The impact of epigenetic information on genome evolution

Soojin V. Yi and Michael A. D. Goodisman

School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

SVY, 0000-0003-1497-1871; MADG, 0000-0002-4842-3956

Epigenetic information affects gene function by interacting with chromatin, while not changing the DNA sequence itself. However, it has become apparent that the interactions between epigenetic information and chromatin can, in fact, indirectly lead to DNA mutations and ultimately influence genome evolution. This review evaluates the ways in which epigenetic information affects genome sequence and evolution. We discuss how DNA methylation has strong and pervasive effects on DNA sequence evolution in eukaryotic organisms. We also review how the physical interactions arising from the connections between histone proteins and DNA affect DNA mutation and repair. We then discuss how a variety of epigenetic mechanisms exert substantial effects on genome evolution by suppressing the movement of transposable elements. Finally, we examine how genome expansion through gene duplication is also partially controlled by epigenetic information. Overall, we conclude that epigenetic information has widespread indirect effects on DNA sequences in eukaryotes and represents a potent cause and constraint of genome evolution.

1. Epigenetic information as a mediator of molecular evolution

The success of multicellular organisms stems from the ability of genetically identical cells within individuals to undertake different functions. Epigenetics is the study of how different cellular functions can arise from cells that possess the same genotype [1,2]. The field of epigenetics has evolved over time [3]. Traditionally, epigenetics specifically focuses on heritable information that does not lead to changes in DNA sequence. Indeed, the very term epigenetics means ‘above the genome’. Thus, epigenetic systems represent modes of heritable information that operated in conjunction with, but distinct from, the traditional DNA-based system of inheritance.

Remarkably, however, epigenetic marks can lead to changes in DNA sequences through a variety of mechanisms. Epigenetic information, in the form of DNA methylation, modifications to histone proteins or non-coding RNAs, directly affects chromatin, which is the DNA–protein complex that makes up chromosomes in eukaryotic taxa [4]. The interactions and modifications to chromatin open the door to mutational processes that lead to alterations in DNA sequence and, potentially, evolutionary changes. The effects of epigenetic marks (i.e. changes to the molecular structure of DNA or histone proteins that impart epigenetic information) on genome sequences are underappreciated specifically because epigenetic information is defined as operating without changes to DNA sequences. This review aims to bring together evidence of how epigenetic marks directly and indirectly lead to DNA sequence changes.

2. DNA methylation is a pervasive mutagenic agent in diverse taxa

DNA methylation is a phylogenetically widespread chemical modification of genomic DNA. Two of the four DNA nucleotides, adenine and cytosine, can
be modified by the addition of a methyl group. Cytosine methylation is extremely widespread in both eukaryotes and prokaryotes, and has been extensively studied for its roles in gene regulation and genome evolution (e.g. [5–8], figure 1). Adenine methylation is relatively common in prokaryotes [11] and also observed in several eukaryotic species [12]. The function and evolutionary consequences of adenine methylation are currently poorly understood [12], especially in comparison to cytosine methylation. Consequently, in this review, we focus on cytosine methylation, and refer to it simply as ‘DNA methylation’.

DNA methylation can expand the information content encoded in genomic sequences and is associated with regulatory functions in diverse taxa [5–7]. Interestingly, DNA methylation is also highly mutagenic [13]. Methylated cytosines undergo spontaneous deamination and can thus transform into thymine. Since thymine is a naturally occurring nucleotide in DNA, these mutations may remain unrepaired, ultimately resulting in a C to T mutation [14]. Therefore, paradoxically, methylation of cytosine can deplete cytosines themselves due to the propensity of methylcytosines to mutate to thymines.

Figure 1. DNA methylation in the DNA double helix. Methylcytosine (yellow) with methyl group circled, alongside unmethylated cytosine (dark blue), guanine (light blue), adenine (red) and thymine (orange) bases. The methylation of DNA is one of the most important factors affecting mutation rate in animals. DNA structure [9] drawn in PyMOL [10].

(a) The effects of DNA methylation on CpG content in the genome

In animal genomes, the majority of DNA methylation occurs at cytosines followed by guanines, or ‘CpG’ dinucleotides. Because methylated cytosines in CpG dinucleotides are prone to mutations toTpG dinucleotides, methylated sequences gradually lose CpGs. In the simplest case, the frequency of CpGs in a sequence may be determined solely by the frequencies of C and G nucleotides in the sequence of interest. More specifically, one can calculate the expected frequency of CpGs as the \((\text{frequency of } C) \times (\text{frequency of } G)\) assuming that random associations of C and G nucleotides determine the CpG frequency. CpG frequencies of sequences from species lacking DNA methylation, such as Drosophila, are well approximated by this estimate. By contrast, CpGs tend to be depleted in species where there is DNA methylation, due to mutations caused by deamination of cytosine bases.

It is convenient to use the observed frequency of CpGs normalized by the expected frequency of CpGs, often referred to as ‘CpG O/E’, to describe the depletion of CpGs. CpG O/E can be calculated as (observed frequency of CpG)/(expected frequency of CG), or \([\text{Number of CG}] / [\text{Number of C} \times \text{Number of G}]\). It should be noted that CpG O/E is calculated from genomic DNA and, therefore, reflects a historical appraisal of the effects of DNA methylation. Moreover, CpG O/E is only affected by DNA methylation in germline, which leads to inherited changes in DNA sequence between generations. Researchers have long noted that CpG O/E is around 0.2 for most genomic regions (figure 2a), indicating that a very high percentage of cytosine bases have mutated to thymine due to the effects of DNA methylation. CpG O/E has been a highly useful metric to infer within-genome variation of DNA methylation and served as an important aid to functional studies [17,18].

(b) CpG islands and lessons learned from DNA methylation of vertebrate genomes

A prime example of how analysis of CpG frequency has guided epigenetic studies is the investigation of so-called CpG islands in vertebrates. CpG islands are so named to indicate the presence of a relatively high frequency of CpGs relative to surrounding regions displaying low CpG frequencies [19,20]. CpG islands were initially identified as genomic regions that possessed low levels of DNA methylation [19]. Because these regions are ‘hypomethylated’, they are likely to be immune to the mutagenic effect of DNA methylation, and thus retain CpGs at frequencies higher than methylated regions (figure 2a). CpG islands stand out because most human and other vertebrate genomes are heavily methylated.
Interestingly, CpG islands tend to be located near the 5' end of genes and are associated with active transcription [20,21]. Numerous studies have revealed the functional importance of CpG islands. For example, genes harbouring CpG islands near their transcription start sites tend to be broadly expressed [21–23]. The use of CpG islands has been critical in functional annotation of the human genome in the pre-genomic era [24] and to this day; CpG islands continue to be among the most widely studied epigenetic regulatory markers.

CpG islands offer a rare opportunity to investigate the relationship between DNA methylation and sequence evolution in the human genome. Efforts to define CpG islands from genomic sequences rely heavily on the observation that CpG O/E and GC contents are often strongly correlated [25]. This pattern is in large part driven by the depletion of CpG dinucleotides affecting GC-poor regions more dramatically than GC-rich regions [26].

Another mechanism that can affect both CpG O/E and GC content is biased gene conversion. Briefly, gene conversion occurs when homologous recombination results in a mismatched base pair. In many genomes, when such a mismatch is corrected, the correction is biased to a G/C allele over an A/T allele, a process referred to as ‘biased gene conversion’ [27,28]. Biased gene conversion can increase GC content, and could directly generate new CpGs. For example, Cohen et al. [29] showed that biased gene conversion could maintain high GC and high CpG O/E regions during primate genome evolution. Therefore, CpG O/E, while being extremely useful to identify putatively hypomethylated regions of vertebrate genomes, can also increase due to biased gene conversion [29,30]. In addition, since biased gene conversion increases with recombination, it is important to consider recombination rates in comparative analyses of GC content and CpG O/E.

(c) DNA methylation and isochore evolution of vertebrate genomes

DNA methylation may have affected the evolution of large regions of DNA showing uniform GC content; such regions are known as ‘isochores’. GC content of genome sequences...
shows considerable variation in many species, particularly in warm-blooded vertebrates including mammals and birds. It was initially suggested that GC content homogeneity stretched thousands of base pairs and formed isochores [31]. Several hypotheses have been put forward to explain the evolutionary origin and maintenance of isochores. For example, a selective hypothesis [32] posits that GC-rich isochores confer selective advantages in warm-blooded species because they are thermally more stable than GC-poor isochores. By contrast, two other hypotheses focus on variation in mutation rates and recombination as causal mechanisms. First, GC content could vary across the genome according to variation in rates and patterns of mutations [33]. Second, isochores may originate through biased gene conversion associated with variation in recombination rate across the genome [27,34].

DNA methylation can influence isochores by impacting mutation and recombination. As discussed above, DNA methylation can cause C to T (or G to A on the complementary strand) mutations due to spontaneous deamination. Fryxell & Zuckerkandl [35] noted that deamination requires a temporary melting of double-stranded DNA, and that more thermodynamic energy is required to melt GC-rich DNA than to melt AT-rich DNA. Consequently, genomic regions already high in GC content may experience lower rates of mutations originating from deamination of methylated DNA. Low-GC regions, on the other hand, can more easily undergo deamination and subsequent GC to AT mutations. This positive feedback loop can drive the evolution of GC-poor regions and further reduce GC content. By contrast, GC-rich regions may gain more GC nucleotides, thus generating and maintaining isochores [35]. This hypothesis has been supported by empirical data demonstrating the melting effects on substitutions [36], prevalence of single nucleotide polymorphisms in human genomes [37], and germline DNA methylation and substitution patterns in avian genomes [38].

DNA methylation may also have a direct link to recombination. Experimental studies have demonstrated that recombination following double-strand breaks leads to the recruitment of DNA methyltransferases, which are the key enzymes involved in methylating DNA, and the generation of methylated regions of the double-strand break [39,40]. At the genome scale, germline methylation levels are positively correlated with inferred recombination rates [41]. Furthermore, recombination can directly cause mutation in some cases [42]. However, how these processes are causatively linked to each other on an evolutionary timescale requires further research.

(d) Identifying methylated regions using CpG contents
It was long recognized that CpG O/E values of different animal genomes showed substantial variation, and that many invertebrate genomes generally had much higher CpG O/E values than other animal taxa [15,18]. The prevailing notion until the last decade was that this pattern arose because invertebrates generally lacked DNA methylation. However, several studies have since discovered that many invertebrate genomes contain all the genes necessary for a functional DNA methylation system and contain methylated CpGs [16,18,43–45]. Furthermore, taxonomically expanded sampling of invertebrate methylomes is revealing a great deal of diversity in genomic DNA methylation. Invertebrate genomes show nearly two orders of magnitude difference in DNA methylation. For example, honeybee genomes show less than 1% methylation [6]; by contrast, sponge genomes show approximately 80% methylation [46]. Some invertebrate genomes even exhibit DNA methylation of promoters and transposable elements (TEs) [46–48].

Studies of DNA methylation in invertebrates may benefit from the comparison of CpG O/E and experimentally measured DNA methylation, as the former reflects historical levels of DNA methylation in the germine, while the latter reveals current DNA methylation levels. These two measures are generally negatively correlated to each other across diverse plant and animal taxa [18,46,49,50] (although some exceptions warrant further study [48]). For example, in the human genome, most genomic regions are heavily methylated, and CpG O/E is reduced overall (figure 2a). Hypomethylated regions, represented by CpG islands (with high CpG O/E), are exceptions to this general pattern (figure 2a). By contrast, the honeybee shows little methylation throughout its genome. In terms of CpG O/E, methylated regions (methylated islands) appear as dips of CpG O/E in an otherwise high CpG O/E genome [51] (figure 2b). Overall, CpG O/E measures continue to play an important role in identifying putatively methylated regions. However, it should be noted that analysis of CpG O/E alone could miss regions that have recently acquired DNA methylation, as those regions may have not experienced sufficient mutational input to alter CpG contents.

3. Histone proteins as mediators of mutational effects in the genome
Most of the eukaryotic genome is compacted into nucleosomes. Nucleosomes consist of approximately 147 bp of DNA wrapped around an octamer of two copies of each of the four core histone proteins, H2A, H2B, H3 and H4 [52] (figure 3a). DNA bound into nucleosomes is less accessible to regulatory proteins than unbound DNA. Therefore, many regulatory processes in eukaryotes have been linked to interactions between histone proteins and DNA.

Histone proteins undergo a variety of histone post-translational modifications (HPTMs) that affect gene function. These HPTMs are carried out by an array of ‘writer’ enzymes that modify histone proteins based on histone structural features and interactions with other regulatory proteins [54]. HPTMs affect gene regulation in two ways [55]. First, HPTMs may cause changes to the properties of the histone proteins themselves resulting in, for example, changes in the positions or stability of nucleosomes. Such changes in nucleosome properties may lead to differences in the compaction of the DNA and, consequently, to variation in transcription of genes in the region. Second, HPTMs may affect gene function by determining which transcription factors interact with DNA. The association of different transcription factors may increase or decrease transcription rates of genes. Thus, HPTMs hold important epigenetic information that has strong, heritable influences on the function of associated genes [55].

(a) Histone proteins affect DNA mutation
The packaging of DNA into nucleosomes has complex effects on the processes that generate and repair mutations [56,57].
DNA that is bound in nucleosomes is less prone to DNA ‘breathing’, whereby DNA opens into a single-stranded state. Single-stranded DNA mutates at a higher frequency than double-stranded DNA bound into nucleosomes. Consequently, genomic DNA that is bound by nucleosomes displays substantially lower mutation probabilities than DNA that resides in unoccupied locations [58]. The mechanistic basis for the effects of nucleosomes on mutation arises from changes to the three-dimensional structure of DNA when it is compacted [59,60]. DNA bases that are ‘outward’ facing from the nucleosome may experience higher mutational pressures. Moreover, DNA bases that reside in the minor or major grooves of the double helix also mutate at different rates [61] (figure 3b).

DNA compacted into heterochromatic and euchromatic regions of the genome possesses different epigenetic information [62] and can also be subject to different levels of DNA repair. Certain DNA repair mechanisms, such as the DNA mismatch or nucleotide excision repair pathways, operate differently in heterochromatic and euchromatic regions [63,64]. Moreover, bases near the edge of nucleosomes are more accessible to DNA repair machinery, thereby leading to nucleosome-related, nucleotide-position effects on DNA repair [59,60].

4. Transposable elements, genome structure and epigenetic information

TEs are DNA sequences that are able to move or be copied to new locations in the genome [70]. TEs comprise a very large fraction of eukaryotic genomes; they make upwards of 33% of the entire genome of many mammals and 75% of the genome of some plants [70,71]. Importantly, a high percentage of spontaneous mutations in plants and animals result directly from TE movement in the genome [70,72].

TEs can disrupt gene activity if they transpose into an existing, functional gene [73,74]. However, TEs can also introduce new exons into genes, thereby generating new substrate for evolution. TE movement may affect patterns of gene expression if a TE absorbs a host promoter or transcription factor binding sequence, and then transposes next to an existing gene [75]. TEs are also implicated in generating large deletions or duplications by facilitating ectopic recombination [70,71,76]. Moreover, insertion of TEs can change the sequence composition as well as the epigenetic landscape of the target region and thus affect the mutational spectrum. Thus, TEs cause a wide variety of mutagenic events and are among the most important factors affecting genome evolution [71,76].

(a) Epigenetic information represses mutations caused by transposable elements

Epigenetic mechanisms have important effects on TE activity and, therefore, on controlling the generation of mutations within eukaryotic genomes. Indeed, some epigenetic mechanisms are believed to have evolved specifically to control TE activity [74]. Therefore, epigenetic mechanisms are
implicated in affecting genome mutation by suppressing or releasing TE movement within the genome [77].

DNA methylation is the best-documented epigenetic system associated with TE activity [78]. Methylation of TEs in mammals differs between sexes, age and tissues, and these differences correlate with the activity of TEs [78]. Animals and plants with experimentally deactivated DNA methylation systems experience an upregulation of TEs, thereby indicating a direct role of DNA methylation on suppressing TE activity [62,74]. Finally, a specific DNA methyltransferase gene, Dmnt3C, was recently discovered in some mammals. Dmnt3C apparently specializes in methylating TEs in the germline, therefore providing strong evidence that DNA methylation plays an important role in suppressing TE movement [79].

Other epigenetic systems also act to suppress TE activity and, consequently, genome mutations. For example, TE RNA can be processed through the small-interfering RNA (siRNA) pathway. siRNAs produced from TE RNA can be targeted by the RNA interference system (RNAi) [80]. The RNAi system can then degrade the target TE RNAs, thereby controlling the spread of TEs in the genome [81]. In addition, the siRNA pathway can lead to subsequent modifications to chromatin structure that ultimately act to silence TE genes [81].

PIWI-interacting RNAs (piRNAs) represent another epigenetic information system that affects the behaviour of TEs [74,82]. piRNAs are short, single-stranded, non-coding RNAs that are specifically defined as interacting with Piwi proteins. piRNAs are believed to have evolved in response to the effects of TEs in animals [83]. piRNAs silence TEs and, therefore, prevent harmful TE-induced mutations [82]. Specifically, the Piwi complex, which interacts with focal piRNAs (figure 4), can promote the generation of heterochromatic marks around TE DNA. The Piwi protein may lead to the silencing of TEs by allowing for the addition of repressive histone modification H3K9me3 or by preventing the deposition of the activating histone modification H3K4me2. piRNAs also facilitate the degradation of TE RNAs and prevent TE multiplication in animals. piRNAs complementary to sequences of TE mRNAs act as guides to protein complexes that either act to destroy the TE mRNA or assist in mRNA degradation [72]. Moreover, new types of piRNAs may be derived directly from invading TEs [85]. In this way, the piRNA system acts as a type of inherited immunity against TEs [82,83].

Epigenetic systems interact in a variety of ways to suppress TEs. The epigenetic silencing of TEs by DNA methylation and non-coding RNAs is often linked to other epigenetic changes [62]. In plants and animals, suppression of TEs is associated with HPMTs that suppress TE activity [62,86–88]. Interestingly, the types of histone modifications associated with TEs are quite complex and depend on the TE sequence itself [86]. Thus, epigenetic systems may be highly tuned to dampening the effects of different TEs and preserving the integrity of the genome.

5. Evolution and epigenetics of duplicated genes

Epigenetic mechanism can affect genome evolution by influencing the fate of duplicate genes. Gene duplication is a fundamental source of genomic variance. Gene duplication can lead to the production of genes with new functions via neofunctionalization [89] or contribute to biological complexity by promoting separation of ancestral functions in a tissue, developmental stage or cell type via subfunctionalization [90]. A duplicated gene is likely to be functionally redundant to the original gene. Thus, loss-of-function mutations are expected to quickly accumulate, rendering the duplicate gene non-functional [91]. Therefore, the most critical step in the preservation of duplicated genes is the initial retention of both gene copies following duplication (e.g. [92]).

Epigenetic mechanisms can offer molecular pathways for duplicate gene retention and functional divergence. Epigenetic silencing of duplicate genes among tissues or developmental stages effectively ‘shields’ duplicate genes from natural selection and increases the probability of subfunctionalization and neofunctionalization [93]. For example, DNA methylation can reduce gene expression, thereby preventing a gene from producing a negative phenotype [94,95]. Indeed, human and yeast duplicate genes showed reduction of expression following duplication, which could facilitate the retention of gene copies [96].

Empirical data from diverse taxa demonstrate that epigenetic patterns of duplicate genes and singletons are distinct. Promoters of duplicate genes in mammals have higher DNA methylation compared to those of singletons [94,95]. Typically, the gene copy with a higher level of promoter DNA methylation tends to exhibit reduced expression compared to its sister copy, indicating that DNA methylation differences may contribute to expression differences [95]. Data from other species such as zebrafish [97] are consistent with this model.

Data from invertebrates and plants also support the role of DNA methylation in regulating duplicate genes [98,99]. Gene body methylation, which is the most common form of DNA methylation in invertebrates, is known to be positively associated with gene expression [4]. Interestingly, duplicated genes have lower levels of gene body methylation than singletons in honeybees [99], which is consistent with the idea that DNA methylation may affect duplicate gene function. Data from plants also support the idea that divergence of gene body methylation correlates with expression divergence [100]. Therefore, a model of DNA methylation and expression reduction may be applicable to divergent animal and plant taxa.
6. Conclusion

Epigenetic information is critical to organismal development, behaviour and function. In addition, many types of epigenetic information can affect genome evolution and structure. Epigenetic marks shape genome evolution through their direct and indirect interactions with chromatin. In particular, DNA methylation leads to a high frequency of single-base substitutions of cytosines in the genome. The packaging of eukaryotic DNA into nucleosomes also has strong effects on genome evolution, because the structure of the nucleosome affects mutation probabilities and DNA repair mechanisms. Epigenetic information in the form of DNA methylation, non-coding RNAs or HPTMs also strongly influences genome mutation by suppressing TE movement. Finally, the evolution of gene content within eukaryotic genomes is affected by epigenetic mechanisms that facilitate the process of gene duplication. Thus, overall, we suggest a greater appreciation of epigenetic information as a mediator of genome evolution. Epigenetic information is ultimately a cause of both organismal plasticity and molecular evolution in eukaryotes.

Data accessibility. This article has no additional data.

Authors’ contributions. S.V.Y. and M.A.D.C. developed and co-wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This research was partially supported by National Science Foundation grant nos IIS-2019799 and EF-2021635.

Acknowledgements. We thank Xin Wu for help with figure 2, and CJ Dyson and anonymous reviewers for comments on earlier versions of this manuscript.

References


5. Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Yi SV. 2012 Birds do it, bees do it, worms and their direct and indirect interactions with chromatin. In particular, DNA methylation leads to a high frequency of single-base substitutions of cytosines in the genome. The packaging of eukaryotic DNA into nucleosomes also has strong effects on genome evolution, because the structure of the nucleosome affects mutation probabilities and DNA repair mechanisms. Epigenetic information in the form of DNA methylation, non-coding RNAs or HPTMs also strongly influences genome mutation by suppressing TE movement. Finally, the evolution of gene content within eukaryotic genomes is affected by epigenetic mechanisms that facilitate the process of gene duplication. Thus, overall, we suggest a greater appreciation of epigenetic information as a mediator of genome evolution. Epigenetic information is ultimately a cause of both organismal plasticity and molecular evolution in eukaryotes.

Data accessibility. This article has no additional data.

Authors’ contributions. S.V.Y. and M.A.D.C. developed and co-wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This research was partially supported by National Science Foundation grant nos IIS-2019799 and EF-2021635.

Acknowledgements. We thank Xin Wu for help with figure 2, and CJ Dyson and anonymous reviewers for comments on earlier versions of this manuscript.

References


5. Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Yi SV. 2012 Birds do it, bees do it, worms and their direct and indirect interactions with chromatin. In particular, DNA methylation leads to a high frequency of single-base substitutions of cytosines in the genome. The packaging of eukaryotic DNA into nucleosomes also has strong effects on genome evolution, because the structure of the nucleosome affects mutation probabilities and DNA repair mechanisms. Epigenetic information in the form of DNA methylation, non-coding RNAs or HPTMs also strongly influences genome mutation by suppressing TE movement. Finally, the evolution of gene content within eukaryotic genomes is affected by epigenetic mechanisms that facilitate the process of gene duplication. Thus, overall, we suggest a greater appreciation of epigenetic information as a mediator of genome evolution. Epigenetic information is ultimately a cause of both organismal plasticity and molecular evolution in eukaryotes.

Data accessibility. This article has no additional data.

Authors’ contributions. S.V.Y. and M.A.D.C. developed and co-wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This research was partially supported by National Science Foundation grant nos IIS-2019799 and EF-2021635.

Acknowledgements. We thank Xin Wu for help with figure 2, and CJ Dyson and anonymous reviewers for comments on earlier versions of this manuscript.

References


5. Bewick AJ, Hofmeister BT, Powers RA, Mondo SJ, Yi SV. 2012 Birds do it, bees do it, worms and their direct and indirect interactions with chromatin. In particular, DNA methylation leads to a high frequency of single-base substitutions of cytosines in the genome. The packaging of eukaryotic DNA into nucleosomes also has strong effects on genome evolution, because the structure of the nucleosome affects mutation probabilities and DNA repair mechanisms. Epigenetic information in the form of DNA methylation, non-coding RNAs or HPTMs also strongly influences genome mutation by suppressing TE movement. Finally, the evolution of gene content within eukaryotic genomes is affected by epigenetic mechanisms that facilitate the process of gene duplication. Thus, overall, we suggest a greater appreciation of epigenetic information as a mediator of genome evolution. Epigenetic information is ultimately a cause of both organismal plasticity and molecular evolution in eukaryotes.

Data accessibility. This article has no additional data.

Authors’ contributions. S.V.Y. and M.A.D.C. developed and co-wrote the manuscript.

Competing interests. We declare we have no competing interests.

Funding. This research was partially supported by National Science Foundation grant nos IIS-2019799 and EF-2021635.

Acknowledgements. We thank Xin Wu for help with figure 2, and CJ Dyson and anonymous reviewers for comments on earlier versions of this manuscript.

References


