

Letters

The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers

Introduction

For nearly a century, biologists, and botanists in particular, have been interested in the determination and documentation of chromosome numbers for extant taxa (reviewed in Goldblatt & Lowry, 2011) as well as extinct ones (Laane & Hoiland, 1986; Masterson, 1994). These data have been widely used to evaluate the evolutionary pattern of chromosome number change and to estimate the base chromosome number of clades of interest. Chromosome numbers have also been extensively utilized as an important phylogenetic character in the context of cytotaxonomy (Chatterjee & Kumar Sharma, 1969; Schlarbaum & Tsuchiya, 1984; Guerra, 2012). Perhaps the most influential use of chromosome number data has been in the inference of major genomic events such as whole genome duplications (polyploidy), as well as changes in single chromosome numbers (e.g. dysploidy). Early researchers analyzed the distribution of chromosome numbers within a group of interest and employed various threshold techniques to estimate ploidy levels for the analyzed taxa (Stebbins, 1938; Grant, 1963; Goldblatt, 1980). More recently, phylogenetic information was incorporated into the analyses, allowing researchers to infer transitions in chromosome numbers along branches of the tree using either the maximum parsimony principle (Schultheis, 2001; Hansen *et al.*, 2006; Ohi-Toma *et al.*, 2006; Wood *et al.*, 2009) or by using a probabilistic evolutionary model within the likelihood paradigm (Mayrose *et al.*, 2010; Cusimano *et al.*, 2012; Glick & Mayrose, 2014).

Due to their significance and the relative ease by which chromosome numbers can be obtained, it is not surprising that chromosome number is the most extensively and consistently recorded cytological property in most plant families and genera (Guerra, 2008). These data have been documented along the years in an array of journal manuscripts, printed books (Löve & Löve, 1948; Darlington & Wylie, 1955; Fedorov, 1969) and, more recently, in the form of online databases (Goldblatt & Johnson, 1979; Watanabe, 2002; Bennett & Leitch, 2011). To date, the most comprehensive data source is the Index to Plant Chromosome Numbers (IPCN; Goldblatt & Johnson, 1979), which provides reference point to original chromosome counts reported in the literature. IPCN was initially established at the University of

California Berkeley in the 1950s and was later maintained by Canada Department of Agriculture, Missouri Botanical Garden, and currently by the International Association for Plant Taxonomy (IAPT). A large portion of the counts referenced during 1979–2006, the years that IPCN has been housed in the Missouri Botanical Garden, can be accessed and searched online. Counts reported in more recent years are currently published under IAPT/IOPB Chromosome Data series (Marhold, 2006) but are not stored within a central, easily searched, database. In addition to IPCN, several other online data sources are available, most of which are dedicated to either a specific geographical region (Slovakia – Marhold *et al.*, 2007; Poland – Góralski *et al.*, 2009 onwards) or to a certain taxonomic group (e.g. Hieracium – Schuhwerk, 1996; Asteraceae – Watanabe, 2002).

The amount of chromosome counts that exist to date is extensive, and searching the large number of resources that contain such information is a daunting task, particularly when a large number of taxa is examined. Consequently, many researchers search for chromosome number information only through the largest online database(s), while smaller but nonetheless valuable sources are ignored. This usually results in missing data for some of the species in question, which may lead to erroneous conclusions drawn from the analysis. Obviously, a large accessible database that unifies all currently known databases, including both printed and online sources, would be of great value to the botanical community and would make the task of data collection much easier. In addition, such a central resource would enable researchers to add new counts as soon as they are being reported, facilitating the task of data sharing.

Here, we present the Chromosome Counts Database (CCDB), as a community resource of plant chromosome numbers. The database incorporates data from dozens of sources, more than doubling the amount of data available within any single resource. The online database additionally enables researchers to add new counts or to comment on existing data entries, thereby facilitating data sharing. The extensive amount of data currently available in CCDB further allowed us to analyze the patterns of chromosome number distribution among major plant groups. We estimate the percentage of plant species exhibiting intraspecific variation in chromosome numbers as well as in their ploidy levels.

Materials and Methods

Data collection

Chromosome counts were collected from a large number of electronic resources, older chromosome counts compendiums in the form of printed books, and an array of miscellaneous sources such as floras, monographs and other scientific manuscripts. The full list of resources is given in Table 1. Data from these sources were collected using the following procedures:

Online chromosome counts databases Data from several online databases were retrieved directly from the database curator via personal communication in the form of comma-separated value (CSV) files. These include data from the *Plant DNA C-values database* (Bennett & Leitch, 2011; obtained from Ilia Leitch) and *Chromosome number database of Polish plants* (Góralski *et al.*, 2009 onwards; obtained from Grzegorz Góralski). Other online chromosome counts databases were downloaded and processed using Perl/Python scripts. The following online sources were retrieved: *IPCN* (Goldblatt & Johnson, 1979–), *Chilean plants cytogenetic database* (Jara-Seguel & Urrutia, 2011), *CHROBASE – Chromosome numbers for the Italian flora* (Bedini *et al.*, 2010 onwards), *BSBI cytology database* [accessed 20 June 2013] (<http://rbg-web2.rbge.org.uk/BSBI/cytsearch.php>), *Index to chromosome*

numbers in Asteraceae (Watanabe, 2002), *Published chromosome counts in Hieracium* (Schuhwerk, 1996), *ChromoPar – Paraguay chromosome counts database* [accessed 12 June 2013] (<http://www.ub.edu/botanica/cromopar/>), *Karyological database of the genus Cardamine* (Kucera *et al.*, 2005) and *Chromosome number survey of the ferns and flowering plants of Slovakia* (Marhold *et al.*, 2007).

Chromosome counts compendiums available as hard copy In addition to online sources as already described, we have obtained well-known and widely used printed books containing chromosome counts indexes. The data in these books were retrieved in the following way: first, the books were scanned to generate image files. Then, using the optical character recognition (OCR) tool of Adobe Pro the files were converted to 'textable' PDF files. This OCR tool

Table 1 Chromosome counts resources incorporated in Chromosome Counts Database (CCDB)

Resource name	Data entries	Unique species names	Unique resolved names
(a) Online resources			
IPCN www.tropicos.org/Project/IPCN	111 224	60 167	48 829
Plant DNA C-values database http://data.kew.org/cvalues	5889	5614	5306
BSBI cytology (British Isles) http://rbg-web2.rbge.org.uk/BSBI/cytsearch.php	2766	1430	1356
Chilean Plants Cytogenetic Database http://www.chileanpcd.com	224	165	154
ChromoPar (Paraguay) http://www.ub.edu/botanica/cromopar	1278	513	303
CHROBASE (Italy) http://www.biologia.unipi.it/chrobase	6517	2983	2563
Slovakia's Karyological database http://www.chromosomes.sav.sk	7734	2541	2322
Chromosome number database of Polish plants http://chromosomes.binoz.uj.edu.pl	2183	1615	1517
Index to chromosome numbers in Asteraceae http://www.lib.kobe-u.ac.jp/infolib/meta_pub/G0000003asteraceae_e	23 800	20 087	13 102
Karyological database of the genus <i>Cardamine</i> http://www.cardamine.sav.sk/	2966	107	90
Published chromosome-counts in <i>Hieracium</i> http://www.botanischestaatsammlung.de/projects/chrzlit.html	356	261	206
(b) Hard-copy resources			
Cytotaxonomical atlas of the Pteridophyta (Löve <i>et al.</i> , 1977)	4216	1826	1763
Chromosome atlas of flowering plants (Darlington & Wylie, 1955)	11 741	10 773	10 009
Chromosome Numbers of Flowering Plants (Fedorov, 1969)	30 544	16 117	14 326
Chromosome atlas of flowering plants of the Indian subcontinent Vol1 (Kumar & Subramaniam, 1987a)	7409	4902	4099
Chromosome Atlas of Flowering Plants of the Indian Subcontinent Vol2 (Kumar & Subramaniam, 1987b)	4497	2367	1780
Index to plant chromosome numbers, 1965 (Ornduff, 1967)	3900	3776	3478
Index to plant chromosome numbers, 1966 (Ornduff, 1968)	3836	3771	3490
Index to plant chromosome numbers, 1967–1971 (Moore, 1973)	21 996	18 833	15 002
Index to plant chromosome numbers, 1972 (Moore, 1974)	3622	3457	3185
Index to plant chromosome numbers, 1973–1974 (Moore, 1977)	10 243	9140	7768
Index to plant chromosome numbers, 1975–1978 (Goldblatt, 1981)	12 525	9982	8696
Chromosome numbers of northern plant species (Löve & Löve, 1948)	1352	1125	1058
Flora Europaea – checklist and chromosome index (Moore, 1982)	4490	3974	3870
(c) Miscellaneous resources^a			
IAPT/IOPB chromosome data (Marhold, 2006)	4123	3182	2707
Eflora (http://www.efloras.org/)	11 405	11 405	10 890
Flora iberica (http://www.floraiberica.es/)	9603	5242	4118
Interactive flora of NW Europe (http://wbd.etibioinformatics.nl/bis/flora.php?)	2764	2657	2640
Araceae chromosome numbers (Cusimano <i>et al.</i> , 2012)	1026	844	740
Brassicaceae chromosome numbers (Warwick & Al-Shehbaz, 2006)	8685	1805	1687
Cyperaceae chromosome numbers (Roalson, 2008)	2818	814	698
Veroniceae chromosome numbers (Albach <i>et al.</i> , 2008)	2491	404	334
Chromosome atlas of the New Zealand flora (M. Dawson, pers. comm.)	2255	1947	1736
Solaneaceae chromosome numbers (E. Goldberg, pers. comm.)	2001	1438	846

^aOnly major resources are given. A full list of resources is available at <http://ccdb.tau.ac.il/about/>.

was chosen because it exhibited the most accurate performance compared to five other OCR tools in an initial screen of several books. In the next step we used 'Some PDF to Text Converter' (available through www.somepdf.com), which converted the PDF files into plain text files that could be parsed automatically using Python scripts. Because this whole automated process suffers from some inaccuracies – particularly due to errors related to the OCR conversion (e.g. occasional confusion between 'l', '1', and '!') – thousands of counts were manually verified. In addition, our general approach in processing such sources was to maximize retrieval accuracy rather than data completeness. Consequently, not all data available through the target source were retrieved. It should be emphasized that occasional errors may still remain (this is particularly so for the compendium published by Fedorov, 1969, for which OCR errors are more abundant due to the Cyrillic font and tables\columns included within the text) and CCDB allows users to report such cases. The following sources were retrieved this way: *Chromosome numbers of northern plant species* (Löve & Löve, 1948), *Chromosome atlas of flowering plants* (Darlington & Wylie, 1955), *Cytotaxonomical atlas of the Pteridophyta* (Löve *et al.*, 1977), *Chromosome numbers of flowering plants* (Fedorov, 1969), *Flora Europaea – checklist and chromosome index* (Moore, 1982), *Chromosome atlas of flowering plants of the Indian subcontinent*, volumes 1 and 2 (Kumar & Subramaniam, 1987a) and *Index to plant chromosome numbers for the years 1965–1974* (Ornduff, 1967, 1968; Moore, 1970, 1971, 1973, 1974, 1977). The IPCN volume for the years 1975–1978 (Goldblatt, 1981) was also parsed but counts were inserted into the database only in case the online IPCN database did not already contain them.

Floras, journal manuscripts and other resources In addition to dedicated chromosome counts databases and hard copy books, a large number of other sources exist that contain information regarding the chromosome number for a given taxon. These resources include floras, monographs and an array of scientific manuscripts. However, automatic retrieval of chromosome number data from such resources is not a trivial task because the data are organized in a source-specific manner (e.g. the botanical description of a given species as appears in its relevant flora obtained through <http://www.efloras.org>). Hence, the downloading and processing of each data source were performed using dedicated Perl/Python scripts written specifically for each data source, followed by a manual verification of hundreds of records. As mentioned above, we preferred to maximize data accuracy over data completeness and therefore some fraction of the data available in these sources was not used.

Thousands of chromosome counts were acquired from online floras – eflora [accessed 20 October 2013] (<http://www.efloras.org>), Flora Iberica [accessed 20 June 2013] (<http://www.floraiberica.es>), and from the Interactive flora of NW Europe [accessed 20 June 2013] (<http://wbd.etibioinformatics.nl/bis/flora.php>). In addition to floras, chromosome counts that appear within several Systematic Botany Monographs were retrieved (Saunders, 2000; Bohs, 2001; Freire-Fierro, 2002; Aldasoro *et al.*, 2004; Zuloaga *et al.*, 2004; Thompson, 2005; Wagner *et al.*, 2005; Meudt, 2006; Miller & Chambers, 2006).

Scientific manuscripts that contain large amounts of chromosome counts were parsed in a source-specific manner and incorporated into the database. IAPT/IOPB Chromosome Data reports 1–16 (Marhold, 2006) were obtained from the International Organization of Plant Biosystematists website (<http://www.iopb.org/>) as PDF files, converted to text files and parsed using Perl scripts. In addition, a large number of journal manuscripts that contain counts for a given taxonomic group or geographic region were obtained and parsed in a source-specific procedure. These include data reported in a large number of Mediterranean chromosome number reports (Kamari *et al.*, 1991), as well as large collections available for Araceae (Cusimano *et al.*, 2012), Brassicaceae (Warwick & Al-Shehbaz, 2006), Colchicaceae (Chacón *et al.*, 2014), Cyperaceae (Roalson, 2008), *Pinguicula* (Casper & Stimper, 2009), and Veroniceae (Albach *et al.*, 2008). The full list of scientific manuscripts that were incorporated into CCDB is available through the database help pages (<http://ccdb.tau.ac.il/about/>).

Finally, chromosome counts datasets that were compiled by individual researchers were obtained via personal communication. These include chromosome numbers of indigenous New Zealand plants obtained from Murray Dawson and chromosome numbers for a large number of Solanaceae species obtained from Emma Goldberg.

Name resolution

Combining data from multiple sources required a method for standardization of the information, especially regarding the taxonomy of the records. Many plant species have been given different names by different authors. Some of these names are considered synonyms, others are recognized as accepted names, while another fraction is still unresolved. Another common problem is differences in spelling conventions between sources, or simply spelling mistakes, resulting from either manual typing errors in the original source, or incorrect processing of our automatic pipelines. To overcome these difficulties, we used Taxonome (Kluyver & Osborne, 2013), a taxonomic name resolution software that provides the ability to match synonymous taxon names to accepted names while accounting for differences in naming conventions and likely misspellings. As the underlying database for names, we used a local repository of synonymous and accepted names that was created based on The Plant List (TPL) v1.1 (<http://www.theplantlist.org/>) with some modifications (i.e. for Solanaceae we used Solanaceae Source (<http://solanaceaesource.org/>) as the primary taxonomic source supplemented with The Plant List for missing taxon names). In case a taxon name could not be matched to a recognized plant name (e.g. due to erroneous OCR processing), the corresponding data entry was excluded from the database.

Data access

CCDB is available through <http://ccdb.tau.ac.il/>. Users can access the data by browsing through the taxonomic hierarchy or by searching for a specific genus or species. At each level, all counts can

be retrieved as a CSV file. Additionally, users can access the data through the dedicated application programming interface (API), available through <http://ccdb.tau.ac.il/services/>. Researchers are invited to contribute to the completeness and correctness of the resource. This can be achieved by submitting new data, originating from resources not yet incorporated into the database as well as reporting errors found in the database. We note that unlike in IPCN, new data entries will not be thoroughly reviewed. Thus, data contributors are strongly encouraged to include supporting information such as voucher specimen or an image file of the cells analyzed.

Results

CCDB encompasses a wide array of resources, the majority of which were unavailable before in a digitized format. At present, CCDB contains 334 963 data entries, encompassing chromosome counts for 171 338 unique taxon names, including species names and infraspecific names. Following a taxonomic name resolution process that collapsed synonymous names to their accepted names, the number of unique names in CCDB is 77 958 (of these 68 146 are accepted names and 9812 are unresolved according to TPL V1.1). This represents a substantial increase in data coverage compared to IPCN – the largest online resource to date – that has information for a total of 60 167 plant names (48 829 following name resolution). Table 1 specifies the number of counts extracted from each source, as well as the number of unique names before and after name resolution.

CCDB includes a total of 8750 genera from 539 families. The coverage of CCDB varies widely across the major plant groups. The current coverage for angiosperms is 19% (58 980 out of 304 419 accepted species as reported in TPL V1.1 – not including data available for infraspecific names). The exact coverage may, however, vary between 12% and 23% depending on the assumed number of angiosperm species, with estimates ranging from 261 750 (Stevens, 2012) to 500 000 if yet undiscovered species are considered (as discussed in Galbraith *et al.*, 2011). The estimated coverage for pteridophytes (here and in the online database referred to as the monilophytes and lycophytes clades), bryophytes and gymnosperms is 22% (2350/10 620), 4% (1436/34 556) and 38% (427/1104), respectively. Within the 20 largest angiosperm families (Supporting Information Table S1), the best covered one is Apiaceae, with counts available for 42% of the taxa (1474 out of 3509), while the coverage for the largest plant family, the Compositae, is 32% (11 776 out of 36 700). Of the 20 largest families, the least covered one is Bromeliaceae with 7%. Our compilation also highlights some additional families where chromosome count data are particularly lacking and where additional efforts should be particularly beneficial. Some of the least represented families in CCDB include the Daltoniaceae (having only one count out of 328 accepted names), Vochysiaceae (1/225) and Calophyllaceae (1/131).

In order to estimate the completeness of the data obtained through CCDB compared to the maximal availability of chromosome count information (i.e. all counts ever reported in the literature), we compared the coverage of CCDB relative to that

obtained in five previous studies. Each of these studies assembled chromosome-number information in a detailed manner for a specific plant clade, and we thus regard those as approximately representing all available data for these groups (*Pinguicula* – Casper & Stimper, 2009; Araceae – Cusimano *et al.*, 2012; Solanaceae – E. Goldberg, pers. comm.; Colchicaceae – Chacón *et al.*, 2014; Danthonioideae – Linder & Barker, 2014). In these comparisons, we calculated the fraction of species in the reference dataset for which information exists in CCDB while considering data entries obtained from other resources only (because the data obtained from the above five studies were already incorporated in CCDB). As demonstrated in Table 2, for several clades, such as Araceae and Colchicaceae, data completeness of CCDB is very high, nearly reaching that obtained by meticulous manual searches. However, for other clades (i.e. *Pinguicula*) our data retrieval was not as complete, missing roughly half of the data that have been previously reported. Notably, even for the least covered group, data availability in CCDB constitutes a major improvement compared to what is currently available through IPCN (Table 2). These results emphasize the need for a community effort aimed towards improving accessibility to the vast amount of chromosome number information that has been determined over the years, but appears sporadically within scientific manuscripts and thus is regularly missed.

Using the chromosome counts data assembled in CCDB, we next examined the distribution of the haploid chromosome numbers within each of the major plant groups. In case more than one count was available for a certain taxon, the median was taken as the representative count. As has been previously observed in ferns (Otto & Whitton, 2000), there are more even haploid numbers than odd ones (across the whole database the median chromosome number for 42 161 taxa is even and for 33 317 it is odd; Table 3), resulting in a ‘saw-toothed’ pattern (Fig. S1a). As noted by Otto & Whitton (2000), this pattern can be explained by frequent polyploidization events, because a genome duplication will always result in an even number while other changes in chromosome numbers (e.g. via dysploidy) can lead to both even and odd

Table 2 Chromosome Counts Database (CCDB) coverage in five manually curated plant clades

Clade	Number of taxa with data according to original study ^a	Number (percentage) of taxa in CCDB ^b	Number (percentage) of taxa in IPCN
Araceae	740	649 (88%)	524 (71%)
Colchicaceae	144	137 (95%)	118 (82%)
<i>Pinguicula</i>	68	36 (53%)	22 (32%)
Solanaceae	846	617 (73%)	474 (56%)
Poaceae subfamily Danthonioideae	161	127 (79%)	112 (70%)

^aFor comparison, the number of taxa in the original study is reported following the name resolution process described in the Materials and Methods section (i.e. omitting those that were not successfully resolved).

^bThe number of species reported in CCDB excluding counts obtained from the original studies.

IPCN, Index to Plant Chromosome Numbers.

Table 3 Major groups even vs odd chromosome counts^a

Clade	Number of total counts	Percentage of even counts
Angiosperm	70 338	56%
Monocots	15 528	58%
Eudicots	53 492	55%
Monilophytes	2986	63%
Lycophytes	220	53%
Gymnosperms	488	59%
Bryophytes	1446	48%

^aIn these analyses ambiguous counts were not considered (e.g. odd $2n$ counts, those appearing within parentheses, or those containing Roman numerals due to the higher OCR error rate).

numbers. Interestingly, the chromosome number distribution varies markedly between the major plant groups (Fig. 1). In monilophytes (Fig. 1a), a clade known to possess particularly high chromosome numbers (reviewed in Barker, 2013), the most common haploid number is 41, followed by 36 with two additional peaks at 82 and 72 that are exact duplications of the two most common numbers. Additionally, while 63% (1887 out of 2986) of the species possess an even chromosome number, the even-to-odd

ratio increases substantially considering counts larger than the modal number (for counts above 41, 79% of the species have an even haploid number), suggesting that chromosome number increases are mainly the result of polyploidy transitions. In lycophytes, three distinct peaks are observed (Fig. 1b): the lowest peak *c.* 9–11 comprises mostly chromosome counts originating from Isoetales and Selaginellales, a second peak *c.* 22–23 that includes counts from Isoetales and Lycopodiales, and a third peak *c.* 34 of Lycopodiales species. In angiosperms (Fig. S1b), as is also reflected in the distribution obtained for eudicots (Fig. 1c), the modal number is more diffused and is centered *c.* 7–12 and the saw-toothed pattern is noticeable for chromosome numbers larger than 12. While 56% of angiosperms have an even haploid number, the even-to-odd ratio changes substantially above the major mode—the ratio between even and odd numbers below 12 is 0.95 (i.e. slightly more odds than evens), whereas for 13 and over it is 1.7. As far as chromosome numbers are concerned, it seems that plants possessing low chromosome numbers have undergone a polyploidy event so long ago that its signal has been eroded by subsequent dysploidy events. When considering the two main angiosperm clades, monocots were shown to have undergone more frequent polyploidy events compared to eudicots (Otto & Whitton, 2000). Indeed, the saw-toothed pattern for monocots (Fig. 1d) is

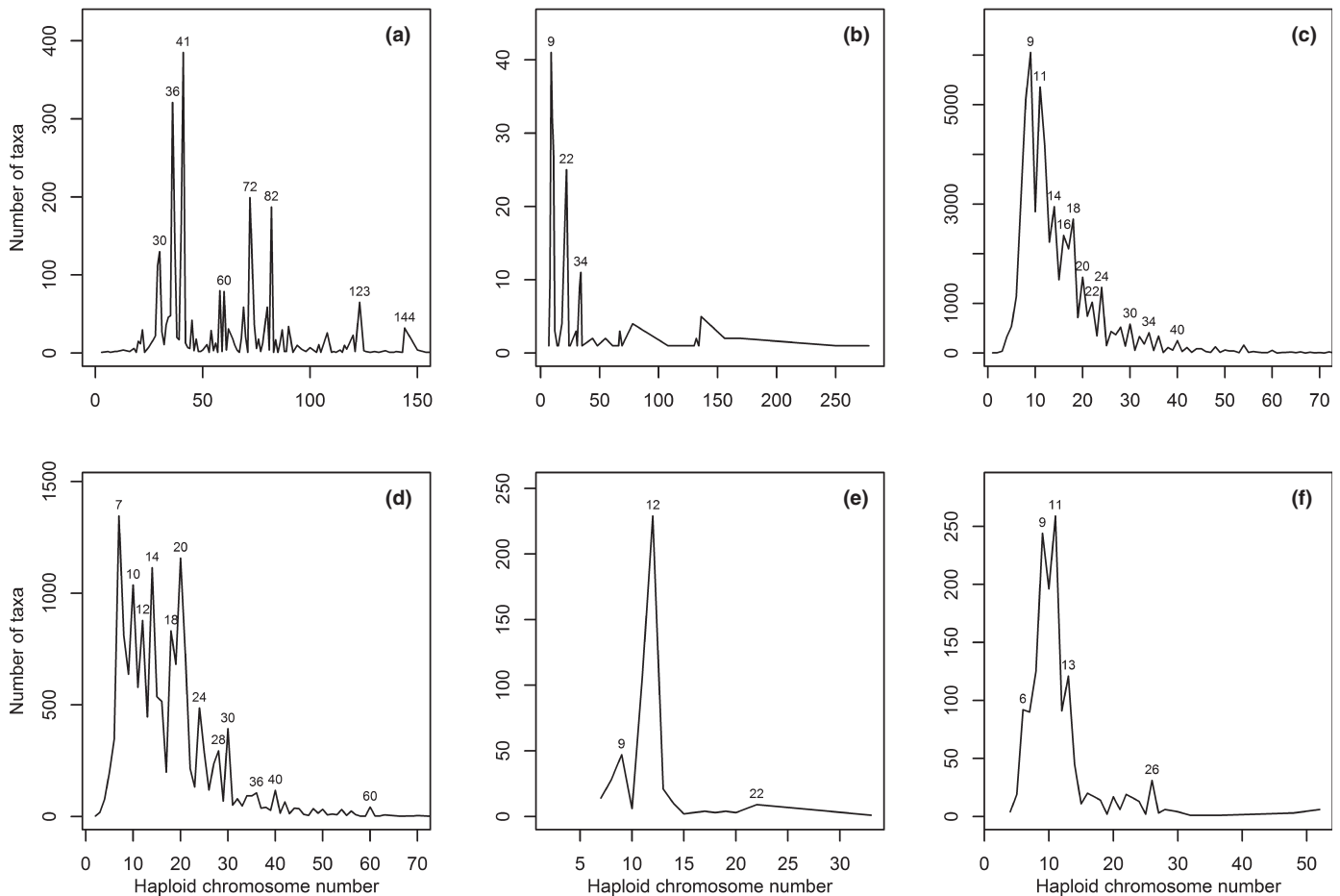


Fig. 1 The distribution of haploid chromosome numbers across the major plant groups. The distribution is calculated across (a) monilophytes, (b) lycophytes, (c) eudicots, (d) monocots, (e) gymnosperms and (f) bryophytes.

particularly apparent with an even-to-odd ratio of 1.7 above the modal count of 7. In gymnosperms (Fig. 1e) – a group in which polyploidy is considered rare (Husband *et al.*, 2013) – there is a high percentage of even counts (59%). However, this is due to the modal count of 12 (47% of all species) and the saw-toothed pattern is not apparent. In bryophytes, no apparent saw-toothed pattern was observed (Fig. 1f), with a relatively diffused mode between 6 and 13.

Next, we examined the extent by which chromosome number varies within resolved named species and infraspecific taxa (i.e. considering subspecies and varieties distinct from the corresponding species). Our analysis revealed that cytotype polymorphism is frequent within named species and infraspecific taxa, existing in 22.7% of taxa in our database; 15% of taxa were reported with two distinct counts and 7.7% with three or more cytotypes. Moreover, repeating this analysis at the species level (i.e. by collapsing all infraspecific names to their corresponding binomials), revealed that intraspecific variation in chromosome numbers exists in 23.5% (16 379 out of 69 639) of species in our database (15.2% of species were reported with two distinct counts and 8.3% with three or more). With the exception of gymnosperms, the frequency of species with multiple counts is relatively similar across the major lineages (23.6%, 26.5%, 22.1%, 20.1% and 12.1% for angiosperms, monilophytes, lycophytes, bryophytes and gymnosperms, respectively). These frequencies are obviously an underestimation due to the incompleteness of the database (i.e. not all reported cytotypes are included in CCDB) and since the karyotypes of many distinct cytotypes were not determined.

The multiple cytotypes that exist within nearly one quarter of named plant species encompass cases that affect only the karyotype but not the genomic content (e.g. chromosome fusion) and those that affect both as a result of major genomic processes such as polyploidy. As suggested by Soltis *et al.* (2007), a significant fraction of such intraspecific ploidal variants arose through autopolyploidy events. In many cases, these autopolyploids should be treated as distinct species under most commonly used species concepts. Thus, we examined the extent to which intraspecific variation in chromosome numbers can be attributed to ploidal variants using a simple nonphylogenetic approach. To this end, for each polymorphic species the *ploidy index* for all its cytotypes was defined as the multiplication factor relative to the lowest chromosome number found in that species (e.g. if the reported gametophytic counts for a certain species were 10, 15 and 20 the respective multiplication factors were 1.5 and 2). As shown in Fig. 2, a very large fraction of the observed intraspecific variation is due to polyploidy. Clearly, the most common factor is 2, which corresponds to a single whole genome duplication; next are the factors 3, 4, 5 and 6, each corresponding to chromosome number changes due to polyploidy. In addition, the frequency *c.* 1 is relatively high and could be explained by dysploidy events (such as chromosome fission and fusion) while another peak is observed *c.* 1.5, corresponding to the occurrence of triploid taxa. In order to evaluate the relative contribution of polyploidy to intraspecific chromosome number variation compared to other processes of chromosome-number change, a threshold of 1.4 was used. Assuming that this threshold can be used to distinguish polyploidy

events (including transition to triploids) from dysploidy transitions, 69% of the observed intraspecific variation is due to polyploidy, whereas 31% are due to other types of chromosome number transition. In total, our analysis revealed that 16.2% of plant species harbor intraspecific variation in their ploidy levels – higher than the estimate provided by Wood *et al.* (2009) who reported that 12–13% of angiosperms and 17% of fern species (in that study including the lycophytes) harbor multiple ploidy levels (compared to 16.2% in angiosperms and 19.7% in pteridophytes: 20.1% in monilophytes and 12.9% in lycophytes observed in our analysis). The difference in estimates stems mainly from the different cutoff used by Wood *et al.* who disregarded triploids in their estimate (by using a threshold of 1.75) but also due to the additional data incorporated in CCDB.

Discussion

Here, we presented the Chromosome Counts Database, as a community resource for plant researchers. While CCDB represents a step towards enhanced data coverage and accessibility, for certain clades data completeness is still lacking. CCDB may thus guide other global initiatives, such as those concerning the collection of C-values (Galbraith *et al.*, 2011) by pointing out taxonomic groups where collection efforts could be particularly rewarding. The current coverage for angiosperms in CCDB is *c.* 20% while Bennett

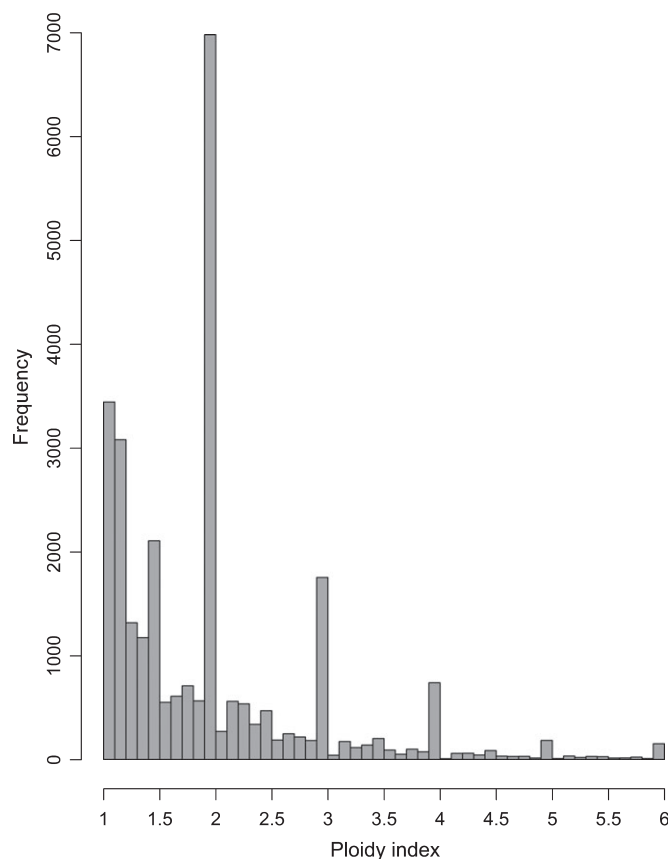


Fig. 2 Ploidy index distribution. The distribution of ploidy index in species that harbor intraspecific chromosome number variation.

(1998) estimated this number to be *c.* 25%. While the difference in these estimates may also stem from the number of angiosperm species assumed, there are obviously additional data that CCDB does not contain. For example, in bryophytes merely 4% of the species have chromosome-number information in CCDB, while Husband *et al.* (2013) estimated the coverage for bryophytes to be three times higher. Importantly, the estimation reported by Husband *et al.* (2013) was based on two printed sources (Fritsch, 1991; Przywara & Kuta, 1995), which in the current compilation of CCDB were not available. Such gaps in coverage can be readily filled by the community either by uploading directly through the CCDB website or by providing data in the form of a printed/scanned copy, which can be automatically processed using the developed procedures. Our goal in the construction of CCDB was to provide an extensive, yet flexible framework within which additional data can be added by the community, thus facilitating data sharing for a wide array of in-depth studies concerning the pattern of chromosome number change.

Acknowledgements

We thank Michael S. Barker, Emma Goldberg and Murray Dawson for providing us with extensive chromosome number collections; Ilia Leitch for providing a CSV file of the Plant DNA C-value database and Grzegorz Góralski for providing a CSV file of the chromosome number database of Polish plants; Sarah P. Otto for providing a scanned copy of the book *Cytotaxonomical atlas of the Pteridophyta*, and Aretuza Sousa, Susanne S. Renner, and an anonymous reviewer for constructive comments. This study was supported by a fellowship from the Manna Center Program in Food Safety & Security to L.G. and by the Israel Science Foundation grant number 1265/12.

**Anna Rice[†], Lior Glick[†], Shiran Abadi[†], Moshe Einhorn,
Naama M. Kopelman, Ayelet Salman-Minkov,
Jonathan Mayzel, Ofer Chay and Itay Mayrose***

Department of Molecular Biology and Ecology of Plants, George S. Wise Faculty of Life Sciences, Tel Aviv University,
Tel Aviv 69978, Israel

(*Author for correspondence: tel +972 3 640 7212;
email itaymay@post.tau.ac.il)

[†]These authors contributed equally to this work.

References

- Albach D, Martínez-Ortega M, Delgado L, Weiss-Schneeweiss H, Özgökcü F, Fischer M. 2008. Chromosome numbers in Veroniceae (Plantaginaceae): review and several new counts 1. *Annals of the Missouri Botanical Garden* **95**: 543–566.
- Aldasoro JJ, Aedo C, Garmendia FM, de la Hoz FP, Navarro C. 2004. Revision of *Sorbus* subgenera *Aria* and *Torminaria* (Rosaceae-Maloideae). *Systematic Botany Monographs* **69**: 1–148.
- Barker MS. 2013. Karyotype and genome evolution in pteridophytes. In: Greilhuber J, Dolezel J, Wendel J, eds. *Plant genome diversity, vol. 2*. Vienna, Austria: Springer, 245–253.
- Bedini G, Garbari F, Peruzzi L. 2010 onwards. Chrobase.it – Chromosome numbers for the Italian flora. [WWW document] URL <http://www.biologia.uniipi.it/chrobase/> [accessed 20 June 2013].
- Bennett MD. 1998. Plant genome values: how much do we know? *Proceedings of the National Academy of Sciences, USA* **95**: 2011–2016.
- Bennett MD, Leitch IJ. 2011. Nuclear DNA amounts in angiosperms: targets, trends and tomorrow. *Annals of Botany* **107**: 467–590.
- Bohs L. 2001. Revision of *Solanum* section *Cyphomandropsis* (Solanaceae). *Systematic Botany Monographs* **61**: 1–85.
- Casper SJ, Stimpert R. 2009. Chromosome numbers in *Pinguicula* (Lentibulariaceae): survey, atlas, and taxonomic conclusions. *Plant Systematics and Evolution* **277**: 21–60.
- Chacón J, Cusimano N, Renner SS. 2014. The evolution of Colchicaceae, with a focus on changes in chromosome numbers. *Systematic Botany* **39**: 415–427.
- Chatterjee T, Kumar Sharma A. 1969. Cytotaxonomy of cichorieae. *Genetica* **40**: 577–590.
- Cusimano N, Sousa A, Renner SS. 2012. Maximum likelihood inference implies a high, not a low, ancestral haploid chromosome number in Araceae, with a critique of the bias introduced by 'x'. *Annals of Botany* **109**: 681–692.
- Darlington CD, Wylie AP. 1955. *Chromosome atlas of flowering plants*. London, UK: George Allen and Unwin Ltd.
- Fedorov AA. 1969. *Chromosome numbers of flowering plants*. Leningrad, USSR: Academy of Natural Sciences of the USSR.
- Freire-Fierro A. 2002. Monograph of *Aciotis* (Melastomataceae). *Systematic Botany Monographs* **62**: 1–99.
- Fritsch R. 1991. *Index to bryophyte chromosome counts*. Berlin, Germany: J. Cramer.
- Galbraith DW, Bennett JL, Kellogg EA, Pires JC, Soltis PS. 2011. The genomes of all angiosperms: a call for a coordinated global census. *Journal of Botany* **2011**: Article 646198.
- Glick L, Mayrose I. 2014. ChromEvol: assessing the pattern of chromosome-number evolution and the inference of polyploidy along a phylogeny. *Molecular Biology and Evolution* **31**: 1914–1922.
- Goldblatt P. 1980. Polyploidy in angiosperms: monocotyledons. In: Lewis WH, ed. *Polyploidy – biological relevance*. New York, NY, USA: Springer/Plenum, 219–239.
- Goldblatt P. 1981. Index to plant chromosome numbers 1975–1978. *Monographs in systematic botany from the Missouri Botanical Garden* **6**: 1–553.
- Goldblatt P, Johnson D. 1979–. *Index to plant chromosome numbers*. St Louis, MO, USA: Missouri Botanical Garden.
- Goldblatt P, Lowry PP. 2011. The Index to Plant Chromosome Numbers (IPCN): three decades of publication by the Missouri Botanical Garden come to an end. *Annals of the Missouri Botanical Garden* **98**: 226–227.
- Góralski G, Lubczynska P, Joachimiak A. 2009 onwards. Chromosome number database. [WWW document] URL <http://chromosomes.binoz.uj.edu.pl> [accessed June 2013].
- Grant V. 1963. *The origin of adaptations*. New York, NY, USA: Columbia University Press.
- Guerra M. 2008. Chromosome numbers in plant cytotoxicity: concepts and implications. *Cytogenetic and Genome Research* **120**: 339–350.
- Guerra M. 2012. Cytotoxicity: the end of childhood. *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* **146**: 703–710.
- Hansen AK, Gilbert LE, Simpson BB, Downie SR, Cervi AC, Jansen RK. 2006. Phylogenetic relationships and chromosome number evolution in *Passiflora*. *Systematic Botany* **31**: 138–150.
- Husband BC, Baldwin SJ, Suda J. 2013. The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: Greilhuber J, Dolezel J, Wendel J, eds. *Plant genome diversity, vol. 2*. Vienna, Austria: Springer, 255–276.
- Jara-Seguel P, Urrutia J. 2011. Chilean plants cytogenetic database. Jardín Botánico Nacional, Chile. [WWW document] URL <http://www.Chileanpcd.com> [accessed 25 June 2013].
- Kamari G, Felber F, Garbari F. 1991. Mediterranean chromosome number reports - 1. *Flora Mediterranea* **1**: 223–245.
- Kluyver TA, Osborne CP. 2013. Taxonome: a software package for linking biological species data. *Ecology and Evolution* **3**: 1262–1265.

- Kucera J, Valko I, Marhold K. 2005. On-line database of the chromosome numbers of the genus *Cardamine* (Brassicaceae). *Biologia* 60: 473–476.
- Kumar V, Subramaniam B. 1987a. *Chromosome atlas of flowering plants of the Indian subcontinent, vol. I*. Calcutta, India: Botanical Survey of India.
- Kumar V, Subramaniam B. 1987b. *Chromosome atlas of flowering plants of the Indian subcontinent, vol. II*. Calcutta, India: Botanical Survey of India.
- Laane MM, Hoiland K. 1986. Chromosome number and meiosis in herbarium specimens from the extinct Scandinavian population of *Crepis multicaulis*. *Hereditas* 105: 187–192.
- Linder HP, Barker NP. 2014. Does polyploidy facilitate long-distance dispersal? *Annals of Botany* 113: 1175–1183.
- Löve A, Löve D. 1948. *Chromosome numbers of northern plant species*. Reykjavik, Iceland: University Institute of Applied Sciences.
- Löve AS, Löve D, Pichi-Sermolli RE. 1977. Cytotaxonomical atlas of the Pteridophyta. Vaduz, Liechtenstein: J. Cramer.
- Marhold K. 2006. IAPT/IOPB chromosome data 1. *Taxon* 55: 443–445.
- Marhold K, Martonfi P, Mered'a P, Mraz P, Hodálová I, Kolník M, Kucera J, Lihová J, Mrázová V, Perný M *et al.* 2007. *Chromosome number survey of the ferns and flowering plants of Slovakia*. www.chromosomes.sav.sk. Bratislava, Slovakia: VEDA.
- Masterson J. 1994. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* 264: 421–424.
- Mayrose I, Barker MS, Otto SP. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. *Systematic Biology* 59: 132–144.
- Meudt HM. 2006. Monograph of *Ourisia* (Plantaginaceae). *Systematic Botany Monographs* 77: 1–188.
- Miller JM, Chambers KL. 2006. Systematics of *Claytonia* (Portulacaceae). *Systematic Botany Monographs* 78: 1–236.
- Moore R. 1977. Index to plant chromosome numbers for 1973–74. *Regnum Vegetabile* 96: 1–257.
- Moore DM. 1982. *Flora Europaea check-list and chromosome index*. Cambridge, UK: Cambridge University Press.
- Moore RJ. 1970. Index to plant chromosome numbers for 1968. *Regnum Vegetabile* 68: 1–115.
- Moore RJ. 1971. Index to plant chromosome numbers for 1969. *Regnum Vegetabile* 77: 1–112.
- Moore RJ. 1973. Index to plant chromosome numbers, 1967–1971. *Regnum Vegetabile* 90: 1–539.
- Moore RJ. 1974. Index to plant chromosome numbers for 1972. *Regnum Vegetabile* 91: 1–108.
- Ohi-Toma T, Sugawara T, Murata H, Wanke S, Neinhuis C, Murata J. 2006. Molecular phylogeny of *Aristolochia* sensu lato (Aristolochiaceae) based on sequences of *rbcl*, *matK*, and *phyA* genes, with special reference to differentiation of chromosome numbers. *Systematic Botany* 31: 481–492.
- Ornduff R. 1967. Index to plant chromosome numbers for 1965. *Regnum Vegetabile* 50: 1–112.
- Ornduff R. 1968. Index to plant chromosome numbers for 1966: International Bureau for Plant Taxonomy and Nomenclature of the International Association for Plant Taxonomy. *Regnum Vegetabile* 55: 1–126.
- Otto SP, Whitton J. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401–437.
- Przywara L, Kuta E. 1995. Karyology of bryophytes. *Polish Botanical Studies* 9: 1–83.
- Rosalson EH. 2008. A synopsis of chromosome number variation in the Cyperaceae. *The Botanical Review* 74: 209–393.
- Saunders RM. 2000. Monograph of *Schisandra* (Schisandraceae). *Systematic Botany Monographs* 58: 1–146.
- Schlarbaum SE, Tsuchiya T. 1984. Cytotaxonomy and phylogeny in certain species of Taxodiaceae. *Plant Systematics and Evolution* 147: 29–54.
- Schuhwerk F. 1996. Published chromosome-counts in *Hieracium*. [WWW document] <http://www.botanischestaatssammlung.de/projects/chrzit.html> [accessed 12 June 2013].
- Schultheis LM. 2001. Systematics of *Downingia* (Campanulaceae) based on molecular sequence data: implications for floral and chromosome evolution. *Systematic Botany* 26: 603–621.
- Soltis DE, Soltis PS, Schemske DW, Hancock JF, Thompson JN, Husband BC, Judd WS. 2007. Autopolyploidy in angiosperms: have we grossly underestimated the number of species? *Taxon* 56: 13–30.
- Stebbins GL. 1938. Cytological characteristics associated with the different growth habits in the dicotyledons. *American Journal of Botany* 25: 189–198.
- Stevens PF. 2012. Angiosperm Phylogeny Website. Version 13, [WWW document] URL <http://www.mobot.org/MOBOT/research/APweb> [accessed October 2014].
- Thompson DM. 2005. Systematics of *Mimulus* subgenus *Schizoplacus* (Scrophulariaceae). *Systematic Botany Monographs* 75: 1–213.
- Wagner WL, Weller SG, Sakai A. 2005. Monograph of *Schiedea* (Caryophyllaceae subfam. Alsinoideae). *Systematic Botany Monographs* 72: 1–169.
- Warwick SI, Al-Shehbaz IA. 2006. Brassicaceae: chromosome number index and database on CD-Rom. *Plant Systematics and Evolution* 259: 237–248.
- Watanabe K. 2002. Index to chromosome numbers in Asteraceae. [WWW document] URL http://www.lib.kobe-u.ac.jp/infolib/meta_pub/G0000003asteraceae_e [accessed 20 June 2013].
- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* 106: 13 875–13 879.
- Zuloaga FO, Pensiero J, Morrone O. 2004. Systematics of *Paspalum* Group *Notata* (Poaceae-Panicoideae-Paniceae). *Systematic Botany Monographs* 71: 1–75.

Supporting Information

Additional supporting information may be found in the online version of this article.

Fig. S1 The distribution of haploid chromosome numbers in CCDB across all taxa and angiosperms.

Table S1 Data coverage in CCDB for the 20 largest plant families

Please note: Wiley Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Key words: chromosome counts, Chromosome Counts Database (CCDB), chromosome numbers, database, dysploidy, intraspecific variation, polyploidy.