

RESEARCH HIGHLIGHT

An enigmatic satellite

William Amos*

See research article: <http://genomebiology.com/2011/12/4/R37>

Abstract

McLaughlin and Chadwick describe an X-linked tandem repeat that is transcribed, conserved and defies X inactivation. Is it selfish, functional or just an oddity?

Satellite DNA has always tended towards the enigmatic and the controversial. Originally named after ‘satellite’ bands that appeared when genomic DNA was separated by density gradient centrifugation, it was shown to comprise highly repeated DNA with unusually high or low GC content, hence its different buoyant densities. Early studies revealed unexpected patterns that came to be known as ‘concerted evolution’, where individual repeats within a tandem array appear to evolve cohesively rather than as independent units. A spectacular example occurs in whales, where a single 1.73-kb satellite sequence is present in about 100,000 copies in most species. However, in certain dolphins the overwhelming majority of repeats carry a 150-base deletion, a pattern that is clearly not the result of 100,000 independent deletion events [1]! In a recent paper in *Genome Biology*, McLaughlin and Chadwick [2] provide an exciting continuation of this story. They study a human ‘macro-satellite’, *DXZ4*, an X-linked 3-kb tandem repeat that might previously have been dismissed as repetitive junk. They show that it is anything but junk, having instead a remarkable range of properties that include being transcribed but probably not translated, being conserved across many primates and remaining active while almost every other gene around it is shut down on the inactivated X chromosome.

Classical satellite DNA includes both highly repeated sequences whose function is largely unclear and gene families such as the ribosomal RNA genes (rDNA). However, the tandemly repeated format inspired derivative names coined for shorter motifs. Thus, when Jeffreys and colleagues [3] discovered a class of highly unstable

tandem repeats capable of generating complicated, individual-specific banding patterns, they named them ‘minisatellites’, as the repeat unit was of the order of a few tens rather than thousands of bases long. The resulting technique, DNA fingerprinting, revolutionized the world of genetic markers and led to a new appreciation of how certain sequences can be highly mutable and highly recombinogenic. Later, it was realized that even shorter motifs, such as (AC)_n, were also highly mutable, providing a wealth of ‘microsatellite’ genetic markers that dominated genetic studies during the 1990s [4].

While a few tandem repeats comprise recognizable genes (for example, the rDNA), most do not, and it remains an important challenge to discover whether this majority are merely by-products of the genome’s tendency to various forms of slippage and sequence duplication or instead have some as yet undiscovered function(s). One class that certainly influences fitness comprises triplet repeats that occur within exons. Here, elongation of the repeat tract eventually disrupts or modifies the gene’s function, causing disease, classic examples being Huntington’s disease, myotonic dystrophy and fragile X [5]. However, triplet repeats in coding regions can also offer added evolutionary flexibility, a hypothesis advanced by Fondon and Garner [6]. They argued that domesticated dogs are too variable in form simply to be accounted for by new substitutions and that the higher mutation rate of slippage mutations might provide an enhanced source of variability. Remarkably, dogs do have unusually pure triplet repeats in a number of developmental genes, and this purity allows higher levels of slippage, which are then translated into inherited morphological flexibility. Over the longer term, such flexibility may give rise to evolutionary novelty, such as the triplet expansion that causes digital deformities and webbing in humans but which seems to have helped whales to evolve flippers [7].

In their recent paper, McLaughlin and Chadwick [2] move back up the size scale to explore ‘macro-satellites’, focusing on a locus called *DXZ4*. *DXZ4* is undoubtedly enigmatic, being located on the X chromosome and comprising a tandemly repeated array of a 3-kb monomer that contains several short open reading frames (ORFs), even though none has any homology to known proteins. Lying on the X chromosome, one might expect it to

*Correspondence: wa100@hermes.cam.ac.uk
Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK

participate in dosage compensation, which in mammals involves the random inactivation of one X chromosome in females. However, *DXZ4* seems to be one of the few regions on the X that bypasses this process because, in contrast to inactivated regions, it is hypo- rather than hypermethylated at CpG islands and lacks certain covalent changes to histones that are associated with the formation of heterochromatin.

McLaughlin and Chadwick [2] go on to investigate the evolutionary origins of *DXZ4* by studying homologous regions in other primates. They find high levels of conservation, ranging from 77% in New World monkeys up to 97% in great apes. Moreover, at least two of three microsatellites that exist within the macrosatellite are also conserved across primates, the third changing motif but remaining AT-rich. Perhaps more significantly, although *DXZ4* does not code for an obvious protein, a generally sharp fall-off in sequence conservation in the more distant primate branches that carry lemurs, galagos and tarsiers does not affect the ORFs, which seem largely conserved, suggesting function. Intriguingly, both sense and antisense transcripts can be detected, the latter being apparently specific to females.

A potentially key observation is the relationship between *DXZ4* and the multifunctional zinc-finger protein CCCTC-binding factor (CTCF). This protein has a wide range of reported roles that include both activation and repression of transcription and the separation of chromatin regions with activating and inactivating modifications. CTCF has also been implicated in the maintenance of monoallelic expression of imprinted genes in mice and humans and in regulating which members of a gene family are transcribed [8]. In this context, the fact that CTCF binds to *DXZ4* within the most conserved 400-bp region adds intriguing confirmation of an evolutionarily maintained function. Moreover, it is not merely the sequence that is conserved: using an immunoprecipitation approach, McLaughlin and Chadwick demonstrate that function is too. Thus, although their most recent common ancestor lived maybe 30 million years ago, macaques and humans both show a pattern of CTCF binding associated with euchromatin in females but not males, and by implication on the inactivated X [9].

Together, these observations raise many more questions than they answer. Why does *DXZ4* produce transcripts but no protein, while the best-studied autosomal macrosatellite, *D4Z4*, produces a protein and causes disease when repeat number declines? What is the role of the antisense RNA? Is its relative novelty, apparently being new to primates, due more to functional innovation or because *DXZ4* is no more than a transient hitchhiker? For me, perhaps the strongest parallels one might speculatively draw are with various aspects of intragenomic

conflict, such as meiotic drive. In *Drosophila*, the gene pair *Stellate* and *Suppressor of Stellate* has no obvious function beyond distorting the normal 50:50 ratio of sperm carrying X and Y chromosomes. Both genes are found as tandemly repeated arrays, and copy number correlates with strength of effect, with mild imbalances between the two causing sex-ratio distortion and larger imbalances leading to sterility [10]. The general features of this system are the basic arms race, where repeat number is effectively titrated against evolutionary consequence, unusual patterns of germline transcription, and an unknown mechanism in which one locus somehow inflicts specific damage on the cells destined to carry its homolog. In this context, the generation of antisense RNA could be understandable, either as an aggressive act aimed at disrupting other cells or processes, or as the genome's attempt to damp down a destructive conflict. Wouldn't it be nice to know whether *DXZ4* ever shows non-Mendelian segregation and, if so, whether this is linked to copy number? Sadly, the expected intensity of intragenomic conflict tends to make active interactions rather short-lived, so even if *DXZ4* was involved at one time, what we see today seems most likely to be at best a smoking gun.

Published: 13 April 2011

References

1. Árnason Ú, Allerdice PW, Lien J, Widegren B: Highly repetitive DNA in the baleen whale genera *Balaenoptera* and *Megaptera*. *J Mol Evol* 1988, **27**:217-221.
2. McLaughlin CR, Chadwick BP: Characterization of *DXZ4* conservation in primates implies important functional roles for CTCF binding, array expression and tandem repeat organization on the X chromosome. *Genome Biol* 2011, **12**:R37.
3. Jeffreys AJ, Wilson V, Thein SL: Hypervariable 'minisatellite' regions in human DNA. *Nature* 1985, **314**:67-73.
4. Tautz D: Hypervariability of simple sequences as a general source of polymorphic DNA markers. *Nucleic Acids Res* 1989, **17**:6463-6471.
5. The Huntington's Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993, **72**:971-983.
6. Fondon III JW, Garner HR: Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci USA* 2004, **101**:18058-18063.
7. Wang Z, Yuan L, Rossiter SJ, Zuo X, Ru B, Zhong H, Han N, Jones J, Jepson PD, Zhang S: Adaptive evolution of 5' HoxD genes in the origin and diversification of the cetacean flipper. *Mol Biol Evol* 2008, **26**:613-622.
8. Guerrero PA, Maggert KA: The CCCTC-binding factor (CTCF) of *Drosophila* contributes to the regulation of the ribosomal DNA and nucleolar stability. *PLoS ONE* 2011, **6**:e16401.
9. Chadwick BP: *DXZ4* chromatin adopts an opposing conformation to that of the surrounding chromosome and adopts a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res* 2008, **18**:1259-1269.
10. Hurst LD: Further evidence consistent with *Stellate*'s involvement in meiotic drive. *Genetics* 1996, **142**:641-643.

doi:10.1186/gb-2011-12-4-110

Cite this article as: Amos W: An enigmatic satellite. *Genome Biology* 2011, **12**:110.