The C-value Enigma in Plants and Animals: A Review of Parallels and an Appeal for Partnership

T. RYAN GREGORY*

Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

Received: 5 November 2003 Returned for revision: 18 November 2003 Accepted: 25 November 2003

Aim: Plants and animals represent the first two kingdoms recognized, and remain the two best-studied groups in terms of nuclear DNA content variation. Unfortunately, the traditional chasm between botanists and zoologists has done much to prevent an integrated approach to resolving the C-value enigma, the long-standing puzzle surrounding the evolution of genome size. This grand division is both unnecessary and counterproductive, and the present review aims to illustrate the numerous links between the patterns and processes found in plants and animals so that a stronger unity can be developed in the future.

Scope: This review discusses the numerous parallels that exist in genome size evolution between plants and animals, including (i) the construction of large databases, (ii) the patterns of DNA content variation among taxa, (iii) the cytological, morphological, physiological and evolutionary impacts of genome size, (iv) the mechanisms by which genomes change in size, and (v) the development of new methodologies for estimating DNA contents.

Conclusions: The fundamental questions of the C-value enigma clearly transcend taxonomic boundaries, and increased communication is therefore urged among those who study genome size evolution, whether in plants, animals or other organisms.

Key words: Body size, cell size, complexity, C-value enigma, C-value paradox, database, development, DNA content, genome size, introns, physiology, selfish DNA, sequencing, transposable elements.

INTRODUCTION

In the beginning (of taxonomy), there were two kingdoms of life: green things were plants and moving things were animals. Today, thanks in no small part to comparative genomic analyses, four kingdoms of eukaryotes are commonly recognized (Animalia, Plantae, Fungi, Protista), all of them contained in the Eukarya, which many authors argue is just one of three ‘empires’ (along with Bacteria and Archaea). The process of splitting is still not complete, given that ‘protists’ probably comprise two kingdoms of their own (Protozoa and Chromista) while also having representatives in the others (Cavalier-Smith, 1998). To be sure, an acknowledgement of the deeply branching diversity of living things has been a great achievement in evolutionary biology, but somehow the equally important underlying unity—most notably, the simple fact that everything alive has a genome which can be compared—has been forgotten. It is a mere truism that for every genome there is a size, defined by either mass (in picograms, pg) or number of base pairs (bp). It is far from a given, however, that these sizes should be (mostly) constant species-specific characters, that they should vary over several orders of magnitude among eukaryotes, or that they should bear no connection to organismal complexity (Gregory, 2001a). In fact, each of these features came as a major surprise to early researchers. Thus, Vendrely and Vendrely (1948) were struck by ‘a remarkable constancy in the nuclear DNA content of all the cells in all the individuals within a given animal species’ (my translation), and Comings (1972) lamented that ‘the lowly lillywort has 18 times as much DNA as we, and the slimy, dull salamander known as Amphiuma has 26 times our complement of DNA’.

The C-value enigma: a cross-kingdom puzzle

As part of a defence of the Vendrelys’ ‘DNA constancy hypothesis’, Hewson Swift (1950a, b) studied DNA contents in different tissues of both animals (frog, mouse and grasshopper) and plants (Tradescantia and Zea) and developed the concept of the ‘C-value’ in reference to the haploid, or 1C, DNA amount. (In diploid organisms, including the vast majority of animals but probably a minority of plants, ‘genome size’ and ‘C-value’ will be identical; in recent polyploids, the situation is more complex because the C-value will comprise more than one genome.) DNA constancy, upon which the C-value concept was based, was in turn taken as evidence that DNA, and not the highly variable proteins, serves as the hereditary material (e.g. Swift, 1950a).

The expression of the combination of demonstrable DNA constancy (within species) and profound genome size variation (among species) as a ‘C-value paradox’ (Thomas, 1971) is easy to comprehend. However defined—as ‘simple’ organisms having more DNA than ‘complex’ ones, in terms of some closely related species displaying highly divergent DNA contents, or by noting that any given organism contains more DNA than would be expected based on its presumed number of genes—the basic ‘paradox’ was that DNA amount is constant because it is the stuff of genes, and yet is unrelated to expected gene number.

The solution to the paradox is now well known: most DNA is non-coding, so the size of a genome need not imply...
anything at all about the number of genes it contains. The term ‘C-value paradox’ persists, but only by virtue of historical entrenchment. Certainly, the discovery of non-coding DNA, which ended the paradox, raised a number of questions of its own. Whence this non-coding DNA? Does it have any phenotypic effects (or even functions)? How is it gained and lost? What are the patterns of its distribution among taxa? Why do some species contain so much of it and others so little?

Because a ‘paradox’ begs a one-dimensional solution, there is a tendency to deal with only one of these questions and to present an answer as ‘the’ solution to the problem of genome size evolution. A much more productive approach is to recognize these issues, taken together, as components of a complex puzzle—the ‘C-value enigma’ (Gregory, 2001a, 2005). Importantly, this distinction makes it immediately clear that the enigma applies to all eukaryotes, but that the particulars will vary according to the biology of the groups in question.

Most of the existing genome size information comes from plants and animals, and much progress has been made in exploring the C-value enigma’s various questions. Unfortunately, this work has often proceeded in parallel, with very limited interaction, by botanists on one side of the classic divide and zoologists on the other. However, and as Comings’s (1972) disdain for liverwort genomes clearly attests, the puzzle of genome size variation (now past its 50th year in existence) clearly transcends these taxonomic boundaries. The purpose of this paper is to further the construction of bridges across the original taxonomic chasm which still divides departments, literature and, too often, communication.

THE STATE OF KNOWLEDGE REGARDING PLANT AND ANIMAL GENOME SIZES

Plant DNA C-values Database

One of the first broad comparisons of available genome size data was made three decades ago by Sparrow et al. (1972), and included values from plants, animals, fungi and prokaryotes. Around this time, a large number of additional C-values began to be measured and compiled in order to compare DNA content with features of practical interest, such as cell and life cycle duration and geographic distribution. As data began to accumulate, it was recognized that additional published lists could be of great value, and a few years later Bennett and Smith (1976) presented the first in a series of plant DNA C-value compilations that extends to the present day (Bennett et al., 1982; Bennett and Smith, 1991; Bennett and Leitch, 1995, 1997, 2001; Murray, 1998; Bennett et al., 2000a; Voglmayr, 2000).

The first electronic version of the Angiosperm DNA C-values Database was made available in April 1997, in anticipation of the plant genome size meetings at Kew in September of that year. Since 2001, it has been presented as an expanded Plant DNA C-values Database that includes all the major groups of land plants. Release 2.0 of the Plant DNA C-values Database, launched in January of 2003, currently includes data from nearly 4000 species consisting of 3493 angiosperms, 181 gymnosperms, 63 monilophytes (members of the horsetail-fern clade; see Pryer et al., 2001), and 171 byrophytes. This covers about 1.5% of known land plants, and in gymnosperms in particular the coverage is nearly 25% (Bennett and Leitch, 2003). Plans are also under way to add more than 200 values for algae (see Kapraun, 2005) within the next year.

The first of ten key recommendations made during the 1997 Kew meeting was to increase the coverage of angiosperms by a further 1% and to obtain estimates from at least one representative of each family. Following the second Kew meeting in September 2003, revised targets of 75% familial, 10% generic and an additional 1% species coverage for angiosperms, a level of 2% species representation for pteridophytes, and improved geographic sampling for bryophytes were proposed. This will involve a considerable number of new estimates, but is expected to be completed within the next 5 years. In addition, groups of specific taxonomic and/or biological interest may be targeted to allow detailed comparative study.

Animal Genome Size Database

Unlike the situation with plants, the initial effort of Sparrow et al. (1972) to compile animal genome sizes had not been followed up in any comprehensive way until quite recently. As such, much of the work on animal genome size evolution has necessarily followed in the footsteps of botanists. Like its botanical predecessor, the Animal Genome Size Database began as part of an investigation of the patterns and phenotypic implications of genome size variation. In particular, the initial animal compilation was made for a study of the relationship between genome size and red blood cell size in mammals (Gregory, 2000). This was subsequently expanded to cover birds (Gregory, 2002a), and eventually grew to include all animals. The database was launched in January of 2001, and as of October 2004 contains data for more than 3700 animals, including roughly 2470 vertebrates and 1260 invertebrates (Gregory, 2001b). Obviously, this coverage is highly biased towards the 50 000 or so species of vertebrates, given that it includes 20% of jawless fishes, about 12% of cartilaginous fishes, 7% each of amphibians and mammals, roughly 4% of ray-finned fishes and reptiles, and nearly 2% of birds, but an abysmally tiny percentage of invertebrates (which undoubtedly total in the millions). First-time assays of major invertebrate groups over the past few years have begun to lessen this discrepancy, but in truth have only scratched the surface (Gregory, 2005).

In absolute terms, the plant and animal databases are fairly similar in size, but clearly the relative coverage is far superior in plants (in part simply because there are already five times as many described animal species as plants). There are currently no plans to gather data from 1% of animals, which would require an enormous effort—another 2500 estimates will be required just to get 1% of beetles, counting only described species. For the foreseeable future, the primary goal with the animal dataset will be simply to fill in some of the more glaring gaps, including
several classes (or even phyla!) that remain largely or entirely unknown, as well as numerous orders and/or families of insects, mammals, fishes and birds that have not yet been studied (as a quick reminder, the zoological taxonomic hierarchy is: Kingdom, Phylum, Class, Order, Family, Genus, Species). Of course, the identification of these gaps itself represents a major step forward that was not possible prior to the assembly of an animal database comparable to that for plants.

PATTERNS OF VARIATION IN PLANTS AND ANIMALS

Interspecific variation

The angiosperms encompass the entire range of C-values found among land plants, in total around a 1000-fold variation. Other groups of land plants vary considerably less: the monilophytes about 95-fold, lycophytes around 75-fold, gymnosperms roughly 14-fold and bryophytes only 12-fold (Bennett and Leitch, 2003). Taking the Chlorophyta, Phaeophyta and Rhodophyta together, algae display a level of variation exceeding 1300-fold, although within any of these groups the range is between only 9- and 200-fold (Kapraun, 2005). From the smallest alga to the largest angiosperm, plants as a whole range nearly 8500-fold in their C-values.

Simple ranges can be somewhat misleading, however, since the majority of plants in all of these groups have relatively small genomes. Thus, the modal genome sizes in all but the gymnosperms (~10 pg) and monilophytes (~8 pg) are 0-6 pg or less, including for the hypervariable angiosperms and algae. This tendency for small genome sizes is also apparent in Fig. 1, which shows the C-value ranges and means for the major groups of plants and animals so far studied. With the exception of the gymnosperms, in all cases the mean is near the bottom end of the overall range.

Interspecific variation is higher in animals than in the land plants, with an overall range of about 3300-fold (Gregory, 2001b). The smallest animal genome size (0.104 pg) is found in the placozoan Trichoplax adhaerens, which, being constructed of only four cell types and essentially resembling a giant ciliated amoeba, is also by far the simplest member of the kingdom. The largest animal genome size so far reported (~132 pg) is that of the marbled lungfish, Protopterus aethiopicus. The tunicate Oikopleura dioica has a genome size of about 0.07 pg, making the range in chordates around 1800-fold, and several pufferfishes of the family Tetraodontidae exhibit C-values around 0.4 pg, for a vertebrate range of roughly 330-fold. Thus, even the vertebrates alone are considerably more variable than any one group of plants besides angiosperms. Ranges among some invertebrate groups may also approach this level, as with flatworms (340-fold), crustaceans (240-fold) and insects (190-fold), but in many cases are considerably smaller, as among annelids (125-fold), arachnids (70-fold), nematodes (40-fold), molluscs (15-fold) and echinoderms (9-fold).

The general pattern among animals, as with plants, is for most members of each major group to be rather constrained in their genome size variation, with only one or a few subset(s) exhibiting large genomes (Fig. 1). In plants, certain
Quantum shifts in genome size

In 1976, Sparrow and Nauman suggested that the minimum genome sizes of groups as wide-ranging as viruses, bacteria, plants, fungi and animals varied discontinuously by following a doubling series within and among taxa. Since this apparent series of multiples did not correspond to differences in chromosome numbers, they considered this to represent a process of ‘cryptopolyploidy’ (as they put it, ‘polyploidy results in more chromosomes; cryptopolyploidy results in larger chromosomes’; Sparrow and Nauman, 1976). Overall, this pattern is very rough, and since it applies only to minimum genome sizes is of limited interest to the C-value enigma.

In a less expansive (and more realistic) context, quantum shifts in genome size have been reported within numerous genera of plants (see Sparrow and Nauman, 1973; Narayan, 1985, 1988, 1998) and also in algae (Maszewski and Kołodziejczyk, 1991). In these cases, it is not minimal genome size across broad groups that varies by a series of doublings, but rather the C-values of congeneric species that differ by multiples of the lowest genome in the group. For example, in a sample of 20 species of the plant genus *Tephrosia*, Raina et al. (1986) found genome sizes to vary from 1.3 to 7.4 pg by increments of about 0.74 pg (approximately half the value of the smallest genome in the group). In some cases, these discontinuous patterns have been explained as an artifact of incomplete taxonomic sampling (e.g. Ohri et al., 1998; Greilhuber and Obermayer, 1999), but it would require a remarkably fortuitous series of collections to account for all the known examples. Other authors have ascribed genome size discontinuity to the existence of ‘steady states’ in DNA content maintained by stabilizing selection (Narayan, 1998), but the question remains as to why the steady states should be found at exact multiples.

Similar patterns of discontinuous variation have been reported for groups of animals as diverse as aphids (Finston et al., 1995), polychaete annelids (Sella et al., 1993; Gambi et al., 1997) and turbellarian flatworms (Gregory et al., 2000). In one of the most striking animal examples, genome sizes in copepod crustaceans of the genera *Calanus* and *Pseudocalanus* vary by intervals of about 2 pg, from 2.25 to 12.5 pg (McLaren et al., 1988, 1989), and similar patterns may hold in other copepods as well (Gregory et al., 2000). Certainly, quantum shifts are not the dominant mode of change in either plants or animals, but are nevertheless sufficiently common in both groups to be of special interest.

Intraspecific variation: real and artificial

Even if it proceeds by less than quantum jumps, genome size change may still be restricted in its occurrence to speciation events (i.e. punctuational). However, it is also conceivable that much change occurs in a truly gradualistic (anagenetic) mode, beginning with differences among con-specifics and ending with differences across species. Pronounced intraspecific variation would, of course, pose a major challenge to the DNA constancy hypothesis upon which the C-value concept and all indirect DNA quantification methods (see below) are based, and is therefore of both theoretical and pragmatic importance in genome size study.

The genomes of plants have frequently been labelled as ‘fluid’, ‘dynamic’, and ‘in constant flux’, due in large part to the seemingly common observation of pronounced intraspecific variation in their nuclear DNA contents. In some cases, real variation within species can be explained by the differential presence of supernumerary B chromosomes. Strictly speaking, this does not refute the notion of DNA constancy because the A chromosome complement remains unchanged. In other examples, however, intraspecific variation in DNA content can be attributed to recognizable polymorphisms in the A chromosomes themselves, as with heterochromatic knobs in maize (e.g. Poggio et al., 1998) or differentially deleted transposable element remnants in barley (Kalendar et al., 2000).

Over the past few years, it has become necessary to abandon many of the most celebrated examples of intraspecific variation in plants as they have been attributed to experimental error (e.g. Greilhuber, 1988, 1997, 1998, 2005; Greilhuber and Obermayer, 1998; Bennett and Leitch, 2005). Even the most careful study can be subject to unanticipated sources of error, as illustrated by the case of apparent environmentally induced variation in DNA content in the sunflower, *Helianthus annuus*. Although it initially appeared that differences in light exposure could alter DNA content (Price and Johnston, 1996), it was later realized that sunflowers generate compounds in the presence of light which interfere with propidium iodide staining and therefore give a false impression of DNA content variation (Price et al., 2000). Similar effects have since been observed in coffee (Noirot et al., 2002), and the presence of stain inhibitors has become a major concern for DNA estimation in plants (see below). In other cases, extensive sampling has revealed striking stability in plant genome sizes. Most notably, it has recently been demonstrated that populations of the onion (*Allium cepa*), a species often used as a standard in plant studies, maintain a constant genome size across four continents (Bennett et al., 2000b).

Fluidity has also been attributed to the genomes of salamanders, although more because of their large size than...
any detailed evidence of dynamic behaviour (Vignali and Nardi, 1996). Examples of intraspecific variation in genome size have been reported for many animal groups, including insects (e.g. Kumar and Rai, 1990), mammals (e.g. Garagna et al., 1999), copepod crustaceans (Escribano et al., 1992), fishes (e.g. Johnson et al., 1987; Lockwood and Bickham, 1991, 1992; Lockwood and Derr, 1992; Collares-Pereira and Moreira da Costa, 1999), molluscs (Rodriguez-Juiz et al., 1991) and reptiles and amphibians (Lockwood et al., 1991; MacCulloch et al., 1996). Specimens of fish and amphibians taken from radioactively contaminated or otherwise polluted areas also show some apparent fluctuations in DNA content (Lingenfelser et al., 1997; Dallas et al., 1998; Vinogradov and Chubinishvili, 1999).

These examples, although interesting, do not provide conclusive evidence of serious violations of DNA constancy. Some of these cases could be based on the presence of cryptic subspecies (Lockwood and Bickham, 1992; MacCulloch et al., 1996); in most instances, the variation in genome size is associated with differences in geography, either at the inter-populational level, or at least along some geographic cline. This issue of inadequate species delineation has also been pointed out for some apparent cases of intraspecific variation in plants (e.g. Ebert et al., 1996). The possibility of experimental error is also ever present in animals, as with plants. For example, Thinwala et al. (1994) reported that specimens of the aphid Schizaphis graminum raised on sorghum had lower DNA contents than individuals from the same biotypes reared on wheat or johnsongrass, but the fact that the animals were simply homogenized prior to flow cytometric analysis raises the possibility that some botanical compound in their food confused staining with propidium iodide.

It must also be noted that most of the examples of intraspecific variation in fishes involve either salmonids or cyprinids, which are well known to exhibit more dynamic chromosome-level changes than most teleost groups. For example, differences in chromosome numbers can be observed among populations of rainbow trout (Thorgaard, 1983). Thus, these families may be atypical in this regard (Johnson et al., 1987; Lockwood and Bickham, 1991), making them of special interest in their own right, but not generally indicative of the situation in fishes (let alone animals or all eukaryotes, as some authors suggest). Perhaps tellingly, a detailed sampling from various wild and domesticated stocks of the channel catfish, Ictalurus punctatus (family Ictaluridae), revealed a high level of genome size stability (Tiersch et al., 1990).

In some groups of animals, it is entirely unclear whether genomes are remarkably flexible or highly stable. For example, Escribano et al. (1992) found significant variation within the copepod crustaceans Calanus glacialis and Pseudocalanus acuspes based on geographical differences, between P. elongatus reared for 96 generations in the laboratory versus wild-caught specimens (and also according to season in one year, but not in others), and within these various species according to differences in rearing conditions including food availability and temperature. On the other hand, no disparity in C-values was found among populations of the cyclopoid copepod Mesocyclops edax collected along the east coast of North America from Nova Scotia to Florida (Wyngaard and Rasch, 2000). In this instance it is worthy of note that the cyclopoid copepods, which apparently have very stable C-values intraspecifically, do not vary greatly among species, whereas the calanoids, which may be more flexible within species, also exhibit marked interspecific variability (Wyngaard and Rasch, 2000).

In any event, B chromosomes are known from many animals (Camacho et al., 2000), and groups with chromosomal sex determination systems are expected to evince differences among males and females in absolute DNA content, although this variation is usually (but not always) minor. Heterochromatin polymorphisms in sex chromosomes may also provide some real intraspecific variation in certain cases (e.g. Garagna et al., 1999). It therefore seems prudent to maintain a certain level of agnosticism regarding the extent of intraspecific variation in genome size for both plants and animals. Obviously, DNA contents must change by some mechanism(s), whether strictly gradualistic or more punctuated. As with most important mechanistic questions in evolutionary biology, this is probably an issue of relative frequencies, not exclusive absolutes.

**MECHANISMS OF GENOME SIZE CHANGE IN PLANTS AND ANIMALS**

There are many ways in which genome sizes can change over time. In general, each of the most commonly recognized mechanisms applies to both plants and animals, although their specific nature and importance may vary considerably both within and between the two groups. The more notable mechanisms common to plants and animals are discussed in the following sections, with special emphasis on the points of divergence between the kingdoms.

**Transposable elements**

Under the classic ‘selfish DNA’ theory, sequences such as transposable elements (TEs), which are capable of their own propagation, spread within the genome until their activity is halted by selection against replicational costs (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Although the modern view of transposable elements is much more complex, with TE–host interactions recognized as ranging from parasitic to mutualistic (Kidwell and Lisch, 2001), the basic notion that TEs contribute substantially to eukaryotic genomes has been borne out. For example, roughly 45% of the human genome (International Human Genome Sequencing Consortium, 2001), and more than 60% of some plant genomes (Bennetzen, 2002), is comprised of TEs and (mostly) their defunct remnants. In a most striking example, it appears that the genome of maize has doubled in size over only a few million years by a surge in TE activity, indicating that this can be a mechanism of rapid change nearly on par with whole-scale duplication (SanMiguel and Bennetzen, 1998).
To be sure, the exact types of TE sequences present may vary considerably among genomes, even within kingdoms. For example, whileLTRretrotransposons have been suggested to predominate in plant genomes (Kumar and Bennetzen, 1999), this may in fact be restricted to grasses (Wendel and Wessler, 2000). Likewise, LTR retrotransposons may be among the most common sequences in fruit fly and mosquito genomes (Adams et al., 2000; Holt et al., 2002), but LINEs and SINEs are much more common in mammals and pufferfishes (International Human Genome Sequencing Consortium, 2001; Aparicio et al., 2002; Mouse Genome Sequencing Consortium, 2002), while DNA transposons predominate in nematodes (C. elegans Sequencing Consortium, 1998). Nevertheless, it is evident that TEs in general play a very important role in shaping genome size diversity among eukaryotes. In broad terms, among those eukaryotes whose genomes have been sequenced in detail (mostly animals, but also some plants) there is an approximately linear relationship between genome size and the total amount of transposable element DNA present, although TEs generally contribute a higher percentage of DNA in larger genomes (Kidwell, 2002).

**Introns and other sequences**

The abundances of many other sequence types also correlate positively with genome size in eukaryotes, which is in keeping with the notion of global genomic forces acting to shape DNA content (Gregory and Hebert, 1999; Petrov, 2001). Simple sequence repeat content (Hancock, 2002), rDNA gene copy number (Prokopowich et al., 2003), and intron size (Vinogradov, 1999) are all positively associated with genome size across taxonomic samples. However, again these overall similarities persist despite important differences. For example, the relationship with rDNA multiplicity is slightly weaker in plants than in animals, and plants clearly maintain more rDNA gene copies per unit genome size than animals (Prokopowich et al., 2003).

The differences between plants and animals may be especially pronounced with regard to introns. Whereas correlations between intron size and genome size are apparent among related animals like *Drosophila* species (Moriyama et al., 1998), such relationships appear to be absent in plants such as *Gossypium* (Wendel et al., 2002a). Perhaps most significantly, it has been claimed that the majority of non-coding DNA is intronic in animals, but that this is not true of plants, in which most non-coding DNA is apparently intergenic (Wong et al., 2000).

**Polyploidy**

Technically, polyploidy itself does not represent a change in genome size *per se* because it actually involves the addition of a second genome. That is to say, so long as the two concellular genomes remain distinct (i.e. until rediploidization occurs), C-value will change but genome size will not. As it turns out, the new polyploid C-value may be less than the expected sum of the parental genomes (e.g. Ozkan et al., 2003; Bennett and Leitch, 2005), but obviously this would nevertheless be a means of adding significant amounts of DNA to the nucleus. In any case, variation in C-value (and probably genome size) has been much more greatly affected by polyploidy in plants than in animals (Gregory and Mable, 2005; Tate et al., 2005). It is often stated that polyploidy is common in plants, but the frequency with which it occurs actually varies considerably among taxa. For example, most (perhaps all) angiosperms and ferns and about 60% of mosses have polyploidy in their ancestry, whereas it is rather uncommon in liverworts and gymnosperms (e.g. Averett, 1980; Delevoryas, 1980; Masterson, 1994; Otto and Whitton, 2000; Wendel, 2000). And although its occurrence is usually under-appreciated in animals (given that examples can be found in every major phylum), it is true that cases are comparatively quite scarce among the metazoa. Nevertheless, an ancient polyploidization event appears to have played an important role in early vertebrate genome evolution (e.g. McLysaght et al., 2002). In teleost fishes, there may have been a second round of genome duplication (van de Peer et al., 2003), and it is notable that in general the only extant teleosts that exceed a fairly small genome size range are those such as the salmonids, cyprinids and catostomids whose more recent ancestry includes polyploidy. On the other hand, polyploidy is much more common in frogs than in salamanders, despite the invariably larger genomes of the latter.

**DNA loss**

Until recently, all of the known mechanisms of genome size change had involved DNA gain. This fact provided cause to wonder whether plants have ‘a one-way ticket to genomic obesity’ (Bennetzen and Kellogg, 1997a). While it is likely that the ancestral genome size was small in plants (Leitch et al., 1998, 2005), phylogenetic reconstructions of specific taxa have shown both increases and decreases to have occurred (e.g. Watanabe et al., 1999; Wendel et al., 2002b). The same can be said of animals, so there is a clear need for a mechanism of DNA loss in both groups. Fortunately, mutational mechanisms of DNA loss have been increasingly emphasized of late (Hartl, 2000; Petrov, 2001; Bennetzen, 2002). Unfortunately, in some cases these have been taken to an undue extreme by the development of purely neutralist models in which differences in DNA loss rate are considered the prime determinants of variation in genome size (Petrov, 2002).

The mutational mechanisms in question operate on (at least) three very different scales. The first, which forms the basis of the ‘mutational equilibrium model’, is based on a predominance of deletions over insertions on scales less than 400 bp. All of the relatively limited data presented in support of this mechanism are derived from animals, but it has been suggested that this should also apply to plants (Petrov, 1997). However, at this scale the deletional mechanism is extremely weak, and is unlikely to play a major role in the large-scale genome size evolution in either plants or animals (Bennetzen and Kellogg, 1997b; Gregory, 2003a, 2004). As a prime example, by this mechanism it would take more than 600 million years to delete half of the newly acquired TEs in the maize genome, provided that no
other insertions or duplications took place in the meantime (Gregory, 2003a, 2004). Since the entire family Poaceae is only about 75 million years old and exhibits particularly dynamic genomes (Gaut, 2002), maize’s recent genomic expansion would hardly seem to be ‘noise around the long-term equilibrium value’ as Petrov (2002) suggests.

The second scale of deletional mechanisms is considerably more powerful, and involves recombination between homologous copies of the long terminal repeats characteristic of LTR retrotransposons. In this case, all of the available evidence comes from plants, most notably Hordeum spp. (Kalander et al., 2000). When this recombination occurs, most of the element is lost, leaving behind only a ‘solo LTR’. Of course, this mechanism can only slow genomic growth by TE insertions, because at the very least a solo LTR is retained each time (Devos et al., 2002).

The third scale, which also involves LTR retrotransposons and is also so far only studied in plants, is the only one of the three capable of producing extensive genomic shrinkage over reasonable timescales. Here, it is not recombination between homologous LTRs, but rather ‘illegitimate recombination’ between LTRs either on the same or different chromosomes that leads to DNA loss. And since these are non-homologous elements, the DNA lost will be all that is located in between the two elements and may therefore involve much larger amounts than are added by each TE insertion (Bennetzen, 2002, Bennetzen et al., 2005). In theory, this mechanism should apply to both plants and animals, although a great deal more data are needed before any general conclusions can be drawn regarding its actual role in genome size evolution. Moreover, when such large deletions are involved, the role of selection cannot be so easily dismissed as in the mutational equilibrium model.

PARALLEL IMPACTS ON PLANT AND ANIMAL PHENOTYPES

Cell size

A general association between nucleus size and cell size in vertebrate red blood cells has long been recognized (e.g. Gulliver, 1875), and even in one of the very first comparative surveys of animal genome size, Mirsky and Ris (1951) noted that ‘in the nucleated red cells of vertebrates . . . there is an approximately direct relationship between cell mass and DNA content’. This general relationship applies to plants and unicellular eukaryotes as well, and, as noted by Cavalier-Smith (1982) more than 20 years ago, is perhaps ‘the most reliably established fact about genome evolution’.

In plants, seed size and genome size are linked by something of a triangular relationship, meaning that small genomes can be associated with small or large seeds, but that large genomes are not found in small seeds (Bennett, 1987; Thompson, 1990; Knight and Ackerly, 2002). Egg size, the zoological counterpart of seed size, has not been nearly as well studied, but is known to correlate positively with genome size in cladoceran crustaceans (which all have very small genomes; Beaton, 1995) and plethodontid salamanders (which have large genomes; data from Jockusch, 1997). Given that egg size, like seed size, often has important fitness consequences, this is a parameter worthy of much more study.

It can be difficult to find correlations between C-value and the sizes of gametes besides eggs in both kingdoms because both pollen and sperm can be greatly modified according to specific reproductive requirements. In plants, a correlation is observed so long as similar types (e.g. wind-dispersed) of pollen are compared across related species (e.g. Bennett, 1972). In animals, the situation may be even more complex, with a correlation appearing when comparing very closely related samples, but not across larger groups. For example, the red viscacha rat (Tympanoctomys barrerae), the only polyploid mammal known, has considerably larger sperm than its diploid relatives (Gallardo et al., 1999), and even the sizes of X- vs Y-chromosome carrying sperm differ in direct proportion to their varying DNA contents (about 3–4 %) in both bulls and humans (Cui, 1997; van Munster et al., 1999). However, across mammals at large there is no relationship between genome size and sperm size (Gage, 1998), presumably because sperm morphology is adaptively modified to a considerable degree in mammals while genome size remains constrained. Other groups with larger genome size ranges may reveal a correlation with sperm size, but this has yet to be investigated.

Relationships with somatic cells are well established in both plants and animals, although obviously the pertinent cell types differ greatly between the two groups. In plants, the best studied cell type from this perspective is meristems (Price et al., 1973), although leaf guard cell size is also associated with DNA content (Masterson, 1994). Cell size is also correlated with genome size in algae (Holm-Hansen, 1969; Kapraun and Dunwoody, 2002). The sizes of various cell types, including neurons, liver cells (hepatocytes) and epithelial cells, all appear to correlate positively with genome size in vertebrates (Gregory, 2001b). However, by far the best known relationship in animals involves red blood cells (erythrocytes), which differ fundamentally from plant meristems in that they are highly compact and non-dividing (Gregory, 2001b, c). Yet, in all of these cell types from both kingdoms, genome size and cell size are probably linked causally by the influence of DNA content on the cell cycle, such that larger genomes delay division and result in the production of larger daughter cells (Gregory, 2001b, c). The fact that erythrocyte size correlates positively with genome size in mammals, even though their mature red blood cells are enucleated (i.e. genome-free), strongly supports this hypothesis (Gregory, 2000, 2001b, c, 2005).

Morphology

Because bodies are composed of cells, the obvious possibility exists that a change in genome size will result in a change in body size. However, a general correlation between genome size and overall body size has not been reported to occur in plants. Instead, the sizes of various structures, such as leaves, tend to correlate with genome
size. Even here the situation is somewhat complex, given that leaf size may actually correlate positively or negatively with genome size, depending on the taxonomic sample (Knight et al., 2005). A similar situation exists in animals, with positive correspondences found between genome size and body size in numerous groups of invertebrates, such as aphids (Finston et al., 1995), flies (Ferrari and Rai, 1989), copepod crustaceans (McLaren et al., 1988; Gregory et al., 2000), polychaete annelids (Soldi et al., 1994) and turbellarian flatworms (Gregory et al., 2000). In nematodes, body size is influenced not by genome size per se, but by the level of somatic endopolyploidy (Flemming et al., 2000). On the other hand, such relationships are not observed in groups like oligochaete annelids (Gregory and Hebert, 2002), moths (Gregory and Shorthouse, 2003) or within the few families of beetles that have been studied to date (Juan and Petitpierre, 1991; Petitpierre and Juan, 1994; Gregory et al., 2003). In most vertebrates, cell (and therefore genome) size does not correlate in any way with body size; notably, the difference in mass between the smallest shrew and the blue whale, which covers more than seven orders of magnitude, is due almost entirely to variation in cell number. However, in birds and rodents, where cell number variation is much more limited, cell and genome size do correlate positively with body size (Gregory, 2002a, b). Some authors have suggested a negative correlation with body size within the beetle genus Pimelia (Palmer et al., 2003), but this is only a phylogenetic correlation and may not have much real biological significance (Gregory et al., 2003).

**Physiology**

The primary role of red blood cells in vertebrates is in gas exchange, a process strongly dependent on cellular surface area to volume ratios. As such, it has long been suggested that genome sizes are assorted according to metabolic parameters, with ‘wasteful’ groups like mammals and birds having small cells and genomes and ‘frugal’ ones like amphibians having large cells and genomes (Szarski, 1983). Indeed, there is a significant negative correlation between mass-corrected oxygen consumption rate and genome size in both mammals (Vinogradov, 1995) and birds (Gregory, 2002a). However, and contrary to many previous assumptions, this has not been found to extend to amphibians except insofar as frogs as a group are more active and have smaller genomes than salamanders (Gregory, 2003b).

In the simplest terms, the respiratory physiology of plants is the opposite of that in animals. However, this may nonetheless relate to genome size, given that stomatal size and specific leaf area correlate positively with DNA content (Masterson, 1994; Chung et al., 1998; Knight et al., 2005), just as do erythrocyte sizes in vertebrates. In fact, it has recently been shown that photosynthetic rate in plants, like metabolic rate in homeothermic vertebrates, correlates negatively with C-value (Knight et al., 2005).

With regard to physiological stress, it has been noted that increased tolerance to both droughts (e.g. Castro-Jimenez et al., 1989; Wakamiya et al., 1993, 1996) and frost (e.g. MacGillivray and Grime, 1995) are associated with larger genome size in plants. As an interesting parallel, Shahbasov and Ganchenko (1990) reported a positive association between DNA content and non-specific thermal and hypoxic stress tolerance in frogs and salamanders. On the other hand, it is becoming apparent that large-genomed plants are generally excluded from the extremes of climatic ranges, whether cold or hot (Knight and Ackerly, 2002; Knight et al., 2005). At least one example of this is found in animals as well: In polychaete annelids, macrobenthic species inhabiting stable environments have larger genome sizes, while those found in harsh interstitial environments invariably have small C-values (Sella et al., 1993; Soldi et al., 1994; Gambi et al., 1997).

**Development**

It is not only cell size, but also both mitotic and meiotic division rates that correlate with genome size. Most of the demonstrations of this inverse correlation have come from plants, but this has also been documented experimentally in some animals (reviewed in Gregory, 2001a). At the organismal level, this often translates into a negative correlation between genome size and developmental rate. Indeed, such correlations have been found within and among numerous groups of plants (e.g. Bennett, 1972; Grime et al., 1985; Mowforth and Grime, 1989). There is even evidence that experimental selection for earlier flowering time may result in a reduction in genome size (Rayburn et al., 1994). More generally, there are the well-known patterns whereby plants with large genomes cannot adopt an annual or ephemeral lifestyle and in which weeds tend to have small genomes (Bennett, 1987; Bennett et al., 1998). However, this developmental correlation is not universal in plants, and in some cases a positive relationship can be observed, again depending on the taxa being compared (Knight et al., 2005).

As with plants, genome size and developmental rate are inversely correlated in many but not all animal taxa. Thus, negative correlations are found in amphibians, insects and crustaceans (see Gregory, 2002c), but not in mammals and birds (Gregory, 2002b). In amphibians and insects, there is a direct parallel between the annual versus perennial lifestyle threshold seen in plants, in this case involving constraints related to metamorphosis. To appreciate this, it is necessary to draw a distinction between developmental rate (the time taken to develop) and developmental complexity (the amount of developing to be done in a limited amount of time), which are in fact two sides of the same coin (Gregory, 2002c, 2005). Since metamorphosis involves a strongly time-limited period of intense tissue differentiation, it requires rapid cell divisions and therefore small genomes. Thus, frogs inhabiting ephemeral pools have the smallest amphibian genomes (~1 pg) whereas obligately non-metamorphosing (neotenic) salamanders have the largest (up to 120 pg). In like fashion, in insects with complete metamorphosis (holometabolan development), genome sizes almost never exceed 2 pg, whereas those with no or only incomplete metamorphosis (amebolaous or hemimetabolan development) may have genomes up to 17 pg (Gregory, 2001b, 2002a).
There are clearly numerous parallels between animals and plants in terms of the correlations between genome size and cell size and division rate, and in the influence of these on the organismal phenotype. The particular biology of the organisms in question is important, of course, and there are major differences in the expression of these relationships between plants and animals. In fact, this is also the case within kingdoms, as for example with mammals and birds, in which metabolism is important but development is not, versus amphibians, in which the opposite is true. The key point to recognize is that while the specific expression of the phenomena differ among groups, the underlying mechanisms are the same whether in plants or animals.

MEASURING GENOME SIZES IN PLANTS AND ANIMALS

Methodology: similarities and differences

As with all of the issues discussed in this review, the methods employed in genome size measurement are far more similar between plants and animals than they are different. In both groups, genome sizes have traditionally been assessed either by Feulgen microdensitometry or flow cytometry (usually using propidium iodide or DAPI). More recently, and essentially simultaneously, the technique of Feulgen image analysis densitometry has been developed for use with both plant (Vilhar et al., 2001; Vilhar and Dermastia, 2002) and animal (Hardie et al., 2002) specimens.

All of the same basic issues arise when using these techniques regardless of whether the specimen is a plant or an animal. For Feulgen-based methods, the primary issues in both groups are the preparation of a suitable monolayer of cells, the temperature and duration of acid hydrolysis, and the quality of the Schiff reagent used (Greilhuber and Temsch, 2001; Hardie et al., 2002; Bennett and Leitch, 2005). In flow cytometry, the preparation of individual nuclei, the choice of dye and the staining protocols are equally important in both groups.

One of the most notable differences between plant and animal studies is that the former typically involve dividing cells like root or shoot tips or leaves, while the latter almost always make use of non-dividing cells such as blood cells or sperm. In practical terms, this means that plant studies involve a challenge in finding 2C or 4C nuclei among the entire range of partially replicated genomes, whereas in animal studies the cells are all either 2C (blood) or 1C (sperm). The use of blood or sperm also alleviates the problem of endopolyploidy, which is common to both animals and plants. However, with many small invertebrates whole bodies or large segments thereof must be used in flow cytometry, and in this case the problem of locating cells of known ploidy is just as prominent as it is in plants.

Plant and animal studies also tend to differ in the methods of tissue preparation as a result of the very different cell types used. In Feulgen-based methods, plant tissues are often fixed in the field and later prepared as squashes. In animals, blood smears can be made in the field and simply allowed to air-dry, or else dissections must be performed on fresh material in the laboratory (e.g. to get insect sperm); the dehydration caused by fixation makes it impossible to acquire blood cells from either vertebrates or invertebrates and greatly complicates dissection. By contrast, flow cytometry often involves the preparation of fresh tissues in plants, but almost invariably uses fixed or frozen tissues (including blood) for animals. The fixatives used may also differ among kingdoms, since 3 methanol : 1 acetic acid is commonly used in plants, whereas 85 methanol : 10 formalin : 5 acetic acid is recommended for most animal preparations (Hardie et al., 2002). However, with some invertebrates (e.g. copepod crustaceans), the fixation and squash protocol employed may be essentially the same as that used for plants (e.g. Wyngaard and Rasch, 2000).

As noted above, many plants produce phytochemicals that may interfere with either Feulgen or fluorescent staining and produce artifactual examples of intraspecific variation (Greilhuber, 1986; Price et al., 2000; Noirot et al., 2002). In animals, no such problematic chemicals have been identified to date, and are not likely to be found when tissues such as blood are used. However, the possibility does exist that stain inhibitors may be present when using whole-body specimens of invertebrates, and once again zoologists would be well advised to take a lesson from their botanical colleagues.

Choice of standards

One of the largest sources of error in genome size measurements, in both plants and animals and whether using either Feulgen methods or flow cytometry, involves differences in DNA compaction levels, which directly affect the level of stain uptake. This can be dealt with in several ways, but the best by far is to simply choose a standard of the same cell type as the unknown being measured. Even within animals, this is very important (Hardie et al., 2002), and this becomes even more critical when comparing across kingdoms. Animal standards (especially chicken or trout blood) have been used many times for plant studies (Bennett and Leitch, 2003), and very occasionally plant standards have been employed in animal measurements (Greilhuber et al., 1983).

Chicken erythrocytes are not considered a suitable standard for comparison with epithelial cells, leukocytes or sperm from fellow vertebrates, and cannot even be compared accurately to liver cells from the same animal (Hardie et al., 2002). Comparisons with plant cells would obviously be much more problematic than this, and it is therefore not surprising that the use of animal standards in plant studies was strongly discouraged at both the 1997 and 2003 Kew meetings. There is one exception to this that was noted at the most recent meeting, which is that it would be advisable to use a completely sequenced genome whose size is known with absolute certainty to calibrate a series of ‘gold standards’ for use in future plant studies. Since the only genome that has truly been sequenced in its entirety is that of the nematode Caenorhabditis elegans, this remains the only choice currently available. Fortunately, it appears that
C. elegans can be used in flow cytometric comparisons to plants with small genomes, such as Arabidopsis (Bennett et al., 2003).

THE FUTURE OF GENOME SIZE

Links with sequencing projects

While the measurement and compilation of plant genome sizes have consistently exceeded the efforts made for animals, the opposite is true in terms of genome sequencing. To date, roughly five times as many animal genomes have been sequenced as compared with plants. This discrepancy is only likely to increase in the future, given that numerous additional animal projects are under way, whereas only very few such programmes are under development for plants. Moreover, while there is much current talk of attempting to acquire genome sequences for at least one representative of every major animal phylum in addition to all the major experimental models and species of economic and medical importance, the upcoming sequencing (or only mapping) efforts for plants will be mainly restricted to species of agricultural interest (e.g. wheat, barley, maize, soybean, oat, banana and tomato).

In any case, it is well appreciated that one of the first steps in deciding on a subject and strategy for sequencing is to determine genome size. For an unknown reason, there has been a large disconnect between the sequencing and genome size communities for both plant and animal projects. As a result, an incorrect genome size was assumed by the Arabidopsis Genome Initiative (2000), even though the C-value of this species had been estimated in ten different publications before the sequencing results were published (see Bennett et al., 2003). These estimates would have been easily available from the Plant DNA C-values Database. Similarly, the Drosophila genome sequence paper of 2000 made no reference to the five different genome size estimates that had already been published (Adams et al., 2000). Granted, the Animal Genome Size Database did not exist in 2000, but it still should have been possible to locate well-known papers like that by Rasch et al. (1971), which provided a very careful and accurate estimate using Feulgen densitometry. Obviously, a stronger link between sizers and sequencers should help to avoid these problems in the future. More importantly, an improved awareness of the Plant DNA C-values Database will greatly facilitate the selection of the next series of plant sequencing subjects once the obvious choice of crop species is exhausted.

Biologists interested in the evolution of genome size would also benefit from such an integration with sequencing projects. Most obviously, whole-scale genome sequencing information can be used to assess the relative contribution of different sequence types to variation in genome size. This has already been done in a very preliminary way with regard to transposable elements, in this review and elsewhere (Kidwell, 2002). As more sequences become available, it will finally become possible to address the crucial component of the C-value enigma dealing with mechanisms of genome size change. Conversely, an integration with existing information on patterns and implications of genome size variation will allow the data of sequencing projects to be placed in a real biological context.

Talk across the trenches

It is unlikely that genome sequencing programmes will pay due attention to existing studies of genome size so long as there is no unified front presented by botanists and zoologists. On the other hand, once it is recognized that plant and animal genome size are components of a single overarching puzzle, the general importance of the issue will probably become much clearer to those in related fields.

Increased talk across the taxonomic trenches would benefit both sides in other ways as well. Since most (if not all) of the patterns and consequences of genome size variation transcend taxonomic boundaries, the work being carried out by one side cannot help but illuminate that of the other. Correlations with cell size, as one clear example, can be best studied comparatively and mechanistically using both plants and animals (Gregory, 2001a). Based on the survey presented above, it would seem that morphology, physiology, development, methodology and mechanisms of change also fall into this category of mutual overlap. On practical grounds, zoologists would benefit from increased interaction simply because botanists have led the way on many of these issues, making it unnecessary to re-invent the conceptual wheel. Botanists, for their part, should recognize the basic fact that zoologists are often asked to review plant genome size papers (while the reverse rarely occurs), suggesting an obvious pragmatic benefit of increased understanding between the two groups.

Beyond plants and animals

To date, there have been no significant efforts to compile databases of ‘prost’ or fungus genome sizes, and even with prokaryotes most of the readily available information comes from species being used in sequencing projects. Given that the most informative comparisons in plants and animals have only become possible after the assembly of broad datasets, it is obvious that many important insights remain hidden amongst the scattered data from these other kingdoms. More fungi have been sequenced than plants, and the construction of a database for these organisms will also undoubtedly help with the highly desirable link between sequence and size described above.

Much work remains to be done in assessing the patterns of variation within animals and plants, and there is clearly a need to better share the lessons learned from the study of these two original kingdoms. However, green things and moving things hardly encompass the fullness of life’s diversity and, as such, expansion, as well as integration, will be necessary if genome size evolution is to be properly understood. Every living thing has a genome, and for every genome there is a size. Within these simple principles lie the future of genome size study and the eventual resolution of the C-value enigma.
ACKNOWLEDGEMENTS

The author was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) postdoctoral fellowship and the NSERC Howard Alper Post-Doctoral Prize. Sincere thanks are given to the organisers and participants of the 2003 Kew genome size workshop and discussion meeting, and especially to M. D. Bennett, T. Cavalier-Smith, J. S. Johnston, C. A. Knight, I. J. Leitch, H. Price and B. Vilhar for stimulating discussions and helpful comments.

LITERATURE CITED


Bennett MD, Bhandol P, Leitch IJ. 2000. Relationship between genome size and content, cell size, and the C-value enigma. Biological Reviews 75: 203–266.


Greilhuber J. 1986. The bigger the C-value, the larger the cell: genome size and red blood cell size in vertebrates. Blood Cells, Molecules, and Diseases 27: 830–843.


