Did circular DNA shape the evolution of mammalian genomes?

Holt, Sylvester; Arrey, Gerard; Regenberg, Birgitte

Published in:
Trends in Biochemical Sciences

DOI:
10.1016/j.tibs.2022.09.010

Publication date:
2023

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
Did circular DNA shape the evolution of mammalian genomes?

Sylvester Holt,1 Gerard Arrey,1 and Birgitte Regenberg 2,1,*

Extrachromosomal circular DNA (eccDNA) can shape the genomes of somatic cells, but how it impacts genomes across generations is largely unexplored. We propose that genomes can rearrange via circular intermediates across generations and show that up to 6% of a mammalian genome can have changed gene order through eccDNA.

Work over the past decade has demonstrated how DNA from chromosomes can circularize and form eccDNA [1–3]. EccDNA is abundant in the somatic cells of a wide range of eukaryotic organisms, in sizes large enough to carry whole genes, and genes expressed from eccDNA can have direct phenotypic consequences (reviewed in [4]). The phenotypic effects of eccDNA are often transient because they lack the centromeres that guarantee their faithful segregation and any selective advantage of eccDNA is therefore only fixed if they reinsert into chromosomes [1,5].

Whether or not eccDNA has permanently shaped mammalian genomes is largely unexplored. We and others have recently demonstrated the existence of eccDNA in mammalian germ cells [6–8]. This is notable because these eccDNA could potentially reinsert into the germline chromosomes and make stable alterations that are inherited across generations.

We propose the circle transposition theory, where inheritable chromosomal rearrangements can happen through a circular DNA intermediate formed from one part of a genome that reinserts into another part of the genome (Figure 1A). Several mammalian species offer reference genomes in which most the genomic regions are well-annotated with unambiguous gene orders. We explored mammalian genome maps to search for signatures that correspond to transposition of circular intermediates. We also discuss additional studies that support eccDNA transposition and finally review alternative theories.

Circle transposition theory

To test whether circle insertions could have affected genome evolution, we screened for changes in gene order (synteny) that suggest germline insertions of eccDNA (Figure 1), extracting coordinates for synteny blocks with a 150-kb window using the SynBuilder tool at the Synteny Portal [9] for eight different representative mammalian species using human as a reference. From the synteny coordinates, we searched for circle insertion signatures that consisted of two flanking inversions or two flanking blocks in the same direction with a change in order using R code (Figure 1 and see Table S1 and Figure S1 in the supplemental information online).

In total, we saw 95 circle insertion signatures in the eight genomes (Figure 2A). If the patterns were observed in multiple species, we inferred their position in the phylogeny and the ancestral branch from which they originated. For example, events 15, 20, 22, 44, 50, 54, 66, 77, and 88 were found both in the cow and the sheep genomes but not in the dog genome, suggesting that these events took place in the last common ancestor of cow and sheep, while events 1 and 81 were unique for sheep (Figure 2B). Because all eight species were mapped to the human reference genome, we expected mirrored signatures in accordance with the position in the phylogenetic tree. For example, if the circle insertion pattern was unique for human, we expected all the other species to have the pattern mirrored in the maps, unless the pattern was degraded through other events. We did not find any human-specific events, since none of the signatures were mirrored in all eight species. This is likely because requiring the synteny to be identical in eight species is strict, given the evolutionary distance of ~159 million years. However, some patterns were still observed in many species, suggesting primate-specific patterns. We deduced that three events were specific for human and chimpanzee (Figure 2B, 34, 66, and 75), because no circle insertion signature was found in chimpanzee, indicating that human and chimpanzee were identical in the genome (Figure 2B). Similarly, we inferred that three events were specific for all primates as the signature was not found neither in chimpanzee nor in rhesus monkey genomes (Figure 2B, 49, 64, and 67).

The signatures shown in Figure 2 reveal the possibility that parts of all tested mammalian genomes could have evolved through transpositions that are consistent with circular intermediates. Ungulates (cow and sheep) appear to have undergone the largest transformation of their genomes, with 6% being altered through circular intermediates (Figure 2A and Table S1 in the supplemental information online). We next annotated genes in the circle signatures and confirmed that orthologous open reading frames overlapped with respect to position, orientation, and function between the synteny blocks (exemplified by three of the largest signatures in ungulates, Figures S3–S5 in the supplemental information online).

Evidence for circle transposition

To support the circle transposition theory, we would expect: (i) that the germline contains circular DNA intermediates, which (ii) occasionally reinsert into chromosomes, and (iii) are subsequently carried by individuals that inbreed, to ensure homozygous offspring. We recently identified...
thousands of eccDNA from human sperm cells of 29 different donors [8], supporting that eccDNA is common in human germ-line cells. These results confirm the previous identification of eccDNA in a pooled sperm sample with an unknown number of human donors [7] and in mouse and wild boar sperm cells [3,8,10]. How circles form in the germline is less well understood. However, indirect evidence suggest that circularization can happen through nonhomologous end-joining of linear fragments [8] and homologous recombination within a chromosome [11].

Transposable elements (TEs) that involve a circular DNA intermediate are known to integrate back to the chromosome, shaping eukaryotic genomes [12]. TEs move around the genome aided by a self-encoded specialized machinery depending on their class, which may include different enzymes: retrotranscriptases, integrases, and endonucleases, as well as helicases and transposases. While moving around, they can change gene expression, gene order, and create recombination hotspots in the loci where they integrate [12].

Similarly, but not involving any transposition machinery, eccDNA transpositions might also allow other types of DNA to translocate and contribute to similar changes. A few examples are known where both the circle and the potential insertion product are found. Transposition of genes on circular elements are known from unicellular eukaryotic yeast, where circularization, amplification, and insertions back into the original locus serve as a mode of environmental adaptation [5]. Yeast cells carry two glucose transporter genes, HXT6 and HXT7, in direct repeat on chromosome IV that often forms a self-replicating circular intermediate, [HXT6/7cyc]. When exposed to glucose limitation, the fraction of cells carrying the [HXT6/7cyc] increases for several generations until a stable chromosomal HXT6 HXT6/7 HXT7 amplification occurs and outcompetes the wild type HXT6 HXT7 genotype and the [HXT6/7cyc] amplification [5]. This result supports the hypothesis that selective advantages created by the eccDNA intermediate can be fixed in a population through stable insertions.

Signatures of circle transpositions in the germ line have also been identified in fish and mammals. For example, Durkin et al. described how the color sidedness on three cow varieties is associated with syntenic changes that can be linked to circle insertions [13]. The Belgian Blue cow has a 492-kb eccDNA insertion in
Figure 2. Circle insertion signatures in eight representative mammals. (A) Circle insertion signatures for eight mammals, projected onto human genome coordinates. (B) Phylogenetic tree indicating shared and unique circle insertion signatures. The tree is read like a mirror from the human branch (arrow) towards the branches for the remaining organisms. The tree was generated with TimeTree and colors indicate the organisms that have the signatures, shown in (A). The signature numbers shown in panel (A) and (B) are the same. Star indicate that the signatures may be degenerated in a few organisms. See supplemental information online for details. Signatures 22, 56, 77, 66, 67 are shown in Figures S3–S7 in the supplemental information online.
chromosome 29 spanning the tyrosine-protein kinase gene KIT gene from chromosome 6. The KIT gene affects the coat color of Belgian Blue and its descendant, Brown Swiss, where a portion of the gene has circularized again and a 575-kb fragment of mixed chromosome 6 and 29 are inserted back into chromosome 6 [13].

Alternative theories
Besides circle translocation, at least three alternative mechanisms could explain the eccDNA insertion patterns in Figure 1. These are random insertion of two flanking linear DNA fragments in the same location (Figure 1B), two inversions within an insertion (Figure 1C), and replication errors through microhomology-mediated break-induced repair (MMBIR, Figure 1D). The likelihood of two adjacent linear fragments inserting randomly in the same locus is different in an original and flanking each other with inverted direction is presumably low (Figure 1B). We tested whether the signatures observed in our study could be due to the random insertion of two linear fragments by randomizing the synteny blocks for all organisms 10 000 times and extracting circle insertion signatures. In all cases the number of identified signatures in randomized data was much lower than observed in the unscrambled genomes (Figure S2 in the supplemental information online). Similarly, insertion and two inversions requires three mutational events rather than the two events in a circle translocation and therefore also seems less likely (Figure 1C). Finally, the observed eccDNA insertion patterns could be caused by stalled DNA replication forks that are repaired through successive replicative strand replacements in areas with microhomology through the MMBIR (Figure 1D). This mechanism is used to explain complex rearrangements in congenital disorders [14,15], though it would require a sequence of very precise strand replacements in multiple locations in different chromosomes to explain the signatures in Figure 2. Thus, all four mechanisms can potentially contribute to the patterns in Figure 1, though circle insertion appears more parsimonious than MMBIR, serial insertions, and inversions.

Concluding remarks
Mammalian genomes are full of genome rearrangements that likely shape the expression and organization of genes. We have proposed the circle transposition theory that can explain many of the changes in gene order found between mammalian genomes. The theory can also be applied to test the role of human eccDNA in human congenital disorders as well as other biological classes, phyla, and kingdoms.

Acknowledgments
This work is supported by the VILLUM Foundation (00029247 to B.R.). G.A. also received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801199.

Declaration of interests
No interests are declared.

Supplemental Information
Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.tibs.2022.09.010.

References

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).