From Cot Curves to Genomics. How Gene Cloning Established New Concepts in Plant Biology

Robert B. Goldberg*

Department of Molecular, Cell, and Developmental Biology, University of California, Los Angeles, California 90095–1606

It is difficult to imagine carrying out plant research without personal computers, the Internet, GenBank, e-mail, cell phones, gene cloning, microchips, whole genomic sequences, expressed sequence tags, RFLPs, PCR, knock-outs, Arabidopsis, reverse genetics, transgenic plants, and molecular biology "kits" that are ready-made to carry out almost any type of DNA manipulation experiment imaginable. The plant world in 1975 was vastly different from the one in which we, as plant scientists, operate in today. The International Society of Plant Molecular Biologists did not exist. International forums such as the Plant Molecular Biology Gordon Conference, the Plant-Oriented Keystone Symposia, and the Plant Molecular Biology Congress had not been established. One of the largest gatherings of plant scientists occurred at the annual meetings of the American Society of Plant Physiologists and seldom more than 20 or 30 scientists attended the nucleic acids section in which the most exciting plant molecular biology results were presented. The "real world" was different as well. The Vietnam War had just ended, the Cold War with the Soviet Union raged on, the Berlin Wall split Europe into East and West, and the world economy was in an inflationary spiral due to the emergence of the oil cartel that sent the prices of gasoline skyrocketing.

Genetic engineering had been "invented" by Stanley Cohen and Herbert Boyer 2 years earlier (11) and was still limited to an elite number of labs that understood bacterial genetics, had the plasmid vectors for DNA cloning, and had access to the enzymes that we purchase in cloning "kits" today. Procedures for cDNA cloning, creating libraries of large eukaryotic genomes, and isolating structural genes had not yet been published. Genetic engineering was as controversial then as genetically modified organisms are today. The Asilomar Conference took place in 1975, and scientists who wanted to use the emerging tools of genetic engineering were required to follow strict, self-imposed guidelines that specified the conditions under which DNA manipulations could be carried out in the laboratory. Demonstrations occurred across the globe forecasting that "monsters" would be created by the new gene splicing techniques and one city (Cambridge, MA) attempted to ban genetic engineering altogether. Nevertheless, it was a magical time to be studying basic plant processes. For the first time, there was a dream that one could finally "see" a plant gene and begin to unravel the complexity of plant processes at the genome level.

THE PRINCIPLES OF PLANT GENOME ORGANIZATION AND GENE REGULATION WERE LAID DOWN IN THE PRECLONING ERA

Plant genomes were investigated in the mid- to late-1970s by quantitative DNA reassociation tools (i.e. Cot curves) that had their origins in the 1960s when the principles of DNA denaturation and renaturation were pioneered at the Carnegie Institution of Washington by Roy Britten and his associates (7, 8)—principles that are still used today each time a gel blot or microchip experiment is carried out, a primer Tm is calculated, or PCR conditions are punched into a thermocycler. Plant genomes had been shown to contain repetitive DNA sequences in the mid-1960s and were, therefore, considered to be "eukaryotic-like" and similar to animal genomes in that respect (7, 8). In 1975, genome organization was the "code word" for those of us who studied "genomics" and it was determined that plant genomes had many families of repetitive sequences and that these repeats varied in copy number and arrangement in the genome (17, 20). These repeats were shown to be both scattered around the genome and localized in long clusters and they were also shown to be flanked by complex single-copy sequences (17, 20). Neither these repeats nor any flanking single-copy DNA had been cloned or sequenced at this time. In fact, DNA sequencing procedures (29, 33) had not yet been invented and plant DNA sequences had not yet been cloned (3). However, the general concepts of plant genome organization that were derived from DNA reassociation studies have stood the test of time and have been illuminated in great detail by a knowledge of the actual DNA sequences that span each Arabidopsis chromosome (5).

During this same period, important principles of plant gene activity were being established in global

* E-mail bobg@ucla.edu; fax 310–825–8201.

terms by the use of RNA-excess/DNA-RNA hybridization techniques (i.e. Rot curves) with either cDNA or genomic single-copy DNA probes (21, 22, 24, 25). The technique of subtraction hybridization (or cascade hybridization as it was first referred to in the literature) was established by Bill Timberlake in this era using kinetically fractionated cDNA populations (36). Both cDNA and genomic single-copy DNA subtraction procedures were used by many of us to investigate developmental changes in plant mRNA populations (21, 22, 24, 25). Several important concepts emerged about higher plant cells in this precloning population hybridization era. First, it became clear that plants contained a complex set of nuclear RNAs and that only about 25% of this complexity was represented in the corresponding mRNA population (21). Today, we know that the additional complexity in the nuclear RNA represents primarily unprocessed introns in primary transcripts. However, this was not understood at the time because plant genes had not yet been cloned and sequenced, and introns had not yet been discovered in any eukaryotic gene. Second, it became clear that a large number of genes were active in plant cells and that these genes were highly regulated in the plant life cycle (21, 24, 25). Each plant organ system was shown to have a unique set of active genes and it was estimated that approximately 60,000 genes were required to program and maintain the entire life cycle of the tobacco plant (24). This estimate of the number of tobacco genes has stood the test of time for plants with large genomes (i.e. corn) and, considering the "bluntness" of the tools used and assumptions that had to be made (e.g. average mRNA size), is not that far off from the 25,000 genes that has been shown by sequencing to be present in the small Arabidopsis genome (5). Finally, it was established that mRNA populations contained sequences with varying degrees of prevalence and that both transcriptional and posttranscriptional processes established the mRNA sequence sets present in various plant organs and tissue types. By the end of the Cot and Rot curve era (mid-1980s), it was clear that plant cells resembled animal cells with respect to the number of genes and the complexity of gene regulatory processes. It was not known, however, how any individual gene was regulated or how sets of genes were co-expressed in space and time.

PLANT GENES CAN BE CLONED!

By the end of the 1970s, exciting new procedures were developed by Tom Maniatis and others to construct cDNA clones of specific eukaryotic mRNAs and isolate the corresponding genes from the genome (26, 27). In addition, techniques were devised to sequence DNA segments (29, 33), visualize genes directly in the electron microscope in association with their RNAs (i.e. R loops; 38), and detect specific DNA fragments and mRNAs using DNA and RNA gel blots, respectively (1, 34). These procedures established a new revolution in molecular biology because, for the first time, the structures of individual genes could be studied and their expression patterns, mechanisms of regulation, and evolutionary origins analyzed. This was an exciting period and the most surprising and startling observation made with the new DNA cloning techniques was that the coding regions of eukaryotic genes were interrupted by noncoding sequences (23)! New words, intron and exon, were introduced into the molecular biology lexicon (19) and posttranscriptional splicing mechanisms were hypothesized and studied (23).

Only a few plant scientists at that time had any experience with bacterial genetics, the new recombinant DNA techniques, or access to enzymes required for DNA cloning and manipulation. In fact, most of us did not know a restriction enzyme from a ligase and had to learn from "scratch" how to streak and grow bacterial cells in order to attempt to clone plant DNA sequences! In the 1970s and 1980s (as well as today) plant scientists were playing "catch-up" with their animal counterparts and were competing for a meager pot of money. It was during this time that Joe Key played a huge role in establishing the U.S. Department of Agriculture Competitive Research Grants Program after many years of fighting the U.S. Department of Agriculture bureaucracy and Congress. This Program has made a major impact over the past 25 years in keeping plant sciences in the forefront of pioneering research.

Rumors began to circulate in the late 1970s that plant DNA could not be cloned. One well-known plant molecular biologist (who will remain anonymous) went from meeting to meeting like Paul Revere declaring that plant DNA was "different" from animal or bacterial DNA and that it could not be cloned! John Bedbrook and colleagues in Dick Flavell's lab in Cambridge, England soon showed that this was not the case and demonstrated directly that plant DNA could be cloned and replicated in bacteria just like the DNA from other organisms (3). They reported their results in 1979 at a meeting in Minneapolis and the era of plant gene cloning began with the successful cloning of ribosomal DNA and telomeric repeated sequences from wheat (3). A pioneering principle was established—plant DNA was similar to that of all other organisms and could be manipulated using the same enzymes, cells, and vector systems.

Soon thereafter, libraries of many plant genomes were constructed and, in the early 1980s, were made available to plant scientists around the world (16, 35). In addition, the first plant structural genes were

cloned, sequenced, and visualized in the electron microscope (16, 35). These genes, encoding seed storage proteins (16, 35) and the small subunit of ribulose bisphosphate carboxylase (4), were shown to contain introns similar to those in animal genes, which supported the notion that plant cells had genetic processes similar to those in animals. It was also demonstrated that plant genes were located relatively close to each other on plant chromosomes (approximately every 4–6 kb) and that genes with different expression patterns were interspersed among each other, implying that each functioned as an independent unit (16)—a suggestion that was verified during the post-transformation era (9, 31, 32).

During the same period, cDNA libraries were constructed for almost every imaginable plant organ system and developmental state, and cDNA clones representing prevalent plant mRNAs, such as those encoding seed proteins, light-regulated proteins, hormone-induced proteins, and cell wall proteins were identified. These cDNA clones were used to demonstrate directly that both transcriptional and posttranscriptional processes played a role in controlling plant gene expression, but that the primary control for most plant genes was at the level of transcription. In addition, the vast array of cDNAs that became available and were sequenced and studied in the 1980s began to illuminate a range of plant developmental, metabolic, and biochemical processes. The age of understanding "how to make a plant" had begun.

PICKING APART PLANT GENES

As the new era of plant gene cloning began, another revolution was occurring in several labs that were engaged in a fierce competitive battle to be the first to transform plant cells. The laboratories of Jeff Schell and Marc Van Montagu (Gent, Belgium), Rob Schilperoort (Leiden, The Netherlands), Mary-Dell Chilton and Michael Bevan (Washington University, St. Louis; Cambridge University, UK), and Rob Fraley, Steve Rogers, and Rob Horsch (Monsanto, St. Louis) were utilizing the new recombinant DNA techniques to construct *Agrobacterium tumafaciens* T-DNA vectors that could be used to introduce new genes into plant cells. In the mid-1970s, Mary-Dell Chilton had shown that *A. tumafaciens* T-DNA was integrated into the chromosomes of plant cells (10), setting the stage for the revolution in plant genetic engineering that continues to this day.

In 1983, the Gent, Monsanto, and Washington/Cambridge groups showed independently that T-DNA vectors could be used to transfer bacterial antibiotic resistance genes into plant cells and that these genes could be expressed if engineered with the correct promoters (6, 12, 18). Much to the surprise of everyone in the plant research world, a different group, headed by Tim Hall, demonstrated that the phaseolin seed storage protein gene from french beans could be transferred to sunflower cells and be expressed (31). This now-famous (or infamous) "sunbean" plant made the front page of the *New York Times* and was proclaimed in *Time* to be a "glowing achievement. . . the first time a gene from one plant had been inserted into the chromosomes of an unrelated species and made to express itself." The sunbean experiment was reported initially at the first University of California (Los Angeles) Keystone Meeting on Plant Molecular Biology that I organized in April of 1983 and was greeted at the time by a now-famous plant cell biologist (who I will not name) as "nonsense!" Nevertheless, it showed for the first time that gene cloning and *A. tumafaciens* transformation techniques could be combined to transfer foreign genes into plant cells and study their function. The age of plant genetic engineering and gene manipulation had begun!

FROM PHENOTYPE TO GENE

Throughout the 1980s and 1990s, many plant genes were cloned and investigated in transformed cells in order to understand the mechanisms regulating their expression. Numerous plant promoters were characterized and DNA sequence elements programming transcription in specific developmental states were uncovered. The prediction of earlier experiments on the structure and organization of plant genes proved correct and a major new concept emerged—plant genes functioned as independent units and contained regulatory regions that could program their correct expression in foreign cell environments. These experiments set the stage for engineering new crops with novel traits that are produced at specific times during the plant life cycle (28).

A major switch in plant gene cloning occurred in the beginning of the late 1980s and early 1990s. Many interesting genes that produced novel phenotypes were being uncovered in corn and Arabidopsis using genetic approaches that were being adopted rapidly by plant scientists. Because their products were unknown and/or very rare, it was not possible to use conventional cloning methods to isolate these genes. Several pioneering procedures were invented that circumvented this problem and enabled a wide range of plant genes to be cloned. First, T-DNA was shown to act as a mutagen in plant cells and, as such, could be used as a tag to identify and clone genes that specified novel phenotypes (15). Ken Feldmann and his colleagues established a novel seed transformation method to obtain large numbers of T-DNA transformed Arabidopsis lines and this method was used to identify important plant genes, such as those involved in the control of floral organ identity and hormone perception (14, 15). In my opinion, this was one of the most important advances in plant biology in the past 25

years because it allowed, for the first time, a relatively simple way to clone plant genes associated with fascinating mutant phenotypes. The availability of Ken Feldmann's T-DNA lines caused numerous investigators (including myself) to adopt Arabidopsis as a model system and opened up many new problems in plant biology to investigation. It also paved the way to the reverse genetics approaches in use today—identifying mutant lines associated with randomly sequenced genes (30).

A second approach to cloning plant genes was also being developed at the same time. During the 1980s, Nina Fedoroff and Sue Wessler cloned the corn Ac and Ds transposable elements (13). This pioneering experiment paved the way for using transposons to tag and capture novel plant genes for which only a phenotype could be identified. The transposon tagging and gene cloning procedure complemented the T-DNA approach and led to the identification of many important new genes in several plants including corn, snapdragon, and Arabidopsis (37). It also became possible in the 1990s to use map-based cloning strategies to identify and clone plant genes—particularly in Arabidopsis because of its small genome size (2). With the completion of the Arabidopsis Genome Project last fall, and the identification of 30,000 single nucleotide polymorphisms in the Arabidopsis genome, map-based cloning of plant genes should permit the identification of any gene for which there is a mutant phenotype—even those induced by chemical mutagens such as ethyl methane sulfonate.

BACK TO THE FUTURE

Looking back, in 1975 plant molecular biologists were asking questions about the number of genes in plant chromosomes and how these genes are regulated in development. We were using precloning tools of DNA and RNA hybridization that gave precise answers, but which could not focus in on specific genes. The questions addressed then are being addressed once again today in the genomics age. In a sense, we have come full circle in trying to understand how plant chromosomes are constructed and how populations of genes are expressed in various cells, tissues, and organs. We progressed from studying populations of genes and mRNAs to investigating individual cloned genes and mRNAs to using high throughput experiments with arrays of thousands of specific genes in order to uncover the secrets of plant cells. Twenty years after the cloning of the first plant DNA segments (3), the genomes of Arabidopsis and rice have been sequenced and numerous expressed sequence tag sequencing projects have uncovered tens of thousands of mRNAs in a wide range of plants (5). It is remarkable that the era of gene cloning is coming to an end. Nevertheless, the challenges are no less daunting and are even more complex: What are the functions of all plant genes and how is the information in plant genomes utilized in order to program plant development from fertilization to seed dormancy?

LITERATURE CITED

1. **Alwine JC, Kemp DJ, Parker BA, Reiser J, Renart J, Stark GR, Wahl GM** (1979) Methods Enzymol **68:** 220–242
2. **Arondel V, Lemieux B, Hwang I, Gibson S, Goodman H, Somerville CR** (1992) Science **258:** 1353–1355
3. **Bedbrook J, Gerlach W, Thompson. R, Flavell RB** (1980) *In* I Rubenstein, B Gengenbach, R Phillips, CE Green, eds, Emergent Techniques. University of Minnestoa Press, Minneapolis, pp 93–113
4. **Berry-Lowe SL, McKnight TD, Shah DM, Meagher RB** (1982) J Mol Appl Genet **1:** 483–498
5. **Bevan M, Bancroft I, Bent E, Love K, Goodman H, Dean C, Bergkamp R, Dirkse W, Van Staveren M, Stiekema W, et al.** (1998) Nature **391:** 485–488
6. **Bevan MW, Flavell RB, Chilton M** (1983) Nature **304:** 184–187
7. **Bolton ET, Britten RJ, Cowie DB, Roberts RB, Szafranski P, Waring MJ** (1965) Carnegie Inst Wash Year Book **64:** 314–333
8. **Britten RJ, Kohne DE** (1968) Science **161:** 529–540
9. **Broglie R, Coruzzi G, Fraley RT, Rogers SG, Horsch RB, Niedermeyer JG, Fink CL, Flick JS, Chua NH** (1984) Science **224:** 838–843
10. **Chilton MD, Drummond MH, Merio DJ, Sciaky D, Montoya AL, Gordon MP, Nester EW** (1977) Cell **11:** 263–271
11. **Cohen SN, Chang AC, Boyer HW, Helling RB** (1973) Proc Natl Acad Sci USA **70:** 3240–3244
12. **Estrella-Herrera L, Depicker A, Van Montagu M, Schell J** (1983) Nature **303:** 209–213
13. **Fedoroff N, Wessler S, Shure M** (1983) Cell **35:** 235–242
14. **Feldmann KA** (1991) T-DNA insertion mutagenesis in *Arabidopsis*. Plant J **1:** 71–82
15. **Feldmann KA, Marks MD** (1987) Mol Gen Genet **208:** 1–9
16. **Fischer RL, Goldberg RB** (1982) Cell **29:** 651–660
17. **Flavell RB, Bennett MD, Smith JB, Smith DB** (1974) Biochem Genet **12:** 257–269
18. **Fraley RT, Rogers SG, Horsch RB, Sanders PR, Flick JS, Adams SP, Bittner ML, Brand LA, Fink CL, Fry JS, Galluppi GR, Goldberg SB, Hoffmann NL, Woo SC** (1983) Proc Natl Acad Sci USA **80:** 4803–4807
19. **Gilbert W** (1978) Nature **271:** 501
20. **Goldberg RB** (1978) Biochem Genet **16:** 45–68
21. **Goldberg RB, Hoschek G, Kamalay JC, Timberlake WE** (1978) Cell **14:** 123–131
22. **Goldberg RB, Hoschek G, Tam SH, Ditta GS, Breidenbach RW** (1981) Dev Biol **83:** 201–217
23. **Jeffreys AJ, Flavell RA** (1977) Cell **12:** 1097–1108

24. **Kamalay JC, Goldberg RB** (1980) Cell **19:** 935–946
25. **Kamalay JC, Goldberg RB** (1984) Proc Natl Acad Sci USA **81:** 2801–2805
26. **Maniatis T, Hardison RC, Lacy E, Lauer J, O'Connell C, Quon D, Sim GK, Efstratiadis A** (1978) Cell **15:** 687–701
27. **Maniatis T, Kee SG, Efstratiadis A, Kafatos FC** (1976) Cell **8:** 163–182
28. **Mariani C, Debeuckeleer M, Truettner J, Leemans J, Goldberg RB** (1990) Nature **347:** 737–741
29. **Maxam AM, Gilbert W** (1977) Proc Natl Acad Sci USA **74:** 560–564
30. **McKinney EC, Ali N, Traut A, Feldmann KA, Belostotsky DA, McDowell JM, Meagher RB** (1995) Plant J **8:** 613–622
31. **Murai N, Sutton DW, Murray MG, Slightom JL, Merlo DJ, Reichert NA, Sengupta-Gopalan C, Stock CA, Barker RF, Kemp JD, Hall TC** (1983) Science **222:** 476–482
32. **Okamuro JK, Jofuku KD, Goldberg RB** (1986) Proc Natl Acad Sci USA **83:** 8240–8244
33. **Sanger F, Nicklen S, Coulson AR** (1977) Proc Natl Acad Sci USA **74:** 5463–5467
34. **Southern EM** (1975) J Mol Biol **98:** 503–517
35. **Sun SM, Slightom JL, Hall TC** (1981) Nature **289:** 37–41
36. **Timberlake WE** (1980) Dev Biol **78:** 497–510
37. **Walbot V** (1992) Annu Rev Plant Phys Plant Mol Biol **43:** 49–82
38. **White RL, Hogness DS** (1977) Cell **10:** 177–192