

Copy Number Changes of CNV Regions in Intersubspecific Crosses of the House Mouse

Rick J. Scavetta¹ and Diethard Tautz^{*,1}

¹Max-Planck Institut für Evolutionsbiologie, Abteilung Evolutionsgenetik, Plön, Germany

*Corresponding author: E-mail: tautz@evolbio.mpg.de.

Associate editor: Manolo Gouy

Abstract

Copy number variation (CNV) contributes significantly to natural genetic variation within and between populations. However, the mutational mechanisms leading to CNV, as well as the processes that control the size of CNV regions, are so far not well understood. Here, we have analyzed a gene family that forms CNV regions on the X and the Y chromosomes in *Mus musculus*. These CNV regions show copy number differences in two subspecies, *M. musculus domesticus* and *M. musculus musculus*. Assessment of copy numbers at these loci for individuals caught in a natural hybrid zone showed copy number increases and a large variance among individuals. Crosses of natural hybrid animals among each other produced even more extreme variants with major differences in copy number in the offspring from the same parents. To assess the inheritance pattern of the loci further, we have produced F1 and backcross hybrid animals from these subspecies. We found that copy number expansions can already be traced in F1 offspring and they became stronger in the backcross individuals. Specific analysis of hybrid male offspring indicated that neither meiotic recombination nor interchromosomal exchange was required for creating these changes because the X and Y chromosomes have no homologues in males. This suggests that intrachromosomal exchanges can drive CNV and that this can occur at an elevated frequency in interspecific crosses, even within an individual. Accordingly, we find copy number mosaicism in individuals, that is, DNA from different tissues of the same individual can have different copy numbers for the loci studied. A preliminary survey of autosomal loci suggests that these can also be subject to change in hybrids. Hence, we conclude that the effects we see are not only restricted to some specific loci but may also be caused by a general induction of replication-coupled repair processes.

Key words: copy number variation, mouse, hybrids, *Mus musculus*.

Introduction

Recent analyses of structural genetic variation in humans and mice have highlighted the presence of extensive naturally occurring copy number variations (CNVs) in healthy individuals (Iafrate et al. 2004; Li et al. 2004; Sebat et al. 2004; Snijders et al. 2005; Tuzun et al. 2005; Perry et al. 2006, 2008; Redon et al. 2006; Cutler et al. 2007; Graubert et al. 2007; Kidd et al. 2008; McCarroll 2008; She et al. 2008; Watkins-Chow and Pavan 2008). CNV loci range from 1 kb to 1 Mb, encompassing gene duplications/deletions and may also influence surrounding gene expression (Stranger et al. 2007; Henrichsen et al. 2009). In laboratory mice, some recurrent CNVs can arise at rates as high as 10^{-2} to 10^{-3} per generation at some loci (Egan et al. 2007). The proximal (phenotypic) consequences of CNVs are now coming to light, such as rare and de novo CNVs associated with autism, schizophrenia, and mental retardation (de Vries et al. 2005; Jacquemont et al. 2006; Sebat et al. 2007; Walsh et al. 2008). The ultimate (evolutionary) consequences of CNVs have so far been studied for adaptive loci, such as the correlation between *AMY1* copy number and dietary starch content, or in a macroevolutionary context among primates (Perry et al. 2006, 2007, 2008; Dumas et al. 2007; Xue et al. 2008). However, despite their potentially high mutation rate and functional impact, the role of CNVs in speciation genetics has not been considered.

To study the role of CNVs in the evolutionary divergence of populations, we took advantage of two wild *Mus musculus* subspecies (*M. musculus domesticus* of Western Europe and *M. musculus musculus* of Eastern Europe) having diverged less than 1 Ma (Guenet and Bonhomme 2003). In the wild, these subspecies come into secondary contact in an approximately 20-km-wide hybrid zone, which runs from the Jutland peninsula in Denmark through Germany and onto the Black Sea. Hybrid animals suffer from reduced fitness compared with the parental subspecies as evidenced by increased parasite load and sperm head deformations associated with sterility (Sage et al. 1986; Storchová et al. 2004). However, reproductive isolation is not complete and varying amounts of genetic introgression can be observed across the hybrid zone (Teeter et al. 2008).

We focused our study on X- and Y-linked loci of the *Slx* gene family. In the reference genome (containing a *M. musculus domesticus*-derived X chromosome), *Slx* is present with 43 copies on the X chromosome in a proximal 9.2 Mb cluster interspersed with 12 copies of an uncharacterized unrelated gene, *E330016L19Rik* (hereafter called *L19*). About 21 Mb distal, a 2.4 Mb cluster contains the closely related genes *Slx-like* (16 copies) and *Xlr* (two flanking copies; see scheme in supplementary fig. S1, Supplementary Material online, for structural details). Laboratory strains with a *M. musculus musculus*-derived X chromosome

contain a large amplification in these regions (Cutler et al. 2007). *Slx* and *Slx-like* are expressed during spermatogenesis (Calenda et al. 1994; Escalier and Garchon 2005; Reynard et al. 2007; Mueller et al. 2008), and we found in previous microarray experiments that they are differentially expressed between the two mouse subspecies (Voolstra et al. 2007). The Y chromosome harbors yet another high copy number variant of the gene family, termed *Sly* (Ellis et al. 2005). *Sly* acts to globally inhibit postmeiotic sex chromosomes, including the expression of *Slx* (Cocquet et al. 2009). Preliminary observations of sex-ratio distortion in offspring of *Sly* mutants, favoring females, suggest the possible amplification of conflicting sex-ratio distorter and suppressor genes (Ellis et al. 2005; Cocquet et al. 2009).

Copy numbers for these loci were surveyed in animals from the hybrid zone, as well as in offspring of laboratory crosses, using quantitative polymerase chain reaction (qPCR) and Southern blotting. Our data showed that these CNV regions are amplified in hybrid animals and that this amplification occurs already in the F1 offspring. This suggests that these mutations are driven by somatic changes, and indeed, we find evidence for somatic mosaicism in copy number. Preliminary surveys of autosomal loci using whole-genome microarray hybridizations suggest that these effects are not restricted to the sex chromosomes.

Materials and Methods

Materials

The *M. musculus domesticus* (Germany and France) and *M. musculus musculus* (Czech Republic and Kazakhstan) mouse population samples used for the hybrid zone comparisons represent unrelated individuals caught in the wild and have previously been described (Ihle et al. 2006). The wild hybrid mice were collected from a Bavarian transect and the DNA was provided by Ruth Rottscheidt and Bettina Harr.

Crosses

The parental animals for the laboratory-generated hybrids were F1 individuals born from unrelated animals caught in the wild. The source of *M. musculus domesticus* individuals was a Western German population (Cologne/Bonn area), and the *M. musculus musculus* individuals came from a population caught close to Vienna (obtained from K. Musolf, Konrad Lorenz Institute for Ethology).

Six *M. musculus domesticus* control crosses were established (designated numbers 4, 7, 8, 11, 12, and 16; see detailed results in [supplementary tables](#), Supplementary Material online). Offspring from the first litters, consisting of 24 male and 12 female offspring, were analyzed. The mothers in families 11 and 12 were sisters and the fathers in families 8 and 12 were brothers; the same male was used for families 12 and 16.

F1 hybrids with *M. musculus musculus* paternity are represented by three families (numbers 2 [two litters], 3, and 4) of unrelated parents, resulting in eight males and seven females. *Mus musculus domesticus* paternity hybrids are

also represented by three families (designated 5, 7, and 8) of unrelated parents (two sisters were mated with one male in family 8), resulting in a group of eight males and eight females for analysis.

A backcross family was generated with a male hybrid, having *M. musculus musculus* paternity, crossed to a *M. musculus musculus* female. The hybrid father was an offspring from hybrid family 4 and the mothers were the same *M. musculus musculus* sisters used in hybrid family 8. In total, four males and six females from two litters were used in the analysis.

qPCR Assays

We designed qPCR assays to determine copy numbers in the *Slx* and the *Slx-like* (4930527E24Rik) gene regions. To allow optimal quantifications, the assays were targeted to the least polymorphic region in *Slx* exon V. This is homologous to *Slx-like* exon III and *Xlr* exon II ([supplementary fig. S1](#), Supplementary Material online) and is therefore expected to amplify copies of these genes as well. The proximal *Slx* cluster is interspersed with an unrelated gene (E330016L19Rik—called *L19*), which we analyzed in parallel. The *Sly* locus on the Y chromosome could be targeted with a diagnostic qPCR assay in exon VII. The following primer and probe combinations were used (location of the amplified regions within the genes is marked in [supplementary fig. S1](#), Supplementary Material online): *Slx*-forward 5'-ANTCAGAAAACGTAAGTTTCTCAGAGG-3', *Slx*-reverse 5'-TTGCTGTTCACCACTTAACAAATTC-3', *Slx*-probe 5'-ATGGCAGCGTTTTGC-3'; *L19*-forward 5'-GATCCAAAGCATTGCTGCATATT-3', *L19*-reverse 5'-CATCTGCCATTGAGGGATGTGAT-3', *L19*-probe 5'-ATCCCAGGAAATTC-3'; *Sly*-forward 5'-AGAGAAAATGGATGGAAACTTATGTCAAAGA-3', *Sly*-reverse 5'-CTCTCGTTCGTTTCGTTTTGCA-3', *Sly* probe 5'-CAGCAACCAGAAATT-3'. For the normalization, we used a ready-made *etd* assay (TaqMan assay Mm00558327_s1) as single-copy X-linked endogenous control. For the analysis of the autosomal locus *gemin7*, we also used a ready-made assay (TaqMan assay Mm00783268_s1). Genomic DNA samples were treated with RNase A to prevent contamination from RNA (although the tissues that we used for DNA extraction should not express these genes anyway). Most results shown are based on DNA extracted from heart tissue. Only the results for the wild *M. musculus musculus* controls used for setting up the hybrid crosses (results in [figs. 1 and 3](#)) were from ear samples because we could not sacrifice the few animals that we had at that time. However, because somatic variation was mainly a problem seen in hybrids, and because the results are qualitatively comparable with the results from the *M. musculus musculus* samples tested earlier, we do not consider this as a problem. All samples were run in triplicates, with high consistency within runs. If the standard deviation in threshold cycle (C_T) from three technical replicates was higher than 0.2, an outlier was defined and removed. Outliers were removed for less than 18% of the technical triplicates. qPCR assays for the results listed in [table 1](#) were performed using

the TaqMan Universal Master Mix, whereas those performed later (figs. 1 and 3) used the newly available Genotyping Master Mix, which allows for more efficient qPCR reactions, but means of the two data sets are not directly comparable. All assays were run on an ABI Prism 7900HT Sequence Detection System using 384-well plates and running Sequence Detection System (SDS) v2.1.1.

To validate the efficiency of the custom qPCR assays, a dilution series (50, 25, 10, 5, 2.5, 1, and 0.5 ng/ μ l starting concentrations of *M. musculus domesticus* DNA from a single individual) was conducted for each custom assay and the slope of a linear regression of the ΔC_t values, measured against the *etd* endogenous control, calculated. Slopes for *Slx* (0.0312), *L19* (0.0461), and *Sly* (0.0992) are all within the accepted range ($m < 0.1$) for 100% efficient custom TaqMan assays, as suggested by the manufacturer. In all subsequent experiments, the starting concentration of genomic DNA was 20 ng/ μ l.

Copy number was taken as the fold change over the endogenous control and calculated for each individual using $2^{-\Delta C_t \pm \text{standard deviation}}$. ΔC_t and the standard deviation were calculated as described by the manufacturer. Due to the nature of the assay, heterozygous CNVs are codominant and so we could only estimate that the fold change of a locus over *etd* represents an average of the two chromosomes for the autosomal *gemin7* locus. For the hemizygous gonosomes in males, fold change is equivalent to copy number.

In spite of these careful validations and the high consistency of triplicate results from within a run, we noticed that the reproducibility for the same animal between two runs was sometimes poor. Part of this problem was due to the use of different DNA preparations, thus reflecting the somatic variation. But even when using the same DNA preparation some time later, we noticed inconsistencies. We have listed the results for two duplicate runs (each with internal triplicates) for two different tissues each for the experiment with the laboratory-bred hybrids in [supplementary table S2](#) (Supplementary Material online). Overall, we find that about 75% of the measurements are reproducible (difference less than 50% of the lower value). This is the range that we also see now in other experiments, and we do not think that this can be significantly improved. Hence, this has to be taken as inevitable technical noise, which can only be compensated by averaging across a large number of samples and replicates. Accordingly, the values given in [figures 1 and 3](#) represent the averages and standard deviations for a total of six replicates of DNA from one tissue for 8–44 individuals per class (listed in the figures).

Southern Blotting

To obtain an *Slx* probe in the same region as the *Slx* qPCR assay, a 1.2 kb region was amplified from genomic DNA surrounding the qPCR probe on *Slx* exon V (forward primer: 5'-CAGGCCAGGCTGTGTTTATTTATG-3', reverse primer: 5'-AGGCATAGTGCCAACATTAGGTT-3'), cloned into a TOPO cloning vector (Invitrogen, Carlsbad, CA) and sequenced. A Digoxigenin-labeled single-stranded

RNA probe was generated using the T7 transcription start site as per standard protocol provided by Roche Applied Sciences (Indianapolis, IN). Hybridization was carried out overnight at 42°C in 0.5% sodium dodecyl sulfate (SDS), 50% formamide, 6 \times saline-sodium citrate buffer (SSC), 5 \times Dendhardt's reagent, and 100 μ g/ml Salmon sperm single-stranded DNA and washed the following day with progressively higher stringency washes of 2 \times SSC and 0.5% SDS, then 2 \times SSC and 0.1% SDS, and finally 0.2 \times SSC and 0.1% SDS. Detection was conducted by chemiluminescence using CDP-STAR (Roche). Quantification of Southern blot signals was done using the ImageJ application (National Center for Biotechnology Information). About 100 ng of genomic DNA was loaded into each lane, after being digested overnight with *EcoRI* (New England Biolabs, Ipswich, MA).

Array-Based Comparative Genome Hybridization Assay

Hybridizations were performed by Miltenyi Biotec (Bergisch Gladbach, Germany) on the Agilent 244K mouse array-based comparative genome hybridization (aCGH) chip (Agilent, Santa Clara, CA), which contains approximately 240,000 sixty-mer probes at an average density of 1 kb, although gene-rich regions have a higher density and complex repeat regions have a lower density. DNA Analytics 4.0 software (Agilent) was used for statistical analysis and CNV calling with the ADM-2 statistical algorithm. The algorithm was used with the default feature-level filter, expanded nonunique probes, a centralization threshold of 6.0, and a centralization bin size of 10. The default aberration filter was also used, which requires a minimum number of three probes for aberration (CNV) calling. Combining of replicates (either intra- or interarray) was not selected.

Results

Copy Number Changes in a Hybrid Zone

Using qPCR assays, we measured copy numbers for the three CNV regions represented by *Slx*, *L19*, and *Sly*. The *Slx* assay is expected to amplify *Slx*, *Slx-like*, and *Xlr* copies, that is, is a compound measure for both CNV regions on the X chromosome ([supplementary fig. S1](#), Supplementary Material online). *L19* copies are only present in the proximal X chromosomal region, and *Sly* is only present on the Y chromosome. Three groups of animals were assessed: 1) wild mice from the two subspecies, 2) wild hybrid mice from a transect in Southern Germany, and 3) the offspring of the wild hybrid animals reared in our mouse facility ([table 1](#)).

The samples for the first group consisted of wild-caught animals from two populations each of *M. musculus domesticus* (12 individuals from Germany and 11 from France) and *M. musculus musculus* (12 individuals each from the Czech Republic and from Kazakhstan). These individuals were molecularly characterized in a previous study and represent pure samples for the respective subspecies (Ihle et al. 2006). We found within- and between-group differences among these individuals. Copy numbers varied up to 3-fold

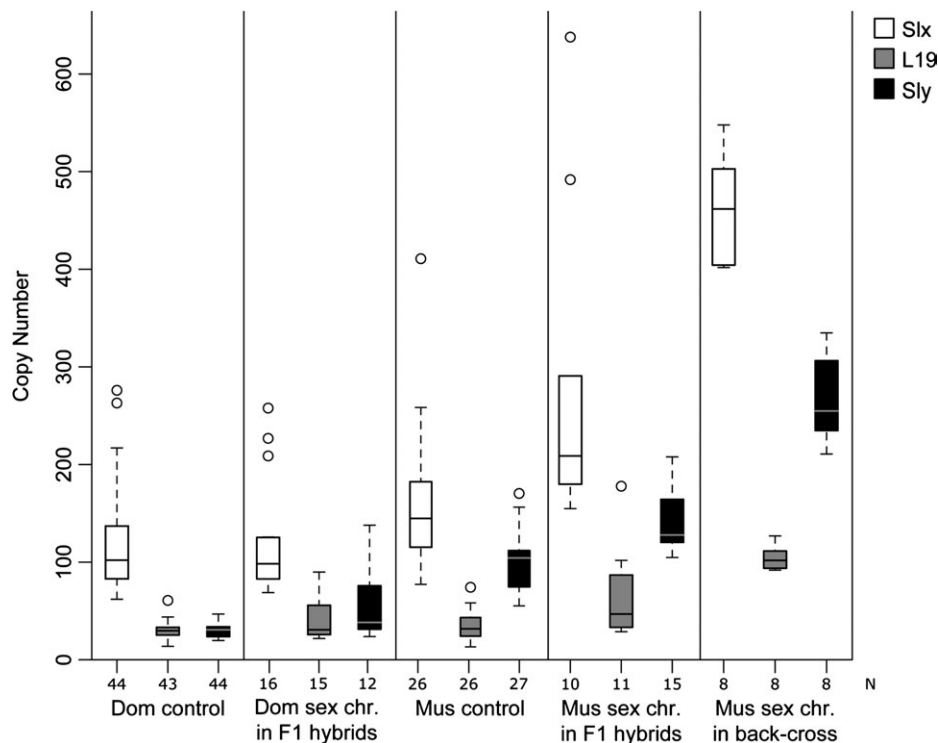


FIG. 1. CNV in pure species and hybrid crosses performed in the laboratory. Results are only shown for males, enabling us to trace the origin of the X and Y chromosomes. Note that these come from different crosses for the respective sex chromosome (see Materials and Methods for crossing scheme). Each box plot represents an offspring cohort of the respective class that is listed below the graph. The number of data points in each cohort is given as *N* below each box plot. Data points combine the results from two independent qPCR experiments with three triplicate replications each for each individual, where available (see [supplementary table S2](#), Supplementary Material online, for individual results), and boxes represent the median and the interquartile range (IQR; between the 25th and 75th percentile); outliers (greater or lesser than $1.5 \times$ the IQR) are represented as open circles. All loci, with the exception of *Slx* on the *Mus musculus domesticus*-derived X chromosome, have a significant ($P < 0.05$) increase in mean copy number over the purebreeding controls, as well as a higher variance.

between individuals of the same population ([supplementary tables S1 and S2](#), Supplementary Material online). This variation can partly be ascribed to the technical limitations of the qPCR assay, as discussed in Materials and Methods. There are no significant differences for *Slx* and *Sly* copy numbers within each subspecies, but there is a significant difference for *L19* between the two *M. musculus musculus* populations ([table 1](#)). When comparing average copy numbers between subspecies, we found highly significant ($P < 0.005$) copy number differences for *Slx* (62 vs. 126, for *M.*

musculus domesticus and *M. musculus musculus*, respectively) and *Sly* (49 vs. 203).

The second group consisted of a total of 39 mice captured at 11 sites in the vicinity of Munich, a known hybrid zone between the two subspecies ([Sage et al. 1986](#); [Teeter et al. 2008](#)). For our analysis, we distinguish Western hybrids and Eastern hybrids, which represent both sides around the center of the hybrid zone ([supplementary fig. S2](#), Supplementary Material online). Western animals consist mainly of *M. musculus domesticus* ancestry, whereas

Table 1. Mean Copy Numbers with Standard Deviations (SD—taken as averages across all individuals of the respective class; see [supplementary table S1](#), Supplementary Material online, for individual measurements) and Independent *t*-Test Comparisons in Wild Animals (see text for group assignments).

| <i>n</i> | Genotypic Classes (group) | Slx | | | L19 | | | Sly | | |
|----------|----------------------------------------|------|----|-------------------|------|-----|-------------------|------|-----|-------------------|
| | | Mean | SD | <i>P</i> | Mean | SD | <i>P</i> | Mean | SD | <i>P</i> |
| 12 | <i>Mus musculus domesticus</i> Ger (1) | 55 | 17 | | 37 | 7 | | 35 | 16 | |
| 11 | <i>M. musculus domesticus</i> Fra (1) | 62 | 13 | n.s. ^a | 34 | 8 | n.s. ^a | 40 | 14 | n.s. ^a |
| 12 | <i>M. musculus musculus</i> Cze (1) | 133 | 24 | | 50 | 10 | | 213 | 38 | |
| 12 | <i>M. musculus musculus</i> Kaz (1) | 118 | 22 | n.s. ^b | 35 | 9 | **b | 191 | 28 | n.s. ^b |
| 11 | Western hybrid zone (2) | 63 | 24 | n.s. ^a | 47 | 12 | *a | 232 | 119 | *a |
| 11 | Offspring Western hybrids (3) | 115 | 44 | **a | 65 | 30 | *a | 113 | 101 | n.s. ^a |
| 28 | Eastern hybrid zone (2) | 182 | 41 | **b | 84 | 21 | **b | 427 | 123 | **b |
| 28 | Offspring Eastern hybrids (3) | 667 | 52 | **b | 414 | 285 | **b | 1055 | 722 | **b |

NOTE.—n.s., nonsignificant (>0.05).

t-test significance in comparison with ^a*M. musculus domesticus* Ger and ^b*M. musculus musculus* Cze.

* $P < 0.05$, ** $P \leq 0.005$.

the Eastern ones consist mainly of *M. musculus musculus* ancestry (Teeter et al. 2008). This was also confirmed for the individuals used in this study by single nucleotide polymorphism (SNP) analysis (information provided by Ruth Rottscheidt and Bettina Harr). For the Western animals, we find a significantly higher copy number for *L19* and *Sly* when compared with the German *M. musculus domesticus* population (the one that is closer to the hybrid zone), whereas the copy number for *Slx* was not (table 1). For the Eastern animals, we found significantly increased copy numbers for all three assays when compared with the Czech *M. musculus musculus* population. It is notable that not only the copy number but also the variance in copy number was increased among these individuals, in particular for *Sly* (table 1).

The third group consisted of 39 offspring obtained from breeding the hybrid zone mice among themselves (breeding partners were from the same or a neighboring sampling location). For the statistical analysis, we also divided these into a Western and an Eastern offspring group (table 1). Intriguingly, we found further increases in copy numbers with even higher variances for most of the offspring of the hybrid zone animals. Breaking this down to the individual families revealed that the copy numbers measured for most of the offspring differed from the parental copy numbers at the respective loci (supplementary fig. S3, Supplementary Material online), suggesting that copy number changes can occur within a single generation. The fact that we did not observe such extreme copy number expansions in the animals caught directly in the hybrid zone (group 2 above) indicates that there may be selection against animals with extreme copy number changes under natural conditions, although a causal connection can currently not be made.

Copy Number Changes in F1 Hybrids

To study the effect of copy number increase in hybrids further, we generated F1 and some backcross hybrids in the laboratory using *M. musculus domesticus* from Germany (animals from the same German population described above) and *M. musculus musculus* derived from animals captured in Austria.

To allow inferences on the parental origin of the sex chromosomes, we focused the analysis on males only. Because males receive the X chromosome from the mother and the Y chromosome from the father, a parental origin can be unequivocally assigned. Furthermore, the results are not a composite for two chromosomes, as would be the case for the X chromosomes in females. Given that we saw no significant differences in copy numbers for males and females in the previous analysis (supplementary table S1, Supplementary Material online), we expect that restricting the analysis to males had no influence on the general outcome.

The results are presented in figure 1 as box plots because this gives the best representation of copy number distribution, including median, range, and outliers. Note that the averages for the *M. musculus domesticus* and *M. musculus*

musculus control groups are mostly slightly higher than for the experiments shown in table 1 because a more efficient PCR protocol was used for these experiments (see Materials and Methods). When comparing the *M. musculus domesticus* pure and F1 hybrids (with their respective sex chromosomes inherited from *M. musculus domesticus*) with each other, we found higher averages and variances for *L19* and *Sly* in the hybrids but not for *Slx* (left columns in fig. 1). For the comparisons of the *M. musculus musculus* pure and F1 hybrids, we found increases of averages and variances for all three loci (middle columns in fig. 1). The males of these hybrids were then backcrossed to pure *M. musculus musculus* females such that the sex chromosomes of the male offspring are both from *M. musculus musculus*. Intriguingly, we found even higher averages and variances for all three loci in these animals (right column in fig. 1). This is particularly noteworthy because the X chromosome in these males did come from a nonhybrid mother, that is, the copy number changes must have occurred during zygotic development of the animals.

Somatic Mosaicism

Based on the observations of the previous experiment, we assumed that somatic mosaicism should be traceable in the hybrids because each zygotic event of a copy number change would be propagated only in the respective cell lineage. We have therefore prepared DNA from a second organ for each of the animals from the laboratory crosses. Comparison of the DNA from heart (derivative of the mesoderm) and liver or lung (derivatives of the endoderm) revealed that copy number differences between organs of the same individual were indeed present (supplementary table S2, Supplementary Material online). Averaging across all individuals showed a stronger effect in DNA extracted from heart for most cases, suggesting a systematic effect for this tissue. However, given the inherent technical noise of qPCR (see Materials and Methods), it was difficult to ascertain this unequivocally for any given individual.

Hence, we next sought out to confirm this finding based on a completely different method, namely Southern blotting. Reliable quantification of Southern blot signals is possible when an internal control is used. In our case, a single-copy gene is not suitable as a control because the signal differences with respect to the multicopy loci in our analysis are too large. We have therefore used the signal ratio between the two *Slx* clusters as an indicator for copy number change. The proximal *Slx*-containing and the distal *Slx-like/Xlr*-containing loci on the X chromosome are separated by almost 20 Mb and can therefore be considered independent in terms of copy number changes. Thus, the signal ratio can be used as an indicator of copy number changes within these clusters. Although this approach does not provide a direct estimate for the absolute copy numbers at these loci, it can be considered as a reliable indicator for asymmetric changes in different organs of the same individual.

The *Slx* and *Slx-like* genes can be differentiated by an *EcoRI* site, producing 5.5 and 8.4 kb fragments, respectively, using a probe overlapping the qPCR target region (see

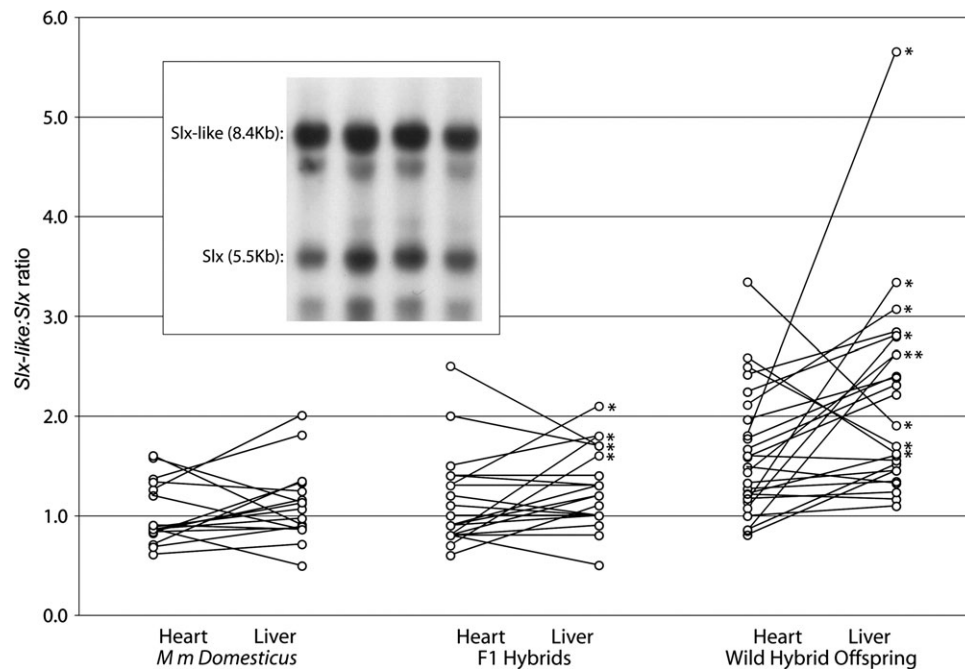


FIG. 2. Southern blot analysis for different tissues of the same individuals. A comparison of *Slx-like:Slx* hybridization intensity ratios was made between heart and liver DNA samples (or on occasion lung, another endodermal organ, when liver was not available). For pure *Mus musculus domesticus* animals, the maximum difference in intensity ratio within an individual was 0.7. Many individuals among the hybrids exceed this ratio (marked by asterisks). Calculating the difference in *Slx-like:Slx* ratios between the two organs, *M. musculus domesticus* and the F1 hybrids have comparable variance (0.2), but the wild hybrid offspring is considerably higher (1.2) even when excluding the most extreme case (0.7). Inset: Example of a Southern blot picture used for obtaining the data. Note that the minor bands seen in these blots were not considered for the analysis, although they also show signs of variation between individuals.

Materials and Methods). Therefore, in case of somatic mosaicism, we expect a difference in intensity between these two bands when comparing different tissues. For this experiment, we used the same DNA preparation from liver and heart as described above and, in addition, DNA from both tissues of the offspring of hybrid zone individuals. We indeed found a variation of signal intensity between the two loci in the two tissues (fig. 2), in particular in the offspring generated from the hybrid zone individuals. Some of these animals show an almost 2-fold difference between the tissues and a considerably higher variance in the ratio difference between tissues (0.2 vs. 1.2 for F1 hybrids and pure controls, even when excluding the most extreme case). Again, it should be emphasized that only a ratio is recorded in this experiment not absolute numbers. Thus, the stronger effects seen for the heart tissue in the qPCR experiments are not necessarily expected to show up in this experiment because animals in which both clusters are enlarged at the same time and to the same extent would not show a ratio difference. However, these experiments substantiate the notion that the hybrid individuals are effective mosaics for copy numbers at these CNV loci.

CNV Destabilization at an Autosomal Locus

Given the striking findings for all three sex chromosomal loci tested, we sought to assess the extent to which such copy number changes occur in autosomal loci as well. To obtain an initial insight into this question, we explored copy number variation on a specific autosomal locus, *gemin7*, for

which we had previous indications that it is duplicated in *M. musculus musculus*. For these experiments, we surveyed the same laboratory-generated F1 and backcross hybrids as described above, but we grouped the hybrids according to their maternity (grouping according to sex chromosome is not relevant for an autosomal locus). For the pure *M. musculus domesticus* mice, we find an average copy number of two (i.e., one per chromosome), although with some variance, which may partly be due to technical noise or segregation of chromosomes with more than one copy (fig. 3). *Mus musculus musculus* mice have an average of three copies, suggesting that the frequent occurrence of chromosomes with more than one copy. If no copy number changes had occurred in the hybrids, we would have expected an intermediate number in them, at least on average. However, we observe again higher copy numbers. The mice with *M. musculus domesticus* maternity have an average number of three, similar as pure *M. musculus musculus*, and the ones with *M. musculus musculus* maternity have an average of five copies, with a large variance (fig. 3, fourth column). The backcross hybrids, which also have an *M. musculus musculus* maternity, have an average of six copies. Again, we also observed somatic variation in these animals when we compared liver and heart DNA (supplementary table S3, Supplementary Material online).

Whole-Genome Survey

To find further candidates for autosomal loci with copy number changes in hybrids, we performed an aCGH

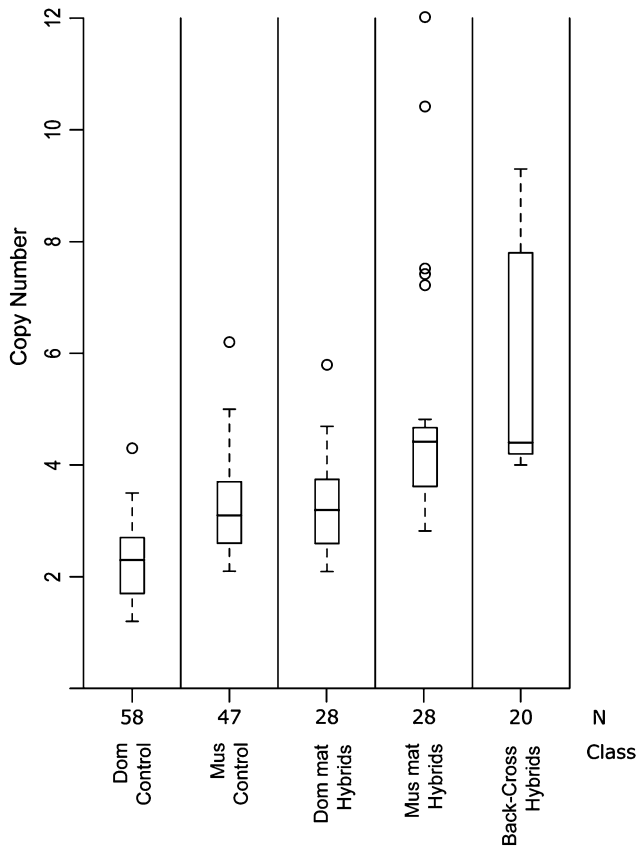


Fig. 3. Analysis of an autosomal CNV locus, *gemin7*, based on qPCR. The animals assayed are the same as those in figure 1 but females were added because the focus here is not on gonosomes. Hence, they were also regrouped—instead of the origin of the sex chromosome, we list maternal contributions for the hybrids. Copy number increases are particularly evident for the hybrids with *Mus musculus musculus* maternity and the backcross hybrids.

analysis for a subset of four animals. The Agilent 244K pre-designed mouse aCGH platform contains approximately 244,000 sixty-mer probes covering the entire genome with an average density of 1 kb, although with a reduced coverage within complex repeat regions. Genomic DNA of four unrelated males was hybridized against control C57Bl/6J genomic DNA: one purebred *M. musculus domesticus*, one purebred *M. musculus musculus*, one F1 hybrid, and one backcross hybrid were analyzed.

Although whole-genome arrays of this type are frequently used for the identification of CNV loci, considerable noise is present due to the often unknown kinetic properties of the hybridizing probes (Pozhitkov et al. 2007). This can partially be compensated for by including many samples for statistical analyses (as we did for the qPCR experiments), but this strategy was beyond the current scope of our study. Hence, because the overall analysis of the arrays resulted in too many inconsistencies, we restricted our analysis of the arrays to previously identified high-confidence CNV regions, which we could reliably trace in our data.

As a first control, we checked whether the probes in the *Slx* and *Slx-like* regions showed the expected higher log₂ ratios with respect to the reference DNA for *M. musculus*

musculus. Seven probes cover the proximal *Slx* region, five of which indeed showed a higher log₂ ratio for *M. musculus musculus*. The same probes showed an even higher ratio for the F1 and backcross hybrids (fig. 4), as expected from the results described above. Of the four probes covering the distal *Slx-like* region, we found that hybrids also represented the highest log₂ ratios for all but one probe (data not shown). The *Sly* region is not covered on the array and could therefore not be assessed. It should be noted that aCGH array results cannot be interpreted in terms of absolute copy number differences because, in addition to the unknown hybridization properties of the probes, there is also a hybridization kinetics effect to be considered and the call for a given CNV region depends on the statistical assessment of the surrounding regions. Hence, the results should only be qualitatively compared with the qPCR results. Thus, we conclude that at least for the two loci that are represented on the array, we can qualitatively confirm the results of the qPCR assays.

To focus on reliable candidates for autosomal CNV loci that might show copy number changes in the hybrids, we followed the results of a previous study that had used the same array platform to survey 42 inbred mouse strains commonly used in the laboratory (Cutler et al. 2007). Two of these mouse strains were of *M. musculus musculus* descent: CZECHII/Eij (hereafter CZE) and PWK/Ph (hereafter PWK). Because both experiments used C57Bl/6J as the reference strain, it was possible to directly compare the results. CNV calls present in both CZE and PWK (i.e., high-confidence loci) are likely to be ancestral and common within *M. musculus musculus* and should, therefore, also be observable in our *M. musculus musculus* sample.

Based on the Cutler et al. (2007) study, CZE and PWK have 26 high-confidence CNVs when compared with C57Bl/6J. Of these 26 loci, 21 were also identifiable as CNVs in the *M. musculus musculus* individual we used for our aCGH analysis. For autosomal CNVs, the hybrid and backcross log₂ ratio should be somewhere intermediate between *M. musculus domesticus* and *M. musculus musculus* if no copy number change has occurred. An example of a locus that fits this pattern is shown in figure 4 (chromosome 2). This pattern was found for most of the high-confidence CNV loci analyzed. However, one of the 21 loci displayed a different pattern in the F1 and backcross individuals. In this locus on chromosome 17, both hybrids had log₂ ratios that were as high or even higher than that for *M. musculus musculus* (fig. 4). This profile is consistent with the assumption that this is indeed an autosomal locus that is subject to copy number change in hybrids. Interestingly, this locus is not annotated as a region including segmental repeats in the current version of the mouse genome (build 37). Hence, it is likely that it constitutes a duplication for *M. musculus musculus* only.

Discussion

Our data provide evidence that the copy number of CNV loci is subject to change in interspecific crosses. The

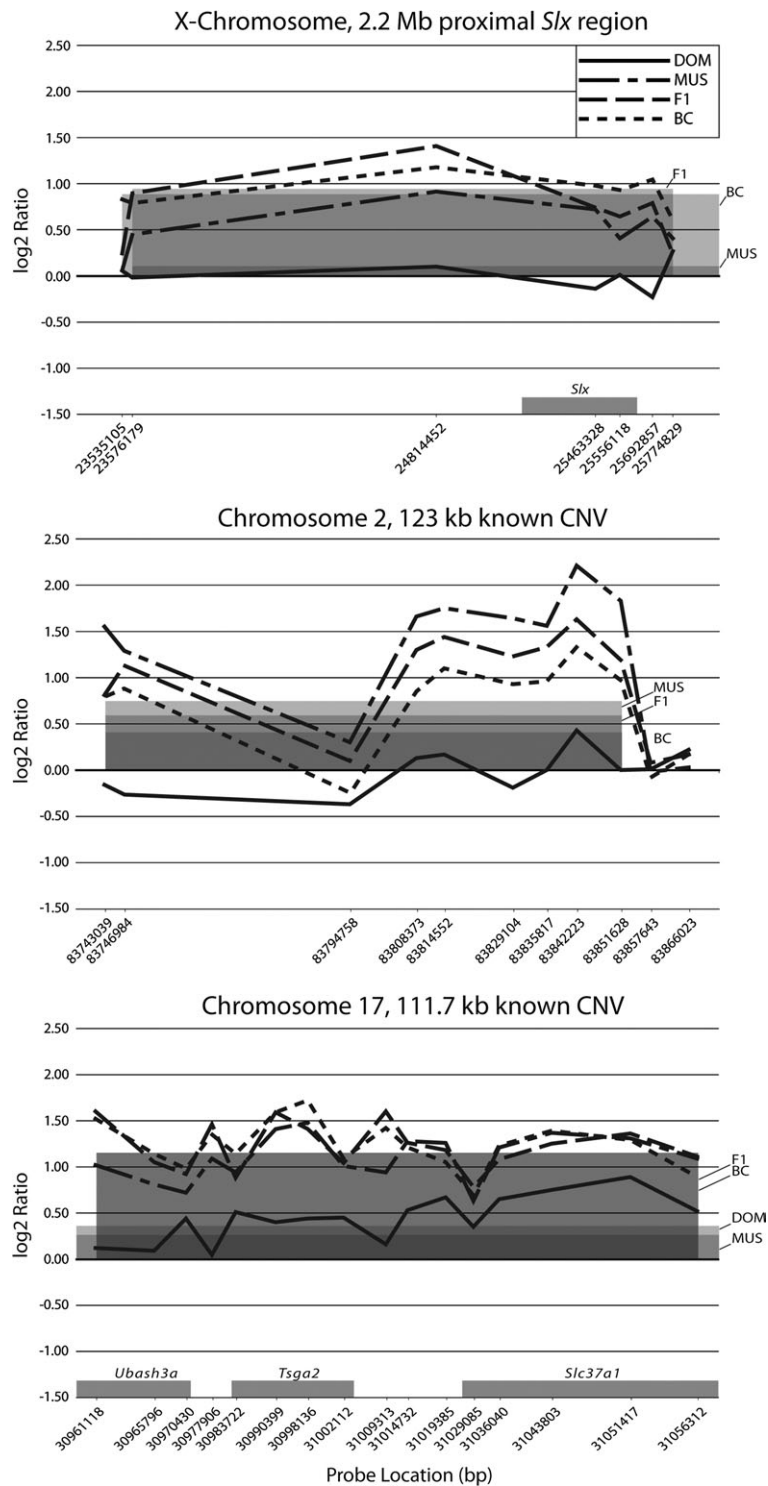


FIG. 4. CNV loci identified by aCGH analysis in four male individuals. A pure *Mus musculus domesticus* (DOM), a pure *M. musculus musculus* (MUS), a hybrid between the two species (F1), and a backcross between a hybrid and *M. musculus musculus* (BC) were hybridized against C57Bl/6j reference DNA. CNVs were identified using the ADM-2 statistical parameter (see Materials and Methods). CNV calls in each individual are indicated by semitransparent boxes. The actual log₂ ratios at each data point labeled on the x axis are depicted in the line plot. The log₂ ratio is calculated as the log₂(signal intensity of experimental DNA/signal intensity of C57Bl/6j), where a higher log₂ ratio related to higher copy number in the experimental sample. The known genes in each region are also indicated on the x axis. Top: log₂ ratios for probes in the proximal *Slx* region of the two hybrid animals (F1 and BC) are much higher than that for the *M. musculus musculus* sample. The CNV call in this region for *M. musculus musculus* is low because surrounding probes are included to form a large CNV region. Middle: an example of expected results for an autosomal CNV without evidence for changes in hybrids. *Mus musculus musculus* probes show the highest ratio, whereas both hybrids are intermediate between the two purebred samples. Bottom: an autosomal locus showing evidence for copy number change in hybrids. The large shaded block indicates overlapping calls for the F1 and the BC individual. Almost all log₂ ratios for individual probes in this region are higher in hybrids than in the purebred samples.

fact that this could already be found in F1 hybrids implies that somatic mutations play a role in these changes. Accordingly, we were able to show that DNA prepared from different organs of the same animal can show copy number differences, implying somatic mosaicism. In the following, we first discuss technical issues associated with copy number detection because none of the available techniques is currently fully reliable. We then speculate regarding potential mechanisms and implications for hybrid fitness.

Technical Considerations

We have used three different techniques to measure copy number differences: qPCR, Southern blotting, and array hybridization. The results are qualitatively comparable, but due to the nature of the techniques, they yield different quantitative results, in particular for the multicopy-gene CNVs that we have studied here.

qPCR is sensitive to polymorphisms in primer binding sites. Although we have targeted our assays to regions of the genes that are particularly highly conserved, we nonetheless noticed in cloning experiments that there are occasional SNPs in the primer binding regions (supplementary fig. S4, Supplementary Material online). Amplification efficiency is expected to be influenced by this to different degrees in different animals, making it difficult to obtain absolute quantifications. Also, we noticed that the change of buffer conditions by the supplier of the assays resulted in differences in quantification, making assays run at later times more efficient than earlier ones. This is particularly evident for the experiments shown in table 1 and figure 2, where absolute quantities for the subspecies were on average higher in the experiments performed with the new buffer conditions (fig. 2). However, because the qualitative differences remained the same (significantly higher copy numbers in *M. musculus musculus* for *Slx* and *Sly* but not for *L19*), we have no reason to assume that the qualitative copy number change effect seen in the hybrids could be a technical artifact (see Materials and Methods for a further discussion of this point). Still, given the general noisiness and dependence on the exact reaction conditions of the qPCR assays, we chose to present the results as averages across the respective groups, including standard deviations to reflect the experimental variances.

The use of Southern blots was indicated for the cases where we needed reliable quantification for the same individual for DNA from two tissues. Because the signals on Southern blots are based on hybridization with long DNA fragments, SNPs are not expected to be a problem. In contrast, there may be problems with signal linearity in these experiments, making it necessary to use a reference signal that lies in the same range as the signal to be measured. Hence, we decided to use the ratio of signal intensity between the two *Slx* clusters on the X chromosome as a measure. The disadvantage of this approach is that it does not allow an absolute quantification and is thus not directly comparable with the qPCR results. However, the approach allows a reliable comparison between the two

tested tissues of the same individual to allow the respective inferences on somatic variation.

Finally, the results from the array hybridization are also not expected to be quantitatively comparable with the results from the other two experiments because the signal strength on arrays depends on the hybridization kinetics of the repeat family, as well as on hybridization characteristics of the probes on the array. However, one should obtain at least qualitatively the same results and this was indeed the case for the two loci that were represented on the array.

Thus, given that three very different techniques confirm our qualitative results, we consider them as reliable, despite the limitations associated with each of the techniques.

Potential Mechanisms

Copy number fluctuations in multicopy-gene families are usually thought to be due to unequal recombination mechanisms. However, for the sex chromosome-linked loci studied here, we can exclude interchromosomal exchanges at least in the males because no homologue is present. Unequal sister chromatid exchange is thus an alternative mechanism. Unequal sister chromatid exchange is known to be connected to replication and aberrant DNA repair. This form of DNA repair, termed replication-coupled single-stranded break repair (Caldecott 2003), is induced when single-strand breaks are encountered during replication and develop into double-strand breaks (Kuzminov 2001). Current estimates suggest that approximately ten double-strand breaks occur per cell division at the replication fork (Haber 1999). Nonallelic (i.e., unequal) sister chromatid exchange can lead to repeat expansion and contraction under these conditions. This has been shown in cell culture models (Read et al. 2004) and is thought to be the major driver of concerted evolution mechanisms in ribosomal genes (Schlötterer and Tautz 1994).

This raises the question why these repair and recombination mechanisms should be more active in the context of a genome of hybrid origin. One possibility is that the increased heterozygosity in hybrids induces repair processes, an explanation that has been proposed for microsatellite length changes (Amos et al. 1996, 2008). It is, however, also possible that incompatibilities between epistatically interacting proteins within hybrids may be responsible. Proteins diverge over time and this is thought to be a major factor in driving species divergence in allopatry (Coyne and Orr 2004). Several such incompatibility loci have been described in *Drosophila* (Orr et al. 2004) and mouse (Storchová et al. 2004; Good et al. 2008; Mihola et al. 2009). Proteins that are part of larger complexes should be particularly sensitive to such incompatibilities. The recombination and repair machinery discussed above forms such complexes (West 2003), and it is therefore possible that even a slight divergence of the proteins could lead to a partial malfunctioning that would lead to frequent occurrences of strand breakage and aberrant repair. Arlt et al. (2009) revealed that putting cells under artificial replication stress indeed leads to copy number variation via nonhomologous end joining.

For the X chromosomal loci in females, as well as the autosomal loci, we cannot exclude the possibility that interchromosomal exchanges play an additional role. However, comparisons of the copy number values in males and females from the hybrid zone animals does not give an indication of large differences between the sexes, suggesting that the mechanisms for change may not be so much different. This question will need to be addressed in future studies.

In our microarray-based whole-genome survey of autosomal CNV loci, we found only one clear case out of 21 loci that could be reliably analyzed as having a hybrid effect. This would suggest that only a fraction of loci display changes in hybrid animals. However, these results must be considered as preliminary, given the limited scope of this study and the conservative evaluation of its results. For example, we considered only examples where a coherent effect was seen in both hybrid animals (F1 and backcross). But given the variability between animals, we would have missed individual effects in this way. Hence, much larger studies will be required before a general conclusion can be drawn on the frequency of the effects. Interestingly, the qPCR assay on one autosomal locus (*gemin7*) did yield evidence for copy number increases in hybrids. Also, given that this locus and the one discovered in the aCGH study are not multicopy loci, in contrast to those studied on the sex chromosomes, it would seem that any CNV locus can potentially become subject to copy number change in hybrids.

We note that the effects on copy number change for *Slx* and *Sly* are much more pronounced in the chromosomes derived from *M. musculus musculus* and in animals with *M. musculus musculus* maternity (table 1 and fig. 2). On the one hand, this could be ascribed to the fact that 2–4 times more copies are present in this subspecies compared with *M. musculus domesticus* (table 1), which may be more subject to destabilization and expansion. On the other hand, the stronger *M. musculus musculus* maternity effect is also seen in the *gemin7* results (fig. 3), indicating that the direction of the cross may indeed play a role, possibly through imprinted loci. Again, elucidation of this effect will require more experiments in the future.

The fact that we mainly find an expansion of the loci is particularly noteworthy. Turner et al. (2008) found an excess of deletions caused by meiotic events and postulate that other mechanisms would have to be biased toward expansion to lead to a relative stability of loci. It seems that the intrachromosomal events that we see in our survey provide such a possible expansion mechanism. In contrast, we have currently no direct evidence that the instability effect would also occur in the mitotically dividing primordial germ cells and would be transmitted to the next generation. Copy numbers in the offspring of the crosses of hybrid zone animals had no direct relation to the parental copy numbers (supplementary fig. S3, Supplementary Material online) and provide, therefore, no clue toward inheritance versus somatic effects. However, the fact that copy numbers are relatively stable within and between populations of the same subspecies suggests that there is a balance

between insertions and deletions keeping the average size under control.

The fact that we can trace somatic mosaicism for our loci is not unprecedented. There are reports on CNV at CNV loci in embryonic stem cells of the mouse (Liang et al. 2008), among monozygotic human twins (Bruder et al. 2008), and in somatic mosaicism within humans (Piotrowski et al. 2008). These observations confirm that somatic mutations can play a role in CNV variability, even when it is not induced through hybridization.

Conclusions

There have been long-standing speculations about the incompatibility of genomes in hybrid situations. Already in her Nobel lecture on transposable elements, Barbara McClintock discussed a possible role of “genome stress” in the formation of new species from hybrids (McClintock 1984). Thus, the effects we observe in CNV regions are well in line with these previous observations. However, CNVs may be of particular relevance because they code for functional RNAs and proteins, and major copy number changes occurring at multiple such loci could have many pleiotropic effects. The fact that we find a much higher variance of copy numbers among the offspring of hybrid zone animals born in the laboratory than among their parents (table 1) suggests that there may be selection against such individuals under natural conditions. Hence, our results suggest that reduced fitness in hybrid zones could partially be due to uncontrolled CNV at multiple loci. This would add to the well-documented hybrid sterility effects in hybrid zones and might therefore contribute to reproductive isolation between subspecies and species.

Supplementary Material

Tables S1–S3 and figures S1–S4 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

We thank Ruth Rottschmidt and Bettina Harr for providing DNA samples from individuals of the hybrid zone and their offspring, Kerstin Musolf for providing *M. musculus musculus* individuals and tissue samples, Christine Pfeifle for help with breeding, Barbara Kleinhenz for expert technical assistance, Bettina Harr and John Baines for comments on the manuscript, and the members of the laboratory for critical discussions. This work was funded by the SFB 680 (University of Cologne) and the Max-Planck Society. R.J.S. was initially supported by a fellowship of the International Graduate School in Genetics and Functional Genomics in Cologne.

References

- Amos W, Flint J, Xu X. 2008. Heterozygosity increases microsatellite mutation rate, linking it to demographic history. *BMC Genet.* 9:72.
- Amos W, Sawcer SJ, Feakes R, Rubinsztein DC. 1996. Microsatellites show mutational bias and heterozygote instability. *Nat Genet.* 13:390–391.

- Arlt MF, Mülle JG, Schaibley VM, Ragland RL, Durkin SG, Warren ST, Glover TW. 2009. Replication stress induces genome-wide copy number changes in human cells that resemble polymorphic and pathogenic variants. *Am J Hum Genet.* 84:1–12.
- Bruder CE, Piotrowski A, Gijsbers AA, et al. (22 co-authors). 2008. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *Am J Hum Genet.* 82:763–771.
- Caldecott KW. 2003. XRCC1 and DNA strand break repair. *DNA Repair Amst.* 2:955–969.
- Calenda A, Allenet B, Escalier D, Bach JF, Garchon HJ. 1994. The meiosis-specific Xmr gene product is homologous to the lymphocyte Xlr protein and is a component of the XY body. *EMBO J.* 13:100–109.
- Cocquet J, Ellis PJ, Yamauchi Y, Mahadevaiah SK, Affara NA, Ward MA, Burgoyne PS. 2009. The multicopy gene Sly represses the sex chromosomes in the male mouse germline after meiosis. *PLoS Biol.* 7:e1000244.
- Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer.
- Cutler G, Marshall LA, Chin N, Baribault H, Kassner PD. 2007. Significant gene content variation characterizes the genomes of inbred mouse strains. *Genome Res.* 17:1743–1754.
- de Vries BB, Pfundt R, Leisink M, et al. (18 co-authors). 2005. Diagnostic genome profiling in mental retardation. *Am J Hum Genet.* 77:606–616.
- Dumas L, Kim YH, Karimpour-Fard A, Cox M, Hopkins J, Pollack JR, Sikela JM. 2007. Gene copy number variation spanning 60 million years of human and primate evolution. *Genome Res.* 17:1266–1277.
- Egan CM, Sridhar S, Wigler M, Hall IM. 2007. Recurrent DNA copy number variation in the laboratory mouse. *Nat Genet.* 39:1384–1389.
- Ellis PJ, Clemente EJ, Ball P, Touré A, Ferguson L, Turner JM, Loveland KL, Affara NA, Burgoyne PS. 2005. Deletions on mouse Yq lead to upregulation of multiple X- and Y-linked transcripts in spermatids. *Hum Mol Genet.* 14:2705–2715.
- Escalier D, Garchon H. 2005. XMR, a dual location protein in the XY pair and in its associated nucleolus in mouse spermatocytes. *Mol Reprod Dev.* 72:105–112.
- Good JM, Dean MD, Nachman MW. 2008. A complex genetic basis to X-linked hybrid male sterility between two species of house mice. *Genetics* 179:2213–2228.
- Graubert TA, Cahan P, Edwin D, et al. (11 co-authors). 2007. A high-resolution map of segmental DNA copy number variation in the mouse genome. *PLoS Genet.* 3:e3.
- Guénet JL, Bonhomme F. 2003. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.* 19:24–31.
- Haber JE. 1999. DNA recombination: the replication connection. *Trends in Biochem Sci.* 24:271–275.
- Henrichsen CN, Vinckenbosch N, Zöllner S, Chaignat E, Pradervand S, Schütz F, Ruedi M, Kaessmann H, Reymond A. 2009. Segmental copy number variation shapes tissue transcriptomes. *Nat Genet.* 41:424–429.
- Iafate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. *Nat Genet.* 36:949–951.
- 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.
- Jacquemont M, Sanlaville D, Redon R, et al. (16 co-authors). 2006. Array-based comparative genomic hybridisation identifies high frequency of cryptic chromosomal rearrangements in patients with syndromic autism spectrum disorders. *J Med Genet.* 43:843–849.
- Kidd J, Cooper G, Donahue W, et al. (46 co-authors). 2008. Mapping and sequencing of structural variation from eight human genomes. *Nature* 453:56–64.
- Kuzminov A. 2001. Single-strand interruptions in replicating chromosomes cause double-strand breaks. *Proc Natl Acad Sci U S A.* 98:8241–8246.
- Li J, Jiang T, Mao JH, Balmain A, Peterson L, Harris C, Rao PH, Haviak P, Gibbs R, Cai WW. 2004. Genomic segmental polymorphisms in inbred mouse strains. *Nat Genet.* 36:952–954.
- Liang Q, Conte N, Skarnes WC, Bradley A. 2008. Extensive genomic copy number variation in embryonic stem cells. *Proc Natl Acad Sci U S A.* 105:17453–17456.
- McCarroll SA. 2008. Extending genome-wide association studies to copy-number variation. *Hum Mol Genet.* 17:R135–R142.
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226:792–801.
- Mihola O, Trachtulec Z, Vlcek C, Schimenti JC, Forejt J. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science* 323:373–375.
- Mueller J, Mahadevaiah SK, Park P, Warburton P, Page DC, Turner JM. 2008. The mouse X chromosome is enriched for multicopy testis genes showing postmeiotic expression. *Nat Genet.* 40:794–799.
- Orr HA, Masly JP, Presgraves DC. 2004. Speciation genes. *Curr Opin Genet Dev.* 14:675–679.
- Perry GH, Dominy NJ, Claw KG, et al. (13 co-authors). 2007. Diet and the evolution of human amylase gene copy number variation. *Nat Genet.* 39:1256–1260.
- Perry GH, Tchinda J, McGrath SD, et al. (12 co-authors). 2006. Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci U S A.* 103:8006–8011.
- Perry GH, Yang F, Marques-Bonet T, et al. (14 co-authors). 2008. Copy number variation and evolution in humans and chimpanzees. *Genome Res.* 18:1698–1710.
- Piotrowski A, Bruder CE, Andersson R, et al. (17 co-authors). 2008. Somatic mosaicism for copy number variation in differentiated human tissues. *Hum Mutat.* 29:1118–1124.
- Pozhitkov AE, Tautz D, Noble PA. 2007. Oligonucleotide microarrays: widely applied–poorly understood. *Brief Funct Genomic Proteomic.* 6:141–148.
- Read LR, Raynard SJ, Ruksc A, Baker MD. 2004. Gene repeat expansion and contraction by spontaneous intrachromosomal homologous recombination in mammalian cells. *Nucleic Acids Res.* 32:1184–1196.
- Redon R, Ishikawa S, Fitch KR, et al. (43 co-authors). 2006. Global variation in copy number in the human genome. *Nature* 444:444–454.
- Reynard LN, Turner JM, Cocquet J, Mahadevaiah SK, Touré A, Hoog C, Burgoyne PS. 2007. Expression analysis of the mouse multi-copy X-linked gene Xlr-related, meiosis-regulated Xmr, reveals that Xmr encodes a spermatid-expressed cytoplasmic protein, SLX/XMR. *Biol Reprod.* 77:329–335.
- Sage RD, Heyneman D, Lim KC, Willson AC. 1986. Wormy mice in a hybrid zone. *Nature* 324:60–63.
- Schlötterer C, Tautz D. 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. *Curr Biol.* 4:777–783.
- Sebat J, Lakshmi B, Malhotra D, et al. (21 co-authors). 2007. Strong association of de novo copy number mutations with autism. *Science* 316:445–449.
- Sebat J, Lakshmi B, Troge J, et al. (32 co-authors). 2004. Large-scale copy number polymorphism in the human genome. *Science* 305:525–528.
- She X, Cheng Z, Zöllner S, Church DM, Eichler EE. 2008. Mouse segmental duplication and copy number variation. *Nat Genet.* 40:909–914.

- Snijders AM, Nowak NJ, Huey B, et al. (15 co-authors). 2005. Mapping segmental and sequence variations among laboratory mice using BAC array CGH. *Genome Res.* 15:302–311.
- Storchová R, Gregorová S, Buckiová D, Kyselová V, Divina P, Forejt J. 2004. Genetic analysis of X-linked hybrid sterility in the house mouse. *Mamm Genome.* 15:515–524.
- Stranger BE, Forrest MS, Dunning M, et al. (17 co-authors). 2007. Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315:848–853.
- Teeter KC, Payseur BA, Harris LW, Bakewell MA, Thibodeau LM, O'Brien JE, Krenz JG, Sans-Fuentes MA, Nachman MW, Tucker PK. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. *Genome Res.* 18:67–76.
- Turner D, Miretti M, Rajan D, Fiegler H, Carter N, Biayney ML, Beck S, Hurles ME. 2008. Germline rates of de novo meiotic deletions and duplications causing several genomic disorders. *Nat Genet.* 40:90–95.
- Tuzun E, Sharp AJ, Bailey JA, et al. (12 co-authors). 2005. Fine-scale structural variation of the human genome. *Nat Genet.* 37:727–732.
- Voolstra C, Tautz D, Farbrother P, Eichinger L, Harr B. 2007. Contrasting evolution of expression differences in the testis between species and subspecies of the house mouse. *Genome Res.* 17:42–49.
- Walsh T, McClellan JM, McCarthy SE, et al. (36 co-authors). 2008. Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia. *Science* 320:539–543.
- Watkins-Chow DE, Pavan WJ. 2008. Genomic copy number and expression variation within the C57BL/6J inbred mouse strain. *Genome Res.* 18:60–66.
- West SC. 2003. Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol.* 4:435–445.
- Xue Y, Sun D, Daly A, et al. (12 co-authors). 2008. Adaptive evolution of UGT2B17 copy-number variation. *Am J Hum Genet.* 83:337–346.