Heslop-Harrison JS. 2012. Genome evolution: extinction, continuation or explosion? Current Opinion in Plant Biology 15:115–121
http://dx.DOI.org/10.1016/j.pbi.2012.03.006

This review comes from a themed issue of COPB on Genome Studies and Molecular Genetics Edited by Yves Van de Peer and J. Chris Pires Published by Elsevier Ltd. -

Author's pre-print[i] version (with Harvard, not numeric, references)

Genome evolution: extinction, continuation or explosion?

JS (Pat) Heslop-Harrison, Department of Biology, University of Leicester, Leicester LE1 7RH, UK. phh4@le.ac.uk

Highlights

- Genome-scale evolution involves mutation, chromosomal rearrangements, hybridization and polyploidy

- Repeated sequences can be localized or dispersed in the genome and make up most of the DNA

- Evolutionary processes may be continuous or episodic and have contrasting long and short-term consequences

- Sequences, synthetic hybrids, comparative genomics and modelling link genome behaviour and consequences

- Understanding genome evolution is critical for biodiversity conservation and breeding sustainable crops

Abstract

Darwin recognized the processes of speciation and the extinctions of species. We now understand many of the genome-scale processes occurring during evolution involving mutations, amplification, loss or homogenization of sequences; rearrangement, fusion and fission of chromosomes; and horizontal transfer of genes or genomes through polyploidy or other mechanisms. DNA sequence information, combined with appropriate informatic tools and experimental approaches including generation of synthetic hybrids, comparison of genotypes across environments, and modelling of genomic responses, is now letting us link genome behaviour with its consequences. The understanding of genome evolution will be of critical value both for conservation of the biodiversity of the plant kingdom and addressing the challenges of breeding new and more sustainable crops to feed the human population.

Evolution and phylogenies

Darwin (1859), in the only figure in "*The Origin of Species*" (fig. 1a) recognized that species arose from existing species and that extinction was occurring continuously, while other species remained extant or gave rise to new species. His diagram indicated the origin of very variable numbers of taxa from a single ancestor, and here I aim to overview some of the events of genome evolution which have the consequence of continuation or extinction. In the last decade, based on reassessment of morphological data in the light of molecular work, the evolutionary relationships of all seed plants have become clear, and the latest revision of angiosperm taxonomy (**APG, 2009) gives the phylogenetic position of almost all taxa down to the level of families. As shown by Darwin, a species and its lineage can become extinct, or a species can diverge into numerous new taxa. Between extremes, a species can remain for an extended evolutionary time without speciation or extinction. Speciation involves a range of changes at the genome or DNA level which are usually heritable. Each evolutionary change in the DNA in a genome is, by its nature, saltatory, involving a jump from one condition to another. When there are large numbers of changes, together they can be treated as a continuous character, widely used in analysis of evolution of single copy DNA from nuclear or organellar genomes (eg *China Plant BOL Group, 2011). Analysis of silent mutations in the third position of codons in coding regions of genes in particular can be treated as fulfilling largely the assumptions of being selectively neutral, and occurring only once during evolution (neither being reiterated nor reverting, so all descendents of the individual with the change carry the new sequence variant). Analysis of such sequences has underpinned the robust and deep phylogenies developed by the Angiosperm Phylogeny Group (APG, 2009), with monophyletic branches, where all families below a node have arisen from a single common ancestor, allowing the order of diversification events to be accurately determined and the ages inferred (Bell et al., 2010).

Genome evolution occurs at a range of different levels. Sequence mutations, involving a base pair change, an insertion or deletion, are used in the analyses discussed above. Where these changes occur in a coding position or regulatory sequence, the mutation will often immediately change the phenotype of the plant. But beyond the relatively simple mutation, other processes including transfer of genes and genomes between lineages, amplification and mobility of DNA elements within genomes, and amplification and homogenization of tandemly repeated DNA sequences play critical roles in generation of new species and genome evolution. All these processes contribute to genome variation, enabling the Darwinian processes leading to "*The Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*".

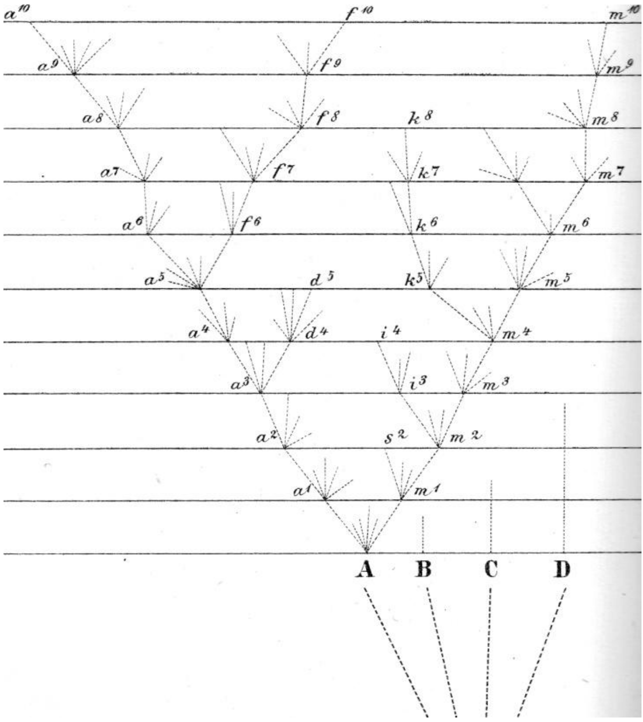

Figure 1. Part of the only figure in *The Origin of Species* (Darwin, 1859). The concepts of separation of species, with many becoming extinct and others separating into many other species, along with the monophyletic origins of branches, is the basis of our current models of phylogeny and genome evolution.

Horizontal gene transfer

Monophyletic classifications cannot account for reticulation events, where genes or genomes from one species are transferred from one branch to another branch, or two branches come together. The common ancestor of all plants includes the mitochondrial and plastid genomes that were incorporated into the cells (*Keeling and Palmer, 2008) in an endosymbiotic relationship. The transfer of functional genes from these organellar genomes into to the nucleus of cells, where they remain functional, is a process that is an ongoing evolutionary process today (*Cullis et al. 2009; Stegemann et al., 2012). The presence of individual transferred genes – a unidirectional process although having the potential to be repeated – provides a valuable marker for lineages with a monophyletic origin. It is also clear that larger or smaller stretches of organelles may be incorporated into the nuclear genomes of plants where they are not functional; given that sequence data is often masked for organellar sequences, the incorporation of these sequences may not be revealed in assemblies, while other phylogenies may be erroneous if nuclear copies rather than true-organelle copies of genes are analysed (Vaughan et al. 1999).

Lateral (or horizontal) gene transfer from one organism to a genome of another is relatively rare although, when it occurs, an important feature of evolution. The best known example is perhaps infection of plants by *Agrobacterium tumefaciens*, involving transfer of the Ti-plasmid with several genes to the host plant; normally the transformed cells proliferate but do not regenerate to new plants and so have no evolutionary consequence. In the plant breeding context, the use of *Agrobacterium* for transformation has been established with incorporation of desirable genes for agriculture or for study of gene function and regulation, as pioneered by Schell, van Montague and colleagues (eg Zambryski et al., 1983). Stevens (*2011) reviews evidence for lateral gene transfer between unrelated plant nuclear genomes, and between parasites and hosts. Horizontal transfer of genes can introduce novel and advantageous characters into the host species; a clear example is perhaps the transfer of segments of DNA of viral origin into the nuclear genome that has been observed in several wild species (Bejarano et al., 1996; Harper et al., 1998). Infection and expression of the virus may be regulated and prevented by an RNA-mediated silencing mechanism (Staginnus et al., 2007), a mechanism exploited in new breeding strategies for virus resistance.

Whole genome transfer and duplication

Hybridization of two different species leading to amphipolyploidy (or allopolyploidy) occurs widely in the plant kingdom, giving new species with the addition of genomes from independent lineages, a reticulation not seen in the diagram of Darwin (Figure 1). Where new species have been generated by the hybridization, it has usually been followed by a doubling of the number of chromosomes so each genome is present in two copies, allowing pairing of chromosomes and regular meiosis. Whole genome duplication within a species, giving autopolyploid plants or species, is also frequent. Autopolyploidy can occur by a range of mechanisms, including non-segregation of chromosomes during mitosis or at meiosis, leading to 2n gamete formation. In establishing monophyly in phylogenetics, Stevens (2011) suggested that genera often can be defined at a level above that at which hybridization is common. Given that hybridization at the level of taxonomic family and above is virtually unknown, the data do not affect monophyly at these levels. It is now well established that whole genome duplications have occurred multiple times through evolution (**van der Peer, 2009). Presumably, at the time of their occurrence, many of these ancient polyploidy events would have been considered as amphiploidy (from hybridization between two recently diverged species), rather than autopolyploidy (chromosome number doubling) events,. Although the evidence of genome duplication remains strong today, subsequent evolution – mutation, gain, loss and homogenization of DNA sequences – will have removed the signature of sub-generic levels of hybridity. In ancient amphiploids, the ancestral genomes become more similar to each other by homogenization of their repetitive DNA sequences (whether directional or not), a phenomenon that can be seen in tetraploid hybrid species of various ages (Patel et al. 2011; Koukalova et al. 2010).

Genome sequencing since 2000 has given novel insight into the occurrence of whole genome duplications (WGDs). With extensive sequence information, the signatures of ancient duplications, back to the origin of angiosperms and beyond, can be detected, but it has been found that there are only a limited number ancient (family level or above) duplications in the angiosperm phylogeny (*Tang et al., 2008). Van de Peer et al. (**2009) have suggested that most genome duplication events are evolutionary dead ends. They show that the signatures of those occurring near times of mass extinctions, such as the Cretaceous-Tertiary KT transition 50 to 60 million years ago (Mya), are widespread in extant plant lineages (van de Peer et al., 2010). Despite the presence of polyploids in most lineages during the last c. 10Mya, and the divergence of families, species and genera throughout the Tertiary and Quaternary, sequence analysis has not identified WGDs between the KT and recent periods.

Over shorter timescales, many authors have discussed the apparent advantages of polyploidy in terms of the genetic fixation of combinations of gene alleles and giving freedom for mutation and hence gain of new functions within duplicated loci, although the data suggest these are mostly of only rare long-term evolutionary significance leading to new lineages. The three most important crop plants for human nutrition are the grasses rice, maize and bread wheat, sharing a whole genome duplication more than 50 Mya, and having no further WGD (rice), a WGD 5 Mya (maize) (Salse et al., 2008), or polyploid hybrid speciation less than 50,000 ya (wheat). . The near-equal success of these three graminaceous crops under intensive selection from the start of agriculture 10,000 years ago until today suggests, perhaps surprisingly, that there are no overwhelming advantages or disadvantages of recent polyploidy over this timescale and selective conditions (Heslop-Harrison and Schwarzacher, 2012).

Chromosome evolution

The type of interspecific hybrids discussed above can also allow horizontal transfer of chromosomes from one species to another. Making such hybrids and backcrossing is a valuable process in plant breeding to allow introgression of characters from a wild species into a cultivar, and the gene alleles from one species are replaced by those from another. Graybosch et al. (2009) give an example of a released wheat variety involving hybridization of wheat with *Thinopyrum intermedium* and backcrossing to wheat, transferring a whole chromosome arm (Figure 2). Other changes in genomes at the chromosomal level involve rearrangements including duplications, deletions, fissions, fusions, translocations and inversions of whole chromosomes, chromosome arms or smaller segments (Heslop-Harrison and Schwarzacher, 2011). These chromosomal events, where they are present in the pollen or egg cells and do not cause sterility, can become homozygous and then generate a new, reproductively isolated line that can become a new species.

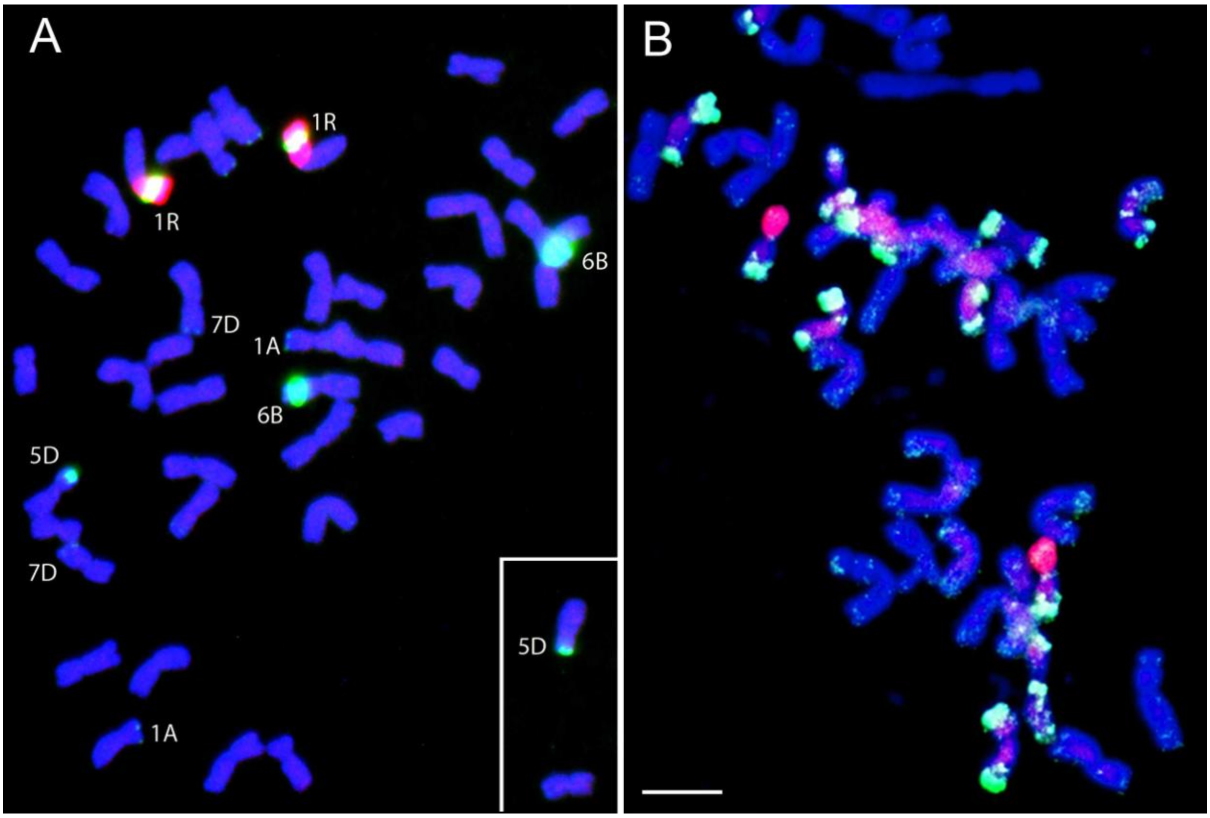

Figure 2. Two metaphases of bread wheat, *Triticum aestivum* (2n=6x=42) with chromosome arms introgressed from other genera in the grass tribe Triticeae, both from successful wheat varieties grown commercially. The alien chromosome arm is labelled in red by in situ hybridization with total genomic DNA from the alien species. A) Wheat variety Beaver with a 1BL.1RS translocation incorporating a chromosome arm from rye, *Secale cereale*; the second green probe labels the rDNA at the nucleolar organizers on chromosomes 1R, 6B, 1A, 5D and 7D. B) The variety Mace (Graybosch et al., 2009) with a 4DL.4AgS translocation. The green probe shows sites of a tandemly repeated DNA sequence dpTa1. Scale bar 10 µm. Figures courtesy of Niaz Ali, Robert Graybosch and Trude Schwarzacher.

The family Brassicaceae, with very variable chromosome numbers, gives excellent examples of the nature and mechanisms of chromosome number changes. Following whole genome duplication Mandáková et al. (2010) have shown the extensive chromosome number reduction and rearrangement that occurs by fusions, translocations and inversions within a group of Australian species. The chromosomal reshuffling over a few generations has been detected in newly made polyploid *Brassica* interspecific hybrids between *B. oleracea* and *B. rapa* (Gaeta et al., 2007), where spontaneous loss of chromosomal material was a feature seen before stabilization of the new hybrid genome combinations, generating new variation in the polyploids.

Repetitive DNA: Tandem repeat evolution

Repetitive DNA is a class of DNA where a particular sequence motif is repeated many times in the genome. Such motifs may represent several percent of all the DNA in the genome (Heslop-Harrison & Schwarzacher, 2011), and they may occur in tandem arrays at a small number of loci in the genome, where copies are adjacent to each other (see green probes in Figure 2). Some such sequences encode genes of importance – the ribosomal

DNA encoding the 45S rRNA and the 5S rRNA are examples, and even in species with very small genomes, such as *Arabidopsis thaliana,* rRNA genes may represent several percent of the genome. Although the sequence of the rRNA genes themselves is highly conserved, the spacers between the genes, both non-transcribed, and, like the internal transcribed spacer (ITS; China Plant BOL Group 2011) evolve rapidly enough that their sequence provides a valuable marker for species divergence. Notably, the consequence of sequence homogenization, presumably involving mechanisms such as slippage replication or recombination with multiplication and deletion of array elements (see Kuhn et al., 2009). leads to identity between monomers within the rDNA arrays in a species.

There may be many different and unrelated non-coding tandemly arrayed DNA motifs in a genome, with each copy present at one or more centromeric, sub-telomeric or intercalary regions of chromosomes. Examples are known where the arrays are substantially different between species, but homogenized within species (eg in the genus *Arabidopsis*; Heslop-Harrison et al., 2003), or represent the same range of variation over a number of different species from both plants (Contento et al., 2009) and animals (where there is more appropriate sequence data available so homogenization processes can be examined in arrays of different sizes; Kuhn et al. 2011). Thus the evolutionary history of different tandem repeat families differs; furthermore, the nature of the repeats may also relate to recombination and facilitate chromosome rearrangements (Molnar et al., 2011).

Dispersed repetitive DNA

The second major group of repeated sequences are those related to transposable elements – pieces of DNA that amplify and move within the plant genome, either through an RNA intermediate (Class I transposable elements or retrotransposons), or through DNA (Class II, DNA transposable elements). Not least because of their mode of amplification, these sequences are usually dispersed throughout the chromosomes, although some families tend to be clustered in centromeric, intercalary or terminal chromosome regions. Sequences are also deleted during evolution of the genome. Transposable elements may diverge and amplify following speciation. Retrotransposons, amplifying through a DNA intermediate, are very widespread in many genomes, and indeed they and their remnants may represent some 50% of all the DNA in many species (eg in banana, see Heslop-Harrison & Schwarzacher 2007). As an example of a DNA transposon, a CACTA element has been amplified in *Brassica oleracea* (CC genome), and in situ hybridization can identify the C-genome chromosomes in the amphiploid species *B. napus* (AACC genomes; Alix et al. 2008) since its amplification is restricted to the C-genome evolutionary lineage. Transposable elements act as drivers of genome diversification or evolution and, by carrying coding DNA or inserting into coding or regulatory sequences, gene evolution.

Genome size evolution

Large scale genome evolution as discussed above occurs at levels from DNA mutation, through the repetitive motifs (motif sequence, number of copies and number of sites in the genome) to chromosomal rearrangement and whole genome duplication or polyploidy. These diverse mechanisms lead to huge changes in the total size of plant genomes. In 2011, the range of known plant genome sizes was known to lie from 63 Mbp (million base pairs) in the carnivorous species *Genlisea margaretae* to 149,000 Mbp in the liliales species *Paris japonica* (*Bennett and Leitch, 2011). *G. margaretae* contains all the genes required for a higher plant, but very little the repetitive DNA, although given the relationship to species with more DNA (as in the Brassicaceae evolution, Mandakova et al., 2010), it is likely that loss of sequences, and particularly the repetitive component of the genome, has occurred from the ancestors. Polyploidy and genome duplication also lead to larger genomes, and, for example within the Triticeae grasses (including barley, rye, wheat and their wild relatives), both chromosome number and genome size gives an indication of ploidy (Bennett and Leitch, 2011). However, even within diploids where molecular maps indicate single copy sequences show conserved synteny and recombination length (cM) is similar between species, genome (and thus average chromosome) sizes may vary by nearly 2-fold because of differences in repetitive DNA content.

Sequencing and genome evolution

Much recent work on plant genomes has come from their study by whole-genome shotgun sequencing. This approach, given the many-fold coverage of the genome that is possible, enables complete assembly of the gene-rich region of the genome, but does not allow assembly of the repetitive parts; in one recent genome assembly, Copia family retrotransposons represented 3.1% of reads but were 50-fold (0.062%) less in the assembly (As-Dous et al. 2011). When a dispersed repetitive element or tandem repeat array has very little variation, it can be extremely difficult to assemble sequences with short read shotgun technologies even using paired end strategies, where tandem arrays or groups of transposons may not be bridged efficiently. Even with sequencing of individual BACs, it has proved difficult to assemble the order of repeat monomers in the reads (Kuhn et al., 2011). BAC-end sequencing, or genomic survey sequencing (GSS) can give an accurate measurement of genome composition, although subcloning or adapter ligation may give unequal representation of repetitive sequence motifs, with different composition (% GC) of DNA or secondary structures. However, use of multiple approaches may assemble large blocks of repetitive DNA, such as those found at the centromeres, and this is essential to identify chromosome-to-chromosome variation, inclusion of genes, and epigenetic marks (*Wu et al., 2011).

Genome extinction, evolution or speciation

Given the range of evolutionary mechanisms at the genome level, can we identify any features of the genomes of species that are successful in continuing for long evolutionary times or show extensive speciation in their lineage, rather than going extinct? Over a shorter period, most current crops were domesticated over a relatively short period between 8,000 and 10,000 years ago, but are there genomic characteristics which mean a few have essentially vanished, and almost none been added, since this earliest period of domestication? Can genomic features of ‘new’ crops be predicted, whether those are currently minor crops, or, for example, biofuel crops where intensive selection has not been practiced (Heslop-Harrison and Schwarzacher, 2012)? Given predictions that species extinction is now occurring at as high rates as during previous mass extinctions, will the extra adaptability of polyploid plants mean they become dominant? The wealth of DNA sequence information, combined with appropriate informatic tools and experimental approaches including generation of synthetic hybrids, comparison of genotypes across environments, and modelling of genomic responses to stress (Kim et al., 2010, 2011), is beginning to generate answers to these important questions. The solutions will be of critical value both for conservation of the biodiversity of the plant kingdom and addressing the challenges of breeding new and more sustainable crops to feed the human population.

Acknowledgements

I am grateful to the many collaborators and co-workers in my laboratory for discussions related to the contents of this paper.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest

** of outstanding interest

Al-Dous EK, George B, et al. (June 2011). *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). Nature Biotechnology 29 (6): 521-528. doi:10.1038/nbt.1860

Alix K, Joets J, Ryder C, Moore J, Barker G, Bailey J, King G, Heslop-Harrison P. 2008. The CACTA transposon Bot1 played a major role in *Brassica* genome divergence and gene proliferation. Plant Journal 56: 1030-1044. http://dx.doi.org/10.1111/j.1365-313X.2008.03660.x .

Alix K, Ryder CR, Moore JM, King G, Heslop-Harrison JS. 2005. The genomic organization of retrotransposons in *Brassica oleracea*. Plant Molecular Biology 59 (6): 839 - 851. http://dx.doi.org/10.1007/s11103-005-1510-1

**APG (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. Botanical Journal of the Linnean Society 161 (2): 105–121, http://dx.doi.org/10.1111/j.1095-8339.2009.00996.x
The Angiosperm Phylogeny Group (APG) establishes the modern framework for the relationship-based (phylogenetic or "natural") taxonomy of all 413 recognized families of flowering plants showing their phylogeny or evolution. It is based on molecular, DNA sequence, variation superimposed on morphological data. There are only about 10 unplaced families, albeit of regional significance. This publication, superseding its 1998 and 2003 predecessors, with an associated paper on all land plants (embryophyta), provides the classification upon which all research on higher plants is based.

Bell CD, Soltis DE, Soltis PS. The age and diversification of the angiosperms re-revisited. American Journal of Botany. 2010 Aug;97(8):1296-1303. http://dx.doi.org/10.3732/ajb.0900346.

*Bennett MD, Leitch IJ. 2011. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. Ann Bot (2011) 107(3): 467-590 http://dx.doi.org/10.1093/aob/mcq258
The latest version of the database of nuclear genome sizes (DNA amounts) in angiosperms, with tables giving DNA contents in Mbp and by weight (picograms, pg), chromosome number and suggested ploidy level. The updates in this manuscript fill some extensive gaps in taxonomic coverage of measurements previously published (including, in particular, basal angiosperms that are sister groups to both monocotyledons and eudicotyledons), extend the range of sizes in both directions, and gives critical assessment of the validity of published measurements in this area where there are many errors in the literature.

Bejarano ER, Khashoggi A, Witty M, Lichtenstein C. 1996 Integration of multiple repeats of geminiviral DNA into the nuclear genome of tobacco during evolution. Proceedings of the National Academy of Sciences of the United States of America. Jan;93(2):759-764. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC40128/ .

*China Plant BOL Group, Li DZ, Gao LM, Li HT, Wang H, Ge XJ, et al. 2011 Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. Proceedings of the National Academy of Sciences 108(49):19641-19646. http://dx.doi.org/10.1073/pnas.1104551108.

A wide survey of relationships between 1,757 species, with a focus on the biodiversity in China, using three plastid genes and the ribosomal internal transcribed spacer (ITS). The paper is part of an international project to generate a robust and universal "Barcode of Life" (BOL) for plant identification, and shows the need to sample multiple individuals with several different markers.

Contento A, Heslop-Harrison JS, Schwarzacher T. 2005. Diversity of a major repetitive DNA sequence in diploid and polyploid Triticeae. Cytogenetic and Genome Research 109 (1-3): 34-42 http://dx.doi.org/10.1159/000082379 .

*Cullis CA, Vorster BJ, Van Der Vyver C, and Kunert KJ. 2009 Transfer of genetic material between the chloroplast and nucleus: how is it related to stress in plants? Ann Bot 103: 625-633.
An overview of the process of gene transfer from the organelle to the nucleus. From the origin of plants by endosymbiosis, gene transfer seems to be a one-way process giving advantages in terms of regulation and segregation of the genes, but must be accompanied by evolution of protein transit and targeting mechanisms,

Darwin C. (1859) *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. pp 495. London: John Murray. 1859

Gaeta RT, Pires JC, Iniguez-Luy F, Leon E, Osborn TC. 2007 Genomic changes in resynthesized *Brassica napus* and their effect on gene expression and phenotype. Plant Cell Nov;19(11):3403-3417. http://dx.doi.org/10.1105/tpc.107.054346.

Graybosch RA, Peterson CJ, Baenziger PS, Baltensperger PD, Nelson LA, Jin V, Kolmer J, Seaborn B, French R, Hein G, Martin TJ, Beecher B, Schwarzacher T, Heslop-Harrison P. 2009. Registration of ‘Mace’ hard red winter wheat. Journal of Plant Registrations 3(1): 51-56. http://dx.doi.org/10.3198/jpr2008.06.0345crc

Harper G, Osuji JO, Heslop-Harrison JS, Hull R. 1998. Integration of banana streak badnavirus into the *Musa* genome: molecular and cytogenetic evidence. Virology 255: 207-213. http://dx.doi.org/10.1006/viro.1998.9581

Heslop-Harrison JS. 2010. Genes in evolution: the control of diversity and speciation. Annals of Botany 106(3):437-438. http://dx.doi.org/10.1093/aob/mcq168.

Heslop-Harrison JS, Brandes A, Schwarzacher T. 2003. Tandemly repeated DNA sequences and centromeric chromosomal regions of *Arabidopsis* species. Chromosome Research 11: 241-253 http://dx.doi.org/10.1023/A:1022998709969

Heslop-Harrison JS, Schwarzacher T. 2011. Organization of the plant genome in chromosomes. Plant J. 66: 18-33. http://dx.doi.org/10.1111/j.1365-313x.2011.04544.x

Heslop-Harrison JS, Schwarzacher T. 2007. Domestication, genomics and the future for banana. Annals of Botany 100(5):1073-1084. http://dx.doi.org/10.1093/aob/mcm191

Heslop-Harrison JS, Schwarzacher T. Genetics and genomics of crop domestication. In: Plant biotechnology and agriculture: Prospects for the 21st century Ed Arie Altman, Paul Michael Hasegawa. Elsevier; 2012. p. 3-18. http://dx.doi.org/10.1016/B978-0-12-381466-1.00001-8 .

*Keeling PJ and Palmer JD 2008 Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics 9 605-618.
Although rare, transfer of genes between lineages of plants, fungi, animals, bacteria and viruses is a regular and evolutionarily important feature of evolution; large-scale DNA sequencing approaches are showing more horizontal transfer than had been suspected previously. This paper reviews well-supported cases of transfer from both prokaryotes and eukaryotes, and discussed the role of the transfers in species adaptation.

Kim J-R, Kim J, Kwon Y-K, Lee H-Y, Heslop-Harrison P, Cho K-H. 2011. Reduction of complex signaling networks to a representative kernel. Science Signaling 4, ra35. http://dx.doi.org/10.1126/scisignal.2001390

Kim TH, Kim J, Heslop-Harrison P, Cho KH. 2010. Evolutionary design principles and functional characteristics based on kingdom-specific network motifs. Bioinformatics 27: 245-251. http://dx.doi.org/10.1093/bioinformatics/btq633 .

Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. Fall and rise of satellite repeats in allopolyploids of Nicotiana over c. 5 million years. New Phytologist. 2010;186(1):148-160. http://dx.doi.org/10.1111/j.1469-8137.2009.03101.x .

Kuhn GCS, Teo CH, Schwarzacher T, Heslop-Harrison JS. Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species. Heredity. 2009 Mar;102(5):453-464. http://dx.doi.org/10.1038/hdy.2009.9.

Kuhn GCS, Küttler H, Moreira-Filho O, Heslop-Harrison JS. 2011. The 1.688 Repetitive DNA of Drosophila: Concerted evolution at different genomic scales and association with genes. Molecular Biology and Evolution. 2011 in press; http://dx.doi.org/10.1093/molbev/msr173 .

**Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA. Fast diploidization in close mesopolyploid relatives of Arabidopsis. Plant Cell. 2010 Jul;22(7):2277-2290. http://dx.doi.org/10.1105/tpc.110.074526.
The most recent of a series of papers from Lysak and his collaborators using molecular cytogenetic and DNA sequence information to reconstruct the changes that have occurred in the chromosomes during evolution in the Brassicaceae (Cruciferae). Their detailed analysis show that chromosome fusions, inversions and translocations happen sequentially, and lead to rapid recovery of diploid behaviour and reproductive isolation in new taxa.

Molnár I, Cifuentes M, Schneider A, Benavente E, Molnár-Láng M. Association between simple sequence repeat-rich chromosome regions and intergenomic translocation breakpoints in natural populations of allopolyploid wild wheats. Annals of Botany. 2011 Jan;107(1):65-76. http://dx.doi.org/10.1093/aob/mcq215 .

Patel D, Power JB, Anthony P, Badakshi F, Pat Heslop-Harrison JS, Davey MR. 2011. Somatic hybrid plants of *Nicotiana* × *sanderae* (+) *N. debneyi* with fungal resistance to *Peronospora tabacina*. Annals of Botany 108(5): 809-819. http://dx.doi.org/10.1093/aob/mcr197 .

Salse J, Bolot S, Throude M, Jouffe V, Piegu B, Quraishi UM, et al. 2008 Identification and characterization of shared duplications between rice and wheat provide new insight into grass genome evolution. Plant Cell Jan 20(1):11-24. http://dx.doi.org/10.1105/tpc.107.056309 .

Severin AJ, Cannon SB, Graham MM, Grant D, Shoemaker RC. 2011. Changes in twelve homoeologous genomic regions in soybean following three rounds of polyploidy. Plant Cell 23: 3129–3136

Staginnus C, Gregor W, Mette MF, Teo C, Fernandez EB, Machado ML, et al. Endogenous pararetroviral sequences in tomato (Solanum lycopersicum) and related species. BMC Plant Biology. 2007 May;7(1):24+. http://dx.doi.org/10.1186/1471-2229-7-24 .

Stegemann S, Keuthe M, Greiner S, Bock R. Horizontal transfer of chloroplast genomes between plant species. Proceedings of the National Academy of Sciences. 2012 Jan;109(7):2434-2438. http://dx.doi.org/10.1073/pnas.1114076109.

*Stevens, P. F. (2011). Angiosperm Phylogeny Website. [23 Oct 2011 update]. http://www.mobot.org/MOBOT/research/APweb/
Although neither a refereed source nor database, this up-to-date and regularly updated website gives an authoritative overview of aspects of genome evolution and the more tricky aspects of defining angiosperm

evolution, phylogentics and monophyly. The text here gives some of the reasons why groups such as aquatic plants and parasites, may be difficult to place, despite general congruence between morphological and molecular data.

Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. Science. 2008 Apr;320(5875):486-488. http://dx.doi.org/10.1126/science.1153917 .

**Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. Nat Rev Genet. 2009 Oct;10(10):725-732. http://dx.doi.org/10.1038/nrg2600 .
An insightful paper overviewing the evidence that most WGD/polyploidy events are evolutionary dead-ends. The few WGDs that are successful enable the lineage to avoid extinction and lead to its evolutionary success.

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K. The flowering world: a tale of duplications. Trends Plant Sci. 2009 Dec;14(12):680-688. http://dx.doi.org/10.1016/j.tplants.2009.09.001.

Vaughan HE, Heslop-Harrison JS, Hewitt GM. 1999. The localisation of mitochondrial sequences to chromosomal DNA in Orthopterans. Genome 42: 874-880.

*Wu Y, Kikuchi S, Yan H, Zhang W, Rosenbaum H, Iniguez AL, et al. Euchromatic Subdomains in Rice Centromeres Are Associated with Genes and Transcription. Plant Cell 2011 Nov; http://dx.doi.org/10.1105/tpc.111.090043.
A careful assembly of the centromeres of four of the 12 pairs of rice chromosomes shows the organization and variation of the repetitive DNA, the single copy genic DNA, and histone modifications in these important chromosome regions.

Zambryski P, Joos H, Genetello C, Leemans J, Montagu MV, Schell J. 1983. Ti plasmid vector for the introduction of DNA into plant cells without alteration of their normal regeneration capacity. The EMBO Journal. 1983;2(12):2143-2150. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC555426/ .

Legends to figures

Figure 1. Part of the only figure in *The Origin of Species* (Darwin, 1859). The concepts of separation of species, with many becoming extinct and others separating into many other species, along with the monophyletic origins of branches, is the basis of our current models of phylogeny and genome evolution.

Figure 2. Two metaphases of bread wheat, *Triticum aestivum* (2n=6x=42) with chromosome arms introgressed from other genera in the grass tribe Triticeae, both from successful wheat varieties grown commercially. The alien chromosome arm is labelled in red by in situ hybridization with total genomic DNA from the alien species. A) Wheat variety Beaver with a 1BL.1RS translocation incorporating a chromosome arm from rye, *Secale cereale*; the second green probe labels the rDNA at the nucleolar organizers on chromosomes 1R, 6B, 1A, 5D and 7D. B) The variety Mace (Graybosch et al., 2009) with a 4DL.4AgS translocation. The green probe shows sites of a tandemly repeated DNA sequence dpTa1. Scale bar 10 µm. Figures courtesy of Niaz Ali, Robert Graybosch and Trude Schwarzacher.

---

[i] NOTICE from http://libraryconnect.elsevier.com/sites/default/files/lcp0404.pdf: "This is the author's version of a work that was accepted for publication in "Current Opinion in Plant Biology". Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms, may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Genome evolution: extinction, continuation or explosion? Current Opinion in Plant Biology 15:115–121 http://dx.DOI.org/10.1016/j.pbi.2012.03.006"