

NEWS AND COMMENTARY

Back in business

‘Satellite DNA transcripts have diverse biological roles in *Drosophila*’

GCS Kuhn

Heredity (2015) 115, 1–2; doi:10.1038/hdy.2015.12; published online 25 March 2015

It has been known for several decades that a large fraction (>50%) of most eukaryotic genomes corresponds to repetitive DNA sequences, mainly represented by dispersed transposable elements (TEs) and tandemly repeated satellite DNAs (satDNAs). Both classes have been traditionally included into the non-coding fraction of the genome, because they were considered unlikely to encode any protein product useful for the cell. They have been often referred to as ‘selfish DNA’, ‘parasitic DNA’ or ‘junk DNA’, terms usually applied to DNA sequences that spread in the genome by the multiplication of copies that conferred neither advantage nor disadvantage to the fitness of organisms (Orgel and Crick, 1980). Their abundance and ubiquitous presence in eukaryotes was traditionally explained by their ability to amplify (intragenomic selection) and as a result of genomic tolerance for such extra and useless genetic material. There is growing evidence showing that TEs may have important functional roles in a genome, participating in gene regulation, chromatin modulation or as functional components of important chromosome structures such as telomeres and centromeres (Kidwell and Lisch, 2001). Nevertheless, our knowledge about the function of satDNAs is far less extensive and understood when compared with TEs.

SatDNAs consist of large and homogeneous arrays of tandem repeats located in the heterochromatic regions of the chromosomes. Sometimes, members of the same satDNA family can be found dispersed throughout euchromatic regions, but in the

form of small arrays (approximately five tandem repeats; Kuhn *et al.*, 2012). SatDNAs do not code proteins and have been traditionally viewed as ‘monotonous and useless material, able to accumulate until they become a too heavy load for a genome’ (Pohl *et al.*, 2012). Moreover, the abundance and high similarity of repeats have long been a challenge for algorithms used in the assembly of whole-genome shotgun sequences. Putting it all together, satDNAs are considered as a poorly understood and neglected genomic component (Pohl *et al.*, 2012). In fact, satDNAs have been barely mentioned in most papers reporting whole-sequenced genomes, despite their obvious high representativeness in the genome of many organisms (>30%) and transcriptional activity (Pezer *et al.*, 2011). Although transcription alone does not directly imply function (Graur *et al.*, 2013), the biological status of such transcripts remained unclear until very recently.

Herein, I would like to highlight recent and important discoveries concerning the biological utility of transcripts derived from the most abundant and most studied satDNA of *Drosophila melanogaster*, named by different authors as 1.688 (this number reflects its g ml^{-1} value in a cesium chloride density gradient), 359bp satellite or satDNA III. Repeats of this satellite family have been previously shown to be located in the centromeric and pericentromeric heterochromatin of the X chromosome, and in the pericentromeric heterochromatin of chromosomes 2 and 3. Moreover, several short arrays made of approximately five tandem repeats have been found widespread at several euchromatic regions, mainly on the X chromosome and to a lesser extent on chromosomes 2 and 3. Several 1.688 arrays underwent different homogenization for

chromosome-specific or array-specific repeat variants (that is, satDNA subfamilies), as a result of intragenomic concerted evolution. Consequently, 1.688 arrays contribute to provide individual structural identities to chromosomes (see Kuhn *et al.*, 2012; Gallach, 2014; and references therein).

Usakin *et al.* (2007) were the first to show that the repeats from the 1.688 satDNA family are transcribed in embryos and adult flies. They found that 1.688 double-stranded RNAs derived from subfamilies located on chromosomes 2 and 3 are processed into siRNAs (small interfering RNAs) that in turn participate in the heterochromatin formation of both chromosomes. The authors further noted that these siRNAs do not act on the X chromosome, where a different 1.688 subfamily is present. The utility of 1.688 transcripts from chromosome X, where the bulk of 1.688 repeats are found, remained obscured. Two exciting works published in November 2014 shed new light on the function of the X-linked satDNA transcripts and, to a wider extent, on the diversity of functional roles that satRNAs may play (Menon *et al.*, 2014; Rošić *et al.*, 2014).

Rošić *et al.* (2014) showed that 1.688 heterochromatic repeats residing on the X chromosome produce long sense and anti-sense polyadenylated RNAs comprising approximately four repeats. Depletion of these transcripts in cell culture by RNA knockdown leads to chromosome segregation defects, including the presence of lagging chromosomes in anaphase. Interestingly, the fact that chromosomes 2 and 3 showed similar mitotic defects suggests that satRNAs from the X chromosome have also the ability to act *in trans*. The authors observed the same phenotype in early embryos carrying the *Zygotic hybrid rescue* (*Zhr¹*) mutation,

Professor GCS Kuhn is at Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
E-mail: gcskuhn@ufmg.br

characterized by a X–Y translocation where most of the 1.688 heterochromatic block from chromosome X has been deleted (Sawamura *et al.*, 1993). RNA-immunoprecipitation experiments revealed that 1.688 satRNAs specifically binds to CENP-C, a centromeric protein that together with CENP-A has a key role in kinetochore assembly and function during cell division. SatRNA knockdown led to a significant reduction of CENP-C in the centromeric regions during mitosis. The authors concluded that an interaction between CENP-C and satRNA is required for proper localization of CENP-C (and consequently CENP-A) to centromeres. Disruption of this interaction may compromise the correct assembly of a functional centromere that in turn leads to genome instability (that is, segregation defects). However, it is important to mention that *Zhr*¹ flies may be surprisingly viable and fertile, despite the mitotic defects observed during the embryonic stage. The authors suggested that, in addition to satRNAs, other mechanisms are likely involved for proper chromosome segregation.

Interactions between satRNAs and centromeric proteins have also been reported in maize and humans. Similarly, inhibition of these satRNAs led to chromosomal segregation defects (Topp *et al.*, 2004; Chan *et al.*, 2012). The fact that satDNA transcripts contribute for centromeric function in flies, humans and plants suggests that they have an evolutionary conserved role, as rightly suggested by Rošić *et al.* (2014). I expect that more examples showing such association will become available in a wide range of species in the near future.

In a different approach, Menon *et al.* (2014) showed that ‘1.688’ satDNA transcripts have a third biological role in *D. melanogaster*. In *Drosophila*, genes on the male X chromosome are doubly expressed, a situation that leads to the balance of gene products between males (XY) and females (XX), a mechanism known as dosage compensation. In this mechanism, the male specific lethal (MSL) complex binds to specific sites on the male X chromosome and promotes an approximately twofold increase in gene expression. Previous work revealed the participation of siRNAs in this process, but their parental sequences remained unknown. Now, the authors showed that siRNAs produced from a subset of X-linked euchromatic 1.688 arrays contribute for the

localization of the MSL complex to the X chromosome. However, instead of interacting directly with MSL, it was suggested these siRNAs act in *cis* creating an X-specific chromatin environment that in turn allows association with MSL.

The data above show that satDNA transcripts may be involved in at least three important biological functions as follows: (i) centromere function; (ii) chromatin silencing/heterochromatin formation and (iii) chromatin modulation and global up regulation of X-linked genes. Considering functional role on one hand, and the fast rate of satDNA turnover in the other (for example, Zhang *et al.*, 2014; Gallach, 2014), it is possible that incompatibilities between satRNAs from one species and chromatin remodeling proteins or chromosome-specific targets (as those involved in centromere function or dosage compensation) from the other may cause genomic instabilities in hybrids, thus contributing for the speciation process (Henikoff *et al.*, 2001; Ferree and Barbash, 2009; Barbash 2010; Brown and O’Neill, 2010).

More than one satDNA family is usually present in the genome. In *D. melanogaster*, for example, 16 satDNAs have been described. It is likely that not all satDNAs in a species are functional. Some of them may indeed be considered junk. But the observation that at least a subset of satDNA families are involved in important biological roles makes the study of these highly abundant and fast-evolving components of the eukaryote genome highly relevant for structural, functional and evolutionary genomics. It is also worth mentioning that in the past few years, new and efficient bioinformatic tools are becoming available for the identification of satDNAs from sequenced genomes (Novák *et al.*, 2014), whereas long-template sequencing and new computational approaches (Altemose *et al.*, 2014) have been fostering the assembly of satDNA repeats. With all these new findings and tools, genomic studies, including those related to whole-sequenced genomes, will certainly benefit by a ‘satDNA recall’.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGEMENTS

I thank Dr Marta Svartman and two anonymous reviewers for the valuable comments made on the first version of the manuscript.

- Altemose N, Miga KH, Maggioni M, Willard HF (2014). Genomic characterization of large heterochromatic gaps in the human genome assembly. *PLoS Comput Biol* **10**: e1003628.
- Barbash DA (2010). Genetic testing of the hypothesis that hybrid male lethality results from a failure in dosage compensation. *Genetics* **184**: 313–316.
- Brown JD, O’Neill RJ. (2010). Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. *Annu Rev Genomics Hum Genet* **11**: 291–316.
- Chan FL, Marshall OJ, Saffery R, Kim BW, Earle E, Choo KH *et al.* (2012). Active transcription and essential role of RNA polymerase II at the centromere during mitosis. *Proc Natl Acad Sci USA* **109**: 1979–1984.
- Ferree PM, Barbash DA (2009). Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. *PLoS Biol* **7**: e1000234.
- Gallach M (2014). Recurrent turnover of chromosome-specific satellites in *Drosophila*. *Genome Biol Evol* **6**: 1279–1286.
- Graur D, Zheng Y, Price N, Azevedo RB, Zufall RA, Elhaik E *et al.* (2013). On the immortality of television sets: “function” in the human genome according to the evolution-free gospel of ENCODE. *Genome Biol Evol* **5**: 578–590.
- Henikoff S, Ahmad K, Malik HS (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**: 1098–1102.
- Kidwell MG, Lisch DR (2001). Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution* **55**: 1–24.
- Kuhn GCS, Küttler H, Moreira-Filho O, Heslop-Harrison JS (2012). The 1.688 repetitive DNA of *Drosophila*: concerted evolution at different genomic scales and association with genes. *Mol Biol Evol* **29**: 7–11.
- Menon DU, Coarfa C, Xiao W, Gunaratne PH, Meller VH (2014). siRNAs from an X-linked satellite repeat promote X-chromosome recognition in *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **111**: 16460–16465.
- Novák P, Hříbová E, Neumann P, Koblížková A, Doležel J, Macas J *et al.* (2014). Genome-wide analysis of repeat diversity across the family Musaceae. *PLoS One* **9**: e98918.
- Orgel LE, Crick FHC (1980). Selfish DNA: the ultimate parasite. *Nature* **284**: 604–607.
- Pezer Z, Brajković J, Feliciello I, Ugarković D (2011). Transcription of satellite DNAs in insects. *Prog Mol Subcell Biol* **51**: 161–178.
- Plohl M, Mestrovic N, Mravinac B (2012). Satellite DNA evolution. *Genome Dynamic* **7**: 126–152.
- Rošić S, Köhler F, Erhardt S (2014). Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division. *J Cell Biol* **207**: 335–349.
- Sawamura K, Yamamoto MT, Watanabe TK (1993). Hybrid lethal systems in the *Drosophila melanogaster* species complex. II. The Zygotic hybrid rescue (*Zhr*) gene of *D. melanogaster*. *Genetics* **133**: 307–313.
- Topp CN, Zhong CX, Dawe RK (2004). Centromere-encoded RNAs are integral components of the maize kinetochore. *Proc Natl Acad Sci USA* **101**: 15986–15991.
- Usakin L, Abad J, Vagin VV, De Pablos B, Villasante A, Gvozdev VA *et al.* (2007). Transcription of the 1.688 satellite DNA family is under the control of RNA interference machinery in *Drosophila melanogaster* ovaries. *Genetics* **176**: 1343–1349.
- Zhang H, Koblížková A, Wang K, Gong Z, Oliveira L, Torres GA *et al.* (2014). Boom-bust turnovers of megabase-sized centromeric DNA in *Solanum* species: rapid evolution of DNA sequences associated with centromeres. *Plant Cell* **26**: 1436–1447.