PNAS RESEARCH ARTICLE | EVOLUTION

Repeat sequences limit the effectiveness of lateral gene transfer and favored the evolution of meiotic sex in early eukaryotes

Marco Colnaghi[a,b], Nick Lane[a,b], and Andrew Pomiankowski[a,b,1]

Edited by Harmit Malik, Fred Hutchinson Cancer Research Center, Seattle, WA; received March 22, 2022; accepted July 28, 2022

The transition from prokaryotic lateral gene transfer to eukaryotic meiotic sex is poorly understood. Phylogenetic evidence suggests that it was tightly linked to eukaryogenesis, which involved an unprecedented rise in both genome size and the density of genetic repeats. Expansion of genome size raised the severity of Muller's ratchet, while limiting the effectiveness of lateral gene transfer (LGT) at purging deleterious mutations. In principle, an increase in recombination length combined with higher rates of LGT could solve this problem. Here, we show using a computational model that this solution fails in the presence of genetic repeats prevalent in early eukaryotes. The model demonstrates that dispersed repeat sequences allow ectopic recombination, which leads to the loss of genetic information and curtails the capacity of LGT to prevent mutation accumulation. Increasing recombination length in the presence of repeat sequences exacerbates the problem. Mutational decay can only be resisted with homology along extended sequences of DNA. We conclude that the transition to homologous pairing along linear chromosomes was a key innovation in meiotic sex, which was instrumental in the expansion of eukaryotic genomes and morphological complexity.

eukaryogenesis | lateral gene transfer | Muller's ratchet | sexual reproduction | mutation accumulation

Significance

The origin of meiotic sex is a long-standing evolutionary enigma. This novel mechanism of reproduction replaced lateral gene transfer (LGT), the uptake and recombination of pieces of environmental DNA seen in bacteria and archaea. We link its origin to the expanded genome size and proliferation of genetic repeats found in early eukaryotes. Both factors led to high levels of mutation accumulation and gene loss under LGT, which could not be retarded through increases in the rate of LGT or the length of DNA recombined. Meiotic sex with homologous pairing of long-chromosome-sized pieces of DNA promoted purifying selection and suppressed ectopic recombination. It permitted the evolution of the expanded genome needed for the evolution of complex eukaryotic life.

The genes for meiosis are universal among eukaryotes, indicating that sex evolved before the divergence of the first eukaryotic clades (1, 2). It evolved from the molecular machinery for lateral gene transfer (LGT), which facilitates genetic exchange in archaea and bacteria (1, 3, 4). Prokaryotes possess homologs of the canonical molecular machinery for meiotic sex, including proteins of the *SMC* gene family of adenosine triphosphatases necessary for chromosome cohesion and condensation (5), as well as actin and tubulin, required for daughter cell separation and the movement of chromosomes (6). The *Rad51/Dcm1* gene family, which plays a central role in meiosis, also has high protein sequence similarity with *RecA*, responsible for homologous search and recombination in prokaryotes (7, 8). But why eukaryotes requisitioned this existing molecular machinery to evolve a completely new mechanism of reproduction, inheritance, and genetic exchange—meiotic sex—remains obscure.

Transformation is one of the major routes of genetic exchange via LGT in bacteria and involves the acquisition of environmental DNA (eDNA), followed by recombination into the host genome (8, 9). By allowing genetic exchange between lineages, transformation can restore genes that have been disrupted through mutation or deletion (10–12), counter the effects of genetic drift and reverse Muller's ratchet (11, 13), and accelerate adaptation by reducing selective interference (14, 15). Previous modeling work has shown that the expansion of early eukaryote genome size was likely to have caused the failure of LGT (13). While LGT via transformation helps to purge deleterious mutations (11), this benefit rapidly wanes as genome size increases because of the difficulty of matching individual mutations with eDNA (13). LGT can resist mutation accumulation in larger genomes by combining more frequent recombination with increased recombination length, the mean length of DNA picked up from the environment and recombined into the host cell genome (13). But the distribution of recombination length in bacteria is skewed toward short eDNA sequences, with a median length that encompasses at most just a few genes (16–18). In addition, bacteria typically cleave eDNA, shortening recombination length. While there are constraints on the rate of uptake and recombination through limited eDNA availability and sequence homology (12, 18), prokaryotes plainly did not follow the eukaryotic trajectory toward recombination across whole chromosomes.

After the endosymbiotic event that gave rise to the first eukaryotes, the archaeal host's genome greatly expanded with genes of bacterial endosymbiotic origin and through gene duplication and divergence, which enabled a range of novel functions (19, 20). This is estimated to have doubled gene number in the last eukaryotic common ancestor (LECA)

Author affiliations: [a]Centre for Computation, Mathematics and Physics in the Life Sciences and Experimental Biology, University College London, London WC1E 6BT, United Kingdom; and [b]Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT, United Kingdom

Author contributions: M.C., N.L., and A.P. designed research; M.C. performed research; M.C., N.L., and A.P. analyzed data; and M.C., N.L., and A.P. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2022 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]To whom correspondence may be addressed. Email: ucbhpom@ucl.ac.uk.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2205041119/-/DCSupplemental.

Published August 22, 2022.

Downloaded from https://www.pnas.org by 147.94.12.7 on August 23, 2022 from IP address 147.94.12.7.

(21, 22). The extra energetic availability provided by the proto-mitochondrial endosymbiont released bioenergetic constraints over prokaryotic cell and genome size (23, 24). But this came with the cost of maintaining a larger genome (19, 21, 24, 25). Early eukaryote genome size expansion also reflected an increase in the density of repeat sequences, arising from gene duplication and the spread of mobile genetic elements (25, 26). Mobile retroelements of endosymbiotic origin are thought to have spread widely through the protoeukaryote host genome, leading to a proliferation of self-splicing introns (27–29). These selfish elements are present in many bacterial species, almost always at low copy numbers (<10 per genome) (30), but likely increased in a more uninhibited manner, perhaps exploiting the nonhomologous end-joining mechanism of DNA repair found throughout eukaryotes (31). Novel intron density is thought to have reached a density comparable to that seen in modern eukaryote species (29).

The need to restrict ectopic recombination caused by increased repeat density might have played a pivotal role in determining the evolution of meiosis (32). However, the possible involvement of such repeat sequences has not been investigated in previous quantitative models of LGT or the transition to meiotic sex (10, 11, 13). In prokaryotes, high repeat density is associated with a high probability of ectopic recombination, increasing the rates of deletions, insertions, and other genomic rearrangements (33, 34). Recombination errors caused by the presence of repeat sequences introduce an additional cost to LGT and potentially constrain the benefits of increased recombination length and LGT frequency. Here, we investigate whether the sharp increase in repeat density in early eukaryotes could have forced them to abandon LGT in favor of meiosis. To investigate this hypothesis, we develop a computational model of mutation and selection in a population undergoing LGT via transformation in the presence of genetic repeats. The model highlights a tradeoff between the benefits of LGT (greater genetic variance, enhancing purifying selection) and its cost (loss of genetic information through ectopic recombination). This leads to the view that the transition to meiotic sex was driven by the need for purifying selection in the expanding and repeat-rich genomes of early eukaryotes, which could not be met by increases in recombination length or LGT rate.

Materials and Methods

We use a Fisher-Wright process with nonoverlapping generations to model the evolution of a population of N haploid individuals with a circular genome composed of g unique protein-coding genes (*SI Appendix*, Fig. S1 and Table 1 and *SI Appendix, Methods*). The genome is interspersed at random intervals with a generic repeat sequence, at an initial density ρ (i.e., with ρg repeats). Each generation consists of a replication stage using a multiplicative fitness function without epistasis, based on the number of functional wild-type alleles. Each individual in the new population then has a probability λ of acquiring a fragment of eDNA of length L. Recombination requires matching of the terminal loci of the eDNA sequence and the host genome, either to a protein-coding gene or a repeat sequence. When there is multiple matching, one of the homologous sequences is randomly selected, with weights inversely proportional to the difference in length between the eDNA and possible matching host genomic sequences. After LGT, individuals accumulate a random number of deleterious mutations, drawn from a Poisson distribution with mean $U = \mu g'$ (where g' is the number of wild-type protein-coding genes), with μ being the per locus mutation rate. For specific details about the model implementation, we refer the reader to *SI Appendix, Methods*. A list of model parameters and variables is given in Table 1.

The evolutionary process is studied with a population initially free of mutants, over $t_{max} = 5,000$ generations, with 100 independent iterations for a given set of parameter values. For each replicate, we evaluate the gene loss rate per generation from deletions ($\Delta M_d/\Delta t$) and mutations ($\Delta M_m/\Delta t$) as the average load of deletions and mutations, respectively, divided by the number of generations. The total gene loss rate per generation $\Delta M/\Delta t$ is calculated as the sum of these two components. The change in some key model variables (genome size, repeat content, total gene loss, mutation load, and duplication content) during the course of a standard simulation and the final distribution of the same variables at the end of a simulation are shown in *SI Appendix*, Figs. S3 and S4, respectively.

Table 1. Model parameters and variables

μ	Deleterious point mutation rate per locus
U	Genome-wide deleterious point mutation rate
λ	LGT rate
L	Recombination length
N	Population size
t_{max}	Length of simulations (generations)
ρ	Initial repeat density
g	number of protein-coding genes
$\Delta M_d/\Delta t$	Gene loss rate per generation due to deletions
$\Delta M_m/\Delta t$	Gene loss rate per generation due to point mutation
$\Delta M/\Delta t$	Total gene loss rate (mutations plus deletions)

Summary of the key parameters and variables used in the numerical simulations.

Results

Muller's Ratchet. The mean mutation load of the whole population and the number of mutations in the least-loaded class (LLC) reflect the interplay between genetic drift and Muller's ratchet (Fig. 1). In a clonal population with no LGT ($\lambda = 0$), random fluctuations can cause the LLC to go extinct (a "click" of the ratchet), determining an ever-increasing mutation load baseline (Fig. 1*A*). In the absence of LGT, the fittest class cannot be restored and the increase in mutation load is irreversible. The introduction of LGT ($\lambda = 0.1$) favors the elimination of mutations by increasing genetic variation and strengthening purifying selection, reducing the frequency at which the LLC is lost (Fig. 1*B*). In addition, LGT permits the reversal of Muller's ratchet and the reduction in mutation number in the LLC. But in the presence of a high repeat density ($\rho = 0.1$), the benefit of LGT comes at the price of deletions due to ectopic recombination, making LGT less obviously beneficial (Fig. 1*C*).

LGT and Repeats. Repeat density strongly influences the benefit of LGT. If repeat density is low ($\rho = 10^{-2}$), increasing LGT (λ) is advantageous and reduces the total gene loss rate (Fig. 2*A*). But as repeat density rises ($\rho \approx 5 \times 10^{-2}$), this benefit is eroded, and higher levels of LGT provide little or no benefit (Fig. 2*A*). At high levels of repeats ($\rho \geq 10^{-1}$), LGT is always detrimental and elevates total gene loss (Fig. 2*A*). Splitting gene loss into its components, it becomes evident that the likelihood of ectopic recombination increases with higher density of repeats, leading to a sharp increase in gene loss through deletions (Fig. 2*C*). In addition, high repeat density limits the effectiveness of LGT in purging deleterious mutations due to selective interference (Fig. 2*D*). The net effect is that LGT ceases to have a beneficial effect beyond a threshold repeat density (Fig. 2*B*).

The lower efficiency of LGT at removing deleterious mutations arises because repeats make homologous recombination less likely and ectopic recombination more likely. This can be seen by adding a requirement for full homology throughout the whole eDNA (i.e., not only at the terminal loci; *SI Appendix, Methods*). This eliminates ectopic recombination, and mutation accumulation then closely follows the case without repeats, showing an accelerating decline with LGT rate (Fig. 3*A*). In contrast, with recombination based on end homology alone, there is only a monotonic decline in mutation accumulation as the rate of LGT increases (Fig. 3*A*). This is because the presence of repeats,

Downloaded from https://www.pnas.org by 147.94.12.7 on August 23, 2022 from IP address 147.94.12.7.

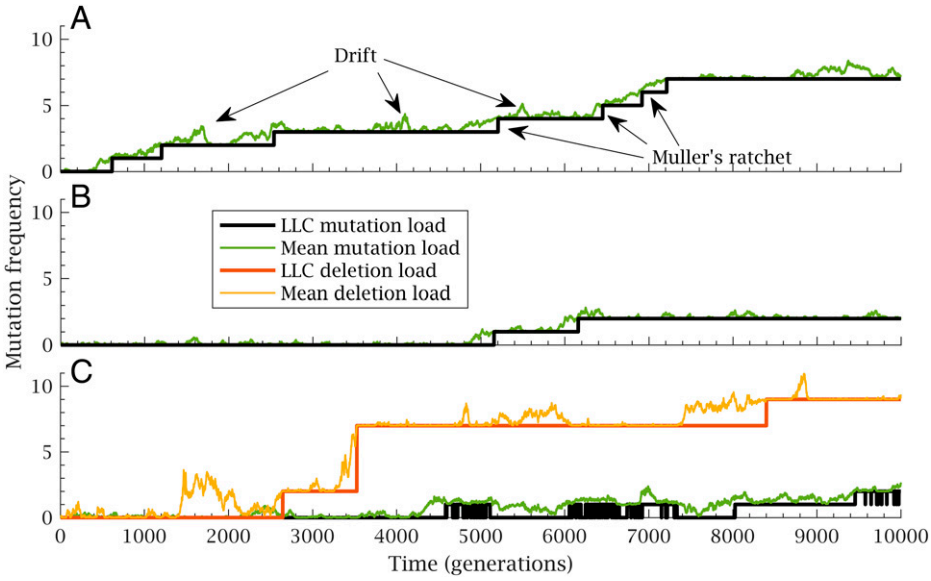

Fig. 1. Impact of LGT on Muller's ratchet. (*A*) Mean mutation load (green) and least-loaded class (LLC) mutation load (black) of a repeat-free ($\rho = 0$), nonrecombining population that does not undergo LGT ($\lambda = 0$), across $t_{max} = 10,000$ generations. Random fluctuations in allele frequencies due to genetic drift lead to irreversible increases of the LLC mutation load (Muller's ratchet). (*B*) In a repeat-free ($\rho = 0$) population undergoing LGT ($\lambda = 0.1$, $L = 5$), recombination via LGT increases purifying selection, countering the ratchet. (*C*) In a population undergoing LGT ($\lambda = 0.1$, $L = 5$) in the presence of repeats ($\rho = 0.1$), LGT allows the ratchet to be reversed, reducing the LLC mutation load, but the presence of repeats leads to a high rate of deletion, resulting in a high mean deletion load (yellow) and LLC deletion load (red). Other simulation parameters: $g = 100$, $N = 2,500$, and $\mu = 3 \times 10^{-5}$.

together with the buildup of deletions, raises the probability that one or both ends of the eDNA lack homology to any genomic sequence (i.e., if the matching sequence has been deleted from the genome or if one end binds to a repeat sequence, but the other end either lacks a homologous sequence or is too far away from it to recombine). Either scenario increases the probability that no recombination takes place (Fig. 3*B*). These effects combine to reduce the rate of homologous recombination as repeat density increases (Fig. 3*B*), constraining LGT's ability to purge deleterious mutations.

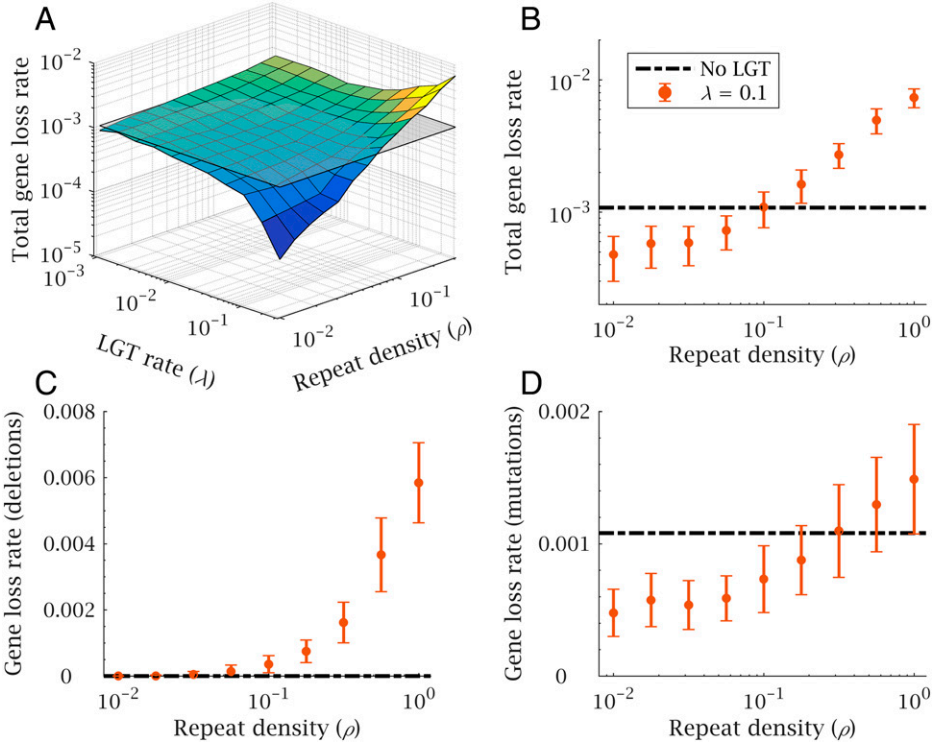

Fig. 2. Repeat sequences cause the failure of LGT. (*A*) The total gene loss rate $\Delta M/\Delta t$ (through both mutations and deletions) is shown for a range of initial repeat densities (ρ) and LGT rates (λ). For comparison, the gray plane shows the total gene loss rate in a repeat-free population not undergoing LGT (null model). Each data point is the average of 100 independent simulations. (*B*) The total gene loss rate ($\Delta M/\Delta t$) is due to (*C*) ectopic recombination leading to deletions ($\Delta M_d/\Delta t$) and (*D*) recurrent deleterious mutations ($\Delta M_m/\Delta t$), both of which increase with initial repeat density (ρ). Simulations in (*B*–*D*) were carried out with a high rate of LGT ($\lambda = 0.1$). The dotted line represents the null model of mutation accumulation in a nonrecombining, repeat-free population. Error bars indicate the SD over 100 independent simulations. Other simulation parameters: $g = 100$, $N = 2,500$, $\mu = 3 \times 10^{-5}$, $t_{max} = 5,000$, and $L = 10$.

Downloaded from https://www.pnas.org by 147.94.12.7 on August 23, 2022 from IP address 147.94.12.7.

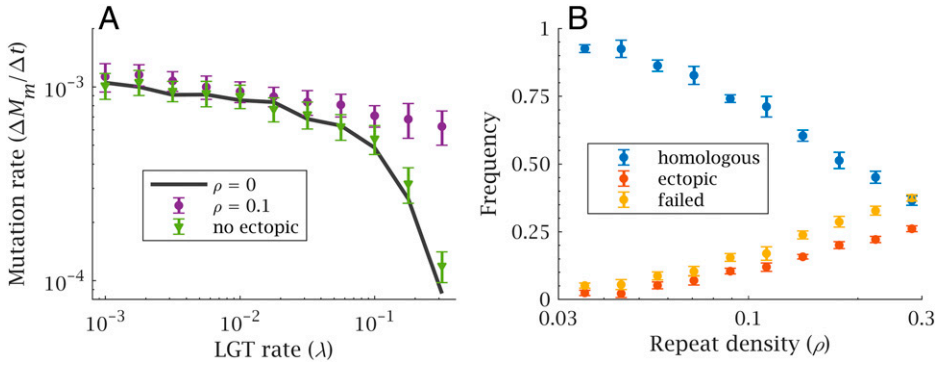

Fig. 3. Ectopic recombination. (*A*) The rate of mutation accumulation is shown for the null model without repeats ($\rho = 0$, black line), a genome with high initial repeat density ($\rho = 0.1$, purple circles), and a genome with the same repeat content but where ectopic recombination is suppressed due to a requirement for homology throughout the eDNA ($\rho = 0.1$, green triangles, no ectopic). (*B*) The frequency of homologous (blue circles), ectopic (red circles), and failed recombination (yellow circles) as a function of initial repeat density. Rate and frequency were calculated over $t_{max} = 5,000$ generations. Error bars indicate the SD over 100 independent simulations. Other parameters: $g = 100$, $N = 2,500$, $L = 10$, $\mu = 3 \times 10^{-5}$ and, in (*B*), $\lambda = 0.1$.

Changes in Genome Size. The limitations of LGT are greater in large genomes (13). This is partly because the probability that eDNA matches a particular mutated sequence (reducing the mutation load) decreases with genome size. This effect can be overcome through increases in L, the size of eDNA selected for recombination (13). But that previous analysis neglected the effect of repeat sequences. If repeat density is low ($\rho = 0.01$), increases in L are favorable and help populations with large genome size resist the ratchet (compare $L = 2$ and $L = 10$; Fig. 4*A*). However, with higher repeat density ($\rho = 0.3$), larger L is deleterious. It increases the total rate of gene loss and does nothing to stop the ratchet in large genomes (Fig. 4*B*). This transition arises for two reasons. In genomes with few repeats, deletions through ectopic recombination are negligible and increasing L has only a very minor effect on their occurrence as genome size increases (*SI Appendix*, Fig. S2*A*). As almost all recombination events are homologous, increasing L is beneficial and facilitates the removal of deleterious mutations (*SI Appendix*, Fig. S2*B*). In contrast, when genomes are repeat rich ($\rho = 0.3$), the benefits of LGT are offset by an elevated rate of deletion (*SI Appendix*, Fig. S2*C*). This compromises the efficiency of eDNA repair of mutations, which becomes almost independent of L (*SI Appendix*, Fig. S2*D*).

Unlike repeat-free populations (13), increasing eDNA length proportionally to genome size provides little or no benefit in the presence of repeats. A large genome ($g = 1,000$) cannot be sustained by LGT above a critical repeat density, as this engenders a rate of total gene loss comparable to that of a nonrecombining population (Fig. 5*A*, red triangles). The way out of this dilemma is to require sequence homology throughout the eDNA. This ensures that recombination is homologous and allows a reduction of the mutation load with large L without incurring the associated increase in gene deletions. As a consequence, recombination is able to lower the total gene loss rate in large ($g = 1,000$) genomes even in the presence of a high density of repeated sequences (Fig. 5*A*, yellow circles). Recombination lengths proportional to genome size ($L = 0.1g$), homologous recombination across the entire length of the eDNA, and a high LGT rate ($\lambda = 1$) are all needed to prevent a sharp increase in total gene loss associated with a large genome size (Fig. 5*B*, yellow circles).

Discussion

Asexual organisms are often portrayed as destined to accumulate mutations via Muller's ratchet, on an inevitable decline to extinction through mutational meltdown (35–37). This view emanates from a eukaryotic perspective seeking to explain the maintenance of sexual reproduction in the face of the twofold cost of sex and other costs relating to meiosis, finding a mate, and cell fusion (38–40). Prokaryotes, both bacteria and archaea, lack meiotic sex and typically reproduce through asexual division, but they nonetheless have a number of mechanisms for achieving genetic recombination (12, 41–43). In particular, LGT through transformation allows competent cells to pick up eDNA released from related lineages and to recombine it into their genome. In agreement with previous studies (10, 11, 13), our modeling shows that LGT generates genetic variation, strengthens purifying

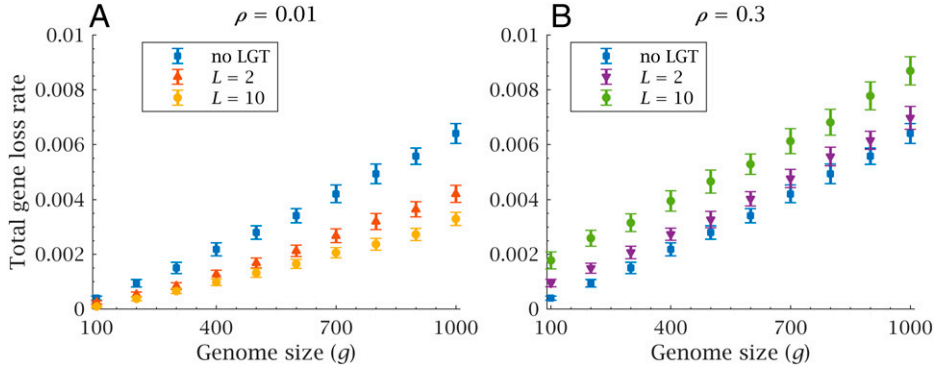

Fig. 4. Genome size and recombination length. The rate of gene loss increases with genome size (g). (*A*) At low repeat density ($\rho = 0.01$), higher values of recombination length (L) minimize the rise in gene loss rate as genome size increases. (*B*) This benefit is reversed at high initial repeat density ($\rho = 0.3$), where higher L increases the total gene loss rate. Error bars show the SD over 100 independent simulations. Note the null models (blue points) are identical in (*A* and *B*). Gene loss rate was calculated over $t_{max} = 5,000$ generations. Other parameters: $N = 2,500$, $\mu = 10^{-5}$, and $\lambda = 0.1$.

Downloaded from https://www.pnas.org by 147.94.12.7 on August 23, 2022 from IP address 147.94.12.7.

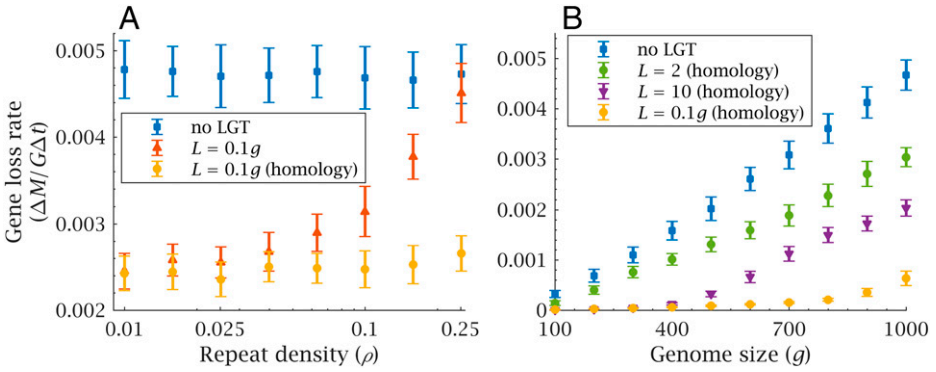

Fig. 5. Advantage of homologous recombination. (*A*) Total gene loss rate as a function of initial repeat density (ρ). The blue squares correspond to the null model of a nonrecombining population, where the total gene loss rate is independent from repeat density. This is compared with LGT of long eDNA sequences ($L = 0.1g$, red triangles) and recombination with full-sequence homology throughout the eDNA (yellow circles) with the same recombination length. (*B*) The impact of recombination length (L) on total gene loss as a function of genome size (g) in a repeat-rich genome (initial repeat density $\rho = 0.3$) under the requirement for full-sequence homology throughout the eDNA. To limit gene loss in large genomes, recombination length must increase proportionally to genome size ($L = 0.1g$, yellow circles). Error bars show the SD over 100 independent simulations. Gene loss rate was calculated over $t_{max} = 5,000$ generations. Other parameters: $N = 2,500$ and $\mu = 10^{-5}$; in (*A*) $\lambda = 0.1$, $g = 1,000$ and (*B*) $\lambda = 1$.

selection, and reduces the rate of mutation accumulation, performing a similar function to meiotic sex in eukaryotes (39, 44, 45). This leads to a simple question: why did the first eukaryotes abandon LGT and replace it with meiotic sex?

Our analysis shows that the benefits of LGT are curtailed by the presence of genetic repeats (Fig. 2*A*). Repeat sequences enable ectopic recombination, causing gene loss through deletions proportional to their density in the host genome (Fig. 2*C*). Repeats also make LGT less effective at purging deleterious mutations (Fig. 2*D*) by reducing the rate of correct homologous recombination (Fig. 3). This is mainly due to selective interference from the presence of deletions and mutations segregating at the same time (46–48). A population in which ectopic recombination has removed a portion of the genome will have lower N_e and weaker selection for any deleterious mutations in that region. In addition, as the frequency of the wild-type allele is reduced, it will be less likely to be in the eDNA, reducing the potential benefits of LGT. Moreover, repeats also make homologous recombination less likely, reducing the potency of LGT. Ectopic recombination events might restore deleted genes, but this effect is second order and outweighed by the detrimental effects of repeat sequences. Repeats cause a higher deletion rate, weaker selection, and lower homologous recombination rate, which amplify the loss of genetic information as their density increases (Figs. 2 and 3).

The acquisition of new genes through endosymbiotic gene transfer (including mobile self-splicing introns), plus duplication and divergence, led to massive genome expansion in the evolution of early eukaryotes (21, 25, 26). This made the first eukaryotes more vulnerable to the accumulation of mutations caused by Muller's ratchet. As genome size rises (g), the homologous recombination rate per locus declines (assuming the probability of LGT remains constant) simply because the probability that an eDNA piece matches to a particular locus is inversely proportional to genome size (13). A solution to this is to increase the recombination length (L) (Fig. 4*A*). Other things being equal, picking up larger pieces of eDNA (larger L) increases the recombination rate per locus and thereby facilitates the elimination of deleterious mutations (13). But other things may not be equal, as environmental fragmentation and deterioration (as well as physical constraints on import) are likely to limit the length of eDNA that can be retrieved from the environment (49, 50). Of greater importance in our analysis, this potential solution is compromised by the presence of genetic repeats. If repeats are common, larger recombination length (L) is associated with higher, rather than lower, loss of genetic information (Fig. 4*B*). A higher recombination length elevates the rate of ectopic recombination between dispersed repeated sequences, resulting in a greater rate of gene deletions. Similar disadvantages have been reported in extant prokaryotes, where higher repeat density is associated with greater genomic instability, increasing deletions, inversions, and other genomic rearrangements (33, 34). All of these issues are amplified as genome size increases, and beyond a threshold in repeat density, LGT brings no benefit to a large genome, even if the recombination length scales with genome size (Fig. 5*A*).

These considerations suggest that, in order to support an expanded genome rich in repeat sequences, the first eukaryotes had to abandon LGT for syngamy and homologous recombination. The fusion of cells (syngamy) and pooling of their genomes is a simple way to achieve larger L without potential fragmentation and deterioration of DNA through environmental release. In agreement with this, some of the largest LGT events in prokaryotes are observed in archaea that exchange genetic material after cell fusion. For example, archaea from the *Haloferax* genus are known to undergo reciprocal exchange of genetic material with L = 200,000–500,000 bp after cell fusion, against $L = 2,000$–$6,000$ bp transfer events typically observed in competent bacteria (18, 51, 52). In addition, there was a requirement for these large pieces of DNA to undergo homologous pairing in order to retain the benefits of recombination without losing genetic information through ectopic end pairing in the presence of repeats. We simulated this in our LGT model by adding a requirement for homology throughout the eDNA as well as homology at the ends. Homology matching eliminates gene deletion through ectopic recombination and allows mutation accumulation to be resisted even as repeat density increases (Fig. 5*A*). Considerable expansion of genome size is now permissible without catastrophic loss of genetic information through Muller's ratchet, provided that recombination length (L) scales with genome size (Fig. 5*B*). This is equivalent to homologous recombination across aligned chromosomes as seen in meiosis, though this neglects reciprocal exchange in meiosis rather than replacement in LGT.

For simplicity, our model considers only one type of repeat sequence. In reality, there will have been numerous repeats at different densities. Frequent gene duplications in early eukaryotes contributed to the increase in repeat density, and it is

Downloaded from https://www.pnas.org by 147.94.12.7 on August 23, 2022 from IP address 147.94.12.7.

estimated that the average copy number per gene in LECA was around 1.8 (25). Another source of repeat sequence in prokaryotic genomes is transposable elements (TEs) and other selfish genetic elements. TEs can promote their own spread and reduce host fitness in other ways, through gene function disruption or gene inactivation (53, 54), and are thought to play a major role in the streamlining of prokaryotic genomes (55). The density of mobile intron-derived sequences in ancestral eukaryotic genomes is estimated to be high (accounting for up to 80% of genome size), making the choice of $\rho = 0.1$ in most of our simulations a conservative one (56). As the focus of this study is on recombination and genetic information loss, we did not explicitly model the population dynamics of repeats but evaluated the effect of variation in their initial density. Future studies should consider a diversity of repeats and include fitness penalties at the individual level associated with their movement and density, as well as the expansion of selfish genetic elements through replication within genomes. In itself, this raises interesting questions about the distribution of repeats in extant prokaryotes—for instance, why the distribution of recombination length in prokaryotes is skewed toward shorter sequences (18), why gram-positive bacteria cleave eDNA sequences before recombination (57), and why the number of TEs and other mobile elements in prokaryote genomes is so tightly constrained (30, 58).

A further possibility not covered by our modeling is the presence of beneficial ectopic recombination events, in particular, those that contribute to the acquisition of novel genes. LGT via plasmids is the main source of acquisition of accessory genes from distant lineages (59–61) and has been shown to provide adaptive benefits (62). In contrast, transformation is mainly limited to sequences from closely related lineages, requiring a high degree of sequence homology, and so less likely to import foreign genes across large taxonomic distances (63–66). As such, the main advantage of transformation is believed to be maintaining local adaptation rather than import of novel functions (9). In our model, gene loading can occur via the acquisition of genes lost through deletion. However, in the presence of a high repeat density, this effect is negligible compared with the loss of genes via deletions. It seems unlikely that repeats are retained in order to enhance gene turnover from the pangenome, but a proper analysis of this question would require a different modeling approach (67, 68).

Wilkins and Holliday (32) suggested that meiosis could arise from mitosis in a single evolutionary step, the evolution of homologous pairing during prophase. They also suggested that the main function of meiosis is to limit the rate of ectopic recombination caused by the spread of repeated sequences in ancestral eukaryotic genomes (32). Our analysis here supports their idea that homologous pairing could have arisen because of the need to evade the deleterious effect of pervasive genomic repeats. But Wilkins and Holliday did not explicitly address the impact of genome size expansion on mutation accumulation. We demonstrate that this is a crucial factor in escalating the severity of Muller's ratchet and determining the failure of LGT. Our results complement their hypothesis by showing that, in addition to restricting ectopic recombination, meiotic sex is also necessary to increase the total rate of recombination per base pair in expanding early eukaryotic genomes.

While our analysis demonstrates the selective advantage of meiotic sex, we do not explicitly address the multifaceted question of its evolution. The picture emerging from phylogenetic studies is that ancestral eukaryotes underwent a massive genome size expansion, but the precise timing of the events leading to eukaryogenesis is still a matter of debate (25, 26, 69). Did meiotic sex evolve after the acquisition of mitochondrial symbionts allowed eukaryotic genome size expansion (23, 24) or were the first eukaryotes able to expand their genomes because they could already undergo syngamy (cell fusion) and reciprocal exchange of genetic material, as observed in some extant archaea (51, 52, 70)? These archaea do not have unusually large genomes (~4,000 genes), which suggests that mitochondrial acquisition was needed to drive the expansion in eukaryotic genome sizes, but wider sampling of archaeal genomes will throw more light on this question. When did premeiotic doubling and a two-step meiosis evolve, and what selective forces imposed the haploid-diploid system of reduction division (32, 71)? Why was meiosis the solution for an expanded eukaryotic nuclear genome (72, 73), while the endosymbiotic bacterial genome shrank almost to oblivion, lost capacity for LGT, and became a multiploid, asexual, uniparentally transmitted mitochondrial genome (74–76)? All those steps were crucial for the survival and evolution of early eukaryotes. Without them, complex life as we know it could not have survived its inception. Nonetheless, our work here shows why early eukaryotes had to take up whole chromosome-sized pieces of DNA and align them along their full length, rather than simply end matching, clarifying the first necessary step from LGT toward meiosis.

Data, Materials, and Software Availability. Simulation code has been deposited in Github: https://github.com/MarcoColnaghi1990/LGT-repeat-sequences (77).

ACKNOWLEDGMENTS. We are grateful to two anonymous reviewers for their helpful comments and suggestions. This work was supported by funding from the Engineering and Physical Sciences Research Council (EP/F500351/1 and EP/I017909/1) and Natural Environment Research Council (NE/R010579/1) to A.P., the Biotechnology and Biological Sciences Research Council (BB/S003681/1) and bgc3 to N.L., and a joint grant to A.P. and N.L. from the Biotechnology and Biological Sciences Research Council (BB/V003542/1).

1. A. M. Schurko, J. M. Logsdon, Jr., Using a meiosis detection toolkit to investigate ancient asexual "scandals" and the evolution of sex. *BioEssays* **30**, 579–589 (2008).
2. D. Speijer, J. Lukeš, M. Eliáš, Sex is a ubiquitous, ancient, and inherent attribute of eukaryotic life. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8827–8834 (2015).
3. H. Bernstein, C. Bernstein, Evolutionary origin and adaptive function of meiosis. *Meiosis* **1**, 56557 (2013).
4. G. Mirzaghaderi, E. Hörandl, The evolution of meiotic sex and its alternatives. *Proc. Biol. Sci.* **283**, 20161221 (2016).
5. T. Hirano, SMC proteins and chromosome mechanics: From bacteria to humans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **360**, 507–514 (2005).
6. H. P. Erickson, Evolution of the cytoskeleton. *BioEssays* **29**, 668–677 (2007).
7. Z. Lin, H. Kong, M. Nei, H. Ma, Origins and evolution of the recA/RAD51 gene family: Evidence for ancient gene duplication and endosymbiotic gene transfer. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10328–10333 (2006).
8. C. Johnston, B. Martin, G. Fichant, P. Polard, J.-P. Claverys, Bacterial transformation: Distribution, shared mechanisms and divergent control. *Nat. Rev. Microbiol.* **12**, 181–196 (2014).
9. O. H. Ambur, J. Engelstädter, P. J. Johnsen, E. L. Miller, D. E. Rozen, Steady at the wheel: Conservative sex and the benefits of bacterial transformation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **371**, 20150528 (2016).
10. B. R. Levin, O. E. Cornejo, The population and evolutionary dynamics of homologous gene recombination in bacterial populations. *PLoS Genet.* **5**, e1000601 (2009).
11. N. Takeuchi, K. Kaneko, E. V. Koonin, Horizontal gene transfer can rescue prokaryotes from Muller's ratchet: Benefit of DNA from dead cells and population subdivision. *G3 (Bethesda)* **4**, 325–339 (2014).
12. M. Vos, M. C. Hesselman, T. A. Te Beek, M. W. J. van Passel, A. Eyre-Walker, Rates of lateral gene transfer in prokaryotes: High but why? *Trends Microbiol.* **23**, 598–605 (2015).
13. M. Colnaghi, N. Lane, A. Pomiankowski, Genome expansion in early eukaryotes drove the transition from lateral gene transfer to meiotic sex. *eLife* **9**, e58873 (2020).
14. D. A. Baltrus, K. Guillemin, P. C. Phillips, Natural transformation increases the rate of adaptation in the human pathogen Helicobacter pylori. *Evolution* **62**, 39–49 (2008).
15. D. J. Engelmoer, I. Donaldson, D. E. Rozen, Conservative sex and the benefits of transformation in Streptococcus pneumoniae. *PLoS Pathog.* **9**, e1003758 (2013).
16. N. L. Hiller *et al.*, Generation of genic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal pediatric infection. *PLoS Pathog.* **6**, e1001108 (2010).
17. J. P. Claverys, B. Martin, P. Polard, The genetic transformation machinery: Composition, localization, and mechanism. *FEMS Microbiol. Rev.* **33**, 643–656 (2009).
18. N. J. Croucher, S. R. Harris, L. Barquist, J. Parkhill, S. D. Bentley, A high-resolution view of genome-wide pneumococcal transformation. *PLoS Pathog.* **8**, e1002745 (2012).

Downloaded from https://www.pnas.org by 147.94.12.7 on August 23, 2022 from IP address 147.94.12.7.

19. J. N. Timmis, M. A. Ayliffe, C. Y. Huang, W. Martin, Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
20. W. F. Martin, S. Garg, V. Zimorski, Endosymbiotic theories for eukaryote origin. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140330 (2015).
21. E. V. Koonin *et al.*, A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* **5**, R7 (2004).
22. K. S. Makarova, Y. I. Wolf, S. L. Mekhedov, B. G. Mirkin, E. V. Koonin, Ancestral paralogs and pseudoparalogs and their role in the emergence of the eukaryotic cell. *Nucleic Acids Res.* **33**, 4626–4638 (2005).
23. N. Lane, How energy flow shapes cell evolution. *Curr. Biol.* **30**, R471–R476 (2020).
24. N. Lane, W. Martin, The energetics of genome complexity. *Nature* **467**, 929–934 (2010).
25. J. Vosseberg *et al.*, Timing the origin of eukaryotic cellular complexity with ancient duplications. *Nat. Ecol. Evol.* **5**, 92–100 (2021).
26. F. D. Tria *et al.*, Gene duplications trace mitochondria to the onset of eukaryote complexity. *Genome. Biol. Evol.* **13**, evab055 (2021).
27. E. V. Koonin, The origin of introns and their role in eukaryogenesis: A compromise solution to the introns-early versus introns-late debate? *Biol. Direct* **1**, 22 (2006).
28. W. Martin, E. V. Koonin, Introns and the origin of nucleus-cytosol compartmentalization. *Nature* **440**, 41–45 (2006).
29. I. B. Rogozin, L. Carmel, M. Csuros, E. V. Koonin, Origin and evolution of spliceosomal introns. *Biol. Direct* **7**, 11 (2012).
30. A. M. Lambowitz, M. Belfort, Mobile bacterial group II introns at the crux of eukaryotic evolution. *Microbiol. Spectrum.* 3, MDNA3-0050-2014 (2015).
31. G. Lee *et al.*, Testing the retroelement invasion hypothesis for the emergence of the ancestral eukaryotic cell. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 12465–12470 (2018).
32. A. S. Wilkins, R. Holliday, The evolution of meiosis from mitosis. *Genetics* **181**, 3–12 (2009).
33. J. Vandecraen, M. Chandler, A. Aertsen, R. Van Houdt, The impact of insertion sequences on bacterial genome plasticity and adaptability. *Crit. Rev. Microbiol.* **43**, 709–730 (2017).
34. P. Siguier, E. Gourbeyre, M. Chandler, Bacterial insertion sequences: Their genomic impact and diversity. *FEMS Microbiol. Rev.* **38**, 865–891 (2014).
35. H. J. Muller, The relation of recombination to mutational advance. *Mutat. Res.* **106**, 2–9 (1964).
36. J. Haigh, The accumulation of deleterious genes in a population–Muller's Ratchet. *Theor. Popul. Biol.* **14**, 251–267 (1978).
37. M. Lynch, R. Bürger, D. Butcher, W. Gabriel, The mutational meltdown in asexual populations. *J. Hered.* **84**, 339–344 (1993).
38. S. P. Otto, T. Lenormand, Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3**, 252–261 (2002).
39. S. P. Otto, The evolutionary enigma of sex. *Am. Nat.* **174** (Suppl. 1), S1–S14 (2009).
40. G. Bell, *The Masterpiece of Nature: The Evolution and Genetics of Sexuality* (Routledge, 1982).
41. M. F. Polz, E. J. Alm, W. P. Hanage, Horizontal gene transfer and the evolution of bacterial and archaeal population structure. *Trends Genet.* **29**, 170–175 (2013).
42. J. P. Gogarten, J. P. Townsend, Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3**, 679–687 (2005).
43. H. Ochman, J. G. Lawrence, E. A. Groisman, Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304 (2000).
44. D. Roze, Disentangling the benefits of sex. *PLoS Biol.* **10**, e1001321 (2012).
45. J. Maynard Smith, *The Evolution of Sex* (Cambridge University Press, Cambridge, 1978), **vol. 4**.
46. W. G. Hill, A. Robertson, The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294 (1966).
47. D. Roze, N. H. Barton, The Hill-Robertson effect and the evolution of recombination. *Genetics* **173**, 1793–1811 (2006).
48. S. P. Otto, Selective interference and the evolution of sex. *J. Hered.* **112**, 9–18 (2021).
49. C. M. Thomas, K. M. Nielsen, Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.* **3**, 711–721 (2005).
50. B. J. Arnold, I. T. Huang, W. P. Hanage, Horizontal gene transfer and adaptive evolution in bacteria. *Nat. Rev. Microbiol.* **20**, 206–218 (2022).
51. A. Naor, U. Gophna, Cell fusion and hybrids in Archaea: Prospects for genome shuffling and accelerated strain development for biotechnology. *Bioengineered* **4**, 126–129 (2013).
52. A. Naor, P. Lapierre, M. Mevarech, R. T. Papke, U. Gophna, Low species barriers in halophilic archaea and the formation of recombinant hybrids. *Curr. Biol.* **22**, 1444–1448 (2012).
53. M. Touchon, E. P. Rocha, Causes of insertion sequences abundance in prokaryotic genomes. *Mol. Biol. Evol.* **24**, 969–981 (2007).
54. A. Wagner, Periodic extinctions of transposable elements in bacterial lineages: Evidence from intragenomic variation in multiple genomes. *Mol. Biol. Evol.* **23**, 723–733 (2006).
55. B. van Dijk, F. Bertels, L. Stolk, N. Takeuchi, P. B. Rainey, Transposable elements promote the evolution of genome streamlining. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200477 (2022).
56. E. V. Koonin, Intron-dominated genomes of early ancestors of eukaryotes. *J. Hered.* **100**, 618–623 (2009).
57. J. C. Mell, J. Y. Lee, M. Firme, S. Sinha, R. J. Redfield, Extensive cotransformation of natural variation into chromosomes of naturally competent *Haemophilus influenzae*. *G3 (Bethesda)* **4**, 717–731 (2014).
58. P. Puigbò, A. E. Lobkovsky, D. M. Kristensen, Y. I. Wolf, E. V. Koonin, Genomes in turmoil: Quantification of genome dynamics in prokaryote supergenomes. *BMC Biol.* **12**, 66 (2014).
59. A. Norman, L. H. Hansen, S. J. Sørensen, Conjugative plasmids: Vessels of the communal gene pool. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **364**, 2275–2289 (2009).
60. H. Heuer, K. Smalla, Plasmids foster diversification and adaptation of bacterial populations in soil. *FEMS Microbiol. Rev.* **36**, 1083–1104 (2012).
61. E. Harrison, M. A. Brockhurst, Plasmid-mediated horizontal gene transfer is a coevolutionary process. *Trends Microbiol.* **20**, 262–267 (2012).
62. T. F. Cooper, Recombination speeds adaptation by reducing competition between beneficial mutations in populations of Escherichia coli. *PLoS Biol.* **5**, e225 (2007).
63. J. C. Mell, I. M. Hall, R. J. Redfield, Defining the DNA uptake specificity of naturally competent Haemophilus influenzae cells. *Nucleic Acids Res.* **40**, 8536–8549 (2012).
64. J. de Vries, W. Wackernagel, Integration of foreign DNA during natural transformation of Acinetobacter sp. by homology-facilitated illegitimate recombination. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2094–2099 (2002).
65. N. Hülter, W. Wackernagel, Double illegitimate recombination events integrate DNA segments through two different mechanisms during natural transformation of Acinetobacter baylyi. *Mol. Microbiol.* **67**, 984–995 (2008).
66. S. A. Frye, M. Nilsen, T. Tønjum, O. H. Ambur, Dialects of the DNA uptake sequence in Neisseriaceae. *PLoS Genet.* **9**, e1003458 (2013).
67. M. A. Brockhurst *et al.*, The ecology and evolution of pangenomes. *Curr. Biol.* **29**, R1094–R1103 (2019).
68. J. O. McInerney, A. McNally, M. J. O'Connell, Why prokaryotes have pangenomes. *Nat. Microbiol.* **2**, 17040 (2017).
69. A. J. Roger, E. Susko, M. M. Leger, Evolution: Reconstructing the Timeline of Eukaryogenesis. *Curr. Biol.* **31**, R193–R196 (2021).
70. D. Moi *et al.*, Discovery of archaeal Fusexins homologous to eukaryotic HAP2/GCS1 gamete fusion proteins. *Nat. Genet.* **13**, 3880 (2022).
71. M. Archetti, Loss of complementation and the logic of two-step meiosis. *J. Evol. Biol.* **17**, 1098–1105 (2004).
72. A. S. Kondrashov, The asexual ploidy cycle and the origin of sex. *Nature* **370**, 213–216 (1994).
73. A. V. Markov, I. S. Kaznacheev, Evolutionary consequences of polyploidy in prokaryotes and the origin of mitosis and meiosis. *Biol. Direct* **11**, 28 (2016).
74. A. J. Roger, S. A. Muñoz-Gómez, R. Kamikawa, The origin and diversification of mitochondria. *Curr. Biol.* **27**, R1177–R1192 (2017).
75. M. W. Gray, G. Burger, B. F. Lang, The origin and early evolution of mitochondria. *Genome Biol.* **2**, S1018 (2001).
76. G. Burger, M. W. Gray, B. F. Lang, Mitochondrial genomes: Anything goes. *Trends Genet.* **19**, 709–716 (2003).
77. M. Colnaghi, N. Lane, A. Pomiankowski, Code from "Repeat sequences limit the effectiveness of lateral gene transfer and favoured the evolution of meiotic sex in early eukaryotes". Github. https://github.com/MarcoColnaghi1990/LGT-repeat-sequences. Deposited 31 January 2022.

Downloaded from https://www.pnas.org by 147.94.12.7 on August 23, 2022 from IP address 147.94.12.7.