The Evolution of Chromosome Numbers: Mechanistic Models and Experimental Approaches

Itay Mayrose [1],* and Martin A. Lysak[2],*

[1]School of Plant Sciences and Food Security, George S. Wise Faculty of Life Sciences, Tel Aviv University, Israel
[2]CEITEC—Central European Institute of Technology, Masaryk University, Brno, Czech Republic

*Corresponding authors: E-mails: itaymay@tauex.tau.ac.il; martin.lysak@ceitec.muni.cz.

Abstract

Chromosome numbers have been widely used to describe the most fundamental genomic attribute of an organism or a lineage. Although providing strong phylogenetic signal, chromosome numbers vary remarkably among eukaryotes at all levels of taxonomic resolution. Changes in chromosome numbers regularly serve as indication of major genomic events, most notably polyploidy and dysploidy. Here, we review recent advancements in our ability to make inferences regarding historical events that led to alterations in the number of chromosomes of a lineage. We first describe the mechanistic processes underlying changes in chromosome numbers, focusing on structural chromosomal rearrangements. Then, we focus on experimental procedures, encompassing comparative cytogenomics and genomics approaches, and on computational methodologies that are based on explicit models of chromosome-number evolution. Together, these tools offer valuable predictions regarding historical events that have changed chromosome numbers and genome structures, as well as their phylogenetic and temporal placements.

Key words: chromosome numbers, polyploidy, dysploidy, cytogenomics, phylogenetic models, genome evolution.

Significance

Chromosome numbers have been widely used to describe the karyotype of eukaryote genomes. Changes in chromosome numbers regularly serve as indication of major genomic events, most notably polyploidy and dysploidy. Here, we describe mechanistic processes underlying changes in chromosome numbers and review recent advancements in the field that increase our ability to make inferences regarding historical events that led to alterations in the number of chromosomes of a lineage.

Chromosome Number Variation in Land Plants

The chromosome number of a taxon is perhaps the single most informative character that describes the genomic organization of a lineage. For over a century, chromosome numbers have been used as informative phylogenetic indicators (Guerra 2008) and their importance in macroevolutionary processes has been repeatedly discussed (Mayr 1982; Clark and Donoghue 2018). Although considerable variation exists in animals, the extent of this variation has been particularly well appreciated in land plants, inspiring botanists to inspect and document chromosome numbers of many thousands of species (Rice et al. 2015). The lowest number of chromosomes of a plant genome was reported for six angiosperm species that are known to possess merely two chromosome pairs ($n = 2$; Castiglione and Cremonini 2012). At the other end, extremely high chromosome numbers were reported in the fern *Ophioglossum reticulatum* ($n = 720$; Khandelwal 1990), the dicot tree *Strasburgeria robusta* ($n = 250$; Oginuma et al. 2006), the succulent *Sedum suaveolens* ($n = \sim320$; Uhl 1978), and the forest coconut palm (*Voanioala gerardii*, $n = 303$; Röser 2015).

The extant chromosome-number variation is external manifestation of the underlying dynamic genomic processes,

encompassing structural chromosomal rearrangements and changes in the DNA content. The most recognizable chromosome-number change is through a whole-genome duplication (WGD), or more generally polyploidization, which describes the acquisition of one or more complete chromosome sets to the genome. Single-chromosome changes represent another common pathway underlying chromosome-number variation. These transitions include the gain/loss of a single chromosome(s)—a process known as aneuploidy, and processes such as chromosome fission and fusion (ascending and descending dysploidy, respectively), which change the chromosome number while preserving most of the genomic content.

The size and morphology of chromosomes change through double-strand breaks (DSBs) in chromosomal DNA and by subsequent miss-repair at these breakpoints. Duplications, deletions, inversions, and translocations, and sometimes combination of these rearrangements, have the potential to alter the length of chromosome arms, change the centromere position, as well as the order and position of genes on chromosomes (gene linkage). Chromosome translocations mediate the reduction of chromosome numbers through recombination between at least two nonhomologous chromosomes (descending dysploidy). Conversely, chromosome breakage not followed by DSB (miss-)repair can potentially result in chromosome-number increase (ascending dysploidy).

During the last decade, revolutionary advancements have enhanced our ability to make inferences regarding historical events that led to chromosome-number changes. These include both experimental procedures, encompassing novel comparative genomics approaches, and computational methodologies that offer more robust and flexible predictions of ancestral chromosome numbers and their phylogenetic placements. Here, we first describe mechanistic processes underlying changes in chromosome numbers. Then, we focus on state-of-the-art experimental and computational methodologies that are applied to uncover such changes and to estimate their timings.

Processes Governing Chromosome-Number Variation

Mechanisms of Chromosome Number Increase

Polyploidy

Mechanisms underlying polyploidization events have been thoroughly discussed in the literature and so we only briefly describe these here and refer the readers to many excellent reviews (Ramsey and Schemske 1998, 2002; Comai 2005; Otto 2007; Soltis et al. 2016; Van De Peer et al. 2017). Chromosome number sets can be multiplied through somatic doubling, polyspermy, or most frequently, through unreduced gametes that contain the same number of

chromosomes as somatic cells. Two types of polyploidy, differentiated by the extent of similarity between the duplicated chromosome sets, have been widely recognized. Autopolyploidy is referred to the condition where the duplicated genomes are nearly identical, having originated from a single species, whereas allopolyploidy is the result of a hybridization of two different lineages (typically species) whose genomes have already diverged to some extent. In case of these euploid changes, the least increment of chromosome number is triploidy $(2n + n)$, typically produced through fusion of reduced and unreduced gametes or by crossing diploids and tetraploids. Although triploidy is an unstable evolutionary condition, usually characterized by low fertility, it may serve as a route to tetraploidy via the triploid bridge pathway. Tetraploids may originate from the union of $2n$ gametes in the triploid via self-fertilization or by backcrossing of the triploid with a closely related diploid individual (Ramsey and Schemske 1998). Tetraploids can also be formed via the fusion of two unreduced gametes, whose frequency is particularly high in hybrids (27.5% compared with 0.56% in nonhybrids). From these, higher ploidy individuals and populations may be generated via a variety of mechanisms, including production of unreduced gametes in polyploid individuals and hybridizations involving one or more polyploid lineages (Ramsey and Schemske 1998).

Aneuploidy

Aside from euploid changes, chromosome numbers may increase in a step-wise manner by one or few chromosomes, a process generally referred to as aneuploidy. Aneuploidy refers to extra homologous chromosome(s) being present within a chromosome set during a life span of their carriers or in one or a few subsequent generations (e.g., trisomy of a single chromosome leading to a $2n + 1$ genome). Aneuploids may originate by several ways, with nondisjunction in meiosis or mitosis being the most frequent pathway. Meiotic nondisjunction leads to formation of aneuploid gametes. The subsequent union of aneuploid and euploid gamete may result in the origin of trisomic $[n + (n + 1) \rightarrow 2n + 1]$ and monosomic $[n + (n - 1) \rightarrow 2n - 1]$ individuals. In plants, products of mitotic nondisjunction can enter the germline and become transmitted to the progeny due to pluripotency of plant cells.

Frequently, the extra chromosome(s) and/or their carriers are eliminated from the population. However, the union of two $n + 1$ gametes may increase the chromosome number and restore the euploidy in the offspring $(2n + 2)$. The resulting tetrasomic progeny will suffer from reduced fertility due to four homologs forming a quadrivalent, followed by irregular chromosome segregation. These individuals are expected to suffer from additional fitness disadvantage due to imbalanced gene content (Otto 2007). Accumulation of structural and DNA changes among the four homologs will eventually
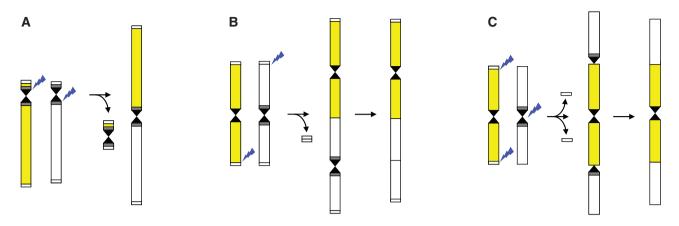
Fig. 1.—Mechanisms of descending dysploidy in plants. (*A*) Robertsonian translocation. (*B*) End-to-end translocation. (*C*) Nested chromosome insertion. The blue lightning symbols denote the location of double-strand breaks (breakpoints), the black sandglass symbols represent centromeres, and the small white/gray rectangles stand for (sub)telomeric/pericentromeric repeats.

restore regular (bivalent) pairing. Still, long-term establishment of such lineages is an exceedingly rare event.

Ascending Dysploidy

Mutations causing the increase of chromosome numbers while preserving the genomic content are referred to as ascending dysploidy. Centric fission is traditionally thought to be the most common type of ascending dysploidy. The breakage within a functional centromere or centromere miss-division during chromosome segregation (Birchler and Han 2018) produces two telocentric chromosomes. For this mutation to be stable, that is, an inheritable chromosome-number increase by one ($n+1$), the compromised centromeres should retain the capacity to form a kinetochore and the centric ends of the two telocentric chromosomes should be healed by de novo telomere formation (Jankowska et al. 2015; Kurzhals et al. 2017). Although ascending dysploidy mediated by centric fissions is assumed to be frequent only in a few plant groups with monocentric chromosomes (e.g., in cycad genus *Zamia*; Rastogi and Ohri 2020), land plant genome evolution is dominated by descending dysploidy (Carta et al. 2020).

Mechanisms of Chromosome Number Decrease (Descending Dysploidy)

Mitotic or meiotic nondisjunction will render some cells aneuploid (e.g., $2n-1$; a monosomy of one chromosome). Such aneuploidies are usually not tolerated due to the loss of essential genes. However, in rare circumstances, gene redundancy of polyploid genomes may allow the accidental droppage of some chromosomes without fatal consequences for carrier's fertility (Clausen and Cameron 1944).

A more common reduction in chromosome numbers, referred to as descending dysploidy, occurs via chromosome fusion. The basis of all descending dysploidies is the mis-repair of DSBs on two or more nonhomologous chromosomes, that is, a chromosome translocation(s). The so-called translocation or "fusion" chromosomes can be transmitted to the offspring and become fixed only at the condition that no housekeeping, or other essential genes, are lost during the process. In organisms with monocentric chromosomes (i.e., with a localized centromere), descending dysploidy is usually accompanied by elimination of one of the two centromeres of the ancestral chromosomes. However, if two nonhomologous centromeres of a fusion chromosome are at a short physical distance, both centromeres may retain their functionality (Page and Shaffer 1998). If more than two chromosomes have contributed to the origin of a fusion chromosome, a more parsimonious step-wise reconstruction, involving the merging of two chromosomes at a time, is usually assumed. However, it is well plausible that in some instances concurrent DSBs on three or more nonhomologous chromosomes would result in structurally complex fusion chromosomes (e.g., Zhang et al. 2008; Mandáková et al. 2019; reviewed by Pellestor 2019). With the advent of high-throughput sequencing technologies and modern cytogenomics, several mechanisms of descending dysploidy were characterized in greater detail or newly identified. The most common types of chromosome-number-reducing mechanisms are illustrated in figure 1 and are briefly described below.

Robertsonian translocations (ROB; fig. 1A). A typical ROB combines long arms of two telo- or acrocentric chromosomes by recombination between their centric ends or short arms. The first translocation product is the fusion chromosome containing two long arms and one or two centromeres (monocentric or dicentric ROB). The second product is either an acentric fragment or centromere-containing mini-chromosome, whose elimination due to its small size and absence of essential genes can be tolerated by the translocation carrier. Thus, ROBs reduce the number of short arms, while

increasing the number of metacentric chromosomes, and hence the karyotype symmetry.

End-to-end translocation (EET; fig. 1B). EETs were deduced from two ancestral chromosomes being tandemly arranged within a single evolutionary younger chromosome. As the head-to-tail collinearity of both ancestral chromosomes remains conserved within the fusion chromosome, recombination between uncapped terminal sequences of two non-homologous chromosomes is the most plausible pathway of these chromosome fusions. Except for instances where the two centromeres are in a tight spatial proximity (i.e., in case of telocentric chromosomes), the fixation of the fusion chromosome depends on elimination of one of the ancestral centromeres. The elimination process is not well characterized and epigenetic modifications together with recombination-dependent deletion of centromere-specific nucleosomes and DNA sequences are purported to instantly restore monocentricity of the fusion chromosome (Lysak 2014).

Nested chromosome insertion (NCI; fig. 1C). NCIs combine two ancestral nonhomologous chromosomes in a peculiar fashion, superficially appearing as an insertion of one chromosome into the (peri)centromere of another chromosome (Luo et al. 2009). The "recipient" chromosome undergoes a centric fission followed by recombination between its two centric ends and sequences on both ends of the "insertion" chromosome; providing that both ends of the insertion chromosome lose their protective telomere structures. Thus, the centromere of the insertion chromosome serves as the functional centromere, whereas the centromere of the recipient chromosome loses its function.

Dysploidy in Groups with Holocentric Chromosomes

In organisms with monocentric chromosomes, all chromosomal rearrangements increasing or reducing the number of chromosomes must comply with the persistence or elimination of a functional centromere. In contrast, in organisms with holocentric or holokinetic chromosomes, that is, chromosomes with a kinetochore assembling along the entire chromosome length, dysploid changes are more easily fixed. The frequent fission and fusion events acting on holocentric chromosomes result in long dysploid series of chromosome numbers in some plant genera, such as *Carex* and *Luzula* (Guerra 2016 and references therein), as well as in butterflies (Lepidoptera; e.g., Ahola et al. 2014).

Methodological Approaches to Analyze Chromosome-Number Variation and Chromosome Collinearity

As all inferences of chromosome-number evolution are based on extant chromosome-number variation, all predictions essentially start on a laboratory workbench by establishing chromosome numbers for species in a group of interest. However,

chromosome number itself usually does not provide sufficient information on how the individual chromosomes and the whole karyotypes originated, and how they are related to chromosome complements of other species. Hence, the chromosome structure and cross-species chromosome/genome collinearity are being examined by comparative cytogenomic and genomic approaches.

Chromosome Counts and Ploidy Estimation

Chromosome counting from floral (e.g., anthers, pistills) and vegetative (most frequently root tip meristems) plant tissues still remains the most reliable method to establish the chromosome number of an investigated individual. This is often a laborious, time-consuming, and occasionally unsatisfactory procedure (e.g., in polyploid species with high chromosome numbers) that cannot be applied to large population samples. Although flow-cytometric DNA content estimation cannot substitute chromosome counting (Suda et al. 2006), it may represent a practical variant of deducing chromosome numbers and ploidy levels, unless genome size in the examined clade varies considerably.

Comparative Cytogenomic Approaches

Although chromosome counting and ploidy estimation methods provide the information on the number of chromosomes, chromosome structure and rearrangements underlying chromosome number variation have to be determined through comparative analysis of individual chromosomal DNA molecules. Direct localization of DNA sequences on chromosomes provides an observable evidence of their physical position. The development of cytogenetic methods used in comparative plant genomics was driven by three principal requirements. First, DNA sequences (called DNA probes) should be chromosome-specific, with none or minimal off-target localization. Second, a DNA probe should identify long stretches of chromosomal DNA (i.e., from hundreds of kb- to Mb-long regions, up to entire chromosomes). Third, comparative analyses require cross-hybridization of chromosome-specific DNA probes among two or more genomes being compared. This means that intergenome sequence homology must be sufficiently high to ensure that a probe originating from one genome will target a homeologous chromosomal region in other genomes. Thus, the level of phylogenetic relatedness among the compared genomes is a critical parameter.

With the advent of DNA sequencing and DNA cloning technologies, methods of classical comparative cytogenetics have changed (for a recent review, see Hu et al. 2020), enabling researchers to target and compare DNA sequences of individual chromosomes. This conceptual shift transformed classical cytogenetics research into comparative cytogenomics—the discipline of chromosome research capitalizing on multiple whole-genome sequences. Some of these approaches are detailed below.

Comparative Cytogenomics

Comparative Chromosome Painting Based on Bacterial Artificial Chromosomes

Bacterial artificial chromosomes (BACs) are large-insert vectors containing fragments of chromosomal DNA (~100 kb or larger). A set of BAC clones covering an entire chromosome or chromosome arm is referred to as a chromosome-specific library, whereas a continuous array of overlapping DNA clones is called a BAC contig (a BAC tiling path). Initial attempts to localize individual chromosome-specific BACs and short BAC contigs on plant chromosomes using fluorescence in situ hybridization (FISH) date to the mid-1990s (Woo et al. 1994; Hanson et al. 1995; Jiang et al. 1995), and were followed by numerous studies applying BAC FISH (reviewed by Jiang and Gill 2006; Jiang 2019). Large-scale chromosome painting to identify entire chromosomes using BAC contigs was established only for crucifers (Brassicaceae) and grasses (Poaceae). Lysak et al. (2001) published a proof-of-concept study showing BAC painting of the shortest *Arabidopsis thaliana* chromosome using 139 clones covering almost 16 Mb of the chromosome. Later, Pecinka et al. (2004) were able to BAC paint all five Arabidopsis chromosomes along their entire length. The true power of BAC-based chromosome painting for plant comparative genomics was demonstrated by application of Arabidopsis chromosome-level BAC contigs to analyze karyotypes of other crucifer species (fig. 2A and B) (Lysak et al. 2005, 2006). This approach, exploiting low- and single-copy (orthologous) sequences ancestrally shared among genomes, allows for identification of homeologous chromosome regions and chromosomes. Although comparative cytogenomics in Brassicaceae essentially explored the collinearity between chromosomes of *A. thaliana* and those of other crucifer species, in the grasses the sequenced genome of *Brachypodium distachyon* served as the source genome for cross-species chromosome painting in *Brachypodium* (Idziak et al. 2011; Betekhtin et al. 2014).

The abundance of dispersed repetitive sequences along plant chromosomes (Schubert et al. 2001) precludes a wider application of BAC-based chromosome painting in comparative plant cytogenomics. BAC-based chromosome painting relies on a BAC library derived from a genome with a low percentage of dispersed repeats and/or their localization is confined to a specific chromosomal region (typically around centromeres). The low repeat abundance decreases the probability that *bona fide* chromosome-specific BAC clones will cross-hybridize to nontarget chromosomes. Thus, BAC FISH and BAC-based chromosome painting was successfully established in plant species with small genome sizes, whose repeat content is low, such as *Amborella trichopoda* (870 Mb), *A. thaliana* (160 Mb), *B. distachyon* (355 Mb), *Genlisea margaretae* (185 Mb), or *Spirodela polyrhiza* (150–165 Mb) (Lysak et al. 2001; Idziak et al. 2011; Chamala et al. 2013; Cao et al. 2016; Tran et al. 2017; Hoang et al. 2018).

Oligo Painting

The ever-increasing number and quality of sequenced genomes opened up a new avenue for comparative cytogenomics. Although most plant genomes are dominated by repeat sequences, usually unsuitable for identification of specific chromosome regions, single- and low-copy (coding) sequences are chromosome-specific. The oligo painting approach (Han et al. 2015) is based on designing a library of short synthetic oligonucleotides (e.g., 45–50 bp in length) along megabase-long chromosomes. Such an oligo library is amplified, labeled by haptens or fluorochromes, and single-stranded labeled oligomers (oligo probes) are hybridized to the target chromosomes by FISH (see Jiang [2019] for a recent review). Compared with high-capacity BAC vectors, oligo painting does not require construction of chromosome-specific BAC libraries and subsequent screening to eliminate repeat-rich BAC clones. On the other hand, this methodology requires the availability of a chromosome-level genome sequence, synthesis of high-cost chromosome-specific oligo libraries, and entails several challenging preparatory steps (Han et al. 2015). In addition, oligo libraries offer less flexibility in targeting particular (shorter) chromosome regions compared with chromosome-specific BAC libraries. Oligo painting was successfully used to explore cross-species chromosome collinearity in several model and crop species, such as banana, cucumbers (fig. 2C and D), maize, or poplar (Han et al. 2015; Filiault et al. 2018; Albert et al. 2019; Šimoníková et al. 2019; Bi et al. 2020; Xin et al. 2020), and its popularity will likely continue to increase. Recently, Bi et al. (2020) modified the original approach by using segment-specific PCR primers that specifically amplify predefined subregions from a single synthetic oligo library (multiplex PCR-based oligo painting, MP-OP; fig. 2C and D). Such double-stranded oligo probes also generate stronger fluorescent signals compared with single-stranded oligomere probes (Han et al. 2015). MP-OP represents a cost- and time-effective approach to pinpoint complex collinearity-corrupting chromosomal rearrangements.

Comparative Plant Genomics

In the last two decades, the number and quality of sequenced plant genomes have increased sharply, from the first plant genome (*A. thaliana*) sequenced by the Sanger method (The Arabidopsis Genome Initiative 2000), to about 330 sequenced genomes of vascular plants available today (Chen et al. 2019; Kersey 2019). These sequencing projects have enlightened our view on complex patterns underlying chromosomal evolution. Already the pioneering Arabidopsis sequencing project has identified segmental duplications that pointed to a previously not recognized ancient WGD, followed by genome shuffling and descending dysploidy. The constantly improving quality of genome assemblies allows for unprecedentedly detailed comparisons of chromosome
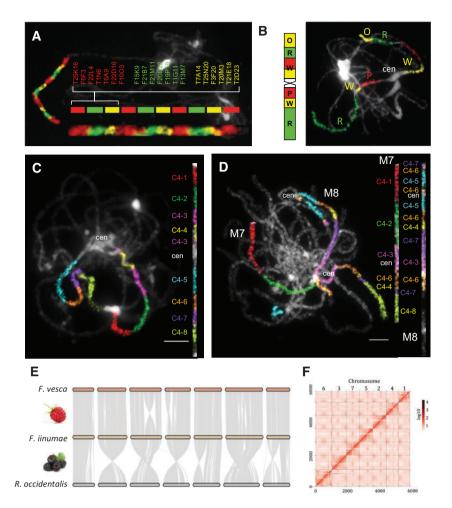
Fig. 2.—Complementary approaches of comparative plant genomics. (*A* and *B*) BAC-based chromosome painting (CP). (*A*) CP using chromosome-1-specific BAC clones of *A. thaliana* on pachytene chromosomes of this species. The set of 66 BACs (~6.7 Mb) was arbitrarily divided into ten alternatively labeled BAC contigs. (*B*) Comparative BAC-based CP. Pachytene chromosome 6 of *Noccaea caerulescens* (Alpine Penny-cress) painted using chromosome-specific BAC clones of *A. thaliana*. Capital letters refer to ancestral genomic blocks. cen, centromere. (*C* and *D*) Oligo painting. (*C*) Multiplex PCR-based oligo painting (MP-OP) using eight oligo probes specific for cucumber (*Cucumis sativus*) chromosome 4 on pachytene chromosomes of this species. (*D*) Comparative MP-OP using the same probes as in (*C*) revealing two homeologous chromosomes (M7 and M8) in the melon genome (*Cucumis melo*). (*E*) Chromosome-scale genome comparison among two strawberry (*Fragaria*) and one black raspberry (*Rubus*) species revealing the conserved versus corrupted intergenome chromosome collinearity. All three genomes were sequenced and assembled using the PacBio single-molecule sequencing technique (Edger et al. 2018, 2020; VanBuren et al. 2018). (*F*) High-throughput chromosome conformation capture (Hi-C) map of the *Rubus occidentalis* genome. Putative locations of centromeres are visible for some of the seven chromosomes. Figures were contributed by the authors of Mandáková and Lysak 2016 (*A*), Mandáková et al. 2015 (*B*), Bi et al. (2020) (*C* and *D*), Hardigan et al. (2020) (*E*), and VanBuren et al. (2018) (*F*).

structures among closely related species and across phyla. This was achieved particularly through single-molecule sequencing platforms producing longer sequence reads (e.g., PacBio—Pacific Biosciences, Nanopore—Oxford Nanopore Technologies; fig. 2*E*) and by novel methodologies for anchoring contigs and scaffolds into chromosome-scale pseudomolecules (i.e., Hi-C, high-throughput chromosome conformation capture, fig. 2*F*, and optical mapping) (for recent reviews, see Belser et al. 2018; Han et al. 2018; Ho et al. 2020; Michael and VanBuren 2020).

Chromosome-scale genome assemblies provide information on the structure of each chromosome within an organisms's karyotype, whereas contiguous genome sequences of multiple species allow for a high-resolution detection of intra- and inter-chromosomal rearrangements and structural variations (i.e., sequence variants >50 bp in size, Ho et al. 2020). Aligning genome assemblies of two or more species enables to discover collinearity-corrupting rearrangements and to characterize the corresponding breakpoints with a single-nucleotide accuracy (fig. 2*E*). Thus, sequence-based comparative genomics provides detailed insights into the principles underlying chromosomal evolution. In comparison with comparative cytogenomics, genome sequencing and assembly manifest an increased capability of detecting near-

complete spectrum of structural variations, including rearrangements, which were below the detection limit of microscopy-based methodologies, in a time-saving high-throughput manner.

Comparative Genomics and Cytogenomics

As stated by Hu et al. (2020), "cytogenetics and genomics are commonly used as complementary methods to provide synergistic information regarding chromosome structure." High-throughput genome sequencing generates data that can be utilized to develop chromosome-specific DNA probes (oligo painting) and novel chromosomal markers, such as tandem repeats (Emadzade et al. 2014; Macas et al. 2015; Báez et al. 2019; Hloušková et al. 2019; Schmidt et al. 2019; Liu et al. 2020). This is a time-saving and cost-effective approach for identifying the most abundant repeats even from low-coverage whole-genome sequence data. Conversely, (comparative) chromosome painting may guide genome assembly (fig. 2A–D), particularly when a genetic linkage map or reference sequence is lacking. Chromosome-specific probes may help to resolve ambiguities during anchoring sequence contigs and scaffolds to pseudochromosomes. Whole-chromosome comparative cytogenomic maps, such as these based on cross-species hybridization of Arabidopsis BAC contigs, guided genome assembly in several Brassicaceae species (Dassanayake et al. 2011; Willing et al. 2015; Geiser et al. 2016; Nowak et al. 2020), and BAC-FISH navigated genome assembly in a duckweed genome (S. polyrhiza; Wang et al. 2014; Hoang et al. 2018). FISH localization of chromosome-specific oligo-paints guided and cross-validated anchoring reference genome sequence to chromosomes in three banana species (Musa; Šimoníková et al. 2019).

Evolutionary Models of Chromosome Number Change

The extensive variation in plant chromosome numbers has been extensively exploited for inferring major genomic events, with particular interest toward determining which species are polyploids and which are diploids. Early work examined the distribution of chromosome numbers within a focal group of species and identified one or more denominators that are common to most chromosome counts. This number, commonly termed x, was regarded as the base number and taken to represent the ancestral haploid genome. Consequently, multiplications of this number were treated as the inferred ploidy level of the species. For example, given the following distribution of haploid chromosome numbers {8, 9, 9, 9, 14, 17, 18,18, 18, 20, 27, 27}, nine can be inferred as the base number, species with eight or nine chromosomes may be regarded as diploids, whereas the remaining ones as polyploids. Alternatively, others have designated a species as polyploid if its haploid number was a multiple of the lowest count

found in the examined clade by a predefined factor (Stebbins 1938; Wood et al. 2009). Clearly, such threshold methods suffered from extrapolated *ad hoc* assumptions, disregarded the relative frequencies of polyploid and dysploid transitions, and frequently disregarded the phylogenetic relationships among the species.

More recently, chromosome numbers were analyzed within a phylogenetic context following the maximum parsimony principle (Schultheis 2001; Hansen et al. 2006; Ohi-toma et al. 2006). The use of parsimony allows the reconstruction of chromosome numbers at ancestral nodes and the identification of putative transition events along particular branches of the phylogeny. However, as has been well discussed in the literature in the context of molecular sequences, the maximum parsimony approach suffers from several drawbacks (Felenstein 2004). Parsimony does not make use of an explicit model of evolution and thus the same weight is assigned to all state changes: whether they indicate a dysploidy (e.g., 10→11) or a polyploidy (e.g., 10→20) transition, or whether they include one or more transitions (e.g., 10→11, that could occur via a single step, compared with 10→12 that involves at least two transitions). Parsimony ignores branch length information, and thus changes along short branches would be treated similarly as those occurring along very long ones. Similarly, parsimony ignores the possibility of multiple and back transitions occurring along the same branch (e.g., 10→11→10) and thus only provides a lower bound on the number of events that practically occurred. Additionally, the use of parsimony implicitly assumes that the chromosome numbers of extinct ancestral taxa must also be presented in the extant taxa—an assumption that is not necessarily sensible, particularly if rates of chromosome-number change within the group are high.

Over the last decade, methods based on probabilistic models of chromosome number evolution have emerged. These methods are more powerful, as they emulate the evolutionary process along the phylogeny as a stochastic process, while taking into account the mechanisms by which chromosome numbers change. Consequently, the use of such models allows researchers to form and test assumptions regarding the most plausible evolutionary pathways by which the evolution of chromosome numbers have proceeded, while relying on the well-developed machinery of probabilistic statistical inference. For example, the likelihood ratio test can be used to compare the fit of alternative models, each containing a different set of parameters, to a specific data set at hand (Huelsenbeck and Crandall 1997). Additionally, once the evolution of chromosome numbers was casted within a probabilistic framework, a generic modeling scheme was created, allowing modeling extensions to be easily implemented and compared.

Several studies have employed general models of character evolution to model the evolution of chromosome numbers in clades in which chromosome fusion and fission events are the

main drivers of karyotype change. For example, Hipp (2007) had employed a series of Brownian Motion and Ornstein–Uhlenbeck processes to examine the evolution of chromosome numbers in the Cyperaceae (sedges), a group characterized by holocentric chromosomes, in which chromosome fusion and fission events are thought to be common and polyploidizations rare (see Dysploidy in Groups with Holocentric Chromosomes above). Rockman and Rowell (2002) have examined the evolution of chromosome numbers in *Planipapillus* (velvet worms), a group characterized by frequent centric fusion events, using the Poisson process. In both these groups, in which the dynamics of chromosomal evolution vary across subclades of the phylogeny, the evolutionary patterns of chromosome numbers better fitted a heterogeneous process. In these studies, chromosome numbers were modeled either as ordered categorical variables or as additive quantitative traits, and thus the possibility of integrating biological phenomena reflecting the mechanisms of chromosome-number change into the models was lacking.

The *chromEvol* Model

A model with a specific focus on the evolution of chromosome numbers was first formulated by Mayrose et al. (2010). The *chromEvol* model is based on a continuous time Markov process, which is defined by a rate matrix that describes the instantaneous rate of change from a genome with i haploid chromosomes to a genome with j haploid chromosomes. The entries in this matrix are determined based on several parameters that define the rate of change for different types of events (fig. 3). The most basic model assumes that three events are possible: WGD (an exact duplication of the number of chromosomes, with rate parameter ρ), a single-chromosome increase (ascending dysploidy at rate λ), or a single-chromosome decrease (descending dysploidy at rate δ). The rate matrix allows for the likelihood function to be computed, given a specified phylogeny and assignment of chromosome numbers to the tip taxa.

The use of the above simple model already allows for several inference tasks: 1) To obtain the maximum likelihood (ML) estimates of the rate parameters, allowing the relative frequencies of the different types of events to be compared; 2) To compare the fit of different model variants, each with different constraints on the free parameters, to a particular data set. For example, one can compare the basic model presented above with three free parameters (ρ, λ, and δ) to a null model, which assumes that polyploidizations are not possible and thus variations in chromosome numbers are only the result of dysploidy events (i.e., λ, and δ are the two free parameters, constraining $\rho = 0$). Model selection criteria, such as the likelihood ratio test or the Akaike Information Criterion, can then be used to test whether there is a significant evidence for polyploidization in the examined data set; 3) To reconstruct the ancestral
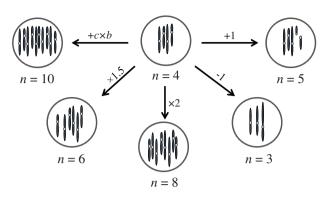

FIG. 3.—The possible transitions allowed in the *chromEvol* model. The models implemented in *chromEvol* allow for several types of transitions from a genome with i haploid chromosomes to a genome with j chromosomes: ascending dysploidy ($j = i + 1$), descending dysploidy ($j = i - 1$), genome duplication ($j = 2 \times i$), demi-ploidy ($j = 1.5 \times i$), and base-number transitions, in which the addition of any multiplication of a basic number b is allowed ($j = i + c \times b$; here $c = 1$ and $b = 6$).

chromosome numbers at internal nodes of the tree, including that of the root node. This can be done either using an ML approach (Pupko et al. 2000), in which the single most likely set of ancestral states is inferred, or using a Bayesian approach (Koshi and Goldstein 1996), in which the probability of each chromosome number occurring at each ancestral node is computed; 4) To estimate the expected number of polyploidy and dysploidy transitions that have occurred along each branch of the phylogeny; 5) To assign ploidy levels to extant taxa. A tip taxon can be classified as either diploid or polyploid, with respect to the state at the root of the phylogeny, if the expected number of diploid-to-polyploid transitions from the root to the tip is above (or below) a certain threshold. In initial applications of *chromEvol*, these thresholds were arbitrarily set as fixed values (e.g., 0.9, Mayrose et al. 2011). A more sensitive alternative was developed in Glick and Mayrose (2014), in which a simulation-based approach was used to compute the thresholds that are most suitable to the analyzed data.

We note that a probabilistic model of chromosome-number change was also developed by Hallinan and Lindberg (2011). This model is based on a background birth–death process (allowing for dysploidy transitions) that operates along branches of a species tree combined with the possibility of strict doubling events (i.e., WGD). This model sums over all possible assignments to ancestral states and over all possible number of dysploidy events, while allowing for the possibility of a single polyploidy transition to occur per branch, and can be used to compute the posterior probability that a polyploidy event occurred on each branch of the phylogeny. Extension of this model, to include polyploid transitions aside from exact duplications or multiple polyploidy events, is not trivial since the computation over all possible number of duplications and across all types of transitions per branch need to be explicitly formulated. In the following sections, we thus

describe modeling extensions that were developed in the context of the more general *chromEvol* probabilistic framework.

Variations of the *chromEvol* Model

Polyploidy Transitions Other Than Exact Duplications

The basic *chromEvol* model incorporates WGDs that involve exact duplications of the chromosome number. However, polyploid transitions also involve the fusion of gametes with different ploidies. Two types of transitions were incorporated into model variants that allow for such possibilities. In the first, "demi-polyploidy," occurring at rate μ, permits multiplications of the number of chromosomes by 1.5 (Mayrose et al. 2010). This allows, for example, the generation of a hexaploid from a tetraploid lineage via the fusion of reduced and unreduced gametes, or from a diploid lineage in a two-step process via a triploid bridge followed by genome duplication. Note that demi-polyploidy transitions are well defined only for even haploid numbers, whereas for odd numbers, the transition rate is split between the two alternative possibilities. For example, a demi-polyploidy transition from a genome with $n = 9$ can either lead to $n = 13$ or to $n = 14$, both occurring at rate $\mu/2$. In this case, the triplication event would unrealistically entail a dysploidy event (either $9 \to 13 \to 26 \to 27$ or $9 \to 14 \to 28 \to 27$). Furthermore, this modeling scheme is inadequate for some polyploid transitions that involve the combination of genomes with high ploidy levels. For example, in a polyploid series ($n = 9, 18, 27, 36, 45$), such as that in *Chrysanthemum* (Liu et al. 2012), intercytotype matings that result in $18 \to 45$ or $27 \to 36$ transitions could not be obtained solely by any combination of demi-polyploidy and WGD events and would erroneously predict some additional dysploidy events (e.g., $18 \to 27 \to 54 \to 53 \to 52 \ldots \to 45$).

To overcome these shortcomings, a more general approach was introduced by Glick and Mayrose (2014) and is particularly beneficial for the analysis of clades that exhibit a wide range of ploidy levels, such as the plant genera *Festuca* ($n = 7, 14, 21, 28, 35$) or *Achillea* ($n = 9, 18, 27, 36$), whose chromosome numbers are linked by a common denominator that often represents the base number of the group. By incorporating two free parameters: β, the base number and ν, its respective transition rate, this model allows for any multiplication of an inferred base number to be added to the genome (fig. 3). For example, if $\beta = 9$, the transitions from a genome with $n = 9$ chromosomes to $n = 18$, $n = 27$, and $n = 36$ are allowed in a single step. Notably, this modeling scheme also comes with some shortcomings of its own. For example, it assumes that a clade is defined by a single base number. However, it is possible that due to dysploidy transitions, each subclade in an analyzed phylogeny would be characterized by its own base number or that some subclades would exhibit multiple base numbers. This is in contrast to demi-polyploidy transitions that explicitly account for the current chromosome number of a lineage. Thus, in clades where dysploidy transitions are common, it is conceivable that models that incorporate demi-polyploidy transitions would be better supported than those that include only transitions by an inferred base number. However, because the transitions allowed by the two modeling approaches do not entirely overlap, it is possible that in large clades, in which a large number of parameters could be supported, the inclusion of both transition types would be beneficial.

Dependency of the Transition Rates on the Current Number of Chromosomes

The basic *chromEvol* model assumes that dysploidy and polyploidy transitions occur at rates that are identical for all lineages. This implicitly assumes, for example, that a descending dysploidy transition occurring in an $n = 20$ lineage is equally likely as that of $n = 10$, whereas in reality, fusion events may be more likely to occur in genomes with large numbers of chromosomes and, accordingly, that the transition rates are related to the number of chromosomes in a lineage. Although this possibility was not yet incorporated for modeling polyploidy transitions, several possibilities were suggested for dysploidy transitions. For example, as implemented in the model by Hallinan and Lindberg (2011), dysploidy rates are forced to linearly depend on the current number of chromosomes (i.e., that the ascending dysploidy rate is λi, where i is the number of chromosomes in the genome and λ is the ascending dysploidy rate). An alternative implementation, incorporated within *chromEvol*, allows this dependency to be tuned using additional free parameters: constant rate parameters, λ and μ (for ascending and descending dysploidy, respectively), and rate modifier parameters, λ_l and μ_l, that describes the extent of linear dependency. In this case, for example, the ascending dysploidy rate in lineages with i chromosomes is $\lambda + \lambda_l(i - 1)$. A different representation, which allows dysploidy rates to vary exponentially, rather than linearly, as a function of the current chromosome number, was suggested by Freyman and Höhna (2018). Although these options allow for more flexibility in modeling dysploidy transitions, in an analysis of 100 plant genera models that incorporated linear dependency of dysploidy rates on the current number of chromosomes were chosen in merely 2% of the data sets (Glick and Mayrose 2014). This result suggests that the probability of chromosome fission and fusion events is comparable across genomes with different number of chromosomes. However, it is also possible that the latter analysis suffered from low statistical power since all genera analyzed were fairly small (<100 species). It is thus possible that analyses with larger clades would support the additional model parameters and would allow to determine whether dysploidy rates are more likely to increase in genomes with higher number of chromosomes.
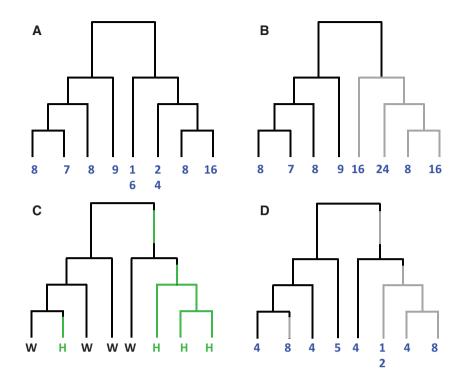
FIG. 4.—Heterogeneous models of chromosome-number evolution. (*A*) The standard *chromEvol* model assumes that chromosome-number dynamics are similar throughout the phylogeny. In this example, however, the subtree on the left is characterized by low rate of polyploidization and high dysploidy rates, whereas the subtree on the right is representative of a hot spot of polyploidizations. If this clade is analyzed using a single rate matrix (represented by a single color to all branches) the model would not fit the data well and would possibly result in erroneous inferences. (*B*) The use of a split model allows examining whether distinct patterns of chromosome-number change are exhibited in different taxonomic clades. Here, the subtree on the right is a priori classified to have a rate matrix that is distinct from that assumed for the rest of the tree, as represented by gray and black branches, respectively. (*C* and *D*) The effect of a character trait on rates of chromosome-number change. (*C*) A hypothetical mapping describing the evolution of a character trait (here, growth form). W and H denote woody and herbaceous states, respectively, whereas the black and green branches represent the corresponding times spent in each state. In (*D*), this mapping induces distinct patterns of chromosome-number change, represented by black and gray lines, respectively. This could capture a scenario where the polyploidization rate of herbaceous lineages is higher than that of woody lineages and could be modeled using a joint model for the evolution of both chromosome numbers and discrete character traits. Numbers at the tips represent chromosome-number assignments.

Nonhomogeneous Processes

The *chromEvol* models detailed thus far assume that the transition pattern is identical throughout the phylogeny. This time-homogeneity assumption is rather unlikely, especially when large phylogenies that include several distinct subclades are analyzed. In such cases, a more realistic approach would allow shifts in the transition pattern: either when different dynamics of chromosome number change are dictated by the presence of a certain organismal trait or when different transition patterns characterize different subclades of the phylogeny (fig. 4).

Márquez-Corro et al. (2019) tested for clade-specific shifts in the pattern of chromosome-number evolution in the large plant family Cyperaceae. The authors found three prominent shifts in the transition process: in the species-rich "non-Siderostictae *Carex*" clade, dysploidy transitions were found to be very frequent; the FAEC clade was typified by karyotype stability with very low rates of dysploidy and polyploidy; the C4 *Cyperus* clade, nested within the FAEC clade, exhibited

high rates of both dysploidy and polyploidy transitions (allowing for both demi-polyploidy and exact duplications). In the remaining phylogeny, the model inferred frequent descending dysploidies, low rates of ascending dysploidy, and additionally allowed for base-number transitions.

Notably, in the analysis of Márquez-Corro et al. (2019), the original phylogeny was artificially pruned into several smaller trees, and likelihood computations were then performed on each one independently. Thus, information that can be obtained from the pruned trees, which could affect the likelihood computation (i.e., probabilities of ancestral states at the base of each pruned tree) was lost. A more general approach would incorporate the possibility of rate heterogeneity across the phylogeny using a "split model" where different parts of the phylogeny evolve according to different transition patterns of chromosome-number change. Though not yet implemented in any existing modeling environment, this is certainly an area of future development. In addition, such an implementation will allow to pinpoint subclades whose transition

patterns deviate the most from the rest of the phylogeny without the need to a priori assign subclades to unique rate matrices. This may be done in a sequential testing approach, similar to branch-site codon models that aim to identify episodic positive selection along certain lineages (Anisimova and Yang 2007), or methods that identify groups with altered diversification patterns along the tree (Alfaro et al. 2009).

Two studies developed modeling extensions that associate patterns of chromosome-number change with the evolution of a discrete character trait (Zenil-Ferguson et al. 2017; Blackmon et al. 2019). Both of these models were developed using the coevolutionary model of Pagel (1994), such that the state space and the data of tip taxa are represented by the pair <chromosome number, character state>. Under the joint model, each character state induces a unique pattern of chromosome-number change with the respective rate parameters, and thus several free parameters are added to the model. For example, assuming a character trait with two possible states (denoted 0 and 1), and the basic *chromEvol* model that allows for dysploidy and WGD transitions, the joint model includes a total of eight free parameters: $(\rho_0, \lambda_0, \mu_0)$ and $(\rho_1, \lambda_1, \mu_1)$ that specify the chromosome-number transition pattern under the two states, and two parameters that indicate the rate of change between the two character states. Hypothesis testing can then be derived by defining reduced models with various constraints on the free parameters. For example, a null model in which $\rho_0 = \rho_1$ allows statistical testing of the hypothesis that the polyploidization rate is associated with the examined trait. Using such an approach, Zenil-Ferguson et al. (2017) inferred that in eudicots the frequency of both polyploidy and dysploidy is far more frequent in herbaceous plants compared with woody lineages. Notably, in this modeling extension, the transition matrix is largely expanded, which leads to a substantial increase in computational demands and practically limits analyses to character traits with two states and to clades in which the range of observed chromosome numbers is not large. To allow manageable computing times, Zenil-Ferguson et al. (2017) have reduced the size of the transition matrix by mapping any species with >50 chromosomes to an auxiliary "50+" state and further included two additional parameters to represent transition rates between lineages with "50+" chromosomes under the two trait states. Potentially, this limitation may be overcome by using a joint model that first samples probable histories of the character trait and, conditioned on the sampled history, computes the likelihood of the chromosome-number data using a *chromEvol* transition matrix of standard size, as has been developed in the context of joint genotype–phenotype models (Mayrose and Otto 2011; Levy Karin et al. 2017).

All probabilistic models detailed above assume that chromosome-number changes occur continuously in time at rates that are proportional to the branch lengths of the phylogeny. However, polyploidy and dysploidy transitions could frequently lead to reproductive incompatibilities and thus their occurrence should be coupled in time with speciation events (Coyne and Orr 2004), implying their clustering around branching events of the phylogeny. Zhan et al. (2016) have examined over a large cohort of plant genera whether polyploid transitions are temporally associated with speciation by comparing models that allow polyploid transitions to occur alongside speciation events (cladogenesis transitions) to those that assume that they occur continuously over time within a lineage (anagenesis transitions), providing some support for the former. This study was based on a two-step analysis, whereby the ploidy levels of extant lineages were first inferred using *chromEvol* and then probabilistic state speciation and extinction (SSE) models (Goldberg and Igic 2012) were applied to infer the proportion of ploidy shifts that are cladogenetic or anagenetic. A unified model, termed *ChromoSSE*, that distinguishes between cladogenetic and anagenetic chromosome-number transitions for both polyploid and dysploid events was presented by Freyman and Höhna (2018), and implemented within the RevBayes phylogenetic framework (Höhna et al. 2016). This model directly allows determining whether chromosome-number transitions in a specific clade occur primarily within lineages, primarily around speciation events, or in a combination of both processes. The *ChromoSSE* model further employs a reversible jump Markov chain Monte Carlo technique that simultaneously considers the space of all possible models (i.e., any combination of parameters governing the process of chromosome-number change). This alleviates the need to choose a particular set of parameters or to perform model selection, and has the additional benefit that the final inference inheritably incorporates model uncertainty and is not condition on a single model. Applying this model to five plant groups, clade-specific combinations of cladogenetic and anagenetic processes were observed (Freyman and Höhna 2018). Notably, similar to other SSE methods, the performance of the *ChromoSSE* model heavily relies on the accuracy and completeness of the phylogeny (Rabosky and Goldberg 2015; Beaulieu and O'Meara 2016; Louca and Pennell 2020; Shafir et al. 2020) and, since the model includes a large number of parameters, its inferences are expected to be reliable only when large clades are examined. Still, extensions of such models should be particularly exciting as they would allow testing long-standing hypotheses regarding the relationship between chromosome-number changes and patterns of lineage diversification.

Future Perspectives

Mechanistic phylogenetic models and comparative genomics offer a powerful combination to infer historical genomic processes. To date, a handful of studies have made use of both approaches to better understand the pathways by which the evolution of chromosome number proceeds

(Sousa and Renner 2015; Mandáková et al. 2017). Yet, considering that increasingly sophisticated methodologies are constantly being developed, it is expected that more and more studies will make use of a combined strategy in which computational predictions of karyotype evolution are contrasted with empirical data acquired using comparative (cyto)genomics. Here, we have summarized recent progress and highlighted several possibilities that can be incorporated in future modeling developments. Still, current models are focused on changes in chromosome numbers and thus explore only one aspect of karyotype evolution. Modeling other chromosomal characteristics would allow for finer analyses. For example, the *chromEvol* model could be applied to track the number of chromosome arms (the fundamental chromosome number) rather than simply the number of chromosomes. This has the potential to better differentiate between karyotype changes that are due to polyploidization versus multiple dysploidy events (Souza et al. 2015; Mandáková et al. 2017). Furthermore, the karyotype may be represented by the number of chromosomes belonging to different morphological classes, determined by the centromere position (e.g., metacentric, acrocentric, or telocentric). A probabilistic model over this state space could then be formulated by constructing the basic set of allowed transitions (e.g., a Robertsonian translocation would decrease the number of acrocentric chromosomes by two and increase the number of metacentrics by one). A more detailed representation could distinguish the chromosome arms by denoting each arm with a letter. White et al. (2010) used the latter representation to reconstruct intraspecific phylogenetic networks among multiple chromosomal races of the house mouse and common shrew. Their method is built upon the computation of the distance between any two examined karyotypes (defined by the number of whole-arm reciprocal translocations and Robertsonian fusions and fissions that are required to transform one karyotype to another). The usability of such an approach for among-species analyses is still unclear, but a main limitation of its widespread use is data availability, since detailed karyotype representations are not available for many species. Moreover, this approach requires that homeology relationships among chromosome arms are resolved, which is not trivial when increasingly distant lineages are compared. Another fruitful endeavor should be to model the evolution of chromosome numbers together with other informative genomic attributes. One obvious candidate is genome size, which is expected to increase simultaneously with chromosome number after a polyploidization event, but could take on different trajectories at longer time intervals. For example, genome size itself could influence patterns of chromosome-number change as additional rounds of polyploidizations are less likely to occur in genomes of larger size (Zenil-Ferguson et al. 2016), whereas increased repeat copy numbers theoretically provide more abundant substrates for DSB misrepair underlying dysploid changes. Certainly, such future methodological advancements would allow for refined understanding of how major genomic events, such as dysploidy and polyploidy, have shaped the karyotype of extant and ancestral lineages.

Acknowledgments

Literature Cited

Ahola V, et al. 2014. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nat Commun. 5:4737.

Albert P, et al. 2019. Whole-chromosome paints in maize reveal rearrangements, nuclear domains, and chromosomal relationships. Proc Natl Acad Sci U S A. 116(5):1679–1685.

Alfaro ME, et al. 2009. Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. Proc Natl Acad Sci U S A. 106(32):13410–13414.

Anisimova M, Yang Z. 2007. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol. 24(5):1219–1228.

Báez M, et al. 2019. Together but different: the subgenomes of the bimodal eleutherine karyotypes are differentially organized. Front Plant Sci. 10:1170.

Beaulieu JM, O'Meara BC. 2016. Detecting hidden diversification shifts in models of trait-dependent speciation and extinction. Syst Biol. 65(4):583–601.

Belser C, et al. 2018. Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. Nat Plants. 4(11):879–887.

Betekhtin A, Jenkins G, Hasterok R. 2014. Reconstructing the evolution of *Brachypodium* genomes using comparative chromosome painting. PLoS One 9(12):e115108.

Bi Y, et al. 2020. Flexible chromosome painting based on multiplex PCR of oligonucleotides and its application for comparative chromosome analyses in *Cucumis*. Plant J. 102(1):178–186.

Birchler JA, Han F. 2018. Barbara McClintock's unsolved chromosomal mysteries: parallels to common rearrangements and karyotype evolution. Plant Cell 30(4):771–779.

Blackmon H, Justison J, Mayrose I, Goldberg EE. 2019. Meiotic drive shapes rates of karyotype evolution in mammals. Evolution 73(3):511–523.

Cao HX, et al. 2016. The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. New Phytol. 209(1):354–363.

Carta A, Bedini G, Peruzzi L. 2020. A deep dive into the ancestral chromosome number and genome size of flowering plants. New Phytol. 228: 1097–1106.

Castiglione M, Cremonini R. 2012. A fascinating island: 2n= 4. Plant Biosyst. 146:711–726.

Chamala S, et al. 2013. Assembly and validation of the genome of the nonmodel basal angiosperm *Amborella*. Science 342(6165):1516–1517.

Chen F, et al. 2019. Genome sequences of horticultural plants: past, present, and future. Hortic Res. 6:112.

Clark JW, Donoghue PCJ. 2018. Whole-genome duplication and plant macroevolution. Trends Plant Sci. 23(10):933–945.

Clausen RE, Cameron DR. 1944. Inheritance in *Nicotiana tabacum*. XVIII. Monosomic analysis. Genetics 29:447–44777.

Comai L. 2005. The advantages and disadvantages of being polyploid. Nat Rev Genet. 6(11):836–846.

Coyne JA, Orr HA. 2004. Speciation. Sunderland (MA): Sinauer Associates.

Dassanayake M, et al. 2011. The genome of the extremophile crucifer *Thellungiella parvula*. Nat Genet. 43(9):913–918.

Edger PP, et al. 2018. Single-molecule sequencing and optical mapping yields an improved genome of woodland strawberry (*Fragaria vesca*) with chromosome-scale contiguity. GigaScience 7(2):1–7.

Edger PP, et al. 2020. Reply to: revisiting the origin of octoploid strawberry. Nat Genet. 52(1):5–7.

Emadzade K, et al. 2014. Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). Ann Bot. 114(8):1597–1608.

Felenstein J. 2004. Inferring phylogenies. Sundeland (MA): Sinauer Associates.

Filiault DL, et al. 2018. The *Aquilegia* genome provides insight into adaptive radiation and reveals an extraordinarily polymorphic chromosome with a unique history. Elife 7:e36426.

Freyman WA, Höhna S. 2018. Cladogenetic and anagenetic models of chromosome number evolution: a Bayesian model averaging approach. Syst Biol. 67(2):195–215.

Geiser C, Mandáková T, Arrigo N, Lysak MA, Parisod C. 2016. Repeated whole-genome duplication, karyotype reshuffling, and biased retention of stress-responding genes in buckler mustard. Plant Cell 28(1):17–27.

Glick L, Mayrose I. 2014. ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Mol Biol Evol. 31(7):1914–1922.

Goldberg EE, Igic B. 2012. Tempo and mode in plant breeding system evolution. Evolution 66(12):3701–3709.

Guerra M. 2016. Agmatoploidy and symploidy: a critical review. Genet Mol Biol. 39(4):492–496.

Guerra M. 2008. Chromosome numbers in plant cytotaxonomy: concepts and implications. Cytogenet Genome Res. 120(3–4):339–350.

Hallinan NM, Lindberg DR. 2011. Comparative analysis of chromosome counts infers three paleopolyploidies in the mollusca. Genome Biol Evol. 3:1150–1163.

Han J, Zhang Z, Wang K. 2018. 3C and 3C-based techniques: the powerful tools for spatial genome organization deciphering. Mol Cytogenet. 11:21.

Han Y, Zhang T, Thammapichai P, Weng Y, Jiang J. 2015. Chromosome-specific painting in *Cucumis* species using bulked oligonucleotides. Genetics 200(3):771–779.

Hansen AK, et al. 2006. Phylogenetic relationships and chromosome number evolution in *Passiflora*. Syst Bot. 31(1):138–150.

Hanson RE, et al. 1995. Fluorescent in situ hybridization of a bacterial artificial chromosome. Genome 38(4):646–651.

Hardigan MA, et al. 2020. Genome synteny has been conserved among the octoploid progenitors of cultivated strawberry over millions of years of evolution. Front Plant Sci. 10:1789.

Hipp AL. 2007. Nonuniform processes of chromosome evolution in sedges (Carex: Cyperaceae). Evolution 61(9):2175–2194.

Hloušková P, Mandáková T, Pouch M, Trávníček P, Lysak MA. 2019. The large genome size variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. Ann Bot. 124(1):103–120.

Ho SS, Urban AE, Mills RE. 2020. Structural variation in the sequencing era. Nat Rev Genet. 21(3):171–189.

Hoang PNT, et al. 2018. Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping, and Oxford Nanopore technologies. Plant J. 96(3):670–684.

Höhna S, et al. 2016. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. Syst Biol. 65(4):726–736.

Hu Q, Maurais EG, Ly P. 2020. Cellular and genomic approaches for exploring structural chromosomal rearrangements. Chromosome Res. 2020:1–12.

Huelsenbeck JP, Crandall KA. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu Rev Ecol Syst. 28(1):437–466.

Idziak D, et al. 2011. Painting the chromosomes of *Brachypodium* – current status and future prospects. Chromosoma 120(5):469–479.

Jankowska M, et al. 2015. Holokinetic centromeres and efficient telomere healing enable rapid karyotype evolution. Chromosoma 124(4):519–528.

Jiang J. 2019. Fluorescence in situ hybridization in plants: recent developments and future applications. Chromosome Res. 27(3):153–165.

Jiang J, Gill BS. 2006. Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome 49(9):1057–1068.

Jiang J, Gill BS, Wang GL, Ronald PC, Ward DC. 1995. Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci U S A. 92(10):4487–4491.

Kersey P. 2019. Plant genome sequences: past, present, future. Curr Opin Plant Biol. 48:1–8.

Khandelwal S. 1990. Chromosome evolution in the genus *Ophioglossum* L. Bot J Linn Soc. 102(3):205–217.

Koshi JM, Goldstein RA. 1996. Probabilistic reconstruction of ancestral protein sequences. J Mol Evol. 42(2):313–320.

Kurzhals RL, et al. 2017. Chromosome healing is promoted by the telomere cap component Hiphop in *Drosophila*. Genetics 207(3):949–959.

Levy Karin E, Wicke S, Pupko T, Mayrose I. 2017. An integrated model of phenotypic trait changes and site-specific sequence evolution. Syst Biol. 66(6):917–933.

Liu PL, Wan Q, Guo YP, Yang J, Rao GY. 2012. Phylogeny of the genus *Chrysanthemum* L.: evidence from single-copy nuclear gene and chloroplast DNA sequences. PLoS One 7(11):e48970.

Liu X, et al. 2020. Dual-color oligo-FISH can reveal chromosomal variations and evolution in *Oryza* species. Plant J. 101(1):112–121.

Louca S, Pennell MW. 2020. Extant timetrees are consistent with a myriad of diversification histories. Nature 580(7804):502–505.

Luo MC, et al. 2009. Genome comparisons reveal a dominant mechanism of chromosome number reduction in grasses and accelerated genome evolution in Triticeae. Proc Natl Acad Sci U S A. 106(37):15780–15785.

Lysak MA. 2014. Live and let die: centromere loss during evolution of plant chromosomes. New Phytol. 203(4):1082–1089.

Lysak MA, et al. 2006. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. Proc Natl Acad Sci U S A. 103(13):5224–5229.

Lysak MA, Fransz PF, Ali HBM, Schubert I. 2001. Chromosome painting in *Arabidopsis thaliana*. Plant J. 28(6):689–697.

Lysak MA, Koch MA, Pecinka A, Schubert I. 2005. Chromosome triplication found across the tribe Brassiceae. Genome Res. 15(4):516–525.

Macas J, et al. 2015. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. PLoS One 10(11):e0143424.

Mandáková T, Lysak MA. 2016. Painting of *Arabidopsis* chromosomes with chromosome-specific BAC clones. Curr Protoc Plant Biol. 1(2):359–371.

Mandáková T, et al. 2017. Multispeed genome diploidization and diversification after an ancient allopolyploidization. Mol Ecol. 26(22):6445–6462.

Mandáková T, Pouch M, Brock JR, Al-Shehbaz IA, Lysak MA. 2019. Origin and evolution of diploid and allopolyploid *Camelina* genomes were accompanied by chromosome shattering. Plant Cell 31:2596–2612.

Mandáková T, Singh V, Kraemer U, Lysak MA. 2015. Genome structure of the heavy metal hyperaccumulator *Noccaea caerulescens* and its stability on metalliferous and nonmetalliferous soils. Plant Physiol. 169(1):674–689.

Márquez-Corro JI, Martín-Bravo S, Spalink D, Luceño M, Escudero M. 2019. Inferring hypothesis-based transitions in clade-specific models of chromosome number evolution in sedges (Cyperaceae). Mol Phylogenet Evol. 135:203–209.

Mayr E. 1982. Speciation and macroevolution. Evolution 36(6):1119–1132.

Mayrose I, et al. 2011. Recently formed polyploid plants diversify at lower rates. Science 333(6047):1257–1257.

Mayrose I, Barker MS, Otto SP. 2010. Probabilistic models of chromosome number evolution and the inference of polyploidy. Syst Biol. 59(2):132–144.

Mayrose I, Otto SP. 2011. A likelihood method for detecting trait-dependent shifts in the rate of molecular evolution. Mol Biol Evol. 28(1):759–770.

Michael TP, VanBuren R. 2020. Building near-complete plant genomes. Curr Opin Plant Biol. 54:26–33.

Nowak M, et al. 2020. The genome of *Draba nivalis* shows signatures of adaptation to the extreme environmental stresses of the Arctic. Mol Ecol Resour. Advance Access published October 15, 2020, doi: 10.1111/1755-0998.13280.

Oginuma K, Munzinger J, Tobe H. 2006. Exceedingly high chromosome number in Strasburgeriaceae, a monotypic family endemic to New Caledonia. Plant Syst Evol. 262(1–2):97–101.

Ohi-toma T, et al. 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. Syst Bot. 31(3):481–492.

Otto SP. 2007. The evolutionary consequences of polyploidy. Cell 131(3):452–462.

Page SL, Shaffer LG. 1998. Chromosome stability is maintained by short intercentromeric distance in functionally dicentric human Robertsonian translocations. Chromosome Res. 6(2):115–122.

Pagel M. 1994. Detecting correlated evolution on phylogenies – a general-method for the comparative-analysis of discrete characters. Proc R Soc Lond B Biol Sci. 255:37–45.

Pecinka A, et al. 2004. Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana* are predominantly random except for NOR-bearing chromosomes. Chromosoma 113(5):258–269.

Van De Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. Nat Rev Genet. 18(7):411–424.

Pellestor F. 2019. Chromoanagenesis: cataclysms behind complex chromosomal rearrangements. Mol Cytogenet. 12:6.

Pupko T, Pe'er I, Shamir R, Graur D. 2000. A fast algorithm for joint reconstruction of ancestral amino acid sequences. Mol Biol Evol. 17(6):890–896.

Rabosky DL, Goldberg EE. 2015. Model inadequacy and mistaken inferences of trait-dependent speciation. Syst Biol. 64(2):340–355.

Ramsey J, Schemske DW. 2002. Neopolyploidy in flowering plants. Annu Rev Ecol Syst. 33(1):589–639.

Ramsey J, Schemske DW. 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu Rev Ecol Syst. 29(1):467–501.

Rastogi S, Ohri D. 2020. Karyotype evolution in cycads. Nucleus 63(2):131–141.

Rice A, et al. 2015. The Chromosome Counts Database (CCDB) – a community resource of plant chromosome numbers. New Phytol. 206(1):19–26.

Rockman MV, Rowell DM. 2002. Episodic chromosomal evolution in *Planipapillus* (Onychophora: Peripatopsidae): a phylogenetic approach to evolutionary dynamics and speciation. Evolution 56(1):58–69.

Röser M. 2015. Mitosis and interphase of the highly polyploid palm *Voanioala gerardii* ($2n = 606 \pm 3$). Cytogenet Genome Res. 147:70–79.

Schmidt T, Heitkam T, Liedtke S, Schubert V, Menzel G. 2019. Adding color to a century-old enigma: multi-color chromosome identification unravels the autotriploid nature of saffron (*Crocus sativus*) as a hybrid of wild *Crocus cartwrightianus* cytotypes. New Phytol. 222(4):1965–1980.

Schubert I, Fransz PF, Fuchs J, De Jong JH. 2001. Chromosome painting in plants. Methods Cell Sci. 23(1/3):57–69.

Schultheis LM. 2001. Systematics of *Downingia* (Campanulaceae) based on molecular sequence data: implications for floral and chromosome evolution. Syst Bot. 26:603–621.

Shafir A, Azouri D, Goldberg EE, Mayrose I. 2020. Heterogeneity in the rate of molecular sequence evolution substantially impacts the accuracy of detecting shifts in diversification rates. Evolution 74(8):1620–1639.

Šimoníková D, et al. 2019. Chromosome painting facilitates anchoring reference genome sequence to chromosomes in situ and integrated karyotyping in banana (*Musa* spp.). Front Plant Sci. 10:1503.

Soltis DE, Visger CJ, Marchant DB, Soltis PS. 2016. Polyploidy: pitfalls and paths to a paradigm. Am J Bot. 103(7):1146–1166.

Sousa A, Renner SS. 2015. Interstitial telomere-like repeats in the monocot family Araceae. Bot J Linn Soc. 177(1):15–26.

Souza G, Crosa O, Guerra M. 2015. Karyological, morphological, and phylogenetic diversification in Leucocoryne Lindl (Allioideae, Amaryllidaceae). Plant Syst Evol. 301(8):2013–2023.

Stebbins GL. 1938. Cytological characteristics associated with the different growth habits in the dicotyledons. Am J Bot. 25(3):189–198.

Suda J, Krahulcová A, Trávníček P, Krahulec F. 2006. Ploidy level versus DNA ploidy level: an appeal for consistent terminology. Taxon 55(2):447–450.

The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814):796–815.

Tran TD, et al. 2017. Chromosome identification for the carnivorous plant *Genlisea margaretae*. Chromosoma 126(3):389–397.

Uhl C. 1978. Chromosomes of Mexican *Sedum*, II. Section *Pachysedum*. Rhodora 80:491–512.

VanBuren R, et al. 2018. A near complete, chromosome-scale assembly of the black raspberry (*Rubus occidentalis*) genome. Gigascience 7(8):giy094.

Wang W, et al. 2014. The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nat Commun. 5(1):1–13.

White TA, Bordewich M, Searle JB. 2010. A network approach to study karyotypic evolution: the chromosomal races of the common shrew

(*Sorex araneus*) and house mouse (*Mus musculus*) as model systems. Syst Biol. 59(3):262–276.

Willing E, et al. 2015. Genome expansion of *Arabis alpina* linked with retrotransposition and reduced symmetric DNA methylation. Nature 1:14023.

Woo S, et al. 1994. Construction and characterization of bacterial artificial chromosome library of *Sorghum bicolor*. Nucleic Acids Res. 22(23):4922–4931.

Wood TE, et al. 2009. The frequency of polyploid speciation in vascular plants. Proc Natl Acad Sci U S A. 106(33): 13875–13879.

Xin H, et al. 2020. An extraordinarily stable karyotype of the woody *Populus* species revealed by chromosome painting. Plant J. 101(2):253–264.

Zenil-Ferguson R, Ponciano JM, Burleigh JG. 2016. Evaluating the role of genome downsizing and size thresholds from genome size distributions in angiosperms. Am J Bot. 103(7):1175–1186.

Zenil-Ferguson R, Ponciano JM, Burleigh JG. 2017. Testing the association of phenotypes with polyploidy: an example using herbaceous and woody eudicots. Evolution 71(5):1138–1148.

Zhan SH, Drori M, Goldberg EE, Otto SP, Mayrose I. 2016. Phylogenetic evidence for cladogenetic polyploidization in land plants. Am J Bot. 103(7):1252–1258.

Zhang P, Li W, Friebe B, Gill BS. 2008. The origin of a 'zebra' chromosome in wheat suggests nonhomologous recombination as a novel mechanism for new chromosome evolution and step changes in chromosome number. Genetics 179(3):1169–1177.

Associate editor: Yves Van De Peer