the androgen-binding protein in the Danish mouse hybrid zone. *Biological Journal of the Linnean Society* **84:** 447–459.

Forejt J, Ivanyi P. 1975. Genetic studies on male sterility of hybrids between laboratory and wild mice (*Mus musculus* L.). *Genetical Research* **24:** 189–206.

Forejt J, Vincek V, Klein J, Lehrach H, Loudovamickova M. 1991. Genetic mapping of the t-complex region on mouse chromosome 17 including the hybrid sterility-1 gene. *Mammalian Genome* **1:** 84–91.

Ganem G, Ginane C, Ostrowski M-F, Orth A. 2005. Assessment of mate preference in the house mouse with reference

to investigations on assortative mating. *Biological Journal of the Linnean Society* **84:** 461–471.

Hagen RH. 1990. Population struture and host use in hybridizing subspecies of *Papilio glaucus* (Lepidoptera: Papilionidae). *Evolution* **44:** 1914–1930.

Hunt WG, Selander RK. 1973. Biochemical genetics of hybridisation in European house mice. *Heredity* **31:** 11–33.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409:** 860–921.

Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16:** 111–120.

Li WH. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution* **36:** 96–99.

Moulia C, LeBrun N, Dallas J, Orth A, Renaud F. 1993. Experimental evidence of genetic determinism in high susceptibility to intestinal pinworm infection in mice: a hybrid zone model. *Parasitology* **106:** 387–393.

Moulia C, LeBrun N, Loubes C, Marin R, Renaud F. 1995. Hybrid vigor in parasites of interspecific crosses between two mice species. *Heredity* **74:** 48–52.

Mouse Genome Sequencing Consortium. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420:** 520–562.

Muller HJ. 1940. Bearing of the *Drosophila* work on systematics. In: Huxley JS, ed. *The new systematics*. Oxford: Clarendon Press, 185–268.

Munclinger P, Božíková E, Šugerková M, Piálek J, Macholán M. 2002. Genetic variation in house mice (*Mus*, Muridae, Rodentia) from the Czech and Slovak Republics. *Folia Zoologica* **51:** 81–92.

Navarro A, Barton NH. 2003a. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* **57:** 447–459.

Navarro A, Barton NH. 2003b. Chromosomal speciation and molecular divergence – Accelerated evolution in rearranged chromosomes. *Science* **300:** 321–324.

Noor MAF, Grams KL, Bertucci LA, Reiland J. 2001. Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences, USA* **98:** 12084–12088.

Orr HA. 1996. Dobzhansky, Bateson, and the genetics of speciation. *Genetics* **144:** 1331–1335.

Orr HA. 1997. Haldane's rule. *Annual Review of Ecology and Systematics* **28:** 195–218.

Payseur BA, Krenz JG, Nachman MW. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution* **58:** 2064–2078.

Porter AH, Wenger R, Geiger H, Scholl A, Shapiro AM. 1997. The *Pontia daplidice–edusa* hybrid zone in northwestern Italy. *Evolution* **51:** 1561–1573.

Presgraves DC, Balagopalan L, Abymayr SM, Orr HA. 2003. Adaptive evolution drives divergence of a hybrid inviability gene between two species of *Drosophila*. *Nature* **423:** 715–719.

Raufaste N, Orth A, Belkhir K, Senet D, Smadja C, Baird SJE, Bonhomme F, Dod B, Boursot P. 2005. Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society* **84:** 593–616.

Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends in Ecology and Evolution* **16:** 351–358.

Rieseberg LH, Whitton J, Gardner K. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152:** 713–727.

Sage RD, Atchley WR, Capanna E. 1993. House mice as models in systematic biology. *Systematic Biology* **42:** 523–561.

Sage RD, Heyneman D, Lim KC, Wilson AC. 1986. Wormy mice in a hybrid zone. *Nature* **324:** 60–63.

Schultz N, Hamra FK, Garbers DL. 2003. A multitude of genes expressed solely in meiotic or postmeiotic spermatogenic cells offers a myriad of contraceptive targets. *Proceedings of the National Academy of Sciences, USA* **100:** 12201–12206.

She JX, Bonhomme F, Boursot P, Thaler L, Catzeflis F. 1990. Molecular phylogenies in the genus *Mus*: comparative analysis of electrophoretic, scnDNA hybridization, and mtDNA RFLP data. *Biological Journal of the Linnean Society* **41:** 83–103.

Sperling FAH, Spence JR. 1991. Structure of an asymmetric hybrid zone between two water strider species (Hemiptera: Gerridae: *Limnoporus*). *Evolution* **45:** 1370–1383.

Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A, Patapoutian A, Hampton GM, Schultz PG, Hogenesch JB. 2002. Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the National Academy of Sciences, USA* **99:** 4465–4470.

Swanson WJ, Aquadro CF, Vacquier VD. 2001. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive darwinian selection of sperm lysin. *Molecular Biology and Evolution* **18:** 376–383.

Swanson WJ, Vacquier VD. 2002. The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3:** 137–144.

Szymura JM, Barton NH. 1986. Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *Bombina variegata*, near Cracow in southern Poland. *Evolution* **40:** 1141–1159.

Ting CT, Tsaur SC, Wu ML, Wu CI. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. *Science* **282:** 1501–1504.

Torgerson DG, Kulathinal RJ, Singh RS. 2002. Mammalian sperm proteins are rapidly evolving: evidence of positive selection in functionally diverse genes. *Molecular Biology and Evolution* **19:** 1973–1980.

Torgerson DG, Singh RS. 2003. Sex-linked mammalian sperm proteins evolve faster than autosomal ones. *Molecular Biology and Evolution* **20:** 1705–1709.

Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM. 1992. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. *Evolution* **46:** 1146–1163.

Vanlerberghe F, Boursot P, Catalan J, Guerasimov S, Bonhomme F, Botev B, Thaler L. 1988. Analyse génétique de la zone d'hybridation entre les deux sous-espèces de souris *M. m. domesticus* et *M. m. musculus* en Bulgarie. *Genome* **30:** 427–437.

Vanlerberghe F, Dod B, Boursot P, Bellis M, Bonhomme F. 1986. Absence of Y-chromosome introgression across the hybrid zone between *Mus musculus domesticus* and *Mus musculus musculus*. *Genetical Research* **48:** 191–197.

Wang PJ, McCarrey JR, Wang F, Page DC. 2001. An abundance of X-linked genes in spermatogonia. *Nature Genetics* **27:** 422–426.

Wyckoff GJ, Wang W, Wu CI. 2000. Rapid evolution of male reproductive genes in the descent of man. *Nature* **403:** 304–309.