Gene Reports 4 (2016) 169–176

Contents lists available at ScienceDirect

Gene Reports

journal homepage: www.elsevier.com/locate/genrep

The canine X chromosome is a sink for canine endogenous retrovirus transposition☆

Rachelle Mariano [a], Bridgett vonHoldt [b,*]

[a] Computer Science, University of Miami, P.O. Box 248154, Coral Gables, FL 33124-4245, USA
[b] Ecology and Evolutionary Biology, Princeton University, 106A Guyot Hall, Princeton, NJ 08544-2016, USA

ARTICLE INFO

Article history:
Received 10 March 2016
Received in revised form 2 May 2016
Accepted 20 May 2016
Available online 26 May 2016

Keywords:
Transposable elements
Methylation
Domestication
Canine endogenous retrovirus
Canid genome evolution
Retrotransposons

ABSTRACT

The domestic dog has experienced intensive artificial selection to preserve breed phenotypes. We predicted this stress activates transposition of canine endogenous retroviruses (CERVs) to restore host genomic variation. Subsequent epigenetic silencing via DNA methylation would prevent rampant amplification. Consequently, recently inserted CERVs would show breed- and species-specific patterns of insertion and methylation. We annotated 477 CERVs and employed a Kimura two-parameter model to estimate their insertion date. We found the X chromosome to be enriched with CERV insertions, primarily in non-coding regions. With reduced representation bisulfite sequencing data from 55 dogs and 34 gray wolves, we assessed CERV methylation. Our results suggest that most CERVs inserted before species divergence and are hyper-methylated. CERV methylation is neither breed- nor species-specific and has a significant inverse correlation with age. Therefore, CERVs may impact gene expression via cis-regulation, especially on the X chromosome, but not active transposition.

1. Introduction

Transposable elements (TEs) are emerging as potent regulators of gene expression that participate in symbiotic relationships with their host genomes (Suzuki and Bird, 2008; Hollister et al., 2011). Although TEs comprise nearly half of mammalian genomes, they are irregularly distributed and often inhabit intergenic regions (Lander et al., 2001; Akagi et al., 2013). However, neutral or beneficial TE insertions may modulate nearby genes via their regulatory elements and interactions with transcriptional machinery. For example, long terminal repeat retrotransposons (LTR-RTs) are known to engage in RNA editing and translational regulation, as their long terminal repeats harbor promoters, termination signals, and alternative splice sites that can modify transcripts in response to environmental signals (Muotri et al., 2007; Rebollo et al., 2012). LTR-RTs can also bind transcription factors that modulate multiple stress-induced pathways, and subsequent transcriptional activation and transposition of these elements may improve host fitness (Muotri et al., 2007; Grandbastien, 1998; McCue et al., 2012). This phenomenon has been well documented in plants, bacteria, and *Drosophila*, where RT movement generates tractable phenotypes (Grandbastien, 1998, 2004; Capy et al., 2000; Civáň et al., 2011).

As genomic stress from inbreeding can physiologically imitate environmental stress and potentially stimulate RT mobilization, we selected the domestic dog as our model to study the potential effects of inbreeding on RT activity (Kristensen et al., 2005). Dog domestication occurred approximately 11–16,000 years ago and imposed a genetic bottleneck in the founding population (Freedman et al., 2014). Breeding for functional traits intensified during the Victorian era (c. 1800) when the aesthetic appeal of different dog breeds inspired strict closed breeding programs (vonHoldt et al., 2010). Today, the domestic dog genome possesses 94.6% homogeneity within individual breeds as a result of this intense artificial selection (Ostrander and Wayne, 2005). TE activity accounts for phenotypic differences across dog breeds, with 34% of the dog genome composed of TEs, some of which are highly active and account for over 10,000 bimorphic sites (e.g. SINEs) and linked phenotypes (e.g. coat color, body size) (Kirkness et al., 2003; Lindblad-toh et al., 2005; Clark et al., 2006; Gray et al., 2010).

We speculated that LTR-RT amplification in the dog genome combats the stress imposed by rigorous artificial selection and increased homozygosity due to line breeding practices. This activation would be characterized by recent LTR-RT transposition into euchromatic regions near genes. However, uncontrolled transposition can lead to deleterious genomic instability (Hedges and Deininger, 2007). As such, we proposed that cytosine methylation accumulates on canine LTR-RTs to suppress their further relocation, as cytosine methylation causes distortion of the DNA that prevents transcription factor binding and produces epigenetic silencing (Jones and Takai, 2001). We assumed that *de novo* methylation of newly integrated LTR-RTs would cause a higher percent

☆ Data Access: All data are provided as Supplemental Materials.

* Corresponding author at: Ecology & Evolutionary Biology, Princeton University, 328 Moffett Hall, Princeton, NJ 08544-2016, USA.
E-mail addresses: r.mariano@umiami.edu (R. Mariano), vonholdt@princeton.edu (B. vonHoldt).

http://dx.doi.org/10.1016/j.genrep.2016.05.003
2452-0144/

methylation in these species compared to older TEs, as changes in chromatin architecture would have permanently silenced transcription in the latter (Hollister and Gaut, 2009).

To test these propositions, we performed a *de novo* annotation of canine endogenous retroviruses (CERVs) in the boxer CanFam3.1 genome assembly and compared our CERV candidates against those that were previously published (Tarlinton et al., 2013; Jo et al., 2012; Barrio et al., 2011). We assessed the methylation status and age of our candidates and performed principal component analysis (PCA) to evaluate breed-specific methylation. We aimed to better understand the regulatory dynamics of the domestic dog genome, as this is crucial for determining the sustainability of current purebred breeding practices. Our results also provide insight into a potential driver of the incredible phenotypic diversity seen in domestic dogs.

2. Results

2.1. CERV annotation and comparison to literature-curated CERVs

The conserved viral protein coding sequences and distinct structural motifs of surrounding LTRs permit the high quality computational prediction of CERVs. We conducted a *de novo* annotation of CERVs in the canine genome via a two-step process that identified the canonical motifs then coding sequence structures of LTR-RT elements (Ellinghaus et al., 2008; Steinbiss et al., 2009). Our methods yielded 477 CERV candidates, of which 392 were novel compared to 805 previously published CERV candidates identified by degenerate PCR with *pro/pol* primers and varied computational analyses (Tarlinton et al., 2013; Jo et al., 2012; Barrio et al., 2011) (see Supplementary Materials 3). Our candidates represent a complementary set from the newest genome build, and the differences in the observed candidates from the literature-curated set stem from the use of different technologies and filtration conditions.

Our candidates covered approximately 0.2% of the genome (Fig. 1) and were significantly larger (mean ± SD = 10.2 ± 1.4 kb) than the aggregated literature-curated set (5.0 ± 4.8 kb; two-tailed *t*-test $p < 0.0001$). This reflects our bias for larger, intact CERVs due to our filtration conditions for CERVs between 3.5 and 16 kb in size, and is similar to the 9.19 kb and 8.427 kb mean sizes of CERV-β elements found by Barrio et al. and Jo et al. respectively (Tarlinton et al., 2013; Jo et al., 2012; Barrio et al., 2011). Our candidates presumably underrepresent smaller Spuma-like and γ family CERVs.

Overlapping candidates came from the works of Barrio et al. (2011) (51/85) and Jo et al. (2012) (34/85) and had similar size profiles (shared mean ± SD = 6.9 ± 3.0 kb and 6.9 ± 3.1 kb for our and literature-curated shared CERVs respectively). 48 overlapping candidates had methylation data, of which 35 were considered hyper-methylated (~73%), 12 normally methylated (25%), and 1 hypo-methylated (2%). This suggests that shared candidates may be enriched with similar retroviral motifs that are targeted by the silencing machinery.

2.2. Genomic neighborhood

We calculated the CERV-derived proportion for each chromosome, and the X chromosome emerged as a transposon sink containing the most annotated CERVs ($n_{Xchr} = 61/477$, 13%) and the highest CERV composition (0.5%; genome mean ± SD = 0.20 ± 0.09%) of any single chromosome. Canine chromosome 32 also contained a significantly high proportion of CERVs (0.42%; Fig. 2). Of the 477 annotated CERVs, 27.7% were genic (n = 132), 4.8% putatively regulatory ($n_{promoter} = 12$; $n_{downstream} = 11$), and 67.5% inter-genic (n = 322). 76/416 (18.2%) autosomal CERVs localized within pericentomeric regions corresponding to the first 5 Mb of their respective chromosome. 12/61 (19.7%) X chromosomal CERVs lied within the X's metacentric centromere, defined as the region from 40 Mb–55 Mb (Lindblad-toh et al., 2005).

2.3. Single nucleotide resolution methylation analysis

Dogs and wolves differed slightly in their aggregated genome-wide CpG methylation frequency profiles across both the autosomes (dog: 66.6%, wolf: 65.5%) and X-chromosome, with higher levels of CpG methylation on the X chromosome (dog: 70.6%, wolf: 69.2%, intra-species autosomal vs. X chromosomal two-tailed *t*-tests $p < 0.0001$), conforming to previous reports that CpGs in mammal genomes are highly methylated (e.g. 70%) (Robertson and Jones, 1999; Bird, 2002). Our CERV candidates showed significantly higher levels of CpG methylation compared to their corresponding genome-wide datasets (averages = 72.7–84.0%; all intra-species two-tailed *t*-tests $p < 0.0001$; see Supplementary Materials 1). Furthermore, CpGs in CERVs had marginally reduced levels

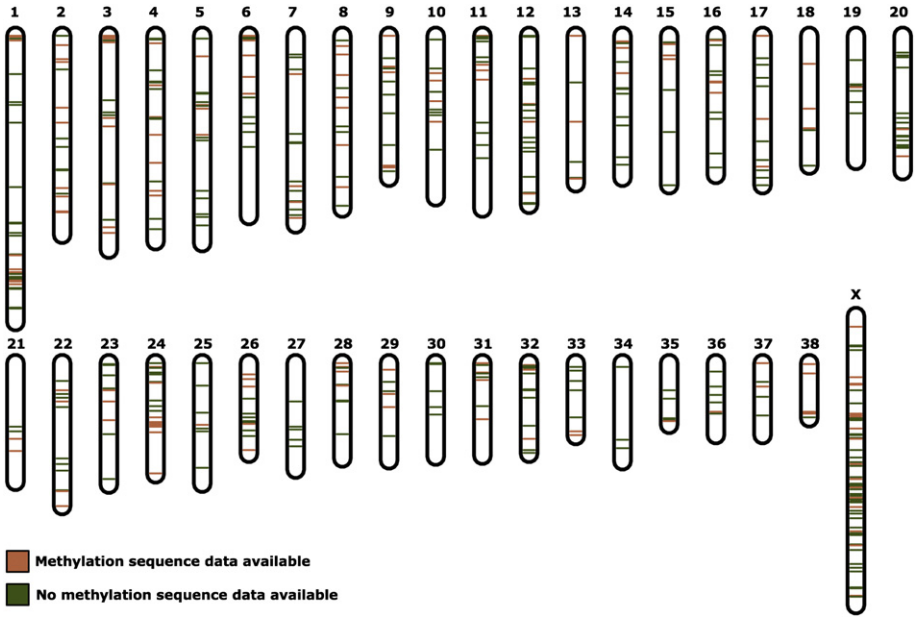

Fig. 1. Chromosomal locations and methylation data of Boxer CERV candidates. Genome coordinates were reduced 400,000-fold, adjusted for minimal overlap, and plotted against a scaled genome.