Minireview

# The many hues of plant heterochromatin

## Jeffrey L Bennetzen

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907-1392 , USA.
E-mail: maize@bilbo.bio.purdue.edu

## Abstract

Recent sequence and cytogenetic analyses of heterochromatin in *Arabidopsis*, together with other results from *Arabidopsis* and maize, indicate that plant heterochromatin can have very different origins, compositions and dynamics. Shared features that must determine and/or be a result of its unique biological properties are also revealed.

Despite comprising the majority of many eukaryotic chromosomes, heterochromatin remains the least characterized and least understood portion of the higher eukaryotic genome. Traditionally, heterochromatin (literally, different in 'color' or staining properties) has been defined by the cytogenetic criterion that it has a different appearance from the more open euchromatic regions of eukaryotic chromosomes that contain most of the expressed genes. Even the first staining of higher eukaryotic chromosomes, however, indicated many different degrees of heterochromatic condensation and different sizes of heterochromatic blocks. Moreover, some heterochromatin has traditionally been associated with centromere or neocentromere function, while other regions appear not to have such activities [1]. In *Drosophila*, some heterochromatic regions (β heterochromatin) contain active genes, whereas other regions of heterochromatin (α heterochromatin) apparently lack active genes. Most euchromatic genes in *Drosophila* show abolished or attenuated expression when translocated into or near heterochromatic regions. In contrast, β heterochromatin genes like *light* show a reversal of this position effect leading to decreased activity when they are moved into euchromatin [2]. Hence, heterochromatin is a term that describes many different types of relatively condensed chromatin, perhaps with many different features and roles.

Recent advances in genomic sequencing technologies now make it possible (albeit challenging) to characterize fully the DNA composition of heterochromatic regions. In addition, more powerful cytogenetic tools now allow more fine-tuned studies of the nature and dynamic properties of heterochromatin *in situ*. Two articles in the February 4 issue of *Cell* [3,4] describe complementary DNA sequence and cytogenetic characterizations of a heterochromatic knob present on *Arabidopsis* chromosome 4.

## Characterizing plant heterochromatin

Paul Fransz and coworkers [3] provide a uniquely detailed and comprehensive cytogenetic analysis of a 9.5 megabase (Mb) segment covering the centromere regions and short arm of *Arabidopsis* chromosome 4. The existence of a contiguous series of clones covering this region made available a wide variety of probes for fluorescent *in situ* hybridization (FISH) analysis. Among the heterochromatic regions investigated were the nucleolar organizer (NOR), the centromeric region (comprising less than 1.8 Mb [5]), an associated pericentric heterochromatin region of 2 Mb or less, and a heterochromatic knob on the short arm of chromosome 4 (hk4S) that the authors estimated to be approximately 700 kilobases [3]. As had been seen by previous researchers for this and all other *Arabidopsis* chromosomes [5,6], a specific 180 base-pair (bp) tandem repeat (pAL1) was almost exclusively associated with the centromeric region. The pericentromeric heterochromatin appeared to lack these repeats, but contained many other repetitive DNAs, including a long tandem array of 5S rDNA genes and many retrotransposons [3]. The hk4S knob contained multiple repeats as well, including retrotransposons. Interestingly, different *Arabidopsis* ecotypes were found to be polymorphic for the presence of the hk4S knob, and structural analysis suggested

that this knob may be derived from an inversion event that moved a portion of the 4S pericentric heterochromatin to the current hk4S position [3].

Because they could align the cytogenetic and physical maps, Fransz *et al.* [3] were able to measure the degree of chromatin condensation (in kilobases per micrometer) on the short arm of *Arabidopsis* chromosome 4 during the stages of meiosis. Their results not only indicated that the euchromatin was much less compacted than the heterochromatin (as expected), but that different regions of heterochromatin showed very different degrees of condensation. For instance, NOR heterochromatin was over threefold more compacted than was the hk4S knob. Moreover, the degree and timing of decondensation was very different for different heterochromatic regions. The centromeric heterochromatin was found to decondense very early in meiosis (during prometaphase I), while adjacent pericentromeric heterochromatin exhibited little or no decondensation throughout meiosis [3]. Hence, these studies indicate that different classes and locations of heterochromatin have distinct structural dynamics as well as different sequence compositions.

The laboratories of Rob Martienssen and Dick McCombie at Cold Spring Harbor Laboratory and Rick Wilson at Washington University (St Louis) describe the first completed sequence of a known heterochromatic region, the hk4S knob, as an approximate 0.5 to 0.7 Mb segment of a contiguous 2.1 Mb of sequence [4]. Their results indicate very different compositions of genic (that is, euchromatic and gene-rich) versus heterochromatic regions. Moreover, the central 183 kb of knob hk4S had a very different sequence composition than was seen in the two regions of a few hundred kilobases on either side, both of which are still seen cytogenetically to be components of the knob [3,4] (Table 1). The core portion of hk4S contains 22.5 direct tandem repeats of a previously undescribed 1950 bp sequence that has some transposon-like properties, including 31 bp direct terminal repeats [4]. This core 183 kb was also enriched in both retrotransposons and DNA transposons, although they were all non-tandemly dispersed and none was inserted within the 1950 bp tandem array. The flanking few hundred kilobases were similarly enriched in retrotransposons and DNA transposons, but had only a few small tandem repeats [4] (Table 1). Only three hypothetical genes were predicted by search programs in the core 183 kb of the hk4S knob, and none of these showed strong homology with any identified *Arabidopsis* EST (expressed sequence tag from a cDNA) sequence. Hence, these hypothetical genes may be pseudogenes or may be expressed at a very low level. Many more genes were predicted within the flanking few hundred kilobases thought to be within hk4S, and two of these were highly homologous to known *Arabidopsis* ESTs. In fact, McCombie *et al.* [4] made the general observation that there was a variably steep gradient of predicted genes that matched ESTs as the sequence extended away from the core

region, with a reciprocal gradient showing the frequency of predicted genes that did not show homology to *Arabidopsis* ESTs. This suggests, then, a gradient of degrees of heterochromaticity as sequences extend away from a core heterochromatic region.

Copenhaver *et al.* [5] identified a similar uneven gradient of predicted gene density as sequences extended away from the centromeres of *Arabidopsis* chromosomes 2 and 4 (summarized in Table 1). As sequence analysis approaches the centromeres, more transposons (particularly retrotransposons with long terminal repeats, LTRs), more tandem repeats and more predicted pseudogenes were observed, but fewer functional genes were predicted [5].

Numerous studies have suggested that most or all meiotic recombination in plant genomes occurs near or within genes (reviewed in [7]). Hence, the low rate of recombination in heterochromatic regions could be caused primarily by the paucity of genes that are found in heterochromatin. McCombie *et al.* compared their sequence results to known recombination frequencies in the hk4S region [8]. As expected, the frequency of recombination in a particular subregion appeared to increase in proportion to the frequency of known functional genes (those with EST matches) and in inverse proportion to the local abundance of tandem repeats and interspersed transposable elements [4].

To date, relatively little genomic sequence information exists for plant species other than *Arabidopsis*, particularly for regions that might contain some heterochromatic potential. Of course, the ribosomal DNA genes of all eukaryotes, often associated with at least a partially heterochromatic status, occur as lengthy direct tandem repeat arrays of the 28S and/or 5S rDNA genes. A highly condensed knob on the short arm of chromosome 9 in maize has been investigated by sample sequencing, and found to be composed primarily of many tandem direct copies of short (180 bp and 350 bp) repeats interrupted by a fairly large number of LTR retrotransposons [9], though at a lower frequency than these elements are represented in the total maize nuclear genome [10] (Table 1).

Taken in sum, these results in *Arabidopsis* and other plant species suggest a situation very much like that for α and β heterochromatin of *Drosophila*. As in the centromeric regions and the core portion of hk4S in *Arabidopsis* and the maize 9S knob, α heterochromatin has long arrays of tandem repeats with occasional transposable element interruptions, and few or no functional genes. As in the pericentromeric and hk4S-flanking regions of *Arabidopsis* heterochromatin, *Drosophila* has β heterochromatin that is less condensed, largely deficient in tandem arrays, filled with copies of numerous transposable element families, and containing a few functional genes [11]. The increasingly-detailed studies that are now appearing for the heterochromatin of

**Table 1**

**Compositions of different DNA types in euchromatic and heterochromatic regions of plant genomes.**

| Region | Approximate proportions (%) | | | | | |
|---|---|---|---|---|---|---|
| | Tandem repeats* | Retroelements | DNA transposons | Predicted genes | Predicted pseudogenes | [Ref] |
| *Arabidopsis* 183 kb central hk4S knob | 25 | 30 | 15 | [0 to 4]† | | [4] |
| *Arabidopsis* 520 kb hk4S knob | 10 | 30 | 10 | [1 to 25]† | | [4] |
| *Arabidopsis* genic | 0 to 4 | 0 to 4 | 0 to 4 | 45 to 75 | 0 to 1 | [5] |
| *Arabidopsis* 'centromeric' (II and IV)‡ | 0 to >50 | 0 to 50 | 0 to 30 | 0 to 20 | 0 to 4 | [5] |
| *Arabidopsis* pericentromeric (II and IV) | 0 to 8 | 0 to 65 | 0 to 8 | 0 to 25 | 0 to 6 | [5] |
| Rice genic (*Adh1/Adh2*) | 0 | 14 | 14 | 26 | 4 | [20] |
| Maize genic (*adh1*) | 0 | 77 | 5 | 10 to 12 | 2 | [12] |
| Maize 9S knob | 70 | 30 | 0 | 0 | 0 | [9] |

*Does not include simple sequence repeats. †Did not distinguish between predicted genes and predicted pseudogenes. ‡Within functionally defined centromeric region, but not including central (most tandem repeat rich) region of the centromeres, which have not yet been sequenced.

*Arabidopsis*, *Drosophila* and other organisms, however, suggest that the α and β heterochromatin designations are actually arbitrary categorizations within a continuum of heterochromatin types.

## Towards a biological definition of heterochromatin

Does the repetitive DNA composition of a region on its own determine whether the region can be classified as heterochromatic? The answer may lie in how one defines heterochromatin. The original definition of heterochromatin is less biological than technical, depending upon the sensitivity of *in situ* chromatin characterization. Perhaps a better definition would be one based on the unique biology of heterochromatin, related to the presence and epigenetic regulation of any genes in the region. In maize, for instance, the sometimes small (5-70 kb) retrotransposon blocks that separate most genes are in a methylated, largely unexpressed state [10,12,13]. Hence, at a functional level, these could be considered to be heterochromatic blocks, but would not be ones detected by any current technology applied to a standard chromosome spread. In this regard, most genic regions of the maize genome may be more comparable in biology to *Drosophila* β heterochromatin than to standard euchromatin. Moreover, small tandem

expansions, such as adjacent duplication of a transposable element from a monomer to a dimer [14] or even an increase in the copy number within a simple sequence repeat (as in human fragile X syndrome and myotonic dystrophy [15]), can lead to altered epigenetic status within and surrounding these sequences. Hence, they might also be considered heterochromatin.

If heterochromatin is redefined as a largely gene poor, epigenetically inactivated region, then examples from *Drosophila*, maize and humans suggest that either direct or inverted repeats can lead to the creation of a heterochromatic state [12-15]. Transposable elements may commonly be found in these regions, but they are not likely to be obligatory. More probably, some transposable elements appear to insert preferentially in these regions: for example, the LTR retrotransposons Athila of *Arabidopsis* [16], Ty5 of yeast [17], and *Grande* of maize [9]. Maize shows severe insertion/retention biases for transposable element of all types. In a 225 kb region around the maize *adh1* locus, 13 of the 14 high copy number LTR retrotransposons are found inserted in other LTR retrotransposons (primarily into their LTRs), while the three LINE retroposons have inserted into introns, and the 33 miniature inverted repeat transposable elements (MITEs) [18] are all in the 23% of the region that does not contain LTR retrotransposons [10,12].

Even if tandemly repeated DNA alone can lead to a local heterochromatic state, this does not explain why many tandem repeats (like some rDNA genes and the centromere repeats) can retain biological function. Some rare genes buried within heterochromatin are active. Understanding the different structures and function of these heterochromatic regions in plants may yield many insights that are pertinent to the epigenetic regulation of euchromatic regions of the genome [19]. There is clearly a lot more to learn about heterochromatin, in all its many colors. More studies merging sequence analysis, epigenetic characterization, cytogenetics and other approaches are needed to understand heterochromatic systems fully. If recent results [3,4] are any indication, some common rules will be uncovered along with a great variety of phenomena unique to different types and assemblages of heterochromatin.

## References

1. Richards EJ, Dawe RK: **Plant centromeres: structure and control.** *Curr Opin Plant Biol* 1998, **1**:130-135.
2. Lohe AR, Hilliker AJ: **Return of the H-word (heterochromatin).** *Curr Opin Genet Dev* 1995, **5**:746-55.
3. Fransz PF, Armstrong S, de Jong JH, Parnell LD, van Drunen C, Dean C, Zabel P, Bisseling T, Jones GH: **Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region.** *Cell* 2000, **100**:367-376.
4. McCombie WR, de la Bastide M, Habermann K, Parnell LD, Dedhia N, Gnoj L, Schutz K, Huang E, Spiegel L, Yordan C *et al.*: **The complete sequence of a heterochromatic island from a higher eukaryote.** *Cell* 2000, **100**:377-386.
5. Copenhaver GP, Nickel K, Kuromori T, Benito M-I, Kaul S, Lin X, Bevan M, Murphy G, Harris B, Parnell LD *et al.*: **Genetic definition and sequence analysis of *Arabidopsis* centromeres.** *Science* 1999, **286**:2468-2474.
6. Round EK, Flowers SK, Richards EJ: ***Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure.** *Genome Res* 1997, **7**:1045-1053.
7. Schnable PS, Hsia AP, Nikolau BJ: **Genetic recombination in plants.** *Curr Opin Plant Biol* 1998, **1**:123-129.
8. Schmidt R, West J, Love K, Lenehan Z, Lister C, Thompson J, Bouchez D, Dean C: **Physical map and organization of *Arabidopsis thaliana* chromosome 4.** *Science* 1995, **270**:480-483.
9. Ananiev EV, Phillips RL, Rines HW: **Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin.** *Genetics* 1998, **149**:2025-2037.
10. SanMiguel PJ, Tikhonov A, Jin Y-K, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z *et al.*: **Nested retrotransposons in the intergenic regions of the maize genome.** *Science* 1996, **274**:765-768.
11. Holmquist GP, Kapitonov VV, Jurka J: **Mobile genetic elements, chiasmata and the unique organization of beta-heterochromatin.** *Cytogenet Cell Genet* 1998, **80**:113-116.
12. Tikhonov AP, SanMiguel PJ, Nakajima Y, Gorenstein N, Bennetzen JL, Avramova Z: **Colinearity and its exceptions in orthologous *adh* regions of maize and sorghum.** *Proc Natl Acad Sci USA* 1999, **96**:7409-7414.
13. Bennetzen JL, Schrick K, Springer PS, Brown WE, SanMiguel P: **Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA.** *Genome* 1994, **37**:565-576.
14. Dorer DR, Henikoff S: **Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*.** *Cell* 1994, **77**:993-1002
15. Tapscott SJ, Klesart TR, Widrow RJ, Stoger R, Laird CD: **Fragile-X syndrome and myotonic dystrophy: parallels and paradoxes.** *Curr Opin Genet Dev* 1998, **8**:245-253.
16. Pelissier T, Tutois S, Tourmente S, Deragon JM, Pickard G: **DNA regions flanking the major *Arabidopsis thaliana* satellite are principally enriched in Athila retroelement sequences.** *Genetica* 1996, **97**:141-151.
17. Zou S, Voytas DF: **Silent chromatin determines target preference of the *Saccharomyces* retrotransposon Ty5.** *Proc Natl Acad Sci USA* 1997, **94**:7412-7416.
18. Wessler SR, Bureaur TE, White SE: **LTR-retrotransposons and MITEs: important players in the evolution of plant genomes.** *Curr Opin Genet Dev* 1995, **5**:814-821.
19. Martienssen R: **Epigenetic phenomena: paramutation and gene silencing in plants.** *Curr Biol* 1996, **6**:810-813.
20. Tarchini R, Biddle P, Wineland R, Tingey S, Rafalski A: **The complete sequence of 340 kbp of DNA around the rice *Adh1-Adh2* region reveals interrupted co-linearity with maize chromosome 4.** *Plant Cell* 2000, **12**:381-92