Meiotic Recombination Strongly Influences GC-Content Evolution in Short Regions in the Mouse Genome

Yves Clément*[,1,2] and Peter F. Arndt[3]

[1]Montpellier SupAgro, Unité Mixte de Recherche 1334, Amélioration Génétique et Adaptation des Plantes Méditerranéennes et Tropicales, Montpellier, France

[2]Institut des Sciences de l'Evolution de Montpellier, Unité Mixte de Recherche 5554, Centre National de la Recherche Scientifique, Université Montpellier 2, Montpellier, France

[3]Max Planck Institute for Molecular Genetics, Ihnestrasse, Berlin, Germany

***Corresponding author:** E-mail: yves.clement@univ-montp2.fr.

Associate editor: Matthew Hahn

Abstract

Meiotic recombination is known to influence GC-content evolution in large regions of mammalian genomes by favoring the fixation of G and C alleles and increasing the rate of A/T to G/C substitutions. This process is known as GC-biased gene conversion (gBGC). Until recently, genome-wide measures of fine-scale recombination activity were unavailable in mice. Additionally, comparative studies focusing on mouse were limited as the closest organism with its genome fully sequenced was rat. Here, we make use of the recent mapping of double strand breaks (DSBs), the first step of meiotic recombination, in the mouse genome and of the sequencing of mouse closely related subspecies to analyze the fine-scale evolutionary signature of meiotic recombination on GC-content evolution in recombination hotspots, short regions that undergo extreme rates of recombination. We measure substitution rates around DSB hotspots and observe that gBGC is affecting a very short region (~1 kbp) in length around these hotspots. Furthermore, we can infer that the locations of hotspots evolved rapidly during mouse evolution.

Key words: **GC-biased gene conversion, double strand breaks, mouse, substitution rates.**

Article

Introduction

By ensuring the correct pairing and migration of chromosomes during the first cell division of meiosis, meiotic recombination is an important biological process for sexually reproducing organisms (Petronczki et al. 2003). Recombination also enhances the efficiency of natural selection by shuffling alleles and creating new allelic combinations (Coop and Przeworski 2007).

After duplication of the chromosomes, meiotic recombination starts with a double strand break (hereafter designated as DSB) in one chromosome of a chromosomal pair. At the breakpoint, the chromosome is resected and builds a heteroduplex with the homologous region on its sister chromosome. The intact chromosome is finally copied to repair the breakpoint. This entire process is known as gene conversion (for reviews on meiotic recombination and gene conversion, see de Massy [2003] and Chen et al. [2007]).

In heteroduplexes during gene conversion, base mismatches occur at heterozygous sites, which are repaired by exchanging one nucleotide. In mammals, this repair process is biased toward G and C bases (Brown and Jiricny 1988; Bill et al. 1998). This has important implications for population genetics as gene conversion can promote the fixation of G and C alleles. This biased fixation process is similar to, but separated from, natural selection (Nagylaki 1983) and has been called GC-biased gene conversion (or gBGC for short [Galtier et al. 2001; Marais 2003; Duret and Galtier 2009]). This process leaves an evolutionary signature easily identifiable as an excess of A/T to G/C substitutions and a decrease of G/C to A/T substitutions, which is proportional to recombination activity.

After repair of the heteroduplexes and filling up the once resected ends, Holliday junctions are formed and resolved leading to either a noncrossover or a crossover event (the latter being the swapping of chromosomal arms between the two chromosomes of each pair) (de Massy 2003; Baudat and de Massy 2007). Current genetic maps in mammals, however, only have enough resolution to detect crossovers. As the majority of recombination events in mammals lead to noncrossovers (Baudat and de Massy 2007), using crossover rates as a proxy measure of recombination will give us an incomplete picture of how recombination influences genome evolution.

In this study, we make use of recent high-throughput mapping of DSB hotspots (regions of a few kbp that exhibit high amounts of DSB) in the *Mus m. musculus* genome (Brick et al. 2012), which allows for the analysis of influences of recombination on substitution pattern at a fine scale. We also take advantage of the recent sequencing and mapping of several mouse subspecies, including *M. m. castaneus* and *M. spretus* to the *M. m. musculus* genome (Keane et al. 2011). This allows us to study genome evolution at a similar time scale as recombination, as meiotic recombination evolves rapidly in mouse species (Dumont et al. 2011).

The influence of meiotic recombination on GC-content evolution was first observed indirectly through an association

between chromosome size and GC-content (Eyre-Walker 1993). Later, through the analysis of substitution patterns in large regions across genomes in murid rodents and primates, it has been shown that gBGC has an effect on substitution patterns and GC-content evolution: Regions with high recombination will have higher A or T to G or C substitution rates and will evolve toward higher GC-content values as a result (Meunier and Duret 2004; Duret and Arndt 2008; Clément and Arndt 2011). This influence is thought to be an important cause of large-scale variations of the base composition in mammalian genomes called isochores (Bernardi 2000; Eyre-Walker and Hurst 2001). The mapping of DSB hotspots in the mouse genome enables us to study at a several hundred base pair scale the consequences of meiotic recombination on substitution patterns and GC-content evolution.

We were able to determine the evolutionary signature of gBGC by analyzing substitution patterns in the close vicinity of DSBs, finding strong signatures of gBGC around DSBs. We could also infer that DSB locations are evolving rapidly in mouse lineages and finally found no evidence for strand-specific mutations associated with meiotic recombination.

Results and Discussion

gBGC Around DSB Hotspots

We studied substitution rates around DSB hotspots by first pooling all hotspots using their middle points as a reference position. We then computed substitution patterns in the *M. m. musculus* lineage in 60 windows of 100-bp long using *M. m. musculus*–*M. m. castaneus*–*M. spretus* triple alignments (for more details, see the Materials and Methods). These pooled windows contain a total of more than 95 Mb of analyzable sites (sites where all three species have a nucleotide), with an average of 1.6 Mb per window. From substitution patterns, we computed in each window GC* values (equilibrium GC-content), a quantity that is correlated to the strength of gBGC: A stronger gBGC will result in an excess of W → S substitutions and a higher GC* value.

Results in the *M. m. musculus* lineage show an increase of GC* relative to the background and centered on DSB hotspots middle points (fig. 1). This increase affects a region of up to 1 kb with an average length of 500 bp (fig. 1). As DSB products are repaired through gene conversion, which will lead to a biased repair of mismatches occurring in heteroduplexes (Duret and Galtier 2009), we infer that this increase of GC* is due to gBGC around DSB hotspots middle point.

To control whether this increase was specific to DSB hotspots inside the *M. m. musculus* lineage, we computed substitution patterns and GC* values in 10,000 randomly chosen regions not overlapping DSB hotspots (hereafter designated as DSB coldspots), using the same method as for DSB hotspots. Results showed no increase of GC* in DSB coldspots (fig. 1).

We compared GC* values in DSB hotspots in the *M. m. musculus* lineage with the *M. m. castaneus* lineage. We did not observe an increase of GC* inside the *M. m. castaneus* lineage (fig. 1). We therefore conclude that the increase of GC* is specific to DSB hotspots in the *M. m. musculus* lineage. Furthermore, this increase of GC* cannot be explained by the fact that GC-content is highly close to DSB hotspots middle points in *M. m. musculus* (supplementary fig. S4, Supplementary Material online): As the *M. m. musculus* and *M. m. castaneus* divergence time is only 500,000 years with a nucleotide divergence of 0.011, GC-content profiles are very similar between the two species (supplementary fig. S4, Supplementary Material online). If the increase of GC* in DSB hotspots was caused solely by an increase of GC-content, we should see high GC* values in the *M. m. castaneus* lineage. The absence of such increase rules out GC-content as a confounding variable.

Because of the short divergence time between *M. m. musculus* and *M. m. castaneus*, cosegregating single-nucleotide polymorphisms in the two species could affect our results. To control for this, we repeated our analyses using different sets of alignments (e.g., comparing *M. m. musculus* and *M. spretus* and using *Rattus norvegicus* as an outgroup) and found that results do not differ (supplementary materials, Supplementary Material online). Thus, our results are robust with respect to effects due to the short divergence time.

This increase of GC* is mainly due to an increase of W → S substitution rates (fig. 2). This is expected, as recombination increases, the fixation bias favoring G and C alleles is stronger than the fixation bias favoring A and T alleles (supplementary fig. S1 and text, Supplementary Material online).

One can estimate the fraction of the genome affected by gBGC for one individual at each generation. The number of DSBs per individual per generation is estimated to be approximately 400 in *M. m. musculus* (using the number of RAD51 foci as a proxy [Baudat and de Massy 2007]). If gBGC affects a region of 500 bp on average in DSB hotspots, we infer this process to affect approximately 200 kb, or less than 0.01% of the genome per generation. This figure is surprisingly low given how important gBGC is in mammalian genome evolution (Duret and Arndt 2008; Duret and Galtier 2009). However, this figure is a low estimate and is likely to be higher, as more than one crossover per chromosomal arm is possible. Moreover, this is calculated per individual per generation: it is likely that this figure gets higher when estimated for populations or over longer evolutionary time scales.

The human genome experiences more DSBs per individual per generation than the mouse genome (Baudat and de Massy 2007). Provided the length of the region around DSBs affected by gBGC has the same length in both human and mouse genomes, gBGC will affect a larger fraction of the genome in human compared with mouse. This could explain the fact that gBGC has less influence on GC-content evolution in mouse compared with human (Clément and Arndt 2011). A recent method based on phylogenetic hidden Markov models estimated that 0.3% of the human genome is affected by gBGC (Capra et al. 2013), which agrees with our prediction that a larger fraction of the human genome will be affected by gBGC compared with the mouse genome. It should be noted that this estimate is likely to be higher than the per generation estimate.

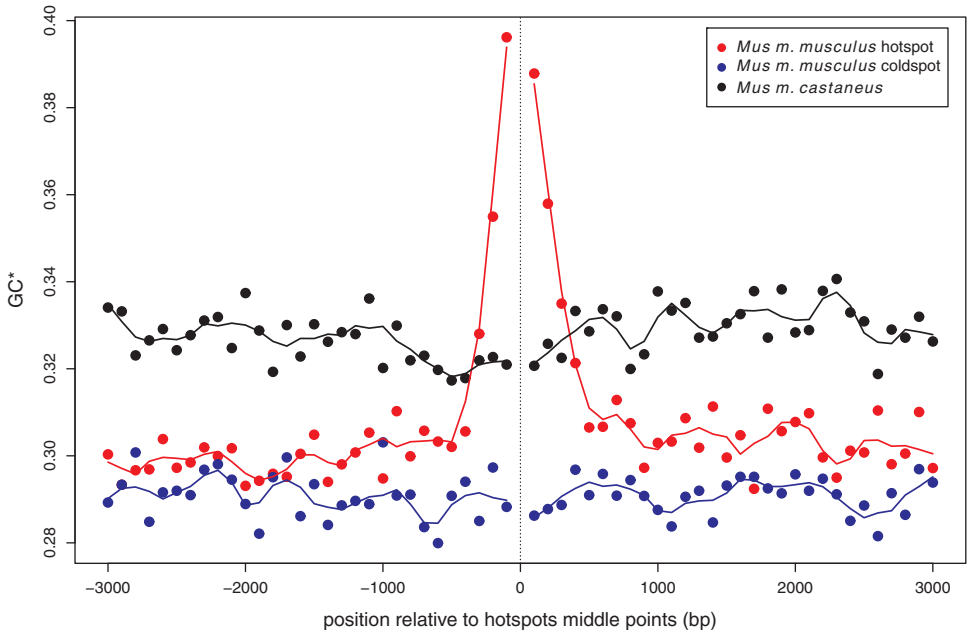

Fig. 1. GC* around DSB hotspots middle points in the *Mus m. musculus* lineage (red), the *Mus m. castaneus* lineage (black), and around *Mus m. musculus* DSB coldspots (blue). Lines represent one-sided local regressions computed over five neighboring windows.

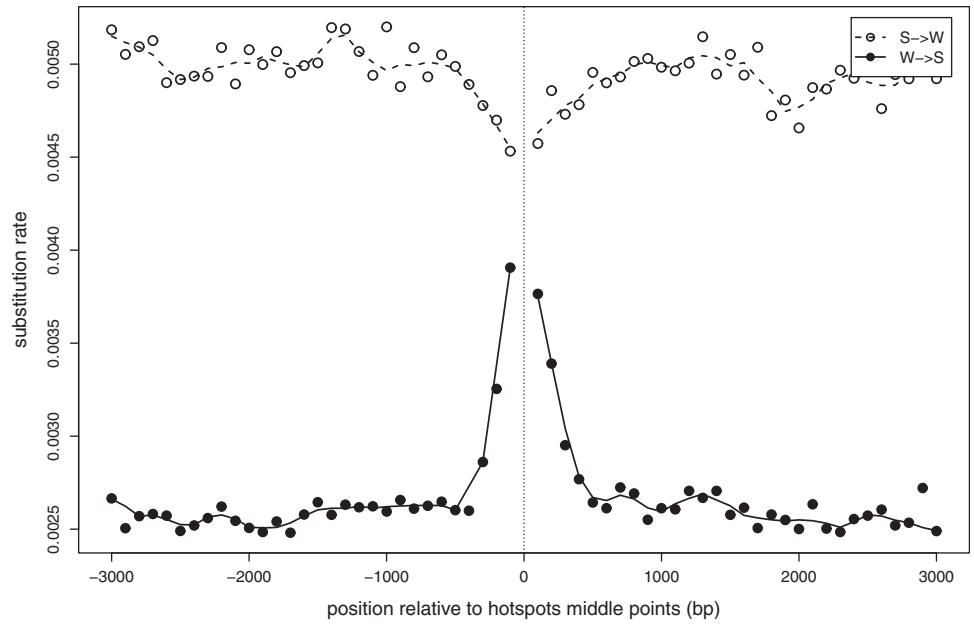

Fig. 2. W → S (solid circles) and S → W (empty circles) substitution rates around DSB hotspots middle points in the *Mus m. musculus* lineage. Lines represent one-sided local regressions computed over five neighboring windows.

The Intensity of gBGC Is Correlated with the Strength of DSB Hotspots

We further analyzed the link between meiotic recombination and GC-content evolution the following way. We used the number of sequencing tags per ChIP-seq peak as a proxy measure of DSB hotspots' strength and divided all hotspots into three groups of same size based on their strength (later designated as high, medium, and low). As the method used to map DSBs specifically targets protein bound to single strand DNA, it makes measures of peak intensities robust to background binding and a good proxy measure of DSB intensity (Khil et al. 2012). We applied the same methodology as indicated before to generate triple alignments in 60 windows of 100 bp around hotspots middle points and compute substitution patterns and GC* values.

GC* values around hotspots middle points for high strength hotspots are higher than for medium and low strength hotspots (P value < 0.015 and 0.004, respectively, one-sided Wilcoxon rank-sum test; fig. 3), whereas GC* values for medium strength hotspots are not significantly

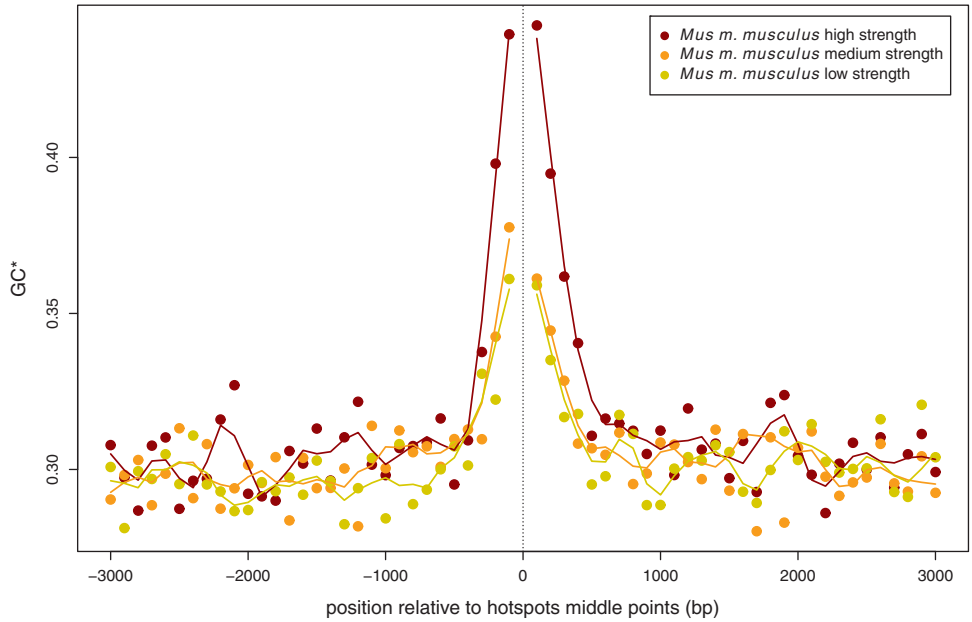

Fig. 3. GC* values around DSB hotspots middle points in the *Mus m. musculus* lineage for hotspots of low strength (yellow), medium strength (orange), and high strength (brown). Lines represent one-sided local regressions computed over five neighboring windows.

higher than low strength hotspots (P value = 0.237, one-sided Wilcoxon rank-sum test; fig. 3). As GC* values are correlated with the intensity of gBGC, we infer that it increases with DSB activity. These results are not confounded by base composition as we observe no link between hotspot intensity and GC-content (supplementary fig. S5, Supplementary Material online).

DSB Locations Are Evolving Rapidly

By analyzing substitution patterns around DSB hotspots in *M. m. musculus* and comparing them with those in corresponding regions in sister species, we can see that DSBs evolve through time. We observe an increase of GC* around DSB hotspots middle points, which is specific to *M. m. musculus*. Because DSBs and their subsequent repair will lead to gBGC and an increase of GC* values, the fact that we cannot observe an increase of GC* at the corresponding locations in *M. m. castaneus* suggests that there is no DSB and recombination occurring at the corresponding locations in *M. m. castaneus*. DSBs locations therefore are very likely different between these two species. Since the *M. m. musculus*–*M. m. castaneus* divergence time is approximately 500,000 years (Geraldes et al. 2008), it shows that this evolution happened very recently.

Such an observation is reminiscent of the fact that meiotic recombination hotspots are poorly conserved in primates (Ptak et al. 2005; Winckler et al. 2005; Coop and Myers 2007; Jeffreys and Neumann 2009). Furthermore, it has been shown that meiotic recombination is controlled in *M. m. musculus* and *Homo sapiens* by a gene called *Prdm9* (Baudat et al. 2010; Myers et al. 2010; Parvanov et al. 2010). This gene exhibits very strong positive selection in metazoans, especially at the DNA-binding residues of its zinc finger domains (Oliver et al. 2009). Models have been proposed to link the fast evolution of *Prdm9* with the fast evolution of recombination hotspots and of binding motifs in the human genome (Hochwagen and Marais 2010; Ponting 2011), and different strains of *M. m. musculus* that use different PRDM9 binding motifs will have different meiotic recombination hotspots (Brick et al. 2012). The fast evolution of DSB hotspots locations in *M. m. musculus* seems to indicate that recombination evolves in a similar way in primates and murids.

No Evidence for Recombination Associated Strand-Specific Mutations

A base composition skew has been reported around DSB hotspots middle points in mouse in an older DSB hotspot data set (Smagulova et al. 2011). When studying the recent data set, a base composition skew can be observed (supplementary fig. S6, Supplementary Material online). As such, skew was interpreted as being the result of strand-specific mutations, which could have been caused by the recombination process or other molecular processes such as transcription (Polak and Arndt 2008) or replication (Chen et al. 2011), we looked for strand asymmetries in substitutions around DSB hotspots middle points.

We first computed around DSB hotspots middle points 14 substitution rates and then compared all pairs of complement rates (2 pairs of transition rates, 4 pairs of transversion rates, and 1 pair of CpG rates; see Materials and Methods for more details). For each pair of complement rates, we computed the log of the ratio of the two rates in each window. For example, for A → G and T → C substitution rates, we computed the following value: $\log_2(A \rightarrow G / T \rightarrow C)$ around DSB hotspots middle points for both the *M. m. musculus* and *M. m. castaneus* lineages.

Results show that strand asymmetries are very weak (fig. 4; supplementary figs. S7–S10, Supplementary Material online).

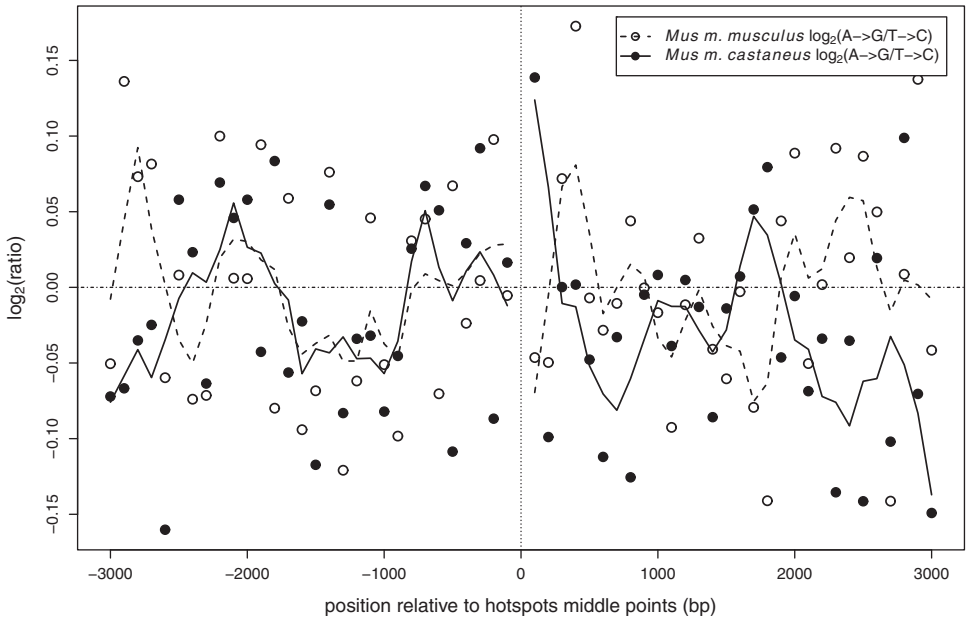

Fig. 4. Strand asymmetries for A → G and C → T substitution rates around DSB hotspots middle points in the *Mus m. musculus* lineage (empty circles) and the *Mus m. castaneus* lineage (solid circles). Lines represent one-sided local regressions computed over five neighboring windows.

Furthermore, strand asymmetries observed in the *M. m. musculus* lineage are of the same magnitude as what is observed in the *M. m. castaneus* lineage (fig. 4; supplementary figs. S7–S10). This shows that strand-specific mutations are not causing the observed base composition skews.

Furthermore, these skews are reverse complement symmetric with respect to DSB hotspots middle points: We observe higher A frequencies on the 5′ end but higher T frequencies on the 3′ end of DSB hotspots, as well as higher G frequencies on the 5′ side but higher C frequencies on the 3′ side (supplementary fig. S6, Supplementary Material online). Base composition skews are therefore independent of the strand orientation of the hotspots. As a result, if they cause the observed base composition skews, strand-specific mutations will also have to be reverse complement symmetric with respect to DSB hotspots middle points: We will be able to observe them regardless of whether we study the + or − strand for the hotspots. This makes our results robust to the fact that all DSB hotspots are oriented in the same manner: The chromosome's centromere is located at the 5′ end, whereas the telomere is located at the 3′ end of DSB hotspots. We therefore conclude that meiotic recombination does not cause strand-specific mutations in *M. m. musculus*. How these skews emerged is still unknown. Nonetheless, there is a bias for DSBs to occur in regions exhibiting such skews. One simple possibility is that regions exhibiting base composition skews are preferentially recruited as DSB hotspots, either directly or indirectly through chromatin opening.

Materials and Methods

We investigated substitution patterns around DSB hotspots, using recently published data (Brick et al. 2012). Hotspots for several strains of mice were published, but we focused on the B6 strain as the published reference *M. m. musculus* genome is of this particular strain.

The method we use to estimate substitution patterns (discussed later) requires approximately 10 kb of aligned sequences to give results that are not dominated by noise effects. To overcome this, given our goal is to study evolutionary signatures of DSBs at a fine scale, we pooled alignment data from all DSB hotspots together, using the hotspot center position as a reference position. We then divided sequences around DSB hotspots middle points into 60 nonoverlapping windows each 100-bp long, 30 on the 5′ side and 30 on the 3′ side. All hotspots are oriented the same way on chromosome, with their 5′ side facing the centromere and the 3′ side facing the telomere. For each window, we built the corresponding *M. m. musculus*–*M. m. castaneus*–*M. spretus* triple alignments by downloading the *M. m. castaneus* and *M. spretus* genomic consensus sequences that were recently published (Keane et al. 2011). As the co-ordinates of those two genomes are identical with those of the *M. m. musculus* (*mm9* version), we could directly compare sequences of these three species. We disregarded insertions and deletions from our alignments. We also masked all exons of the *M. m. musculus* genome from our alignments (Ensembl version 62 annotation, Flicek et al. 2011).

We computed substitution patterns by comparing sequences of several sister species and an outgroup. The divergence time of *M. m. musculus* and *M. spretus* is estimated to be approximately 1 My (She et al. 1990; Suzuki et al. 2004), while their nucleotide divergence (number of positions where bases are different in both species divided by number of positions where both species have a nucleotide) is approximately 0.022. The divergence time of *M. m. musculus* and *M. m. castaneus* is estimated to be approximately 500,000 years (Geraldes et al. 2008, 2011; Duvaux et al. 2011), while

their nucleotide divergence is approximately 0.011 (for comparison, the nucleotide divergence of human and chimpanzee for repeat-masked sequences is ~0.013). Because *M. m. musculus* and *M. m. castaneus* are more closely related to each other than they are to *M. spretus*, we used the latter as an outgroup for the first two species.

We computed substitution rates from these alignments using a maximum likelihood-based method (Arndt et al. 2003; Arndt and Hwa 2005; Duret and Arndt 2008). This method does not assume time reversibility of the substitution process, nor that base composition is at equilibrium and infers one substitution matrix for each branch of the tree. Moreover, it takes into account the hypermutability of methylated cytosines of CpG dinucleotides: C → T and G → A mutations are about 10 times more frequent in CpGs than in non-CpGs (Bird 1978; Giannelli et al. 1999). Finally, this method computes substitution rates independently for each sister species, which allows direct comparison of evolutionary patterns in *M. m. musculus* and *M. m. castaneus*.

As we wanted to compare complementary rates (e.g., the rate of A → G substitutions to the rate of T → C substitutions), we computed 14 rates: 4 transition rates, 8 transversion rates, and 2 CpG rates. We grouped together A/T → G/C substitution rates as Weak (W) → Strong (S) substitution rates, and G/C → A/T substitution rates as S → W substitution rates. A substitution pattern consists of all substitution rates. We finally computed in each window an equilibrium GC-content, or future GC-content (later designated as GC*), which is the expected final GC-content value provided the sequences evolve with a constant substitution pattern through time. GC* values can be viewed as summary values of substitution patterns and shows how strong W → S and S → W rates are relative to each other: high GC* values will indicate high W → S rates relative to S → W rates.

To visualize trends in substitution patterns around DSB hotspots middle points, we performed two independent local polynomial regression on windows on the 5′ side and the 3′ side of reference positions, using the windows' positions as predictor values and GC*, W → S or S → W substitution rates as response values, giving us fitted values. The smoothing was done over 5 neighbor windows, which corresponds to a span parameter of 0.005 for the loess function in R.

To investigate the link between hotspot intensity and substitution patterns, we divided hotspots into three groups of identical size based on their number of ChIP-seq tags per peak and recomputed substitution patterns in each of the three groups using the same methodology as indicated earlier. We performed Wilcoxon rank-sum tests to determine whether GC* values computed in different intensity groups were significantly higher or lower than each other.

Conclusion

We use recently acquired sequence data of closely related species to *M. m. musculus* and fine-scale measures of DSBs to study the signatures of recombination on base composition evolution. Around hotspots of DSB activity, that is, short regions exhibiting a high number of such events, we find that the influence of gBGC is limited to a short region (~1 kbp in size) but has a strong impact. We also present evidence that the location of meiotic recombination along the chromosomes evolves rapidly in mouse lineages, which agrees with previous observations in mouse and primates. Our study quantifies the influence of gBGC on a spatial resolution inside the genome for the first time. It therefore allows for fine-scale comparison of the gBGC process in different species.

Supplementary Material

Supplementary figures S1–S10 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Barbara Wilhelm, Florian Massip, and Laurent Duret for fruitful discussions and comments on previous versions of the manuscript. This work was supported by a stipend from the IMPRS-CBSC.

References

Arndt PF, Burge CB, Hwa T. 2003. DNA sequence evolution with neighbor-dependent mutation. *J Comput Biol.* 10:313–322.

Arndt PF, Hwa T. 2005. Identification and measurement of neighbor-dependent nucleotide substitution processes. *Bioinformatics* 21: 2322–2328.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G, de Massy B. 2010. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* 327:836–840.

Baudat F, de Massy B. 2007. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Res.* 15:565–577.

Bernardi G. 2000. Isochores and the evolutionary genomics of vertebrates. *Gene* 241:3–17.

Bill CA, Duran WA, Miselis NR, Nickoloff JA. 1998. Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese hamster ovary cells. Competition between long-patch and G-T glycosylase-mediated repair of G-T mismatches. *Genetics* 149: 1935–1943.

Bird AP. 1978. Use of restriction enzymes to study eukaryotic DNA methylation: II. The symmetry of methylated sites supports semi-conservative copying of the methylation pattern. *J Mol Biol.* 118:49–60.

Brick K, Smagulova F, Khil P, Camerini-Otero RD, Petukhova GV. 2012. Genetic recombination is directed away from functional genomic elements in mice. *Nature* 485:642–645.

Brown TC, Jiricny J. 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. *Cell* 54: 705–711.

Capra JA, Hubisz MJ, Kostka D, Pollard KS, Siepel A. 2013. A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes. *PLoS Genet.* 9:e1003684.

Chen CL, Duquenne L, Audit B, et al. (11 co-authors). 2011. Replication-associated mutational asymmetry in the human genome. *Mol Biol Evol.* 28:2327–2337.

Chen JM, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. *Nat Rev Genet.* 8:762–775.

Clément Y, Arndt PF. 2011. Substitution patterns are under different influences in primates and rodents. *Genome Biol Evol.* 3:236–245.

Coop G, Myers SR. 2007. Live hot, die young: transmission distortion in recombination hotspots. *PLoS Genet.* 3:e35.

Coop G, Przeworski M. 2007. An evolutionary view of human recombination. *Nat Rev Genet.* 8:23–34.

de Massy B. 2003. Distribution of meiotic recombination sites. *Trends Genet.* 19:514–522.

Dumont BL, White MA, Steffy B, Wiltshire T, Payseur BA. 2011. Extensive recombination rate variation in the house mouse species complex inferred from genetic linkage maps. *Genome Res.* 21:114–125.
Duret L, Arndt PF. 2008. The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* 4:e1000071.
Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu Rev Genomics Hum Genet.* 10:285–311.
Duvaux L, Belkhir K, Boulesteix M, Boursot P. 2011. Isolation and gene flow: inferring the speciation history of European house mice. *Mol Ecol.* 20:5248–5264.
Eyre-Walker A. 1993. Recombination and mammalian genome evolution. *Proc Biol Sci.* 252:237–243.
Eyre-Walker A, Hurst LD. 2001. The evolution of isochores. *Nat Rev Genet.* 2:549–555.
Flicek P, Amode MR, Barrell D, et al. (52 co-authors). 2011. Ensembl 2011. *Nucleic Acids Res.* 39:D800–D806.
Galtier N, Piganeau G, Mouchiroud D, Duret L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159:907–911.
Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu HT, Bulatova N, Ziv Y, Nachman MW. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. *Mol Ecol.* 17:5349–5363.
Geraldes A, Basset P, Smith KL, Nachman MW. 2011. Higher differentiation among subspecies of the house mouse (*Mus musculus*) in genomic regions with low recombination. *Mol Ecol.* 20:4722–4736.
Giannelli F, Anagnostopoulos T, Green PM. 1999. Mutation rates in humans. II. Sporadic mutation-specific rates and rate of detrimental human mutations inferred from hemophilia B. *Am J Hum Genet.* 65: 1580–1587.
Hochwagen A, Marais GAB. 2010. Meiosis: a PRDM9 guide to the hotspots of recombination. *Curr Biol.* 20:R271–R274.
Jeffreys AJ, Neumann R. 2009. The rise and fall of a human recombination hot spot. *Nat Genet.* 41:625–629.
Keane TM, Goodstadt L, Danecek P, et al. (41 co-authors). 2011. Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477:289–294.
Khil PP, Smagulova F, Brick KM, Camerini-Otero RD, Petukhova GV. 2012. Sensitive mapping of recombination hotspots using sequencing-based detection of ssDNA. *Genome Res.* 22:957–965.
Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* 19:330–338.
Meunier J, Duret L. 2004. Recombination drives the evolution of GC-content in the human genome. *Mol Biol Evol.* 21:984–990.
Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G, Donnelly P. 2010. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science* 327:876–879.
Nagylaki T. 1983. Evolution of a finite population under gene conversion. *Proc Natl Acad Sci U S A.* 80:6278–6281.
Oliver PL, Goodstadt L, Bayes JJ, Birtle Z, Roach KC, Phadnis N, Beatson SA, Lunter G, Malik HS, Ponting CP. 2009. Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genet.* 5:e1000753.
Parvanov ED, Petkov PM, Paigen K. 2010. Prdm9 controls activation of mammalian recombination hotspots. *Science* 327:835.
Petronczki M, Siomos MF, Nasmyth K. 2003. Un ménage à quatre: the molecular biology of chromosome segregation in meiosis. *Cell* 112: 423–440.
Polak P, Arndt PF. 2008. Transcription induces strand-specific mutations at the 5′ end of human genes. *Genome Res.* 18:1216–1223.
Ponting CP. 2011. What are the genomic drivers of the rapid evolution of PRDM9? *Trends Genet.* 27:165–171.
Ptak SE, Hinds DA, Koehler K, Nickel B, Patil N, Ballinger DG, Przeworski M, Frazer KA, Pääbo S. 2005. Fine-scale recombination patterns differ between chimpanzees and humans. *Nat Genet.* 37:429–434.
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. *Biol J Linn Soc.* 41:83–103.
Smagulova F, Gregoretti IV, Brick K, Khil P, Camerini-Otero RD, Petukhova GV. 2011. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature* 472: 375–378.
Suzuki H, Shimada T, Terashima M, Tsuchiya K, Aplin K. 2004. Temporal, spatial, and ecological modes of evolution of Eurasian *Mus* based on mitochondrial and nuclear gene sequences. *Mol Phylogenet Evol.* 33: 626–646.
Winckler W, Myers SR, Richter DJ, et al. (11 co-authors). 2005. Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308:107–111.