Cytometry
PART A
Journal of the International Society for Advancement of Cytometry

Nuclear Genome Size: Are We Getting Closer?

Jaroslav Doležel,[1*] Johann Greilhuber[2]

[1]Laboratory of Molecular Cytogenetics and Cytometry, Institute of Experimental Botany, Sokolovská 6, CZ-77200 Olomouc, Czech Republic

[2]Department of Systematic and Evolutionary Botany, Faculty of Life Sciences, University of Vienna, Rennweg 14, A-1030 Vienna, Austria

Received 23 January 2010; Revision Received 10 April 2010; Accepted 14 April 2010

Grant sponsor: Czech Ministry of Education, Youth and Sports; Grant number: LC06004.

*Correspondence to: Jaroslav Doležel, Institute of Experimental Botany, Laboratory of Molecular Cytogenetics and Cytometry, Sokolovská 6, CZ-77200 Olomouc, Czech Republic

Email: dolezel@ueb.cas.cz

Published online 20 May 2010 in Wiley InterScience (www.interscience.wiley.com)

DOI: 10.1002/cyto.a.20915

Abstract

Correct information on genome size is important in many areas of research. For a long time, scientists have been struggling to understand the reason for the huge variation in eukaryotic genome size and its biological significance. More recently, the knowledge on genome size has become important to structure genome sequencing projects as their scale and cost depend on genome size. Despite the fact that the first estimates of genome size in eukaryotes were made more than 50 years ago, we are still not quite sure about the exact genome size in practically all animal and plant species. Moreover, different estimates continue to be published for the same species. These discrepancies compromise data comparison and interpretation and point to methodological problems, which include standardization. This article assesses the current state of DNA reference standards for flow cytometry and the issues related to their calibration. © 2010 International Society for Advancement of Cytometry

Key terms

cytometric techniques; reference standards; genome sequencing; conversion factor; nuclear DNA content; C-value

MOST of the genetic information in eukaryotes is localized in the cell nucleus and numerous attempts have been made to determine the quantity of nuclear DNA, especially after various lines of evidence associated DNA with genes (1). It soon became clear that the amount of DNA per nucleus is relatively constant in somatic cells of a given species and that sperm cells have approximately half the DNA amount found for somatic cells (2,3). These experiments provided independent data to support the hereditary role of DNA and marked the beginning of numerous fruitful lines of research and applications, many of them still important today, all relying on the ability to determine DNA amounts in cell nuclei.

The early determinations were made colorimetrically on DNA extracted from a known number of cells although the presence of cells in different cell cycle phases could compromise the accuracy of these estimates. The colorimetric approach made identification of subpopulations of cells with different DNA amounts impossible. However, cytometric methods suitable for measurement of absorbance of light in individual nuclei (either using UV light with non-stained nuclei and/or monochromatic visible light with nuclei stained by the Feulgen reaction) were already available during the 1940s and were used to discover the presence of cells with different classes of DNA amounts. Such cells were observed in a variety of animal tissues and were associated with polyploidy and polyteny (4,5). The presence of nuclei with 2, 4, 8, 16, or 32 times the haploid value was described in plant tissues by Swift (6) who introduced the "C" terminology to classify nuclear DNA amounts. In this terminology, DNA classes are labeled as C, 2C, 4C, 8C, etc. to characterize DNA amounts of nuclei by multiples of the DNA amount in a complete chromosome set in a non-replicated haploid nucleus, which has the class C DNA amount.

At the beginning of the 1950s, biochemical and cytochemical studies established the constancy of nuclear DNA amount for a given species and described its alterations during the mitotic and meiotic cycle and changes due to polyploidy and poly-

ISAC
International Society for Advancement of Cytometry

teny. An observation with important implications for the understanding of cellular pathology and carcinogenesis and for subsequent biomedical applications of nuclear DNA content measurements concerned the increase of nuclear DNA amounts in carcinomas and lymphocytes of patients with leukemia (7,8).

In line with studies on differences in DNA amounts within various organisms, the pioneering studies discovered differences in DNA amounts between species (9). This work, together with other studies in animals and plants which followed (10–14), revealed a lack of correlation between the evolutionary complexity of an organism and its nuclear DNA content. This implied a lack of correlation between the amount of nuclear DNA and the number of genes and Thomas (15) coined a term "C-value paradox" to describe this phenomenon. Although this term still remains in use today, there is no longer a paradoxical aspect of this discrepancy. Nuclear DNA of various organisms contains various proportions of noncoding and repetitive sequences and hence its quantity needs not to be related to the number of genes. But why do organisms differ in DNA amounts after all? This is a persisting problem and various components of this question have been collectively termed the "C-value enigma" (16).

In a search to reveal the evolutionary trends in the variation of nuclear DNA content and its biological significance, efforts have been made to estimate DNA amounts in as many species as possible. On March 1, 2010, the Animal Genome Size Database contained 6,518 records representing 4,972 species (17) while the Plant DNA C-Values Database comprised 6,744 entries (18). Although these numbers are impressive, the coverage of the existing biodiversity is far from complete and many taxa and important clades are missing (19). Moreover, contrasting C-values can be found for the same species. For example, 1C-values ranging from 3.80 to 5.93 pg DNA can be found for *Pisum sativum* in the Plant DNA C-Values Database (18), see also Table 1. Although it is possible that in some species publications on intraspecific variation of nuclear DNA content reflect a real variability, many reports are suspected to be erroneous (20,21).

The discrepancies in the estimation of nuclear DNA content point to methodological problems and compromise data comparison and interpretation. More than 50 years after the pioneering studies, we are still far from the point of being able to estimate reliably eukaryotic nuclear DNA amounts in absolute units. This article focuses on an important methodological aspect of the estimation of nuclear DNA content in plants, which is the use of reference standards.

Terminology on Nuclear DNA Content and Genome Size

Before proceeding with methodological aspects of DNA content estimation, it is useful to summarize the current terminology. One copy of the genetic information had already been termed genome in 1920 by Winkler (22). However, the term genome size to denote the quantity of DNA in which the information is stored was not coined until the late 1960s (13,23). Unfortunately, the definition of genome size was not clear enough and the terminology remained ambiguous. Thus, the mass of nuclear DNA in the whole chromosome complement (with chromosome number *n*) was often referred to as genome size irrespective of ploidy and number of basic chromosome sets (*x*) the chromosome complement comprised. Only recently, Greilhuber et al. (24,25) proposed a coherent terminology in which the term genome size is used as a covering term in the wide sense, irrespective of ploidy. The necessary distinction of the kinds of genome sizes is made by the adjectives "monoploid" (one chromosome set of an organism having the chromosome base number *x*) and "holoploid" (the whole complement of chromosomes with chromosome number *n* characteristic for the organism). The abbreviated terms (symbols) for monoploid and holoploid genome size are Cx-value and C-value, respectively (24).

Conversion Between DNA Mass and the Number of Base Pairs

Nuclear DNA amounts were traditionally given in picograms of DNA. With the advent of molecular biology the trend was to express genome sizes in the number of base pairs. This necessitated a conversion factor. Britten and Davidson (26) considered 9.13×10^8 nucleotide pairs per pg DNA and this factor was quoted by Nagl (27) without providing a reference and later used by Rasch (28) and other authors (29). When estimating genome size of *Rana pipiens*, Straus (30) mentions 5.6×10^9 nucleotide pairs per 5.8 pg DNA in the haploid genome. These values were most probably used by Bennett and Smith (31) to derive a factor of 0.965×10^9 for the conversion of pg DNA to the number of base pairs. This factor was used extensively, although a different factor of 0.98×10^9 base pairs per pg DNA was used by Cavalier-Smith (32). Unfortunately, derivation of the latter was not explained in the publication.

Probably the most accurate conversion factor was derived by Doležel et al. (33) who were motivated to do so after reading a paper by Thomas et al. (34). In this work, 2C DNA amount of human female and male was estimated to be 3.77 pg and 3.70 pg DNA, respectively. These values were in a sharp contrast to all previous estimates and a careful reading revealed that DNA amounts were calculated assuming 6.162×10^9 nucleotides for the human male nucleus and a mean nucleotide molecular weight of 360 g/mol (the authors erroneously considered nucleotides and not nucleotide pairs). Doležel et al. (33) determined the mean relative molecular weight of one nucleotide pair to be 615.8771 and used this value to derive a formula for converting the number of nucleotide pairs to picograms of DNA: genome size (bp) = $(0.978 \times 10^9) \times$ DNA content (pg). These calculations not only revealed the serious error of Thomas et al. (34), but provided a clear reference for converting DNA mass in picograms to the number of base pairs. This conversion factor has been used in many studies and as of April 30, 2010, the paper of Doležel et al. (33) has been cited 126 times (ISI Web of Knowledge), catapulting it to a position of the third most cited paper in the history of the scientific journal Cytometry Part A (A. Tárnok, pers. comm.).

Table 1. Independent estimates of 2C DNA content in *Pisum sativum*. Two different 2C-values were considered for *Allium cepa*, which was used as reference standard

		RATIO OF 2C-VALUES (*P. SATIVUM*/*A. CEPA*)		*P. SATIVUM* 2C DNA CONTENT (PG)[a]			
SAMPLE NO.	ACCESSION OF *P. SATIVUM*	RATIO	RANK	A	B	METHOD[b]	REFERENCE
1	cv.	0.3119	35	10.45	10.84	FDM	85
2	Minerva Maple	0.2896	34	9.70	10.07	FDM	31
3	cv.	0.2269	1	7.60	7.89	FDM	86
4	cv.	0.2783	31	9.32	9.67	FCM	87
5	cv.	0.2609	17	8.74	9.07	FCM	43
6	Express Long	0.2560	11	8.58	8.90	FCM	88
7	Frühe Maiperle	0.2693	25	9.02	9.36	FDM	75
8	Frühe Maiperle	0.2693	26	9.02	9.36	FDM	75
9	5 cultivars	0.2573	12	8.62	8.94	FDM	75
10	2 cultivars	0.2543	9	8.52	8.84	FDM	75
11	3 cultivars	0.2651	19	8.88	9.21	FDM	75
12	Frühe Maiperle	0.2663	21	8.92	9.26	FDM	75
13	Perfektion	0.2633	18	8.82	9.15	FDM	75
14	Wunder von Kelvedon	0.2675	23	8.96	9.30	FDM	75
15	Frühe Heinrich	0.2651	19	8.88	9.21	FDM	75
16	Minerva Maple	0.2585	14	8.66	8.99	FDM	75
17	19 accessions	0.2714	29	9.09	9.43	FDM	75
18	19 accessions	0.2694	27	9.02	9.36	FDM	75
19	19 accessions	0.2687	24	9.00	9.34	FDM	75
20	Kleine Rheinländerin	0.2535	7	8.49	8.81	FCM	74
21	Kleine Rheinländerin	0.2532	5	8.48	8.80	FCM	74
22	Ctirad	0.2672	22	8.95	9.29	FDM	44
23	Ctirad	0.2587	15	8.67	8.99	FDM	44
24	Ctirad	0.2532	6	8.48	8.80	FDM	44
25	Ctirad	0.2589	16	8.67	9.00	FDM	44
26	Ctirad	0.2696	28	9.03	9.37	FCM	44
27	Ctirad	0.2543	8	8.52	8.84	FCM	44
28	Ctirad	0.2448	2	8.20	8.51	FCM	44
29	Ctirad	0.2758	30	9.24	9.59	FCM	44
30	Minerva Maple	0.2848	32	9.54	9.90	FCM	77
31	Ctirad	0.2549	10	8.54	8.86	FIDM	89
32	Ctirad	0.2574	13	8.62	8.95	FIDM	89
33	Ctirad	0.2867	33	9.60	9.97	FIDM	89
34	Frison	0.2500	3	8.38	8.69	FCM	76
35	Minerva Maple	0.2500	4	8.38	8.69	FCM	76
Mean		0.2641	–	8.85	9.18	–	–
Standard Deviation		0.0148	–	0.496	0.516	–	–

[a] Two different 2C values were considered for the reference standard *Allium cepa*: A - 2C = 33.5 pg (31) and B - 2C = 34.76 pg (43).
[b] FCM, flow cytometry; FDM, Feulgen scanning densitometry; FIDM, Feulgen image densitometry.

Estimation of Genome Size Using Flow Cytometry

Although it may come as a surprise, estimation of genome size is not a trivial task even in the beginning of the third millennium. As most of animal and plant tissues are comprised of cells at various stages of the cell cycle, it is not advisable to isolate DNA from many cells and calculate mean DNA content. Instead, various cytometric methods have been used to estimate DNA amounts in single cells and nuclei whose cell cycle position is known. Feulgen microdensitometry dominated the field until the beginning of 1980s, when it has been gradually replaced by flow cytometry, which offers a convenient approach and higher throughput. A trend towards using flow cytometry was stimulated by the cessation of the production of scanning microdensitometers—"the obsolescence time bomb" (35). Although image cytometry can replace scanning microdensitometry, it has been used rather exceptionally in the plant sciences. On the other hand, it plays a significant role in the animal genome size research (17).

Flow cytometric estimation of nuclear DNA content in plants is typically done on nuclei isolated by homogenization of fresh tissues (36). Although various modifications of the original protocol have been introduced (37,38), tissue chop-

ping with a sharp razor blade remains the prevailing approach. The composition of the nuclei isolation buffer is critical to mechanically stabilize isolated nuclei, protect their DNA from degradation and provide conditions needed for specific and stoichiometric DNA staining (39,40). Recent data have further revealed a role for the isolation buffer—to suppress the negative effect of cytoplasmic compounds on DNA staining and avoid precipitation of cellular debris (41). A list of the most frequently used buffers can be found in the FLOWer database (42). After the initial uncertainty on the suitability of DNA fluorochromes for the estimation of genome size, it became clear that only DNA intercalators such as ethidium bromide or propidium iodide, which do not exhibit preferential binding to AT or GC base pairs, should be used (43,44). Once the DNA of nuclei in the solution is stained, the sample can be analyzed using flow cytometry for relative nuclear fluorescence intensity.

Cytometric measurements are always relative and a standard with known genome size is needed to estimate genome size of the unknown sample. In principle, this can be done so that the sample and the standard are measured independently and the means of fluorescence intensity of nuclei in the same phase of the cell cycle (typically in G1) are compared. However, this so called external standardization is not recommended, especially in high-precision measurements, as both samples are not treated under identical condition. For the same reason, it is not advisable to mix two independently prepared samples prior to the analysis (45). It has been generally accepted that only the so called internal standardization, when the nuclei of the standard and the unknown sample are isolated, stained, and analyzed simultaneously, yields reliable results (40,45–47).

DNA Reference Standards

As genome size of the unknown sample is estimated after a comparison with a reference standard, calibration of the standard's genome size is of prime importance. It will establish the absolute precision of the estimate and a possibility to compare the estimate with those obtained in other laboratories and with other samples. In principle, nuclei isolated from animals and plants should be used interchangeably. However, as the staining of nuclear DNA is influenced by chromatin structure and DNA accessibility, it has been recommended not to use animal standards, such as chicken (*Gallus gallus*) red blood cells, for plants (Key recommendation 3 at the Angiosperm Genome Size Discussion Meeting and Workshop 1997, see 35). But, as it is useful to compare genome size in animals and plants, there is a need for calibrating their standards against each other and a common primary reference standard is desirable.

As there has never been an agreement on DNA reference standards, various animal and plant species have been used with a risk of producing data, which cannot be compared. Even worse, as there is no agreement on the genome size of reference standards, the same standards have been used in different studies with different DNA amounts assigned to them (see 48). Ideally, only one reference standard should be used for all estimations. However, this is not realistic as the range of known genome sizes in plants and animals extends three orders of magnitude (17,18). As the use of only one standard for the whole range would bring a risk of error due to nonlinearity (49), a set of reference standards is needed to cover the range. A logical strategy is to calibrate a primary ("gold") reference standard and then perform a series of experiments to calibrate other (secondary) reference standards. As the calibration needs to be done in a step-wise manner, the number of steps from the primary reference to the secondary standard is important (44) and the genome size of the primary reference should be in the middle of the range. In land plants, 2C values range from 0.12 to 254.8 pg, with the median 2C-value 5.06 pg and mean 2C-value 12.38 pg (18, I. J. Leitch, pers. comm.).

Human as a Primary Reference Standard

A number of studies indicate that the genome of man falls close to the median 2C-value of plants and thus would appear to be an excellent primary reference standard. However, do we know its genome size with enough precision? Vendrely and Vendrely (50) extracted DNA from a known number of human liver nuclei by modified Schmidt-Thannhauser and Schneider procedures and determined purin content colorimetrically with the Dische diphenylamine reaction. Estimated nuclear DNA contents (supposedly representing 2C-values) were 6.3, 5.9, and 5.8 pg, with the average of 6.0 pg DNA. Using similar methods, Mandel et al. (51) investigated nuclear DNA content of human leucocytes, which are diploid and unreplicated, and reported a value of 6.848 pg/2C. Although this value differs from that of Vendrely and Vendrely (50), it seems to indicate that the liver nuclei analyzed by these authors were mainly diploid and unreplicated. During the same period, Métais et al. (52) obtained a value of 6.98 pg (2C) from 14 determinations using leucocytes from normal humans. When considering these estimates, one should bear in mind that their precision depends not only on the reliability of the chemical assays themselves, but also on the presence of cells in different phases of the cell cycle and the precision with which the number of cells is determined.

On the basis of the analysis of a number of previous estimates for human genome size, Tiersch et al. (53) assigned a value of 7.0 pg DNA for fresh male human leukocytes. Using this value they calibrated a set of 45 animal species with 2C-values ranging from 1.5 pg to 110.0 pg DNA as reference standards for flow cytometry. Although the human 2C-value of 7.0 pg was set rather arbitrarily, it provided a solid platform to calibrate a range of secondary standards. The 2C-value of 2.5 pg determined by Tiersch et al. (53) for chicken was in a perfect agreement with the value determined by Rasch et al. (54) and other authors cited by them and independently confirmed the 2C-value of 7.0 pg for man. This 2C-value is commonly assumed today as a consensual 2C-value of the human (17,55–57) and was used to calibrate a set of recommended plant DNA reference standards (58, Table 2).

However, a detailed analysis of published data indicates that 2C = 7.0 pg for human seems to be close to the upper realistic limit. For example, a circumstantial evidence for a lower human 2C-value comes from three independent studies, in

Table 2. A set of plant DNA reference standards calibrated using human male leukocytes (2C = 7.0 pg) as primary reference standard[a]

SPECIES	CULTIVAR	2C DNA CONTENT	REFERENCE
Raphanus sativus L.	Saxa	1.11	43
Solanum lycopersicum L.	Stupické polní rané	1.96	43
Glycine max Merr.	Polanka	2.50	90
Zea mays L.	CE-777	5.43	84
Pisum sativum L.	Ctirad	9.09	44
Secale cereale L.	Daňkovské	16.19	44
Vicia faba L.	Inovec	26.90	43
Allium cepa L.	Alice	34.89	44

[a] Seeds may be obtained free of charge by contacting the corresponding author at dolezel@ueb.cas.cz.

which a model plant *Arabidopsis thaliana*, a model nematode *Caenorhabditis elegans*, chicken, and human were used. Galbraith et al. (36) chemically determined the chicken 2C-value as 2.33 ± 0.22 pg. After simultaneous analysis of *A. thaliana* and chicken nuclei, Bennett et al. (59) arrived at a 2C-value of 2.233 pg for chicken (in this study *A. thaliana* was calibrated against *C. elegans* assuming its 1C value = 0.102 pg). Given the human/chicken ratio of 0.357 as determined by Tiersch et al. (53), one arrives at a 2C-value of 6.252 pg for human (Ref. 45).

In addition to colorimetry, Feulgen microdensitometry and flow cytometry, other methods have been employed to determine the genome size of human and are worth mentioning. For example, Sandritter et al. (60) determined 3.12 pg DNA for RNase-treated human sperm using UV-cytophotometry, which fits within the range of acceptable genome size estimates. In contrast, genome size estimated using DNA reassociation kinetics seems to be too low to be considered reliable: 1.8 Gbp or 1.84 pg/2C (61,62). The error was almost certainly due to sequence complexity of the human genome with the presence of various classes of repetitive DNA. More recently, Wilhelm et al. (29) compared genome sizes of yeast, fish *Xiphophorus maculatus* and human (of Caucasian race, sex not given) and estimated the human genome size to be 2.9 Gbp/1C, which amounts to 2.97 pg. However, as the conversion factor by Britten and Davidson (26) was used, Wilhelm et al. (29) actually reported 1C = 3.178 pg. Altogether, these results may indicate that the human 2C-value is lower than 7.0 pg and higher than 6.0 pg.

Using a Sequenced Reference Standard

An ideal scenario is to use a DNA reference standard whose genome has been sequenced to completion. The progress in sequencing technologies seems to make this proposal realistic and an uninformed observer may get the impression that in species, which were announced to be fully sequenced, the complete sequences are known. Unfortunately, this is not true. No matter what reasons prompted research teams to announce the completion of their projects, the reality is that the largest eukaryotic genome for which an essentially complete sequence is known is that of the model nematode *C. elegans*. One copy (1C genome size) occupies 100.3 millions base pairs (63) and this seems to be too low to qualify *C. elegans* as the primary reference standard to be used for calibration of secondary reference standards.

The human genome sequence was declared complete in 2004 (64). However, due to the presence of peculiar DNA sequences, mainly repetitive DNA which are difficult to map and sequence, the sequence was interrupted by 341 gaps, which were estimated to represent 225 Mb. Together with the 2.85 Gb finished sequence, this gives an estimated human genome size 3.1 Gb. However, the efforts to close the gaps continue (65,66) and the exact human genome size will not be known until this work is completed. At the same time, one has to be aware of the variation in the human genome, which may result in differences in genome size between individuals (67).

The genome of chicken is another relevant animal genome, which has been sequenced. However, once again the sequence is far from complete and contains a large number of gaps (68). Thus, the genome size of 1.05 Gb (1C) as reported by the International Chicken Genome Sequencing Consortium (69) is certainly an underestimate. A comparison with the situation in the human genome indicates that it is not realistic to expect that the gaps in the chicken genome will be closed in the near future. The same holds true for other animal species potentially useful as primary reference standards.

Calibrating Plant Reference Standards

The FLOWer database developed by Loureiro et al. (48) provides a list of reference standards which have been used to determine genome size in plants using flow cytometry. However, only a few were calibrated using an independent method. Chemical genome size determination in multicellular plants is complicated by the cumbersome determination of cell number and by DNA synthetic activity and the occurrence of endopolyploidy. Onion (*Allium cepa*) is suitable for chemical determination of nuclear DNA content because of a near absence of endopolyploidy in root tip meristems and adjacent parts of the root. Sparrow and Miksche (70) determined cell numbers in root tips and performed chemical determination of DNA using essentially the same methods as previously used for human (50–52, see also Ref. 71 for describing the methods used in plants). The initial value of 54.3 pg DNA / cell, which was determined on DNA isolated from a population of cells containing nuclei in G1, S and G2-phase, was subsequently calibrated by Van't Hof (72) after considering the relative duration of cell cycle phases. The resulting 2C-value of 33.55 pg, or 33.5 pg for simplicity (31), holds till today as a reasonable estimate for this species. Greilhuber et al. (73) compared *A. cepa* with the human and a number of animal species using Feulgen two-wavelength cytophotometry. The estimates for mitotic blood cells of a male and a female were 6.08 and 6.38 pg (2C), respectively, or 6.22 pg on average. This result was in excellent agreement with the early chemical determinations of the human (50–52) and provided a reciprocal confirmation of the correctness of the 2C value for *A. cepa*.

For a long time, *A. cepa* was the preferred standard for DNA measurements in plants, but is now increasingly replaced by other species such as garden pea (*Pisum sativum*), which has desirable characteristics such as moderate genome size, genome size constancy over all cultivars, absence of interfering metabolites in leaf tissues, continuous availability and ease of growth. Although the genome size of *P. sativum* was reported to be constant (74,75), published C-values are surprisingly variable. For the purpose of this review, we consider only those studies in which *P. sativum* was compared with *A. cepa* (Table 1). Obviously, the published C-values are of different weight, depending on the number of lines studied, number of replicates and the methods used. The estimates for genome size of *P. sativum* vary by 1.375-fold and the median of genome size ratio *P. sativum*/*A. cepa* is 0.2633 (rank 18 of 35). Using the traditional 2C-value of 33.5 pg for *A. cepa* (31) this gives 2C = 8.82 pg for *P. sativum*, a value close to 8.84 pg estimated using Feulgen densitometry (75).

Like the densitometric estimates of genome size ratio *P. sativum*/*A. cepa*, also the flow cytometric estimates vary between laboratories (76,77, Table 1). Considering the human-based 2C-value of 9.09 pg for *P. sativum* (44), 2C-value of *A. cepa* can be estimated to be 34.52 pg, i.e. only 1.030-fold higher than the traditional value of 33.5 pg. These data seem to indicate that the ratio of genome size *P. sativum*/*A. cepa* is close to 0.2633 and that the 2C-value of *A. cepa* is probably slightly higher than 33.5 pg, i.e., about 34.52 pg, if 2C = 9.09 pg is considered for *P. sativum*—a value determined assuming 2C = 7.0 pg DNA for human (44). The use of 2C = 9.09 pg or 8.89 Gbp for *P. sativum* makes the estimates of genome size obtained in plants directly comparable with those in animals, which are based mostly directly or indirectly on the human value of 2C = 7.0 pg. Irrespective of which C-value authors consider for their plant standard, experimentally determined C-value should always be accompanied by C-value of reference standard to allow for subsequent recalculations.

Will Genome Sequencing in Plants Help?

A considerable number of 160 plant genomes (out of them 149 land plants) and 228 animal (metazoa) genomes are reported to be presently sequenced or are projected to be (National Center for Biotechnology Information http://www.ncbi.nlm.nih.gov/). However, none of them can be considered fully sequenced, including the small genome of the model plant *A. thaliana*. The Arabidopsis Genome Initiative (78) estimated its genome size to be 125Mb. This value was based on the size of all sequenced regions together with an estimate of the size of nonsequenced regions. The current estimate, which again includes an estimate of nonsequenced regions like centromeres is 135 Mb (The Arabidopsis Information Resource http://www.arabidopsis.org/). It is interesting to note that this estimate differs only by 8.6% from 157 Mb as determined after the comparison with *C. elegans* (59). The reality is that a complete sequence and genome size of *A. thaliana* is currently not known and one cannot exclude a possibility that it is close to 150 Mb.

A similar situation is found in other sequenced plant genomes, which include rice (*Oryza sativa*). The International Rice Genome Sequencing Project (79) produced 389 Mbp sequence, which is supposed to cover virtually all the euchromatin, but only some heterochromatin regions. The same is true for sorghum (*Sorghum bicolor*) with an estimated 1C-value of 730 Mbp (80,81). No matter if the genomes of rice and sorghum are sequenced or not, their genome sizes appear at the lower end to be ideal primary reference standards. This role could be played by maize (*Zea mays*). However, its large and complex genome makes efforts to obtain a complete sequence especially difficult (82). It is interesting to note that the progress of sequencing in maize has been measured against the genome size of 1C = 2,300 Mbp estimated by Rayburn et al. (83) using flow cytometry. However, higher estimates have been published (1C = 2,628 Mbp, 84) indicating that a larger part of the maize genome is yet to be sequenced.

Conclusion

In the near future, genome size will be known with sufficient precision only for a minority of eukaryotes, mainly those with very small genomes. As they are at the low end of the range of genome size of plants and animals, they are not ideal primary reference standards. Thus, it seems reasonable to continue using human as a primary reference in the animal kingdom and *P. sativum* as a primary reference in plant kingdom. To prevent introduction of another source of variation and a chaos in the literature, it is advisable to continue using reference standards calibrated against the human male with an assigned 2C-value of 7.00 pg DNA. Although this 2C-value is probably overestimated by 5–10%, this approach offers an important advantage by providing a link and permitting comparisons between the estimates made both in the plant and animal kingdom. Once the human genome size is known with enough precision, all values assigned to reference standards and estimations made with them can be easily recalculated. There is a demand to obtain the full genome sequence for human and it is realistic to expect its availability in the not too distant future. On the other hand, economic constraints will hamper the attempts to obtain complete genome sequences in other eukaryotes with larger genomes in the near future.

Acknowledgments

We thank Dr. I. J. Leich for the statistical analysis of data deposited in the Plant DNA C-values Database (18) and we appreciate useful suggestions of two anonymous reviewers.

Literature Cited

1. Avery OT, MacLeod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med 1944;79:137–158.
2. Boivin A, Vendrely R, Vendrely C. L'acide désoxyribonucléique du noyau cellulaire, dépositaire des caractères héréditaires; arguments d'ordre analytique. Compt Rend Acad Sci Paris 1948;226:1061–1063.
3. Vendrely R, Vendrely C. La teneur du noyau cellulaire en acide désoxyribonucléique à travers les organes, les individus et les espèces animales. Techniques et premiers résultants. Experientia 1948;4:434–436.
4. Ris H, Mirsky AE. Quantitative cytochemical determination of desoxyribonucleic acid with the Feulgen nucleal reaction. J Gen Physiol 1949;33:125–146.
5. Swift HH. The desoxyribose nucleic acid content of animal nuclei. Physiol Zool 1950;23:169–198.

6. Swift HH. The constancy of desoxyribose nucleic acid in plant nuclei. Proc Natl Acad Sci USA 1950;36:643–654.
7. Stowell RE. Thymonucleic acid in tumors. Cancer Res 1945;5:283–294.
8. Stowell RE. Nucleic acids in human tumors. Cancer Res 1946;6:426–435.
9. Mirsky AE, Ris H. The deoxyribonucleic acid content of animal cells and its evolutionary significance. J Gen Physiol 1951;34:451–462.
10. Macgregor HC, Uzzell TM. Gynogenesis in salamanders related to *Ambystoma jeffersonianum*. Science 1964;143:1043–1045.
11. Rees H, Cameron FM, Hazarika MH, Jones GH. Nuclear variation between diploid angiosperms. Nature 1966;211:828–830.
12. Southern DI. Species relationships in genus *Tulipa*. Chromosoma 1967;23:80–94.
13. Hinegardner R. Evolution of cellular DNA content in teleost fishes. Am Nat 1968;102:517–523.
14. Laird CD, McCarthy BJ. Molecular characterization of the Drosophila genome. Genetics 1969;63:865–882.
15. Thomas CA Jr. The genetic organization of chromosomes. Ann Rev Genet 1971; 5:237–256.
16. Gregory TR. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol Rev 2001;76:65–101.
17. Gregory TR. Animal genome size database. 2009. Available at: http://www.genomesize.com.
18. Bennett MD, Leitch IJ. Plant DNA C-values database (release 4.0, October 2005). Available at: http://data.kew.org/cvalues/.
19. Bennett MD. Plant genome values: How much do we know? Proc Natl Acad Sci USA 1998;95:2011–2016.
20. Greilhuber J. Intraspecific variation in genome size: A critical reassessment. Ann Bot 1998;82(Suppl A):27–35.
21. Greilhuber J. Intraspecific variation in genome size in angiosperms: Identifying its existence. Ann Bot 2005;95:91–98.
22. Winkler H. Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche. Jena: Gustav Fischer Verlag; 1920.
23. Wolf U, Ritter H, Atkin NB, Ohno S. Polyploidization in the fish family Cyprinidae. Order Cypriniformes. I. DNA-content and chromosome sets in various species of Cyprinidae. Humangenetik 1969;7:240–244.
24. Greilhuber J, Doležel J, Lysák MA., Bennett MD. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. Ann Bot 2005;95:255–260.
25. Greilhuber J, Doležel J. 2C or not 2C: A closer look at cell nuclei and their DNA content. Chromosoma 2009;118:391–400.
26. Britten RJ, Davidson EH. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. Quart Rev Biol 1971;40:111–133.
27. Nagl W. Zellkern and Zellzyklen. Stuttgart: Verlag Eugen Ulmer; 1976. p 248.
28. Rasch EM. DNA "standards" and the range of accurate DNA estimates by Feulgen absorption microspectrophotometry. Adv Microbiol 1985;196:137–166.
29. Wilhelm J, Pingoud A, Hahn M. Real-time PCR-based method for the estimation of genome sizes. Nucl Acids Res 2003;31:e56.
30. Straus NA. Comparative DNA renaturation kinetics in amphibians. Proc Natl Acad Sci USA 1971;68:799–802.
31. Bennett MD, Smith JB. Nuclear DNA amounts in angiosperms. Phil Trans Roy Soc London B 1976;274:227–274.
32. Cavalier-Smith T. The Evolution of Genome Size. Chichester, New York, Brisbane: Wiley; 1985.
33. Doležel J, Bartoš J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. Cytometry Part A 2003;51A:127–128.
34. Thomas RA, Krishan A, Robinson DM, Sams C, Costa F. NASA/American Cancer Society high-resolution flow cytometry project. I. Cytometry 2001;43:2–11.
35. Bennett MD, Leitch IJ. Nuclear DNA amounts in angiosperms: Progress, problems and prospects. Ann Bot 2005;95:45–90.
36. Galbraith DW, Harkins KR, Maddox JM, Ayres NM, Sharma DP, Firoozabady E. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Science 1983;220:1049–1051.
37. Matzk F, Meister A, Schubert I. An efficient screen for reproductive pathways using mature seeds of monocots and dicots. Plant J 2000;21:97–108.
38. Roberts AV. The use of bead beating to prepare suspensions of nuclei for flow cytometry from fresh leaves, herbarium leaves, petals and pollen. Cytometry Part A 2007;71A:1039–1044.
39. Doležel J. Flow cytometric analysis of nuclear DNA content in higher plants. Phytochem Anal 1991;2:143–154.
40. Doležel J, Bartoš J. Plant DNA flow cytometry and estimation of nuclear genome size. Ann Bot 2005;95:99–110.
41. Loureiro J, Rodriguez E, Doležel J, Santos C. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. Ann Bot 2006;98:679–689.
42. Loureiro J, Rodriguez E, Santos C, Doležel J, Suda J. FLOWer: A plant DNA flow cytometry database (release 1.0, May 2008). Available at: http://flower.web.ua.pt/.
43. Doležel J, Sgorbati S, Lucretti S. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content estimation in plants. Physiol Plant 1992;85:625–631.
44. Doležel J, Lucretti S, Meister A, Lysák M, Nardi L, Obermayer R. Plant genome size estimation by flow cytometry: Inter-laboratory comparison. Ann Bot 1998;82(Suppl A): 17–26.
45. Greilhuber J, Temsch EM, Loureiro JCM. Nuclear DNA content measurement. In: Doležel J, Greilhuber J, Suda J, editors. Flow Cytometry with Plant Cells. Weinheim: Wiley-VCH Verlag; 2007. pp 67–101.
46. Greilhuber J. Cytochemistry and C-values: The less-well-known world of nuclear DNA amounts. Ann Bot 2008;101:791–804.
47. Loureiro J, Rodriguez E, Doležel J, Santos C. Flow cytometric and microscopic analysis of the effect of tannic acid on plant nuclei and estimation of DNA content. Ann Bot 2006;98:515–527.
48. Loureiro J, Suda J, Doležel J, Santos C. FLOWer: A plant DNA flow cytometry database. In: Doležel J, Greilhuber J, Suda J, editors. Flow cytometry with plant cells. Weinheim: Wiley-VCH Verlag; 2007. pp 423–438.
49. Vindelov LL, Christensen IJ, Nissen NI. Standardization of high-resolution flow cytometric DNA analysis by the simultaneous use of chicken and trout red blood cells as internal reference standards. Cytometry 1983;3:328–331.
50. Vendrely R, Vendrely C. La teneur du noyau cellulaire en acide désoxyribonucléique à travers les organes, les individus et les espèces animales. Etude particulière des Mammifères. Experientia 1949;5:327–329.
51. Mandel P, Métais P, Cuny S. Les quantités d'acide désoxypentose-nucléique par leucocyte chez diverses espèces de Mammifères. C R Acad Sci Paris 1950;231:1172–1174.
52. Métais P, Cuny S, Mandel P. La teneur en acide pentose-nucléique des leucocytes chez diverses espèces de Mammifères et en particulier chez l'Homme. C R Soc Biol 1951;145:1235–1238.
53. Tiersch TR, Chandler RW, Wachtel SS, Elias S. Reference standards for flow cytometry and application in comparative studies of nuclear DNA content. Cytometry 1989;10:706–710.
54. Rasch EM, Barr HJ, Rasch RW. DNA content of sperm of *Drosophila melanogaster*. Chromosoma 1971;33:1–18.
55. Rasch EM. Feulgen-DNA cytophotometry for estimating C values. In: Henderson DS, editor: Drosophila Cytogenetic Protocols. Totowa, NJ: Humana Press. Meth Mol Biol 2004;247:163–201.
56. Vinogradov AE. Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. Cytometry 1998;31:100–109.
57. Krishan A, Dandekar P, Nathan N, Hamelik R, Miller C, Shaw J. DNA index, genome size, and electronic nuclear volume of vertebrates from the Miami Metro Zoo. Cytometry Part A 2005;65A:26–34.
58. Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. Nature Protocols 2007;2:2233–2244.
59. Bennett MD, Leitch IJ, Price HJ, Johnston JS. 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 2003;91:547–557.
60. Sandritter W, Müller D, Gensecke O. Ultraviolettspektrophotometrische Messungen des Nukleinsäuregehaltes von Spermien und diploiden Zellen. Acta Histochem 1960;10:139–154.
61. Schmid CW, Deininger PL. Sequence organization of the human genome. Cell 1975;6:345–358.
62. Lewin B.Gene expression. Eucaryotic Chromosomes, 2nd ed., Vol. 2. New York, Chichester, Brisbane, Toronto: Wiley; 1980.
63. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, Coulson A, D'Eustachio P, Fitch DH, Fulton LA, Fulton RE, Griffiths-Jones S, Harris TW, Hillier LW, Kamath R, Kuwabara PE, Mardis ER, Marra MA, Miner TL, Minx P, Mullikin JC, Plumb RW, Rogers J, Schein JE, Sohrmann M, Spieth J, Stajich JE, Wei C, Willey D, Wilson RK, Durbin R, Waterston RH. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. PLoS Biol 2003;1:166–192.E45.
64. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature 2004;431:931–945.
65. Bovee D, Zhou Y, Haugen E, Wu ZN, Hayden HS, Gillett W, Tuzun E, Cooper GM, Sampas N, Phelps K, Levy R, Morrison VA, Sprague J, Jewett D, Buckley D, Subramaniam S, Chang J, Smith DR, Olson MV, Eichler EE, Kaul R. Closing gaps in the human genome with fosmid resources generated from multiple individuals. Nature Genet 2008;40:96–101.
66. Cole CG, McCann OT, Collins JE, Oliver K, Willey D, Gribble SM, Yang FT, McLaren K, Rogers J, Ning ZM, Beare DM, Dunham I. Finishing the finished human chromosome 22 sequence. Genome Biol 2008;9:R78.
67. Zhang F, Gu WL, Hurles ME, Lupski JR. Copy number variation in human health, disease, and evolution. Ann Rev Genomics Hum Genet 2009;10:451–481.
68. Gao F, Zhang CT. Isochore structures in the chicken genome. FEBS J 2006;273:1637–1648.
69. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature 2004;432:695–716.
70. Sparrow AH, Miksche JP. Correlation of nuclear volume and DNA content with higher plant tolerance to chronic radiation. Science 1961;134:282–283.
71. Baetcke KP, Sparrow AH, Nauman CH, Schwemmer SS. The relationship of DNA content to nuclear and chromosome volumes and to radiosensitivity (LD_{50}). Proc Natl Acad Sci USA 1967;58:533–540.
72. Van't Hof J. Relationships between mitotic cycle duration. S period duration and the average rate of DNA synthesis in the root meristem cells of several plants. Exp Cell Res 1965;39:48–58.
73. Greilhuber J, Volleth M, Loidl J. Genome size of man and animals relative to the plant *Allium cepa*. Can J Genet Cytol 1983;25:554–560.
74. Baranyi M, Greilhuber J. Flow cytometric and Feulgen densitometric analysis of genome size variation in *Pisum*. Theor Appl Genet 1996;92:297–307.

75. Greilhuber J, Ebert I. Genome size variation in *Pisum sativum*. Genome 1994;37:646–655.

76. Zonneveld BJM, Leitch IJ, Bennett MD. First nuclear DNA amounts in more than 300 angiosperms. Ann Bot 2005;96:229–244.

77. Johnston JS, Bennett MD, Rayburn AL, Galbraith DW, Price HJ. Reference standards for determination of DNA content of plant nuclei. Am J Bot 1999;86:609–613.

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

79. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. Nature 2005;436:793–800.

80. Sasaki T, Antonio BA. Plant genomics: Sorghum in sequence. Nature 2009;457:547–548.

81. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS. The *Sorghum bicolor* genome and the diversification of grasses. Nature 2009;457:551–556.

82. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK. The B73 Maize genome: Complexity, diversity, and dynamics. Science 2009;326:1112–1115.

83. Rayburn AL, Biradar DP, Bullock DG, McMurphy LM. Nuclear-DNA content in F1 hybrids of maize. Heredity 1993;70:294–300.

84. Lysák MA, Doležel J. Estimation of nuclear DNA content in *Sesleria* (Poaceae). Caryologia 1998;52:123–132.

85. Ingle J, Sinclair J. Ribosomal RNA genes and plant development. Nature 1972;235: 30–32.

86. Olszewska MJ, Osiecka R. The relationship between 2C DNA content, life cycle type, systematic position and the dynamics of DNA endoreplication in parenchyma nuclei during growth and differentiation of roots in some dicotyledonous herbaceous species. Biochem Physiol Pflanzen 1983;178:581–599.

87. Arumuganathan K, Earle ED. Nuclear DNA content of some important plant species. Plant Molec Biol Rep 1991;9:208–218.

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

89. Vilhar B, Greilhuber J, Dolenc Koce J, Temsch EM, Dermastia M. Plant genome size measurement with DNA image cytometry. Ann Bot 2001;87:719–728.

90. Doležel J, Doleželová M, Novák FJ. Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). Biol Plant 1994;36:351–357.