Conserved Gene Order at the Nuclear Periphery in Drosophila

José M. Ranz,*1 Carlos Díaz-Castillo,1 and Rita Petersen2
1Department of Ecology and Evolutionary Biology, University of California, Irvine
2Department of Statistics, University of California, Irvine
*Corresponding author: E-mail: jranz@uci.edu.
Associate editor: John Parsch

Abstract

Whether higher-order chromatin organization is related to genome stability over evolutionary time remains elusive. We find that regions of conserved gene order across the genus Drosophila are larger if they harbor genes bound by B-type lamin (Lam) and Suppressor of Under-Replication (SUUR), two proteins located at the nuclear periphery. Low recombination rates and coexpression of genes in regions of conserved gene order do not explain the lower probability of disruption in these regions by genome rearrangements. Instead, we find a significant colocalization between evolutionarily stable genomic regions associated with Lam and sequences thought to regulate local gene expression, which have the potential to impose constraints on genome rearrangement. At least in the genus Drosophila, localization of particular genomic regions at the nuclear periphery is intimately associated with their long-term integrity during evolution.

Key words: genome stability, nuclear periphery, Drosophila, lamin, Nup98, SUUR.

Analysis of genome-wide binding maps of 30 chromatin-associated proteins in Drosophila melanogaster revealed that its genome is organized into functional chromatin domains (de Wit et al. 2008). Remarkably, chromatin domains associated with the nuclear periphery via B-type lamin (Lam) and Suppressor of Under-Replication (SUUR) were two of the three with the lowest probability of being disrupted by the rearrangement breakpoints that have differentiated the chromosomes of D. melanogaster and D. pseudoobscura, two species that shared a common ancestor 55 Ma (Tamura et al. 2004). The molecular underpinnings of the relationship between Lam and SUUR chromatin domains and evolutionary stability, and whether this relationship extends across a wider evolutionary timescale, are unknown.

Lam and SUUR often bind to the same genes at the nuclear periphery (Pickersgill et al. 2006; Pindyurin et al. 2007). Lam is one of the two types of lamins (Munoz-Alarcon et al. 2007) that forms the nuclear lamina, which is associated with the inner nuclear membrane of metazoan cells (Akhtar and Gasser 2007). SUUR locally modulates replication in Drosophila chromosomes (Pindyurin et al. 2007). Dysfunctional lamin genes have been shown to be associated with impaired DNA repair activity, increased number of chromosomal aberrations, and altered nuclear morphology (Liu et al. 2005; Brandt et al. 2008). Thus, one explanation for the reported trend of conservation of domains associated with Lam and SUUR could be that their interaction with the nuclear periphery somehow decreases their exposure to break-prone molecular environments. Alternatively, the presence of genes that require complex regulation may impose a constraint on the occurrence of chromosomal rearrangements that affect the evolutionary stability of these genomic regions (Kikuta et al. 2007; de Wit et al. 2008).

A recent comparison of nine Drosophila species that accumulate ~380 My divergence indicates that the fruit fly genome has undergone at least 2,688 disruptions of gene order (von Grotthuss et al. 2010). This figure represents a ~3-fold increase in relation to the number of breakages previously examined (de Wit et al. 2008). In light of this more representative picture of gene rearrangement in the genus Drosophila, we investigated the relationship between conservation of gene organization and its association with the nuclear periphery at the level of the whole genus Drosophila and subsequently the role played by the different mechanisms that might explain this relationship.

We first mapped Lam and SUUR-associated genes, as identified in D. melanogaster Kc cells, to chromosome regions referred to as orthologous landmarks (OLs), within which gene order is unaltered across nine Drosophila species (von Grotthuss et al. 2010). Approximately 93% (430/463) of all Lam target genes and 80% (1,840/2,299) of all SUUR target genes were mapped to 240 and 896 OLs, respectively (supplementary data set S1 and text, Supplementary Material online). The size of OLs is measured in a unit termed independent gene anchor, which ignores gene size and collapses physically related genes such as antisense overlapping genes, facilitating the comparison across distantly related species. If the presence of Lam and SUUR targets correlates with higher genome stability, OLs harboring these targets should have a larger size than expected by chance. Table 1 shows the observed size of OLs containing at least one target gene for the Lam and SUUR proteins as compared with OLs containing no target. Monte Carlo simulations indicated that it is very unlikely to find OLs containing targets for either protein with an average size equal to or higher than that observed. Therefore, the size of genomic regions with conserved gene organization across the Drosophila genus is heavily associated with the presence of genes bound by Lam and SUUR.
We sought to determine whether this pattern is also shown by other proteins located at the nuclear periphery. Nucleoporin 98 (Nup98) is one of the components of the nuclear pore complex at the nuclear envelope (Kalverda et al. 2010). Nup98 has been shown to have a very different chromatin-binding profile to that of Lam at the nuclear periphery. Importantly, unlike the nuclear lamina (Pickersgill et al. 2006), the nuclear pore complex is more closely associated with active chromatin (Vaquerizas et al. 2010), which can result in a differential exposure to break-prone environments (Lin et al. 2009). We distinguished between nuclear periphery–nucleoplasmic–chromatin interactions, which have different targets, and focused on the former. Approximately 86% (344/400) of all Nup98 target genes were mapped to 295 OLs. The size of OLs containing Nup98 targets genes was not significantly different from that expected by chance (table 1).

If the presence of Lam and SUUR targets indeed correlates with higher-genome stability, and Nup98 targets do not, then this pattern should be especially evident for the 1% largest OLs, that is, 22 OLs termed ultraconserved regions (UCRs; von Grothuss et al. 2010). Taking into account how targets and nontargets of the three proteins are distributed across UCRs and the rest of the genome, we found that the fraction of the genome represented by UCRs indeed contains 187% and 74% more Lam and SUUR targets, respectively, and 25% fewer Nup98 targets, than expected by chance (table 1). This result substantiates the tendency of Lam and SUUR targets to aggregate in OLs of relatively large size and therefore to be underrepresented in OLs of small size, a pattern that contrasts with that of Nup98 targets (supplementary fig. S1, Supplementary Material online). The pattern observed for Nup98 does not result from the lower number of targets analyzed as compared with that for Lam and SUUR (table 1). Therefore, despite the limited number of proteins associated with the nuclear periphery for which this type of data exists, only some components of the nuclear periphery seem to be tightly associated with the pattern of genome stability over evolutionary time, which is prominently observed in the UCRs of the Drosophila genome.

We examined different factors that might have contributed to the conservation of Lam chromatin domains over evolutionary time. High recombination rates have been found to correlate with the occurrence of chromosomal rearrangements in several metazoans (Poyatos and Hurst 2006; Volker et al. 2010; Weber and Hurst 2011). This pattern would result from nonallelic homologous recombination events (Stankiewicz and Lupski 2002), which should be less frequent in regions of low recombination. In fact, recombination rates seem to be negatively correlated with the size of OLs in Drosophila (Weber and Hurst 2011). Using a less restrictive definition of the maintenance of gene organization than that used by Weber and Hurst (2011), and the available estimates of recombination rates (Larracuente et al. 2008), we found the same trend (Spearman’s $r = -0.0410, P = 0.0477; n = 2,328$). Then, we tested for differences in the median recombination rates between OLs with and without Lam targets and found no evidence that the former have lower recombination rates ($1.097 \ VS. 1.21, P = 0.4359; Mann–Whitney U$). Subsequently, we examined whether UCRs harboring Lam targets are located in chromosomal regions exposed to low recombination rates and found only three, all located close to centromeric regions (supplementary fig. S2 and text, Supplementary Material online). In contrast, nine UCRs were found to be exposed to high median recombination rates, four of them harboring Lam targets. Therefore, it does not seem that a more restricted accessibility of the recombination machinery to particular regions of the Drosophila genome associated with the nuclear periphery can account for the evolutionary stability exhibited by many of the largest OLs.

Table 1. Salient Features of Genomic Distribution of Putative Targets for Several Chromatin-Binding Proteins in Relation to the OLs That Make Up the Drosophila Genome.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Observed With No Target</th>
<th>Expected With ≥1 Target</th>
<th>Distribution</th>
<th>Targets Mapped (OLs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>With No Target</td>
<td>With ≥1 Target</td>
<td>With ≥1 Target</td>
<td>$P$ (expected ≥ observed)</td>
</tr>
<tr>
<td>Lam</td>
<td>3.2 ± 2.9/2</td>
<td>10 ± 7.9/8</td>
<td>6.45 ± 0.22/4.55</td>
<td>&lt;0.00001</td>
</tr>
<tr>
<td>Nup98</td>
<td>3.1 ± 3.2/2</td>
<td>6.54 ± 6.3/5</td>
<td>6.58 ± 0.26/4.66</td>
<td>0.57</td>
</tr>
<tr>
<td>SUUR</td>
<td>2.33 ± 1.7/8</td>
<td>5.92 ± 5.28/5</td>
<td>5.06 ± 0.57/3.72</td>
<td>&lt;0.00001</td>
</tr>
</tbody>
</table>

NOTE.—IGA, independent gene anchor.  
<sup>a</sup> As identified by DNA adenine methyltransferase identification technology in Kc cells.  
<sup>b</sup> Mean ± SD/median number of IGAs.  
<sup>c</sup> Using the average size of OLs with ≥1 target according to Monte Carlo simulations; identical results were obtained for the median size and when genes, instead of IGAs, were used as a proxy for the size of OLs (data not shown).  
<sup>d</sup> Two-tailed Fisher’s exact test.  
<sup>e</sup> If the distribution of Nup98 targets mirrored the proportion of Lam targets in UCRs and in the rest of the genome, the result would have been statistically significant ($P = 2.27 \times 10^{-13}$, supplementary data set S1 and text, Supplementary Material online).
Table 2. Phylogenetic Status of Lam Clusters as Defined in Drosophila melanogaster across the Genus Drosophila for Two Expression Profile Features.

<table>
<thead>
<tr>
<th>Phylogenetic Statusa</th>
<th>Life Cycle Expressionb</th>
<th>Testis Expressionc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r \leq 0.5$</td>
<td>$r &gt; 0.5$</td>
</tr>
<tr>
<td>Disrupted</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Nondisrupted</td>
<td>18</td>
<td>19</td>
</tr>
</tbody>
</table>

Note.—TB, testis biased.

a Upon comparing the cluster organization in eight other Drosophila species (von Grothehus et al. 2010).
b Pickersgill et al. (2006); c intraclass expression Pearson's correlation coefficient across six developmental stages.

Pickersgill et al. (2006), HCNEs also tend to cluster in the genome and, more importantly, both genomic features tend to colocalize. Since this colocalization does not result from the association of genomic regions of low gene density with the nuclear periphery (supplementary text, Supplementary Material online), the notion that it could represent a type of functional constraint that has contributed to the evolutionary stability of the OLs in which both genomic features are embedded is reinforced.

We have documented a genus-wide positive correlation between genome stability, using the size of collinear blocks as a proxy, and intranuclear localization in D. melanogaster, specifically the nuclear periphery. Both limited exposure to break-prone molecular environments and selection for particular gene arrangements might account for genome stability. We did not find differences in recombination rates between OLs with and without Lam targets and only three out of 16 UCRs harboring Lam targets are exposed to low recombination environments, at least in D. melanogaster. Therefore, although local recombination rates can play a role (Weber and Hurst 2011), they do not account for the overall larger size of OLs localized in the periphery, which points to the action of other mechanisms. Further, coexpression of clustered Lam targets was not found to correlate with phylogenetic conservation in good agreement with previous observations (Liao and Zhang 2008; Weber and Hurst 2011). This pattern would conform with the notion of transcriptional interference between neighboring coexpressed genes (Liao and Zhang 2008) or, alternatively, with the possibility that much of this coexpression reflects colocalization in the same chromatin domain (Weber and Hurst 2011). Nevertheless, the strong association found between OLs harboring Lam targets and HCNEs is suggestive of the presence of some kind of functional constraints (Kikuta et al. 2007). This association is speculatively preeminent in UCRs, including four that are exposed to high recombination rates and therefore to potentially break-prone molecular environments. We cannot discount though that the aggregation of peaks of HCNEs in UCRs has not been facilitated by their evolutionary stability.

Table 3. Relationship between Genomic Distributions of Lam Targets and HCNEs.

<table>
<thead>
<tr>
<th>OL Typea</th>
<th>Number of OLs (UCRs)</th>
<th>Number of HCNEs (in UCRs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed</td>
<td>Expectedb</td>
</tr>
<tr>
<td>Lam target and HCNE</td>
<td>62** (11**)</td>
<td>20–48 (1–10)</td>
</tr>
<tr>
<td>Lam target only</td>
<td>178** (5)</td>
<td>325–365 (4–16)</td>
</tr>
<tr>
<td>HCNE only</td>
<td>61** (2)</td>
<td>94–120 (0–6)</td>
</tr>
</tbody>
</table>

a Based on the resident genomic features.
b 0.5th–99.5th percentiles of the distribution obtained using Monte Carlo simulations (supplementary text, Supplementary Material online).

**P < 0.01.
the intimate association among conservation of gene organization, localization at the nuclear periphery, and presence of HCNEs in the genus *Drosophila* holds in other taxa remains to be elucidated.

**Supplementary Material**

Supplementary text, data sets S1 and S2, and figures S1–S3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We thank A. Pindyurin and A. Larracuente for providing data and members of the Ranz laboratory for comments. This work was supported by start-up funds and a National Science Foundation grant (DEB-0949365) to J. R.

**References**


