LINE-1-Derived Poly(A) Microsatellites Undergo Rapid Shortening and Create Somatic and Germline Mosaicism in Mice

Fiorella C. Grandi, James M. Rosser, and Wenfeng An*
School of Molecular Biosciences and Center for Reproductive Biology, Washington State University

*Corresponding author: E-mail: wenfengan@vetmed.wsu.edu.
Associate editor: Norihiro Okada

Abstract

Interspersed and tandem repeat sequences comprise the bulk of mammalian genomes. Interspersed repeats result from successive replication by transposable elements, such as Alu and long interspersed element type 1 (L1). Microsatellites are tandem repeats of 1–6 base pairs, among which poly(A) microsatellites are the most abundant in the human genome. The rise and fall of a microsatellite has been depicted as a life cycle. Previous studies have demonstrated that Alu and L1 insertions are a major source of A-rich microsatellites owing to the concurrent formation of a poly(A) DNA tract at the 3′-end of each insertion. The fate of such poly(A) tracts has been studied by surveying the length distribution of genomic resident Alu and L1 insertions. However, these cross-sectional studies provide no information about the tempo of mutation immediately after birth. In this study, de novo L1 insertions were created using a transgenic L1 mouse model and traced through generations to investigate the early life of poly(A) microsatellites. High frequencies of intra-individual and intergenerational shortening were observed for long poly(A) tracts, creating somatic and germline mosaicism at the insertion site, whereas little variation was observed for short poly(A) alleles. As poly(A) microsatellites are the major intrinsic signal for nucleosome positioning, their remarkable abundance and variability make them a significant source of epigenetic variation. Thus, the birth of poly(A) microsatellites from retrotransposons and the subsequent rapid and variable shortening represent a new way with which retrotransposons can modify the genetic and epigenetic architecture of our genome.

Key words: development, LINE-1, mononucleotide microsatellite repeat, mosaicism, mouse model, poly(A) tract shortening.

Introduction

Repetitive sequences are ubiquitous in mammalian genomes. A recent survey indicates that they may comprise over two-thirds of the human genome (de Koning et al. 2011), a startling contrast to the 1% of DNA that codes for protein products (Church et al. 2009). The majority of these repetitive sequences belong to transposable elements. As the result of repeated insertion into new genomic locations, transposable elements are typically present in the genome as discrete individual copies, therefore referred to as interspersed repeats. Mammals possess four types of transposable elements: long interspersed elements (LINEs), short interspersed elements (SINEs), long-terminal repeat (LTR) retrotransposons, and DNA transposons. Retrotransposons, which encompass LINEs, SINEs and LTR-retrotransposons, replicate through an RNA intermediate, and have played an important role in shaping the architecture of all mammalian genomes (Kazazian 2004; Cordaux and Batzer 2009). LINE-1 (or L1) is an autonomous, non-LTR retrotransposon. There are over half a million copies of L1s in human and mouse genomes, although the majority are 5′-truncated and inactive. Nevertheless, approximately 100 active, full-length L1 copies exist in a diploid human genome (Brouha et al. 2003; Beck et al. 2010); they are capable of creating new copies in the genome through a “copy and paste” mechanism as well as mobilizing non-autonomous SINEs such as Alu and SVA (Moran et al. 1996; Dewannieux et al. 2003; Hancks et al. 2011; Raiz et al. 2012).

Tandem repeats, another class of repetitive sequences, are iterations of repeat units of any size. Based on the size of the repeat unit and the overall length of the repeat, tandem repeats are subgrouped into microsatellites, minisatellites, and satellites. Microsatellites are typically defined as tandem repeats of 1–6 base pairs (bp) in unit length (Ellegren 2004). For repeats that are 12 bp or longer in overall length, mononucleotide repeats are the most abundant microsatellites in the human genome (accounting for ~4 kb per million bp) (Subramanian et al. 2003). Not all nucleotides are equally represented. Over 99% of all mononucleotide microsatellites in the human genome are homopolymeric A repeats, or poly(A) tracts [poly(T) in the reverse strand] (Nadir et al. 1996; Subramanian et al. 2003). The preponderance of poly(A) tracts over poly(C) tracts [poly(G) in the reverse strand] is not a mere consequence of the modest 41% GC content in the human genome (Lander et al. 2001). In fact, A-rich microsatellite repeats are frequently located immediately downstream and contiguous to the 3′-end of Alu and L1 retrotransposons (Economou et al. 1990; Zuliani and Hobbs 1990; Beckman and Weber 1992; Nadir et al. 1996). Furthermore, comparative genomic studies on human and non-human primates indicate that Alu and L1 insertions are a

The rise and fall of a microsatellite is depicted as a life cycle (Buschiazzo and Gemmell 2006). A-rich microsatellites can be born in non-LTR retrotransposons either concurrent or subsequent to their insertion into the genome. The L1 protein machinery mobilizes its own transcripts as well as SINEs and, occasionally, other cellular RNAs. A defining characteristic of these transcripts is a poly(A) tail (Boeke 1997). L1 is transcribed by RNA polymerase II and the poly(A) tail originates from 3’ polyadenylation of L1 mRNA. SINEs, on the other hand, are primarily transcribed by RNA polymerase III (Okada 1991). Their poly(A) tails are partially encoded in the DNA copy and can be lengthened through retrotransposition (Dewannieux et al. 2003; Odom et al. 2004). In either case, a new poly(A) DNA tract is formed at the 3’-end of an insertion concurrent to each retrotransposition event. Subsequent to the retrotransposition event, nucleotide substitutions, deletions, and insertions are responsible for the birth and/or death of other A-rich microsatellites (Kelkar et al. 2011).

The fate of retrotransposition-associated poly(A) tracts has been investigated by surveying the length distribution of genomic resident Alu and L1 insertions. These studies indicate that the length of poly(A) tract is inversely correlated with the evolutionary age of Alu and L1 subfamilies (Ovchinnikov et al. 2002; Roy-Engel et al. 2002; Chen et al. 2005). For example, the mean size of poly(A) tract for evolutionarily old (AluJ and AluS), young (AluYa5), and very recent (disease-associated) Alu insertions is 21, 26, and 56 bp, respectively (Roy-Engel et al. 2002; Chen et al. 2005). In comparison, the mean size of poly(A) tract for primate-specific, human-specific (L1Ta), and disease-associated L1 insertions is 18, 27, and 90 bp, respectively (Symer et al. 2002; Chen et al. 2005). Together, these cross-sectional analyses paint a picture of poly(A) length shortening as a function of time over an evolutionary scale. The early life of retrotransposition-derived poly(A) microsatellites remains to be examined as cross-sectional studies provide no information about the tempo of mutation immediately after birth. In this study, we took advantage of the high frequency of de novo germline insertions in a transgenic L1 mouse model and tracked the intra-individual and intergenerational dynamics of poly(A) tracts at multiple independent L1 insertions. Our results indicate that long poly(A) tracts are highly unstable both in vivo and in vitro and such shortening occurs in both somatic and germline lineages, and thus create somatic and germline mosaicism.

**Results**

**The L1 Poly(A) Tract Undergoes Shortening in a Single Generation**

An L1 transgenic mouse model (An et al. 2006; Rosser and An 2010) was used to generate new germline insertions in the mouse genome (fig. 1A). An intron-disrupted green fluorescent protein reporter cassette was incorporated in the 3’-untranslated region. The intronic sequence would be removed through splicing and absent in resulting L1 insertions. Thus, the presence/absence of the donor L1 transgene and insertions could be diagnosed by a simple polymerase chain reaction (PCR) with a pair of intron-flanking primers (fig. 1B). For example, B1692 was hemizygous for the donor transgene; it had both a large intron-containing band (i.e., the donor) and a small intronless band (i.e., insertions). It is important to note that this assay does not distinguish between germline and somatic insertions in donor-containing animals. In this work, we define germline insertions as inheritable insertions that can be passed to the next generation. They were empirically identified by screening the offspring from a cross between a hemizygous female and a wild-type (WT) male or a reciprocal cross. For example, two of B1692’s offspring, B1712 and B1718, lacked the donor band but were positive for the intronless band, signifying the presence of a germline insertion that had segregated from the donor transgene through meiosis (fig. 1B). To facilitate the tracking of individual germline insertions through generations, we designate animals from which the segregated insertions were first identified as generation zero (G0), animals from subsequent generations as G1 and G2, and so forth.

The germline insertion was independently mapped for G0 animals B1712 and B1718. Both were mapped to the same genomic location, an intergenic area of chromosome (Chr) 2 located 16 kb from the nearest gene (supplementary fig. S1, Supplementary Material online). Although it is possible to observe two independent insertions at the same genomic location (Hancks and Kazazian 2012), we suspected that B1712 and B1718 insertions originated from a single retrotransposition event in the donor-positive parent. To verify the genomic location, we designed a 3’-junction PCR with an L1 primer and a primer specific to the flanking genomic DNA (gDNA) (fig. 1C). The expected size of the amplicon, excluding the poly(A) tract between the L1 sequence and the 3’-genomic sequence, is 129 bp. However, a striking difference in the size of the amplicon was observed between B1712 (251 bp) and B1718 (152 bp). The respective PCR products from B1712 and B1718 were recovered and sequenced. Sequencing results confirmed the same genomic location at Chr 2 but revealed variation in the number of adenosine bases in the poly(A) tracts. Subsequent GeneScan analysis confirmed the difference noted on the gel and allowed for an accurate estimation of poly(A) tract length for B1712 (251–129 = 122 As) and B1718 (152–129 = 23 As) (fig. 1D).

The observed poly(A) length polymorphism in the Chr 2 insertion of B1712 and B1718 could reflect either an authentic allelic difference in vivo or a PCR artifact in vitro. When amplifying microsatellite repeats, there is a concern of creating artificial, PCR-derived shortened amplicons (Hauge and Litt 1993; Clarke et al. 2001). To distinguish between these two possibilities, we set up independent PCR reactions with either single or mixed gDNA samples at varied ratios. The observed difference in poly(A) tract was reproducible in separate PCR reactions when visualized by gel electrophoresis (supplementary fig. S2A, Supplementary Material online). Further genotyping indicated that the B1712 and B1718 insertions had identical 5’-junction between the flanking gDNA and the
L1 sequence (supplementary fig. S2B, Supplementary Material online). Thus, our data support the notion that the B1712 and B1718 insertions originated from the same retrotransposition event and the L1-associated poly(A) tract subsequently shortened in the gamete leading to animal B1718. Additional experiments indicate that the parent B1692 was also positive for the insertion and that a donor-positive littermate B1715 carried a shortened allele with 41 As (supplementary fig. S2B–D, Supplementary Material online).

Two additional germline insertions, located in Chr 8 and Chr 10, were also characterized. The insertion in Chr 8 was derived from a cross between a donor-positive female mouse B1720 and a WT male (supplementary fig. S3, Supplementary Material online). It was first mapped in one of the donor-negative offspring, B1919 (G0), to the second intron of the insulin receptor gene (supplementary fig. S1, Supplementary Material online). The insertion in Chr 10 was isolated in donor-negative G0 animal B1498 (supplementary fig. S4, Supplementary Material online) and mapped to an intergenic area, 27 kb from the nearest annotated transcript (supplementary fig. S1, Supplementary Material online).

Poly(A) Tract Shortening Occurs in Both Somatic and Germline Lineages over Multiple Generations

To investigate the timing of poly(A) tract shortening, we traced the long and short alleles of the Chr 2 insertion over several generations. This was achieved by crossing one of the animals carrying a variant allele with a WT mouse. If poly(A) shortening occurred only during the development of the gamete destined to become the offspring, it would manifest as a single allele of the same size in all tissues. In contrast, if shortening occurred during embryonic and/or postnatal development, tissue mosaicism consisting of both the original allele and the shortened allele(s) would be expected. Gel electrophoresis showed that the long poly(A) tract (122 As) passed from B1712 was unstable and prone to variable shortening in the G1 generation (fig. 2A; the full set in supplementary fig. S5A, Supplementary Material online). GeneScan analysis revealed a surprisingly large number of allelic variants within the G1 generation (fig. 2B; the full set in supplementary fig. S5B, Supplementary Material online). To analyze the allelic variants in more detail, the following convention was established: 1) an allele was defined as a cluster of stutter bands with a distinct peak; 2) the presence or absence of the parental allele was determined, designated either as parental or shortened, respectively; 3) the presence or absence of additional allelic variants was determined, designated as either multiple or single, respectively; and 4) finally, each animal was classified into one of four possible categories (parental single, parental multiple, shortened single, and...
shortened multiple; fig. 2B and supplementary fig. SS8, Supplementary Material online).

The four distinct categories of allelic variations observed in the tail gDNA of G1 offspring indicate that poly(A) tract shortening can occur in both somatic and germline lineages (table 1 and supplementary fig. SS8, Supplementary Material online). Four of the 27 G1 animals lacked the parental allele and were assigned to either shortened single or shortened multiple category. We reasoned that, if the gamete had started with a parental allele, the parental allele would have been preserved in at least a fraction of the cells even if a shortening event had occurred as early as at the two-cell stage (it is possible, although less likely, that remaining cells containing the parental allele undergo independent shortening episodes, which effectively dilute the presence of any parental alleles). In addition, each of the four animals had a dominant shortened allele of different length. Thus, these animals may represent independent poly(A) shortening events in the germline, with a frequency of 14.8% (4/27).

The remaining 85.2% (23/27) of the animals carried the parental allele and were assigned to either parental single or parental multiple category, suggesting no shortening occurred in the gamete before fertilization. In contrast, somatic shortening appeared to be much more prevalent since the majority of the G1 animals, that is, 85.2% (23/27), had multiple alleles, including 21/23 carrying the parental allele and 2/4 lacking the parental allele. Similar levels of somatic/germline shortening were obtained when an independent insertion on Chr 10 with 111 As was propagated from G0 animal B1498; 2 of 15 G1 animals (13.3%) showed germline shortening and the majority of animals carrying the parental allele showed somatic mosaicism (supplementary fig. S4C, Supplementary Material online). It appears that the frequency of poly(A) shortening is positively correlated with the length of the poly(A) tract in the parental allele. When the same Chr 2 insertion was propagated as a shortened allele of 23 As from B1718, all its G1 offspring carried the parental size (supplementary fig. S6A and B, Supplementary Material online). Similar results were obtained when an independent insertion on Chr 8 with 57 As was propagated from B1919 to its G1 offspring (supplementary fig. S6C and D, Supplementary Material online).

Variable poly(A) shortening in somatic cells was further confirmed by analyzing additional tissues from individual G0 and G1 animals. Among G0 animals, additional alleles were observed in heart, liver, and ovarian tissues from B1712 (122 As; fig. 3A), but not from B1718 (23 As; fig. 3B). The overall banding pattern for each tissue was highly reproducible among multiple independent PCR reactions (supplementary fig. S7, Supplementary Material online), suggesting the minor alleles detected are not PCR artifacts. Data from two additional independent insertions confirmed the tendency of having more somatic variation for long poly(A) tracts (i.e., B1498, supplementary fig. S8A, Supplementary Material online) but less variation for short poly(A) tracts (i.e., B1919, supplementary fig. S9A, Supplementary Material online). Similar patterns were observed in somatic tissues of G1 offspring (B1712 G1 tissues in supplementary fig. S10, Supplementary Material online; B1498 G1 tissues in supplementary fig. S8B, Supplementary Material online; and B1919 G1 tissues in supplementary fig. S9B, Supplementary Material online). Animals carrying long poly(A) tracts (>100 bp) displayed multiple allelic variants in all tissues examined; conversely, animals inheriting short poly(A) tracts (<60 bp) displayed few or no variants. No major difference was detected between the number of variants in the heart, liver, and testis among B1712 G1 offspring (supplementary fig. S10, Supplementary Material online).

The amount of allelic variation in both somatic and germline tissues reported here is unprecedented. It is possible that the observed high frequency of poly(A) shortening may only
apply to nascent long poly(A) tracts present in the genome in a hemizygous state. To examine this possibility, we obtained homozygous G3 animals by intercrossing G2 animals. The zygosity of the offspring was determined by PCR genotyping with primers flanking the Chr 2 insertion. Hemizygous and homozygous animals were differentiated by the presence and absence of the empty allele, respectively (examples given in supplementary fig. S2, Supplementary Material online). The dynamics of poly(A) shortening was tracked in homozygous G4 animals (fig. 4A). The maternal G3 harbored a long allele of 122 As and a short allele of 13 As whereas the paternal G3 carried two alleles of 122 As and 98 As, respectively (fig. 4B). Both showed multiple minor alleles, indicative of somatic shortening events. The four parental alleles were re-assorted in the G4 offspring as expected (fig. 4C). It appeared that R387 retained the long maternal allele and a shortened version of a paternal allele. The latter likely arose during spermatogenesis in R292. Additionally, all G4 offspring showed multiple minor alleles, confirming the presence of somatic shortening. In sum, the analysis of two homozygous generations indicates that the poly(A) shortening occurs regardless of zygosity.

**Poly(A) Tract Shortening Occurs During Clonal Expansion of Bone Marrow Progenitor Cells**

The extensive somatic mosaicism revealed by GeneScan analysis, although remarkable, provides no temporal information about the shortening events. To obtain such information, we performed a clonal analysis of adult bone marrow progenitor cells (Pereira et al. 2007) (fig. 5A). Briefly, bone marrow was isolated from B1712 G1 animals and cultured in vitro. Under the culture conditions used, only bone marrow progenitor cells would proliferate and form distinct colonies. An example is shown for B2031, which carried the original allele with 122 As (fig. 5B). Gel electrophoresis showed the presence of the parental allele in 16 out of the 17 colonies examined (supplementary fig. S11A, Supplementary Material online). GeneScan analysis revealed the presence of additional minor alleles in each of these 16 colonies (fig. 5C; additional
colonies in supplementary fig. S11B, Supplementary Material online). Each colony was formed from a single progenitor cell after an average of 12 mitotic divisions. The presence of additional alleles in all long poly(A) containing colonies indicates that multiple shortening events had occurred in vitro during this time span. One of the 17 colonies showed a single dominant allele with 23 As (fig. 5C, colony 9), which likely originated from a progenitor cell already carrying this allele at the time of plating; indeed, a minor allele of the same size was present in the pre-plating bone marrow cells (fig. 5B). Such pre-existing shortened alleles were observed from two additional G1 animals (supplementary fig. S12A and B, Supplementary Material online: colony 17 from B2032, and colonies 2, 13, and 14 from B2033). In total, 5 out of 53 colonies from the three biological replicates lacked the long parental allele, suggesting a frequency of somatic shortening at 9.4% in the bone marrow progenitor cells.

Poly(A) Tract Shortening Occurs When Propagated in Bacteria on a Plasmid

Thus far, we have shown that long poly(A) tracts undergo frequent shortening both in vivo and in vitro, likely resulting from stochastic replication slippage by mouse DNA polymerases. To test whether replication slippage of long poly(A) tracts is a shared property between eukaryotic and prokaryotic DNA polymerases, the long poly(A) allele of a B1712 G1 animal was subcloned into a plasmid and propagated in Escherichia coli (fig. 6A). Ten colonies picked from the initial bacterial transformation showed length variation (two shown in fig. 6B). Each of these 10 colonies was further streaked and grown into discrete colonies. Five subclones were analyzed for each of the 10 original colonies; most displayed additional allelic variants that were not present in the parental colony (fig. 6C).

Discussion

Using a transgenic L1 model, we investigated the early life of mononucleotide repeats birthed into the genome by retrotransposons and the dynamics of their length variation in both somatic and germ cell lineages. High frequencies of germline and somatic shortening were observed for long poly(A) microsatellites, such as those >100 bp alleles in B1712 and B1498, creating somatic and germ cell mosaicism, whereas little variation was observed for short alleles, such as those <60 bp in B1919 and B1718. These observations...
are consistent with the previously reported mutational bias for long di- and tetra-nucleotide microsatellites toward contractions (Ellegren 2000; Xu et al. 2000; Huang et al. 2002; Calabrese and Durrett 2003). In our study, the intergenerational/germline shortening of long > 100 bp poly(A) tracts occurred at a frequency ranging between 14.8% (as for the Chr 2 insertion in B1712) and 13.3% per generation (as for the Chr 10 insertion in B1498). The intra-individual/somatic shortening was even more prevalent: most animals carrying long > 100 bp poly(A) tracts displayed one or more shortened alleles in every tissue examined. Clonal analysis revealed multiple pre-existing shortened alleles in the isolated bone marrow progenitor cells; additional shortening events were evident during in vitro proliferation. Thus, our data provide a rare glimpse of the rapid shortening events immediately following the birth of long poly(A) microsatellites.

The relationship between microsatellite mutability and the number of repeat units has been previously modeled with microsatellites catalogued in human and chimpanzee genomes. Overall, the mutability grows exponentially as the number of repeat units increases (Webster et al. 2002; Lai and Sun 2003; Kelkar et al. 2008). For example, the average mutation rates for 10 and 45 bp poly(A) tracts are ~1/50,000 and ~1/1,000 per locus per generation, respectively (Kelkar et al. 2008). The mutation rates for much longer poly(A) microsatellites are unknown because of the scarcity of such poly(A) tracts in the sequenced genomes. An extrapolation from Kelkar et al.’s exponential model predicts an average mutability of ~1/8 per locus per generation for 100 bp poly(A) tracts. This rate is in line with our germline data, indicating that mutability continues to increase exponentially even when the length of poly(A) microsatellites reaches beyond 100 bp. A potential caveat is that an engineered L1 containing the codon-optimized ORFeus element was used to launch new insertions but we believe that the findings would be the same for poly(A) tracts associated with endogenous L1s. In fact, long > 100 bp poly(A) tracts have been reported for many disease-causing endogenous L1 insertions (Chen et al. 2005).

Length changes in microsatellites are generally attributed to replication slippage, which involves transient strand dissociation and subsequent out-of-register strand realignment during DNA replication (Levinson and Gutman 1987; Ellegren 2004). The role of recombination in microsatellite mutability is considered secondary (Ellegren 2004; Kelkar et al. 2008). In this study, we observed similar frequencies of shortening for long poly(A) tracts when they were present as either hemizygous or homozygous alleles, suggesting homologous recombination does not play a large role in the contraction of long mononucleotide microsatellite repeats. Overall, our in vivo data support a model in which poly(A) shortening occurs in both germline and somatic lineages (fig. 7). It appears that the only prerequisite for poly(A) shortening is multiple rounds of DNA replication and cell divisions. Furthermore, our in vitro data indicate that long poly(A) tracts are highly instable and can be shortened within a few cell divisions, regardless whether they are replicated by eukaryotic or prokaryotic DNA polymerases.

Microsatellites provide a dynamic source of genetic variation and are thought to play a variety of important functions in the genome (Hannan 2010). They are perhaps best known for their roles in many neuromuscular diseases and cancers. For example, the expansion of trinucleotide repeats inside coding regions, promoters and introns alters gene expression and is responsible for Huntington’s disease, fragile X syndrome and Friedreich’s ataxia (Li et al. 2004). Approximately 15% of colorectal cancers are associated with microsatellite instability (MSI) owing to a deficient mismatch repair system (Vilar and Gruber 2010). Several intragenic poly(A) tracts have been used as preferred microsatellite markers to diagnose MSI in patients with cancer (Laghi et al. 2008). MSI-High cancers show elevated frequencies of mononucleotide repeat mutations in coding as well as noncoding regions of many candidate genes (Duval and Hamelin 2002). The consequence of mononucleotide repeat variation in coding regions is easy to understand as it frequently causes frameshift mutations, but our understanding of the role of poly(A) length changes in noncoding or intergenic regions is very limited. In two separate cases, shortening of intronic poly(A) microsatellites has been shown to cause exon skipping and impair normal cellular function (Giannini et al. 2004; Dorard et al. 2011). The largest functional impact of poly(A) length changes

![Fig. 7. Developmental timing of poly(A) tract shortening. Microsatellite length expansion/contraction mainly depends on DNA replication, and thus poly(A) tract shortening is coupled with mitotic divisions. Our data support a model in which shortening occurs in both germline and somatic lineages. In the germline, shortening primarily occurs in pre-meiotic, mitotically dividing germ cells, generating mature gametes with differing allelic sizes of the poly(A) tract and germline mosaicism. In somatic tissues, shortening can occur during embryonic as well as postnatal development (e.g., in actively dividing cell lineages such as bone marrow progenitors), resulting in somatic mosaicism in all tissues. Thus, in a single generation, animals can possess poly(A) alleles of different lengths at the same genomic locus. The poly(A) allelic variants are represented by the number of As. The relative frequency of shortening events is indicated by the thickness of the arrows.](image-url)
may stem from a unique property of poly(A) microsatellites. They serve as the major intrinsic signal for nucleosome positioning: nucleosomes are strongly depleted from poly(A) tracts owing to their unusual structural, dynamic, and mechanical properties, and the depletion of nucleosomes has been shown to extend for considerable distances into the flanking DNA (Segal and Widom 2009). Accordingly, poly(A)-mediated nucleosome depletion affects chromatin organization and influences genomic function by regulating local DNA accessibility to other DNA binding proteins. Given their remarkable abundance and variability, we propose that poly(A) microsatellites constitute a significant source of epigenetic variation in the genome. Thus, the birth of poly(A) microsatellites from retrotransposition and their subsequent rapid and variable shortening in both germline and somatic lineages represent yet another way with which non-LTR retrotransposons can modify both the genetic and epigenetic architecture of our genome.

**Materials and Methods**

**Mouse Genotyping**

All animal work was conducted in accordance to protocols approved by the Institutional Animal Care and Use Committee at Washington State University. gDNA was isolated from tail biopsies using the Gentra PureGene Tissue Kit (Qiagen). Tissues were collected from euthanized animals and kept frozen in RNAlater (Ambion). gDNA was isolated from tissues using the DNeasy Blood and Tissue Kit (Qiagen) after homogenization with TissueLyser II (Qiagen). PCR genotyping of the donor transgene and insertion(s) was performed as previously described (An et al. 2006). Once an insertion was mapped, primers were designed in the flanking region(s) to amplify the 5′- or 3′-junction of the insertion. All primers are listed in supplementary table S1, Supplementary Material online.

**L1 Insertion Mapping**

Both inverse PCR (iPCR) (An et al. 2009) and splinkerette PCR (Uren et al. 2009) were used for initial mapping of the genomic location of each L1 insertion. For splinkerette PCR, DNA from intronless animals was divided into 2 mg aliquots and digested overnight by five different enzymes: MspI, Msel, EcoRI, CviQI, and Sau3AI (New England Biolabs). After digestion, the DNA was ligated to synthetic adapters with compatible ends using T4 DNA ligase (New England Biolabs). The ligated DNA was amplified in two rounds of PCR reactions. The first round of PCR was performed with a biotin-tagged L1 forward primer (WA198) and a reverse primer annealing to the adapter (WA053). The Dynal kilobaseBINDER Kit (Invitrogen) was used to enrich for PCR products containing the biotin forward primer according to the manufacturer’s specifications except that the incubation step was overnight. This purified product was subjected to the second round of PCR with L1 forward primer J88939 and reverse primer WA054 in the adapter sequence. The products were visualized on agarose gels. Positive bands were extracted using the QIAquick Gel Extraction Kit (Qiagen). If the concentration was sufficient, the product was directly sequenced. Low concentration bands were cloned using the StrataClone PCR Cloning Kit (Agilent Technologies) prior to Sanger sequencing.

**Gel Electrophoresis**

All PCR products were resolved by gel electrophoresis using LE agarose (GeneMate) or high-resolution MetaPhor agarose (Lonza). LE agarose gels were prepared at 1% or 2% w/v using 1X Tris–Taurine–EDTA (TTE) buffer and solidified at room temperature. Gels were run at room temperature for 1–2 h at 4.4 V/cm. MetaPhor gels were prepared at 2% w/v with cold 1X TTE buffer and were heated in increments of 30 s until fully dissolved. Gels were allowed to solidify at room temperature and then chilled for 20 min to 1 h. The samples were run in 1X TTE buffer at 4.4 V/cm for 3–4 h. To prevent overheating, the gel apparatus was placed inside a 4°C chamber and chilled with an ice bath. The length of poly(A) size was estimated by comparing with the molecular weight standards, which were consistently annotated across all gel images.

**GeneScan Fragment Analysis**

For GeneScan analysis, 3′-junction PCR was performed with a 6-FAM-labeled L1 primer (WA350) common to all insertions and the respective 3′-flanking gDNA primer. Approximately 0.25–1 μL of each PCR reaction was mixed with 0.5 μL of GeneScan 500 LIZ Size Standard and variable amount of Hi-Di formamide (Applied Biosystems) up to a total volume of 15 μL. The samples in sets of 48 or 96 were heat denatured and fluorescent data from PCR fragments and the size standard were collected with an Applied Biosystems 3730 DNA Analyzer. The internal size standard provided 16 single-stranded labeled fragment (35, 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, 350, 400, 450, 490, and 500 nucleotides) and was used by the GeneMapper Software version 3.7 (Applied Biosystems) to size PCR fragments in each sample. An allele was defined as a cluster of stutter bands; the length of the poly(A) tract was scored as the size of the highest peak as previously reported for trinucleotide repeats (Mangiarini et al. 1997).

**Mouse Bone Marrow Culture**

Hematopoietic colony-forming cell assays were performed as previously described (Pereira et al. 2007). Briefly, mouse femurs were dissected and bone marrow was isolated by flushing with 1 mL of Iscove’s modified Dulbecco’s medium (IMDM) supplemented with 2% fetal bovine serum (FBS). Cells were then pelleted at 500 × g for 7 min and resuspended in 100 μL of DNase (1 mg/mL). Red blood cells were lysed with 1 mL of Ammonium Chloride Solution (Stemcell Technologies) at 4°C for 5 min. The lysis reaction was stopped by adding 3 mL of Dulbecco’s phosphate-buffered saline (DPBS) + 2% FBS and cells were next pelleted at 500 × g for 7 min. Cells were resuspended with 100 μL of DNase and 500 μL of IMDM + 2% FBS. Cell and viability counts were performed using glacial acetic acid and trypan blue. Bone marrow cells were plated in ColonyGEL Complete Medium
Without Epo (ReachBio LLC) at a density of 2 × 10^4 or 1 × 10^4 cells in 35 mm cell culture dishes and cultured at 37°C with 5% CO₂ for 14 days. The remaining pre-plating cells were pelleted and cell pellets were frozen at −20°C. After 14-day culture, colonies were picked and washed in 200 μL of DPBS + 2% FBS. Cells were then pelleted and stored frozen at −20°C. Additional colonies were pooled and counted; the average number of cells per colony was 4,038. Assuming no cell death, the average number of cell divisions was estimated to be 12. Culture components were purchased from Hyclone unless indicated otherwise. To amplify the 3'-junction of Chr 2 insertion in cultured bone marrow cells, we utilized a direct PCR approach. Frozen colonies were resuspended in 5 μL of Dilution Buffer and 0.125 μL of DNARelease Additive (Finnzyme). The cells were vortexed for 5 to 10 s, briefly centrifuged, and subsequently incubated at 98°C/C 14° for 5 min. Cells were subsequently incubated at room temperature for 5 min. Cells were subsequently incubated at 98°C for 2 min and placed on ice. 0.5 μL of released template from each colony was used in each subsequent PCR reaction. As an endogenous control for DNA quality, the Hprt gene was amplified as previously described (An et al. 2006).

Clonal Analysis in Bacteria
The main band from 3'-junction PCR of B1769 (supplementary fig. S5A, Supplementary Material online) was gel purified using the QIAquick Gel Extraction Kit (Qiagen) and cloned using the StrataClone PCR Cloning Kit (Agilent Technologies). Purified plasmid DNA prep was used to transform chemically competent TOP10 E. coli cells by heat shock. Transformants were plated on a fresh LB–carbenicillin plate and allowed to grow overnight. Ten distinct colonies were picked and amplified by colony PCR with the same 3'-junction PCR primers. Each of these 10 colonies was preserved by replicating on a fresh LB-carbenicillin plate and allowed to grow overnight. Single subclones were obtained by dilution streaking on additional LB–carbenicillin plates. Five subclones were analyzed for each original colony by colony PCR.

Supplementary Material
Supplementary table S1 and figures S1–S12 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
The authors thank Haig Kazazian and Dustin Hancks for helpful discussions, Richard Badge and Peter Freeman for advice on target enrichment with biotinylated oligos, and Weiwei Du and Derek Pouchnik for assistance in GeneScan analysis. They are very grateful to the two anonymous reviewers for their constructive comments and insightful suggestions, which greatly improved this article. This work was supported by the American Cancer Society (IRG-77-003-32) and start-up funds from Washington State University. F.C.G. is supported by the WSU STARS Program and an Auvil Fellowship. J.M.R. is supported by the National Institutes of Health Training Program (5T32GM008336-22).

References


