PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Review

Cite this article: Blommaert J. 2020 Genome size evolution: towards new model systems for old questions. *Proc. R. Soc. B* **287**: 20201441. http://dx.doi.org/10.1098/rspb.2020.1441

Received: 18 June 2020
Accepted: 29 July 2020

Subject Category:
Evolution

Subject Areas:
evolution, genetics, genomics

Keywords:
genomics, next-generation sequencing, long-read sequencing, experimental evolution

Author for correspondence:
Julie Blommaert
e-mail: julie.blommaert@ebc.uu.se

THE ROYAL SOCIETY
PUBLISHING

Genome size evolution: towards new model systems for old questions

Julie Blommaert

Department of Organismal Biology, Uppsala University, Uppsala, Sweden

JB, 0000-0003-1411-2313

Genome size (GS) variation is a fundamental biological characteristic; however, its evolutionary causes and consequences are the topic of ongoing debate. Whether GS is a neutral trait or one subject to selective pressures, and how strong these selective pressures are, may remain open questions. Fundamentally, the genomic sequences responsible for this variation directly impact the potential evolutionary outcomes and, equally, are the targets of different evolutionary pressures. For example, duplications and deletions of genic regions (large or small) can have immediate and drastic phenotypic effects, while an expansion or contraction of non-coding DNA is less likely to cause catastrophic phenotypic effects. However, in the long term, the accumulation or deletion of ncDNA is likely to have larger effects. Modern sequencing technologies are allowing for the dissection of these proximate causes, but a combination of these new technologies with more traditional evolutionary experiments and approaches could revolutionize this debate and potentially resolve many of these arguments. Here, I discuss an ambitious way forward for GS research, putting it in context of historical debates, theories and sometimes contradictory evidence, and highlighting the promise of combining new sequencing technologies and analytical developments with more traditional experimental evolution approaches.

1. Background

Genome size (GS) varies tremendously across eukaryotes (by at least five orders of magnitude), but is thought of as a stable trait within species [1]. Across animals, GS ranges from 19 Mbp in a parasitic nematode (*Pratylenchus coffeae*) to 130 Gbp in the marbled lungfish [2], from 10 Mbp (*Pterothamnion plumula*, an algae) to 149 Gbp (*Paris japonica*) in plants [3], and from 9 Mbp to 178 Mbp in fungi [4]. This variation is not linked with any measure of organismal complexity [1,5]. This observation was originally termed the 'C-value paradox', but is now more commonly known as the 'C-value enigma', since the proximate cause of GS variation generally seems to be ncDNA, especially repetitive elements [6,7]. Additionally, other examples of GS variation, such as those associated with developmental stages [8], cell types (e.g. [9]) or sex (e.g. [10]), are undoubtedly interesting and biologically important, but fall outside the scope of this review and could be extensively discussed in reviews of their own.

(a) Terminology

The terminology around GS, non-coding DNA (ncDNA) and genomic function has been fraught with changes and controversies, as well as efforts to unite terminologies. GS itself has been defined in different ways by different authors, but is often thought of as the '1C' value [6,7,11,12], or the amount of DNA, in picograms, in a haploid nucleus of an organism. However, depending on the GS measurement method and the ploidy level of the organism, this cannot always be defined. Often, the total nuclear DNA content is simply divided by two and reported as the 1C value, though this assumes diploidy when there may not be evidence to support this. For a full review of GS terminology as well as recommendations for general usage, see Greilhuber *et al.* [12].

Perhaps more controversy has surrounded the term 'junk' DNA and definitions of function, especially of ncDNA. The term 'junk DNA' was first used by Susumo Ohno [13] and is sometimes considered a catch-all term for ncDNA, including repetitive elements and regulatory regions [14–17]. Some argue that the term junk is unsuitable for ncDNA since it implies no past, present or future function, but junk in everyday use often refers to objects which one day had or might in future have a use [14]. This junk is often kept since it is not actively harmful, so there is no perceived need to eliminate it. The same could be said of genomic sequences referred to as junk [18]. Since some ncDNA has regulatory functions [19,20], there is debate on how to define the function of a given sequence, and whether or not function should influence whether a sequence should be considered junk [14]. The most controversial attempted definition of the function of ncDNA was by the ENCODE foundation, which asserted that most of the human genome is functional since a large part is transcribed to RNA under at least one condition in at least one cell type [15,21]. Evolutionary biologists tend to have different considerations regarding function, based mostly around sequence conservation over time [14]. Both of these approaches have pitfalls, namely that many of the transcription events described by ENCODE could be transcriptional noise [15], and that the evolutionary approach may miss novel sequences or misclassify others as functional despite undetected loss of function [14]. Another consideration when considering function is whether that function is beneficial to the organism or the sequence alone. In the strictest sense, evolution involves only the transmission of DNA, so on a genomic level, it is possible to identify elements which are functional in that they promote their own transmission, but may actually be detrimental to the organism they are found in. These types of sequences are often called 'selfish DNA' [22]. For a thorough and recent discussion of the different considerations of 'functions' readers should consult the recent review by Linquist *et al.* [23].

In summary, many terms around GS and its evolution are as hotly debated as the genomic causes and evolutionary consequences of this trait. Each of these terms has a useful place in this discussion, but it is important to be clear about their usage in whatever context they are being used.

(b) Evolutionary theories of genome size variation

The evolutionary causes and consequences of GS change are widely discussed and contentious. Some argue that GS is an adaptive trait, while others prefer, in the absence of overwhelming evidence of selection, the null hypothesis of neutral evolution of GS.

The neutral theories of GS evolution posit that GS is mainly a product of genetic drift and that selective pressures play no role, or a minimal role, in the accumulation or loss of DNA. The two main (nearly) neutral theories of GS evolution are the mutational hazard hypothesis (MHH) and the mutational equilibrium hypothesis (MEH) [24,25]. Both suggest that DNA accumulation occurs only by drift, but have different explanations of DNA loss. The MHH, considered a nearly neutral theory, suggests that 'extra' DNA is very slightly deleterious, and that mutation rates are higher in larger genomes. It predicts, specifically, that organisms with larger effective population size (N_e) have smaller genomes since selection on slightly deleterious 'excess' DNA is more effective as population size increases [24]. On the other hand, the MEH argues that GS reflects a balance between insertions and deletions into the genome, with different rates in different genomes. It suggests that genome expansion happens in 'bursts' through duplications or transposon activity, while a more constant rate of small deletions mediates genome shrinkage [25]. Testing these hypotheses is challenging, and sometimes studies reach opposing conclusions. For example, an examination of GS and mutational rates in salamanders found that even though salamanders have larger genomes than frogs, they have a lower mutation rate, the exact opposite of the predictions made by the MHH [26]. However, on broader evolutionary scales, the predictions of the MHH do find some support (e.g. [27,28]). The MEH garners perhaps more support, with rapid genome expansions often being caused by 'bursts' of activity by transposable elements (TEs), which have been slowly counteracted by constant species-specific rates of small deletions [29,30]. An analysis across birds and mammals has also offered support for the MEH and suggested that GS evolution follows an 'accordion model' whereby TE-driven genome expansions are soon followed by DNA losses, the rates of which are driven mainly by differences in life history and N_e [31]. These two theories, the MHH and MEH, suggest that GS itself is not a trait under selection, but rather that it is influenced largely by genetic drift, weak selective forces on other processes or gradual processes of the genome. It seems neutral evolutionary forces play a role in GS evolution; however, their importance (especially at the species level) remains unclear [32,33].

Other hypotheses suggest GS can be a (mal)adaptive trait, through impacting phenotypic traits including body size, developmental time and other cell size-related effects. Two of these selective hypotheses, the nucleotypic [34,35] and the nucleoskeletal hypothesis [36,37], focus explicitly on GS impacts on cell biology and size. Both of these hypotheses are based on the correlation of GS with nuclear and cell volume, but make slightly different predictions. The nucleoskeletal hypothesis focuses on the ratio between nuclear and cytoplasmic volume and the effects of this on cell division times and metabolic rates [36,37]. On the other hand, the nucleotypic hypothesis focuses on only the association between GS and cell size, and the implications this has, especially on organismal growth [34,35]. In general, there is a relationship between GS and overall cell volume (e.g. [17,38,39]). In support of the nucleotypic hypotheses, studies have found correlations between GS and traits such as general cell size, reproductive cell size, body size and growth rates, although these have been across large phylogenetic distances (e.g. [40,41]). Support for the nucleotypic hypothesis can be found in examples of parasites with minimal genomes and strong selective pressures for rapid cell division and fast metabolism [36,42]. A related hypothesis, the genome streamlining hypothesis, suggests that metabolic resources such as phosphorus (P) and nitrogen (N) are important in GS selection [43]. This hypothesis assumes that under P and N limitation, large GSs will be at a disadvantage because these are major components of DNA. This seems to be true in environments with extremely limited amounts of P and N [44,45]. There is some evidence for selective pressures influencing GS under extreme environmental conditions and over broad evolutionary scales; however, it is not apparent if these forces are strong at the species level or under less extreme conditions.

There is evidence for the roles of both neutral and selective forces in GS evolution, and resolving these seemingly

conflicting hypotheses has relied mostly on comparisons across large phylogenetic distances. By focusing on population-level differences in GS, it is possible to draw conclusions about the importance of selection versus drift without the confounding effects of large phylogenetic distances, but there are few examples of intraspecific GS variation and its phenotypic effects. Causes and consequences of GS variation are particularly well understood in maize, with a recent study finding that GS was selected for via its effects on flowering time at different altitudinal clines, which is consistent with the nucleotypic hypothesis [46]. There are examples of intraspecific GS variation in animals (e.g. stick insects [47], snapping shrimp [48], flour beetles [49] and snails [50]), but investigations into the evolutionary causes and consequences of such variation remain sparse. One example suggests that small intraspecific GS changes are linked with reproductive fitness in seed beetles [51]. Since the evolutionary history and genomic basis of GS variation can influence the phenotypic responses (see next section), understanding this is crucial to interpreting experiments aimed at testing evolutionary hypotheses robustly.

(c) Causes of genome size variation

GS changes occur through various mechanisms, each of which has its own impact on organismal phenotype and fitness.

GS, as nuclear DNA content, can increase by a change in ploidy (i.e. whole genome duplications) or smaller duplication events. Such discrete changes in DNA content come with changes outlined above regarding the nucleotypic and nucleoskeletal hypotheses, and other phenotypic disruptions including meiotic disruptions [52], incorrect gene dosage and cytoplasmic incompatibility leading to speciation [53–55]. However, gen(om)e duplication can facilitate the evolution of duplicated genes because one copy can maintain normal function, while the other evolves to perform new functions with minimal phenotypic consequences [56]. Polyploidy can have drastic impacts on organisms in the short and long term, and is an evolutionary boon in many cases, but not all [55,57,58].

Other examples of chromosome-level impacts on GS include supernumerary chromosomes (B-chromosomes) and heterochromatic knobs. B-chromosomes are usually smaller than regular A-chromosomes, segregate independently at meiosis, often exhibit meiotic drive and sometimes mitotic instability, and are usually derived from regular A-chromosomes [59]. Although B-chromosomes can contain genes, they are more often highly repetitive, consisting mostly of TEs and satellite repeats [60–62]. Because of their tendency towards the meiotic drive, B-chromosomes can quickly spread through populations causing changes in GS and are usually considered as selfish genetic elements [63]. In many cases, there seems to be a mechanism by which meiotic drive is suppressed or reversed [64]. In some fungi, there are large accessory regions which seem to harbour beneficial sequences [65]. Additionally, B-chromosomes could face selective pressure because of their high repetitive element load and the nucleotypic or mutational hazard effects discussed above. Similar patterns of drive and phenotypic effects can be caused by heterochromatic knobs [66,67], which are found on regular chromosomes. They tend to be densely packed with TEs, though some genes, especially those associated with centromeres [67,68], can be found in heterochromatic knobs [69].

More gradual GS changes are usually due to amplification of repetitive DNA (either tandem repeat sequences or TEs) [70]. Each of these changes is individually quite small, but they can accumulate and cause drastic genome expansions over relatively short evolutionary time scales (e.g. [71]). Satellite DNA can play structural roles in chromosomes, such as telomeres and centromeres [19], and its proliferation is passive, often through DNA polymerase slippage [72]. Most tandemly repeated sequences are found in short arrays randomly distributed through the genome. Transposons are another type of repetitive DNA and can actively amplify themselves throughout the genome, either via their own transposase enzymes, or by recruiting those of other elements, and are also considered to be selfish genetic elements [73–75]. Harmful effects from transposon activity include interruption of functional genes and loss of function, overall increased mutation rates and disruption of gene expression [75,76]. However, there can be positive influences of TEs on organisms (e.g. placental gene regulation [77]). Repetitive elements can be found in most genomes, even the most streamlined [6], and can cause negative and positive impacts on the host organism through their sequence alone, and contribute to broader evolutionary impacts on genomes, their size and dynamics.

Many mechanisms, which sometimes interact, are responsible for GS change. For example, polyploidization can trigger bursts of TE activity via loss of epigenetic repression, leading to further genome expansion and instability, but eventually this is counteracted by increased DNA loss [78]. Recently, in catfish, the proliferation of TEs was associated with two whole genome duplications [79]. These dynamics can influence direct phenotypic outcomes of GS changes and long-term evolutionary consequences. It is therefore essential to understand the genomic basis of genome expansion or contraction before testing evolutionary hypotheses and extrapolating across broader evolutionary scales.

2. Methodology

(a) Genome size estimates by sequencing

Measuring GS accurately is obviously important for understanding GS evolution. Many recent studies use whole genome sequencing outputs such as genome assembly size or kmer counts [80,81], but these methods have some drawbacks. Despite decades of progress with genome sequencing and assembly, there are very few examples of complete eukaryotic genome assemblies [82,83]. Genome assemblies tend to be shorter than the genomes they represent, and especially lacking in repetitive regions because these regions are difficult to assemble, especially with sort-read sequencing. Genome assembly size should not be treated as synonymous with GS, though it often is in [84–86]. Estimates based on kmer frequencies can be more accurate, but polyploidy and highly repetitive regions can lead to mis-estimations [81,87]. There are also documented sequencing biases depending on GC content and technology, which can lead to no representation or under-representation of large parts of the genome [88–90]. Sequencing data can give information on GS, but it is not always precise or accurate, and certain genomic features can be underestimated with each method. Many limitations can be overcome with long-read and long-range sequencing to resolve ‘genomic dark regions’ that harbour the repetitive sequences which so often cause GS changes while simultaneously hindering sequencing, assembly and analysis [91–93].

(b) Laboratory-based genome size estimates

Measuring DNA amounts directly is a reliable method which can avoid many of the above issues, and gold standards already exist. This involves staining the DNA in cells and measuring the intensity of the staining in comparison to a known standard. Feulgen densitometry [94,95] and GS estimates by flow cytometry (FCS) are the most common [96]. Dye selection is important as nucleotide-binding specificity can lead to inaccurate estimations [97]. Feulgen densitometry can lack accuracy, and FCS involves specialist, expensive equipment. Since accurate GS measurements are of utmost importance when considering this trait, FCS with an internal standard of known GS is the current gold standard for these measurements [96–98]. These methods, while offering precise and reliable GS estimates, can only provide current insights, and not show whether or not genomes are expanding or contracting.

(c) Genome content analyses

Understanding the forces which have shaped the measured GS relies on annotating genome content and features. In non-model organisms, and those with no close relatives sequenced, genome analyses can be difficult [99,100]. Determining gene content requires transcriptomic data. If these data are from related species, gene annotations can proceed using homology, but some genes (especially species-specific genes) may be falsely missed [100]. This may lead to incorrect conclusions about missing genes' contribution to genome contraction [101,102]. Accurate gene annotation is also crucial for measuring intron size, which is predicted to be correlated with GS [24].

Correctly annotating repetitive regions of the genome is even more challenging. Some regions will simply not be present in the genome assembly, or will be in small fragments which are difficult to analyse [84]. The repetitive regions which are present in the genome assembly could be incorrectly assembled, fragmented or represent multiple copies collapsed into a single copy. Many highly repetitive regions, such as centromeres and other long tracts of repeats, are being resolved by new, long-read technologies [103], but problems still exist [104]. Additionally, such technologies are still more expensive than short-read technologies, which can be especially costly for larger genomes that might harbour more repetitive regions [85]. Aside from this, repeats are annotated using databases, lineage-specific, novel repetitive elements could be missed [105]. A number of tools have been developed for *de novo* repeat annotation. Some of these assemble only high-copy sequences from the sequencing reads, annotate these sequences via homology to repeat databases and then quantify their numbers or proportion in the genome [106,107]. Other strategies identify high-copy regions by aligning the genome assembly to itself and again classifying by homology to known repeats [108–110]. This requires a high-quality genome assembly and can be time-consuming on large or highly repetitive genomes. All automatic classification methods are prone to mis- or non-classification of repeat families especially for species distantly related to well-annotated genomes, and are often still likely to underestimate repetitive content of genomes, although time-consuming manual curation can help address these issues [105,111]. Some validation can be attempted through laboratory-based methods such as qPCR or chromosome painting [112,113]. There are some efforts focusing on TE discovery and annotation using machine learning, and these may address many of the current difficulties (e.g. [114–116]). Once repeat families are identified and classified, their accumulation or deletion within or between populations with differing GS can be considered.

In addition to gene and TE annotations to identify GS changes, one must consider how genomes contract. Thorough repeat annotation allows for the identification of signatures of DNA loss through transposon inactivation (e.g. LTR elements reduced to Solo-LTRs [117]). In addition, whole-genome and local alignments of closely related species can identify deletions (both large and small) in relation to the other genomes. Such approaches have identified varying rates of deletions across birds and mammals [31,118]. The balance between DNA loss and gain is pivotal in GS evolution, so comparative genome analyses to answer these questions should consider both.

Understanding GS variation, its causes and consequences requires reliable methodology, but there are still clear challenges to be overcome. From the accuracy of laboratory- and sequence-based GS measurements, to genome annotation and content analysis, there are pitfalls and best practices to consider at each point. There is an ongoing discussion about this regarding genome sequencing, assembly and analyses (e.g. [84,109]), but gold standards and best practices are still evolving, even among large consortia producing high-quality genome assemblies [119,120].

3. Potential model organisms for genome size evolution

Most studies outlined in §1 attempting to support or refute the evolutionary theories of GS evolution focus on either theoretical evolutionary models or correlation studies between GS and a variety of other traits and statistics, at varying levels of evolutionary distance, sometimes focusing on the extremes of GS. However, the combination of genomics and evolutionary experiments utilizing examples of intraspecific GS variation can offer a fruitful platform for resolving some of the debates about GS evolution. By selecting model systems which have intraspecific GS variation and are tractable to both modern and traditional evolutionary biology approaches, we can test many of the debated hypotheses without the confounding effects often seen across large phylogenetic distances. In fact, such integrative approaches have strong proponents across evolutionary biology [121,122]. While there are many examples of intraspecific GS variation in eukaryotes, both in plants and animals, there are relatively few recent examples of genomic dissection of such variation. Further, even fewer of these present opportunity for experimentation which will lead to a deeper and more concrete understanding of the evolution of GS. Such an ambitious approach, which has been successfully employed in at least one pathogenic fungal species, revealed that accessory chromosomes which exhibit meiotic drive [123] are also associated with genome instability and an increase in virulence and fitness [124]. Here, I consider a further few promising model systems for testing the causality of these correlations through direct manipulation of both the environment and GS, to see the outcome on a phenotypic level.

(a) Selected examples of intraspecific genome size variation

There is undoubtedly a multitude of examples of intraspecific GS variation in eukaryotes, but here I select three which show

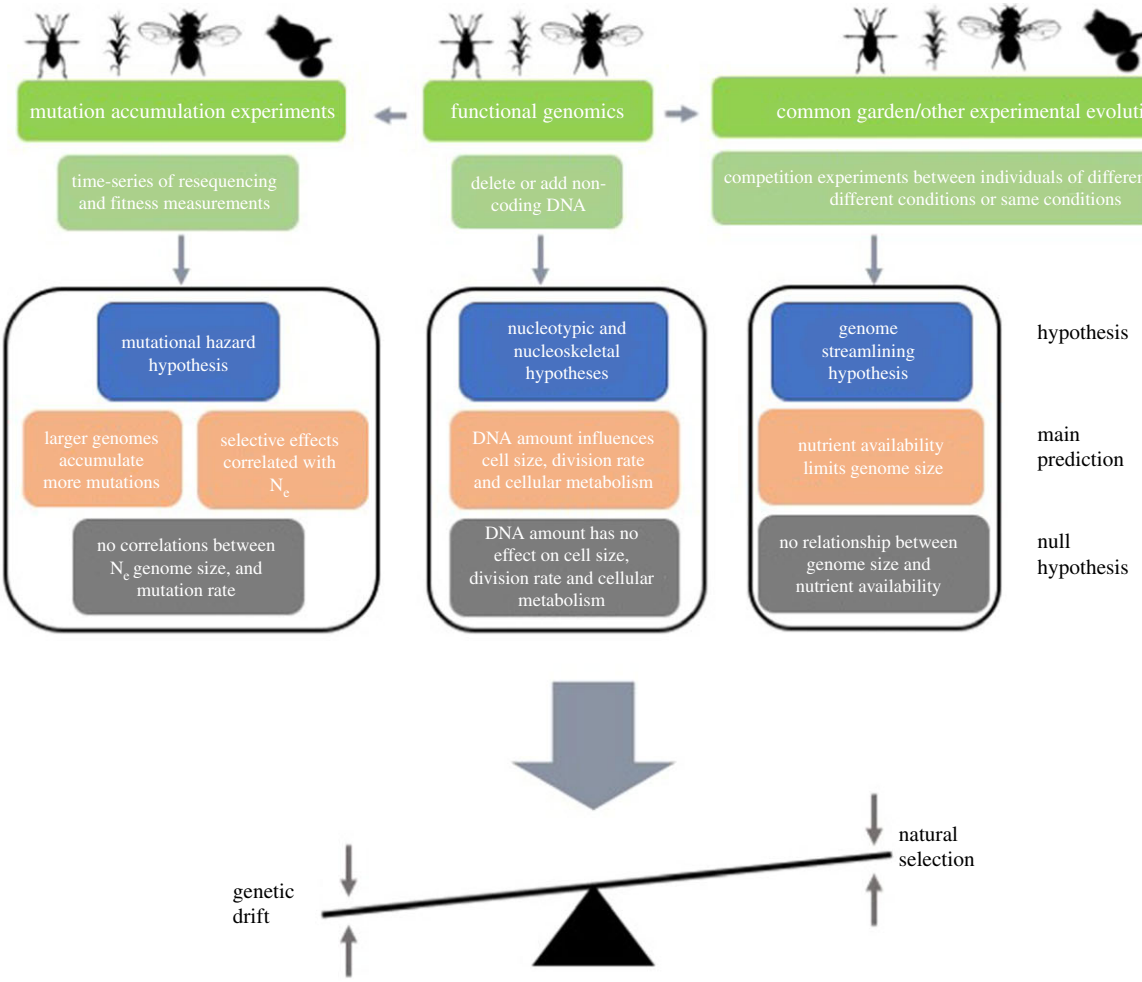

Figure 1. An overview of the evolutionary hypotheses explaining GS, their predictions and null hypotheses along with proposed experiments to test these in new model systems and develop a new overview of the influences of genetic drift and natural selection on GS. Each model species is represented by a silhouette of itself from phylopic.org. Green boxes are general experimental categories, with more detailed experiments in light green; blue contains the previously proposed evolutionary theories, along with their main predictions (peach) and null hypotheses (grey). (Online version in colour.)

exceeding promise for becoming model study systems. This section should not be seen as a prescriptive list, but as a means to open discussion and the development of new study systems, questions and approaches.

In regard to GS and content and intraspecific changes, my first example, maize (*Zea mays*) is possibly the closest to a model organism. It is perhaps not a coincidence that the organism in which TEs were first described [125,126] also displays GS change caused by transposons within and between species [127,128]. Additionally, maize is tractable to experimental evolution, and especially common garden experiments and resequencing [129], as well as functional genomics approaches such as CRISPR/Cas9 [130]. These particular features together with the vast resources and community involved in maize research means it is well placed for experimental evolution approaches to resolving the debates around GS evolution.

The rotifer species complex *Brachionus plicatilis* [131,132] exhibits dramatic sevenfold GS variations [133,134]. Rotifers have been ecological study systems for a long time, but have recently also become the focus of genetic and genomic studies, as well as model systems for evolutionary questions. The GS changes observed in *B. plicatilis* are largely due to transposable element accumulation [135]. Additionally, within one population of one species in this complex, *B. asplanchnoidis*, GS varies up to 1.9-fold. Initial work suggests the presence of accessory or B-chromosomes causing these variations [136]. The genomic content and the biological and evolutionary effects of these elements may provide insights into their impacts on the genome and organism. While functional genomics approaches, including CRISPR/Cas9, remain out of reach for rotifer biologists, the natural variation in GS, and the ability to cross rotifer clones of differing GSs [137], results in a continuous complement of GS variants across a presumably homogenous genetic background, providing an ideal platform for testing evolutionary hypotheses about GS. In fact, such approaches have already revealed a seeming lower limit to GS in this species, but no apparent upper limit [136].

Another example of within- and between-species GS variation can be found in the genus *Tribolium*, or flour beetles [49], though it is not as dramatic as in rotifers. Despite a similar pattern being found in related beetles [51], correlations between GS and reproductive fitness traits in both cases, and extensive use of *Tribolium* as a genetic and developmental model system [138], little focus has been devoted to understanding the genomic basis of this GS variation. However, comprehensive analyses of the repeat content of the *Tribolium castaneum* genome found high-repeat content [139] and variations in satellite DNA between populations [140]. These data combined with the position of *Tribolium* as a developmental and genetic model organism with functional genetics tools [141] mean it is well placed for further investigations into GS variation and its proximate causes and ultimate consequences.

(b) How do we pick new model organisms for genome size studies?

Existing model organisms in biology tend to be tractable to standardized protocols for rearing and analyses, and have built up infrastructure and a research community with extensive resources [142]. Some of these model organisms may be suitable for GS evolution studies, while other non-model species may present larger opportunities, even without the infrastructure, community and resources of traditional model systems. Choosing appropriate model systems requires thought about the specific question being addressed, and with new technologies and resources, one is no longer limited to traditional systems [143].

When considering the questions regarding GS evolution, one must return to the evolutionary hypotheses (outlined in §1) and consider ways to answer the outstanding questions combining traditional evolutionary experiments and modern techniques. Such an approach requires the identification of ideal model systems. While many correlational studies have found support for the nucleotypic [34,35] and nucleoskeletal [36,37] hypotheses, perhaps more resounding support could come from functional genetics approaches, where GS could be manipulated by tools such as CRISPR/Cas9 or cross-breeding to see the impact on nuclear and cell size, and growth rates of cells and whole organisms. These manipulated GSs could be used in further experiments. While this may sound ambitious, CRISPR/Cas9 has been used to remove entire chromosome arms [144] and inactivate TE copies throughout genomes [145,146], though not with the express aim of manipulating GS. The mutational hazard hypothesis [26] could be tested by mutation accumulation experiments. Since N_e can be difficult to measure in natural populations, mutation accumulation experiments in species with intraspecific GS variation can be used to decrease N_e drastically and magnify the effects [147] of drift in relation to GS. The dramatic GS differences and ease of laboratory culture of *Brachionus* rotifers make them an ideal model in this case, but both maize and flour beetles would probably also be suitable. The genome streamlining hypothesis [43] could be tested with experimental approaches where nutrient conditions are manipulated to analyse the impact on organismal growth rate in various natural and laboratory-manipulated GS variants. Clearly, with examples of intraspecific GS variation, genomic resources to understand the basis of these changes and their potential direct phenotypic effects, traditional evolutionary experiments and newer technologies, many of these controversies may become more (or less) clear. The integrative approach outlined here and are summarized in figure 1, targeting multiple hypotheses in these systems would probably not result in the support of any individual hypothesis, but offer varying levels of support for all, leading to a new view of GS evolution and the importance of both neutral and selective forces.

Ultimately, answering these questions is unlikely with just one technique in one model system. The ongoing revolutions in sequencing technologies and analyses, combined with the other techniques outlined in this review, will only reveal more genomic oddities and the mutational forces underlying these. The future may yet hold more, not fewer controversies regarding the influences of drift and selection on genome evolution, which will doubtless be the focus of many discussions among evolutionary biologists.

Data accessibility. This article has no additional data.

Competing interests. I declare that I have no competing interests.

Funding. I acknowledge the support by a Postdoc Grant of SciLifeLab Uppsala via Alexander Suh.

Acknowledgements. I would like to thank the members of the Suh lab, including Alexander Suh, Valentina Peona, Augustin Chen, Octavio Palacios-Gimenez, Paco Ruiz-Ruano, Ivar Westberg and Roberto Rossini, for their feedback on the review at all stages, including proposal. Additionally, I would also like to thank my PhD supervisor, Claus-Peter Stelzer and the rest of his lab for the fruitful discussions during my PhD which led to this review article.

References

1. Gregory TR, DeSalle R. 2005 *The evolution of the genome*, pp. 585–675. London, UK: Academic Press. (doi:10.1016/B978-012301463-4/50012-7)
2. Gregory TR. 2005 Animal genome size database. http://www.genomesize.com (accessed 23 August 2019).
3. Leitch I, Johnston E, Pellicer J, Hidalgo O, Bennett MD. 2019 Plant DNA C-values database. Release 71 2019. See https://cvalues.science.kew.org/ (accessed 23 August 2019).
4. Mohanta TK, Bae H. 2015 The diversity of fungal genome. *Biol. Proced. Online* **17**, 8. (doi:10.1186/s12575-015-0020-z)
5. Wright SI. 2017 *Evolution of genome size*. Chichester, UK: John Wiley & Sons.
6. Elliott TA, Gregory TR. 2015 What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Phil. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **370**, 20140331. (doi:10.1098/rstb.2014.0331)
7. Gregory TR. 2005 The C-value enigma in plants and animals: a review of parallels and an appeal for partnership. *Ann. Bot.* **95**, 133–146. (doi:10.1093/aob/mci009)
8. Wang J, Davis RE. 2014 Programmed DNA elimination in multicellular organisms. *Curr. Opin Genet. Dev.* **27**, 26–34. (doi:10.1016/j.gde.2014.03.012)
9. Kinsella CM *et al.* 2019 Programmed DNA elimination of germline development genes in songbirds. *Nat. Commun.* **10**, 1–10. (doi:10.1038/s41467-019-13427-4)
10. Hjelmen CE, Blackmon H, Renee Holmes V, Burrus CG, Spencer Johnston J. 2019 Genome size evolution differs between *Drosophila* subgenera with striking differences in male and female genome size in *Sophophora*. *G3 Genes, Genomes, Genet* **9**, 3167–3179. (doi:10.1534/g3.119.400560)
11. Gregory TR. 2005 Genome size evolution in animals. In *The evolution of the genome* (ed. TR Gregory), pp. 3–87. London, UK: Academic Press.
12. Greilhuber J, Doležel J, Lysák MA, Bennett MD. 2005 The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. *Ann. Bot.* **95**, 255–260. (doi:10.1093/aob/mci019)
13. Ohno S. 1972 So much 'junk' DNA in our genome. *Brookhaven Symp. Biol.* **23**, 366–370.
14. Graur D, Zheng Y, Azevedo RBR. 2015 An evolutionary classification of genomic function. *Genome Biol. Evol.* **7**, 642–645. (doi:10.1093/gbe/evv021)
15. Doolittle WF. 2013 Is junk DNA bunk? A critique of ENCODE. *Proc. Natl Acad. Sci. USA* **110**, 5294–5300. (doi:10.1073/pnas.1221376110)
16. Doolittle WF, Brunet TDP. 2017 On causal roles and selected effects: our genome is mostly junk. *BMC Biol.* **15**, 1–9. (doi:10.1186/s12915-017-0460-9)
17. Gregory TR. 2001 Coincidence, coevolution, or causation? DNA content, cellsize, and the C-value enigma. *Biol Rev* **76**, 65–101. (doi:10.1111/j.1469-185X.2000.tb00059.x)
18. Palazzo AF, Gregory TR. 2014 The case for junk DNA. *PLoS Genet.* **10**, e1004351. (doi:10.1371/journal.pgen.1004351)

19. Shapiro JA, von Sternberg R. 2005 Why repetitive DNA is essential to genome function. *Biol. Rev.* **80**, 227–250. (doi:10.1017/S1464793104006657)
20. Huang S *et al.* 2014 Decelerated genome evolution in modern vertebrates revealed by analysis of multiple lancelet genomes. *Nat. Commun.* **5**, 5896. (doi:10.1038/ncomms6896)
21. ENCODE Project Consortium TEP. 2004 The ENCODE (ENCyclopedia Of DNA Elements) project. *Science* **306**, 636–640. (doi:10.1126/science.1105136)
22. Orgel L, Crick F. 1980 Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607. (doi:10.1038/288645a0)
23. Linquist S, Doolittle WF, Palazzo AF. 2020 Getting clear about the F-word in genomics. *PLoS Genet.* **16**, e1008702. (doi:10.1371/journal.pgen.1008702)
24. Lynch M, Conery JS. 2003 The origins of genome complexity. *Science* **302**, 1401–1404. (doi:10.1126/science.1089370)
25. Petrov DA. 2002 Mutational equilibrium model of genome size evolution. *Theor. Popul. Biol.* **61**, 531–544. (doi:10.1006/tpbi.2002.1605)
26. Mohlhenrich ER, Mueller RL. 2016 Genetic drift and mutational hazard in the evolution of salamander genomic gigantism. *Evolution (N Y)* **70**, 2865–2878. (doi:10.1111/evo.13084)
27. Kelkar YD, Ochman H. 2012 Causes and consequences of genome expansion in fungi. *Genome Biol. Evol.* **4**, 13–23. (doi:10.1093/gbe/evr124)
28. Sung W, Ackerman MS, Dillon MM, Platt TG, Fuqua C, Cooper VS, Lynch M. 2016 Evolution of the insertion-deletion mutation rate across the tree of life. *G3; Genes Genomes Genetics* **6**, 2583–2591. (doi:10.1534/g3.116.030890)
29. Mueller RL, Jockusch EL. 2018 Jumping genomic gigantism. *Nat. Ecol. Evol.* **2**, 1687–1688. (doi:10.1038/s41559-018-0703-3)
30. Canapa A, Barucca M, Biscotti MA, Forconi M, Olmo E. 2016 Transposons, genome size, and evolutionary insights in animals. *Cytogenet Genome Res.* **147**, 217–239. (doi:10.1159/000444429)
31. Kapusta A, Suh A, Feschotte C. 2017 Dynamics of genome size evolution in birds and mammals. *PNAS* **114**, 201616702. (doi:10.1073/pnas.1616702114)
32. Arkhipova IR. 2018 Neutral theory, transposable elements, and eukaryotic genome evolution. *Mol. Biol. Evol.* **35**, 1332–1337. (doi:10.1093/molbev/msy083)
33. Linquist S, Cottenie K, Elliott TA, Saylor B, Kremer SC, Gregory TR. 2015 Applying ecological models to communities of genetic elements: the case of neutral theory. *Mol. Ecol.* **24**, 3232–3242. (doi:10.1111/mec.13219)
34. Bennett MD. 1971 The duration of meiosis. *Proc. R. Soc. Lond. B* **178**, 277–299. (doi:10.1098/rspb.1971.0066)
35. Gregory TR, Hebert PDN. 1999 The modulation of DNA content: proximate causes and ultimate consequences. *Genome Res.* **9**, 317–324. (doi:10.1101/gr.9.4.317)
36. Cavalier-Smith T. 2005 Economy, speed and size matter: evolutionary forces driving nuclear genome miniaturization and expansion. *Ann. Bot.* **95**, 147–175. (doi:10.1093/aob/mci010)
37. Cavalier-Smith T. 1978 Nuclear volume control by nucleoskeletal DNA, selection for cell volume and cell growth rate, and the solution of the DNA C-value paradox. *J. Cell Sci.* **34**, 247–278.
38. Hardie DC, Hebert PDN. 2003 The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* **46**, 683–706. (doi:10.1139/g03-040)
39. Henry TA, Bainard JD, Newmaster SG. 2015 Genome size evolution in Ontario ferns (Polypodiidae): evolutionary correlations with cell size, spore size, and habitat type and an absence of genome downsizing. *Genome* **57**, 555–566. (doi:10.1139/gen-2014-0090)
40. Beaulieu JM, Leitch IJ, Patel S, Pendharkar A, Knight CA. 2008 Genome size is a strong predictor of cell size and stomatal density in angiosperms. *New Phytol.* **179**, 975–986. (doi:10.1111/j.1469-8137.2008.02528.x)
41. Šímová I, Herben T. 2012 Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proc. R. Soc. B* **279**, 867–875. (doi:10.1098/rspb.2011.1284)
42. Keeling PJ, Slamovits CH. 2005 Causes and effects of nuclear genome reduction. *Curr. Opin Genet. Dev.* **15**, 601–608. (doi:10.1016/j.gde.2005.09.003)
43. Hessen DO, Jeyasingh PD, Neiman M, Weider LJ. 2010 Genome streamlining and the elemental costs of growth. *Trends Ecol. Evol.* **25**, 75–80. (doi:10.1016/j.tree.2009.08.004)
44. Kang M, Wang J, Huang H. 2015 Nitrogen limitation as a driver of genome size evolution in a group of karst plants. *Sci. Rep.* **5**, 11636. (doi:10.1038/srep11636)
45. Bales AL, Hersch-Green EI. 2019 Effects of soil nitrogen on diploid advantage in fireweed, *Chamerion angustifolium* (Onagraceae). *Ecol. Evol.* **9**, 1095–1109. (doi:10.1002/ece3.4797)
46. Bilinski P *et al.* 2018 Parallel altitudinal clines reveal adaptive evolution of genome size in *Zea mays*. *PLoS Genet.* **14**, e1007162. (doi:10.1101/134528)
47. Marescalchi O, Scali V, Zuccotti M. 1998 Flow-cytometric analyses of intraspecific genome size variations in *Bacillus atticus* (Insecta, Phasmatodea). *Genome* **41**, 629–635. (doi:10.1139/g98-064)
48. Jeffery NW, Hultgren K, Chak STC, Gregory TR, Rubenstein DR. 2016 Patterns of genome size variation in snapping shrimp. *Genome* **59**, 393–402. (doi:10.1139/gen-2015-0206)
49. Alvarez-Fuster A, Juan C, Petitpierre E. 1991 Genome size in tribolium flour-beetles: inter-and intraspecific variation. *Genet. Res.* **58**, 1–5. (doi:10.1017/S0016672300029542)
50. Neiman M, Paczesniak D, Soper DM, Baldwin AT, Hehman G. 2011 wide variation in ploidy level and genome size in a new zealand freshwater snail with coexisting sexual and asexual lineages. *Evolution (N Y)* **65**, 3202–3216. (doi:10.1111/j.1558-5646.2011.01360.x)
51. Arnqvist G, Sayadi A, Immonen E, Hotzy C, Rankin D, Tuda M, Hjelmen CE, Johnston JS. 2015 Genome size correlates with reproductive fitness in seed beetles. *Proc. R. Soc. B* **282**, 20151421. (doi:10.1098/rspb.2015.1421)
52. Yant L, Hollister JD, Wright KM, Arnold BJ, Higgins JD, Franklin FCH, Bomblies K. 2013 Meiotic adaptation to genome duplication in *Arabidopsis arenosa*. *Curr. Biol.* **23**, 2151–2156. (doi:10.1016/j.cub.2013.08.059)
53. Otto SP. 2007 The evolutionary consequences of polyploidy. *Cell* **131**, 452–462. (doi:10.1016/j.cell.2007.10.022)
54. Comai L. 2005 The advantages and disadvantages of being polyploid. *Nat. Rev. Genet.* **6**, 836–846. (doi:10.1038/nrg1711)
55. Baduel P, Bray S, Vallejo-Marin M, Kolář F, Yant L. 2018 The 'polyploid hop': shifting challenges and opportunities over the evolutionary lifespan of genome duplications. *Front. Ecol. Evol.* **6**, 117. (doi:10.3389/fevo.2018.00117)
56. Ohno S. 1970 *Evolution by gene duplication*. Berlin, Germany: Springer Science & Business Media.
57. Parisod C, Holderegger R, Brochmann C. 2010 Evolutionary consequences of autopolyploidy. *New Phytol.* **186**, 5–17. (doi:10.1111/j.1469-8137.2009.03142.x)
58. Gerstein AC, Otto SP. 2009 Ploidy and the causes of genomic evolution. *J. Hered.* **100**, 571–581. (doi:10.1093/jhered/esp057)
59. Camacho JPM. 2005 B chromosomes. In *The evolution of the genome* (ed. TR Gregory), pp. 223–286. New York, NY: Academic Press.
60. Coan RLB, Martins C. 2018 Landscape of transposable elements focusing on the B chromosome of the cichlid fish *Astatotilapia latifasciata*. *Genes* (*Basel*) **9**, 269. (doi:10.3390/genes9060269)
61. Navarro-Domínguez B, Ruiz-Ruano FJ, Cabrero J, Corral JM, López-León MD, Sharbel TF, Camacho JPM. 2017 Protein-coding genes in B chromosomes of the grasshopper *Eyprepocnemis plorans*. *Sci. Rep.* **7**, 45200. (doi:10.1038/srep45200)
62. Ruiz-Ruano FJ, Navarro-Domínguez B, López-León MD, Cabrero J, Camacho JPM. 2019 Evolutionary success of a parasitic B chromosome rests on gene content. *BioRxiv* 683417. (doi:10.1101/683417)
63. Camacho JPM, Sharbel TF, Beukeboom LW. 2000 B-chromosome evolution. *Phil. Trans. R. Soc. Lond. B* **355**, 163–178. (doi:10.1098/rstb.2000.0556)
64. Manrique-Poyato MI, Cabrero J, López-León MD, Perfectti F, Gómez R, Camacho JPM. 2019 Interpopulation spread of a parasitic B chromosome is unlikely through males in the grasshopper *Eyprepocnemis plorans*. *Heredity* (*Edinb*) **124**, 197–206. (doi:10.1038/s41437-019-0248-5)
65. De Jonge R *et al.* 2012 Tomato immune receptor Ve1 recognizes effector of multiple fungal pathogens uncovered by genome and RNA sequencing. *Proc. Natl Acad. Sci. USA* **109**, 5110–5115. (doi:10.1073/pnas.1119623109)
66. Buckler IV ES, Phelps-Durr TL, Buckler CSK, Dawe RK, Doebley JF, Holtsford TP. 1999 Meiotic drive of

chromosomal knobs reshaped the maize genome. *Genetics* **153**, 415–426.

67. Ghaffari R, Cannon EKS, Kanizay LB, Lawrence CJ, Dawe RK. 2013 Maize chromosomal knobs are located in gene-dense areas and suppress local recombination. *Chromosoma* **122**, 67–75. (doi:10.1007/s00412-012-0391-8)
68. Mroczek RJ, Dawe RK. 2003 Distribution of retroelements in centromeres and neocentromeres of maize. *Genetics* **165**, 809–819.
69. Hines PJ. 2000 Heterochromatin knobs revisited. *Science* **287**, 1169e. (doi:10.1126/science.287.5456.1169e)
70. Kidwell MG. 2002 Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**, 49–63. (doi:10.1023/A:1016072014259)
71. Naville M, Henriet S, Warren I, Sumic S, Reeve M, Volff JN, Chourrout D. 2019 Massive changes of genome size driven by expansions of non-autonomous transposable elements. *Curr. Biol.* **29**, 1161–1168. (doi:10.1016/j.cub.2019.01.080)
72. Garrido-Ramos MA. 2015 Satellite DNA in plants: more than just rubbish. *Cytogenet Genome Res.* **146**, 153–170. (doi:10.1159/000437008)
73. Hurst GDD, Werren JH. 2001 The role of selfish genetic elements in eukaryotic evolution. *Nat. Rev. Genet.* **2**, 597–606. (doi:10.1038/35084545)
74. Feschotte C, Pritham EJ. 2007 DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368. (doi:10.1146/annurev.genet.40.110405.090448)
75. Bourque G *et al.* 2018 Ten things you should know about transposable elements. *Genome Biol.* **19**, 199. (doi:10.1186/s13059-018-1577-z)
76. Schrader L, Schmitz J. 2019 The impact of transposable elements in adaptive evolution. *Mol. Ecol.* **28**, 1537–1549. (doi:10.1111/mec.14794)
77. Haig D. 2012 Retroviruses and the placenta. *Curr. Biol.* **22**, R609–R613. (doi:10.1016/J.CUB.2012.06.002)
78. Vicient CM, Casacuberta JM. 2017 Impact of transposable elements on polyploid plant genomes. *Ann. Bot.* **120**, 195–207. (doi:10.1093/aob/mcx078)
79. Marburger S, Alexandrou MA, Taggart JB, Creer S, Carvalho G, Oliveira C, Taylor MI. 2018 Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes. *Proc. R. Soc. B* **285**, 20172732. (doi:10.1098/rspb.2017.2732)
80. Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P. 2007 Prediction of effective genome size in metagenomic samples. *Genome Biol.* **8**, R10. (doi:10.1186/gb-2007-8-1-r10)
81. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Berger B. 2017 GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204. (doi:10.1093/bioinformatics/btx153)
82. Schneider VA *et al.* 2017 Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.* **27**, 849–864. (doi:10.1101/gr.213611.116)
83. Peona V, Weissensteiner MH, Suh A. 2018 How complete are 'complete' genome assemblies?—an avian perspective. *Mol. Ecol. Resour.* **18**, 1188–1195. (doi:10.1111/1755-0998.12933)
84. Liu R, Bennetzen JL. 2008 Enchilada redux: how complete is your genome sequence? *New Phytol.* **179**, 249–250. (doi:10.1111/j.1469-8137.2008.02527.x)
85. Kutter C, Jern P, Suh A. 2018 Bridging gaps in transposable element research with single-molecule and single-cell technologies. *Mob. DNA* **9**, 34. (doi:10.1186/s13100-018-0140-5)
86. Treangen TJ, Salzberg SL. 2012 Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46. (doi:10.1038/nrg3117)
87. Marçais G, Kingsford C. 2011 A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770. (doi:10.1093/bioinformatics/btr011)
88. Chen YC, Liu T, Yu CH, Chiang TY, Hwang CC. 2013 Effects of GC bias in next-generation-sequencing data on de novo genome assembly. *PLoS ONE* **8**, e62856. (doi:10.1371/journal.pone.0062856)
89. Sato MP *et al.* 2019 Comparison of the sequencing bias of currently available library preparation kits for Illumina sequencing of bacterial genomes and metagenomes. *DNA Res.* **26**, 391–398. (doi:10.1093/dnares/dsz017)
90. Tilak M-K, Botero-Castro F, Galtier N, Nabholz B. 2018 Illumina library preparation for sequencing the GC-rich fraction of heterogeneous genomic DNA. *Genome Biol. Evol.* **10**, 616. (doi:10.1093/GBE/EVY022)
91. Rhoads A, Au KF. 2015 PacBio sequencing and its applications. *Genomics, Proteomics Bioinforma* **13**, 278–289. (doi:10.1016/j.gpb.2015.08.002)
92. Sović I, Križanović K, Skala K, Šikić M. 2016 Evaluation of hybrid and non-hybrid methods for de novo assembly of nanopore reads. *Bioinformatics* **32**, 2582–2589. (doi:10.1093/bioinformatics/btw237)
93. Lu H, Giordano F, Ning Z. 2016 Oxford Nanopore MinION sequencing and genome assembly. *Genomics, Proteomics Bioinforma* **14**, 265–279. (doi:10.1016/j.gpb.2016.05.004)
94. Feulgen R, Rossenbeck H. 1924 Mikroskopisch-chemischer Nachweis einer Nucleinsäure vom Typus der Thymonucleinsäure und die-darauf beruhende elektive Färbung von Zellkernen in mikroskopischen Präparaten. *Hoppe Seylers Z Physiol. Chem.* **135**, 203–248. (doi:10.1515/bchm2.1924.135.5-6.203)
95. Hardie DC, Gregory TR, Hebert PDN. 2002 From pixels to picograms: a beginners' guide to genome quantification by Feulgen image analysis densitometry. *J. Histochem. Cytochem.* **50**, 735–749. (doi:10.1177/002215540205000601)
96. Doležel J, Greilhuber J. 2010 Nuclear genome size: are we getting closer? *Cytom Part A* **77**, 635–642. (doi:10.1002/cyto.a.20915)
97. Greilhuber J. 2008 Cytochemistry and C-values: the less-well-known world of nuclear DNA amounts. *Ann. Bot.* **101**, 791–804. (doi:10.1093/aob/mcm250)
98. Greilhuber J. 1998 Intraspecific variation in genome size: a critical reassessment. *Ann. Bot.* **82**, 27–35. (doi:10.1006/anbo.1998.0725)
99. Ekblom R, Wolf JBW. 2014 A field guide to whole-genome sequencing, assembly and annotation. *Evol. Appl.* **7**, 1026–1042. (doi:10.1111/eva.12178)
100. Matz MV. 2018 Fantastic beasts and how to sequence them: ecological genomics for obscure model organisms. *Trends Genet.* **34**, 121–132. (doi:10.1016/j.tig.2017.11.002)
101. Lovell PV, Wirthlin M, Wilhelm L, Minx P, Lazar NH, Carbone L, Warren WC, Mello CV. 2014 Conserved syntenic clusters of protein coding genes are missing in birds. *Genome Biol.* **15**, 565. (doi:10.1186/s13059-014-0565-1)
102. Bornelöv S, Seroussi E, Yosefi S, Pendavis K, Burgess SC, Grabherr M, Warren WC, Mello CV. 2017 Correspondence on Lovell *et al.*: identification of chicken genes previously assumed to be evolutionarily lost. *Genome Biol.* **18**, 112. (doi:10.1186/s13059-017-1231-1)
103. Weissensteiner MH, Pang AWC, Bunikis I, Höijer I, Vinnere-Petterson O, Suh A, Wolf JBW. 2017 Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. *Genome Res.* **27**, 697–708. (doi:10.1101/gr.215095.116)
104. Peona V *et al.* 2019 Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. BioRxiv 2019.12.19.882399. (doi:10.1101/2019.12.19.882399)
105. Goerner-Potvin P, Bourque G. 2018 Computational tools to unmask transposable elements. *Nat. Rev. Genet.* **19**, 688–704. (doi:10.1038/s41576-018-0050-x)
106. Goubert C, Modolo L, Vieira C, Moro CV, Mavingui P, Boulesteix M. 2015 De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). *Genome Biol. Evol.* **7**, 1192–1205. (doi:10.1093/gbe/evv050)
107. Novák P, Neumann P, Pech J, Steinhaisl J, MacAs J. 2013 RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**, 792–793. (doi:10.1093/bioinformatics/btt054)
108. Volfovsky N, Haas BJ, Salzberg SL. 2001 A clustering method for repeat analysis in DNA sequences. *Genome Biol.* **2**, 1. (doi:10.1186/gb-2001-2-8-research0027)
109. Zeng L, Kortschak RD, Raison JM, Bertozzi T, Adelson DL. 2018 Superior ab initio identification, annotation and characterisation of TEs and segmental duplications from genome assemblies. *PLoS ONE* **13**, e0193588. (doi:10.1371/journal.pone.0193588)

110. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020 RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457. (doi:10.1073/pnas.1921046117)
111. Platt RN, Blanco-Berdugo L, Ray DA. 2016 Accurate transposable element annotation is vital when analyzing new genome assemblies. *Genome Biol. Evol.* **8**, 403–410. (doi:10.1093/gbe/evw009)
112. Albert PS, Zhang T, Semrau K, Rouillard J-MM, Kao Y-HH, Wang C-JRJR, Danilova TV, Jiang J, Birchler JA. 2019 Whole-chromosome paints in maize reveal rearrangements, nuclear domains, and chromosomal relationships. *Proc. Natl Acad. Sci. USA* **116**, 1679–1685. (doi:10.1073/pnas.1813957116)
113. Woodard LE, Li X, Malani N, Kaja A, Hice RH, Atkinson PW, Bushman FD, Craig NL, Wilson MH. 2012 Comparative analysis of the recently discovered hAT Transposon TcBuster in human cells. *PLoS ONE* **7**, e42666. (doi:10.1371/journal.pone.0042666)
114. Schietgat L, Vens C, Cerri R, Fischer CN, Costa E, Ramon J, Carareto CMA, Blockeel H. 2018 A machine learning based framework to identify and classify long terminal repeat retrotransposons. *PLoS Comput. Biol.* **14**, e1006097. (doi:10.1371/journal.pcbi.1006097)
115. Nakano FK, Pinto WJ, Pappa GL, Cerri R. 2017 Top-down strategies for hierarchical classification of transposable elements with neural networks. In *Proc. Int. Jt. Conf. Neural Networks*, *2017—May*. See https://lirias.kuleuven.be/2305810.
116. Yan H, Bombarely A, Li S. 2020 DeepTE: a computational method for de novo classification of transposons with convolutional neural network. *Bioinformatics*, btaa519. (doi:10.1093/bioinformatics/btaa519)
117. Vitte C, Panaud O. 2003 Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Mol. Biol. Evol.* **20**, 528–540. (doi:10.1093/molbev/msg055)
118. Nam K, Ellegren H. 2012 Recombination drives vertebrate genome contraction. *PLoS Genet.* **8**, e1002680. (doi:10.1371/journal.pgen.1002680)
119. Papanicolaou A. 2016 The life cycle of a genome project: perspectives and guidelines inspired by insect genome projects. *F1000Research* **5**, 18. (doi:10.12688/f1000research.7559.1)
120. Editorial. 2016 A reference standard for genome biology. *Nat. Biotechnol.* **36**, 1121. (doi:10.1038/nbt.4318)
121. Kawecki TJ, Lenski RE, Ebert D, Hollis B, Olivieri I, Whitlock MC. 2012 Experimental evolution. *Trends Ecol. Evol.* **27**, 547–560. (doi:10.1016/j.tree.2012.06.001)
122. Schlötterer C, Kofler R, Versace E, Tobler R, Franssen SU. 2015 Combining experimental evolution with next-generation sequencing: a powerful tool to study adaptation from standing genetic variation. *Heredity* (*Edinb*) **114**, 431–440. (doi:10.1038/hdy.2014.86)
123. Habig M, Kema GH, Holtgrewe Stukenbrock E. 2018 Meiotic drive of female-inherited supernumerary chromosomes in a pathogenic fungus. *Elife* **7**, e40251. (doi:10.7554/eLife.40251)
124. Möller M, Habig M, Freitag M, Stukenbrock EH. 2018 Extraordinary genome instability and widespread chromosome rearrangements during vegetative growth. *Genetics* **210**, 517–529. (doi:10.1534/genetics.118.301050)
125. McClintock B. 1950 The origin and behavior of mutable loci in maize. *Proc. Natl Acad. Sci. USA* **36**, 344–355. (doi:10.1073/pnas.36.6.344)
126. McClintock B. 1947 Mutable loci in maize. *Carnegie Institution of Washington Yearbook* **47**, 155–169.
127. Muñoz-Diez C, Vitte C, Ross-Ibarra J, Gaut BS, Tenaillon MI. 2012 *Using nextgen sequencing to investigate genome size variation and transposable element content*, pp. 41–58. Berlin, Germany: Springer.
128. Tenaillon MI, Hufford MB, Gaut BS, Ross-Ibarra J. 2011 Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*. *Genome Biol. Evol.* **3**, 219–229. (doi:10.1093/gbe/evr008)
129. Josephs EB, Berg JJ, Ross-Ibarra J, Coop G. 2019 Detecting adaptive differentiation in structured populations with genomic data and common gardens. *Genetics* **211**, 989–1004. (doi:10.1534/genetics.118.301786)
130. Gao H *et al.* 2020 Superior field performance of waxy corn engineered using CRISPR–Cas9. *Nat. Biotechnol.* **38**, 579–581. (doi:10.1038/s41587-020-0444-0)
131. Mills S *et al.* 2016 Fifteen species in one: deciphering the *Brachionus plicatilis* species complex (Rotifera, Monogononta) through DNA taxonomy. *Hydrobiologia* **796**, 39–58. (doi:10.1007/s10750-016-2725-7)
132. Michaloudi E, Mills S, Papakostas S, Stelzer CP, Triantafyllidis A, Kappas I, Vasileiadou K, Proios K, Abatzopoulos TJ. 2016 Morphological and taxonomic demarcation of *Brachionus asplanchnoidis* Charin within the *Brachionus plicatilis* cryptic species complex (Rotifera, Monogononta). *Hydrobiologia* **796**, 19–37. (doi:10.1007/s10750-016-2924-2)
133. Stelzer C-P. 2010 A first assessment of genome size diversity in monogonont rotifers. *Hydrobiologia* **662**, 77–82. (doi:10.1007/s10750-010-0487-1)
134. Stelzer C-P, Riss S, Stadler P. 2011 Genome size evolution at the speciation level: the cryptic species complex *Brachionus plicatilis* (Rotifera). *BMC Evol. Biol.* **11**, 90. (doi:10.1186/1471-2148-11-90)
135. Blommaert J, Riss S, Hecox-Lea B, Mark Welch DB, Stelzer CP. 2019 Small, but surprisingly repetitive genomes: transposon expansion and not polyploidy has driven a doubling in genome size in a metazoan species complex. *BMC Genomics* **20**, 466. (doi:10.1186/s12864-019-5859-y)
136. Stelzer C, Pichler M, Stadler P, Hatheuer A, Riss S. 2019 Within-population genome size variation is mediated by multiple genomic elements that segregate independently during meiosis. *BioRxiv*, 1–40. (doi:10.1101/623470)
137. Riss S, Arthofer W, Steiner FM, Schlick-Steiner BC, Pichler M, Stadler P, Stelzer C-P. 2017 Do genome size differences within *Brachionus asplanchnoidis* (Rotifera, Monogononta) cause reproductive barriers among geographic populations? *Hydrobiologia* **796**, 59–75. (doi:10.1007/s10750-016-2872-x)
138. Denell R. 2008 Establishment of tribolium as a genetic model system and its early contributions to evo-devo. *Genetics* **180**, 1779–1786. (doi:10.1534/genetics.104.98673)
139. Wang S, Lorenzen MD, Beeman RW, Brown SJ. 2008 Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome Biol.* **9**, R61. (doi:10.1186/gb-2008-9-3-r61)
140. Feliciello I, Akrap I, Brajković J, Zlatar I, Ugarković Đ. 2014 Satellite DNA as a driver of population divergence in the red flour beetle *Tribolium castaneum*. *Genome Biol. Evol.* **7**, 228–239. (doi:10.1093/gbe/evu280)
141. Gilles AF, Schinko JB, Averof M. 2015 Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle *Tribolium castaneum*. *Development* **142**, 2832–2839. (doi:10.1242/dev.125054)
142. Leonelli S, Ankeny RA. 2013 What makes a model organism? *Endeavour* **37**, 209–212. (doi:10.1016/j.endeavour.2013.06.001)
143. Russell JJ *et al.* 2017 Non-model model organisms. *BMC Biol.* **15**, 1–31. (doi:10.1186/s12915-017-0391-5)
144. Adikusuma F, Williams N, Grutzner F, Hughes J, Thomas P. 2017 Targeted deletion of an entire chromosome using CRISPR/Cas9. *Mol. Ther.* **25**, 1736–1738. (doi:10.1016/j.ymthe.2017.05.021)
145. Todd CD, Deniz Ö, Taylor D, Branco MR. 2019 Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. *Elife* **8**, e44344. (doi:10.7554/eLife.44344)
146. Yang L *et al.* 2015 Genome-wide inactivation of porcine endogenous retrovirus or other sequences in the pig genome endogenous retroviruses (PERVs). *Science* **350**, 1101–1104. (doi:10.1126/science.aad1191)
147. Halligan DL, Keightley PD. 2009 Spontaneous mutation accumulation studies in evolutionary genetics. *Annu. Rev. Ecol. Evol. Syst.* **40**, 151–172. (doi:10.1146/annurev.ecolsys.39.110707.173437)