the plant journal

SEB Society for Experimental Biology

The Plant Journal (2011) **66**, 18–33 doi: 10.1111/j.1365-313X.2011.04544.x

THE PLANT GENOME: AN EVOLUTIONARY VIEW ON STRUCTURE AND FUNCTION

Organisation of the plant genome in chromosomes

J. S. (Pat) Heslop-Harrison* **and Trude Schwarzacher***

Department of Biology, University of Leicester, Leicester LE1 7RH, UK

Received 31 January 2011; revised 11 February 2011; accepted 15 February 2011.
*For correspondence (fax +44 116 252 3330; e-mail phh4@le.ac.uk or ts32@le.ac.uk).

SUMMARY

The plant genome is organized into chromosomes that provide the structure for the genetic linkage groups and allow faithful replication, transcription and transmission of the hereditary information. Genome sizes in plants are remarkably diverse, with a 2350-fold range from 63 to 149 000 Mb, divided into *n* = 2 to *n* = approximately 600 chromosomes. Despite this huge range, structural features of chromosomes like centromeres, telomeres and chromatin packaging are well-conserved. The smallest genomes consist of mostly coding and regulatory DNA sequences present in low copy, along with highly repeated rDNA (rRNA genes and intergenic spacers), centromeric and telomeric repetitive DNA and some transposable elements. The larger genomes have similar numbers of genes, with abundant tandemly repeated sequence motifs, and transposable elements alone represent more than half the DNA present. Chromosomes evolve by fission, fusion, duplication and insertion events, allowing evolution of chromosome size and chromosome number. A combination of sequence analysis, genetic mapping and molecular cytogenetic methods with comparative analysis, all only becoming widely available in the 21st century, is elucidating the exact nature of the chromosome evolution events at all timescales, from the base of the plant kingdom, to intraspecific or hybridization events associated with recent plant breeding. As well as being of fundamental interest, understanding and exploiting evolutionary mechanisms in plant genomes is likely to be a key to crop development for food production.

Keywords: genome, nucleus, chromosomes, cytogenetics, genome size, evolution, polyploidy, centromeres, plant breeding, heterochromatin.

THE ORGANIZATION OF THE PLANT GENOME

Plant nuclear genomes

The plant nuclear genome, consisting of the DNA and associated proteins, is organized into discrete chromosomes. Each unreplicated chromosome and metaphase chromatid consists of a single DNA molecule that is linear and unbroken from one end to the other (Figure 1). At metaphase of mitosis, the DNA is condensed into mitotic chromosomes – short, rod like bodies – while at interphase, the chromosomes are decondensed within the interphase nucleus (Figure 2). The study of the chromosome and its organization involves cytogenetics, and the field of molecular cytogenetics has developed to understand DNA sequence and the molecular structure of the chromosome and chromatin. Both the size of the plant genome and the number of chromosomes vary widely between species. In this article we will discuss the nature and consequences of these differences in an evolutionary context.

The Arabidopsis genome sequencing initiative was established partially on the basis that the genes and gene sequences found in Arabidopsis would be substantially similar to those in all other plants (Meyerowitz, 1989; Somerville, 1989). Rice, because of its nutritional importance as a crop, was the next target for genomic sequencing following an initiative to identify all genes by sequencing. The similarity of gene sequences across all plants has been found to be true, although an initial surprise was the low total number of genes (27 206 protein-coding genes in Arabidopsis, The Arabidopsis Information Resource website, http://www.arabidopsis.org/portals/genAnnotation/genome_snapshot.jsp, and rice with 37 544 genes; International Rice Genome Sequencing Project, 2005), only half the number estimated before gene sequences were analysed directly (Heslop-Harrison, 1991). Arabidopsis and rice were also selected for genome sequencing in part because of their small genome size. Chromosome biologists have tended to choose species with large chromosomes as their 'model'

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd

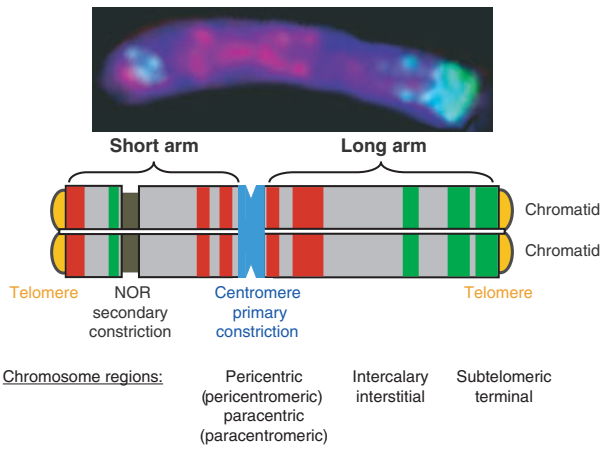

Figure 1. The organization and features of a plant chromosome.
Top: A fluorescent light micrograph of a metaphase chromosome stained blue with the DNA-binding fluorochrome 4′,6-diamidino-2-phenylindole (DAPI). *In situ* hybridization shows the location of two tandemly repeated DNA sequences detected by red and green fluorescence.
Centre: A diagram of the structure of a metaphase chromosome with two chromatids.
Centromeres or primary constrictions are seen as gaps in cytological chromosome preparations (see Figure 3b,d). They nucleate the proteinaceous kinetochore plate to which the spindle microtubules attach and are characterised by specific centromeric histones. DNA sequences at the centromeres are not conserved between species (see text).
The centromere and the regions surrounding it, called paracentromeric regions, contain large arrays of tandem repeats and are often enriched in transposable elements.
Euchromatin. Lightly stained in cytological preparations (Figure 3a); generally gene rich, showing high transcriptional activity and higher levels of recombination at meiosis. Transposable elements may be dispersed widely through euchromatin.
Heterochromatin. Strongly stained in cytological preparations (Figure 3a); rich in highly repeated tandemly organised DNA sequence families and sometimes transposable elements. Heterochromatin generally lacks meiotic recombination and is relatively deficient in genes, and those present often have decreased transcriptional activity.
NOR. Nucleolus organising regions contain the long arrays of 45S rDNA repeat units, including the 18S–5.8S–26S rRNA genes and intergenic spacers. Most genomes have several major and minor rDNA loci (Figure 4c,d). Expression of the rRNA genes generates the nucleolus at interphase (Figure 4b); at metaphase, NORs are often visible as secondary constrictions as the arrays of genes active at the previous interphase remain decondensed.
Subtelomeric or telomere associated sequences (TAS) are long tandem repeats (Figure 3b) sometimes containing degenerate (TTTAGGG)*n* motifs, and are species specific and often chromosome specific.
Telomeres, at the ends of chromosomes, are relatively short arrays usually of the conserved simple repeat (TTTAGGG)*n* (Figure 4c). They maintain chromosome integrity by stabilizing chromosome termini.

species such as *Secale*, *Triticum* (Figures 3 and 4), *Lilium* or *Vicia faba.*

Chromosome organization is related to genome function within the cell nucleus (Spector, 2003), with physical organization relating to regulation and gene expression, cell division, recombination and replication. There are genes involved in aspects of chromosome organization. The Gene Ontology (GO) project aims to generate descriptions of gene products in their database consisting of a controlled vocabulary of terms covering biological concepts (http://amigo.geneontology.org). It defines 'chromosome organization' as 'a process that is carried out at the cellular level that results in the assembly, arrangement of constituent parts, or disassembly of chromosomes, structures composed of a very long molecule of DNA and associated proteins that carries hereditary information'. Many of these genes are related to chromatin (see Fransz and deJong, 2011), or meiosis and recombination, rather than the structural and evolutionary aspects of chromosome organization that are discussed here.

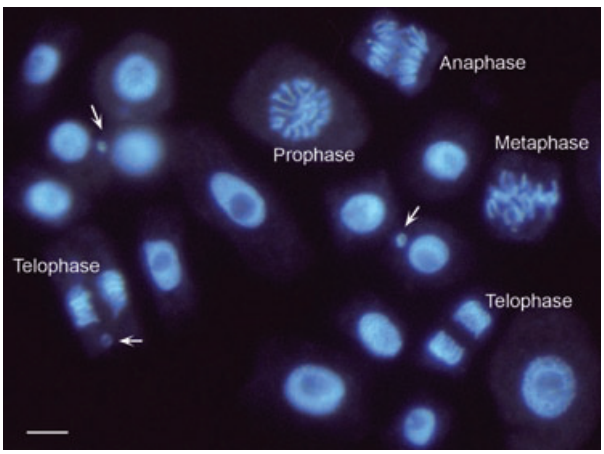

Figure 2. Chromosomes at all stages of the cell cycle.
A spread of a root tip of a hybrid plant of *Hordeum* × *Secale* showing nuclei at all stages of the cell cycle (labelled). The condensation of chromosomes to metaphase and decondensation at telophase is evident, and one or several nucleoli are seen within the interphase nuclei. This hybrid line is unstable and chromosomes are sometimes lost during division, forming micronuclei (arrows). Bar 10 μm.

Non-nuclear genomes and DNA sequences

Along with the nuclear genome, genes are also carried in the organelles (chloroplasts or plastids, and mitochondria) and the genomes of viruses, mycoplasmas, bacteria and fungi may be present within or in close association with plant nuclei or cells. These genomes interact and impact on the organization and evolution of the associated plant nuclear genome. Furthermore, the possible presence and effects of non-nuclear genomes (which may be transmitted to the next generation) must be considered in genomic and evolutionary studies. Increasing amounts of data obtained after the first plant genome sequences were completed have shown that transfer of genes into the plant nuclear genome, while not frequent, is a regular and evolutionarily important occurrence.

Transfer of genes from both mitochondria (see Goremykin *et al.*, 2009) and chloroplasts or other plastids (see Cullis *et al.*, 2009) to the nucleus over evolutionary time has led to the loss of many genes from organelles (see Green, 2011). There is also evidence for transfer of genes from mitochondria to chloroplast (grape: Goremykin *et al.*, 2009). These

© 2011 The Authors

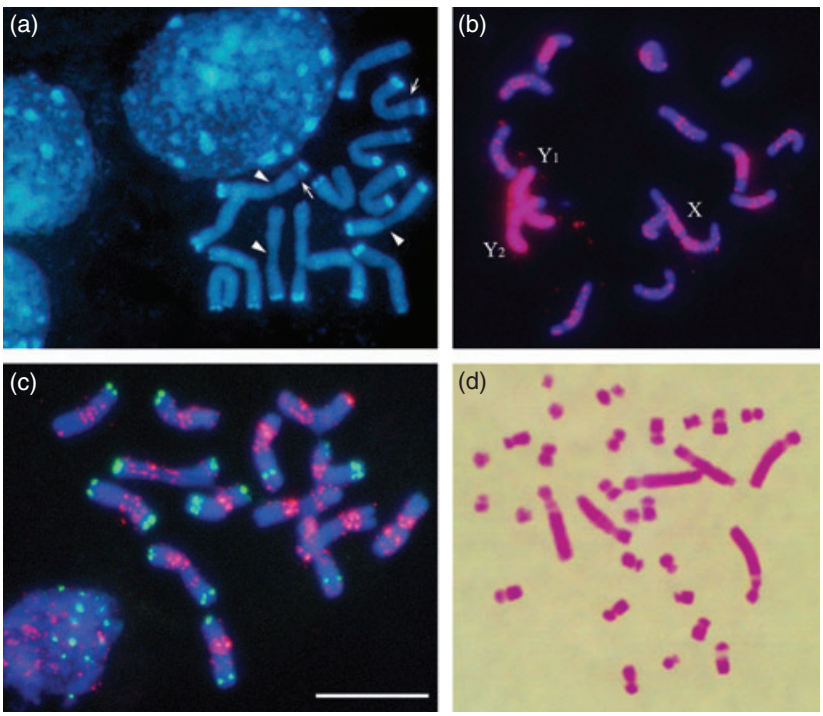

Figure 3. Metaphase chromosomes with centromeres and heterochromatin composed of tandemly repeated sequences.
(a, c) Metaphase and interphase chromosomes of rye, *Secale cereale*, $2n = 2x = 14$ after fluorescent banding with 4′,6-diamidino-2-phenylindole (DAPI) (a); and fluorescence *in situ* hybridization (FISH) (c); with simple sequence repeats (red) and tandemly repeated satellite DNA sequences (green; Cuadrado and Schwarzacher, 1998). Many homologous chromosomes show differences of signal strength indicating polymorphic repeat copy numbers in this heteromorphic and outbreeding species. At interphase the subtelomeric heterochromatin consisting of tandemly repeated satellite DNA (Alkhimova *et al.*, 2004) is strongly stained with DAPI (a); and green FISH signals (c); are on opposite ends to the centromeres identified by the red FISH signals. (b) Chromosomes of *Rumex acetosa* $2n = 12 + XY_1Y_2$. After FISH with the simple sequence repeat $(AAC)_5$. The sequence is amplified on the two Y chromosomes. (d) Feulgen stained chromosomes of *Cephalanthera longifolium* ($2n = 36$). Both large and small chromosomes show clear primary constrictions at the centromere. Bar = 10 μm in (a) and (c); 15 μm in (b) and (d).

authors, and Bock and Timmis (2008), review the continuing nature of transfer of genes into the nucleus, with the increased regulatory ability, and the variation in genes that have been transferred in different evolutionary groups of plants. These gene transfers have led to many incongruent evolutionary trees from analysis of nuclear copies of organellar genes, where short PCR products have not distinguished the origin of the gene. Large insert (e.g. BAC) sequences can identify DNA sequences flanking the organelle-origin genes, or *in situ* hybridization can show their location on chromosomes rather than in organelles (e.g. Vaughan *et al.*, 1999).

Viral genomes, particularly from the pararetroviruses with a DNA genome, can transfer from the episomal virus into the nucleus, and can be expressed as infective virus particles that cause disease. The banana streak virus was the first example to be characterized (Gayral and Iskra-Caruana, 2009; Harper *et al.*, 1998), and petunia and tobacco vein clearing virus was identified soon after (Lockhart *et al.*, 2000; Richert-Pöggeler *et al.*, 2003). Solanaceous species are particularly rich in endogenous pararetroviruses (EPRVs; Hansen *et al.*, 2005), where the majority show homology to *Cavemoviruses* and in some cases reach several thousand integrants. Host genome invasion by pararetroviruses has occurred several times during the evolution of Solanaceae (Staginnus and Richert-Pöggeler, 2006) and repeatedly in banana (Gayral and Iskra-Caruana, 2009). Non-functional sequences as well as complete and inducible integrants have been isolated indicating that integrated sequences decay and can be highly degenerated; they tend to be concentrated in pericentromeric heterochromatin associated with retrotransposable elements (*Metaviridae* sequences; Gregor *et al.*, 2004; Hansen *et al.*, 2005; Staginnus *et al.*, 2007), and may play a role in host defence against virus infection through RNAi silencing (Staginnus and Richert-Pöggeler, 2006).

In the 1970s, *Agrobacterium* species were shown to be able to transfer genes for hormone and opine synthesis into the plant nuclear genome, and Schell and van Montagu showed how this property could be used in plant transformation (see, for example, Zambryski *et al.*, 1983). Subsequently, technology to transfer genes from outside the nucleus into the genome of the host plant has been developed using the *Agrobacterium* or other approaches.

© 2011 The Authors

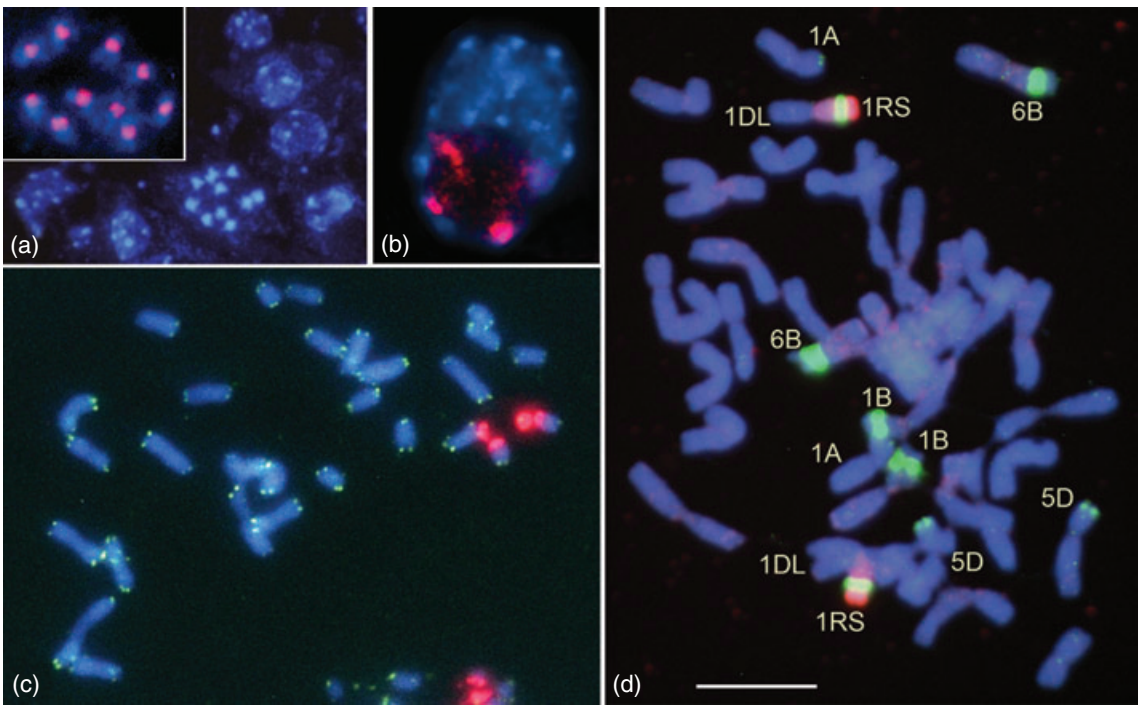

Figure 4. Large and small chromosomes share features of rDNA loci, centromeres and telomeres.
(a) 4′,6-diamidino-2-phenylindole (DAPI)-stained metaphase and interphase chromosomes of *Arabidopsis thaliana;* enlarged inset after fluorescence *in situ* hybridization (FISH) with the abundant centromeric 180-bp repeat (red). (b) Interphase of *Medicago truncatula* after FISH with 45S rDNA probe (red). Unexpressed condensed sites of the rDNA are visible at the periphery of the nucleolus while decondensed, expressed strands run though the lighter volume of the nucleolus with less DAPI-stained DNA (micrograph from Matheus Mondin). (c) A wheat/rye recombinant line carrying a 1DL.1RS translocation as identified by genomic rye DNA (red) and the presence of major 45S rDNA sites (green) on the short arms of chromosome 1R, 1B and 6B and minor sites on 5D and 1A. (d) Oil palm, *Elaeis guineensis* ($2n = 32$) metaphase chromosomes after FISH with the telomeric sequence $(TTTAGGG)_6$ (green) located in variable copy numbers at the end of each chromatid, and the 45S rDNA on one chromosome pair (red). Part of an interphase is visible lower right hand side (Castilho *et al.*, 2000). Bar = 10 μm.

Molecular cytogenetic analysis including fluorescent *in-situ* hybridization is very appropriate to locate the transgene in the genome), and even to determine copy numbers (Fransz *et al.*, 1996; Leggett *et al.*, 2000; Pedersen *et al.*, 1997; Salvo-Garrido *et al.*, 2001; Schwarzacher, 2008; Svitashev and Somers, 2002; Wolters *et al.*, 1998). Considerable efforts are required for analysis of low or single copy sequences, but these are justified as verification of nuclear integration may be difficult by Southern hybridization or PCR, particularly in slow-growing, sterile or non-intercrossable hybrids or polyploids where transmission and segregation analysis is impractical. Chromosomal analysis of transgenic lines can also establish whether the plants have maintained their chromosomal integrity or whether aneuploidy, polyploidy or rearrangements have occurred.

Composition of nuclear DNA

The nuclear DNA of plants consists of the single- or low-copy coding sequences, introns, promoters and regulatory DNA sequences, but also of various classes of repetitive DNA motifs that are present in hundreds or even thousands of copies in the genome (Heslop-Harrison and Schmidt, 1998). Repetitive DNA motifs include characteristic sequences at chromosome centromeres and telomeres (see below; Figures 1 and 4a,c), and the rDNA (rRNA genes and intergenic spacers) at the 45S and 5S loci (Figures 1 and 4b–d). Tandemly repeated or satellite DNA consists of a motif (as short as two bases, a microsatellite or simple sequence repeat, but sometimes 10 000 bp long) that is repeated in many copies at one or more genomic locations (Figure 3b,c). Satellite DNA in plants typically consists of motifs of about 180 bp (Kubis *et al.*, 1998; Contento *et al.*, 2005), and can be seen either as deep-staining heterochromatin that does not decondense during interphase (blue condensed chromatin in Figures 3a and 4b) or by *in situ* hybridization of the sequence after labelling (Figure 3b,c); these satellite sequences are abundant but their function in the genome is not known. Transposable elements are the third class of repetitive DNA sequences; both class I (retrotransposons) and class II (DNA transposons) elements may amplify and the elements and recognizable degraded remnants may represent half or more of the entire DNA present in the genome (Kubis *et al.*, 1998; Pearce *et al.*, 1997). Both classes of transposable elements include sequences that encode enzymes related to their own replication and integration into the nuclear DNA.

© 2011 The Authors

Genome size or nuclear DNA content

Each plant species has a characteristic number of base pairs in its nuclei, known as its genome size or nuclear DNA content. Work from Swift (1950) onwards has shown that the nuclear DNA content is largely constant within a species. Measuring and cataloguing the size of genomes, number of chromosomes and range of chromosome sizes and morphology (karyotypes) has been carried out over many decades. Karyotype data have proven useful for evolutionary and phylogenetic studies at taxonomic levels between the species and family. In contrast to DNA sequence data, karyotype data often do not allow inference of higher levels of relationships. Indeed, the significance, any selective constraints, or other 'reasons' for differences in genome organization above the family level between species groups remain unknown.

Genome sizes are now normally estimated by using flow cytometry, replacing earlier methods of measuring absorbance of stained nuclei (microdensitometry). Nuclear genome size has been widely measured and cited in pg (picograms) of DNA, but in the context of molecular biology is now most frequently given in number of base pairs for the 1C DNA content. A nucleus immediately after meiosis but before DNA replication will have the 1C DNA content, while a replicated nucleus entering mitosis in the vegetative part of an angiosperm would have four times this amount, the 4C DNA content. Bennett and Leitch (2011) have assembled the diverse measurements of plant genome sizes into online databases (http://data.kew.org/cvalues/cvalOrigReference.html); the algae (see Bowler and Tirichine, 2011), pteridophyte and bryophyte data are not considered here. The databases of plant and animal genome sizes have been discussed in a broader context by Gregory *et al.* (2006; http://www.genomesize.com).

Published measurements of genome sizes and chromosome numbers often need critical assessment as they can be made for purposes where rigorous checking and replication is not required, may be field-based, carried out on a large scale, use techniques which are unproven or of limited reliability, or have technical errors (Greilhuber, 2005; Suda and Leitch, 2009). Hence individual reports should be compared with measurements from multiple sources or observation. Reports of extreme values are particularly prone to error. Casual examination of stained chromosome preparations by light microscopy – preferably of metaphase spreads, but even of stained interphase nuclei – will avoid mistakes in measurement of genome size by four-fold or more, and ensure that diploids are separated from polyploids with 50% (3×) or more (4× and above) chromosomes.

Bennett and Leitch (2011) report angiosperm genome sizes as varying from the smallest reported higher plant genome size of 63 Mb in two species of the carnivorous *Genlisea*, *G. aurea* (2*n* = approximately 52) and *G. margaretae* (2*n* = approximately 40), to the largest of *Paris japonica* (2*n* = 8*x* = 40) at 149 000 Mb, a 2350-fold range among measurements of 6288 species. For a diploid rather than polyploid species, *Fritillaria platyptera* (2*n* = 2*x* = 24) has the highest value at 84 150 Mb. Species with the smallest genomes of <200 Mb belong to one monocot and 13 diverse eudicot families. Many species with very large genomes are in the order Liliales (Liliaceae, Melanthiaceae and Alstroemeriaceae), with only nine eudicot families having species with genomes over 15 000 Mb. The average angiosperm genome size is 5800 Mb, with the major groups (Angiosperm Phylogeny Group III, 2009) of basal angiosperms (average 2300 Mb) and eudicots (2800 Mb) being smaller than the monocots (10 200 Mb, reduced to 8500 Mb if the order Liliales is excluded). Interestingly, gymnosperm genomes are larger with an average genome size of 18 200 Mb, and a range from 2200 to 35 200 Mb. Figure 5 illustrates this wide range of nuclear DNA contents in angiosperms.

Among eukaryotic genomes which have been sequenced, the average length of the coding sequences (excluding introns) has been reported as 1346 bp (with little variation between groups; Xu *et al.*, 2006), while the number of genes in diploid higher plants has been found to be about 30 000 (see Ming *et al.*, 2008), accounting for a total of 40 Mb of

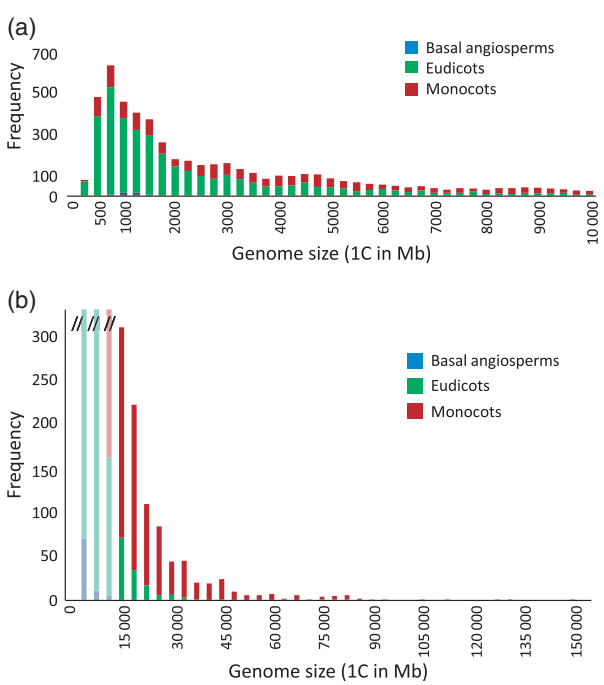

Figure 5. Frequency distribution histograms showing the nuclear DNA content of angiosperms.
(a) Genome sizes up to 10 000 Mb in 250 Mb bins. (b) Genome sizes up to 150 000 Mb in 3750 Mb bins. Vertical axis: frequency; Horizontal axis: alternate bin boundaries in Mb. red: monocots; green: eudicots; blue: basal angiosperms; light: truncated columns. Data from http://data.kew.org/cvalues/cvalOrigReference.html, downloaded 1/2011 (see Bennett and Leitch, 2011; Leitch *et al.*, 2010).

© 2011 The Authors

DNA. With the requirement for structural regions of chromosomes (centromeres and telomeres), rRNA, regulatory sequences and introns, this suggests 60 Mb is close to the minimum genome size. Lysak *et al.* (2009) studied genome size evolution in the Brassicaceae (showing a 16-fold range in 185 taxa studied) in the context of the phylogenetic relationships within the family. They concluded that half the species had a decreased genome size compared with the common ancestor, despite the occurrence of dynamic genomic processes (transposition of transposable elements and polyploidization) that can increase genome size; the mechanisms to eliminate amplified DNA remain to be elucidated. Knowledge of genome size is important for choice of strategies for genomic projects including library construction, cloning, and genome sequencing. In general terms the collection of this data has not revealed general principles related to consequences of variation in genome size, nor suggested constraints, nor the mechanisms or selection pressures that modulate genome size over evolutionary time.

Greilhuber (2005) remarked that the occurrence and extent of genome size variation below the species level is controversial, pointing out faults in a number of studies reporting differences. Nevertheless, unless speciation is driven by genome size changes, differences between species show that intraspecific differences in DNA content are present and have consequences for chromosome behaviour including meiotic pairing. Chromosomal polymorphisms caused by differences in repetitive DNA sequences can occur rapidly. In maize, there are differences in the sizes of terminal heterochromatic knobs, consisting of repetitive DNA sequences (Aguiar-Perecin and de Vosa, 1985; Laurie and Bennett, 1985). The extensive variation in heterochromatin contents in rye – seen as chromosome polymorphisms even within the two homologues (see Figure 1b) – also gives differences in nuclear DNA content (Alkhimova *et al.*, 2004). Under some conditions, repetitive sequences at the terminal regions of chromosomes are lost during mitotic divisions. Özkan *et al.* (2010) have shown limited variation in genome size in wheat, with substantial interspecific variation, due to the activity of retroelements. Copy number variations (CNV) have been demonstrated to arise in the rRNA arrays of flax given different treatments by Cullis (2005). CNVs involving chromosome segments more than 1 kb in size with insertions, deletions and duplications, have been found across all chromosome arms in maize (Beló *et al.*, 2010). Such polymorphisms in the genome, in plants like animals, are likely to have important consequences for populations and their adaptation (Biemont, 2008), disease response and heterosis (Beló *et al.*, 2010).

Chromosome number

Every species has a characteristic number of chromosomes in the nucleus. Numbers vary extensively between species, and examples of both increases and decreases during evolution and speciation are frequent. Within the eudicots, the lowest and highest chromosome numbers, $2n = 4$ and $2n$ = around 640 have both been reported in the single genus *Sedum* (Crassulaceae; in a flora by 't Hart and Bleij, 2003; source and reliability unknown), although few species have more than 200 chromosomes. Several other eudicots and monocots have $2n = 4$, while $2n$ = circa 596 has been reported in the monocot palm *Voanioala gerardii* and $2n$ = around 1200 in the fern *Ophioglossum reticulatum.* In genetic mapping and DNA sequencing projects, chromosome number is critical to know as it defines the number of independent linkage groups.

There are a few exceptions to the constancy of chromosome number within a species where species include several cytotypes, like members with different ploidy levels. For example, individuals of *Hordeum murinum* may be diploid ($2n = 2x = 14$) or tetraploid ($2n = 4x = 28$) plants (Taketa *et al.*, 1999); there are even a few tetraploid populations of *Arabidopsis thaliana* ($2n = 4x = 20$; Heslop-Harrison and Maluszynska, 1994; Steinitz-Sears, 1963). Another source of variation in chromosome number (and genome size) is the presence of supernumerary or B chromosomes (review: Jones *et al.*, 2008) in addition to the normal chromosome complement. These usually small chromosomes are derived from the standard chromosomes in the complement, and apparently lack genes although there is a 'drive' process which ensures their survival and indeed amplification in number within some plants despite having detectable and often negative effects on the phenotype.

In contrast to the wide chromosome number range seen among the angiosperms, gymnosperms (characterized by large genomes; Murray *et al.*, 2002) have no species with extreme chromosome numbers (typically $2n = 2x = 14$–28), and there are very few polyploid species in the group. Chromosome number can be stable across families: of the 232 species in 11 genera in the Pinaceae, all those studied have $2n = 2x = 24$ chromosomes except for Douglas fir (*Pseudotsuga menziesii*, $2n = 26$; Krutovsky *et al.*, 2004). The 400–500 species of grasses (Poaceae) in the subtribe Triticeae, including barley, rye, wheat and a number of forage grasses (Barkworth, 2010), all have a basic chromosome number of $x = 7$ (Figure 3a,c), although many are polyploids (Figure 4d; see below). In contrast, the *Brassica* genus has a wide range in chromosome number, and the changes, discussed below, may be driving speciation.

Chromosome size

Average chromosome size for a species is derived from chromosome number and genome size. Based on Bennett and Leitch (2011), taking unreplicated haploid genome sizes (1C) for angiosperms and dividing by haploid number (n) of

© 2011 The Authors

chromosomes reveals that 18 of the 5163 species have chromosomal DNA molecules (as would, for example, be analysed by pulse field gel electrophoresis, PFGE) <10 Mb in average size, while 118 species have an average size of more than 3000 Mb. The double-stranded DNA molecule in each chromatid of a metaphase chromosome of *Genlisea aurea*, averaging 2.4 Mb, is only half the size of the 4.6 Mb genome of the bacteria *Escherichia coli.* In species with small chromosomes, stained bacteria (where the genome may be replicated several times) can be confused with chromosomes in microscope preparations. Figure 4 shows *A. thaliana* chromosomes averaging 30 Mb in size together with wheat chromosomes averaging 800 Mb and oil palm (*Elaeis guineensis*) chromosomes of 114 Mb.

Despite the stability of chromosome number in the Pinaceae ($2n = 24$), genome size varies over a three-fold range up to 35 000 Mb, and in the Triticeae, the haploid, x, genome size varies from about 3300 to more than 8000 Mb. Like genome size and chromosome number, these differences in average chromosome size, and the nature of the differences involving amplification or DNA and RNA transposable elements, tandemly repeated DNA sequences, and perhaps segmental duplications of the genome, can be described accurately from several complementary methods. Detailed sequence analysis (e.g. International *Brachypodium* Initiative, 2010) indicates that footprints of centromeric repeats and peaks in retroelement frequency are seen at the junctions of ancestral chromosome insertions. Both single-generation chromosomal changes and long-term accumulation of repetitive DNA have evolutionary roles in reproductive isolation and restriction of gene flow between newly evolving species, with consequences for understanding genome and gene evolution, as well as for the population biology, acquisition, loss or modification of gene function, and allele diversity.

As chromosomes within a species can be of different sizes, they can be sorted using flow-cytometry based on their fluorescence. In bread wheat, the first DNA library was made by Wang *et al.* (1992) from wheat chromosome 4A. A flow sorted BAC library of chromosome 1B was made by Janda *et al.* (2006), and many other chromosomes have been sorted and characterized (Dolezel *et al.*, 2007; Paux *et al.*, 2006; Šafář *et al.*, 2004). The International Wheat Genome Sequence Consortium (IWGSC – http://www.wheatgenome.org) is using these flow sorted chromosomes to partition the wheat genome before chromosome-by-chromosome sequencing of the 17 000 Mb genome.

CHROMOSOMAL AND KARYOTYPE EVOLUTION

Chromosome evolution and structural variation

Chromosomes evolve by fission and fusion (leading to a change in chromosome number, or to inversions of segments within one chromosome; Jones, 1998), events that may be accompanied by duplication and inversions of chromosome arms. As an example, the chromosomes of the native European orchid *Cephalanthera* (see Figure 3d) with species having $2n = 32$, 36 or 44, are thought to have evolved by palaeotetraploidy from $x = 9$ followed by centric (Robertsonian) fusions leaving interstitial telomeres (Moscone *et al.*, 2007).

With genomic data involving both genetic mapping and genome sequencing, it is now possible to identify the large scale chromosomal rearrangements that have occurred during evolution. Chromosome numbers in the Brassicaceae vary from $2n = 8$ to $2n = 256$ (Lysak *et al.*, 2005). *A. thaliana*, with $2n = 10$ (Figure 4a), has one of the smallest chromosome numbers, an advanced character representing reduction from its ancestors in the clade including *A. lyrata* and *Capsella rubella* (both $2n = 16$). An impressive use of comparative chromosome painting to meiotic pachytene chromosomes using groups of BAC probes to identify each chromosome segment allowed Lysak *et al.* (2006) to show the origin of each chromosome in *A. thaliana* relative to the ancestral $n = 8$ karyotype, involving four chromosomal inversions, two translocations and three chromosome fusion events. In *Brassica*, Mandakova and Lysak (2008) used multiple selected BACs as probes to reveal the monophyletic origin of the $x = 7$ tribes, some of which included a translocation where chromosomal segments are exchanged between two chromosomes. The results also suggest that structure of the ancestral karyotype of the *Brassica*, with a reduction in chromosome number from $n = 8$ to $n = 7$ has happened more than once, with different fusion and intra-chromosomal inversion events. Xiong and Pires (2011) have developed an *in situ* chromosome painting method to identify all chromosomes in *Brassica napus* and its diploid progenitors, showing a chromosomal translocation in one *B. napus* cultivar. They suggest that this approach will be useful to understand chromosome reorganization, genome evolution and recombination; sequence analysis would not be appropriate for the detection of single translocation breakpoints.

While some of the chromosome number changes occur through doubling of chromosome numbers or polyploidy (see below), many involve fusion or fission of chromosomes, as shown in the Brassicaceae, grasses and many other families. Through sequence comparisons, multiple orthologous gene sequences are found to show a conserved order (synteny) along chromosomes over large taxonomic distances. Data of this nature are accumulating rapidly, and syntenic comparisons are now an essential part of most genome sequence papers. For example, Jaillon *et al.* (2007) compared *Vitis* (grape vine) genomic regions to their orthologues in *Populus trichocarpa*, *A. thaliana* and *Oryza sativa*, a taxonomic range where direct comparisons were hardly conceivable before sequence-based comparisons became possible. In *Vitis*, their analysis showed that the

© 2011 The Authors

genome has been triplicated during its early evolution, before the split of the poplar/Arabidopsis/*Vitis* lineages, but after the monocot/eudicot split as it was not shared with rice. The analysis identified an additional duplication in the poplar lineage, and two whole genome duplication events in the Arabidopsis lineage, as well as global duplications in the rice lineage. In the grass *Brachypodium distachyon* ($2n = 10$), sequencing of the 272 Mb genome (International *Brachypodium* Initiative, 2010) revealed a complex evolutionary history with six major interchromosomal duplications within the genome, the five *Brachypodium* chromosomes originating from a five-chromosome ancestral genome through a 12-chromosome intermediate involving seven major chromosome fusions. Sets of collinear genes along all ten *Brachypodium* chromosome arms can be identified easily in the other grasses where detailed genetic maps are available (rice, barley, wheat, sorghum, and *Aegilops tauschii*). Twelve separate syntenic blocks of orthologous genes from *Brachypodium* are present in rice, sorghum and barley, with nested insertions of some *Brachypodium* ancestral groups into centromeres of the other species. In the Triticeae, a detailed analysis of syntenic regions by Luo *et al.* (2009) has shown how the basic number of $x = 7$ has been derived from $x = 12$ in the ancestral species (represented by rice and sorghum) not through end-to-end chromosome fusions, or translocations and loss of microchromosomes, but by the insertion of four whole chromosomes into breaks in the centromeric region of four other chromosomes, with a further fifth fusion and translocation event.

Analysis of the nature of the rearrangements using whole genome sequence comparisons is enabling the history of genome evolution to be reconstructed with unprecedented accuracy. For plant breeders, knowledge of the nature of the changes shows the types of changes which might be introduced in the future, and suggests strategies and candidate accessions for crossing programmes. Parallel work across the mammals (Nagarajan *et al.*, 2008) is also showing the evolutionary chromosome rearrangements across diverse species. Similar chromosomal fusion, fission and elimination events to those discussed in *Brassica* have been reported in cattle and the Artiodactyla (Chaves *et al.*, 2003). In mammals, *in situ* hybridization and chromosome painting is widely used (Froenicke *et al.*, 2006). Despite some successes (Mandakova and Lysak, 2008), this technique has been less used in plants, presumably because of the more rapid homogenization of DNA sequences from retrotransposons, so probes from large amounts of DNA become genome-specific rather than chromosome- or linkage-group specific. Recent advances in large-insert (BAC or fosmid) hybridization suggest it will be increasingly used to address chromosome evolution (Lysak *et al.*, 2006) and physical linkage mapping of sequences (Anhalt *et al.*, 2008; Han *et al.*, 2011).

Aneuploidy – chromosome loss or gain

Aberrant cell division is relatively frequent, and chromosomes are lost or gained during mitosis or meiosis leading to aneuploidy. Figure 2, an intergeneric hybrid, shows nuclei at all phases of the cell cycle, but includes some cells with micronuclei (arrows) from mis-divisions. In many cases, these cells will not divide further, but the mis-division can occur in gametes or cells which regenerate to a whole organism. In mammals, most such aneuploids do not develop. Many plant aneuploids grow to generate adult plants, not least because plant genomes are often polyploid (see below) and have higher plasticity and mechanisms for gene dosage compensation. Chromosome addition lines, with an extra copy of a chromosome, occur naturally (first found in *Datura* by Blakeslee and Avery, 1919). They are also made by crossing tetraploid and diploid plants, or crossing different species, followed by backcrossing to derive lines with one or a few extra chromosomes. These hybrids have proved valuable to transfer alien chromosomes from wild relatives to crop species; recombination between the alien and crop chromosome can then reduce the chromosome number while still transferring the required characters. Particularly in wheat, such lines (Figure 4d) have a long history of use in breeding programmes (see, e.g. Heslop-Harrison *et al.*, 1991; Schwarzacher *et al.*, 1992; Bardsley *et al.*, 1999), and a number of programmes are exploiting the transfer of important disease resistance genes into wheat (Ayala-Navarrete *et al.*, 2007; Sepsi *et al.*, 2008; Graybosch *et al.*, 2009; Molnár *et al.*, 2011).

Monosomic plants are regularly found in species with a recognizable polyploid ancestry and are missing one (of a pair) of chromosomes. These have proved extremely valuable for genetic analysis, as the phenotype of the plant reflects modified expression of the genes carried by that monosomic chromosome; substantial amounts of genetic analysis in wheat (Sharp *et al.*, 1989) and in maize have involved monosomic analysis (Helentjaris *et al.*, 1986). Trisomic lines, with an additional single chromosome, are also valuable for genetic analysis of diploid species to assign linkage groups to chromosomes (rice: McCouch *et al.*, 1988).

Polyploidy

Whole genome duplication or polyploidy has probably played a major role in the evolution of all angiosperms by enabling fertile interspecific hybrids to be generated with multiple gene alleles at each locus, through freeing duplicated genes to mutate, and through reproductive isolation of new polyploids leading to speciation with limited gene flow (see, for example, Soltis and Burleigh, 2009; Proost *et al.*, 2011). Polyploidy can arise by multiplication of the genome in one plant – autopolyploidy – or through hybridization of two species with doubling of the chromosomes of one or more of the species involved – allopolyploidy. Autopolyp-

© 2011 The Authors

loids may be recognized as a different species from their diploid progenitor, or may be placed in the same taxon, despite usually having some morphological differences including size and pollen morphology, and being reproductively isolated.

Cytological evidence for polyploidy includes the occurrence of a regular series of chromosome numbers within a species group (e.g., *Cephalanthera*; Moscone *et al.*, 2007), the behaviour of hybrids with chromosome pairing at meiosis, and the existence of monosomic plants. In the 1990s, this evidence suggested that perhaps 30% of plants were polyploid, although some questioned whether species such as maize were polyploids or palaeopolyploids. However, with DNA sequence and genetic map data showing the presence of copies of multiple genes in the same order on two or more chromosomes, evidence for whole genome duplications or polyploidy in the ancestry of species becomes unequivocal (Tang *et al.*, 2010). Schnable *et al.* (2009) show that every chromosome arm in maize carries blocks of genes duplicated in order on another chromosome, and the results clearly show chromosomes involved in translocations. It is now obvious that 'diploid' *Brassica* species including *B. oleracea* and *B. rapa* are ancient hexaploids (Lagercrantz and Lydiate, 1996), with three different genomes. The analysis of sequence data in combination with physical and genetic mapping shows the complex nature of the collinear genome segments, translocations and inversions (Trick *et al.*, 2009) and the amplification of repetitive elements after separation of the ancestral species (Alix *et al.*, 2008).

Many of the polyploid events, recent and ancient, have involved autopolyploidy or hybridization of species which are evolutionarily close. For these plants to be fertile, meiotic chromosome pairing must lead to regular formation of bivalents, rather than multivalents involving more than one homologous pair of chromosomes where recombination and segregation would lead to unbalanced gametes. In wheat, Riley and Chapman (1958) described the effect of a single locus, Pairing homoeologous (*Ph*), which ensures strict bivalent formation, showing that homology search mechanisms are under genetic control. We can speculate that the widespread and early occurrence of polyploidy in the angiosperm lineage is due to the group's unique ability to achieve strict bivalent pairing at meiosis, which could be a consequence of very sensitive homology matching (Schwarzacher, 1997). Evidence suggests mediation by cyclin-dependent kinase-like genes (reviewed in Yousafzai *et al.*, 2010).

Recent work by Fawcett *et al.* (2009) and associated commentary by Soltis and Burleigh (2009) has dated whole genome duplication events across 13 diverse angiosperm families to the Cretaceous–Tertiary (K–T) boundary when 60% of plant species went extinct; Fawcett *et al.* (2009) speculate that the new polyploids had a substantial evolutionary advantage over their diploid ancestors (Proost *et al.*, 2011). It will be interesting to see if more recent events are found, or whether polyploidy is ultimately an evolutionary dead-end except following catastrophic climate change. Interestingly, the K–T adaptation through polyploidy seems to be restricted to the angiosperms. The pteridophytes include polyploids and many high chromosome numbers that potentially represent higher ploidies, but the K–T extinction event marked the extinction of the fern forests; in contrast, the gymnosperms survived and remain a very successful group although they include few polyploids (except in the genus *Ephedra*). There are not enough sequence data from these large genomes to identify older polyploids, although the similar and low chromosome number in most gymnosperms provides weak evidence against whole genome duplication.

Chromosome changes and speciation

Occasional chromosomal mutations can become fixed in a population, thus establishing reproductive barriers and leading to the emergence of new species. The diverged species may later form hybrids, often in a limited geographic area, a hybrid or tension zone, where otherwise selectively disadvantaged hybrids with reduced fitness survive in an environment not optimal for either of the parental species (Hewitt, 1988). Analysing the gene flow and differential introgression of genomes in such hybrid zones allows identifying genomic regions involved in speciation (Payseur, 2010). Furthermore, the seemingly random changes found in chromosomal sets of individuals are often of a similar nature to those found between species. They can be seen as the first step in speciation through chromosome evolution.

THE STRUCTURE OF THE CHROMOSOME

Chromosome packaging

The packaging of the double-stranded DNA helix into the nucleosomes is similar in all organisms (Richmond *et al.*, 1984); coiling into the next level of fibre is discussed by Fransz and deJong (2011). Neither the detailed nature nor the consequences of packaging of the DNA fibres into the chromosome at higher levels are clear. Many biology textbooks include diagrams with a hierarchy of coiled-coils, but evidence for this is weak and inconsistent. There are technical reasons why investigation has been difficult, including the fact that the DNA is in a hydrated matrix with salts and proteins which is rapidly disturbed by fixation protocols, while the structures are too polymorphic to be understood by crystallography. However, study of higher levels chromatin packaging, its genetic control and the access by replication, transcription and condensation proteins will lead to better understanding of normal and abnormal nuclear development and the genetic and epigenetic regulation processes.

© 2011 The Authors

Morphological features of chromosomes

In most species, chromosomes have three structural features that have been identified since the earliest microscopy work: the telomeres at the ends of each chromosome, the centromere or primary constriction and, on some chromosomes, a secondary constriction at the nucleolar organizing region (NOR) (Figure 1). Using conventional DNA stain Feulgen these features are particularly well distinguishable (Figure 3d). Chromosome shape is defined by the position of the centromere along its length: it can be at one end of the chromosome (a telocentric chromosome), close to the end (acrocentric), near the middle (metacentric), or somewhere between the physical middle and the end (submetacentric). The description of the chromosome sizes, usually given as measurements of physical length made in a microscope, and the position of the centromeres, gives the karyotype of a species. Karyotypes can include a set of very similar sized chromosomes such as seen in rye and wheat (Figures 3a,c and 4d), but bimodal karyotypes with several large and a number of smaller chromosomes (Figure 3d) are frequently seen.

Telomeres

The Nobel prize-winning work of Blackburn and Szostak discovered that a unique DNA sequence in the telomeres protects the chromosomes from degradation in many species, and confirmed that indeed each chromosome was a single, double-stranded DNA molecule. In work with *A. thaliana*, Richards and Ausubel (1988) showed that chromosomes ended with the repeated 7-bp long DNA motif (TTTAGGG)*n*, which is added by a telomerase enzyme, rather than through semi-conservative replication. This event solves the capping and replication problem of the ends of a DNA double helix (reviews: Fajkus *et al.*, 2005; Watson and Riha, 2010). Because of this mode of addition to chromosomes, the copy number of the repeat unit has been found to vary both between different cells and different chromosomes (Figure 4c; Schwarzacher and Heslop-Harrison, 1991). The repetitive motif is not universal, but a 6 bp motif, as found in many mammals, (TTAGGG)*n* is present in some groups of plants (Sykorova *et al.*, 2003a,b).

Centromeres

The centromere of plant metaphase chromosomes is normally visible as a sharp constriction along its length (Figures 1 and 3a,d), if not present near the end on acrocentric chromosomes. It acts as the focus where the proteinaceous kinetochore plate forms, to which the spindle microtubules attach. The centromeres of most plant species include large arrays of tandemly repeated DNA (Figure 4a; Maluszynska and Heslop-Harrison, 1991; Brandes *et al.*, 1997; Heslop-Harrison *et al.*, 1999) and often retrotransposon sequences (Gindullis *et al.*, 2001; Wolfgruber *et al.*, 2009). Genomic analysis has shown the presence of actively transcribed genes (Jiang *et al.*, 2003; Yan *et al.*, 2006; Mutti *et al.*, 2010). However, despite the conservation of the function, the kinetochore proteins and the CenH3 histone that forms part of the nucleosomes core at centromeres of metaphase chromosomes, the DNA sequences at the centromere in different species are highly diverged and show considerable size variation (Ma *et al.*, 2007). It is now clear that epigenetic mechanisms establish and propagate active centromeres on chromosomes, independent of their sequence (Jiang *et al.*, 2003; Carroll and Straight, 2005; Morris and Moazed, 2007; Wang *et al.*, 2009).

Because of the epigenetic nature of centromeres, it is possible for a chromosome to have a 'neo-centromere' that is not always functional (Carvalho *et al.*, 2008). It is also found that centromeres from one species may not nucleate microtubules strongly in another species background (e.g. Ishi *et al.*, 2010), and hence the chromosomes of one species do not segregate efficiently and are lost (Figure 2). In the hybrid *Hordeum vulgare* × *Hordeum bulbosum*, the chromosomes from many genotypes of *H. bulbosum* are lost during division (Bennett *et al.*, 1976; mechanism investigated by Gernand *et al.*, 2006), giving a haploid *H. vulgare* plant where the chromosome number can be doubled to generate homozygous plants. A very exciting approach to generating haploids came from Ravi and Chan (2010): noting that the centromeres of the eliminated genome were less able to interact with spindle microtubules, they made transgenic Arabidopsis plants with a CenH3 protein modified to be less efficient. When crossed to wild-type plants, chromosomes from the modified genome were eliminated, leading to the formation of haploids.

While the monocentric centromere as above is very widespread in the plant kingdom, two other types of centromere structure have been identified in eukaryotes. The localized point centromere from budding yeast *Saccharomyces cerevisiae*, with a DNA sequence of about 125 bp that provides specific kinetochore protein binding sites (Morris and Moazed, 2007), seems not to have any sequence similarity with the centromeres of plant and animal eukaryotes. The second centromere type is not localized on the chromosome, but functions to allow microtubules to bind along their complete length. The first animal to be fully sequenced, *Caenorhabditis elegans*, had these diffuse or holocentric centromeres, where the microtubules attach along the whole chromosome. Six families of plants (three monocots and three eudicots), have holocentric chromosomes. Nagaki *et al.* (2005) showed that CenH3 was localized along the length of the holocentric chromosomes in *Luzula*. The association of microtubules along the whole chromosome length was observed by Guerra *et al.* (2006) in *Rhynchospora tenuis* (2*n* = 4; Cyperaceae). In this family, chromosome number varies up to 2*n* = circa 200, including many chromosomes <10 Mb in size, suggesting that

© 2011 The Authors

chromosome fragmentation may have occurred during evolution, but the chromosomes are still able to segregate at division by binding microtubules. In contrast to these exceptionally small chromosomes, another genus with holocentric chromosomes, *Cuscuta*, has a large average chromosome size ranging up to 1000 Mb.

The rRNA sites and the nucleolus

As well as the centromeres, another constriction or gap is usually seen on some metaphase chromosomes in a complement – the secondary constriction at the NOR (Figures 1 and 3a, arrow). The NOR corresponds to major sites of the 45S rDNA, consisting of a tandem repeat of a unit with the 18S–5.8S–26S rRNA genes and their transcribed and untranscribed spacer regions (Figure 4b–d). The repeat unit is typically about 10 kb long, and in Arabidopsis it is present about 360 times on two pairs of chromosomes, representing about 5% of the DNA (Copenhaver and Pikaard, 1996; Heslop-Harrison and Maluszynska, 1994). In other species with larger genomes, such as wheat, the rRNA genes are present at a small number of discrete sites on the chromosomes (Figure 4d), with a larger number of copies of the repeat – 1200 at one locus in hexaploid wheat.

At interphase, the nucleolus, the most conspicuous structure within the nucleus, is the site of transcription of the rRNA repeat units and there is little stained DNA within the volume of the nucleolus. Untranscribed copies of the rDNA are often condensed and locate just outside the nucleolus, while *in situ* hybridization shows the transcribed genes as a decondensed thread running through the nucleolus (Figure 4b).

The 18S, 5.8S and 26S rRNA products come together with the 5S rRNA and the ribosomal proteins to make the ribosomes. The 5S rRNA genes, like the 18S–5.8S–26S rRNA genes, are present in the genome as a tandem repeat. Both the 45S and the 5S rRNA loci are often found to have 'rearranged' as blocks during evolution. In *A. thaliana*, the sites of the 5S rDNA are on different chromosomes in the Landsberg and Columbia ecotypes (Murata *et al.*, 1997). In cereals, both the sites of the rDNA and the order of the loci, varies extensively between related species (Castilho and Heslop-Harrison, 1995). Where genetic maps are available, the change in position of the loci is not accompanied by transfer of regions of genes flanking the moved rRNA genes. (Dubcovsky and Dvorak, 1995).

THE CELL CYCLE AND THE INTERPHASE NUCLEUS

The physical structure of the plant cell nucleus changes through the cell cycle (Figure 2). The 'framework' within which these physical events happen can be regarded as the architecture of the nucleus. It is this architecture, in combination with the linear order of genes along the chromosomes, that is responsible for the higher-level organization of the nucleus, and the processes related to interactions between independent molecules or parts of macromolecules. The degree to which this framework involves a physical scaffold or is self-organizing remains uncertain. The processes involved in 'decondensation' of the chromosome to the interphase nucleus are also, in general, poorly understood, although likely to involve loops of chromatin extending from more condensed axes that are visible by light or electron microscopy. During interphase there may be a gradient across the nucleus in the proportion that is filled with chromatin, and chromatin may be more dense adjacent to the nuclear envelope, particularly in species with small genomes. The interphase nucleus itself is a dynamic environment, and both structural components and the DNA move during the interphase. Most obviously, soon after division, rRNA gene expression from multiple chromosomes (the homologous pair if only one pair of sites is present, or sites on several different chromosomes) form individual nucleoli. At later stages of the cell cycle, these have normally moved and fused to a smaller number of larger nucleoli. Interphase nucleus size varies within a single plant: the egg cell is often characterized by a large volume, with the chromatin being much dispersed through the whole volume, while the male sperm cell nucleus is highly condensed (Cao and Russell, 1997; Russell *et al.*, 1996).

In 2003, Cremer and Cremer wrote 'there is increasing agreement that the study of the functional architecture of the eukaryotic nucleus will be one of the most important post-genomic research areas'. Since writing this, chromatin research, involving understanding of the interactions of DNA and proteins has expanded, and the epigenetic consequences of chromatin modification have become clear (see Fransz and deJong, 2011). However, the relationship between nuclear organization, gene expression, higher-order chromatin arrangements and their interactions with other nuclear components, as considered by Cremer and Cremer (2001) remains a challenge to understand. Shopland and Bewersdorf (2008) discuss how recent advances in light microscopy are likely to reveal more information about chromosome structure and function, and point out that relatively little is known about the structural, dynamic, and mechanical properties of these macromolecular assemblies. Figure 6 illustrates the application of super-resolution microscopy to resolve the synaptonemal complex at meiosis, where conventional light microscopy is unable to resolve the two lateral elements that are closer than 300 nm. Gustafsson *et al.* (2008) show that advanced systems have wide application to study chromosomal organization at high resolution, so in great detail.

SEX CHROMOSOMES AND SEX DETERMINATION IN PLANTS

More than 95% of angiosperm and gymnosperm species are hermaphrodite, bearing flowers with both pollen and ovules (as in Arabidopsis or wheat), or monoecious where both

© 2011 The Authors

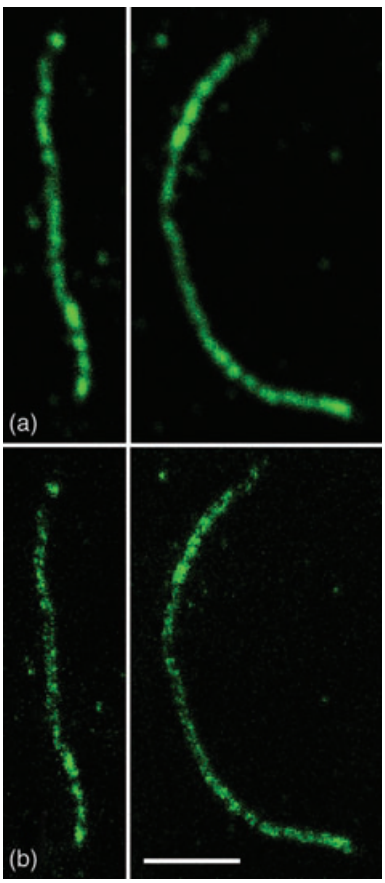

Figure 6. Super resolution microscopy resolves the lateral elements of the synaptonemal complex.
Two synaptonemal complexes of meiotic prophase in the domestic pig (*Sus scrofa domestica)* after immuno-staining with rabbit anti-SCP3 (detected with goat anti-rabbit Alexa 488; green fluorescence) specific for the lateral elements. In imaging with the Leica TCS STED CW in conventional confocal scanning mode (TCS SP5), the two parallel lateral elements that form a gap of 100–300 nm cannot be distinguished (a). Using the same microscope with the super-resolution mode enables imaging below the diffraction limit of light by purely optical methods; the two lateral elements can be seen (b). (Micrographs from Kees Straatman and Trude Schwarzacher who thank Leica Microsystems Milton Keynes UK for use of the microscope).

male and female flowers are carried on the same plant (as in maize) (Dellaporta and Calderon-Urrea, 1993). Some 4% of plants are dioecious, where male and female flowers are carried on different plants and, in most of these, sex is determined genetically. Dioecy is thought to have evolved relatively recently and independently in a number of plant families. In a few cases, dimorphic sex chromosomes were found such as in the 'classic' examples of *Rumex* species and *Silene latifolia*, as well as *Humulus*, *Cannabis* and *Coccinia* (see Figure 3b; Kejnovsky and Vyskot, 2010; Navajas-Pérez *et al.*, 2005, 2009; Vyskot and Hobza, 2004). When cytologically homomorphic sex chromosomes are present, gene differences and sex-determining genes, including a MSY (male specific Y) region are found in male and female plants. Such non-heteromorphic sex-chromosome-like regions have been described in several crop plants whose genomes have been sequenced such as papaya, grape and poplar (grape: Jaillon *et al.*, 2007; papaya, Ming *et al.*, 2008; poplar, Yin *et al.*, 2008), as well as asparagus, kiwi and spinach.

Papaya is trioecious with XX female, XY male, and XYh hermaphrodite (Liu *et al.*, 2004; Zhang *et al.*, 2008). The Y is evolutionarily young and is estimated to have diverged from the X 2–3 million years ago. Within its male specific region, some 13% of the Y, including the centromere and highly methylated heterochromatic knobs have been found (Zhang *et al.*, 2008) and numerous chromosomal rearrangements have been detected (Yu *et al.*, 2008). In poplar, Yin *et al.* (2008) have identified a region of one chromosome showing characteristics of a sex chromosome with a gender-associated locus. Reduced recombination, distorted segregation and haplotype divergence was only observed in the female and consequently sex determination in *Populus* is an incipient ZW chromosome system where males are ZZ and females are the ZW heterogametic sex.

Plant sex chromosome evolution occurred recently, and is still ongoing, so provides an excellent model to study DNA sequence and chromosome evolution. It is believed that the process started with the emergence of sex determining genes (X has male sterility and female fertility; Y has maleness factor and female suppressor) followed by suppression of recombination in their surrounding region (for review see Bergero and Charlesworth, 2009; Kejnovsky and Vyskot, 2010; Navajas-Pérez *et al.*, 2005, 2006). Thus cytological homomorphic sex chromosomes with their heteromorphic DNA regions could represent this first step and are indeed often found to be younger than dimorphic sex chromosomes. The expansion of suppression of recombination to the majority of the chromosome is postulated to lead to accumulation of deleterious mutations, erosion of genes caused by insertion of retroelements or DNA transposons and finally degeneration. As a result heteromorphic sex chromosomes emerge that are often larger than the autosomes in plants (Figure 3b) due to accumulation of repetitive DNA elements (see below) and are in contrast to the small mammalian Ys that are much older and have been allowed to lose genes by rearrangements (Bergero and Charlesworth, 2009).

Molecular investigations have shown that the Y chromosome of *Silene latifolia* estimated to be about 10 million years old shows all of the above signs of sex chromosome evolution including genetic degeneration, reduction of DNA polymorphism, accumulation of mutations at important functional sites coding for proteins, and gene expression changes (see Armstrong and Filatov, 2008; Filatov *et al.*, 2009). Analysis of the repetitive DNA distribution and comparing female and male DNA sequences on *S. latifolia* sex chromosomes, has revealed that parts of the Y

© 2011 The Authors

chromosome have diverged from the X at different times and can be divided into 'strata' similar to the human Y. Different amounts of various DNA sequence families, from almost all classes of repeats known in plants, are present on the Y in large numbers. Cermak *et al.* (2008) undertook a survey of all repeats on the Y of *S. latifolia* and found in decreasing abundance, subtelomeric tandem repeats, gypsy and copia like retroelements, followed by LINEs and SINEs and DNA transposons including hATs and MITES. Interestingly, they and Filatov *et al.* (2009) found a transposable element (TE) abundant on autosomes that is excluded on the Y indicating a divergent evolution of DNA sequences on sex and autosomal chromosomes.

Accumulation of repetitive DNA sequences has also been seen in the genus *Rumex*, which contains several species with dimorphic sex chromosomes and a derived complex XX/XY1Y2 system in *R. acetosa*, *R. papilaris* and *R. hastatulus* (Navajas-Pérez *et al.*, 2006, 2009; see also Figure 3b). The Y degeneration in XX/XY1Y2 system was accompanied by massive accumulation of repetitive DNA followed by chromosomal rearrangements giving rise to the multiple Y chromosomes (Mariotti *et al.*, 2009; Navajas-Pérez *et al.*, 2009). The loss of recombination between X and Y chromosomes would reduce the evolutionary rate of Y-specific satDNAs, but also hinders intra-specific homogenization processes. As a consequence, different rates of evolution have been found for autosomal and sex chromosome variants of repeats, and differential patterns of Y-heterochromatin as well as the presence of different subfamilies and related satDNAs in different regions of the Y chromosomes (Mariotti *et al.*, 2009; Navajas-Pérez *et al.*, 2006, 2009). Further the Y chromosome experienced many inversions of various extents.

Additional evidence of repeat accumulation at different times during the evolution of the Y chromosomes, comes from the studies of simple sequence repeats that have accumulated in the Y chromosome of *Silene* especially in the longer arm which has stopped recombining relatively recently and harbours no other repeats yet (Kejnovsky *et al.*, 2009). In *Rumex acetosa* several simple sequence repeats including (ACC) (see Figure 3b; Karwur, 2001) are found highly amplified throughout both Y chromosomes except towards one telomere, presumably the pseudoautosomal regions. The autosomes and X chromosome show much lower levels with several distinct bands along most chromosomes similar to the pattern found in wheat and rye chromosomes (see Figure 3c; Cuadrado and Schwarzacher, 1998).

THE SIGNIFICANCE OF CHROMOSOME ORGANIZATION

The chromosome is a key level of organization of the plant genome, providing the structure for the genetic linkage groups, allowing replication, transcription and transmission of the genome, and allowing whole genome duplication and physical reorganization. Following completion of the Arabidopsis and other genome sequences, the widespread presence of segmental and whole genome duplications across angiosperms is much more frequent than was suspected from earlier studies. Comparative genomics using whole genome sequencing complemented by molecular cytogenetics has provided new insights into the nature of chromosomal rearrangements including fusions, fissions, inversions, deletions and duplications, across a much wider groups of plants than has been possible with cytogenetic approaches alone. These episodic events combine with continuous processes including sequence mutation, transposable element accumulation, tandem repeat amplification and sequence homogenization. Improved methods of chromosomal analysis with *in-situ* hybridization and use of antibodies are assisting characterization of genome-wide and chromosome-level changes in the genome. The fundamental insights gained from these studies are now showing how genomes evolve and how diversity can be generated.

So far, the controls on many features of chromosome organization and their variability remain to be elucidated. Why should different species have genomes varying in size by more than 2000-fold, and both chromosome number and chromosome sizes vary by 300-fold? The behaviour of these genomes seems to be similar in terms of replication, gene expression, control and evolution, or at least differences do not reflect the huge variation in genome organization. Indeed, it is remarkable that the same genetic, segregation, expression, replication and evolutionary mechanisms seem to be applicable over this large range. Crop plants represent an intensively selected subset of <0.1% of the 400 000 angiosperm species, and fewer than 30 species provide more than 97% of the world's food (FAOstat, 2010). Even among the top crops, the variation in nature of genomes is evident with diploids, recent polyploids, and hybrid species, and genome sizes between 465 Mb in rice to 17 000 Mb in wheat. Exploiting the diversity and evolutionary mechanisms in plant genomes is likely to be a key to crop development for food production.

REFERENCES

Aguiar-Perecin, ML.R. and de Vosa, C.G. (1985) C-banding in maize. II. Identification of somatic chromosomes. *Heredity*, **54**, 37–42.

Alix, K., Joets, J., Ryder, C., Moore, J., Barker, G., Bailey, J., King, G. and Heslop-Harrison, P. (2008) The CACTA transposon Bot1 played a major role in *Brassica* genome divergence and gene proliferation. *Plant J.* **56**, 1030–1044.

Alkhimova, O.G., Mazurok, N.A., Potapova, T.A., Zakian, S.M., Heslop-Harrison, J.S. and Vershinin, A.V. (2004) Diverse patterns of the tandem repeats organization in rye chromosomes. *Chromosoma*, **113**, 42–52.

Angiosperm Phylogeny Group III (2009) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Bot. J. Linn. Soc.* **161**, 105–121.

Anhalt, U.C.M., Heslop-Harrison, J.S., Byrne, S., Guillard, A. and Barth, S. (2008) Segregation distortion in *Lolium:* evidence for genetic effects. *Theor. Appl. Genet.* **117**, 297–306.

© 2011 The Authors

Armstrong, S.J. and Filatov, D.A. (2008) A cytogenetic view of sex chromosome evolution in plants. *Cytogenet. Genome Res.* **120**, 241–246.

Ayala-Navarrete, L., Bariana, H., Singh, R., Gibson, J., Mechanicos, A. and Larkin, P. (2007) Trigenomic chromosomes by recombination of *Thinopyrum intermedium* and *Th. ponticum* translocations in wheat. *Theor. Appl. Genet.* **116**, 63–75.

Bardsley, D., Cuadrado, A., Jack, P., Harrison, G., Castilho, A. and Heslop-Harrison, J.S. (1999) Chromosome markers in the tetraploid wheat *Aegilops ventricosa* analysed by *in situ* hybridization. *Theor. Appl. Genet.* **99**, 300–304.

Barkworth, M. (2010) *Triticeae*. Modified by Barkworth, M. (Barkworth, M. ed.), Flora of North America, Vol. 24/25.

Beló, A., Beatty, M., Hondred, D., Fengler, K., Li, B. and Rafalski, A. (2010) Allelic genome structural variations in maize detected by array comparative genome hybridization. *Theor. Appl. Genet.* **120**, 355–367.

Bennett, M.D. and Leitch, I.J. (2011) Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Ann. Bot.* **107**, 467–590.

Bennett, M.D., Finch, R.A. and Barclay, I.R. (1976) The time, rate and mechanism of chromo some elimination in *Hordeum* hybrids. *Chromosoma*, **54**, 175–200.

Bergero, R. and Charlesworth, D. (2009) The evolution of restricted recombination in sex chromosomes. *Trends Ecol. Evol.* **24**, 94–102.

Biemont, C. (2008) Genome size evolution: within-species variation in genome size. *Heredity*, **101**, 297–298.

Blakeslee, A.F. and Avery, B.T. (1919) Mutations in the jimson weed. *J. Hered.* **10**, 111–120.

Bock, R. and Timmis, J.N. (2008) Reconstructing evolution: gene transfer from plastids to the nucleus. *Bioessays*, **30**, 556–566.

Bowler, C. and Tirichine, L. (2011) Decoding algal genomes: tracing back the history of photosynthetic life on Earth. *Plant J.* **66**, 45–58.

Brandes, A., Thompson, H., Dean, C. and Heslop-Harrison, J.S. (1997) Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. *Chromosome Res.* **5**, 238–246.

Cao, Y. and Russell, S.D. (1997) Mechanical isolation and ultrastructural characterization of viable egg cells in *Plumbago zeylanica*. *Sex. Plant Reprod.* **10**, 368–373.

Carroll, C.W. and Straight, A.F. (2005) Centromere formation: from epigenetics to self-assembly. *Trends Cell Biol.* **16**, 70–78.

Carvalho, A., Guedes-Pinto, H., Heslop-Harrison, J. and Lima-Brito, J. (2008) Wheat neocentromeres found in F1 *Triticale* × *Tritordeum* hybrids (AABBRH) after 5-azacytidine treatment. *Plant Mol. Biol. Rep.* **26**, 46–52.

Castilho, A. and Heslop-Harrison, J.S. (1995) Physical mapping of 5S and 18S–25S rDNA and repetitive DNA sequences in *Aegilops umbellulata*. *Genome*, **38**, 91–96.

Castilho, A., Vershinin, A. and Heslop-Harrison, J.S. (2000) Repetitive DNA and the chromosomes in the genome of oil palm (*Elaeis guineensis*). *Ann. Bot.* **85**, 837–844.

Cermak, T., Kubat, Z., Hobza, R., Koblizkova, A., Widmer, A., Macas, A., Vyskot, B. and Kejnovsky, E. (2008) Survey of repetitive sequences in Silene latifolia with respect to their distribution on sex chromosomes. *Chromosome Res.* **16**, 961–976.

Chaves, R., Adega, F., Heslop-Harrison, J.S., Guedes-Pinto, H. and Wienberg, J. (2003) Complex satellite DNA reshuffling in the polymorphic *t*(1;29) Robertsonian translocation and evolutionarily derived chromosomes in cattle. *Chromosome Res.* **11**, 641–648.

Contento, A., Heslop-Harrison, J.S. and Schwarzacher, T. (2005) Diversity of a major repetitive DNA sequence in diploid and polyploid Triticeae. *Cytogenet. Genome Res.* **109**, 34–42.

Copenhaver, G.P. and Pikaard, C.S. (1996) Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. *Plant J.* **9**, 273–282.

Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292–301.

Cremer, T. and Cremer, C. (2003) Why study nuclear architecture? *Chromosome Res.* **11**, 385–386.

Cuadrado, A. and Schwarzacher, T. (1998) The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosoma*, **107**, 587–594.

Cullis, C.A. (2005) Mechanisms and control of rapid genomic changes in flax. *Ann. Bot.* **95**, 201–206.

Cullis, C.A., Vorster, B.J., Van Der Vyver, C. and Kunert, K.J. (2009) Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? *Ann. Bot.* **103**, 625–633.

Dellaporta, S.L. and Calderon-Urrea, A. (1993) Sex determination in flowering plants. *Plant Cell*, **5**, 1241–1251.

Dolezel, J., Kubalakova, M., Paux, E., Bartos, J. and Feuillet, C. (2007) Chromosome-based genomics in cereals. *Chromosome Res.* **15**, 51–66.

Dubcovsky, J. and Dvorak, J. (1995) Ribosomal RNA multigene loci: nomads of the triticeae genomes. *Genetics*, **140**, 1367–1377.

Fajkus, J., Sýkorová, E. and Leitch, A.R. (2005) Telomeres in evolution and evolution of telomeres. *Chromosome Res.* **13**, 469–479.

FAOstat (2010) *Food and Agricultural Commodities Production Figures*. Food and Agriculture Organization of the United Nations. http://faostat.fao.org.

Fawcett, J.A., Maere, S. and van de Peer, Y. (2009) Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl Acad. Sci. USA*, **106**, 5737–5742.

Filatov, D.A., Howell, E.C., Groutides, C. and Armstrong, S.J. (2009) Recent Spread of a retrotransposon in the *Silene latifolia* genome, apart from the Y chromosome. *Genetics*, **181**, 811–817.

Fransz, P. and deJong, H. (2011) From nucleosome to chromosome: a dynamic organization of genetic information. *Plant J.* **66**, 4–17.

Fransz, P.F., Stam, M., Montijn, B., Hoopen, R.T., Wiegant, J., Kooter, J.M., Oud, O. and Nanniga, N. (1996) Detection of single-copy genes and chromosome rearrangements in *Petunia hybrida* by fluorescence *in situ* hybridization. *Plant J.* **9**, 767–774.

Froenicke, L., Caldés, M.G., Graphodatsky, A. *et al.* (2006) Are molecular cytogenetics and bioinformatics suggesting diverging models of ancestral mammalian genomes? *Genome Res.* **16**, 306–310.

Gayral, P. and Iskra-Caruana, M.L. (2009) Phylogeny of banana streak virus reveals recent and repetitive endogenization in the genome of its banana host (*Musa* spp.). *J. Mol. Evol.* **69**, 65–80.

Gernand, D., Rutten, T., Pickering, R. and Houben, A. (2006) Elimination of chromosomes in *Hordeum vulgare* and *H. bulbosum* crosses at mitosis and interphase involves micronucleus formation and progressive heterochromatinization. *Cytogenet. Genome Res.* **114**, 169–174.

Gindullis, F., Desel, C., Galasso, I. and Schmidt, T. (2001) The large-scale organization of the centromeric region in beta species. *Genome Res.* **11**, 253–265.

Goremykin, V.V., Salamini, F., Velasco, R. and Viola, R. (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol. Biol. Evol.* **26**, 99–110.

Graybosch, R.A., Peterson, C.J., Baenziger, P.S. *et al.* (2009) Registration of 'Mace' hard red winter wheat. *J. Plant Registrations*, **3**, 51–56.

Green, B.R. (2011) Chloroplast genomes of photosynthetic eukaryotes. *Plant J.* **66**, 34–44.

Gregor, W., Mette, M.F., Staginnus, C., Matzke, M.A. and Matzke, A.J. (2004) A distinct endogenous pararetrovirus family in *Nicotiana tomentosiformis*, a diploid progenitor of polyploid tobacco. *Plant Physiol.* **134**, 1191–1199.

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J. and Bennett, M.D. (2006) Eukaryotic genome size databases. *Nucleic Acids Res.* **35**, D332–D338.

Greilhuber, J. (2005) Intraspecific variation in genome size in angiosperms: identifying its existence. *Ann. Bot.* **95**, 91–98.

Guerra, M., Brasileiro-Vidal, A.C., Arana, P. and Puertas, M.J. (2006) Mitotic microtubule development and histone H3 phosphorylation in the holocentric chromosomes of Rhynchospora tenuis (Cyperaceae). *Genetica*, **126**, 33–41.

Gustafsson, M.G.L.G., Shao, L., Carlton, P.M.M. *et al.* (2008) Three-dimensional resolution doubling in widefield fluorescence microscopy by structured illumination. *Biophys. J.* **94**, 4957–4970.

Han, Y., Zhang, Y. and Huang, S. (2011) An integrated molecular cytogenetic map of *Cucumis sativus* L. chromosome 2. *BMC Genet.* **12**, 18.

Hansen, C.N., Harper, G. and Heslop-Harrison, J.S. (2005) Characterization of pararetrovirus-like sequences in the genome of potato (*Solanum tuberosum*). *Cytogenet. Genome Res.* **110**, 559–565.

Harper, G., Osuji, J.O., Heslop-Harrison, J.S. and Hull, R. (1998) Integration of banana streak badnavirus into the *Musa* genome: molecular and cytogenetic evidence. *Virology*, **255**, 207–213.

't Hart, H. and Bleij, B. (2003) Sedum. In *Illustrated Handbook of Succulent Plants: Crassulaceae*, Vol. 4 (Eggli, U., ed.). Berlin: Springer, pp. 235–300.

© 2011 The Authors
The Plant Journal © 2011 Blackwell Publishing Ltd,

Helentjaris, T., Weber, D.F. and Wright, S. (1986) Use of monosomics to map cloned DNA fragments in maize. *Proc. Natl Acad. Sci. USA*, **83**, 6035–6039.

Heslop-Harrison, J.S. (1991) The molecular cytogenetics of plants. *J. Cell Sci.* **100**, 15–21.

Heslop-Harrison, J.S. and Maluszynska, J. (1994) The molecular cytogenetics of *Arabidopsis*. In *Arabidopsis* (Meyerowitz, E.M. and Sommerville, C.R., eds). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, pp. 63–87.

Heslop-Harrison, J.S. and Schmidt, T. (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends Plant Sci.* **3**, 195–199.

Heslop-Harrison, J.S., Murata, M., Ogura, Y., Schwarzacher, T. and Motoyoshi, F. (1999) Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes. *Plant Cell*, **11**, 31–42.

Heslop-Harrison, J.S., Leitch, A.R., Schwarzacher, T., Anamthawat-Jónsson, K. (1990) Detection and characterization of 1B/1R translocations in hexaploid wheat. *Heredity*, **65**, 385–392.

Hewitt, G.M. (1988) Hybrid zones – natural laboratories for evolutionary studies. *Trends Ecol. Evol.* **3**, 158–167.

International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, **463**, 763–768.

Ishii, T., Ueda, T., Tanaka, H. and Tsujimoto, H. (2010) Chromosome elimination by wide hybridization between Triticeae or oat plant and pearl millet: pearl millet chromosome dynamics in hybrid embryo cells. *Chromosome Res.* **18**, 821–831.

Jaillon, O., Aury, J.M.M., Noel, B. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Janda, J., Safar, J., Kubalakova, M. *et al.* (2006) Advanced resources for plant genomics: BAC library specific for the short arm of wheat chromosome 1B. *Plant J.* **47**, 977–986.

Jiang, J., Birchler, J.A., Parrott, W.A. and Dawe, R.K. (2003) A molecular view of plant centromeres. *Trends Plant Sci.* **8**, 570–575.

Jones, K. (1998) Robertsonian fusion and centric fission in karyotype evolution of higher plants. *Bot. Rev.* **64**, 273–289.

Jones, R.N., Viegas, W. and Houben, A. (2008) A century of b chromosomes in plants: so what? *Ann. Bot.* **101**, 767–775.

Karwur, F.F. (2001) *Molecular Biology of Chromosomal Sex Determination in Dioecious* Rumex acetosa*, L.* PhD Thesis. University of London, Wye College.

Kejnovsky, E. and Vyskot, B. (2010) *Silene latifolia*: the classical model to study heteromorphic sex chromosomes. *Cytogenet. Genome Res.* **129**, 250–262.

Kejnovsky, E., Hobza, R., Cermak, T., Kubat, Z. and Vyskot, B. (2009) The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity*, **102**, 533–541.

Krutovsky, K.V., Troggio, M., Brown, G.R., Jermstad, K.D. and Neale, D.B. (2004) Comparative mapping in the *Pinaceae*. *Genetics*, **168**, 447–461.

Kubis, S., Schmidt, T. and Heslop-Harrison, J.S. (1998) Repetitive DNA elements as a major component of plant genomes. *Ann. Bot.* **82S**, 45–55.

Lagercrantz, U. and Lydiate, D.J. (1996) Comparative genome mapping in *Brassica*. *Genetics*, **144**, 1903–1910.

Laurie, D.A. and Bennett, M.D. (1985) Nuclear DNA content in the genera *Zea* and *Sorghum*. Intergeneric, interspecific and intraspecific variation. *Heredity*, **55**, 307–313.

Leggett, J.M., Perret, S.J., Harper, J. and Morris, P. (2000) *Chromosoma*l localization of cotransformed transgenes in the hexaploid cultivated oat *Avena sativa* L. using fluorescence *in situ* hybridization. *Heredity*, **84**, 46–53.

Leitch, I.J., Beaulieu, J.M., Chase, M.W., Leitch, A.R. and Fay, M.F. (2010) Genome size dynamics and evolution in monocots. *J. Bot.* **2010**, 1–19.

Liu, Z., Moore, P.H., Ma, H. *et al.* (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature*, **427**, 348–352.

Lockhart, B.E., Menke, J., Dahal, G. and Olszewski, N.E. (2000) Characterization and genomic analysis of tobacco vein clearing virus, a plant pararetrovirus that is transmitted vertically and related to sequences integrated in the host genome. *J. Gen. Virol.* **81**, 1579–1585.

Luo, M.C., Deal, K.R., Akhunov, E.D. *et al.* (2009) Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. *Proc. Natl Acad. Sci. USA*, **106**, 15780–15785.

Lysak, M.A., Koch, M.A., Pecinka, A. and Schubert, I. (2005) Chromosome triplication found across the tribe Brassicae. *Genome Res.* **15**, 516–525.

Lysak, M.A., Berr, A., Pecinka, A., Schmidt, R., McBreen, K. and Schubert, I. (2006) Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc. Natl Acad. Sci. USA*, **103**, 5224–5229.

Lysak, M.A., Koch, M.A., Beaulieu, J.M., Meister, A. and Leitch, I.J. (2009) The dynamic ups and downs of genome size evolution in Brassicaceae. *Mol. Biol. Evol.* **26**, 85–98.

Ma, J., Wing, R., Bennetzen, J. and Jackson, S. (2007) Plant centromere organization: a dynamic structure with conserved functions. *Trends Genet.* **23**, 134–139.

Maluszynska, J. and Heslop-Harrison, J.S. (1991) Localization of tandemly repeated DNA sequences in *Arabidopsis thaliana*. *Plant J.* **1**, 159–166.

Mandakova, T. and Lysak, M.A. (2008) Chromosomal phylogeny and karyotype evolution in $x = 7$ crucifer species (Brassicaceae). *Plant Cell*, **20**, 2559–2570.

Mariotti, B., Manzano, S., Kejnovský, E., Vyskot, B. and Jamilena, M. (2009) Accumulation of Y-specific satellite DNAs during the evolution of *Rumex acetosa* sex chromosomes. *Mol. Genet. Genomics*, **281**, 249–259.

McCouch, S.R., Kochert, G., Yu, Z.H. *et al.* (1988) Molecular mapping of rice chromosomes. *Theor. Appl. Genet.* **76**, 815–829.

Meyerowitz, E.M. (1989) *Arabidopsis*, a useful weed. *Cell*, **56**, 263–269.

Ming, R., Hou, S., Feng, Y. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* L.). *Nature*, **452**, 991–996.

Molnár, I., Cifuentes, M., Schneider, A., Benavente, E. and Molnár-Láng, M. (2011) Association between simple sequence repeat-rich chromosome regions and intergenomic translocation breakpoints in natural populations of allopolyploid wild wheats. *Ann. Bot.* **107**, 65–76.

Morris, C.A. and Moazed, D. (2007) Centromere assembly and propagation. *Cell*, **128**, 647–650.

Moscone, E.A., Samuel, R., Schwarzacher, T., Schweizer, D. and Pedrosa-Harand, A. (2007) Complex rearrangements are involved in *Cephalanthera* (Orchidaceae) chromosome evolution. *Chromosome Res.* **15**, 931–943.

Murata, M., Heslop-Harrison, J.S. and Motoyoshi, F. (1997) Physical mapping of the 5S ribosomal RNA genes in *Arabidopsis thaliana* by multicolor fluorescence *in situ* hybridization with cosmid clones. *Plant J.* **12**, 31–37.

Murray, B.G., Friesen, N. and Heslop-Harrison, J.S. (2002) Molecular cytogenetic analysis of *Podocarpus* and comparison with other gymnosperm species. *Ann. Bot.* **89**, 483–489.

Mutti, J., Sandhu, D., Sidhu, D. and Gill, K. (2010) Dynamic nature of a wheat centromere with a functional gene. *Mol. Breed.* **26**, 177–187.

Nagaki, K., Kashihara, K. and Murata, M. (2005) Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant *Luzula nivea*. *Plant Cell*, **17**, 1886–1893.

Nagarajan, S., Rens, W., Stalker, J., Cox, T. and Ferguson-Smith, M. (2008) Chromhome: a rich Internet application for accessing comparative chromosome homology maps. *BMC Bioinformatics*, **9**, 168.

Navajas-Pérez, R., de la Herran, R., Lopez Gonzalez, G. *et al.* (2005) The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastidial sequence data. *Mol. Biol. Evol.* **22**, 1929–1939.

Navajas-Pérez, R., Schwarzacher, T., de la Herrán, R., Ruiz Rejón, C., Ruiz Rejón, M. and Garrido-Ramos, M.A. (2006) The origin and evolution of the variability in a Y-specific satellite-DNA of *Rumex acetosa* and its relatives. *Gene*, **368**, 61–71.

Navajas-Pérez, R., Schwarzacher, T., Ruiz Rejón, M.R. and Garrido-Ramos, M.A. (2009) Molecular cytogenetic characterization of *Rumex papillaris*, a dioecious plant with an XX/XY_1Y_2 sex chromosome system. *Genetica*, **135**, 87–93.

Özkan, H., Tuna, M., Kilian, B., Mori, N. and Ohta, S. (2010) Genome size variation in diploid and tetraploid wild wheats. *AoB Plants*, doi:10.1093/aobpla/plq015.

Paux, E., Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P. and Feuillet, C. (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J.* **48**, 463–474.

© 2011 The Authors

Payseur, B.A. (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Mol. Ecol. Resour.* **10**, 806–820.

Pearce, S.R., Harrison, G., Heslop-Harrison, J.S., Flavell, A.J. and Kumar, A. (1997) Characterisation and genomic organisation of Ty1-copia group retrotransposons in rye (*Secale cereale*). *Genome*, **40**, 617–625.

Pedersen, C., Zimny, J., Becker, D., Jähne-Gärtner, A. and Lörz, H. (1997) Localization of introduced genes on the chromosomes of transgenic barley, wheat and triticale by fluorescence *in situ* hybridization. *Theor. Appl. Genet.* **94**, 749–757.

Proost, S., Pattyn, P., Gerats, T. and van der Peer, Y. (2011) Journey through the past: 150 million years of plant genome evolution. *Plant J.* **66**, 58–65.

Ravi, M. and Chan, S.W.L. (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature*, **464**, 615–618.

Richards, E. and Ausubel, F.M. (1988) Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, **53**, 127–136.

Richert-Pöggeler, K.R., Noreen, F., Schwarzacher, T., Harper, G. and Hohn, T. (2003) Induction of infectious *Petunia* vein clearing (pararetro) virus from endogenous provirus in, petunia. *EMBO J.* **22**, 4836–4845.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D. and Klug, A. (1984) Structure of the nucleosome core particle at 7 Å resolution. *Nature*, **311**, 532–537.

Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature*, **182**, 713–715.

Russell, S.D., Strout, G.W., Stramski, A.K., Mislan, T.W., Thompson, R.A. and Schoemann, L.M. (1996) Development polarization and morphogenesis of the generative and sperm cells of *Plumbago zeylanica.* 1. Descriptive cytology and three-dimensional organization. *Am. J. Bot.* **83**, 1435–1453.

Šafář, J., Bartoš, J., Janda, J. *et al.* (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant J.* **39**, 960–968.

Salvo-Garrido, H., Travella, S., Schwarzacher, T., Harwood, W.A. and Snape, J.W. (2001) An efficient method for the physical mapping of transgenes in barley using *in situ* hybridization. *Genome*, **44**, 104–110.

Schnable, P.S., Ware, D., Fulton, R.S. *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112–1115.

Schwarzacher, T. (1997) Three stages of meiotic homologous chromosome pairing in wheat: cognition, alignment and synapsis. *Sex. Plant Reprod.* **10**, 324–331.

Schwarzacher, T. (2008) Fluorescent *in situ* hybridization to detect transgene integration into plant genomes. *Methods Mol. Biol.* **478**, 227–246.

Schwarzacher, T. and Heslop-Harrison, J.S. (1991) *In situ* hybridization to plant telomeres using synthetic oligomers. *Genome*, **34**, 317–323.

Schwarzacher, T., Anamthawat-Jónsson, K., Harrison, G.E. *et al.* (1992) Genomic *in situ* hybridization to identify alien chromosomes and chromosome segments in wheat. *Theor. Appl. Genet.* **84**, 778–786.

Sepsi, A., Molnár, I., Szalay, D. and Molnár-Láng, M. (2008) Characterization of a leaf rust-resistant wheat *Thinopyrum ponticum* partial amphiploid BE-1, using sequential multicolor GISH and FISH. *Theor. Appl. Genet.* **116**, 825–834.

Sharp, P.J., Chao, S., Desai, S. and Gale, M.D. (1989) The isolation, characterization and application in the *Triticeae* of a set of wheat RFLP probes identifying each homoeologous chromosome arm. *Theor. Appl. Genet.* **78**, 342–348.

Shopland, L. and Bewersdorf, J. (2008) Seeing the world through a new set of glasses: emerging technologies for the study of cell nuclei and chromosomes. *Chromosome Res.* **16**, 347–349.

Soltis, D.E. and Burleigh, J.G. (2009) Surviving the K–T mass extinction: new perspectives of polyploidization in angiosperms. *Proc. Natl Acad. Sci. USA*, **106**, 5455–5456.

Somerville, C. (1989) *Arabidopsis* blooms. *Plant Cell*, **1**, 1131–1135.

Spector, D.L. (2003) The dynamics of chromosome organization and gene regulation. *Annu. Rev. Biochem.* **72**, 573–608.

Staginnus, C. and Richert-Pöggeler, K.R. (2006) Endogenous pararetroviruses: two-faced travelers in the plant genome. *Trends Plant Sci.* **11**, 485–491.

Staginnus, C., Gregor, W., Mette, M.F., Teo, C.H., Borroto-Fernández, E.G., Machado, M.L., Matzke, M. and Schwarzacher, T. (2007) Endogenous pararetroviral sequences in tomato (*Solanum lycopersicum*) and related species. *BMC Plant Biol.* **7**, 24.

Steinitz-Sears, L.M. (1963) Chromosome studies in *Arabidopsis thaliana*. *Genetics*, **48**, 483–490.

Suda, J. and Leitch, I.J. (2010) The quest for suitable reference standards in genome size research. *Cytometry*, **77A**, 717–720.

Svitashev, K.S. and Somers, D.A. (2002) Characterization of transgene loci in plants using FISH: a picture is worth a thousand words. *Plant Cell Tissue Organ Cult.* **69**, 205–214.

Swift, H. (1950) The constancy of deoxyribose nucleic acid in plant nuclei. *Proc. Natl Acad. Sci. USA*, **36**, 643–654.

Sykorova, E., Lim, K.Y., Chase, M.W., Knapp, S., Leitch, I.J., Leitch, A.R. and Fajkus, J. (2003a) The absence of *Arabidopsis*-type telomeres in *Cestrum* and closely related genera *Vestia* and *Sessea* (Solanaceae): first evidence from eudicots. *Plant J.* **34**, 283–291.

Sykorova, E., Lim, K.Y., Kunicka, Z. *et al.* (2003b) Telomere variability in the monocotyledonous plant order *Asparagales*. *Proc. R. Soc. Lond. B Biol. Sci.* **270**, 1893–1904.

Taketa, S., Harrison, G.E. and Heslop-Harrison, J.S. (1999) Comparative physical mapping of the 5S and 18S–25S rDNA in nine wild *Hordeum* species and cytotypes. *Theor. Appl. Genet.* **98**, 1–9.

Tang, H., Bowers, J.E., Wang, X. and Paterson, A.H. (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl Acad. Sci. USA*, **107**, 1472–1477.

Trick, M., Kwon, S.J.J., Choi, S.R.R. *et al.* (2009) Complexity of genome evolution by segmental rearrangement in *Brassica rapa* revealed by sequence-level analysis. *BMC Genomics*, **10**, 539.

Vaughan, H.E., Heslop-Harrison, J.S. and Hewitt, G.M. (1999) The localisation of mitochondrial sequences to chromosomal DNA in *Orthopterans*. *Genome*, **42**, 874–880.

Vyskot, B. and Hobza, R. (2004) Gender in plants: sex chromosomes are emerging from the fog. *Trends Genet.* **20**, 432–438.

Wang, M.L., Leitch, A.R., Schwarzacher, T., Heslop-Harrison, J.S. and Moore, G. (1992) Construction of a chromosome-enriched HpaII library from flow-sorted wheat chromosomes. *Nucleic Acids Res.* **20**, 1897–1901.

Wang, G., Zhang, X. and Jin, W. (2009) An overview of plant centromeres. *J. Genet. Genomics*, **36**, 529–537.

Watson, J.M. and Riha, K. (2010) Comparative biology of telomeres: where plants stand. *FEBS Lett.* **584**, 3752–3759.

Wolfgruber, T.K., Sharma, A., Schneider, K.L. *et al.* (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic loci shaped primarily by retrotransposons. *PLoS Genet.* **5**, e1000743.

Wolters, A.-M.A., Trindade, L.M., Jacobsen, E. and Visser, R.G.F. (1998) Fluorescence *in situ* hybridization on extended DNA fibres as a tool to analyse complex T-DNA loci in potato. *Plant J.* **13**, 837–847.

Xiong, Z. and Pires, J.C. (2011) Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. *Genetics*, **187**, 37–49.

Xu, L., Chen, H., Hu, X., Zhang, R., Zhang, Z. and Luo, Z.W. (2006) Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Mol. Biol. Evol.* **23**, 1107–1108.

Yan, H., Ito, H., Nobuta, K. *et al.* (2006) Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. *Plant Cell*, **18**, 2123–2133.

Yin, T., Difazio, S.P., Gunter, L.E. *et al.* (2008) Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res.* **18**, 422–430.

Yousafzai, F., Al-Kaff, N. and Moore, G. (2010) The molecular features of chromosome pairing at meiosis: the polyploid challenge using wheat as a reference. *Funct. Integr. Genomics*, **10**, 147–156.

Yu, Q., Hou, S., Feltus, F.A. *et al.* (2008) Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant J.* **53**, 124–132.

Zambryski, P., Joos, H., Genetello, C., Leemans, J., Montagu, M.V. and Schell, J. (1983) Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. *EMBO J.* **2**, 2143–2150.

Zhang, W., Wang, X., Yu, Q., Ming, R. and Jiang, J. (2008) DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res.* **18**, 1938–1943.

© 2011 The Authors