CHAPTER 2

Genome Size Evolution in Plants

MICHAEL D. BENNETT AND ILIA J. LEITCH

Every cellular organism possesses a genome, and, because of this, the question of genome size evolution is not limited to any one taxon, but rather is of universal biological interest. Along with animals, plants are the best studied group with regard to variation in DNA content, and have played a critical role since the earliest days of genome size study. This chapter provides an overview of the current state of knowledge concerning genome size evolution in plants. As with the previous chapter on animals, this includes a review of the available data, the patterns of variation both within and among species and higher taxa, the major mechanisms involved in generating disparities among groups, and the impacts of differences in genome size at the nuclear, cellular, tissue, and whole-organism levels. It is evident from this discussion that there are both deep parallels and major divergences between plants and animals in terms of genome size evolution. However, on one point the two kingdoms clearly project the same message: that genome size is a highly relevant biological characteristic whose evolution continues to represent a key puzzle in genomics and evolutionary biology.

The Evolution of the Genome, edited by TR Gregory
Copyright 2005, Elsevier Inc. All rights reserved.

A BRIEF HISTORY OF GENOME SIZE STUDY IN PLANTS

The First Estimates of DNA Amounts

As noted in Chapter 1, estimates of nuclear DNA amounts have been made since before the elucidation of the double helix structure in 1953. The earliest approaches were based on analyses of isolated nuclei or cell suspensions, as was first done for several animal species (Boivin *et al.*,1948; Vendrely and Vendrely, 1948). The constancy of nuclear DNA reported in the early animal studies was examined in greater detail almost immediately for both animals (e.g., Mirsky and Ris, 1949; Swift, 1950a) and plants (e.g., Swift, 1950b). Importantly, such work on *Zea mays* and *Tradescantia paludosa* led Swift (1950b) to develop the still widely used term *C-value*[1] to define the DNA content of an unreplicated haploid nuclear genome. However, these studies dealt only with relative DNA contents in different tissues of a few test species, and did not provide estimates of absolute DNA mass (i.e., in picograms, pg; 1 pg = 10^{-12} g). Probably the first estimate of the absolute amount of DNA in the nuclear genome of a plant was done for *Lilium longiflorum* cv. Croft by Ogur *et al*. (1951).

Just 10 years later, published measurements of DNA amount per cell in angiosperms (flowering plants) already ranged more than 50-fold, from 5.5 pg in *Lupinus albus* to 313 pg in *Lilium longiflorum* (McLeish and Sunderland, 1961). Soon after, studies reporting 40-fold interspecific variation between 22 diploid species in the family Ranunculaceae (Rothfels *et al.*,1966), and 5-fold within the genus *Vicia* (Martin and Shanks, 1966) confirmed that extensive variation occurred within families and even individual genera, independently of ploidy level (e.g., Fig. 2.1).

Possession of nucleated erythrocytes and the absence of a cell wall were important factors in the selection of animal materials for pioneering research on genome size, whereas easy availability as laboratory models with well-studied genetics, amenable cytology, and large nuclei influenced the first plant materials chosen. Over the following decades, plants became increasingly important in studies of genome size, often proceeding several steps ahead of equivalent work on animals (Gregory, 2005).

---

[1]There has been considerable confusion regarding the origin of the term "C-value." Many authors have assumed incorrectly that it refers to "content," "complement," or "characteristic." In coining the term, Swift (1950b) did not provide a clear definition, and made reference only to the 1C (haploid) "class" of DNA, leading to the reasonable conclusion that "C" stood for "class" (e.g., Gregory 2001, 2002). However, in a letter to M.D. Bennett dated June 24, 1975, Hewson Swift stated that: "I am afraid the letter C stood for nothing more glamorous than 'constant,' i.e., the amount of DNA that was characteristic of a particular genotype."

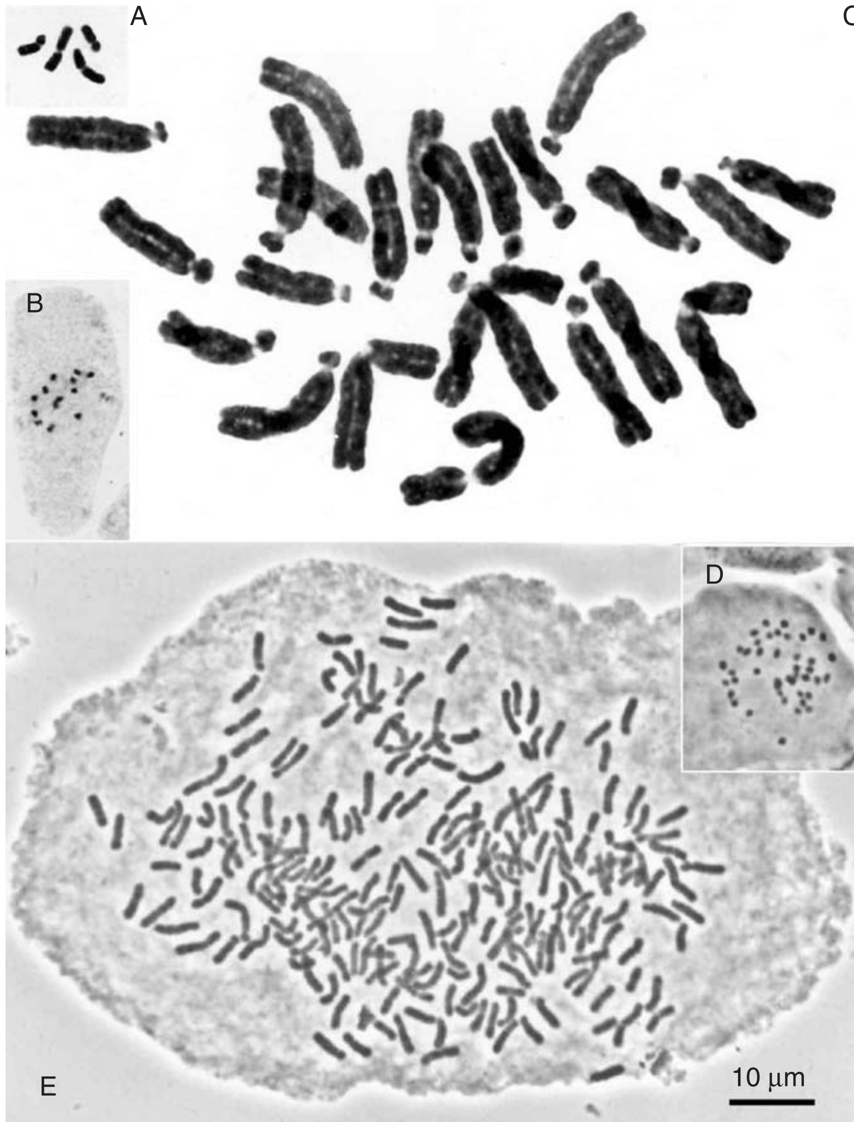

FIGURE 2.1 An example of the extensive variation in DNA amounts and chromosome sizes encountered in plants. **(A)** *Brachyscome dichromosomatica* 2n = 2*x* = 4, 1C = 1.1 pg; **(B)** *Myriophyllum spicatum* 2n = 2*x* = 14, 1C = 0.3 pg; **(C)** *Fritillaria* sp. 2n = 2*x* = 24, 1C ≈ 65 pg; **(D)** *Selaginella kraussiana* 2n = 4*x* = 40, 1C = 0.36 pg; **(E)** *Equisetum variegatum* 2n ≈ 216, 1C = 30.4 pg. **(A)** From Kenton *et al.* (1993), reproduced by permission (© Elsevier Inc.).

THE MAIN AREAS OF FOCUS OF EARLY GENOME SIZE STUDIES

Once its central role in genetics was clear, research on many aspects of DNA became greatly intensified. Naturally, this included an interest in total genome size, which was soon fuelled by the realization that although this was remarkably constant within individual organisms and species, it varied extensively among different species, often in puzzling ways that were quite independent of an organism's complexity.

Most of the early effort in studying plant genome sizes concentrated on angiosperms, especially crop or model laboratory species. Again, plant studies often led work on animals, which was especially true with regard to early cytological research on chromosomes (from the 1930s to the 1950s). For example, the chromosome number for humans (2n = 46) was still uncertain until 1956 (Ford and Hamerton, 1956; Tjio and Levan, 1956), fully 35 years after the chromosome number and ploidal level of allohexaploid[2] wheat, *Triticum aestivum* (2n = 6*x* = 42), was firmly established (Sakamura, 1918).

As in animals, much of the initial research was concerned with testing the notion of DNA constancy within species, but it soon diversified to include three main lines of inquiry (which are all still ongoing today):

1. The technical issue of developing methods for estimating genome size, and testing and improving their accuracy.
2. An exploration of the ranges in genome size in different groups and at various taxonomic levels.
3. An investigation of the meaning of genome size variation in terms of the mechanisms responsible, rates of change, and evolutionary significance. In particular, resolving the seeming contradiction between constancy within species (because DNA is the hereditary material) and diversity among species (with the associated lack of correlation with the number of genes), once termed the "C-value paradox" (Thomas, 1971), became a major theme in the study of genome size for several decades.

Although the very broad third issue was well framed in the 1950s and 1960s, progress toward an answer was blocked by a lack of knowledge about the nature of the DNA sequences responsible for variation in genome size, and of suitable techniques to investigate this. It remained so until the molecular revolution, which allowed the copy numbers and karyotypic distributions of specific DNA sequences to be investigated for the first time. In the absence of such information,

[2]See Chapter 7 for a discussion of polyploidy in plants, including definitions of allopolyploidy, hexaploidy, and other related terms.

much attention was focused instead on the issue of the phenotypic and other consequences of genome size variation, with particular reference to its practical significance (e.g., in agriculture). Studies of the broader ecological relevance of genome size followed, as discussed in more detail in a later section.

Impact of the Molecular Revolution on Genome Size Research

Although it provided much-needed insights into the structure and content of individual genomes, the molecular revolution also had an inhibitory effect on genome size research. As molecular work on DNA sequences filled the limelight, interest in plant DNA C-values per se began to fade. By the 1980s, the strong emphasis on sequence-based studies made it almost impossible to obtain grant funding to estimate genome sizes in their own right. Consequently, such information was obtained either as a by-product of studies focused primarily on other topics, or by a few laboratories or individuals with basic core funding and equipment. Nevertheless, the process of gathering more data on genome size for plant taxa continued at a low level in an uncoordinated way, although occasionally data for larger samples were gathered in order to test particular hypotheses about the patterns and consequences of DNA content variation.

The careful attention to technical detail characteristic of most early work was also abandoned in many cases, leading to reports of substantial intraspecific variation (i.e., violations of the rule of DNA constancy), including some seemingly related to developmental, environmental, or geographical factors. This was perhaps influenced by molecular studies showing that much of the genome consisted of repetitive DNA sequences that had the potential to change in copy number. In this context, some workers suggested that certain "fluid domains" within the genome were capable of undergoing rapid changes in copy number and hence to alter genome size in response to certain developmental events (e.g., Cavallini and Natali, 1991; Frediani *et al.*,1994). Around this time the concept of the "plastic genome" became popular, in direct contrast to the very notion of the "C-value."

The unquestioning assumption by many researchers that all such variation was real necessitated a second and protracted phase of methodological ground-truthing, the development of best practice techniques, and an emphasis on a more critical approach to claims of substantial variation within species (see, for example, the recommendations arising from the first and second Plant Genome Size Workshops held at the Royal Botanic Gardens, Kew in 1997 and 2003 at www.rbgkew.org.uk/cval/workshopreport.html). This second wave of very careful measurements revealed that much supposed intraspecific variation (though not all) was due to technical artifacts (e.g., Greilhuber 1998, 2005) (see later section).

Genome Size Studies in the Post-Genomic Era

The advent of large-scale genome sequencing programs in the 1990s focused new interest on the study of genome size, and facilitated detailed studies of the molecular basis of genome evolution in plants. The availability of DNA sequence information for entire genomes (or at least substantial segments thereof) of different taxa has allowed detailed comparisons at several taxonomic levels, including within species (e.g., between the subspecies *indica* and *japonica* of rice, *Oryza sativa*) (Goff *et al.,* 2002), between species within a family (such as the grasses *Oryza, Sorghum,* and *Zea*) (Ilic *et al.,* 2003) and between families (such as Poaceae and Brassicaceae) (Bennetzen and Ma, 2003). These provided the first insights into the sorts of intra- and interspecific variation at the DNA sequence level that together result in changes in genome size.

Such work is still in its infancy, and it remains unclear how typical findings based on the study of a small segment of one linkage group really are. Nevertheless, it has already been confirmed that, as expected, changes in genome size mostly involve the gain or loss of families of repeated DNA sequences (especially transposable elements, see Chapter 3) located primarily in intergenic regions. Contrary to what was once thought, it now seems that speciation may have as much to do with changes in such regions as with alterations in the sequences or arrangements of coding regions (Kubis *et al.,* 1998).

Such comparisons will form a main element of future work on plant genome size, and promise to reveal the key molecular mechanisms involved in the gain and/or loss of DNA, and the rate at which such changes can occur either in genomes at large or in different components thereof. This, in turn, will allow a more detailed investigation of the types of sequences involved in generating patterns of variation involving phenotypic and ecological correlates of DNA content. The development and increasingly broad application of new techniques for sequencing, assembling, and comparing genomes make the early 21st century a truly exciting time for all forms of genomics, including the holistic variety concerned with genome size.

THE STATE OF KNOWLEDGE REGARDING PLANT GENOME SIZES

Since 1950, more than 10,000 quantitative estimates of plant C-values have been made, covering roughly 4000 species of plants (Bennett and Leitch, 2003). Such a broad sampling can allow the key questions listed previously to be addressed in a more comprehensive way—but only if the data are accessible. In plants, unlike in animals, there has long been an effort to compile genome size data for the purposes of comparative study. Of course, there is an important feedback process involved in this case, because the compilation of data for one purpose tends to reveal gaps in the dataset and to stimulate the targeted measurement of new values.

In 1972, Bennett collected data for 273 angiosperm species to test a possible relationship between DNA amount and minimum generation time, and incidentally created the largest list of its type available at the time. Although Bennett's (1972) list proved to be a valuable reference resource, it became clear within a few years that the majority of information on plant DNA amounts remained difficult to locate, in part because much of it was widely scattered in a diverse range of journals, and still worse, because a significant proportion of existing estimates were not published anywhere. In recognition of this, Bennett and colleagues began compiling large lists of published and unpublished C-values in plants. To date, seven such lists have been published, providing extensive coverage of the available angiosperm data. Taken together, these lists give DNA amounts for more than 4000 species (1.6% of the global angiosperm flora) derived from 465 original sources (Bennett and Smith, 1976, 1991; Bennett *et al.,* 1982a, 2000a; Bennett and Leitch, 1995, 1997, 2005).

Work in compiling lists of C-values for other plant groups lagged behind that of angiosperms and it was not until 1998 that the first reference list for a non-angiosperm group was published (Murray, 1998). This contained estimates for 117 gymnosperm species (corresponding to 16% of described species), cited from 24 original sources. Since then, C-values for a further 64 gymnosperm taxa have been published from seven new reference sources (e.g., Leitch *et al.,* 2001), but they have not been compiled into a second published list.

For pteridophytes—that is, lycophytes (clubmosses) and monilophytes (including ferns and horsetails)—estimates of DNA amounts were pooled into one list by Bennett and Leitch (2001). The list contained DNA C-values for just 48 species from eight original sources and highlighted the ongoing need for work to increase knowledge in this area. Since then, new pteridophyte C-value data have been published by Obermayer *et al.* (2002) and Hanson and Leitch (2002).

In bryophytes (mosses, liverworts, and hornworts), some limited data are available but remain scattered, and there is no equivalent compilation of C-values combined from different sources into a pooled list. The largest dataset comes from the work by Voglmayr (2000), who estimated C-values in 138 mosses in a carefully targeted study whose aim was to cover a representative spectrum of taxa. Voglmayr's paper also reviewed C-value estimates made by previous workers and is thus the closest approximation to a single printed reference source for C-values in bryophytes currently available.

C-VALUES IN CYBERSPACE: DEVELOPMENT OF THE *PLANT DNA C-VALUES DATABASE*

The collected lists of angiosperm DNA amounts were produced to make data more accessible for both reference and analysis purposes. However, as the number of such lists rose, it became more cumbersome to determine whether an estimate

for a particular species was listed. By 1997, the problem had become acute, with five collected lists published, containing a total of 2802 species. It was therefore decided to pool available data into a single database and make it available on the Internet. The first version of the *Angiosperm DNA C-values Database* went live in April 1997, and was subsequently updated in 1998 to make it more user-friendly.

Following the publication of a sixth list of angiosperm DNA amounts (Bennett *et al.*, 2000a), which included first C-values for another 691 species, a third release of the database was launched in December 2000. As reference lists for other plant groups were published (i.e., gymnosperms and pteridophytes), the option to make these data more widely available electronically became possible. The C-value data for 48 pteridophytes given in Bennett and Leitch (2001) were also released in December 2000 as the online *Pteridophyte DNA C-values Database*.

These online databases proved to be very useful, and thought was therefore given to constructing counterparts for other plant taxa as data became available. Following one of the key recommendations of the first Plant Genome Size Workshop in 1997, it was decided to assemble one overarching database to cover all land plant groups (i.e., the Embryophyta: angiosperms, gymnosperms, pteridophytes, and bryophytes). The resulting *Plant DNA C-values Database* (www.rbg.kew.org.uk/cval/homepage.html) was launched in September 2001.

Since the first release, work has continued to develop and extend the database. Release 2.0 went live in January 2003 (Bennett and Leitch, 2003) and contains data for 3927 species, comprising 3493 angiosperms, 181 gymnosperms, 82 pteridophytes, and 171 bryophytes (Table 2.1).

USES AND USERS OF THE *PLANT DNA C-VALUES DATABASE*

The *Plant DNA C-values Database* is widely used, having received more than 40,000 hits since being launched in September 2001. On average, the database is visited more than 50 times per day, with ~100 C-values commonly taken in a single visit. Not surprisingly, the types of questions for which the database is used are varied. In practical terms, C-value information about individual plant species is important for planning the construction of genomic libraries, for undertaking amplified fragment length polymorphism (AFLP) or microsatellite studies (e.g., Scott *et al.,* 1999; Garner, 2002; Fay *et al.,* 2005), and for deciding which will be the next plant species to have its genome sequenced.

At the other end of the spectrum, the availability of C-value data in one central database has opened up the possibility of carrying out large-scale comparative analyses involving hundreds or even thousands of species. Studies so far reported that have used the database in this way cover diverse fields of biology including ecology, evolution, genomics, and conservation. For example, Knight and Ackerly (2002)

TABLE 2.1 **Minimum (min.), maximum (max.), mean, mode, and range (max./min.) of 1C DNA values[a] in major groups of plants, together with the level of species representation of C-value data**

	Min. (pg)	Max. (pg)	Mean (pg)	Mode (pg)	Range (max./min.)	No. species with DNA C-values	No. of species recognized[b]	Species representation (%)	No. species in *Plant DNA C-values Database*[c]	% Representation in the *Plant DNA C-values Database*[b]
Algae										
Chlorophyta	0.10	19.6	1.75	0.3	196	85	~ 6500	~ 1.3	0	0
Rhodophyta	0.10	1.4	0.43	0.2	28	111	~ 6000	~ 1.9	0	0
Phaeophyta	0.10	0.9	0.42	0.25	9	44	~ 1500	~ 2.9	0	0
Bryophytes	0.17	2.05	0.51	0.45	12.1	171	~18,000	~1.0	171	~1.0
Pteridophytes										
Lycophytes	0.16	11.97	3.81	n/a	74.8	4	~900	~0.4	4	~0.4
Monilophytes	0.77	72.68	13.58	7.8	95.0	63	~9000	~0.7	63	~0.7
Gymnosperms	2.25	32.20	16.99	9.95	14.3	181	~ 730	~ 24.8	181	~ 24.8
Angiosperms	~0.11	127.40	6.30	0.60	~1000	4119	~250,000	~1.6	3493	~1.4
All land plants	~0.11	127.40	6.46	0.60	~1000	4538	~280,000	~1.6	3927	~1.4

[a]C-value data for algae taken from Kapraun (2005), and for bryophytes, lycophytes, monilophytes, gymnosperms and angiosperms from Bennett and Leitch (2003) and Bennett and Leitch (2005).

[b]Numbers of species recognized; taken from Kapraun (2005) for algae, Qiu and Palmer (1999) for bryophytes, lycophytes, and monilophytes, Murray *et al.* (2001) for gymnosperms, and Bennett and Leitch (1995) for angiosperms.

[c]*Plant DNA C-values Database* (release 2.0, January 2003) (Bennett and Leitch, 2003).

used the database to extract 401 angiosperm C-values to ask ecological questions such as how DNA amount varied across environmental gradients. This has been aided by the inclusion of information in addition to genome size and taxonomic classification, such as chromosome number, ploidy level, and the method used to analyze DNA content. Specific information relating to the various plant groups is also provided, such as life cycle type (annual, biennial, or perennial) in angiosperms, sperm type (multiflagellate or none) in gymnosperms, and spore type (homosporous or heterosporous), sporangium type (eusporangiate or leptosporangiate), and sperm flagella number (biflagellate or multiflagellate) in pteridophytes.

In the field of genomics, Leitch and Bennett (2004) used data for 3021 angiosperm species to provide insights into the dynamics of C-value evolution in polyploid species. Vinogradov (2003) extracted C-values for 3036 species from the database to reveal a startling negative relationship between the genome size of a species and its current extinction risk status. Most broadly of all, insights into the evolution of DNA amounts across all angiosperms (Soltis *et al.,* 2003), and indeed all land plants (Leitch *et al.,* 2005), have recently been provided by using C-values for more than 140 families included in the database.

PATTERNS IN PLANT GENOME SIZE EVOLUTION

Understanding the evolution of plant genome size involves at least four major components. First, it is necessary to identify the overall distributional patterns of variation observed within and among extant plant taxa. Second, it is important to determine the historical trends that generated the current patterns, and to establish the basic directionality of genome size change through evolutionary time. Third are the questions relating to the mechanisms by which genomes change in size. Fourth are the phenotypic consequences that may influence both the taxonomic and geographical distribution of genome size variation among species. These issues, all of which deal with variation across species, will be treated in order in the following sections, followed later by a discussion of the related issue of variation within species.

THE EXTENT OF VARIATION ACROSS PLANT TAXA

In photosynthetic organisms, reported C-values vary more than 12,000-fold, from ~0.01 pg in the unicellular alga *Ostreococcus tauri* (Prasinophyceae) (Courties *et al.,* 1998) to more than 127 pg in the angiosperm *Fritillaria assyriaca* (Bennett and Smith, 1991). Yet the range, minimum, maximum, mean, and modal C-values vary considerably between the different groups for which data have been compiled (Table 2.1). These differences are further highlighted when data are

plotted as histograms (Fig. 2.2). Whereas each group for which data are available contains species with small C-values, the upper limit appears to vary greatly. The smallest range in C-values is found in the Phaeophyta (brown algae, 9-fold), and the largest in the angiosperms (~1000-fold) (Tables 2.1 and 2.2).

However, it should be noted from Table 2.1 that the percent species representation for all but gymnosperms is poor, meaning that the range and distribution of C-values reported may not be entirely representative. (This may be especially true for groups in which fewer than 100 C-values are known.) The exception to this might be the angiosperms, as the ~1000-fold range was first reported more than 20 years ago based on C-values for 993 species (Bennett *et al.,* 1982a) but has not changed even after adding a further 3126 species.

GENOME SIZE IN A PHYLOGENETIC CONTEXT

Given the large range in DNA amounts encountered in plants, any attempt to investigate the directionality of genome size evolution requires that the data are viewed within a rigorous phylogenetic framework. In plants, there are only a few cases where this has been done (Bennetzen and Kellogg, 1997; Cox *et al.,* 1998; Leitch *et al.,* 2001; Obermayer *et al.,* 2002; Wendel *et al.,* 2002). In most cases, this has involved examining variation within individual families of angiosperms, but large-scale analyses across the angiosperm phylogeny have also been conducted, first by Leitch *et al.* (1998) and then extended using character-state mapping by Soltis *et al.* (2003). Although an analysis of a continuously varying character such as genome size presents problems when defining character states, it was possible to partially circumvent this difficulty by assigning genome sizes to a series of distinct categories (Soltis *et al.,* 2003): "very small" (C-values ≤1.4 pg), "small" (>1.4 to ≤3.5 pg), "intermediate" (>3.5 to <14.0 pg), "large" (≥14.0 to <35 pg), and "very large" (≥35 pg). With the exception of the "intermediate" category, these size classes were the same as those first defined by Leitch *et al.* (1998) based on the modal C-value of 0.7 pg for a sample of 2802 species available at the time of analysis. In this sense, the very small and small C-value categories were twice and five times the mode, respectively, while large and very large C-value categories were 20 times and 50 times the mode, respectively.

Using gymnosperms as the outgroup, character-state reconstruction data showed that a very small genome was the ancestral state not only at the root of angiosperms, but also for most major clades within angiosperms (e.g., monocots, magnoliids, all core eudicots, Caryophyllales) (Fig. 2.3) in agreement with the earlier analysis of Leitch *et al.* (1998). The evolution of very large genomes was shown to be phylogenetically restricted to a few derived families within the monocot and Santalales clades (Fig. 2.3), suggesting that very large genomes have evolved independently more than once during the evolution of angiosperms (Leitch *et al.,* 1998; Soltis *et al.,* 2003).

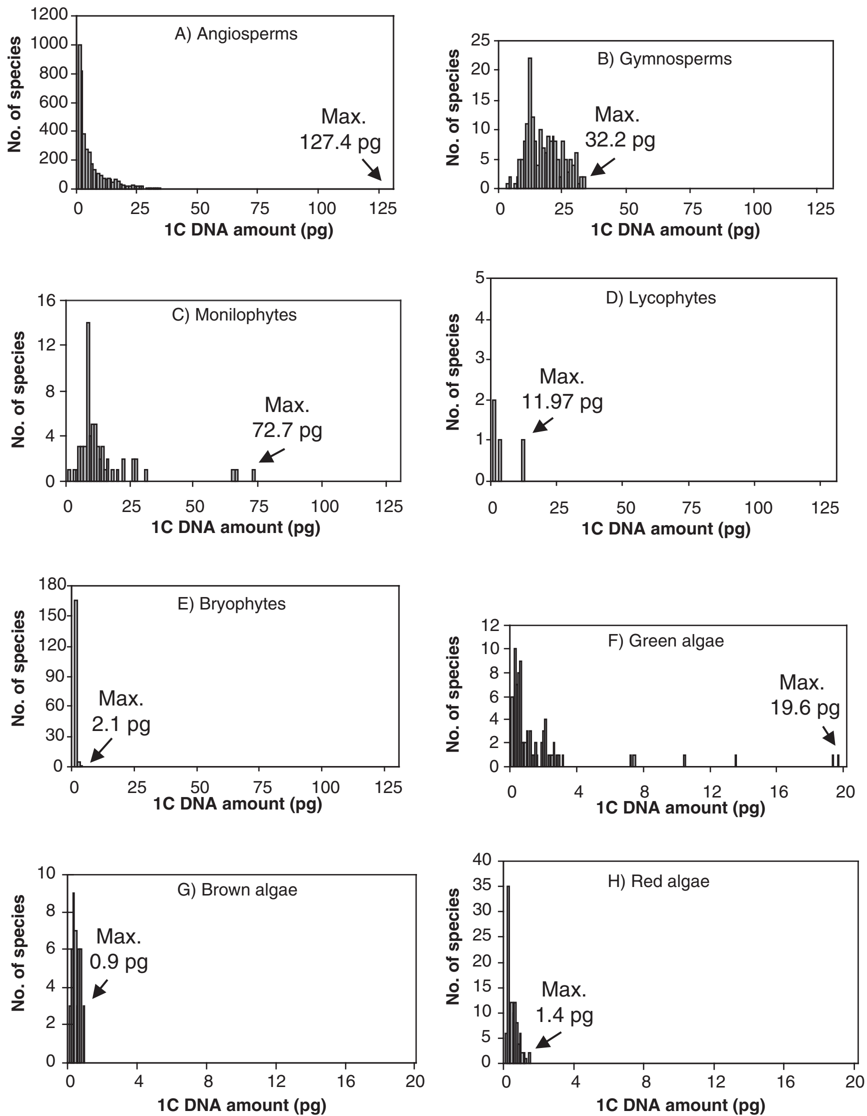

FIGURE 2.2 Histograms showing the distribution of DNA C-values for (A) 4119 angiosperms, (B) 181 gymnosperms, (C) 63 monilophytes, (D) 4 lycophytes, (E) 171 bryophytes, (F) 85 green algae (Chlorophyta), (G) 44 brown algae (Phaeophyta), and (H) 111 red algae (Rhodophyta). The maximum C-value for each group is indicated. (A–E) Redrawn from Leitch *et al.* (2005), reproduced by permission (© Oxford University Press); (F–H) Data from Kapraun (2005).

TABLE 2.2 Some well-known representative species showing the range of 1C DNA amounts in angiosperms

Species	Common name	Chromosome number (2n)	Ploidy level (x)	1C DNA amount	
				pg	Mb[a]
Arabidopsis thaliana	Thale cress	10	2	0.16	157
Oryza sativa	Rice	24	2	0.50	490
Lycopersicon esculentum	Tomato	24	2	1.00	980
Glycine max	Soybean	40	2	1.10	1078
Zea mays	Maize	20	2		
Seneca 60 line				2.50	2450
Zapalote Chico line				3.40	3332
Hordeum vulgare	Barley	14	2	5.55	5400
Secale cereale	Rye	14	2	8.28	8110
Vicia faba	Bean	12	2	13.33	13,060
Allium cepa	Onion	14	2	16.75	16,415
Triticum aestivum	Wheat	42	6	17.32	16,970
Lilium longiflorum	Easter lily	24	2	35.20	34,500
Fritillaria assyriaca	Fritillaria	48	4	127.40	124,850

[a] 1 pg ≈ 980 Mb (see Chapter 1).

FIGURE 2.3 Parsimony reconstruction of C-value diversification in the angiosperms using the “all most parsimonious states” resolving option of MacClade (based on Soltis *et al.*, 2003). SANT = Santalales. From Leitch *et al.* (2005), reproduced by permission (© Oxford University Press).

Extending the analysis to include other land plant groups (i.e., gymnosperms, monilophytes, lycophytes, and bryophytes) has provided both insights into the size of ancestral genomes at different parts of the land plant phylogeny and evidence for the bidirectionality of genome size evolution (Leitch *et al.*, 2005) (Fig. 2.4). The main results can be summarized as follows:

1. Different land plant groups are characterized by different ancestral genome sizes. Whereas angiosperms and bryophytes are reconstructed with very small ancestral genomes (i.e., ≤1.4 pg), in gymnosperms and most branches of monilophytes the ancestral genome size is reconstructed as intermediate (i.e., >3.5 to <14.0 pg).
2. Genome size evolution across land plants has been dynamic, with evidence of several independent increases and decreases taking place. Examples of genome size reductions are evident within the monilophytes at the base of the heterosporous water ferns and within the gymnosperms in the branch leading to Gnetaceae (Gnetales). Evidence of large independent increases is seen in the Ophioglossaceae + Psilotaceae clade (monilophytes) and within heterosporous water ferns in Marsileaceae. Thus, observations made within Malvaceae (angiosperms) that both increases and decreases can take place during genome size evolution (Wendel *et al.*, 2002) appear to form a pattern that is repeated across land plants, except perhaps in bryophytes where all species to date have small or very small genomes (Table 2.1).
3. The differences in C-value profiles and patterns of evolution based on reconstruction data strongly suggest that each major group of land plants has been subject to different evolutionary forces. Conversely, it is likely that genome size has itself influenced the shape of the overall plant phylogeny. As will be discussed in a later section, a small genome size correlates with several developmental phenotypic characters (e.g., rapid seedling establishment, short minimum generation time, reduced cost of reproduction, and increased reproductive rate) (Bennett, 1972, 1987; Midgley and Bond, 1991), which together may permit greater evolutionary and ecological flexibility. Thus, the smaller genome sizes of angiosperms may provide one functional explanation as to why they have become so dominant in the global flora relative to the larger-genomed groups like gymnosperms and many monilophytes (Leitch *et al.*, 1998).

HOW DO PLANT GENOME SIZES EVOLVE?

The absence at the time of any known mechanism for decreasing DNA amount led Bennetzen and Kellogg (1997) to speculate that plants may have a "one-way ticket

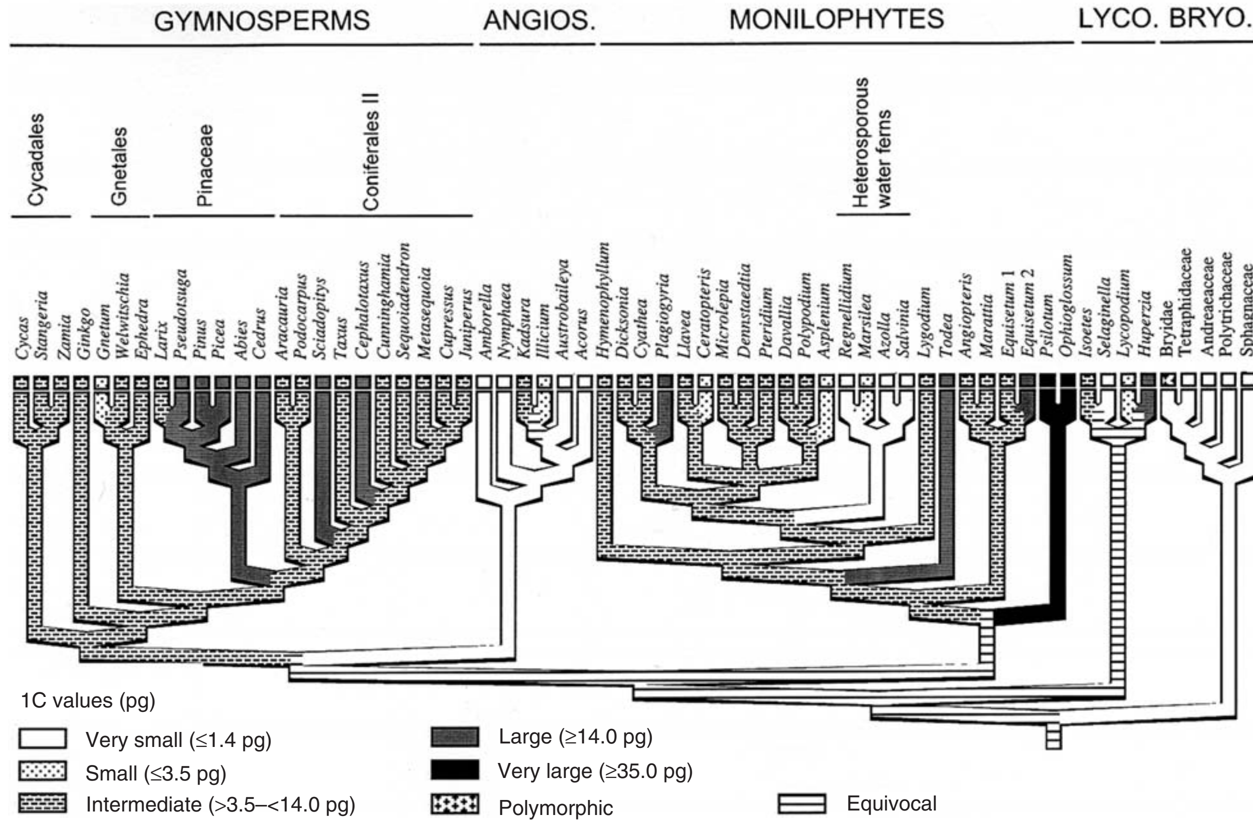

FIGURE 2.4 Parsimony reconstruction of C-value diversification across land plants using the "all most parsimonious states" resolving option of MacClade. *Equisetum* 1 = subgenus *Equisetum*, and *Equisetum* 2 = subgenus *Hippochaete*. ANGIOS. = angiosperms; LYCO. = lycophytes; BRYO. = bryophytes. From Leitch *et al.* (2005), reproduced by permission (© Oxford University Press).

to genomic obesity" through amplification of retrotransposons and polyploidy. However, as noted previously, there is considerable evidence that both increases and decreases may occur in plant lineages. The revised view is that the net DNA amount of an organism reflects the dynamic balance between the opposing forces of expansion and contraction, in terms of both mechanisms (e.g., Petrov, 2001) and selective consequences (e.g., Gregory, 2001a). Thanks to detailed studies at the DNA sequence level, an understanding of the mechanisms by which genomes change in size is finally taking shape. In many ways, plant studies have led the way on this issue, even though most of the large-scale eukaryotic sequencing projects conducted to date have been from animals and fungi.

Sequences Responsible for the Range of Genome Sizes Encountered in Plants

It is generally accepted that different amounts of noncoding, repetitive DNA are primarily responsible for the large range in genome sizes observed in plants (e.g., >70% of some plant genomes are repetitive DNA) (Flavell *et al.,* 1977; Barakat *et al.,* 1997). However, information on the exact nature of the repetitive DNA involved, such as the length of each repeat and their relative contributions, remains elusive in most species (Feschotte *et al.,* 2002).

Based on available data, it is clear that much of the repetitive DNA in plants is composed of transposable elements (TEs). As discussed in detail in Chapter 3, these can be divided into two distinct classes. Class I elements use an RNA-mediated mode of transposition, whereas Class II elements use DNA-mediated transposition mechanisms.

Class I elements are further divided into two subclasses: (1) retrotransposons, which are characterized by direct long terminal repeats (LTRs) (e.g., *Ty1-copia* and *Ty3-gypsy* elements) and appear to be ubiquitous in vascular plants (Voytas *et al.,* 1992; Suoniemi *et al.,* 1998), and (2) retroposons, which lack terminal repeats and are referred to as non-LTR retroelements (e.g., long interspersed nuclear elements [LINEs] such as *Cin4* in *Zea mays,* and *del2* in *Lilium speciosum,* and short interspersed nuclear elements [SINEs] such as the *S1* element in *Brassica*). Both subclasses may reach very high copy numbers in plants.

In grasses (Poaceae), LTR-retrotransposons are clearly the most abundant type of transposable elements, and in some species may comprise more than 60% of the nuclear genome (Vicient *et al.,*1999a; Wicker *et al.,* 2001). In maize, for example, estimates of the size and copy number of retrotransposons in a 240 kilobase (kb) region flanking the *adh1* gene suggested that 33–62% of the genome was composed of high copy number retrotransposons, with an additional 16% of the genome containing low to middle copy number retrotransposons (SanMiguel and Bennetzen, 1998). By comparing the structure of the same 240 kb DNA segment

in maize (*Zea mays*) and sorghum (*Sorghum bicolor*) it was further shown that the approximately 3-fold larger genome of maize was predominantly due to the presence of retrotransposon sequences that had inserted into the maize genome (SanMiguel and Bennetzen, 1998). Thus, whereas the order of the genes remains largely conserved between the two species, extensive differences in genome size have emerged since their split ~20 million years ago, predominantly reflecting the different extent to which retrotransposons have undergone amplification in the two lineages since that time (Fig. 2.5).

It has been speculated that LTR-retrotransposons play an important role in determining the size of plant genomes in general (Kumar and Bennetzen, 1999). However, although this may be the case in grasses, sequences other than these appear to have a greater influence on genome size differences in other organisms (Wendel and Wessler, 2000), especially those with smaller genomes (Kidwell, 2002).

Class II elements include the *Helitrons, Mu,* and *mutator*-like elements (MULEs), which tend to be large (up to 20 kb), and the smaller miniature inverted repeat transposable elements (MITEs) (0.1–0.5 kb). From available sequence data, Class II elements make up ~6% and ~12% of the *Arabidopsis thaliana* and *Oryza sativa* genomes, respectively (Feschotte *et al.,* 2002; Jiang *et al.,* 2004).

What Triggers the Spread of Transposable Elements?

From the previous section, it is evident that transposable element proliferation plays a major role in increasing plant genome sizes through time. However, the factor(s) responsible for triggering amplification in certain lineages are still not clearly understood in most plant systems. In fact, there remains a puzzling discrepancy between the large number of plant TEs characterized to date and their apparent transcriptional silence observed during normal plant development. The few exceptions to this include the *BARE-1* element from barley (Vicient *et al.,* 1999a), the related *OARE-1* element from oats (Kimura *et al.,* 2001), and *IRRE* elements from some *Iris* species (Kentner *et al.,* 2003), which have been shown to be transcriptionally active under normal growing conditions.

In at least some cases, transcriptional activation can be induced by experimental manipulations of various biotic or abiotic stresses such as wounding, tissue culture, and pathogen attack (e.g., Grandbastien, 1998; Feschotte *et al.,* 2002). Similar effects have also been documented in a natural setting, as with the recent report of retrotransposon activation in natural populations of *Hordeum spontaneum* by Kalender *et al.* (2000). These authors showed that the *BARE-1* LTR-retrotransposon, which comprises ~3% of the *H. spontaneum* genome (Vicient *et al.,* 1999a), had been highly insertionally active in recent times within different plant populations from the single location of "Evolution Canyon" in Israel. They also showed that copy number

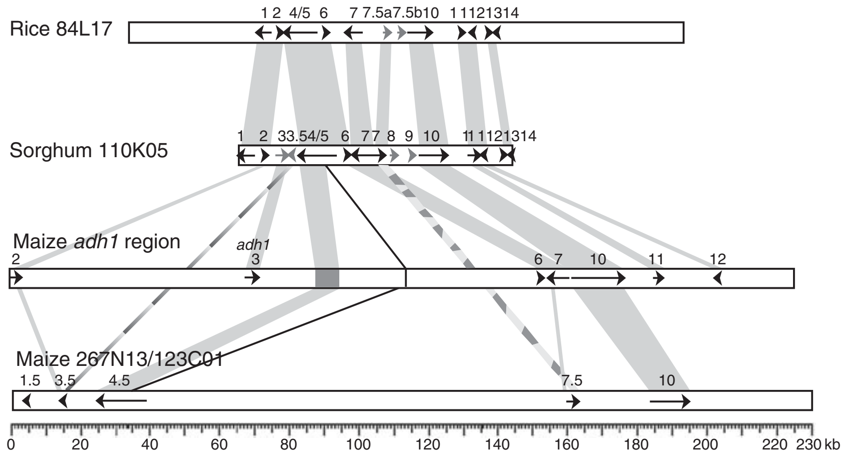

FIGURE 2.5 Comparison between the organization of colinear genomic regions in rice, sorghum, and two homoeologous segments in maize. Shaded areas connecting the regions represent conserved sequences. The length of each segment is drawn to scale and illustrates the large expansion in length, mainly owing to insertion of retrotransposons, that has taken place since *Sorghum* and maize diverged 20 million years ago. From Ilic *et al.* (2003), reproduced by permission (© National Academy of Sciences).

varied 3-fold between different populations and that there was a correlation between copy number and various ecological variables, most notably water availability. Specifically, the population with the highest copy number of *BARE-1* elements occupied the driest site in the canyon, suggesting the possibility that water stress induced the activation of *BARE-1*. Although the question as to the role *BARE-1* might play in the physiological stress response remains to be determined, the data highlight the potential for natural environmental cues to trigger retrotransposon activity. Interestingly, analysis of the draft sequence of the rice genome has also shown that a DNA transposon (*mPing* MITE) has been preferentially amplified in cultivars adapted to environmental extremes (Jiang *et al.*, 2003).

A few studies have suggested that polyploidization and interspecific hybridization may also trigger TE amplification in plants, for example in *Nicotiana* (Fig. 2.6A) (Comai, 2000), *Aegilops-Triticum* allopolyploids (Kashkush *et al.*, 2002), and *Spartina anglica* (Baumel *et al.*, 2002). However, this is clearly not always the case. The doubling of the maize genome as a result of LTR-retrotransposon activity was estimated to have taken place primarily in the last three million years (SanMiguel *et al.*, 1998), roughly eight million years after the polyploidization event that gave rise to the species (Gaut and Doebley, 1997). Similarly, bursts of retrotransposon amplification estimated to have occurred in the last two million years in *Arabidopsis thaliana* (Devos *et al.*, 2002) and the last five million years in *Oryza sativa* (Vitte and Panaud, 2003)

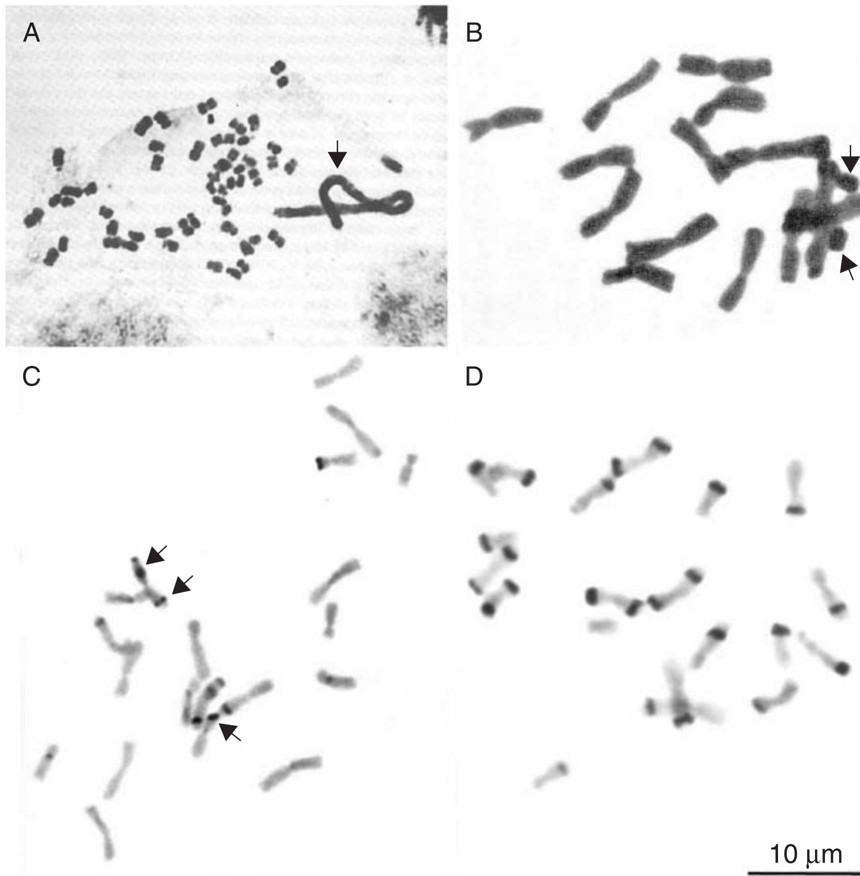

FIGURE 2.6 Examples of intraspecific variation in DNA amount visible at the chromosome level. **(A)** Megachromosome (arrowed) observed in a corolla metaphase cell from a *Nicotiana tabacum* × *N. otophora* hybrid made by Gerstel and Burns (1966). **(B)** Loss of heterochromatin blocks of DNA (arrowed) from chromosomes of *Secale kuprijanovii*. **(C,D)** Intraspecific differences in amount of heterochromatic DNA visible as C-bands (some of which are arrowed) in maize, *Zea mays* ssp. *mays*. **(C)** Commercial hybrid Seneca 60 (1C = 2.5 pg), a line of maize with few heterochromatic C-bands; **(D)** Race Zapalote Chico Oaxaca 50 (1C = 3.4 pg), a line of maize with many C-bands. **(A)** From Reed (1991), reproduced by permission (© Elsevier Inc.), **(B)** From Gustafson *et al.* (1983), reproduced by permission (© Springer-Verlag), **(C,D)** From Laurie and Bennett (1985), reproduced by permission (© Nature Publishing Group).

do not coincide with any known hybridization or polyploidization event. Further, where TE amplification has been shown to take place following hybridization, this may be quickly silenced by other genetic and epigenetic events such as methylation (e.g., Liu and Wendel, 2000), so the effects of TE amplification owing to hybridization

and polyploidization may be minimal. The recent studies in *Arabidopsis thaliana* showing that TE amplification may be under epigenetic control, together with the development of tools such as transposon display or sequence-specific amplification polymorphism (S-SAP) for detecting TE activity, offer the potential for significant progress to be made in this field in the near future (Feschotte *et al.*, 2002).

In summary, whereas TEs contribute greatly to the large variation in genome size observed among many plant species, especially the grasses, the factors responsible for triggering their amplification still remain unknown in the majority of cases. Moreover, it remains to be determined to what extent TE amplification contributes to changes in total genome size within a species. The study by Kalendar *et al*. (2000) discussed earlier showed that although copy number of the *BARE-1* retrotransposon varied considerably between different ecological sites, no correlation was found between genome size and copy number of the repeat. Such results suggest that although TE amplification clearly has the *potential* to increase genome size, whether or not it actually does so, or instead is compensated for by a reduction in sequence repeats elsewhere in the genome, may depend on other factors within the cell. The possibility that genome size is maintained by internal stabilizing mechanisms will be an important consideration in the discussion of intraspecific variation given later.

Satellite DNA

Tandemly arranged repeats of identical or similar sequences (satellite DNA) can also comprise large fractions of plant genomes. Indeed, some of the first satellite DNAs to be isolated were from rye (*Secale cereale*), with one type comprising ~6% of the genome (Bedbrook *et al.,* 1980). Although satellite sequences are variable in size, the most common monomeric units are 150–180 bp and 320–380 bp. The structure of the repeat may be highly complex, and in some cases may include DNA from other repeat classes. For example, a centromeric satellite sequence in *Brassica campestris* was shown to contain DNA sequences related to both tRNA genes and SINEs (Kubis *et al.,* 1998). Minisatellites (10–40 bp repeats) and microsatellites (2–6 bp, also called simple sequence repeats or SSRs) may also represent an appreciable amount of DNA, such as in telomeric sequences. These are believed to occur ubiquitously in all eukaryotic genomes, but their number and the proportion of the genome they occupy vary significantly among species (Ellegren, 2002; Morgante *et al.,* 2002).

Genome Size Increase by Polyploidy

Polyploidy, resulting from combining three or more basic chromosome sets or genomes in a single nucleus, is a prominent mode of speciation, especially

in angiosperms and monilophytes (see Chapter 7), and results in an instant increase in the DNA content of the nucleus (i.e., its C-value). Its prominence in plant evolution has been brought into even sharper focus by recent large-scale genomic analyses that have uncovered the polyploid nature of many plant genomes that were traditionally considered to be diploid, including maize (Moore *et al.,* 1995) and *Arabidopsis thaliana* (Bowers *et al.,* 2003). The fact that even small-genomed plants such as these appear to be ancient polyploids has led to the suggestion that *all* angiosperms may have experienced polyploidization at some point in their evolutionary history (Wendel, 2000) (see Chapter 7).

The issue of polyploidy leads to terminological complications when referring to "C-value" versus "genome size." Specifically, for a diploid with just one genome in each gametic nucleus, C-value and genome size are interchangeable. However, in a polyploid, the C-value will represent the total DNA amount of all genomes within the nucleus (e.g., in a tetraploid, each gametic nucleus will have two genomes, whereas a hexaploid will have three). Thus in polyploids, C-value and "basic" genome size *sensu stricto* are not equivalent: in a polyploid with more than two genomes in the gametic nucleus, basic genome size will always be smaller than the C-value. In general, the basic genome size of a polyploid can be estimated by dividing C-value by the number of genomes in the gametic nucleus (i.e., half the ploidy) but it should be recognized that this only gives an accurate estimate in taxa with equal genome sizes. In taxa with genomes of different sizes (e.g., some allopolyploids) it gives only a mean genome size. The value will be close to the actual genome sizes in most but not all species (e.g., some taxa with bimodal karyotypes). As polyploidy is so prevalent, especially within angiosperms and monilophytes, this has led some authors to suggest that terminology should distinguish between the 1C-value, representing the original definition of C-value (which is independent of ploidy level), and the $1C_x$-value, which indicates the basic genome size (Greilhuber *et al*., 2005).

In recognizing the difference between basic genome size and C-value it becomes clear that polyploidy will only result in an increase in C-value and not basic genome size. In simple terms, the expectation in new polyploids is that C-value will increase in direct proportion with ploidal level, and that the basic size of the individual genomes included will be unchanged. This expectation is observed in some polyploid series, especially those newly formed (e.g., see Pires *et al.,* 2004). However, there are many examples suggesting that C-values in particular polyploids are less than the expected sum of parental genomes (e.g., Ozkan *et al.,* 2003; Leitch and Bennett, 2004). On a larger scale, a recent analysis of 3008 angiosperms revealed that mean 1C DNA amount did not increase in direct proportion with ploidy (Fig. 2.7A), and thus that mean basic genome size (calculated by dividing 1C value by half the ploidy) tended to decrease with increasing ploidy (Fig. 2.7B) (Leitch and Bennett, 2004). These results suggest that "genome downsizing" following polyploid formation may be a widespread phenomenon of considerable biological significance.

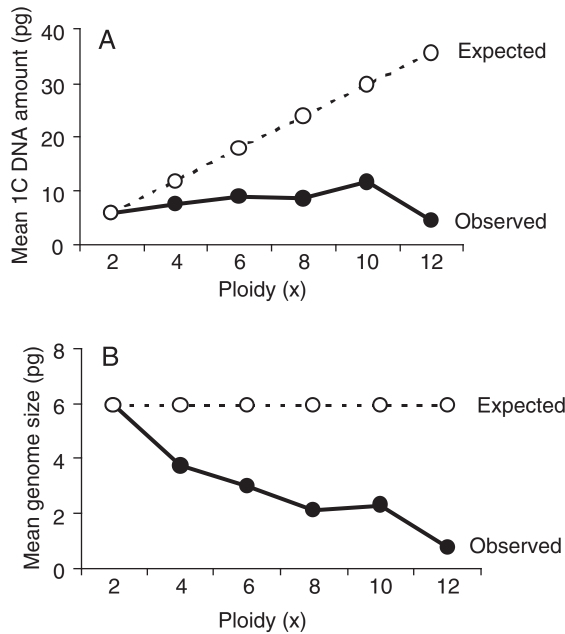

FIGURE 2.7 DNA C-values and basic genome sizes (2C value divided by ploidy) in an "all angiosperms" sample with known C-values and even ploidies between 2*x* and 12*x*. **(A)** Mean 1C DNA values observed (—●—) compared with expectation (- -○- -) assuming C-value and ploidy are directly proportional. **(B)** Mean basic genome size values observed (—●—) compared with expectation (- -○- -). From Leitch and Bennett (2004), reproduced by permission (© Blackwell Publishing).

In fact, the loss of DNA following polyploidy has been reported both at the gene level (e.g., loss of ribosomal DNA genes, reviewed by Wendel, 2000) and on a genomewide scale (e.g., Shaked *et al.,* 2001; Ilic *et al.,* 2003; Ozkan *et al.,* 2003). For example, a comparison of DNA sequences in orthologous regions of maize, sorghum, and rice suggested that maize may have lost up to 40% of its genes in the analyzed regions since the polyploid event that gave rise to it (Ilic *et al.,* 2003) (see Fig. 2.5). Mechanisms responsible for the loss remain ambiguous, but there are some indications that this involves unequal crossing over.

In some cases, polyploid formation involves the bringing together of genomes that may be similar enough to undergo homoeologous chromosome pairing during meiosis (i.e., between the equivalent chromosomes derived from the different parental species; see Chapter 7). This may lead indirectly to DNA loss as a result of a breakdown in the postreplicative mismatch repair system (Comai, 2000), which corrects any mismatched base pairs or displaced loops in double-stranded DNA following DNA replication and normally also blocks

homoeologous recombination by the binding of mismatch repair proteins. In a newly formed polyploid, the excessive number of mismatches caused by homoeologous recombination may lead to the depletion of the mismatch repair proteins, and consequently to a rise in the amount of homoeologous recombination and associated deletion of DNA. A positive feedback loop could be envisaged with increasing homoeologous recombination and associated DNA deletions as the mismatch repair proteins are used up (Comai, 2000). How much DNA is lost from polyploids by this mechanism may be species-dependent, because the extent of homoeologous recombination in different polyploids may vary. It should be noted that this is probably not a ubiquitous mechanism, as there are cases in which polyploidy is associated with only a minimal level of homoeologous recombination (e.g., in *Brassica* polyploids) (Axelsson *et al.*, 2000).

In short, polyploidy will undoubtedly lead to an increase in C-value, but the extent to which this persists over evolutionary time may vary considerably from one case to the next. Also, as discussed in Chapter 1, this will constitute a change in genome size, strictly defined, only following rediploidization.

Mechanisms of Genome Size Decrease

Mechanisms responsible for bringing about a reduction in DNA amount have only recently come to light, and there is still much to learn regarding their significance in genome evolution. However, already the available studies suggest that deletional mechanisms may play a more prominent role in genome size evolution than previously recognized. Several such mechanisms have been identified in plants.

Unequal Intrastrand Homologous Recombination

The process of unequal intrastrand homologous recombination occurs between the long terminal repeats of LTR-retrotransposons, and can lead to deletion of the internal DNA segment and one LTR, leaving behind only one "solo LTR." If this occurs between adjacent LTR-retrotransposons of the same family, several elements may be lost in a single step. That this mechanism could provide a way to counteract retroelement-driven genome expansion was first proposed by Vicient *et al.* (1999b), based on an observed 16-fold excess of solo LTRs to intact *BARE-1* elements in barley.

Subsequent research has shown that the extent to which unequal homologous recombination takes place between individual LTRs, and hence its contribution to genome downsizing, may depend on the particular LTR-retrotransposon in question. Notably, recent studies by Vicient and Schulman (2002) and Vitte and Panaud (2003) found that the proportion of solo LTRs varied considerably among retroelement types in rice. For example, the ratio of solo LTRs to complete retroelements ranged from 6.3 : 1 for the *RIRE1 Ty1-copia*–like element to 0.1 : 1

for the *Retrosat1 Ty3-gypsy*–like element. It was speculated that these differences may be determined by the specific characteristics of the retroelement such as the preferential insertion site (e.g., centromeric versus dispersed along the chromosome).

The length of the LTR element may also be a crucial factor in determining the efficacy of this deletional mechanism. Thus Shirasu *et al*. (2000) suggested that the lower frequency of solo LTRs in maize arose from a lower recombination efficiency between the comparatively short LTRs of maize retroelements (average 450 bp) compared with the longer LTRs of five barley retroelements studied (1.5–4.9 kb). Interestingly, the only two solo LTRs identified in the analysis of a 240 kb segment of the maize genome had longer than average LTRs (1.1 kb) (SanMiguel *et al*., 1996). Further, Vitte and Panaud (2003) noted that the proportion of solo-LTRs increased with increasing LTR size in three rice retroelements. Overall, the extent to which DNA is lost via unequal homologous recombination may be determined, at least in part, by the characteristics of the LTR-retrotransposons that comprise a species' genome.

Whether unequal homologous recombination occurs continuously through time or is triggered by bursts of retrotransposon amplification (as proposed by Rabinowicz, 2000) remains to be determined, although preliminary data from a study of three *Ty3-gypsy*–like retroelements in rice suggest that solo LTR formation seems to be concomitant with the amplification of active copies of the retroelements (Vitte and Panaud, 2003).

Illegitimate Recombination

Illegitimate recombination, or recombination that does not require the participation of a recA protein or large (>50 bp) stretches of sequence homology, can be the product of many different mechanisms including slipped strand repair and double strand break repair. Because it does not require such long stretches of homologous sequences to work as unequal homologous recombination, it has the potential to occur within any region of the genome. Indeed, an analysis by Devos *et al.* (2002) of LTR-retrotransposons concluded that illegitimate recombination was the main driving force behind genome size decrease in *Arabidopsis thaliana,* removing at least five times more DNA than unequal homologous recombination because it can act on a larger fraction of the genome. A role for illegitimate recombination as a mechanism to remove DNA has also been suggested by Bennetzen *et al*. (2005) and Ma *et al*. (2004) in rice, and by Wicker *et al.* (2003) in their study of the much larger genomes of *Triticum* species.

Loss of DNA During the Repair of Double Stranded Breaks

The repair of double stranded breaks (DSBs) in plant DNA is often accompanied by DNA deletions, although insertions may also occur in some species.

Kirik *et al.* (2000) compared the products of DSBs between *Arabidopsis thaliana* and *Nicotiana tabacum* (tobacco), two species differing more than 20-fold in DNA amount. They found that the size of the deletions differed markedly between the two species: *Arabidopsis* deletions were on average larger than in tobacco, and were not associated with insertions. The apparent negative correlation between the size of the deletions and genome size led to the speculation that species-specific differences in DSB repair pathways may contribute significantly to the evolution of genome size. This has been supported by recent data showing differences in the mechanisms used by *Arabidopsis* and tobacco to repair DSBs (Filkowski *et al.*, 2004). Another component, namely differences in the stability of the free DNA ends resulting from DSBs, was assessed by Orel and Puchta (2003). They established that free DNA ends were more stable in tobacco than *Arabidopsis,* owing to lower DNA exonuclease activity and/or better protection of the DNA break ends. The implication is that if such patterns were observed to occur on a wider range of species, then differences in the degree of exonucleolytic degradation of DNA ends might prove to be an important force in the evolution of genome size (Orel and Puchta, 2003).

KEY CORRELATES OF GENOME SIZE ACROSS PLANT SPECIES

The recognition that DNA amount in eukaryotes varies over several orders of magnitude, even in related groups of organisms of similar complexity, has provoked considerable interest in the biological effects and other consequences of such differences. The first studies indicating that this is an important line of enquiry were in animals (see Chapter 1), but, as seen in the following section, work on plants has also played a major role in developing this field.

EARLY WORK ON THE PHENOTYPIC CONSEQUENCES OF GENOME SIZE VARIATION IN PLANTS

Much of the earliest work on the phenotypic effects of genome size variation in plants was motivated by strong practical, and even political, interests. During the early Cold War days of the 1950s and 1960s, much attention was focused on the relationship between genome size and radiosensitivity in plants (e.g., Sparrow and Miksche, 1961; Abrahamson *et al.*, 1973) (Fig. 2.8). That genome size studies were seen as potentially important for national security interests explains the generous funding obtained for such work by a group at the Brookhaven National Laboratory (Upton, NY) led by Arnold Sparrow.

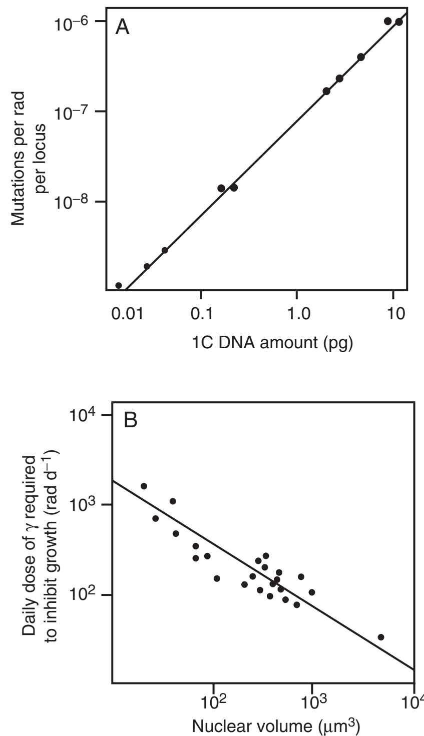

FIGURE 2.8 **(A)** Relationship between forward mutation rate per locus per rad and the 1C DNA amount; **(B)** The relationship between nuclear volume and radiosensitivity in 23 species of plants. **(A)** From a figure redrawn in Bennett (1987) based on a figure originally presented by Abrahamson *et al.* (1973), reproduced by permission (© Blackwell Publishing); **(B)** Data from Sparrow and Miksche (1961).

Sparrow's analyses included a major emphasis on plant materials to develop a theoretical understanding of the principles involved in radiation sensitivity. This was primarily because plants (1) display a very wide range of DNA content, in terms of both polyploidy and basic genome size; (2) represent staple food sources of national and global strategic significance; (3) are readily available and remain immobile during prolonged exposure to various types of radiation treatment; and (4) are socially more acceptable than animal subjects for work that involves killing large numbers of specimens. Such studies first revealed close correlations between radiosensitivity and nuclear volume (Sparrow and Miksche, 1961), then showed a positive relationship between nuclear volume and DNA amount (Baetcke *et al.*, 1967), and hence revealed an association between DNA content and sensitivity to

nuclear radiation. There was considerable discussion as to which was the most meaningful and determining character given the possibly confusing effects of ploidy level (*x*), heterochromatin, and chromosome number (2n), and it was concluded that interphase chromosome volume (i.e., mean DNA amount per chromosome, equal to nuclear DNA amount ÷ 2n) gave the closest correlation (Baetcke *et al.,* 1967).

From the perspective of genome size evolution, the correlations with sensitivity to intensive radiation were not particularly important. Fortunately, other experimental work undertaken by members of Sparrow's group (e.g., Van't Hof, Price), though still focused on the strategic questions, was of much broader significance and led to seminal discoveries relating to fundamental aspects of cell development (Van't Hof and Sparrow, 1963) and overall patterns of genome size variation (Sparrow and Nauman, 1976).

Meanwhile, other laboratories began parallel studies on the effects of genome size on various additional characters of practical significance. This included a British group of cytogeneticists led by Hugh Rees at University College of Wales, Aberystwyth, whose primary interest was plant breeding. Such work focused on relationships between DNA amount and physical characters (length, volume, mass) at the chromosomal, nuclear, and cellular levels, especially in crops and their close relatives (Rees *et al.,* 1966).

Today, relationships are known between nuclear DNA amount and more than 40 widely different phenotypic characters at the nuclear, cellular, tissue, and organismal levels. Moreover, these extend to all types of plants, and are no longer limited to species of practical interest (e.g., crops or common experimental subjects).

Chromosome Size

Early work revealed highly significant positive correlations between nuclear DNA amount and chromosomal characters, such as total mitotic metaphase volume in samples of species in the genera *Lathyrus, Vicia, Lolium* (Rees *et al.,* 1966), and *Allium* (Jones and Rees, 1968). Such findings have since been confirmed in many other comparisons both within and across genera. For example, Figure 2.9A shows the relationship between DNA amount and total mitotic chromosome volume per somatic metaphase cell in 14 angiosperm species estimated from reconstructed cells using quantitative electron microscopy (Bennett *et al.,* 1983), and more recently chromosome area and length were shown to be positively correlated with DNA amount among 12 diploid rice species (Uozo *et al.,* 1997). A positive relationship was also found to apply to specialized chromosome regions; Figure 2.9B shows the tight relationship between DNA C-value and the total volume of centromeres per cell in 11 angiosperm species (Bennett *et al.,* 1981).

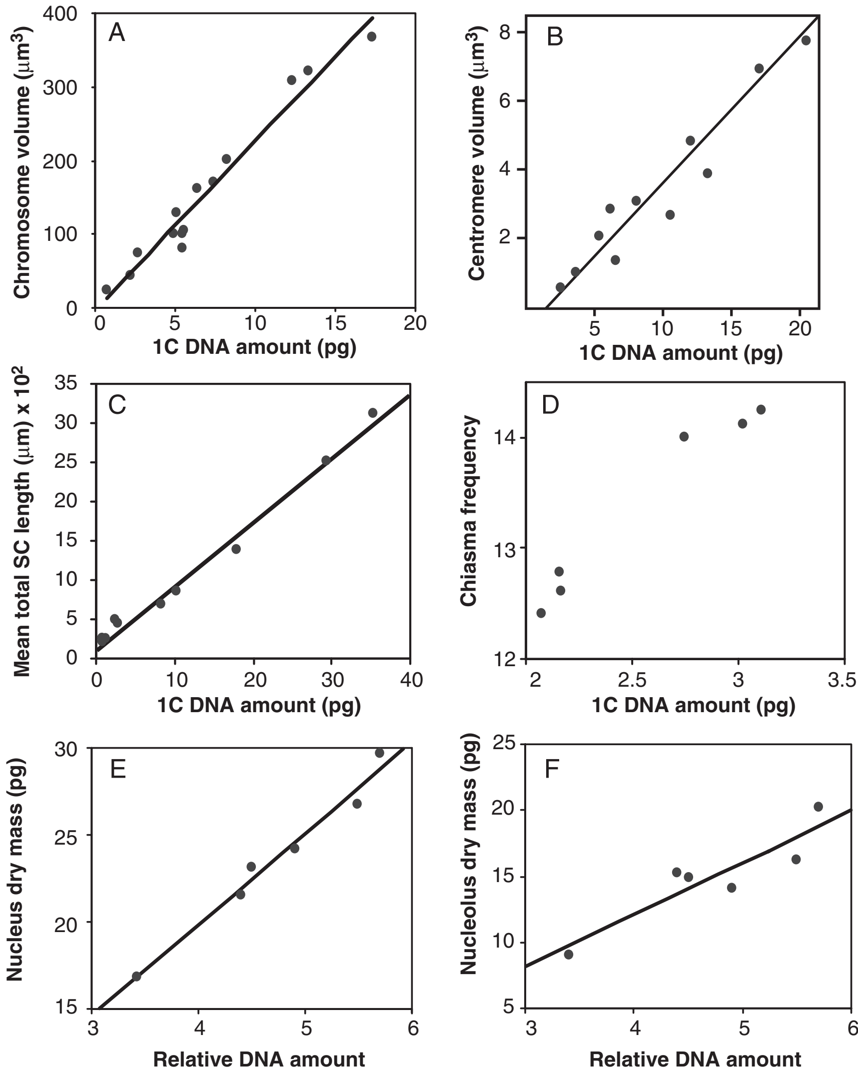

FIGURE 2.9 Relationship between DNA amount and six nuclear characters in angiosperms. (**A**) Total somatic chromosome volume (μm^3) at mitiotic metaphase in 14 species; (**B**) Total volume of centromeres per nucleus in 11 species; (**C**) Total length of the synaptonemal complex (SC) per cell at pachytene in 10 species; (**D**) Mean chiasma frequency in pollen mother cells of six *Lolium* species; (**E**) Nuclear dry mass and (**F**) nucleolar dry mass in six species of *Sorghum*. (**A**) Redrawn from Bennett *et al.* (1983), reproduced by permission (© Company of Biologists Ltd.); (**B**) From Bennett *et al.* (1981), reproduced by permission (© Company of Biologists Ltd.); (**C**) Modified from Anderson *et al.* (1985), reproduced by permission (© Elsevier Inc.); (**D**) From Rees and Narayan (1988), reproduced by permission of the authors; (**E**, **F**) Modified from Paroda and Rees (1971), reproduced by permission (© Springer-Verlag).

In addition to volume, total mitotic metaphase chromosome length has been shown to correlate positively with genome size (e.g., Rothfels *et al.,* 1966). However, these relationships are not typically as close as with volume because of the differential condensation of various chromosomal segments (especially euchromatin versus heterochromatin). During meiosis, when chromatin condensation is more relaxed, such differences may be minimized. Thus Anderson *et al.* (1985) showed a tight correlation between genome size and the total length of the haploid chromosome complement at meiosis (measured by tracing the lengths of synaptomental complexes in spread pachytene cells of 10 angiosperms with a roughly 30-fold range of DNA C-values) (see Fig. 2.9C). In addition, total chiasma frequency (i.e., recombination rate) per chromosome and per complete complement are also positively related to genome size (Rees and Narayan, 1988) (see Fig. 2.9D).

Nucleus Size

Examples of other size-related correlations at the nuclear level include the positive relationship observed between genome size and interphase nuclear volume mentioned in the previous section (e.g., Baetcke *et al.,* 1967). Many of the studies reporting positive relationships between genome size and total chromosome volume also noted similar relationships between genome size and total nuclear dry mass (measured using interference microscopy) in the same plant genera (Rees and Hazarika, 1969; Pegington and Rees, 1970; Paroda and Rees, 1971). For example, Figure 2.9E shows a close relationship between genome size and total nuclear dry mass in six *Sorghum* species. White and Rees (1987) also reported a positive relationship between nuclear dry mass and DNA content in six *Petunia* species. Similar positive relationships have been found with total nuclear histone content (Rasch and Woodard, 1959) and total nucleolar dry mass (Paroda and Rees, 1971) (see Fig. 2.9F). More recently, rRNA gene copy number has been shown to correlate positively with DNA amount in both plants and animals across a wide range of taxa (Prokopowich *et al.,* 2003).

Cell Size

As in the well-known example of vertebrate erythrocytes (see Chapter 1), the effects of bulk DNA content extend to characters at the cellular level in plants. For example, Martin (1966) noted a positive correlation between DNA amount and cell mass in root tip cells of 12 angiosperm species, and Holm-Hanson (1969) reported a relationship between DNA content per cell and the total

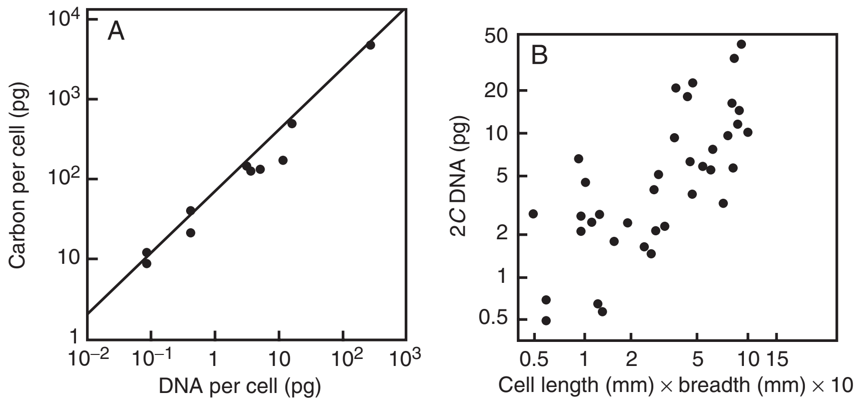

FIGURE 2.10 Relationship between DNA amount and cellular characters. **(A)** Total inorganic carbon per cell in 10 species of unicellular algae; **(B)** Mean length × breadth of epidermal cells in mature leaves of a range of herbaceous species. **(A)** Redrawn from Holm-Hanson (1969), reproduced by permission (© American Association for the Advancement of Science); **(B)** From a figure presented by Bennett (1987), based on data from Grime (1983), reproduced by permission (© Blackwell Publishing).

weight of carbon per cell in 10 unicellular algae (Fig. 2.10A). Grime (1983) showed a strong positive relationship between nuclear DNA content and the linear dimensions of leaf epidermal cells for 37 British herbaceous angiosperms (Fig. 2.10B).

Cell Division Rate

It has long been appreciated that nucleus size, cell size, and cell division rate are closely linked. As part of the radiosensitivity research program, Van't Hof and Sparrow (1963) noted that "for diploid plants a relationship does exist between the minimum mitotic cycle time, the interphase nuclear volume, and the DNA content per cell. Moreover, the relationship is such that if any one of the three variables is known, an estimate can be made for the remaining two." Figure 2.11A shows their data for a sample of root-tip meristem cells from six angiosperm species, all grown at 23°C, and reveals a strikingly close association between DNA content and mitotic cycle time. Additional experiments showed that a relationship could even be identified between DNA amount and the duration of the DNA synthesis phase (S-phase) in particular (Van't Hof, 1965). These results were first obtained using only small numbers of species, but larger subsequent

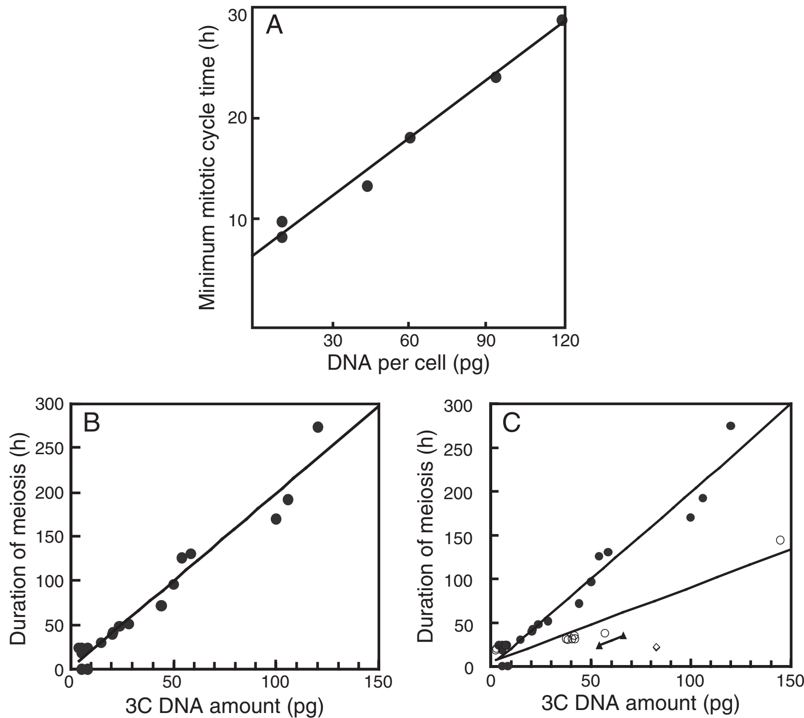

FIGURE 2.11 Relationships between DNA amount and duration of mitosis and meiosis. **(A)** The relationship between minimum duration of the mitotic cycle in root-tip cells of six angiosperms grown at 23°C; **(B)** Duration of meiosis in 18 diploid angiosperm species grown at 20°C; **(C)** Duration of meiosis in 18 diploid (2*x*) (•), ten tetraploid (4*x*) (○), three hexaploid (6*x*) (▲) and two octoploid (8*x*) (◇) angiosperms. **(A)** Data from Van't Hof and Sparrow (1963); **(B)** From Bennett (1977), reproduced by permission (© The Royal Society); **(C)** Data from Bennett (1977).

comparisons strongly confirmed the original conclusions (Van't Hof, 1974). It has also since been demonstrated that the durations of *all* phases of the cell cycle, not only of S-phase, are positively related to C-value (e.g., Evans and Rees, 1971; Evans *et al.*, 1972).

Reproductive cell development is both protracted (compared with other stages) and highly canalized, and hence is particularly useful for showing relationships between genome size and other characters. Thus Bennett (1971) found a strong positive correlation between genome size and the duration of meiosis in diploid angiosperms when he examined reproductive cell development in shoots. This was later confirmed for a larger sample of species (Bennett, 1977). By way of example, Figure 2.11B shows the remarkably precise relationship between DNA

C-value and the duration of meiosis in 18 unrelated diploid angiosperm species, all grown at 20°C.

Interestingly, comparisons for plants at different ploidal levels revealed close positive correlations between nuclear DNA C-value and the duration of meiosis, but also showed highly significant differences in the slope of the relationship according to ploidy (see Fig. 2.11C). Surprisingly, the duration of meiosis in plants grown at 20°C was reduced as ploidy increased. In wheat, for example, meiosis was found to take 42 hours in diploids, 30 hours in tetraploids, but only 24 hours in hexaploids, despite the proportionately higher DNA contents of the polyploids (Bennett and Smith, 1972). This shows conclusively that such relationships are not determined by DNA amount alone. Although this is undoubtedly one key factor, its effects can be modified by genetic factors such as genotypic differences and variation in gene dosage owing to polyploidy (Bennett *et al.,* 1974), as well as abiotic factors such as temperature (Bennett *et al.,* 1972). The duration of meiosis in plants can even differ between male and female meiocytes of the same species (Bennett, 1977).

CAUSATION AT THE CELLULAR LEVEL: THE NUCLEOTYPE CONCEPT

It is clear that cell size and division rate are not determined solely by C-value, given that both parameters can vary considerably within and among organisms even with constant nuclear DNA amount. Nevertheless, many relationships discussed in the previous section are strikingly close, a fact that is especially clear when several plots for widely different characters are viewed together, as in Figures 2.9 through 2.11. Indeed, some of these linkages are so strong as to be more reminiscent of physical or chemical relationships than biological ones (Bennett, 1987).

Taken together, these considerations suggest that the relationship between DNA content and chromosomal, nuclear, and cellular parameters is causal in nature, albeit with other factors involved. This is quite evident at the chromosomal level, in which the amount of DNA necessarily impacts upon total chromosome volume and mass, but in which proteins (e.g., histones) also play a role. Figure 2.1, which shows the chromosome complements from species differing ~220-fold in DNA amount all taken at the same magnification, illustrates this point. Similarly, as the nucleus is assembled around a scaffolding of DNA, C-value must causally influence nuclear volume (Cavalier-Smith, 1985).

For biophysical reasons, it is impossible to increase greatly the C-value without also increasing the minimum time needed for cell division. This is because more DNA not only takes longer to replicate (i.e., prolongs S-phase), it also impacts on all stages of the cell cycle (see Gregory, 2001a, and Chapter 1). According to the model developed by Gregory (2001a) this delay in cell division is ultimately

responsible for generating the positive correlation between DNA content and cell size (see Chapter 1). At the very least, it is clear that one cannot increase DNA content beyond a certain point without also increasing nucleus and cell size. For example, a comparison between *Fritillaria* sp. (1C ≈ 65 pg) and *Myriophyllum spicatum* (1C = 0.3 pg) shows the impossibility of containing the nucleus, chromosomes, and even the DNA of the former in the small cells of the latter (see Fig. 2.1B,C).

The realization that the relationships discussed in the previous sections were influenced causally by the amount of DNA led to the development of the "nucleotype" concept (Bennett, 1971). Specifically, the nucleotype describes those conditions of the nuclear DNA, most notably its amount, that affect the phenotype independently of its encoded informational content. Whereas the combined set of genes defines the "genotype," the nucleotype consists of all the DNA, both genic and nongenic (Bennett, 1971). The nucleotype can be considered as setting the minimum conditions, or perhaps as exerting a very coarse control, of parameters at the cell level, whereas the genotype is responsible for fine control of these features within these limits (Bennett, 1972; Karp *et al.,* 1982; Gregory, 2001a). For example, while nuclear volume is subject to variation by genetic control during development despite a constant DNA amount (Bennett, 1970), such control can operate only at or above the minimum volume determined by the DNA C-value. The same is true of parameters at the cell level and above.

Pollen and Seeds

There is now considerable evidence that these nucleotypic effects at the cellular level are additive and extend to higher level features of direct relevance to fitness. For example, size relationships are known for reproductive structures such as the male gametophyte (pollen grains) and early sporophyte stages (seeds) in angiosperms. As examples, Figure 2.12A shows the positive relationship between DNA amount and mean pollen grain volume in wind pollinated grasses (Bennett, 1972), and Figure 2.12B shows the correlation between DNA content and minimum seed weight in 24 British legumes (Mowforth, 1985). Similar relationships with pollen size have been shown for species of *Ranunculus, Vicia* (Bennett, 1973), and *Petunia* (White and Rees, 1987). Positive relationships between DNA amount and seed weight have likewise been reported for comparisons within populations of the same species (e.g., Caceres *et al.,* 1998), within genera, including *Vicia* and *Allium* (Fig. 2.12C) (Bennett, 1972, 1973; see also Knight and Ackerly, 2002), and across large numbers of species (e.g., Thompson, 1990; Knight *et al.,* 2005). Interestingly, the use of quantile regression analysis by Knight *et al.* (2005) suggests that the relationship between seed weight and genome size may not be linear. Thus it appears that whereas species with small genomes

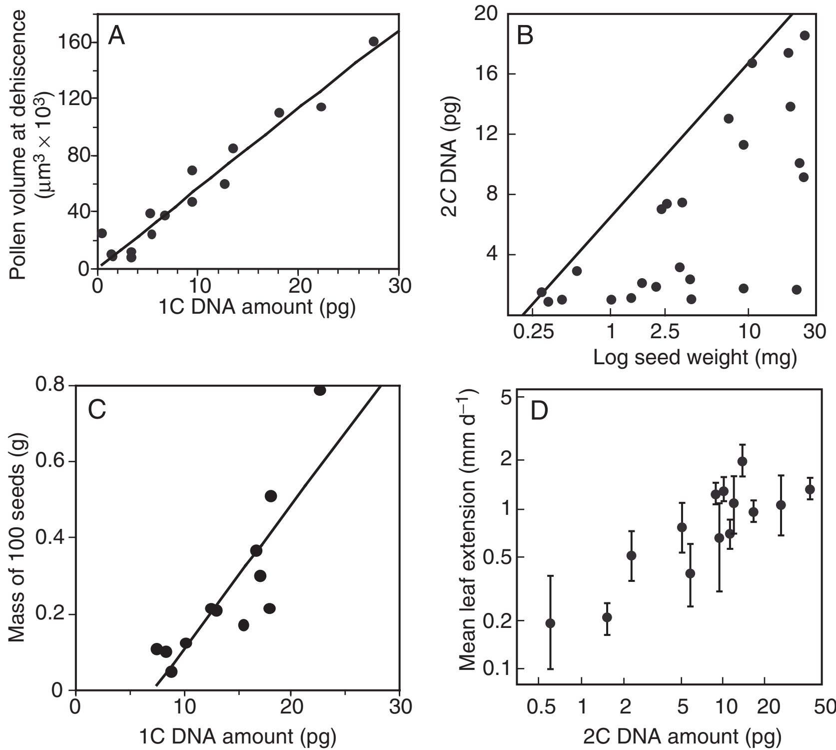

FIGURE 2.12 Relationships between DNA amount and reproductive and growth parameters. **(A)** Pollen volume in 16 grass and cereal species; **(B)** Seed weight in 24 British legume species; **(C)** Seed dry mass in 12 *Allium* species; **(D)** Mean range of leaf extension over the period March 25 to April 5 in 14 grassland species coexisting in the same turf. **(A)** Redrawn from Bennett (1972), reproduced by permission (© The Royal Society); **(B)** From Bennett (1987), originally based on Mowforth (1985), reproduced by permission (© Blackwell Publishing); **(C)** Redrawn from Bennett (1972), reproduced by permission (© The Royal Society); **(D)** From Bennett (1987), based on a figure first presented by Grime (1983), reproduced by permission (© Blackwell Publishing).

display a wide range of seed weights, those with large genomes rarely produce small seeds.

Of course, these relationships are complex, and again it should not be imagined that genome size alone determines parameters such as seed weight or pollen size. Nevertheless, genome size may have a major effect in determining the minimum values possible for these features—an important consideration, given that seed weight is believed to affect many other ecologically relevant attributes such as invasiveness and survivability (Westoby *et al.,* 1992; Rejmanek, 1996).

Minimum Generation Time and Developmental Lifestyle

Development at the organism level consists of division and growth at the cell level, suggesting that it can be strongly influenced by variation in genome size. However, because genome size is correlated negatively with division rate, but positively with cell size, the nature of the relationships may vary according to which cellular parameter dominates in a particular developmental process. Thus, although it is generally the case that a large genome size limits the rate at which plants can develop, DNA content may correlate positively with growth rate under certain conditions (Grime and Mowforth, 1982; Knight *et al.,* 2005).

In 1972, as part of the early work on the nucleotype concept, Bennett asked whether DNA content might place a lower limit on the duration of the period from germination until the production of the first mature seed—that is, minimum generation time (MGT). To test this, Bennett (1972) compared the mean and the ranges of nuclear DNA contents for 271 angiosperm species with different life cycle types and different ranges of MGT. For this comparison the species were divided into one of four types of life cycles: (1) ephemerals, which can complete their life cycle in a very short period of time (arbitrarily defined as seven weeks or less), (2) annuals, which by definition complete their life cycle within 52 weeks, (3) facultative perennials, which can potentially set fertile seed within 52 weeks of germination, and (4) obligate perennials, which require more than 52 weeks to produce mature seed. The following intriguing results were obtained:

1. The mean 1C nuclear DNA content for ephemeral species (1.5 pg) was less than for annuals (7.0 pg), which in turn was less than for perennial species (24.6 pg). This was true irrespective of whether the comparison included all species or only diploids.
2. The maximum 1C DNA content was lower for ephemerals (3.4 pg) than for annual species (27.6 pg), which was much lower than the maximum for perennials (127.4 pg).
3. The mean and range of DNA amounts for facultative perennials and annuals (both of which have the same maximum MGT of 52 weeks) were very similar and both were much less than for obligate perennials.
4. Species with very low DNA amounts (i.e., ≤3.4 pg) had life cycle types ranging from ephemeral to long lived perennials.
5. With increasing nuclear DNA content, the MGT increased and the range of life cycle types decreased, such that above 3.4 pg no ephemeral species were found, and above 27.6 pg no annual species or facultative perennials were found. Consequently, all species with 1C values greater than 27.6 pg were obligate perennials.

Taken together, the results clearly suggested a positive relationship between nuclear DNA content and MGT. Based on this, Bennett (1973) developed a model to show the linear relationship between C-value and MGT in which threshold effects play an important role (Fig. 2.13A). As C-value increases, so does MGT, and at various points this results in an inescapable shift in developmental lifestyle. To be an ephemeral, a species would have to have a C-value of less than the limiting value for a MGT of seven weeks. At the opposite extreme, species with C-values greater than the threshold value for a MGT of 52 weeks will be obligate perennials.

Of course, many limiting factors (such as availability of phosphorus, essential for DNA synthesis) can act to slow or delay a plant's generation time. Such effects can occur at many points in the development of multicellular organisms, and therefore can easily confound a relationship between genome size and generation time. However, it is important to note that their effect is only to delay and hence is unidirectional in always increasing generation time above the minimum. Thus the model (see Fig. 2.13A) shows a plot with points distributed either along a linear relationship between genome size and minimum generation time (line A-B), or scattered in the triangle for generation times longer than the minimum (triangle ABC). In this sense, species with a very small genome may include both ephemerals (e.g., *Arabidopsis thaliana,* 1C = 0.16 pg), where the point is expected to fall close to the line A-B, and trees (e.g., birch *Betula populifera,* 1C = 0.2 pg), where the point will fall well to the right of the line A-B. Taxa with very large genomes all have long MGTs, and so are obligate perennials and never ephemerals or annuals. Nevertheless, they still show variation due to delays in development, with some taxa falling on or close to the line A-B (with generation times of just over one year), while others require several or many years to complete a generation and are shifted to the right of the line.

The concept of thresholds is clearly illustrated by the absence of any points in the outer triangle (i.e., to the left of the line A-B; compare this with Fig. 2.12B for seed weight versus DNA amount). However, it should be noted that although it is easy to find generation time data for plant species in general, records of true *minimum* generation times are difficult to obtain. Figure 2.13B shows such "record" minimum generation times for 10 herbaceous angiosperm species (ranging from 31 days in *Arabidopsis thaliana* to at least eight months in *Lilium longiflorum*) plotted against 1C DNA amount. Clearly, the shortest minimum generation time increases with increasing C-value, and the plot suggests that the MGT may be slightly shorter in polyploids (plotted as open circles) than in diploids with the same DNA amount. Once again, the important point is that DNA content sets limits on cellular and organismal development, but does not determine these features by itself. Species with low DNA amounts (which exhibit rapid mitosis and can complete meiosis within one to two days) have the option to express a wide range of life cycle types, from ephemeral to perennial—subject to further genic control. By contrast, species with very high C-values spend so long growing and

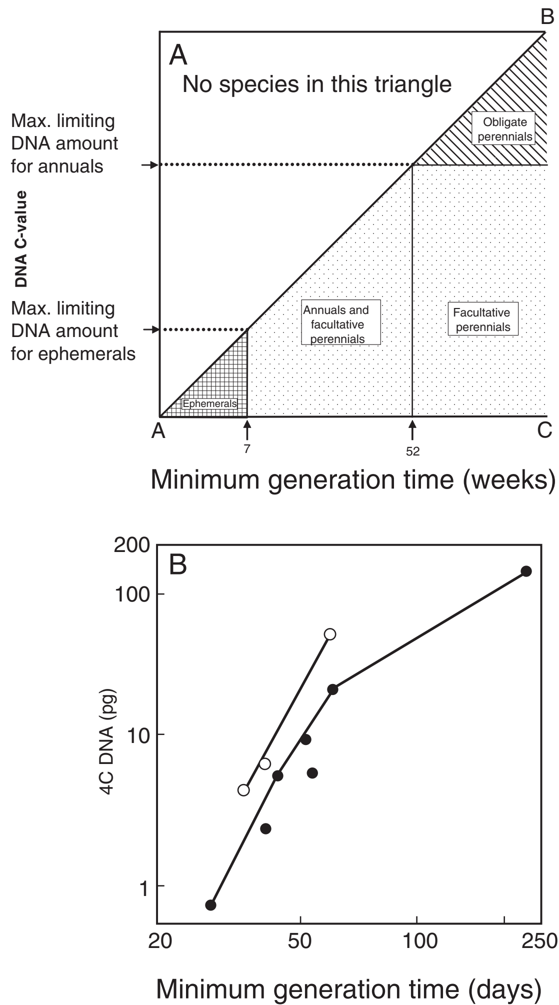

FIGURE 2.13 **(A)** Diagrammatic illustration of a model for a simple relationship between DNA amount and minimum generation time (MGT) in a temperate environment. Points for the shortest MGT for a given DNA amount lie on the line A-B. All points lie on or to the right of line A-B with none in upper triangle. Species with MGTs of seven or fewer weeks lie in the triangle with cross hatching and hence all have low DNA amounts. Species whose DNA amounts exceed the maximum limiting DNA amount for annuals have MGTs of 52 weeks or more and are therefore nucleotypically determined obligate perennials (striped triangle). **(B)** Relationship between 4C DNA amount and the shortest MGT known for three polyploid (○) and seven diploid (●) angiosperms. **(A)** Redrawn based on Bennett (1987), reproduced by permission (© Blackwell Publishing); **(B)** From Bennett (1987), reproduced by permission (© Blackwell Publishing).

completing meiosis and sporogenesis that they are simply precluded from expressing an ephemeral (or possibly even an annual) habit, regardless of genotype.

For this reason, there tends to be a change in genome size when plants shift developmental lifestyles from perennial to annual (though not necessarily vice versa). Specifically, such changes from perennial to annual lifestyle appear to occur often when plants move into harsh new environments with short growing seasons, with this derived developmental condition associated with a smaller genome. Such has been the case in species from numerous genera, including *Arachis, Brachyscome, Calotis, Crepis, Haplopappus, Helianthus, Lathyrus, Papaver,* and *Vicia* (Resslar *et al.,* 1981; Sims and Price, 1985; Srivastava and Lavania, 1991; Singh *et al.,* 1996; Naranjo *et al.,* 1998; Watanabe *et al.,* 1999). In one particularly informative example, Watanabe *et al*. (1999) examined the shift from perennial to annual lifestyle in *Brachyscome* using a phylogenetic framework, and were able to provide a clear demonstration that a reduction in genome size was part of the shift in developmental program, and therefore in the adaptation to a new environment.

There is another way of classifying the developmental lifestyles of plants that is of more practical significance to humans—namely, weeds versus nonweeds. Weeds represent a taxonomically eclectic group of plants defined by their annoying habit of growing with great success in places where they are not wanted. Key factors suggested to be important for the success of many weeds include rapid establishment and completion of reproductive development, short minimum generation times, and fast production of many small seeds. All of these factors correlate with low DNA amount, thereby raising the possibility that life as a weed imposes significant constraints on genome size.

Indeed, this is just what is found. In a detailed analysis of the DNA amounts of 156 angiosperm weed species (including 97 recognized as important world weeds) versus 2685 other angiosperms, Bennett *et al*. (1998) provided strong evidence that small genome size is a requirement for "weediness." Specifically, the weed sample had a significantly smaller mean C-value than the nonweed sample, with DNA amounts in weed species restricted to the bottom 20% of the range known for angiosperms. Moreover, a highly significant negative relationship was found between DNA amount and the proportion of species recognized as successful weeds (Fig. 2.14A), demonstrating that the probability of being a recognized weed decreased with increasing DNA amount, reaching zero at a cutoff value of just above 25 pg. In other words, while many angiosperm species have DNA amounts greater than 25 pg, none of them is (or, most likely, could ever become) a weed.

Other unique requirements of developmental lifestyle may impose limits on genome size. An interesting example is provided by the orchid *Erycina pusilla,* which grows as an epiphyte on leaves of tropical trees in South America. In order to survive, it must complete its life cycle before the leaf of the tree falls. This small plant can reach reproductive maturation in only four months, compared with the one to five years commonly taken by other orchid species. Notably, it has

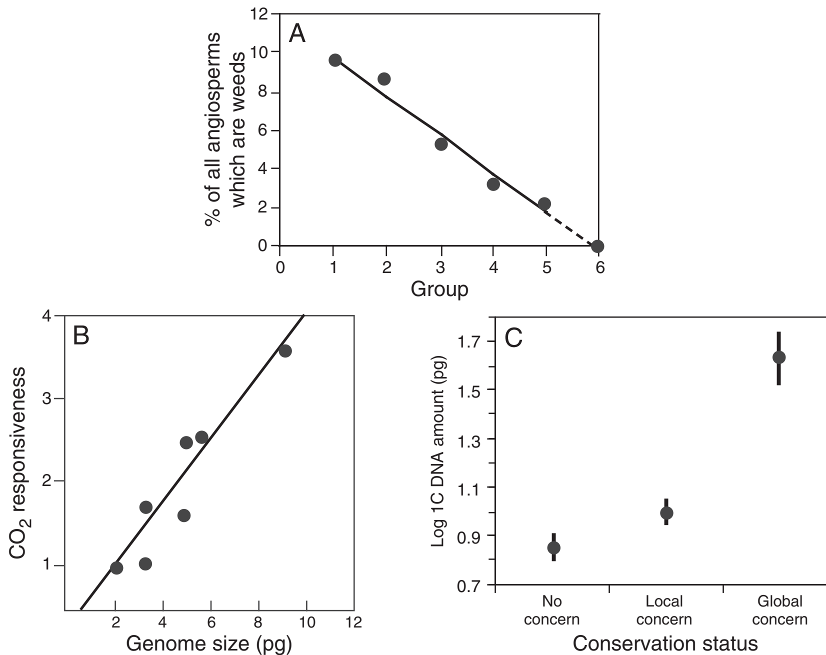

FIGURE 2.14 **(A)** Relationship between DNA amount and the probability of being a weed. The analysis of the percentage of all angiosperms that are weeds was based on data for 2841 species of known DNA amount (comprising 156 weeds and 2685 other species) ranked in order of increasing size, including 2719 species with 1C values of ≤25.0 pg (the maximum for weeds) divided into five groups with 544 (groups 1–4) and 543 (group 5) and 122 species ≥25.0 pg (group 6). **(B)** Relationship between the responsiveness (= ratio of average final biomass attained at elevated and ambient CO_2) to elevated CO_2 and 1C nuclear DNA amount among seven annual species of grasses. **(C)** Relationship between DNA amount and conservation status of 3036 species of angiosperms. Species with known DNA amounts were checked against the United Nations Environment Programme World Conservation and Monitoring Centre Species Database. From the total of 3036 species with known DNA amounts, 305 species were found to be of global concern and 1329 species were of local concern (i.e., threatened only in particular countries). **(A)** From Bennett *et al.* (1998), reproduced by permission (© Oxford University Press); **(B)** From Jasienski and Bazzaz (1995), reproduced by permission (© Nature Publishing Group); **(C)** From Vinogradov (2003), reproduced by permission (© Elsevier Inc.).

a small genome for an orchid (1C = 1.5 pg) compared with the mean for the family (1C = 7.7 pg).

Physiology and Climate Response

Physiological correlates of genome size are well known for certain animals, most notably mass-specific oxygen consumption rate in mammals and birds

(see Chapter 1). These are based on the effects of genome size on cell size, which also apply to plants. However, other than comparisons between diploids and polyploids, the corresponding implications for plant physiology had remained unaddressed until very recently (see Knight *et al.*, 2005). Importantly, it now appears that mass-specific maximum photosynthetic rate is inversely correlated with genome size (Knight *et al.*, 2005), in parallel with the situation regarding metabolic rate in endothermic animals. This is in keeping with the finding of a significant negative correlation between genome size and specific leaf area (leaf mass ÷ leaf area), which in turn is associated with a range of physiological and ecological traits (Knight *et al.*, 2005).

The more general issue of how plants respond to climate involves both physiological and developmental components, and has therefore been addressed in several different ways. Some of the earliest studies in this area were conducted by Grime and his colleagues. First, Grime and Mowforth (1982) compared genome size with time of shoot expansion and temperature using 24 herbaceous species growing in the Sheffield region of the United Kingdom. They found that the species growing most actively at a given time tended to have smaller and smaller genomes as the seasons progressed from early spring to midsummer. Thus species with large genomes tended to grow early in the spring, and those with smaller genomes started growth later in the year. They suggested that this pattern was related to the fact that cell division is temperature-dependent and cell expansion is not. In this case, species with large genomes grew early in the spring predominantly through the expansion of cells that had divided in the preceding year, whereas small-genomed species growing in the summer did so by normal cell division.

The idea that large DNA contents facilitate growth under moderately cold conditions was reinforced by subsequent studies. In particular, a positive correlation was found between C-value and the mean rate of leaf extension for 14 major plant species in a damp limestone grassland during the cold conditions of an early British spring (Grime, 1983; Grime *et al.*, 1985) (see Fig. 2.12D). More recently, MacGillivray and Grime (1995) also showed larger-genomed plants to be more tolerant to frost.

ECOLOGICAL AND EVOLUTIONARY IMPLICATIONS OF GENOME SIZE VARIATION

Through its effects on cell size and division rate, genome size can influence a wide range of phenotypic characters, including seed and pollen size, developmental lifestyle (e.g., annual versus perennial, weed versus nonweed), physiology, and climate tolerance—in short, variation in genome size can play a major role in determining when, where, and how plants grow. As discussed in the following sections, this is important for both the natural distribution of plant species and their responses to changes in the environment.

Geographical Distribution and the Large Genome Constraint Hypothesis

The first clues that DNA content might have some relationship with geographical distribution long predated the availability of genome size data. In 1931, Avdulov noted that tropical grasses had uniformly small- to medium-sized chromosomes, whereas most grasses of cool temperate regions had large chromosomes. This is in keeping with the positive correlations between genome size and both chromosome size and capacity for growth under cooler conditions.

The expectation that species with large genome sizes might be concentrated in temperate regions was confirmed in two separate studies carried out more than 45 years after Avdulov's (1931) initial report. Levin and Funderberg (1979) analyzed a broad range of herbaceous angiosperms and showed that the mean 1C-value for temperate species (6.8 pg) was more than double that for tropical species (3.0 pg). However, this difference stemmed from the greater range of DNA amounts and the higher frequency of species with large genomes in temperate versus tropical floras, and not from any exclusion of species with small genomes from temperate regions.

In an analysis of herbaceous cultivated pasture grasses, cereal grain crops, and pulses, Bennett (1976b) showed that the cultivation of species with high DNA amounts per diploid genome tended to be localized at temperate latitudes, or to seasons and regions at lower latitudes where the conditions approximate those normally found in temperate areas. Overall, it seems that a natural positive cline of DNA amount with latitude has been reinforced and exaggerated by the tendency of humans to choose species for cultivation with increasing DNA amounts at successively higher latitudes (Bennett, 1976b, 1987). Importantly, this cline is exhibited by crop species with both C_3 and C_4 photosynthesis and by both annuals and perennials (Bennett, 1976b). Obviously, the choice of species for domestication was done independently of knowledge about DNA content. Rather, the prime factor suiting certain species to human requirements would have been their high yield of seed or leaf per unit area and per unit time. That this indirectly involved choosing species with higher DNA amounts at higher latitudes is an intriguing demonstration of the importance of the nucleotype.

However, as with most of the relationships discussed in this chapter, this pattern involves some important additional complexities. For example, while this DNA–latitude cline for crops is particularly clear in comparisons of the northerly limits of cultivation of cereals in the northern hemisphere in the winter, it may take a very different form in the summer (Fig. 2.15) (Bennett, 1987). Specifically, species with high DNA amounts, which show a pronounced DNA amount–latitude cline in winter, tend to be bunched at the northern end of their range in summer. Indeed, the cline appears to reverse its polarity in summer at the very high latitudes; for example, barley (1C = 5.5 pg) is grown nearer the pole in summer than rye

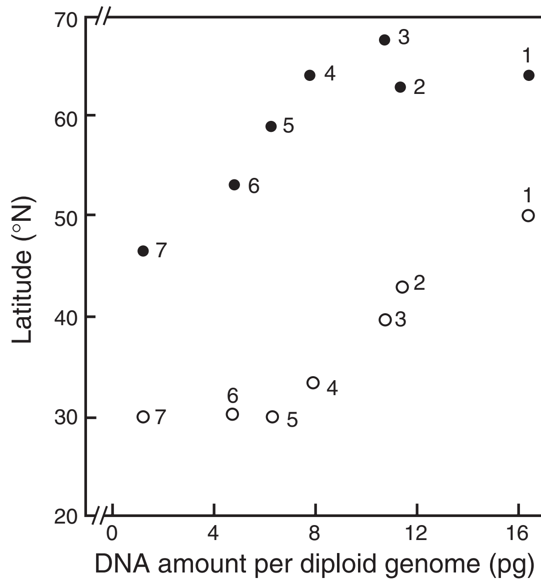

FIGURE 2.15 The relationship between DNA amount per diploid genome and the northern limits of cultivation of several cereal grain species. Key to points: ○ for a transect from Hudson Bay to Key West in Florida (approx. 82°W) in winter; ● for a transect from near Murmansk by the Arctic Ocean to Odessa by the Black Sea (approx. 32°E) in summer. 1. *Secale cereale;* 2. *Triticum aestivum;* 3. *Hordeum vulgare;* 4. *Avena sativa;* 5. *Zea mays;* 6. *Sorghum* spp.; 7. *Oryza sativa.* From Bennett (1987), reproduced by permission (© Blackwell Publishing).

(1C = 8.3 pg), and, similarly, peas (1C = 4.8 pg) are grown closer to the pole in summer than field beans (13.4 pg)—the reverse of their northern limits during the winter.

This reversal at high latitudes in summer agrees with the general trend for DNA C-value and latitude noted for noncrop species at high latitudes. Thus, for angiosperms from South Georgia and the Antarctic Peninsula, DNA amounts per nucleus (of whatever ploidy) and per diploid genome are within the ranges known for angiosperms at temperate latitudes but, significantly, at their lower end (Bennett *et al.,* 1982b). Bennett *et al*. (1982b) concluded that, above a certain high latitude, maximum DNA content per nucleus and per diploid genome decreases with increasing latitude (and with decreasing temperature). It was therefore suggested that at high latitudes, selection against species with high DNA amounts strongly increases toward the poles as the climate becomes progressively harsher and growing seasons shorter. In particular, this probably relates to limitations in establishment ability among large-genomed plants (Bennett *et al.,* 1982a). Further support for this hypothesis comes from Grime and Mowforth's (1982) survey of British angiosperms, which showed a significantly higher mean genome

size for 69 southern species (1C = 6.7 pg) than for 80 widespread species (3.98 pg), with even lower 1C DNA amounts noted in eight northern species (1.13 pg).

Overall, studies of genome size versus latitude, altitude, temperature, and precipitation have tended to show both positive and negative correlations, depending on the particular species and environmental parameters analyzed (reviewed in Knight *et al.,* 2005). In sum, such findings are actually not contradictory, and have recently led to an interesting "large genome constraint hypothesis" (Knight *et al.,* 2005) in which species with low DNA amounts are widespread, whereas those with the highest C-values are progressively excluded from increasingly harsh environments above a midlatitude. Results for both crop and noncrop species fit this view, but the critical latitude above which this exclusion operates is probably higher for crops, owing to the additional input of human selection. Again, within the more intermediate range, large genomes may be favored in the cold where cell expansion is a more efficient means of growth than cell division. Outside this range, however, plants with large genomes appear to be at a disadvantage. This seems to apply to extremes of both temperature and precipitation (Knight and Ackerly, 2002).

Genome Size and Plant Response to Human Environmental Change

The power of human activity to alter the global climate and the species inhabiting it should not be underestimated. Anthropogenic effects on the environment pose a major threat to a great many species, including plants. Given the role that genome size plays in setting limits to development, physiology, and geographical distribution, it is important to consider its potential influence in terms of plant responses to continued human activity.

Global Warming

To predict the effects of human production of greenhouse gasses and the resulting increase in global temperature on plants, Grime (1996) looked at how plants responded to elevated temperatures and showed that plants with small genomes exhibited a greater enhancement of growth than those with larger genomes. This led to the suggestion that global warming would stimulate the expansion of species with small genome sizes in certain floras (Grime, 1996), although this could potentially be checked by occasional frosts favoring species with large genomes (MacGillivray and Grime, 1995). On the other hand, Jasienski and Bazzaz's (1995) study of the responses of annual grasses to elevated CO_2 found that growth rates were enhanced in a way *positively* correlated with genome size (see Fig. 2.14B). So, although the most likely net outcome remains somewhat

ambiguous, it is clear that genome size variation can be expected to influence the response of plant communities to global warming in some fashion.

Radiation

As noted previously, sensitivity to radiation was one of the first correlates of genome size identified (see Fig. 2.8). Although such experiments obviously employed rather severe doses, the general finding that plants with larger genomes are more susceptible to radiation could have real-world applications. For one thing, nuclear radiation continues to be produced and released by various means (e.g., through the use of nuclear power, and the production of nuclear weapons). Moreover, it could be that large-genomed plants are more sensitive to additional mutagens such as ultraviolet light. If so, then the depletion of the ozone layer might be expected to have especially adverse effects on plants with large C-values.

Chemical Pollution

In a recent study, Vilhar and her colleagues (personal communication and Vidic *et al.*, 2003) investigated the distribution of plants along a gradient of heavy metal pollution produced by a former lead smelter at Zerjav, Slovenia. In this case, there was a negative correlation between the concentration of contaminating metals in the soil and the proportion of species with large genomes. Their results provided the first direct evidence that plants with large genomes are at a selective disadvantage under the extreme environmental conditions caused by pollution.

Generalized Extinction Risk

In an important recent analysis, Vinogradov (2003) showed that large-genomed diploid species are significantly more likely to be listed as rare or in danger of extinction than ones with small genomes (see Fig. 2.14C). This relationship was shown to hold when the analyses were also carried out within families, overcoming to some extent complications arising from phylogenetic issues. Further analysis showed that the effect was independent of ploidy, thereby suggesting that the process of polyploidy itself is not associated with increased risk of extinction. Vinogradov also showed that the maladaptive nature of larger genomes was (at least partially) independent of the duration of the plant life cycle (i.e., whether it was annual or perennial). Vinogradov (2003) and Knight *et al*. (2005) both reported a negative relationship between genome size and species richness in different plant groups, suggesting that plants with large genomes may indeed be more prone to extinction, and/or that such groups speciate more slowly and would have more difficulty recovering from a loss in biodiversity.

INTRASPECIFIC VARIATION IN GENOME SIZE

Until now, most of the discussion in this chapter has focused on variation in genome size among species and larger clades. Even when dealing with the processes responsible for generating these patterns, the focus has been on mechanisms at the genomic and interspecific levels, without much consideration of variation within species. In part, this is because intraspecific variation remains one of the most controversial topics in the study of plant genome size evolution, and because it is not entirely clear how often such variation even occurs. The following sections provide a review of the controversy in this area, in addition to highlighting some of the known cases in which intraspecific variation has important consequences for phenotypic variation among conspecifics.

Overview of Intraspecific Variation

Again, the earliest work on genome size in both plants and animals established the concept of DNA constancy for a species, given a constant basic chromosome number and type (i.e., with no variation due to aneuploidy, sex chromosomes, or supernumerary chromosomes). For plants, this was supported by several studies in crop species and their wild relatives. For example, whereas Bennett was routinely able to distinguish between lines of wheat that differed by less than 1% in DNA amount (M.D. Bennett, unpublished observation), he failed to find any intraspecific variation in lines of *Vicia faba* with very different ("major" and "minor") seed sizes (Bennett and Smith, 1976) or in land races of *Hordeum vulgare* from widely different geographical regions (Finland, England, Ethiopia, and Iraq) (Bennett and Smith, 1971). Furthermore, several studies reporting intraspecific variation proved to be unrepeatable. For example, although Zakirowa and Vakhtina (1974) reported intravarietal variation of up to 77% in *Allium cepa,* Bennett and Smith (1976) noted clear problems with the techniques used in their study.

During these early stages, Evans *et al.* (1966) did show heritable differences in DNA content of up to 16% induced by environmental factors in flax (*Linum usitatissimum*), and Furuta *et al*. (1975) reported DNA variation of greater than 20% in *Aegilops squarrosa,* but these were considered exceptions to the generally accepted view of DNA constancy in angiosperms.

By the 1980s this view had changed as reports of detectable, and in some cases considerable, intraspecific variation became increasingly common. For example, in 1981 genome size variations of 25% and 27% were reported in two *Microseris* species (Price *et al.,* 1981a,b). This variation, independent of any observable differences in heterochromatin, was reported to be possibly related to environmental factors. In 1985 Bennett reviewed the field, listing 24 angiosperm species

in which intraspecific variation was reported, from as low as 4% to 40–50% in *Glycine max,* 80% in *Poa annua,* and a staggering 288% in *Collinsia verna* (Bennett, 1985). In light of such findings, the original concept of the *very* constant genome was replaced by a new view of the *fairly* constant genome.

The 1990s saw the appearance of additional reports (implicit or explicit) of intraspecific variation, including 46% in *Narcissus hedraeanthus* (González-Aguilera *et al.,* 1990), 47% in *Festuca arundinacea* (Ceccarelli *et al.,* 1992), and 287% in *Helianthus annuus* (Johnston *et al.,* 1996; Price and Johnston, 1996). This produced a further shift in perception as the concept of the fairly stable genome was increasingly replaced by the idea of outright genome plasticity (Bassi, 1990). Such studies were often accompanied by molecular data showing that genome size changes could be associated with variability in particular repetitive DNA sequences (e.g., Frediani *et al.,* 1994; Cavallini *et al.,* 1996).

Not everyone was convinced that plant genomes were inherently plastic, however, and some authors began to question whether much of this supposed "intraspecific" variation might be the result of technical artifacts and taxonomic errors. Greilhuber (1998), in particular, sought to distinguish between "orthodox" intraspecific variation owing to genuine chromosomal variation involving duplications and deletions, spontaneous aneuploidy and polyploidy, heterochromatic segments, B chromosomes (see Chapter 4), and in special cases sex chromosomes (see later section) and "unorthodox" intraspecific variation, where genome size variation could not be explained in terms of "orthodox" events but instead was explained by unobserved postulated events such as rapid amplification or deletion of repetitive DNA sequences.

Such concerns stimulated a series of reinvestigations, led by Greilhuber and his colleagues in Vienna, to carefully repeat some of the claims of intraspecific variation using the original material. These studies have resulted in many of the claims for intraspecific variation being rebutted or greatly reduced. For example, several careful studies (Greilhuber and Ebert, 1994; Baranyi and Greilhuber, 1995, 1996) found little or no significant genome size variation in *Pisum sativum,* a species where others (Cavallini and Natali, 1991; Cavallini *et al.,* 1993) had previously reported up to 129% variation. Claims of 15% variation in *Glycine max* related to maturity group (Graham *et al.,* 1994) were unrepeatable according to Greilhuber and Obermayer (1997). A report of 130% variation in *Dactylis glomerata* negatively correlated with altitude (Reeves *et al.,* 1998) was also challenged by Greilhuber and Baranyi (1999), who could not repeat the different hydrolysis curves (see later section) claimed for *Dactylis* compared with the calibration standard *Hordeum vulgare*.

Other reported cases of intraspecific variation were shown to be taxonomic artifacts resulting from insufficient consideration of actual species boundaries and from an incomplete or nonexistent knowledge of the phytogeographical history of the populations in question. For example, *Scilla bifolia* L. *sensu lato* is treated as

one species by *Flora Europaea* (McNeill, 1980) but was split into 18 species (of which five are polyploid) by Speta (1980). So, whereas *Scilla bifolia* treated as one species shows a 2-fold genome size variation, its level of "intraspecific" variation diminishes to hardly more than methodological error when split up taxonomically (Greilhuber and Speta, 1985). Further examples disproving claims of intraspecific variation are given in the critical reviews of Greilhuber (1998, 2005). Unfortunately, these do not include a direct reassessment of the largest claim for intraspecific variation of all (288% for *Colinsia verna*) (Greenlee *et al.,* 1984) because no original research material is available. Nevertheless, this case is considered suspect on the grounds of circumstantial evidence (Greilhuber, 1998).

Recent findings in *Helianthus annuus* (sunflower) provide a particularly relevant lesson concerning intraspecific variation. In 1996, Price and Johnston made the surprising claim that the genome size in *H. annuus* varied 2.8-fold depending on light quantity and quality. They concluded that the major factor responsible for inducing a change in genome size was the ratio of red to far-red light, and suggested that phytochromes might be involved in the stability of genome size in sunflowers. However, in a subsequent reevaluation, Price *et al.* (2000) reported that their previous results were not biologically real, but rather a technical artifact caused by staining inhibitors (whose levels were presumably influenced by the quantity and quality of light) and their lack of an internal standard. Similar staining inhibitors caused by compounds in plant tissue have also been noted in other plant species including coffee, yams, roses, oaks (*Quercus* species), and *Allium* species (Ricroch and Brown, 1997; Zoldos *et al.,* 1998; Noirot *et al.,* 2000, 2002, 2003, 2005; Yokoya *et al.,* 2000).

As a result of the careful recent studies, the current view is that many, and perhaps most, previous examples of "unorthodox" intraspecific genome size variation (*sensu* Greilhuber, 1998) must be interpreted with caution until confirmed by independent studies using modern best practice techniques. To be fully accepted, claims of extensive intraspecific variation should also be accompanied by appropriate, wide-ranging molecular studies of the entire repeated sequence profile of the individual materials claimed to show such variation. At present, cases of orthodox variation and confirmed unorthodox variation are relatively few, although some intriguing examples do exist.

Genuine Intraspecific Variation in Angiosperms

Orthodox intraspecific variation associated with observable chromosomal phenomena has been noted in several cases, as with duplications, aneuploidy, and B chromosomes in *Zea mays* (e.g., Poggio *et al.,* 1998). Although much rarer in plants than in animals, the presence of dimorphic sex chromosomes can also

generate variation between individual plants, as has been found in *Silene latifolia* (e.g., Costich *et al.*, 1991).

Events visible under the light microscope can also generate detectable gains or losses of repeated sequence DNA in heterochromatin. For example, Gerstel and Burns (1966) noted the occurrence of megachromosomes in F_1 hybrids of *Nicotiana tabacum* × *N. otophora,* where one chromosome in a few cells was enlarged by up to 20 times its normal size (see Fig. 2.6A). Megachromosomes, whose size varied from cell to cell, contained additional DNA in proportion to their size (Collins *et al.*, 1970), and appeared to result from the differential replication of a prominent block of heterochromatin. However, it is unknown whether such pronounced changes in chromosomal DNA content are heritable, and/or whether they might be selected against in later generations.

Loss of telomeric heterochromatin from rye (*Secale cereale*) chromosomes was also seen under the light microscope by Gustafson *et al.* (1983) (see Fig. 2.6B). Selection using cytological screening for reduced telomeric C-bands in hexaploid triticale (× *Triticosecale*) (from 23% to 7% of the rye genome length), was accompanied by a detectable reduction in C-value (0.3–0.7 pg, equivalent to one chromosome), and showed that significant (nondeleterious) reductions in DNA amount can be produced in triticale by artificial selection in just a few years. Interestingly, a line with similarly reduced heterochromatin was the first hexaploid triticale variety to be awarded plant breeders' rights in the United Kingdom (Bennett, 1985).

A few additional examples of genuine intraspecific variation involving differences in heterochromatic segments are known for both crops (e.g., in *Secale cereale*) (Bennett *et al.*, 1977) and noncrops. In this latter category, for example, Greilhuber (1998) cited the subspecies pair *Scilla bithynica* Boiss. ssp. *bithynica,* which has many large C-bands (1C = 29.20 pg) and *S. bithynica* ssp. *radkae,* with few small C-bands (1C = 22.90 pg).

The Special Case of Maize

One of the most widely studied angiosperm species that shows clear and significant intraspecific variation in DNA content is maize, *Zea mays* ssp. *mays*. Pachytene chromosomes of maize can have large heterochromatic knobs, and comparative studies have shown that the distribution of knobs is virtually the same as that of heterochromatin detected as C-bands in mitotic chromosomes. Early studies by Brown (1949) and Bennett (1976a) showed that knob number was negatively correlated with both latitude and altitude, but it remained unclear whether any relationship actually existed between knob number and DNA amount. Independent studies by Rayburn *et al*. (1985), Laurie and Bennett (1985), and Tito *et al.* (1991) have since provided an affirmative answer to this question.

For example, Laurie and Bennett (1985) showed 37% variation in 1C DNA amount in 10 accessions of maize ranging from ~3.35 pg in a Mexican race (Zapalote Chico) down to 2.45 pg in the Seneca 60 race from New York State. Further, this was shown to correlate with the number of C-bands observed (see Fig. 2.6C,D). Results from a study of 21 lines of maize from various locations in North America by Rayburn *et al.* (1985) supported this finding, because they showed significant correlations between DNA content and the numbers of C-bands and heterochromatic knobs. By assigning their 21 lines to five relative maturity zones along a south–north axis, Rayburn *et al.* (1985) were able to demonstrate significant negative correlations between DNA C-value and maturity zone (i.e., latitude of origin), and between maturity zone and both heterochromatin amount and knob number. A similar analysis comparing the zones of origin with DNA C-value for the 11 lines examined by Laurie and Bennett (1985) likewise showed a significant negative correlation. Taken together, these studies have provided convincing evidence that the intraspecific variation in genome size in maize is largely caused by differences in the amount of heterochromatin, and that previously reported correlations between geographical location and knob number are actually related to intraspecific differences in nuclear DNA content.

Rayburn *et al.* (1985) suggested that the pattern of distribution of DNA contents in maize "may relate to selection pressures imposed by man … which influence the DNA content via its nucleotypic effects." It is generally agreed that maize originated at low latitude and was taken north by humans until environmental barriers (primarily a shorter growing season) prevented normal maturation. Rayburn *et al.* (1985) speculated that the lower DNA C-values of varieties adapted to high latitudes may have resulted from simultaneous selection by humans for earlier maturation and maximum plant size and yield. This may have involved selection for more cells, which could result from the shorter mitotic cycle time that correlates with reduced DNA C-value (see Fig. 2.11A). Significant positive correlations have also been found in maize between DNA content and altitude of cultivation (Rayburn and Auger, 1990) and effective growing season (Bullock and Rayburn, 1991), whereas negative relationships have been reported between genome size and various growth and yield parameters (Biradar *et al.,* 1994). Clearly, where intraspecific variation is real, it can be of considerable adaptive significance—to people as well as plants.

Genuine Intraspecific Variation in Nonangiosperms

To date, no studies have reported significant intraspecific variation in pteridophytes or bryophytes. In gymnosperms, on the other hand, cases of intraspecific variation have been claimed in some species. For example, Miksche (1968, 1971) reported intraspecific variation in three gymnosperms, including 58% in *Picea glauca* and

92% in *P. sitchensis*, related to their latitude of growth in North America. However, paralleling the work in angiosperms, subsequent reinvestigations of the material collected from the same range sampled by Miksche failed to repeat these observations (e.g., Teoh and Rees, 1976). Other similar examples are reviewed by Murray (1998). Indeed, a number of more recent studies have failed to find evidence of intraspecific variation even when measurements have been made using different methods in different laboratories (e.g., Ohri and Khoshoo, 1986; Wakamiya *et al.,* 1993) or on highly disjunct populations (Auckland *et al.,* 2001). Intraspecific variation was reported in *Ginkgo biloba* by Marie and Brown (1993), but unfortunately this was done without mention of the sexes of the plants sampled. Because *Ginkgo* is one of the relatively few plants with sex chromosomes, this may represent a case of genuine intraspecific variation. In summary, more detailed studies of a wide range of species are needed to obtain a clearer picture of the extent of intraspecific variation in gymnosperms, following the same principles as outlined for work in angiosperms.

Intraspecific Variation and Speciation

Some closely related species may have very similar C-values (e.g., *Hordeum bulbosum, H. glaucum, H. marinum,* and *H. murinum,* which all have 1C DNA values of 5.5 pg) (Bennett and Smith, 1976), whereas in other species individual lines may have distinct C-values (e.g., different lines of *Zea mays* have C-values that vary by 35%) (Laurie and Bennett, 1985). The implication is that there is no absolute link between the process of speciation and changes in genome size. That is, speciation may occur without any detectable change in C-value, and likewise variation in DNA amount (both gain and/or loss, mostly of repeated sequences) can also precede reproductive isolation and morphological diversification.

Speciation was once thought to depend mainly on changes in informational genes. However, comparative genomics has emphasised striking elements of constancy in this part of the genome. Thus Devos and Gale (1997) noted that "gene content and gene orders are highly conserved between species within the grass family, and that the amount and organization of repetitive sequences has diverged considerably." Such results have led to a rethinking of the role of noncoding repeated DNA sequences in determining diversity, and even to the suggestion that they play a major role in plant speciation (Kubis *et al.,* 1998).

To date, there is no research to indicate in any general way whether species normally diverge before detectable variation in genome size occurs, and/or vice versa. Nor, for species displaying variation in DNA amount before distinct morphological divergence, is there any definitive information on how much intraspecific variation in C-value usually occurs before species diverge, or if there is any limit to the amount of variation that may accrue before species divergence

becomes an inevitable consequence. Furthermore, according to Greilhuber (1998), nothing reliable is known about the rate of genome size changes in natural populations. To complicate matters, the answers to these questions will be greatly influenced by the species concept used because amounts of "intraspecific" variation will be much greater for lumpers than for splitters, as illustrated earlier by the example of *Scilla bifolia*.

THE MYSTERY OF DNA CONSTANCY

Part of the current interest in C-values and what determines genome size focuses on a tension between the massive interspecific variation in DNA amounts existing in the angiosperms (Table 2.2), and the surprisingly high degree of genome constancy found in many widely distributed species (e.g., Bennett *et al.,* 2000b). In view of the known molecular mechanisms with the potential to rapidly generate considerable variation in C-value and the clear demonstration in some studies that particular sequences within the genome can fluctuate considerably and often rapidly (e.g., Kalendar *et al.,* 2000), the degree of C-value constancy found in many species is remarkable, and needs explanation. Indeed, it is arguable that such invariance would not be expected without some mechanism(s) to select for constancy (or against changes) in C-value (Bennett *et al.,* 2000b).

Although genome size is widely perceived as free to vary, many results suggest instead that DNA amount may normally be subject to innate controls by "counting mechanisms" that somehow detect, quantify, and regulate genomic size characters within quite tightly defined or preselected limits (Bennett, 1987; Bennett *et al.,* 2000b). For example, several careful studies have shown evidence of "karyotypic orthoselection," whereby large differences in C-value between related plant species involve strictly proportional changes in all chromosomes, which preserves the particular form of a complement (Seal and Rees, 1982; Seal, 1983; Brandham and Doherty, 1998) (Fig. 2.16). This could be a case in which counting mechanisms are at play, acting to produce proportional changes in all linkage groups and/or chromosome arms, presumably driven by some underlying organizational principle.

Clearly, there is still much to be discovered about intraspecific variation in genome size. Fortunately, modern molecular methods can provide powerful new insights into how changes at the DNA sequence level relate to others at the whole genome level, including intraspecific variation in nuclear DNA amount. Comparisons between whole genome sequences for closely related "diploid" subspecies and species with different C-values will soon be possible (see Chapter 9). Meanwhile, complete sequences for homologous chromosome segments in lines of *Zea mays* with different C-values distributed in both knobs and euchromatin should be particularly illuminating. Studies linking DNA information to characters of environmental and ecological interest will provide an important focus for new work

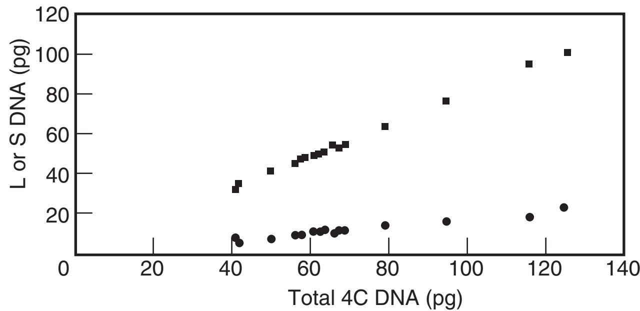

FIGURE 2.16 Karyotypic orthoselection in the genus *Aloe*. The total 4C DNA amounts of 20 *Aloe* species were plotted against the calculated DNA amount for the combined long (L) (■) and combined short (S) (•) chromosomes. For each species the value for the long chromosome is directly above that of the short chromosome. The two highest values are tetraploid, the rest are diploid species. From Brandham and Doherty (1998), reproduced by permission (© Oxford University Press).

on intraspecific variation in plant genome size over the next few years. Such work may also shed light on whether certain types of plants are more predisposed to exhibit gain or loss of repeated sequences, and if this is relevant to understanding patterns of gradualistic versus punctuated changes in genome size over time.

METHODOLOGY FOR ESTIMATING GENOME SIZE IN PLANTS

None of the big questions in plant genome size evolution can be addressed without broad comparative analyses, making the assessment of large numbers of plant genome sizes a crucial first step toward a resolution of the C-value enigma (see also Chapter 1). However, as the previous section showed, it is not the case that any data will do. Rather, these must be collected in such a way as to avoid the numerous technical pitfalls that may otherwise contribute to an erroneous view of genome size variation and evolution. The following sections review the several methods that have been used to estimate genome size in plants, and provide some best practice guidelines for the implementation of the most commonly used techniques. This is meant to facilitate the accurate measurement of plant genome sizes in the future, and to inform critical analyses of estimates reported in the past.

CHEMICAL EXTRACTION AND REASSOCIATION KINETICS

Unlike animals, plants characteristically have cell walls thickened with polymers (e.g., cellulose and lignin) whose presence may complicate the extraction of DNA.

Nevertheless, chemical extraction methods developed for use in animals (Schmidt and Thannhauser, 1945) were later successfully applied to plants (Sunderland and McLeish, 1961; Lyndon, 1963; Rothfels *et al.,* 1966). Under these techniques, the total DNA is extracted from a sample of cells and dissolved in a known volume of solvent. The concentration of DNA is estimated colorimetrically using a modification of the diphenylamine reaction (Burton, 1956, 1968) to give a color reaction whose intensity is proportional to the concentration of deoxyribose sugar, and hence of DNA. The method typically used very large numbers of cells, whose total was estimated using a hemocytometer. Results were usually expressed as mean DNA amount per cell, which consisted mainly of nuclear DNA, but also included small amounts from organelle (mitochondrial and chloroplast) genomes. The results were calibrated in absolute units using either a plot of color intensity against known concentrations of DNA or deoxyribose, or by ultraviolet spectrophotometry with known concentrations of DNA as a standard (Sunderland and McLeish, 1961). Although these methods were usefully reliable, they were complicated and slow, and hence rate-limiting for many larger comparisons; thus these methods have been rarely used since the 1960s.

In the 1970s and 1980s, it became relatively common to employ reassociation kinetics to assess the composition and size of genomes. In this method, DNA is extracted from cells and then denatured by heating. The rates of reassociation of the DNA strands indicate the relative copy number of repeated DNA sequences, and can be calibrated against a standard to give an estimate of absolute DNA content. However, as a method for measuring genome size, reassociation kinetics is very slow (the reassociation of some DNA fragments can last for several days) and not particularly accurate. Not surprisingly, this technique was only rarely used for genome size estimates in plants (Table 2.3). Instead, nearly all modern genome size measurements are made by either Feulgen microdensitometry or flow cytometry.

FEULGEN MICRODENSITOMETRY

The advent and widespread application of photomicrodensitometry represented a major step forward in genome size estimation (Table 2.3). Rather than extracting the DNA from cells, this method involves staining the nuclei and then measuring the amount of light absorbed by the stain. The most commonly employed method of staining is still the Feulgen reaction, first developed by Feulgen and Rössenbeck in 1924. In this case, aldehyde groups are freed by hydrolyzing the DNA with strong acid, followed by staining with Schiff's reagent containing leuco-basic fuchsin, which gives a purple coloration when it complexes with the aldehyde groups. Plant material is typically prepared for Feulgen microdensitometry using various modifications of the method described by McLeish and Sunderland (1961) for plants, itself a variant of the method of Leuchtenberger (1958) for

TABLE 2.3 Summary of the main methods used to estimate genome size in a sample of 5844 angiosperms over the period between 1950 and 2003. Data are grouped into five-year periods (except for 2000–2003). Ch = chemical extraction; Fe = Feulgen microdensitometry; FC = flow cytometry; RK = reassociation kinetics; CGS = complete genome sequencing; Fe/IA = Feulgen image analysis densitometry.

	Estimation method					
Time period	Ch	Fe	FC	RK	CGS	Fe/IA
1950–1954	1	0	0	0	0	0
1955–1959	0	5	0	0	0	0
1960–1964	24	1	0	0	0	0
1965–1969	47	101	0	0	0	0
1970–1974	6	434	0	0	0	0
1975–1979	2	573	0	14	0	0
1980–1984	11	757	26	12	0	0
1985–1989	0	883	36	0	0	0
1990–1994	2	449	479	0	0	0
1995–1999	0	687	550	0	0	0
2000–2003	0	291	438	0	3	12
Total	**93**	**4181**	**1529**	**26**	**3**	**12**

animal tissues. It is generally accepted that Feulgen staining is specific and stoichiometric for DNA, meaning that a measure of nuclear stain concentration (calculated from total optical density), compared against a standard of known DNA amount, provides an accurate estimate of C-value.

Detailed reviews of standardized staining and measurement protocols have been provided recently (Greilhuber and Baranyi, 1999; Greilhuber and Temsch, 2001; Greilhuber, 2005), which the interested reader is urged to consult before performing new estimates of plant genome size. As noted previously, a failure to follow best practice guidelines can lead to substantial errors in results and thence to a false understanding of patterns and mechanisms. Common sources of error include biological, chemical, and physical (optical) factors, some of which are listed here:

1. *Chromatin condensation.* The level of chromatin condensation can vary significantly between different types of plant nuclei (e.g., pollen and egg nuclei), which directly affects the uptake and/or detection of stain, such that DNA amount tends to be underestimated as chromatin condensation increases (Verma and Rees, 1974).
2. *Hydrolysis time.* The acid hydrolysis is perhaps the most sensitive step in the procedure, with both insufficient and excessive hydrolysis resulting in understaining. It is therefore essential that an optimum hydrolysis time is

used to maximize staining of nuclear DNA in all samples being compared (both unknowns and calibration standards). This can be determined by the construction of “hydrolysis curves,” which provide measurements of stain densities of nuclei from the same individual hydrolyzed for different times.

3. *Hydrolysis temperature.* For many years “hot hydrolysis” (~12 minutes in 1M HCl at 60°C) was used, giving only a relatively short (typically 3–5 minutes) plateau of maximum staining. Longer treatment (1–2 hours) in 5M HCl at 20°C (“cold hydrolysis”) gives greater control of this important step and greatly prolongs the plateau of maximum staining. Moreover, it is important that the hydrolysis step is not done simply at “room temperature,” because this may be about 20°C in temperate regions but may be well above 30°C in tropical areas. Failure to recognize the importance of hydrolysis conditions has been a major contributor to false reports of intraspecific variation.
4. *Staining inhibitors.* A quite different, but equally important source of staining error is the effect of cytosolic compounds present in many plant materials that can bind to DNA and greatly reduce its ability to undergo Feulgen staining (“self-tanning”) (Greilhuber, 1988).

Provided that these problems are recognized, and adequate steps taken to apply best practice, Feulgen microdensitometry can give quantitative estimates of nuclear DNA amounts of considerable accuracy, with error variation routinely controlled to within 5%, and sometimes to within 1–3%.

FEULGEN IMAGE ANALYSIS DENSITOMETRY

One major factor likely to limit the future applicability of Feulgen methods is the “obsolescence time bomb” of aging microdensitometers, which are no longer manufactured and are becoming increasingly difficult to repair (see www.rbgkew.org.uk/cval/conference.html#outline) (Bennett *et al.,* 2000a). Fortunately, advances in computing and imaging technology have facilitated the development of inexpensive computer-based image analysis densitometry methods that will allow the continued implementation of the time-tested method of Feulgen staining.

In this technique, DNA is Feulgen-stained as usual, but the density of stain in the nucleus is measured using a microscope-mounted video or digital camera to “grab” images and to display them as composites of pixels on a computer screen. The intensity (gray value) of each pixel can be used to calculate an individual point density, allowing the instant and simultaneous measurement of integrated optical density for all nuclei in the field.

The method was originally developed for DNA quantification in cancer diagnosis (Jarvis, 1986), for which accuracy is obviously of extreme importance, and with the result that scientists have reached an international consensus on the methodology

and best practice (e.g., Chieco *et al.*, 1994; Bocking *et al.*, 1995; Puech and Giroud, 1999). However, its application to DNA quantification in other organisms has been slow to take off. Probably the first reported use of this method in plants was by Temsch *et al.* (1998), who used it to estimate genome sizes in species of the moss genus *Sphagnum*. This was followed by studies of a variety of plant genera, such as *Crepis* (Dimitrova and Greilhuber, 2000), *Gagea* (Greilhuber *et al.*, 2000), *Hedera* (Obermayer and Greilhuber, 2000), and *Arachis* (Temsch and Greilhuber, 2001).

A recent interlaboratory comparison showed the results of Feulgen image analysis densitometry to be very comparable to those obtained by other methods, over a 100-fold range in plant genome sizes (Vilhar *et al.*, 2001). Studies using animal tissues likewise established the validity of the method (Hardie *et al.*, 2002). Further studies by Vilhar and Dermastia (2002) have led to proposals for standardizing the method to maximize accuracy of the data generated in plants. It is expected that image analysis will become one of the most important sources of new genome size estimates in the near future.

FLOW CYTOMETRY

Perhaps the first description of a flow cytometry apparatus was that of Moldavan (1934) for use in counting the number of red blood cells or yeast in a suspension. With extensive development of the equipment and methodology, the technique of flow cytometry has since been adapted to many different applications, including DNA quantification for cancer detection and, more recently, genome sizing.

In plants, this involves mechanically isolating nuclei, usually from leaf tissue, by chopping. The isolated nuclei are stained with a fluorescent dye that binds quantitatively to DNA and then passed through a flow cytometer, which forces nuclei to pass one at a time past a series of lights, lenses, mirrors, and amplifiers that detect and convert the amount of fluorescent light being emitted by each nucleus into a digital signal. By comparing the intensity of fluorescence with that from a plant of a known DNA amount, the absolute DNA content of the plant can be determined. In physical terms, this is the opposite of densitometric methods.

Although it is now one of the primary methods employed (Table 2.3), the application of this technology to plant genome size studies came relatively slowly, limited largely by the difficulty of isolating nuclei from cells with rigid cell walls. The first successful preparations of intact nuclei suitable for plant flow cytometry were made from root tips of *Vicia faba* by Heller (1973), who used the enzymes pectinase and pepsin to digest the cell wall. Although he noted the potential of this method to analyse cell cycle kinetics, the method was time-consuming and was not adopted by other researchers. Alternative approaches were tried in the early 1980s (e.g., use of intact plant protoplasts) (Puite and Tenbroeke, 1983), but these also suffered from being too laborious and time-consuming. The breakthrough

came when Galbraith *et al.* (1983) developed the simple, rapid, and convenient method of chopping to provide isolated nuclei.

Following Galbraith *et al.*'s (1983) paper, various researchers carried out experiments comparing data obtained using flow cytometry with the established method of Feulgen microdensitometry. Generally they found good agreement between the two across a large range of DNA amounts from a broad array of plants (e.g., Hulgenhof *et al.*, 1988; Michaelson *et al.*, 1991; Dickson *et al.*, 1992; Dolezel *et al.*, 1998), so long as best practice techniques were implemented. Two of the most important sources of error turned out to be an inappropriate choice of fluorochrome and the presence of staining inhibitors.

1. *Choice of fluorochromes*. A range of fluorochromes has been used for plant DNA estimations by flow cytometry. These divide into two groups: DNA intercalating dyes (e.g., propidium iodide and ethidium bromide), which bind to DNA independently of the DNA sequence, and base pair–specific dyes, which preferentially bind to AT-rich (e.g., DAPI, Hoechst 33258) or GC-rich (e.g., mithromycin) regions of the DNA. In a comparative study, Dolezel *et al.* (1992) showed that use of base pair–specific dyes could lead to errors approaching 100%, and recommended that only intercalating fluorochromes be used. This recommendation was endorsed at the first Kew Plant Genome Size Workshop in 1997 (see www.rbgkew.org.uk/cval/conference.html#keyrecs).
2. *DNA staining inhibitors and the importance of internal standardization*. Although the use of intercalating fluorochromes overcomes some of the problems encountered in obtaining accurate genome size estimates, in recent years it has become apparent that compounds in the cytoplasm, released during nuclear isolation, can interfere with fluorochrome binding and fluorescence and lead to erroneous genome size data (Noirot *et al.*, 2000, 2002, 2003; Price *et al.*, 2000). Even though the nature of many of these compounds remains elusive, identifying their existence, understanding how their levels may be influenced by environmental and/or genetic factors, and determining how they affect genome size estimations are vital for accurate genome size studies. The importance of testing for the presence of inhibitors was emphasized in the more recent Kew Plant Genome Size Workshop in 2003 (see www.rbgkew.org.uk/cval/workshopreport.html). In practice, this problem can be addressed by chopping, staining, and measuring standards and unknowns together.

If done correctly, flow cytometry has the great advantage of providing rapid and accurate measurements of DNA amount for a large number of nuclei from a small sample of plant tissue. This allows thorough plant population studies to be made *in situ,* enabling the extent and evolutionary significance of intraspecific variation

to be more completely assessed and evaluated. For example, flow cytometry has been the method of choice for comparisons of ecotypes of *Medicago* species (Blondon *et al.,* 1994), and of different populations, F_1 hybrids, and inbred lines of maize (Bullock and Rayburn, 1991; Biradar and Rayburn, 1993).

That said, it is important to recognize that flow cytometry is subject to two major constraints. First, running a flow cytometer can be very expensive and requires high levels of technical support and other infrastructure. Second, using flow cytometry does not remove the need for cytological work on the species being studied, because although it may give a highly accurate DNA value for a taxon, this will be of limited value if the chromosome number (2n) of the individual plant (or even tissue) measured is unknown. Also, if chromosomal variations such as aneuploidy, duplications and deletions, and sex and supernumerary chromosomes are not identified along with flow cytometric measurements, then the interpretation of the results could be flawed. So, although flow cytometry is appealingly fast and highly suited to certain types of studies (e.g., population studies, ploidy screening), the importance of cytological analyses, which can be time-consuming, must not be neglected.

Complete Genome Sequencing

Although thousands of DNA amounts have been determined using the previously discussed techniques over the past 50 years, every one of them is but an *estimate,* inevitably subject to technical errors. For this reason, the need for an exact calibration standard whose C-value is not subject to such errors has long been recognized. Since the mid-1990s, a large number of highly accurate determinations of genome size based on complete genome sequences have been published for numerous microbes (see Chapter 10). Because of the high cost and intensive effort currently required, it is unlikely that complete genome sequencing will become a viable and routine method for determining plant genome size in the near future. However, the first complete genome sequence for a plant was eagerly awaited by those in the plant genome size community, because it was expected to provide a highly accurate baseline reference point for calibrating future estimates by Feulgen densitometry and flow cytometry. By 1997 it was clear that the prime candidate for this honor would almost certainly be *Arabidopsis thaliana* (*Arabidopsis* Genome Initiative, 1997).

December 2000 saw the landmark publication giving the genome sequence for *Arabidopsis thaliana* ecotype Columbia (*Arabidopsis* Genome Initiative, 2000), and an estimate for its genome size of 125 megabases (Mb), based on the size of the sequenced regions (115.4 Mb) plus a rough estimate of 10 Mb for the unsequenced centromere and ribosomal DNA regions. Sadly, this was not the long-awaited benchmark standard, because it was not the result of "complete genome

sequencing" as these words would be understood by most people. Indeed, later work showed that the amount of DNA in the unsequenced regions had been seriously underestimated and that the genome size of *Arabidopsis thaliana* was in fact about 157 Mb (Hosouchi *et al.,* 2002; Bennett *et al.,* 2003).

Some Comments on Plant Genome Size Standards

The early comparative studies in plants used various plant species as calibration standards, including *Pisum sativum* and *Vicia faba* (McLeish and Sunderland, 1961). By the 1970s, onion (*Allium cepa*), had already emerged as the most frequently used calibration standard. Onion has the advantage that it is widely cultivated and globally available as seed or bulbs, can be readily grown to provide a source of actively growing root-tips over long periods, and is highly amenable to cytological techniques to make root-tip squashes. An analysis of 5871 plant taxa in 2004 showed that ~39% were calibrated either against onion (2259 taxa) or a secondary standard that itself was calibrated against onion (52 taxa). It follows from this that the absolute accuracy of genome size estimates for many plants are directly dependent on the accuracy of the value (1C = 16.75 pg) determined for onion by Van't Hof (1965) using chemical methods. Fortunately, this agrees closely with estimates from four independent studies that used animal species (including *Homo sapiens*) as calibration standards. Importantly, recent work found no significant differences in DNA amount between cultivars from widely different environments, confirming that it has the stable genome size required for a key calibration standard (Bennett *et al.,* 2000b).

For technical reasons, however, it is desirable to use a calibration standard whose genome size is within about 2-fold of the unknown taxon being studied. Thus a need was recognized for other calibration standards with smaller genome sizes than *A. cepa,* spread at useful intervals over the range of genome sizes encountered in plants. The first attempt to provide a range of such standards was made by Bennett and Smith (1976), who listed eight species with 1C-values from 1.5 pg to 17 pg, including *Pisum sativum, Hordeum vulgare,* and *Vicia faba,* all calibrated against *A. cepa.* Subsequently, other studies have attempted to refine the values for these taxa, and/or to extend the number and range of standard calibration species to include smaller genomes such as *Vigna radiata, Oryza sativa* (Bennett and Leitch, 1995), *Lycopersicon esculentum* (Obermayer *et al.,* 2002), and *Arabidopsis thaliana* (Bennett *et al.,* 2003).

About 10% of all plant C-value estimates have used an animal standard, 89% of these being chicken red blood cells (CRBC). Chicken (*Gallus domesticus*) is used for several reasons. First, like *Allium cepa,* it is readily accessible across the globe. Second, it has long been used in animal studies (Gregory, 2001b) and was

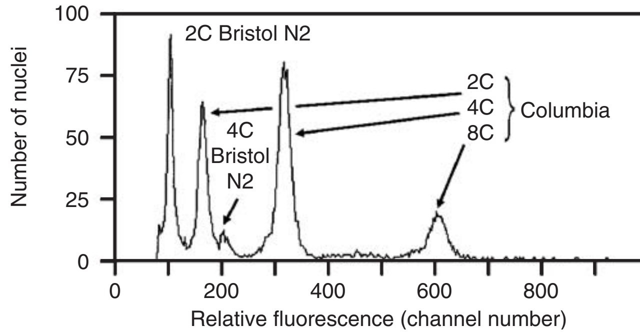

FIGURE 2.17 Flow cytometry histogram showing the relative staining in nuclei of *Arabidopsis thaliana* ecotype Columbia and *Caenorhabditis elegans* variety Bristol N2 (1C = 100 Mb). *Arabidopsis* has approximately 157% of the 2C nuclear DNA fluorescence of *Caenorhabdites*, resulting in a 1C value for *Arabidopsis* of 157 Mb, contrary to the 125 Mb estimate given by the *Arabidopsis* Genome Initiative (2000). From Bennett *et al.* (2003), reproduced by permission (© Oxford University Press).

therefore chosen by Galbraith *et al.* (1983) in their seminal work developing flow cytometry for plant genome size studies. Third, it is an amenable material: one animal can be bled nondestructively at intervals to provide a stable and constant standard for several years. However, in general the use of chicken blood, and animal standards in general, has been discouraged for plant studies because of the different DNA condensation levels and other staining properties of animal versus plant cells. There is one important exception to this, however, namely the potential use of the nematode *Caenorhabditis elegans* (which is the only multicellular organism whose genome has actually been completely sequenced) as the basal calibration standard for genome size comparisons in both plants and animals. Indeed, *C. elegans* was the species used to identify the previous error in the *Arabidopsis thaliana* genome size estimate (Fig. 2.17) (Bennett *et al.*, 2003). Work is already under way to create a ladder of reliable standards, from *A. thalania* on up, using *C. elegans* as the baseline "gold standard."

CONCLUDING REMARKS AND FUTURE PROSPECTS

In an important sense, the field of genome size research in plants may be considered to have reached "the end of the beginning." The first big push for representative and accessible coverage of plant genome size data has been completed by the launch of the *Plant DNA C-values Database* (Bennett and Leitch, 2003). This has allowed at least a basic understanding of the major patterns and consequences of

genome size variation, although much work remains to be done in these areas. With a large and growing dataset, the field is poised to tackle some of the key comparative questions relating to the biological, ecological, and evolutionary importance of genome size variation.

Expansion of the Plant Genome Size Dataset

Although the *Plant DNA C-values Database* includes representatives from each of the major land plant groups, the percent coverage at the species level remains very poor (generally <2%) for all but the gymnosperms (see Table 2.1). Furthermore, Release 2.0 of the database does not currently contain information for any algae. This latter gap will be filled in the next release of the database because Kapraun (2005) recently compiled three reference lists containing C-value data for 240 algal species, including 85 green algae (Chlorophyta), 111 red algae (Rhodophyta), and 44 brown algae (Phaeophyta). It also bears noting in more optimistic terms that the current state of knowledge of plant genome size is largely due to significant progress made in recent years (e.g., Leitch *et al.,* 2001; Leitch and Hanson, 2002; Hanson *et al.,* 2003), indicating that the rate of data acquisition is accelerating.

One of the key driving forces behind the expansion of the genome size dataset has been the targets set at the Plant Genome Size Workshops held at the Royal Botanic Gardens, Kew in 1997 and 2003. These provided a forum for identifying crucial gaps in plant C-value knowledge, and for setting targets to fill them. In September 2003, the progress achieved after the 1997 meeting was reviewed and new five-year goals were set:

1. *Angiosperms:* To estimate first C-values for the next 1% of species (i.e., an additional 2500 species). Within this, targets of achieving 75% familial (i.e., an additional ~114 families) and 10% generic (i.e., an additional ~400 genera) representation were set.
2. *Pteridophytes:* To estimate first C-values for a further 100 species, with particular emphasis on leptosporangiate ferns (the most diverse group of land plants after the angiosperms).
3. *Bryophytes:* To improve geographical representation by targeting species from the tropical and southern hemisphere floras (no data are currently available for species in these regions). Further, for conservation studies targets were set to estimate C-values in rare taxa in the European flora.
4. *Algae:* Two groups were identified as targets for future C-value research: the Micromonadophyceae, which are considered to hold a place close to the origin of plants, and Charophyceae, now recognized as the sister group to land plants.

Mechanistic Questions

One of the most fundamental questions in the field of genome size research is "How and why do DNA amounts change?" In recent years there have been huge advances in understanding how DNA amounts can increase and decrease, but this probably represents only the tip of the iceberg. Future research should continue to search for novel mechanisms capable of generating changes in DNA amount, and use comparative studies to determine the extent to which such mechanisms are either specific to a particular taxon or universal to all plants. In addition, the natures of the evolutionary forces acting on these mechanisms to promote or restrict their activity are largely unknown, but are becoming increasingly amenable to study with the development of large-scale comparative genomics. All of these issues are essential for shedding light on the molecular bases underlying the C-value enigma.

Ecological and Environmental Questions

Another key question for future researchers to address is "What role does genome size play in the response of plants to their environment?" Given that DNA amount is linked to numerous cytological, morphological, and physiological parameters, it will be a critical component of attempts to determine the evolutionary fates of plants subjected to changing ecological and environmental conditions. Already there is evidence that genome size plays an important, yet complex, role in affecting how a plant may respond to anthropogenic changes in the environment such as global warming (Grime, 1996), increased nuclear radiation (Sparrow and Miksche, 1961; Underbrink and Pond, 1976), elevated CO_2 levels (Jasienski and Bazzaz, 1995), and pollution (Vilhar, personal communication). These questions can be approached from various perspectives, including comparative surveys across natural ecological gradients, experimental manipulations in the laboratory, and theoretical modeling designed to examine how organisms will respond to the multifaceted environmental changes induced through human activities. Such work also promises to illuminate the factors that influence the natural distributions of plants with differing genome sizes.

Evolutionary Questions

The ability to track changes in C-value over time using characters such as fossil cell size as a proxy for DNA content (Masterson, 1994) will enable the evolutionary patterns of genome size evolution to be viewed over millions of years. Such approaches will provide insights into the role genome size has played during key evolutionary developments in plants, such as the transition onto land, the development of

a vascular system for efficient water conduction, the evolution of the seed habit found only in angiosperms and gymnosperms, and the evolution of angiosperms themselves, the most diverse and species-rich group of plants on this planet.

Studies of extant taxa can also allow insights into the origin of long-term evolutionary patterns, for example by combining genome size data with detailed information of plant phylogeny (Leitch *et al.,* 1998, 2005; Soltis *et al.,* 2003) and by further investigating the links between genome size and susceptibility to extinction (Vinogradov, 2003) and propensity for speciation (Knight *et al.,* 2005).

In the roughly 50 years that have elapsed since the first plant genome size estimate was recorded, genome size research has moved from simply documenting C-values and expressing confusion over the "C-value paradox" to identifying the major patterns of variation, illuminating some of the molecular mechanisms and evolutionary forces responsible, and framing a specific series of questions as part of the "C-value enigma." With continued surveys of plant genome size and integration with other fields ranging from molecular biology to ecology, it is possible that the future will finally see the emergence of the "C-value solution(s)." Such an effort will clearly require an appreciation of the biological and evolutionary significance of genome size variation and the development of a more holistic framework for genomics.

REFERENCES

Abrahamson S, Bender MA, Conger AD, Wolff S. 1973. Uniformity of radiation-induced mutation rates among different species. *Nature* 245: 460–462.

Anderson LK, Stack SM, Fox MH, Zhang CS. 1985. The relationship between genome size and synaptonemal complex length in higher plants. *Exp Cell Res* 156: 367–378.

Arabidopsis Genome Initiative. 1997. Objective: The complete sequence of a plant genome. *Plant Cell* 9: 476–478.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.

Auckland LD, Johnston JS, Price HJ, Bridgwater FE. 2001. Stability of nuclear DNA content among divergent and isolated populations of Fraser fir. *Can J Bot* 79: 1375–1378.

Avdulov NP. 1931. Karyo-systematische Untersuchungen der Familie Gramineen. *Bull Appl Bot Genet Plant Breed (44th Suppl)* 4: 1–428.

Axelsson T, Bowman CM, Sharpe AG, *et al.* 2000. Amphidiploid *Brassica juncea* contains conserved progenitor genomes. *Genome* 43: 679–688.

Baetcke KP, Sparrow AH, Nauman CH, Schwemmer SS. 1967. The relationship of DNA content to nuclear and chromosome volumes and to radiosensitivity (LD_{50}). *Proc Natl Acad Sci USA* 58: 533–540.

Barakat A, Carels N, Bernardi G. 1997. The distribution of genes in the genomes of Gramineae. *Proc Natl Acad Sci USA* 94: 6857–6861.

Baranyi M, Greilhuber J. 1995. Flow cytometric analysis of genome size variation in cultivated and wild *Pisum sativum* (Fabaceae). *Plant Syst Evol* 194: 231–239.

Baranyi M, Greilhuber J. 1996. Flow cytometric and Feulgen densitometric analysis of genome size variation in *Pisum*. *Theor Appl Genet* 92: 297–307.

Bassi P. 1990. Quantitative variations of nuclear DNA during plant development. A critical approach. *Biol Rev* 65: 185–225.

Baumel A, Ainouche M, Kalendar R, Schulman AH. 2002. Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* CE Hubbard (Poaceae). *Mol Biol Evol* 19: 1218–1227.

Bedbrook JR, Jones J, O'Dell M, *et al*. 1980. A molecular description of telomeric heterochromatin in *Secale* species. *Cell* 19: 545–560.

Bennett MD. 1970. Natural variation in nuclear characters of meristems in *Vicia faba*. *Chromosoma* 29: 317–335.

Bennett MD. 1971. The duration of meiosis. *Proc R Soc Lond B* 178: 277–299.

Bennett MD. 1972. Nuclear DNA content and minimum generation time in herbaceous plants. *Proc R Soc Lond B* 181: 109–135.

Bennett MD. 1973. Nuclear characters in plants. *Brookhaven Symp Biol* 25: 344–366.

Bennett MD. 1976a. DNA amount, latitude and crop plant distribution. In: Jones K, Brandham PE eds. *Current Chromosome Research*. Amsterdam: North Holland Publishing Company, 151–158.

Bennett MD. 1976b. DNA amount, latitute and crop plant distribution. *Env Exp Biol* 16: 93–108.

Bennett MD. 1977. The time and duration of meiosis. *Philos Trans R Soc Lond B* 277: 201–277.

Bennett MD. 1985. Intraspecific variation in DNA amount and the nucleotypic dimension in plant genetics. In: Freeling M ed. *Plant Genetics. UCLA Symposia on Molecular and Cellular Biology*. New York: Alan R Liss Inc., 283–302.

Bennett MD. 1987. Variation in genomic form in plants and its ecological implications. *New Phytol* 106: 177–200.

Bennett MD, Leitch IJ. 1995. Nuclear DNA amounts in angiosperms. *Ann Bot* 76: 113–176.

Bennett MD, Leitch IJ. 1997. Nuclear DNA amounts in angiosperms: 583 new estimates. *Ann Bot* 80: 169–196.

Bennett MD, Leitch IJ. 2001. Nuclear DNA amounts in pteridophytes. *Ann Bot* 87: 335–345.

Bennett MD, Leitch IJ. 2003. Plant DNA C-values database (release 2.0, Jan. 2003). www.rbgkew.org.uk/cval/homepage.html.

Bennett MD, Leitch IJ. 2005. Nuclear DNA amounts in angiosperms—progress, problems and prospects. *Ann Bot* (in press).

Bennett MD, Smith JB. 1971. The 4C nuclear DNA content of several *Hordeum* genotypes. *Can J Genet Cytol* 13: 607–611.

Bennett MD, Smith JB. 1972. The effects of polyploidy on meiotic duration and pollen development in cereal anthers. *Proc R Soc Lond B* 181: 81–107.

Bennett MD, Smith JB. 1976. Nuclear DNA amounts in angiosperms. *Philos Trans R Soc Lond B* 274: 227–274.

Bennett MD, Smith JB. 1991. Nuclear DNA amounts in angiosperms. *Philos Trans R Soc Lond B* 334: 309–345.

Bennett MD, Bhandol P, Leitch IJ. 2000a. Nuclear DNA amounts in angiosperms and their modern uses: 807 new estimates. *Ann Bot* 86: 859–909.

Bennett MD, Dover GA, Riley R. 1974. Meiotic duration in wheat genotypes with or without homoeologous meiotic chromosome pairing. *Proc R Soc Lond B* 187: 191–207.

Bennett MD, Gustafson JP, Smith JB. 1977. Variation in nuclear DNA in the genus *Secale*. *Chromosoma* 61: 149–176.

Bennett MD, Heslop-Harrison JS, Smith JB, Ward JP. 1983. DNA density in mitotic and meiotic metaphase chromosomes of plants and animals. *J Cell Sci* 63: 173–179.

Bennett MD, Johnston S, Hodnett GL, Price HJ. 2000b. *Allium cepa* L. cultivars from four continents compared by flow cytometry show nuclear DNA constancy. *Ann Bot* 85: 351–357.

Bennett MD, Leitch IJ, Hanson L. 1998. DNA amounts in two samples of angiosperm weeds. *Ann Bot* 82: 121–134.

Bennett MD, Leitch IJ, Price HJ, Johnston JS. 2003. Comparison with *Caenorhabditis* (~100 Mb) and *Drosophila* (~175 Mb) using flow cytometry show genome size in *Arabidopsis* to be ~157 Mb and thus 25% larger than the *Arabidopsis* genome initiative estimate of ~125 Mb. *Ann Bot* 91: 547–557.
Bennett MD, Smith JB, Heslop-Harrison JS. 1982a. Nuclear DNA amounts in angiosperms. *Proc R Soc Lond B* 216: 179–199.
Bennett MD, Smith JB, Kemble R. 1972. The effect of temperature on meiosis and pollen development in wheat and rye. *Can J Genet Cytol* 14: 615–624.
Bennett MD, Smith JB, Lewis Smith RI. 1982b. DNA amounts of angiosperms from the Antarctic and South Georgia. *Env Exp Biol* 22: 307–318.
Bennett MD, Smith JB, Ward J, Jenkins G. 1981. The relationship between nuclear DNA content and centromere volume in higher plants. *J Cell Sci* 47: 91–115.
Bennetzen JL, Kellogg EA. 1997. Do plants have a one-way ticket to genomic obesity? *Plant Cell* 9: 1509–1514.
Bennetzen JL, Ma JX. 2003. The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. *Curr Opin Plant Biol* 6: 128–133.
Bennetzen JL, Ma JX, Devos KM. 2005. Mechanisms of recent genome size variation in flowering plants. *Ann Bot* (in press).
Biradar DP, Rayburn AL. 1993. Heterosis and nuclear DNA content in maize. *Heredity* 71: 300–304.
Biradar DP, Bullock D, Rayburn A. 1994. Nuclear DNA amount, growth and yield parameters in maize. *Theor Appl Genet* 88: 557–560.
Blondon F, Marie D, Brown S, Kondorosi A. 1994. Genome size and base composition in *Medicago sativa* and *Medicago truncatula* species. *Genome* 37: 264–270.
Bocking A, Giroud F, Reith A. 1995. Consensus report of the ESCAP task-force on standardization of diagnostic DNA image cytometry. *Analyt Cell Pathol* 8: 67–74.
Boivin A, Vendrely R, Vendrely C. 1948. L'acide desoxyribonucleique du noyau cellulaire dépositaire des caracteres hereditaires: arguments d'ordre analytique. *C R Acad Sci* 226: 1061–1063.
Bowers JE, Chapman BA, Rong JK, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433–438.
Brandham PE, Doherty MJ. 1998. Genome size variation in the Aloaceae, an angiosperm family displaying karyotypic orthoselection. *Ann Bot* 82: 67–73.
Brown WL. 1949. Numbers and distribution of chromosome knobs in United States maize. *Genetics* 34: 524.
Bullock D, Rayburn A. 1991. Genome size variation in southwestern US indian maize populations may be a function of effective growing season. *Maydica* 36: 247–250.
Burton K. 1956. A study of the conditions and mechanism of the diphenylamine reaction for the colorimetric estimation of deoxyribonucleic acid. *Biochem J* 62: 315–323.
Burton K. 1968. Determination of DNA concentration with diphenylamine. *Methods Enzymol* 12B: 163–166.
Caceres ME, De Pace C, Mugnozza GTS, *et al*. 1998. Genome size variations within *Dasypyrum villosum*: correlations with chromosomal traits, environmental factors and plant phenotypic characteristics and behaviour in reproduction. *Theor Appl Genet* 96: 559–567.
Cavalier-Smith T. 1985. *The Evolution of Genome Size*. Chichester: John Wiley & Sons Ltd.
Cavallini A, Natali L. 1991. Intraspecific variation of nuclear DNA content in plant species. *Caryologia* 44: 93–107.
Cavallini A, Natali L, Cionini PG, Gennai D. 1993. Nuclear DNA variability within *Pisum sativum* (Leguminosae): nucleotypic effects on plant growth. *Heredity* 70: 561–565.
Cavallini A, Natali L, Giordani T, *et al*. 1996. Nuclear DNA changes within *Helianthus annuus* L.: variations in the amount and methylation of repetitive DNA within homozygous progenies. *Theor Appl Genet* 92: 285–291.
Ceccarelli M, Falistocco E, Cionini PG. 1992. Variation in genome size and organization within hexaploid *Festuca arundinacea*. *Theor Appl Genet* 83: 273–278.

Chieco P, Jonker A, Melchiorri C, *et al*. 1994. A user's guide for avoiding errors in absorbency image cytometry: a review with original experimental observations. *Histochem J* 26: 1–19.

Collins GB, Anderson MK, Legg PD. 1970. Cytophotometric determination of DNA content in *Nicotiana* megachromosomes. *Can J Genet Cytol* 12: 377–378.

Comai L. 2000. Genetic and epigenetic interactions in allopolyploid plants. *Plant Mol Biol* 43: 387–399.

Costich DE, Meagher TR, Yurkow EJ. 1991. A rapid means of sex identification in *Silene latifolia* by use of flow cytometry. *Plant Mol Biol Report* 9: 359–370.

Courties C, Perasso R, Chretiennot-Dinet MJ, *et al*. 1998. Phylogenetic analysis and genome size of *Ostreococcus tauri* (Chlorophyta, Prasinophyceae). *J Phycol* 34: 844–849.

Cox AV, Abdelnour GJ, Bennett MD, Leitch IJ. 1998. Genome size and karyotype evolution in the slipper orchids (Cypripedioideae : Orchidaceae). *Am J Bot* 85: 681–687.

Devos KM, Gale MD. 1997. Comparative genetics in the grasses. *Plant Mol Biol* 35: 3–15.

Devos KM, Brown JKM, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12: 1075–1079.

Dickson EE, Arumuganathan K, Kresovich S, Doyle JJ. 1992. Nuclear DNA content variation within the Rosaceae. *Am J Bot* 79: 1081–1086.

Dimitrova D, Greilhuber J. 2000. Karyotype and DNA content evolution in ten species of *Crepis* (Asteraceae) distributed in Bulgaria. *Bot J Linn Soc* 132: 281–297.

Dolezel J, Greilhuber J, Lucretti S, *et al*. 1998. Plant genome size estimation by flow cytometry: inter–laboratory comparison. *Ann Bot* 82 (Suppl. A): 17–26.

Dolezel J, Sgorbati S, Lucretti S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. *Physiol Plantarum* 85: 625–631.

Ellegren H. 2002. Mismatch repair and mutational bias in microsatellite DNA. *Trends Genet* 18: 552.

Evans GM, Rees H. 1971. Mitotic cycles in dicotyledons and monocotyledons. *Nature* 212: 697–699.

Evans GM, Durrant A, Rees H. 1966. Associated nuclear changes in the induction of flax genotrophs. *Nature* 212: 697–699.

Evans GM, Rees H, Snell CL, Sun S. 1972. The relationship between nuclear DNA amount and the duration of the mitotic cycle. *Chromosomes Today* 3: 24–31.

Fay MF, Cowan RS, Leitch IJ. 2005. The effects of DNA amount on the quality and utility of AFLP fingerprints. *Ann Bot* (in press).

Feschotte C, Jiang N, Wessler SR. 2002. Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3: 329–341.

Feulgen R, Rössenbeck H. 1924. Mikroskopisch-chemischer Nachweiss einer Nucleinsäure von Typus der Thymonucleinsäure und die darauf beruhende elective Färbung von Zellkernen in mikroskopischen präparaten. *Hoppe-Seyler Z Physiol Chem* 135: 203–248.

Filkowski J, Kovalchuk O, Kovalchuk I. 2004. Dissimilar mutation and recombination rates in *Arabidopsis* and tobacco. *Plant Sci* 166: 265–272.

Flavell RB, Rimpau J, Smith DB. 1977. Repeated sequence DNA relationships in four cereal genomes. *Chromosoma* 63: 205–222.

Ford CE, Hamerton JL. 1956. Chromosomes of Man. *Nature* 178: 1020–1023.

Frediani M, Colonna N, Cremonini R, *et al*. 1994. Redundancy modulation of nuclear DNA sequences in *Dasypyrum villosum*. *Theor Appl Genet* 88: 167–174.

Furuta Y, Nishikawa K, Makino T. 1975. Intraspecific variation in nuclear DNA content in *Aegilops squarrosa*. *Jap J Genet* 50: 257–263.

Galbraith DW, Harkins KR, Maddox JM, *et al*. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. *Science* 220: 1049–1051.

Garner TWJ. 2002. Genome size and microsatellites: the effect of nuclear size on amplification potential. *Genome* 45: 212–215.

Gaut BS, Doebley JF. 1997. DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc Natl Acad Sci USA* 94: 6809–6814.

Gerstel DU, Burns JA. 1966. Chromosomes of unusual length in hybrids between two species of *Nicotiana*. *Chromosomes Today.* 1: 41–56.
Goff SA, Ricke D, Lan TH, *et al.* 2002. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *japonica*). *Science* 296: 92–100.
González-Aguilera JJ, Ludeña Reyes P, Fernández-Peralta AM. 1990. Intra- and interspecific variation in nuclear parameters of two closely related species of *Narcissus* L. *Genetica* 82: 25–31.
Graham MJ, Nickell CD, Rayburn AL. 1994. Relationship between genome size and maturity group in soybean. *Theor Appl Genet* 88: 429–432.
Grandbastien MA. 1998. Activation of plant retrotransposons under stress conditions. *Trends Plant Sci* 3: 181–187.
Greenlee JK, Rai KS, Floyd AD. 1984. Intraspecific variation in nuclear DNA content in *Collinsia verna* Nutt (Scrophulariaceae). *Heredity* 52: 235–242.
Gregory TR. 2001a. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol Rev* 76: 65–101.
Gregory TR. 2001b. *Animal Genome Size Database*. www.genomesize.com.
Gregory TR. 2005. The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann Bot* (in press).
Greilhuber J. 1988. Self-tanning: a new and important source of stoichiometric error in cytophotometric determination of nuclear DNA content in plants. *Plant Syst Evol* 158: 87–96.
Greilhuber J. 1998. Intraspecific variation in genome size: a critical reassessment. *Ann Bot* 82 (Suppl. A): 27–35.
Greilhuber J. 2005. Intraspecific variation in genome size in angiosperms: identifying its existence. *Ann Bot* (in press).
Greilhuber J, Baranyi M. 1999. Feulgen densitometry: importance of a stringent hydrolysis regime. *Plant Biol* 1: 538–540.
Greilhuber J, Ebert I. 1994. Genome size variation in *Pisum sativum*. *Genome* 37: 646–655.
Greilhuber J, Obermayer R. 1997. Genome size and maturity group in *Glycine max* (soybean). *Heredity* 78: 547–551.
Greilhuber J, Speta F. 1985. Geographical variation of genome size at low taxonomic levels in the *Scilla bifolia* alliance (Hyacinthaceae). *Flora* 176: 431–438.
Greilhuber J, Temsch EM. 2001. Feulgen densitometry: some observations relevant to best practice in quantitative nuclear DNA content determination. *Acta Bot Croat* 60: 285–298.
Greilhuber J, Dolezel J, Lysak MA, Bennett MD. 2005. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value.' *Ann Bot* (in press).
Greilhuber J, Ebert I, Lorenz A, Vyskot B. 2000. Origin of facultative heterochromatin in the endosperm of *Gagea lutea* (Liliaceae). *Protoplasma* 212: 217–226.
Grime JP. 1983. Prediction of weed and crop response to climate based upon measurements of nuclear DNA content. *Aspects Appl Biol* 4: 87–98.
Grime JP. 1996. Testing predictions of the impacts of global change on terrestrial ecosystems. *Aspects Appl Biol* 45: 3–13.
Grime JP, Mowforth MA. 1982. Variation in genome size: an ecological interpretation. *Nature* 299: 151–153.
Grime JP, Shacklock JML, Band SR. 1985. Nuclear DNA contents, shoot phenology and species co-existence in a limestone grassland community. *New Phytol* 100: 435–445.
Gustafson JP, Lukaszewski AJ, Bennett MD. 1983. Somatic deletion and redistribution of telomeric heterochromatin in the genus *Secale* and in *Triticale*. *Chromosoma* 88: 293–298.
Hanson L, Leitch IJ. 2002. DNA amounts for five pteridophyte species fill phylogenetic gaps in C-value data. *Bot J Linn Soc* 140: 169–173.
Hanson L, Brown RL, Boyd A, *et al*. 2003. First nuclear DNA C-values for 28 angiosperm genera. *Ann Bot* 91: 1–8.
Hardie DC, Gregory TR, Hebert PDN. 2002. From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. *J Histochem Cytochem* 50: 735–749.

Heller FO. 1973. DNS-Bestimmung an Keimwurzeln von *Vicia faba* L. mit Hilfe der Impulscytophotometrie. *Berichte Deutsch Bot Gesellsch* 86: 437–441.
Holm-Hanson O. 1969. Algae: amounts of DNA and organic carbon in single cells. *Science* 163: 87–88.
Hosouchi T, Kumekawa N, Tsuruoka H, Kotani H. 2002. Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3. *DNA Res* 9: 117–121.
Hulgenhof E, Weidhase RA, Schlegel R, Tewes A. 1988. Flow cytometric determination of DNA content in isolated nuclei of cereals. *Genome* 30: 565–569.
Ilic K, SanMiguel PJ, Bennetzen JL. 2003. A complex history of rearrangement in an orthologous region of the maize, sorghum, and rice genomes. *Proc Natl Acad Sci USA* 100: 12265–12270.
Jarvis LR. 1986. A microcomputer system for video image-analysis and diagnostic microdensitometry. *Analyt Quant Cytol Histol* 8: 201–209.
Jasienski M, Bazzaz FA. 1995. Genome size and high CO_2. *Nature* 376: 559–560.
Jiang N, Bao ZR, Zhang XY, *et al*. 2003. An active DNA transposon family in rice. *Nature* 421: 163–167.
Jiang N, Feschotte C, Zhang X, Wessler SR. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Curr Opin Plant Biol* 7: 115–119.
Johnston JS, Jensen A, Czeschin DG, Price HJ. 1996. Environmentally induced nuclear 2C DNA content instability in *Helianthus annuus* (Asteraceae). *Am J Bot* 83: 1113–1120.
Jones RN, Rees H. 1968. Nuclear DNA variation in *Allium*. *Heredity* 23: 591–605.
Kalendar R, Tanskanen J, Immonen S, *et al*. 2000. Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. *Proc Natl Acad Sci USA* 97: 6603–6607.
Kapraun DF. 2005. Nuclear DNA content estimates in multicellular eukaryotic green, red and brown algae: phylogenetic considerations. *Ann Bot* (in press).
Karp A, Rees H, Jewell AW. 1982. The effects of nucleotype and genotype upon pollen grain development in *Hyacinth* and *Scilla*. *Heredity* 48: 251–261.
Kashkush K, Feldman M, Levy AA. 2002. Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. *Genetics* 160: 1651–1659.
Kentner EK, Arnold ML, Wessler SR. 2003. Characterization of high-copy-number retrotransposons from the large genomes of the Louisiana iris species and their use as molecular markers. *Genetics* 164: 685–697.
Kenton A, Parokonny A, Bennett ST, Bennett MD. 1993. Does genome organization influence speciation? A reappraisal of karyotype studies in evolutionary biology. In: Lees DR, Edwards D eds. *Evolutionary Patterns and Processes*. London: Academic Press, 189–206.
Kidwell MG. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* 115: 49–63.
Kimura Y, Tosa Y, Shimada S, *et al.* 2001. OARE-1, a *Ty1-copia* retrotransposon in oat activated by abiotic and biotic stresses. *Plant Cell Physiol* 42: 1345–1354.
Kirik A, Salomon S, Puchta H. 2000. Species-specific double-strand break repair and genome evolution in plants. *EMBO J* 19: 5562–5566.
Knight CA, Ackerly DD. 2002. Variation in nuclear DNA content across environmental gradients: a quantile regression analysis. *Ecol Lett* 5: 66–76.
Knight CA, Molinari N, Petrov DA. 2005. The large genome constraint hypothesis: evolution, ecology and phenotype. *Ann Bot* (in press).
Kubis S, Schmidt T, Heslop-Harrison JS. 1998. Repetitive DNA elements as a major component of plant genomes. *Ann Bot* 82: 45–55.
Kumar A, Bennetzen JL. 1999. Plant retrotransposons. *Annu Rev Genet* 33: 479–532.
Laurie DA, Bennett MD. 1985. Nuclear DNA content in the genera *Zea* and *Sorghum:* intergeneric, interspecific and intraspecific variation. *Heredity* 55: 307–313.
Leitch IJ, Bennett MD. 2004. Genome downsizing in polyploid plants. *Biol J Linn Soc* 82:651–663.
Leitch IJ, Hanson L. 2002. DNA C-values in seven families fill phylogenetic gaps in the basal angiosperms. *Bot J Linn Soc* 140: 175–179.

Leitch IJ, Chase MW, Bennett MD. 1998. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann Bot* 82: 85–94.

Leitch IJ, Hanson L, Winfield M, *et al*. 2001. Nuclear DNA C-values complete familial representation in gymnosperms. *Ann Bot* 88: 843–849.

Leitch IJ, Soltis DE, Soltis PS, Bennett MD. 2005. Evolution of DNA amounts across land plants (Embryophyta). *Ann Bot* (in press).

Leuchtenberger C. 1958. Quantitative determination of DNA in cells by Feulgen microspectrophotometry. In: Danielli JF ed. *General Cytochemical Methods*. New York: Academic Press, 219–278.

Levin DA, Funderburg SW. 1979. Genome size in angiosperms: temperate versus tropical species. *Am Nat* 114: 784–795.

Liu B, Wendel JF. 2000. Retrotransposon activation followed by rapid repression in introgressed rice plants. *Genome* 43: 874–880.

Lyndon RF. 1963. Changes in the nucleus during cellular development in the pea seedling. *J Exp Bot* 14: 419–430.

Ma J, Devos KM, Bennetzen J. 2004. Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. *Genome Res* 14: 860–869.

MacGillivray CW, Grime JP. 1995. Genome size predicts frost resistance in British herbaceous plants: implications for rates of vegetation response to global warming. *Funct Ecol* 9: 320–325.

Marie D, Brown SC. 1993. A cytometric exercise in plant DNA histograms, with 2C values for 70 species. *Biol Cell* 78: 41–51.

Martin PG. 1966. Variation in amounts of nucleic acids in cells of different species of higher plants. *Exp Cell Res* 44: 84–94.

Martin PG, Shanks R. 1966. Does *Vicia faba* have multi-stranded chromosomes? *Nature* 211: 650–651.

Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in a majority of angiosperms. *Science* 264: 421–424.

McLeish J, Sunderland N. 1961. Measurements of deoxyribosenucleic acid (DNA) in higher plants by Feulgen photometry and chemical methods. *Exp Cell Res* 24: 527–540.

McNeill J. 1980. *Scilla* L. In: Tutin TG, Heywood VH, Burges NA, *et al.* eds. *Flora Europaea*. Cambridge: Cambridge University Press, 41–43.

Michaelson MJ, Price HJ, Ellison JR, Johnston JS. 1991. Comparison of plant DNA contents determined by Feulgen microspectrophotometry and laser flow cytometry. *Am J Bot* 78: 183–188.

Midgley JJ, Bond WJ. 1991. Ecological aspects of the rise of angiosperms: a challenge to the reproductive superiority hypotheses. *Biol J Linn Soc* 44: 81–92.

Miksche JP. 1968. Quantitative study of intraspecific variation of DNA per cell in *Picea glauca* and *Pinus banksiana*. *Can J Genet Cytol* 10: 590–600.

Miksche JP. 1971. Intraspecifc variation of DNA per cell between *Picea sitchensis* (Bong.) Carr. provenances. *Chromosoma* 32: 343–352.

Mirsky AE, Ris H. 1949. Variable and constant components of chromosomes. *Nature* 163: 666–667.

Moldavan A. 1934. Photo-electric technique for the counting of microscopical cells. *Science* 80: 188–189.

Moore G, Devos KM, Wang Z, Gale MD. 1995. Cereal genome evolution: grasses, line up and form a circle. *Curr Biol* 5: 737–739.

Morgante M, Hanafey M, Powell W. 2002. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30: 194–200.

Mowforth MAG. 1985. Variation in nuclear DNA amounts in flowering plants: an ecological analysis. Ph.D. thesis, United Kingdom, University of Sheffield.

Murray BG. 1998. Nuclear DNA amounts in gymnosperms. *Ann Bot* 82 (Suppl. A): 3–15.

Murray BG, Leitch IJ, Bennett MD. 2001. *Gymnosperm DNA C-values Database* (Release 1.0). www.rbgkew.org.uk/cval/homepage.html.

Naranjo CA, Ferrari MR, Palermo AM, Poggio L. 1998. Karyotype, DNA content and meiotic behaviour in five South American species of *Vicia* (Fabaceae). *Ann Bot* 82 (Suppl. A): 757–764.

Noirot M, Barre P, Duperray C, *et al*. 2003. Effects of caffeine and chlorogenic acid on propidium iodide accessibility to DNA: consequences on genome size evaluation in coffee tree. *Ann Bot* 92: 259–264.

Noirot M, Barre P, Duperray C, *et al*. 2005. Investigation on the causes of stoichiometric error in genome size estimation using heat experiments. Consequences on data interpretation. *Ann Bot* (in press).

Noirot M, Barre P, Louarn J, *et al*. 2000. Nucleus-cytosol interactions: a source of stoichiometric error in flow cytometric estimation of nuclear DNA content in plants. *Ann Bot* 86: 309–316.

Noirot M, Barre P, Louarn J, *et al*. 2002. Consequences of stoichiometric error on nuclear DNA content evaluation in *Coffea liberica* var. *dewevrei* using DAPI and propidium iodide. *Ann Bot* 89: 385–389.

Obermayer R, Greilhuber J. 2000. Genome size in *Hedera helix* L.: a clarification. *Caryologia* 53: 1–4.

Obermayer R, Leitch IJ, Hanson L, Bennett MD. 2002. Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Ann Bot* 90: 209–217.

Ogur M, Erickson RO, Rosen GU, *et al*. 1951. Nucleic acids in relation to cell division in *Lilium longiflorum*. *Exp Cell Res* 2: 73–89.

Ohri D, Khoshoo TN. 1986. Genome size in gymnosperms. *Plant Syst Evol* 153: 119–132.

Orel N, Puchta H. 2003. Differences in the processing of DNA ends in *Arabidopsis thaliana* and tobacco: possible implications for genome evolution. *Plant Mol Biol* 51: 523–531.

Ozkan H, Tuna M, Arumuganathan K. 2003. Nonadditive changes in genome size during allopolyploidization in the wheat (*Aegilops-Triticum*) group. *J Hered* 94: 260–264.

Paroda RS, Rees H. 1971. Nuclear DNA variation in eu-sorghums. *Chromosoma* 32: 353–363.

Pegington C, Rees H. 1970. Chromosome weights and measures in the Triticinae. *Heredity* 25: 195–205.

Petrov DA. 2001. Evolution of genome size: new approaches to an old problem. *Trends Genet* 17: 23–28.

Pires JC, Lim KY, Kovarik A, *et al.* 2004. Molecular cytogenetic analysis of recently evolved *Tragopogon* (Asteraceae) allopolyploids reveals a karyotype that is additive of the diploid progenitors. *Am J Bot* 91: 1022–1035.

Poggio L, Rosato M, Chiavarino AM, Naranjo CA. 1998. Genome size and environmental correlations in maize (*Zea mays* ssp. *mays*, Poaceae). *Ann Bot* 82 (Suppl. A): 107–117.

Price HJ, Johnston JS. 1996. Influence of light on DNA content of *Helianthus annuus* Linnaeus. *Proc Natl Acad Sci USA* 93: 11264–11267.

Price HJ, Chambers KL, Bachmann K. 1981a. Genome size variation in diploid *Microseris bigelovii* (Asteraceae). *Bot Gaz* 142: 156–159.

Price HJ, Chambers KL, Bachmann K. 1981b. Geographic and ecological distribution of genomic DNA content variation in *Microseris douglasii* (Asteraceae). *Bot Gaz* 142: 415–426.

Price HJ, Hodnett G, Johnston JS. 2000. Sunflower (*Helianthus annuus*) leaves contain compounds that reduce nuclear propidium iodide fluorescence. *Ann Bot* 86: 929–934.

Prokopowich CD, Gregory TR, Crease TJ. 2003. The correlation between rDNA copy number and genome size in eukaryotes. *Genome* 46: 48–50.

Puech M, Giroud F. 1999. Standardisation of DNA quantitation by image analysis: quality control of instrumentation. *Cytometry* 36: 11–17.

Puite KJ, Tenbroeke WRR. 1983. DNA staining of fixed and non-fixed plant protoplasts for flow cytometry with Hoechst 33342. *Plant Sci Lett* 32: 79–88.

Qiu YL, Palmer JD. 1999. Phylogeny of early land plants: insights from genes and genomes. *Trends Plant Sci* 4: 26–30.

Rabinowicz PD. 2000. Are obese plant genomes on a diet? *Genome Res* 10: 893–894.

Rasch E, Woodard JW. 1959. Basic proteins of plant nuclei during normal and pathological cell growth. *J Biophys Biochem Cytol* 6: 263–276.

Rayburn AL, Auger JA. 1990. Genome size variation in *Zea mays* ssp. *mays* adapted to different altitudes. *Theor Appl Genet* 79: 470–474.
Rayburn AL, Price HJ, Smith JD, Gold JR. 1985. C-band heterochromatin and DNA content in *Zea mays*. *Am J Bot* 72: 1610–1617.
Reed SM. 1991. Cytogenetic evolution and aneuploidy in *Nicotiana*. In: Tsuchiya T, Gupta PK eds. *Chromosome Engineering in Plants: Genetics, Breeding, Evolution (Part B)*. Amsterdam: Elsevier Inc., 483–505.
Rees H, Hazarika MH. 1969. Chromosome evolution in *Lathyrus*. *Chromosomes Today* 2: 158–165.
Rees H, Narayan RKJ. 1988. Chromosome constraints: chiasma frequency and genome size. In: Brandham PE ed. *Kew Chromosome Conference III*. London: HMSO, 231–239.
Rees H, Cameron FM, Hazarika MH, Jones GH. 1966. Nuclear variation between diploid angiosperms. *Nature* 211: 828–830.
Reeves G, Francis D, Davies MS, *et al*. 1998. Genome size is negatively correlated with altitude in natural populations of *Dactylis glomerata*. *Ann Bot* 82: 99–105.
Rejmanek M. 1996. A theory of seed plant invasiveness: the first sketch. *Biol Conserv* 78: 171–181.
Resslar PM, Stucky JM, Miksche JP. 1981. Cytophotometric determination of the amount of DNA in *Arachis* L. sect. *Arachis* (Leguminosae). *Am J Bot* 68: 149–153.
Ricroch A, Brown SC. 1997. DNA base composition of *Allium* genomes with different chromosome numbers. *Gene* 205: 255–260.
Rothfels K, Sexsmith E, Heimburg M, Krause MO. 1966. Chromosome size and DNA content of species of *Anemone* L. and related genera (Ranunculaceae). *Chromosoma* 20: 54–74.
Sakamura T. 1918. Kurze Mitteilung über die Chromosomenzahlen und die Verwandtschaftsverhältnisse der *Triticum*-Arten. *Bot Mag* 32: 149–153.
SanMiguel P, Bennetzen JL. 1998. Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann Bot* 82 (Suppl. A): 37–44.
SanMiguel P, Gaut BS, Tikhonov A, *et al*. 1998. The paleontology of intergene retrotransposons of maize. *Nat Genet* 20: 43–45.
SanMiguel P, Tikhonov A, Jin YK, *et al.* 1996. Nested retrotransposons in the intergenic regions of the maize genome. *Science* 274: 765–768.
Schmidt G, Thannhauser SJ. 1945. A method for the determination of desoxyribonucleic acid, ribonucleic acid and phosphoproteins in animal tissues. *J Biol Chem* 161: 83–89.
Scott LJ, Cross M, Shepherd M, *et al*. 1999. Increasing the efficiency of microsatellite discovery from poorly enriched libraries in coniferous forest species. *Plant Mol Biol Report* 17: 351–354.
Seal AG. 1983. The distribution and consequences of changes in nuclear DNA content. In: Brandham PE, Bennett MD eds. *Kew Chromosome Conference II*. London: George Allen & Unwin, 225–232.
Seal AG, Rees H. 1982. The distribution of quantitative DNA changes associated with the evolution of diploid Festuceae. *Heredity* 49: 179–190.
Shaked H, Kashkush K, Ozkan H, *et al*. 2001. Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. *Plant Cell* 13: 1749–1759.
Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Res* 10: 908–915.
Sims LE, Price HJ. 1985. Nuclear DNA content variation in *Helianthus* (Asteraceae). *Am J Bot* 72: 1213–1219.
Singh KP, S.N.R, Singh AK. 1996. Variation in chromosomal DNA associated with the evolution of *Arachis* species. *Genome* 39: 890–897.
Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003. Evolution of genome size in the angiosperms. *Am J Bot* 90: 1596–1603.
Sparrow AH, Miksche JP. 1961. Correlation of nuclear volume and DNA content with higher plant tolerance to chronic radiation. *Science* 134: 282–283.

Sparrow AH, Nauman AF. 1976. Evolution of genome size by DNA doublings. *Science* 192: 524–529.

Speta F. 1980. Die frühjahrsblühenden *Scilla*-Arten des östlichen Mittelmeerraumes. *Naturkundl JahrB Stadt Linz* 25 (1979): 19–198.

Srivastava S, Lavania UC. 1991. Evolutionary DNA variation in *Papaver*. *Genome* 34: 763–768.

Sunderland N, McLeish J. 1961. Nucleic acid content and concentration in root cells of higher plants. *Exp Cell Res* 24: 541–554.

Suoniemi A, Tanskanen J, Schulman AH. 1998. *Gypsy*-like retrotransposons are widespread in the plant kingdom. *Plant J* 13: 699–705.

Swift H. 1950a. The desoxyribose nucleic acid content of animal nuclei. *Physiol Zool* 23: 169–198.

Swift H. 1950b. The constancy of desoxyribose nucleic acid in plant nuclei. *Proc Natl Acad Sci USA* 36: 643–654.

Temsch EM, Greilhuber J. 2001. Genome size in *Arachis duranensis*: a critical study. *Genome* 44: 826–830.

Temsch EM, Greilhuber J, Krisai R. 1998. Genome size in *Sphagnum* (peat moss). *Bot Acta* 111: 325–330.

Teoh SB, Rees H. 1976. Nuclear DNA amounts in populations of *Picea* and *Pinus* species. *Heredity* 36: 123–137.

Thomas CA. 1971. The genetic organization of chromosomes. *Annu Rev Genet* 5: 237–256.

Thompson K. 1990. Genome size, seed size and germination temperature in herbaceous angiosperms. *Evol Trends Plants* 4: 113–116.

Tito CM, Poggio L, Naranjo CA. 1991. Cytogenetic studies in the genus *Zea*. 3. DNA content and heterochromatin in species and hybrids. *Theor Appl Genet* 83: 58–64.

Tjio JH, Levan A. 1956. The chromosomes of man. *Hereditas* 42: 1–6.

Underbrink AG, Pond V. 1976. Cytological factors and their predictive role in comparative radiosensitivity. *Curr Topics Radiat Res Quart* 11: 251–306.

Uozo S, Ikehashi H, Ohmido N, *et al*. 1997. Repetitive sequences: causes for variation in genome size and chromosome morphology in the genus *Oryza*. *Plant Mol Biol* 35: 791–799.

Van't Hof J. 1965. Relationships between mitotic cycle duration, S period duration and average rate of DNA synthesis in root meristem cells of several plants. *Exp Cell Res* 39: 48–58.

Van't Hof J. 1974. The duration of chromosomal DNA synthesis, of the mitotic cycle, and of meiosis of higher plants. In: King RC ed. *Handbook of Genetics, vol. 2, Plants, Plant Viruses and Protists.* New York: Plenum Press, 181–200.

Van't Hof J, Sparrow AH. 1963. A relationship between DNA content, nuclear volume, and minimum mitotic cycle time. *Proc Natl Acad Sci USA* 49: 897–902.

Vendrely R, Vendrely C. 1948. La teneur du noyau cellulaire en acide désoxyribonucléique à travers les organes, les individus et les espèces animales: Techniques et premiers résultats. *Experientia* 4: 434–436.

Verma SC, Rees H. 1974. Nuclear DNA and evolution of allotetraploid Brassicae. *Heredity* 33: 61–68.

Vicient CM, Schulman A. 2002. *Copia*-like retrotransposons in the rice genome: few and assorted. *Genome Lett* 1: 35–47.

Vicient CM, Kalendar R, Anamthawat-Jonsson K, *et al*. 1999a. Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. *Genetica* 107: 53–63.

Vicient CM, Suoniemi A, Anamthawat-Jonsson K, *et al*. 1999b. Retrotransposon BARE-1 and its role in genome evolution in the genus *Hordeum*. *Plant Cell* 11: 1769–1784.

Vidic T, Greilhuber J, Vilhar B. 2003. Genome size is associated with differential survival of plant species. *Abstracts of the Second Plant Genome Size Meeting held at The Royal Botanic Gardens, Kew, September 2003*.

Vilhar B, Dermastia M. 2002. Standardization of instrumentation in plant DNA image cytometry. *Acta Bot Croat* 61: 11–25.

Vilhar B, Greilhuber J, Koce JD, *et al*. 2001. Plant genome size measurement with DNA image cytometry. *Ann Bot* 87: 719–728.

Vinogradov AE. 2003. Selfish DNA is maladaptive: evidence from the plant Red List. *Trends Genet* 19: 609–614.

Vitte C, Panaud O. 2003. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol Biol Evol* 20: 528–540.

Voglmayr H. 2000. Nuclear DNA amounts in mosses (Musci). *Ann Bot* 85: 531–546.

Voytas DF, Cummings MP, Konieczny A, *et al*. 1992. *Copia*-like retrotransposons are ubiquitous among plants. *Proc Natl Acad Sci USA* 89: 7124–7128.

Wakamiya I, Newton RJ, Johnston JS, Price HJ. 1993. Genome size and environmental factors in the genus *Pinus*. *Am J Bot* 80: 1235–1241.

Watanabe K, Yahara T, Denda T, Kosuge K. 1999. Chromosomal evolution in the genus *Brachyscome* (Asteraceae, Astereae): statistical tests regarding correlation between changes in karyotype and habit using phylogenetic information. *J Plant Res* 112: 145–161.

Wendel JF. 2000. Genome evolution in polyploids. *Plant Mol Biol* 42: 225–249.

Wendel JF, Wessler SR. 2000. Retrotransposon-mediated genome evolution on a local ecological scale. *Proc Natl Acad Sci USA* 97: 6250–6252.

Wendel JF, Cronn RC, Johnston JS, Price HJ. 2002. Feast and famine in plant genomes. *Genetica* 115: 37–47.

Westoby M, Jurado E, Leishman M. 1992. Comparative evolutionary ecology of seed size. *Trends Ecol Evol* 7: 368–372.

White J, Rees H. 1987. Chromosome weights and measures in *Petunia*. *Heredity* 58: 139–143.

Wicker T, Stein N, Albar L, *et al*. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. *Plant J* 26: 307–316.

Wicker T, Yahiaoui N, Guyot R, *et al*. 2003. Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A(m) genomes of wheat. *Plant Cell* 15: 1186–1197.

Yokoya K, Roberts AV, Mottley J, *et al*. 2000. Nuclear DNA amounts in roses. *Ann Bot* 85: 557–562.

Zakirowa RO, Vakhtina LI. 1974. Cytophotometrical and caryological investigation of some species of the genus *Allium* subgenus *Melanocrommyum* (Webb et Berth.) Wendelbo, Sect. *Melanocrommyum*. *Bot J USSR* 12: 1819–1827.

Zoldos V, Papes D, Brown SC, *et al*. 1998. Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. *Genome* 41: 162–168.

PART II

The Evolution of Genomic Parasites

CHAPTER 3

Transposable Elements

MARGARET G. KIDWELL

Transposable elements (TEs) are discrete DNA sequences that move from one location to another within the genome. They are found in nearly all species that have been studied and constitute a large fraction of some genomes, including that of *Homo sapiens*. TEs are potent broad-spectrum mutator elements that are responsible for generating variation in the host genome and have a role as key players in the ecology of the genome. This chapter presents an overview that includes coverage of TE structures, regulation, distribution, and dynamics. A wealth of examples provides many illustrations of the diversity of TE types and behaviors as well as the rich variety of interactions between TEs and their host genomes. It is evident from this that knowledge of these elements is essential for a full understanding of genome evolution.

A BRIEF HISTORY OF THE STUDY OF TRANSPOSABLE ELEMENTS

Transposable elements comprise a group of distinct DNA segments with the capacity to move, or transpose, between many nonhomologous (unrelated) sites

The Evolution of the Genome, edited by TR Gregory
Copyright 2005, Elsevier Inc. All rights reserved.