



ELSEVIER

Genome studies and molecular genetics

The consequences of gene and genome duplication in plants

Editorial overview

Susan R Wessler and James C Carrington

Current Opinion in Plant Biology 2005, 8:119–121

1369-5266/\$ – see front matter

© 2005 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.pbi.2005.01.015

Susan R Wessler

Department of Plant Biology, The University of Georgia, Athens, Georgia 30602, USA
e-mail: sue@plantbio.uga.edu

Susan R Wessler is professor of plant biology at the University of Georgia. Her research has focused on identifying plant transposable elements and determining how they contribute to the diversification of genes and genomes. With the availability of increasing amounts of plant genomic sequence, her laboratory has pioneered the integration of computational analysis into transposon studies.

James C Carrington

Center for Gene Research and Biotechnology, 3021 ALS Building, Oregon State University, Corvallis, Oregon 97331
e-mail: carrington@orst.edu

More information on Jim's work is available at <http://jcclab.science.oregonstate.edu>

The plant community has long suspected that at least some plant genomes have an abundance of duplicate genes. In the pre-genomic era, studies of duplicate genes were restricted to a few plant species in which duplication events could be easily recognized. The notion that maize is the product of an ancient polyploidization event was supported, in part, by the analysis of genes that were located on different chromosomes but that had overlapping functions. The widespread nature of genome duplication in maize became apparent when clusters of linked molecular markers were found to map to different chromosomes. Cytogenetic methods were used to understand the polyploid origins of another member of the grass clade, wheat, by visualizing the chromosomal location of repeat clusters, such as the DNA that encodes 5S rRNA.

The analysis of increasing amounts of genomic sequence has not only confirmed the prevalence of gene duplication in these plant species but, surprisingly, has also revealed that gene and genome duplication events are common occurrences in the evolution of plant genomes. Computational analyses of the complete DNA sequence of *Arabidopsis thaliana* and the nearly complete sequence of *Oryza sativa* have demonstrated that even these small model plant genomes have experienced numerous episodes of polyploidization and segmental duplication. And plants are not unusual in this regard. This year brought news that the genome of the model unicellular eukaryote, *Saccharomyces cerevisiae* is the product of an ancient whole-genome duplication event and, that a significant proportion of the human genome is derived from segmental duplications.

Why has it taken so long to recognize the prevalence of polyploidy and segmental duplication events in the evolution of both plant and animal genomes? For plants at least, the answer to this question seems to be that the duplicated genome returns to a functionally diploid state by mechanisms that begin, in some cases, almost immediately after the polyploid is formed. Just how this happens and what the evolutionary consequences might be are discussed, to varying degrees, in all of the reviews in this section.

The evolutionary consequences of genome duplication events have been the subject of theoretical studies for decades. Moore and Purugganan (pp. 122–128) discuss how the ideas of Ohno in his 1970 book, *Evolution by Gene Duplication*, are faring in light of the flood of genomic sequence. Ohno proposed that there are two alternative fates for duplicate genes: either one of the duplicates is lost (pseudogenization) or it acquires a new function (neofunctionalization). However, as discussed by Moore and Purugganan, plant duplicate genes are retained for longer than previously predicted.

This situation has led to the formulation of a new paradigm, which posits that a third fate of duplicate genes, called subfunctionalization, is an important force, at least in sequenced plant genomes. Subfunctionalization is synonymous with a process called duplicate, degenerate, complement (DDC) in which mutations in duplicate genes result in partial loss-of-function and the maintenance of gene duplicates because of functional complementation. Examples from the well-characterized MADS-box genes provide support for this model. Importantly, regulatory genes (such as transcription factors) appear particularly well-suited for functional and regulatory diversification in duplicated gene families.

Another intriguing set of gene duplication events involves disease resistance (*R*) genes. As Meyers *et al.* (pp. 129–134) describe, these genes belong to several structural classes, although most contain a conserved set of modules within the NBS-LRR domains. *Arabidopsis* contains at least 149 such genes, many of which are organized as clusters of paralogs that exchange sequences frequently. These *R* gene clusters evolve by tandem and segmental duplications, rearrangement between genes in a cluster, point mutations and frequent loss of genes. The obvious benefit of organizing an *R* gene repertoire in this manner is the potential to spawn new functionality by gene duplication and multiple diversification mechanisms. However, proliferation and functionalization of *R* genes appears to be subject to balancing selection, where the beneficial contributions of new genes that have novel pathogen specificities are countered by fitness costs to the plant. The result appears to be perpetual sampling of *R* gene activities but with a limited number of *R* gene units at any given time.

Although tandem and segmental duplications probably provide the major mechanism to proliferate *R* genes, polyploidy might offer spectacular episodes of *R* gene diversification and reshuffling. Although the impact of polyploidy on *R* gene diversification will certainly be a fascinating area for future studies, the review of the current status of polyploid research by Adams and Wendel (pp. 135–141) highlights the rapid progress and new insights that are being fueled by the availability of genomic resources for an increasing number of plant species. The fact that polyploidization has occurred throughout the evolution of eukaryotes and is still an ongoing process means that experimental systems are available to analyze all stages of polyploidization; from those in the distant past to synthetic polyploids created in laboratories. At one end of the evolutionary spectrum, ancient polyploids such as *Arabidopsis* have been analyzed to determine whether gene loss is random or whether certain gene functions are preferentially retained. At the other end of the spectrum, analyses of new natural polyploids in cotton and synthetic polyploids in both cotton and wheat have demonstrated that DNA loss and/or changes in the epigenetic regulation

of duplicate genes are initiated immediately or soon after polyploidization.

The striking effects of polyploidy on gene expression can be partially explained by gene-specific and regional gene silencing. Gendrel and Colot (pp. 142–147) provide a mechanistic basis, chromatin-associated RNAi, to explain at least some of the silencing that is associated with genome-merger events. This mechanism also explains the silencing of highly repeated elements (e.g. transposons and retroelements) that comprise heterochromatin domains, tandem duplications and transgenes with repeated structure. The emerging model states that transcripts from duplicated sequences are funneled toward a specialized, chromatin-associated branch of the RNAi pathway, resulting in small RNAs that guide cytosine methylation and histone methylation. These heterochromatic marks inactivate genes at the transcriptional level and provide the chemical basis for epigenetic silencing in subsequent generations. Further, the RNAi apparatus offers a mechanism to renew gene silencing in the event that DNA methylation and chromatin marks are lost, and to recognize the subsequent emergence of related repeat sequences within the genome. Although it is clear that sequence-complementarity rules govern where DNA (and associated histones) get methylated, how small RNAs guide covalent modifications of DNA and histone subunits remains to be determined.

If a picture is really worth a thousand words, then the new cytogenetic techniques reviewed by Kato *et al.* (pp. 148–154) will facilitate more rapid advances in our understanding of epigenetic regulation and the origins of polyploid genomes. Techniques to gently stretch chromatin while maintaining DNA-protein associations have already improved our resolution of plant centromeres. Similarly, the ability to dissect the closely related genomes that contribute to wheat polyploids has been greatly facilitated by improvements in the probes used to visualize individual chromosomes.

Another picture that has been of extraordinary value in the comparative analysis of grass genomes is the so-called 'Crop Circle' diagram; first published by Gale and colleagues in 1995. The Crop Circle visualizes the chromosomal features that unify the grasses and serves to facilitate their study as a single genetic system. As discussed by Devos (pp. 155–162), the Crop Circle is a work in progress that must be continually updated through the addition of new species and the incorporation of new information, such as that derived from the analyses of genomic resources such as the rice genome sequence and expressed sequence tag (EST) collections from a growing list of grass species. Here we come full-circle, so to speak. For it turns out that the greatest challenge in updating the Crop Circle is figuring out how to incorporate genome duplication events that both shuffle the deck and add new

cards, making it even more difficult to distinguish orthologs and paralogs.

The remarkable findings and ideas summarized in these articles represent only the initial wave of discovery about how plants, as well as other organisms, evolve their genomes and generate useful diversity. Plant biologists are in a unique position to study genome diversification. Plants appear to be especially well suited to

deal with the 'negative' consequences of small-scale and large-scale duplication (by rapid genome rearrangements and gene silencing), but they can simultaneously exploit the rich opportunities that arise through the expansion of genomes (by retaining duplicates for long periods). It is now time to consider questions that depend on the availability of hundreds of plant genome sequences, and to prepare for the next era of discovery in plant genome evolution.