Recombination Rate Variation and Infrequent Sex Influence Genetic Diversity in *Chlamydomonas reinhardtii*

Ahmed R. Hasan [ID] [1,2,*] and Rob W. Ness[1,2]

[1]Department of Cell and Systems Biology, University of Toronto, Ontario, Canada

[2]Department of Biology, University of Toronto Mississauga, Ontario, Canada

*Corresponding author: E-mail: ahmed.hasan@mail.utoronto.ca.

Abstract

Recombination confers a major evolutionary advantage by breaking up linkage disequilibrium between harmful and beneficial mutations, thereby facilitating selection. However, in species that are only periodically sexual, such as many microbial eukaryotes, the realized rate of recombination is also affected by the frequency of sex, meaning that infrequent sex can increase the effects of selection at linked sites despite high recombination rates. Despite this, the rate of sex of most facultatively sexual species is unknown. Here, we use genomewide patterns of linkage disequilibrium to infer fine-scale recombination rate variation in the genome of the facultatively sexual green alga *Chlamydomonas reinhardtii*. We observe recombination rate variation of up to two orders of magnitude and find evidence of recombination hotspots across the genome. Recombination rate is highest flanking genes, consistent with trends observed in other nonmammalian organisms, though intergenic recombination rates vary by intergenic tract length. We also find a positive relationship between nucleotide diversity and physical recombination rate, suggesting a widespread influence of selection at linked sites in the genome. Finally, we use estimates of the effective rate of recombination to calculate the rate of sex that occurs in natural populations, estimating a sexual cycle roughly every 840 generations. We argue that the relatively infrequent rate of sex and large effective population size creates a population genetic environment that increases the influence of selection on linked sites across the genome.

Key words: recombination rate variation, selection at linked sites, frequency of sex, *Chlamydomonas*.

Introduction

Recombination is both a fundamental evolutionary process and required to ensure proper disjunction of chromosomes during meiosis. Meiotic recombination has two possible outcomes: crossing over (CO) and noncrossing over, also known as gene conversion. At the population level, recombination reduces interference between linked adaptive and harmful mutations and is therefore an important determinant of how well natural selection can act (Hill and Robertson 1966). There is clear evidence that recombination rate varies at multiple scales across nature, with variability observed within and between taxa (de Massy 2013; Stapley et al. 2017) as well as within the genome (Nachman 2002; Choi and Henderson 2015). Across the genome, COs often cluster in localized elevations known as "hotspots" (Choi and Henderson 2015). This fine-scale variation means that local recombination rate may substantially affect the rate of

adaptation and accumulation of harmful mutations at a locus (Felsenstein 1974; McVean et al. 2004).

A major determinant of recombination rate in a population is the frequency with which sex occurs. Even if physical recombination rate is high, the effective rate of recombination in a population may be lower due to infrequent sex or high rates of inbreeding (Charlesworth and Wright 2001; Wright et al. 2008). Reduced opportunity for recombination between heterozygous chromosomes should extend linkage over larger tracts of the genome than would be expected for species with large effective population sizes and high physical recombination rates. In species that switch between asexual and sexual generations, as is the case in many eukaryotic microbes (Dacks and Roger 1999; Weedall and Hall 2015), the influence of selection at linked sites in determining patterns of diversity will vary with not only the physical rate of recombination but also the frequency of sex in the population

(Hartfield 2016). The influence of selection at linked sites may explain why genetic diversity is constrained to a narrow range in nature relative to the scale of variation in population size (Maynard Smith and Haigh 1974; Coop and Ralph 2012; Corbett-Detig et al. 2015; but see Coop 2016), especially when considering protists with extremely large census population sizes (e.g., Filatov 2019).

Despite the fact that protists represent the majority of eukaryotic diversity (Burki et al. 2008), recombination in protists has seen little investigation. A recent review reports only nine linkage maps from species in the Stramenopiles–Alveolates–Rhizaria supergroup, as opposed to over 300 across animals and plants (Stapley et al. 2017). Although laboratory crosses indicate generally high rates of recombination in protists (∼10–50 cM/Mb, e.g., Heesch et al. 2010; Blake et al. 2011), these estimated rates do not consider the influence of frequency of sex in nature. It is estimated that the frequency of sex is unknown in the vast majority (>99%) of free-living protist species (Weisse 2008), to the point where it is unknown whether many are sexual at all (Schurko et al. 2009; D'Souza and Michiels 2010). Although very difficult to measure in nature, the rate of sex can be estimated as the relative frequency of meioses to mitoses by combining direct estimates of the recombination (r) and mutation (μ) rate with population estimates of genetic diversity ($4N_e\mu$) and the effective recombination rate ($4N_er$) (Tsai et al. 2008). To date, this technique has only been used in a limited set of organisms (e.g., Tsai et al. 2008; Grimsley et al. 2010) because there are very few in which diversity (θ), population recombination rate (ρ), physical recombination rate (r), and mutation rate (μ) are known.

Here, we use population genomic data from the unicellular, facultatively sexual green alga *Chlamydomonas reinhardtii* to examine fine-scale variation in the population recombination rate across the genome. We also integrate laboratory-based estimates of mutation and physical recombination rate to infer the frequency of sex in nature. Specifically, we address the following questions: 1) What is the landscape of recombination rate variation across the genome of *C. reinhardtii*? 2) What genomic features predict recombination rate variation in the genome? 3) What is the rate of sex in natural populations of *C. reinhardtii*? and 4) How does recombination rate affect patterns of neutral diversity?

Materials and Methods

Strains, Sequencing, and Alignment

Here we use genome sequence from 19 (9 *MT+*, 10 *MT−*) natural strains of *C. reinhardtii*, sampled from Quebec, Canada. For strains CC-2935, CC-2936, CC-2937, and CC-2938, we obtained publicly available sequencing data (Flowers et al. 2015), whereas the remainder of the strains were originally published in Ness et al. (2016). These 19

haploid strains (supplementary table S1, Supplementary Material online) are all sampled from two nearby localities in Quebec and show no evidence of population structure (Craig et al. 2019). We aligned 100-bp paired-end reads with the Burrows-Wheeler Aligner 0.7.4-r385 (Li and Durbin 2009); as the *C. reinhardtii* reference genome is derived from an *MT+* individual and contains no organelle genomes, we appended the *MT−* locus and organelle genomes to the reference to allow mapping of reads derived from these regions. The GATK v3.3 tools HaplotypeCaller and GenotypeGVCFs were then used to call single nucleotide polymorphisms (SNPs) and short indels and stored in Variant Call Format files (nondefault settings: ploidy = 1, includeNonVariantSites=true, heterozygosity = 0.02, indel_heterozygosity = 0.002). Ploidy was set to 1 because all strains included are haploid. Heterozygosity values were approximately equal to the genomewide average rate of polymorphism (θ_π) across silent and coding sites in the Quebec subpopulation (Craig et al. 2019).

Of the 6,497,950 SNPs identified, we retained 4,736,814 high-quality SNPs for recombination rate estimation after using the following site filters: 1) mean genotype quality across all calls >30, 2) only one alternate allele identified (i.e., diallelic), and 3) minor allele frequency >0.1 (to exclude singletons). All filtering was performed with bcftools (https://github.com/samtools/bcftools; last accessed April 1st, 2020) and cyvcf2 (Pedersen and Quinlan 2017).

Estimation of Recombination Rate Variation

To obtain chromosomewide maps of recombination rate variation in the *C. reinhardtii* genome, we used LDhelmet 1.9 (Chan et al. 2012), which calculates fine-scale estimates of population recombination in intervals bounded by adjacent SNPs. The coalescent-based approach of LDhelmet allows for inference of ancestral recombination rate variation. LDhelmet reports the population recombination rate $\rho = 2N_e r$ which reflects the size of the recombining population (N_e) and the physical recombination rate in sexual generations (r, recombination events bp^{-1} generation^{-1}).

Mutation Matrix Estimation

LDhelmet incorporates a quadra-allelic mutation model, allowing for transition probabilities between the four nucleotides to be specified and taken into account in recombination rate inferences. The method for estimating this transition matrix specified by Chan et al. (2012) involves comparison with an outgroup to designate ancestral alleles. However, because no outgroup genome is currently available for *C. reinhardtii*, we instead used data from a prior mutation accumulation study in *C. reinhardtii* (Ness et al. 2015) to estimate transition probabilities. Using a data set of 5,710 single nucleotide mutations, we designated the prior states of mutated sites within each strain as "ancestral" and otherwise followed

the method of Chan et al. (2012) to estimate the mutation matrix, which was as follows (with both rows and columns ordered as A, C, G, T):

$$\begin{bmatrix} 0.61 & 0.11 & 0.19 & 0.09 \\ 0.26 & 0 & 0.22 & 0.52 \\ 0.5 & 0.22 & 0.03 & 0.25 \\ 0.11 & 0.2 & 0.11 & 0.57 \end{bmatrix}.$$

Block Penalty

The block penalty parameter in LDhelmet determines the level of smoothing of ρ estimates along the chromosomes. To ascertain the optimal block penalty for our data set, we adapted a Python script from Singhal et al. (2015) to perform simulations with the coalescent-based simulator macs-0.5d (Chen et al. 2009). We simulated haplotypes under a range of six background recombination rates ($\rho = 0.0001$, 0.001, 0.01, 0.1, 1.0, and 2.5). These values were based on the range of values observed in preliminary LDhelmet runs with default parameters. For each of these background recombination rates, we simulated 19 1-Mb sequences, assigning bases to derived alleles using the mutation matrix estimated above. Additionally, following Singhal et al. (2015), we placed eight 2-kb recombination hotspots along the length of each sequence, with two hotspots each of relative fold increases 10×, 20×, 40×, and 60×. We used a window size of 2 kb since hotspots are 2 kb or larger in many species (Kim et al. 2007; Coop et al. 2008; Singhal et al. 2015; Booker et al. 2017). To test for the effect of variation in hotspot length, or proportion of hotspots in the genome, we repeated these simulations with 4 and 6 kb hotspot sizes. Each simulation was repeated ten times. We then ran LDhelmet across block penalties 5, 10, 50, and 100 on all simulated haplotypes and compared the observed LDhelmet recombination estimates with expected recombination rates. For our LDhelmet runs, we used a θ value of 0.03, based on a prior empirical estimates of neutral diversity (Flowers et al. 2015; Craig et al. 2019). Comparing observed ρ with our known recombination rates, we found that a block penalty of 100 yielded a ratio of observed/expected ρ closest to 1 across all simulated recombination rates (supplementary fig. S1, Supplementary Material online).

Power to Detect Hotspots

To detect recombination hotspots from LDhelmet ρ estimates, we modified the "find_hotspots.py" Python script from Singhal et al. (2015) that summarizes ρ values in nonoverlapping windows. Hotspots are elevations in recombination rate at a locus compared with its surrounding regions; however, whether or not a region is classified as a hotspot will be affected by both the block penalty as well as the size of the

flanking regions. To determine the optimal block penalty and flank size to best detect recombination hotspots in our data, we averaged the LDhelmet outputs from our simulations above into nonoverlapping 2-kb windows. For each window, we also calculated recombination rates on either flank over various flank sizes (ranging from 20 to 100 kb in increments of 20 kb). Then, we examined how many of our known simulated hotspots across all combinations of block penalty values and flank sizes demonstrated at least a 5-fold increase in ρ compared with the flanks. We found that a block penalty of 10 and a flank size of 100 kb provided the highest power (1 − number of false negatives) to detect hotspots across all simulated background ρ values, including $\rho = 0.001$, which is the closest to expected recombination rate in *C. reinhardtii* following a preliminary LDhelmet run (supplementary fig. S2, Supplementary Material online). Based on these simulations, our power to detect hotspots is relatively high (∼0.9). Following these simulations, we defined hotspots as regions that 1) are at least 2 kb in length and 2) exhibit a mean 5-fold increase in ρ compared with the surrounding 200 kb of sequence, similar to previous approaches (International HapMap Consortium 2005; Singhal et al. 2015; Booker et al. 2017).

Estimating the Recombination Landscape

To estimate the genomic landscape of recombination, we ran LDhelmet twice for 1,000,000 iterations on each chromosome of *C. reinhardtii* following 100,000 iterations of burn-in. The two runs were performed with the same parameters save for the block penalty. In the first run, which was used to estimate the landscape of genomewide ρ, we used a block penalty of 100. For the second run, we changed the block penalty to ten for better hotspot detection given the results of our simulations (see above). As with our simulations, we used a θ value of 0.03 (Craig et al. 2019) and the mutation matrix shown above in both runs. To detect hotspots, we used our adaptation of the find_hotspots.py Python script (see Data Availability) from Singhal et al. (2015) to summarize LDhelmet output into nonoverlapping 2-kb windows, while also summarizing ρ over the surrounding 200 kb for each window. We used the LDhelmet run with a block penalty of 100 for all analyses except for hotspot detection.

LD Decay across Chromosomes

Pairwise calculations of linkage disequilibrium (r^2) between SNPs within each of *C. reinhardtii*'s 17 chromosomes were conducted using plink 1.90 (Chang et al. 2015). For all pairs of SNPs, plink calculates LD statistics with a maximum likelihood approach described in Gaunt et al. (2007). By default, plink filters out pairs of SNPs with an r^2 value below 0.2; we disabled this filtering. We calculated r^2 for all pairs of SNPs within 100 kb of one another, and modeled the expected decay of LD with distance for each chromosome with nonlinear least

squares regression in R (R Core Team, 2017) using the following equation from Appendix 2 of Hill and Weir (1988):

$$E(r^2) = \frac{10 + \Gamma}{22 + 13\Gamma + \Gamma^2} \times \left[1 + \frac{(3 + \Gamma)(12 + 12\Gamma + \Gamma^2)}{n(22 + 13\Gamma + \Gamma^2)}\right],$$
(1)

where Γ is the product of population recombination rate ρ and distance between sites d.

Genomic Correlates of Recombination Rate

We subclassified genic sites in the reference genome of *C. reinhardtii* as protein-coding sequence (CDS), introns, and UTRs. Intergenic sites were subclassified as being within 2 kb upstream of a gene ("flank", i.e., gene proximate) or more than 2 kb from the nearest gene ("nonflank"). Upon finding that ρ in sites upstream of a gene varied based on intergenic tract length (see Results), we binned intergenic tracts by size into tracts <2 kb (i.e., flanked by genes within 2 kb on either side) and tracts >2 kb. For each of the above, average ρ was calculated from every corresponding site in the genome. Recombination rates for each annotation (and each bin, in the case of intergenic sequence) were bootstrapped for 1,000 replicates to obtain 95% confidence intervals. For the correlation of GC content with recombination rate, we used a custom Python script to compute GC content in nonoverlapping 2-kb windows.

LD-based estimates of ρ are expected to be correlated with SNP density due to the action of background selection and selective sweeps. Although we know that the resolution of the recombination landscape is affected by SNP density, ρ estimates from methods such as LDhelmet should be unbiased with respect to SNP density. However, subsampling SNPs has been shown to sometimes cause reductions in ρ estimates (Chan et al. 2012). Thus, to ascertain whether there were differences in recombination rate between annotations while controlling for variation in SNP density, we split the genome into tracts of annotations of interest and calculated ρ and SNP density for each tract. We then fit a multiple regression model with predictors annotation, SNP density, and their interaction, whereas ρ was set as the response variable.

Recombination and Nucleotide Diversity

Although a correlation between recombination rate and neutral diversity is suggestive of the effects of selection at linked sites, LD-based estimates of ρ are inappropriate for correlations with nucleotide diversity because both statistics are scaled by N_e. This means that the two measures will be autocorrelated due to demography, selection, and variance in coalescence times across the genome. We therefore tested the relationship between recombination and diversity using crossover data from Liu et al. (2018), who sequenced 108 offspring from 27 tetrads. First, we examined the concordance

between our ρ estimates and these COs by testing for an enrichment of COs in genomic regions with high ρ. We binned our ρ estimates for genomic intervals into 50 equal-sized bins between ρ values of 0 and 0.06. For each bin, we counted the number of COs found in regions of the genome corresponding to that range of ρ values. The count of COs was converted to an estimate of physical recombination rate in cM/Mb by dividing the number of COs in each bin by the number of individuals in the data set (108), multiplying by 100 (since 1 cM represents a 1% chance of a CO), and then dividing the resulting per individual measure by the number of sites falling within each ρ bin.

Next, to examine the relationship between diversity and CO density, we calculated θ_π across the genome in 10-kb windows, and then binned these values across 50 bins ranging from θ_π values of 0 to 0.1. Windows with <500 silent sites were discarded to reduce noise in diversity estimates, whereas bins with <200 kb were discarded because with so few COs in the data set these bins tended to have zero COs. Then, we assigned COs to bins based on local diversity in the region bounded by each CO and calculated CO rates in cM/Mb as above.

Results

The *C. reinhardtii* Recombination Landscape Is Variable and Hotspot-Punctuated

From our population sample of 19 individuals (supplementary table S1, Supplementary Material online), we calculated fine-scale recombination landscapes across the genome of *C. reinhardtii* with LDhelmet 1.9 (Chan et al. 2012). The genomewide average population recombination rate was $\rho = 0.0041$/bp, and mean recombination rate for each chromosome varied >4.5-fold from 0.0024 to 0.0109, inversely scaling with chromosome lengths (supplementary fig. S3, Supplementary Material online, $R^2 = 0.365$, $P = 0.01$). ρ estimates across the genome were then summarized in nonoverlapping 2-kb windows for fine-scale analysis, with 95% of windows ranging from $\rho = 2 \times 10^{-5}$ to 0.017; the distribution of recombination rates is shown in figure 1A. Analysis of LD decay measured as r^2 dropped to half of their starting value within a mean distance of 577 bp (fig. 1B; range 247–1,682 bp). Moreover, the decay of r^2 approached baseline levels within a mean distance of 9,700 bp, where "baseline" was defined as the point at which the instantaneous rate of change in r^2 with physical distance approximates zero to five significant digits (mean r^2 at level off point = 0.074 ± 0.0009).

To examine recombination hotspots, defined as a region that was: 1) >2 kb in length and 2) exhibited a >5-fold increase in ρ relative to the flanking 200 kb of sequence. Our hotspot definition was based on a power analysis using simulated sequence data (see Materials and Methods). Under this definition, we found hotspots in all chromosomes, with 875
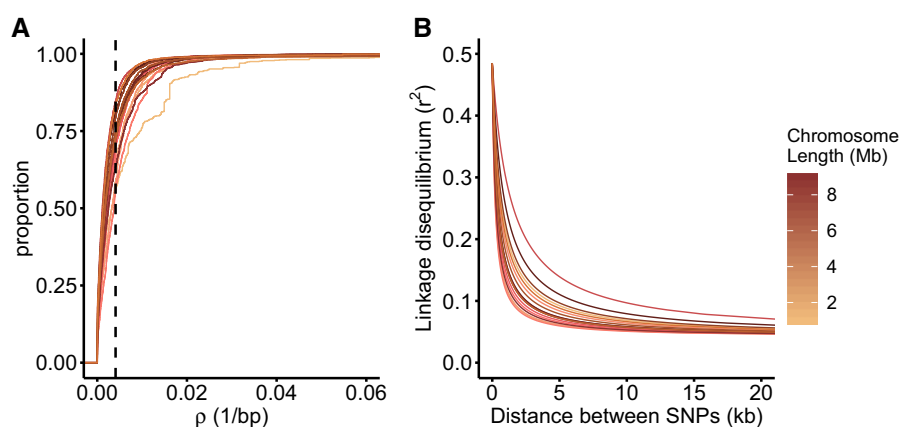
FIG. 1.—(A) Cumulative frequency distribution of population recombination rate (ρ) for each chromosome of *Chlamydomonas reinhardtii*. Each curve represents one of the 17 chromosomes, shaded by chromosome length. ρ values were summarized in 2-kb windows. The vertical dashed line indicates the genomewide mean ρ value. (B) Decay of linkage disequilibrium (r^2) across the 17 chromosomes of *C. reinhardtii*, modeled using the equation provided in Appendix 2 of Hill and Weir (1988).

hotspot regions in total representing 2.75% of the genome, where the average ρ within hotspots was more than 26 times the genome average (mean ρ at hotspots = 0.1046, mean length = 3,428 bp, mean ρ fold increase over local background = 20.6×, mean distance between adjacent hotspots = 122.1 kb).

Recombination Rates Are Highest Immediately Surrounding Genes

To investigate the correlates of recombination across the genome, we examined how ρ varied with different functional annotations in the *C. reinhardtii* reference genome (Merchant et al. 2007). Within genes, we found that ρ was significantly higher in coding sequence than introns (fig. 2A and supplementary table S2, Supplementary Material online; β_{coding} = 0.31, $P = 2.2 \times 10^{-16}$). UTRs displayed the lowest recombination rates of any annotation (fig. 2A, mean ρ 5' UTR = 0.00313, mean ρ 3' UTR = 0.00351). By contrast, we found that intergenic regions had the highest mean ρ (= 0.00452) of all annotations.

We found that recombination rate was 12.8% higher in intergenic regions (mean $\rho = 0.00452$) than genic regions (mean $\rho = 0.00401$, Mann–Whitney U test, $P < 2.2 \times 10^{-16}$). Within genic DNA, coding exons had significantly higher rates of recombination than introns (Mann–Whitney U test, $P = 3.41 \times 10^{-9}$; ρ CDS = 0.0044, ρ introns = 0.0039).

Within intergenic regions, recombination rate was highest in sequence proximate to genes, with sites within 2 kb of genes displaying 11.7% higher mean ρ than the genome background (mean ρ of sites within 2 kb of genes = 0.0046). However, we observed that patterns of intergenic recombination varied by intergenic tract length. In the *C. reinhardtii* genome, most intergenic tracts are short, with

50% of all intergenic tracts below 142 bp in size, whereas 91% are below 2 kb (supplementary fig. S4, Supplementary Material online). We observed that in intergenic tracts <2 kb recombination rate was lower than the genomewide average (fig. 2B, $\rho = 0.0036$). Conversely, in longer (>2 kb) intergenic tracts, which are less common but represent 9.9% of all genome sequence, recombination rate was slightly higher than the genome average (mean $\rho = 0.0050$). After controlling for SNP density (see above), recombination rate was still significantly higher in longer intergenic tracts than in tracts <2 kb (supplementary table S3, Supplementary Material online; $\beta_{long} = 0.24$, $P < 2.2 \times 10^{-16}$). At the ends of these longer tracts, recombination was highest and nearly 1.5× the genome average (mean ρ in 2 kb upstream of genes = 0.0057, mean ρ in 2 kb downstream of genes = 0.0061). This trend is also reflected in the UTRs flanking these intergenic tracts: ρ in 5' UTRs adjacent to shorter (<2 kb) intergenic tracts is 4.7% lower than those adjacent to tracts >2 kb.

Intergenic sequence upstream of genes in >4 kb tracts was enriched for hotspots (Fisher's exact test, odds ratio = 3.25, $P < 2.2 \times 10^{-16}$). 5.6% of all hotspots occur in the 1.9% of the genome corresponding to these sites. We also found a positive correlation between recombination rate and GC content at fine (2 kb) scales (Spearman's $\rho = 0.289$, $P < 2.2 \times 10^{-16}$) but not at broad (1 Mb) scales (Spearman's $\rho = 0.031$, $P = 0.737$).

Recombination Rate Is Correlated with Nucleotide Diversity

If background selection and selective sweeps are common, we expect to observe reduced neutral diversity in regions of low recombination. Thus, to examine the relationship between recombination and neutral diversity in the *C. reinhardtii* genome, we tested whether COs detected in a prior tetrad sequencing experiment (Liu et al. 2018) tended
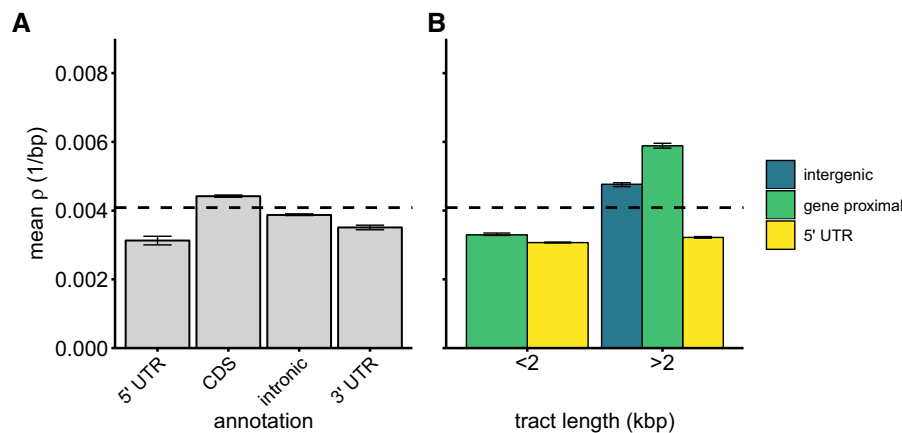
FIG. 2.—Population recombination rate (ρ) in different annotation categories across the *Chlamydomonas reinhardtii* genome. Error bars represent bootstrapped 95% confidence intervals ($n = 1,000$). The dashed horizontal line represents the mean genomewide ρ value. (A) Mean ρ in genic annotations. (B) Mean ρ in intergenic sequence and 5′ UTRs by intergenic tract length. The "gene-proximal" annotation represents the 2 kb of sequence at the 3′ end of a given tract (i.e., upstream of the nearest gene), whereas "intergenic" represents the remainder of the tract.
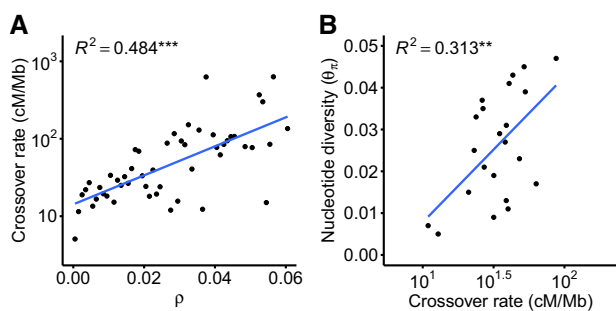

FIG. 3.—(A) Population recombination rate (ρ) is correlated with cross-over density, as obtained from the crossover data set of Liu et al. (2018). (B) Crossover rates from the Liu et al. (2018) data set correlate with silent site diversity. Nucleotide diversity was calculated at intronic, intergenic, and 4-fold degenerate sites. **$P<0.001$, ***$P<0.0001$.

to colocalize with regions of high ρ. Then, we binned our genomewide ρ values and assigned COs to each bin to create a measure of physical recombination from CO density (see Materials and Methods) and found a positive and significant relationship between ρ and crossover recombination rates (fig. 3A, $R^2 = 0.484$, $P = 1.46 \times 10^{-8}$), indicating that our LDhelmet ρ values reflected real variation in the locations of COs in the genome. Using a similar binning method (see Materials and Methods), we found physical recombination rate and θ_π were also significantly correlated (fig. 3B, $R^2 = 0.313$, $P = 0.0068$), consistent with widespread reduction of genetic diversity due to the action of selection at linked sites.

Estimating the Frequency of Sex in *C. reinhardtii*

Due to the fact that mutations can arise each cell division (each meiosis and mitosis) yet recombination only occurs during the fraction of cell divisions that are sexual (f), we can use

estimates of neutral diversity ($\theta = 2N_e\mu$) and population recombination rate ($\rho = 2N_e r$) combined with lab estimates of μ and r to roughly estimate the relative frequency of meiosis to mitosis, or the frequency of sexual relative to clonal reproduction (Ruderfer et al. 2006; Tsai et al. 2008; Grimsley et al. 2010), as follows:

$$\frac{\text{Number of meioses}}{\text{Number of mitoses}} \sim \frac{\rho}{\theta} = \frac{2N_e r f}{2N_e \mu}.$$

In facultatively sexual species, the realized rate of recombination per generation is the product of physical recombination rate (r) and the fraction of generations that are sexual f (=meioses/mitoses). We can therefore express the above in terms of f:

$$f = \frac{\text{Number of meioses}}{\text{Number of mitoses}} \sim \frac{\rho/r}{\theta/\mu}.$$

Thus, our genomewide ρ estimate of 4.09×10^{-3} can be used in tandem with previous estimates of the *C. reinhardtii* recombination rate ($r = 12$ cM/Mb) (Liu et al. 2018), the mutation rate ($\mu = 9.63 \times 10^{-10}$) (Ness et al. 2015), and neutral diversity ($\theta = 2.75 \times 10^{-2}$) (Ness et al. 2016) to solve for f and estimate the frequency of sex in natural populations of *C. reinhardtii*. With this approach, we obtain $f = 0.001194$, corresponding to one sexual generation for every $1/f = \sim$840 asexual generations.

Discussion

In this study, we have estimated fine-scale recombination rate variation *C. reinhardtii* using patterns of linkage disequilibrium, revealing a recombination landscape punctuated with frequent hotspots. We report a genomewide recombination map for this model species, which offers much higher

resolution than a genetic mapping approach would. We found an enrichment of hotspots near genes that leads to an overall increase in recombination rate in intergenic sequence, in concordance with observations in other nonmammalian eukaryotes. Variation in recombination rate across the genome is correlated with nucleotide diversity, suggesting that the influence of selection at linked sites is widespread in the genome and that recombination is a major driver of genetic variation. We have used our estimate of the population recombination rate to estimate the frequency of sex as being once every ~840 generations in *C. reinhardtii*, which may in part explain why a species with a relatively high rate of physical recombination and large effective population size experiences such strong effects of selection at linked sites.

Assuming an N_e of 1.4×10^7 (Ness et al. 2016), a genomewide per bp ρ of 4.09×10^{-3} corresponds to an estimate of $r = \rho/2N_e = 0.015$ cM/Mb. This estimate of genomewide r on its own is two orders of magnitude below the estimate of r from the genetic map of *C. reinhardtii* (9.15 cM/Mb, Kathir et al. 2003) and most plants (Henderson 2012). The discrepancy between the two estimates of r is likely driven by the fact that the frequency of sex is not accounted for in laboratory crosses, since the genetic map measure of r is otherwise in line with estimates in other protists (Heesch et al. 2010; Blake et al. 2011; Henderson 2012). Our result shows that in facultatively sexual species, the rate of sex needs to be accounted for in studies of recombination, since laboratory crosses can overestimate the realized r (i.e., $r'f$) in nature.

Between chromosomes, we observe 4.5-fold variation in mean recombination rates, and also find that recombination rate inversely correlates with chromosome length (supplementary fig. S3, Supplementary Material online), a relationship consistent with prior studies in a variety of organisms (Kaback et al. 1992; Kawakami, Smeds, et al. 2014). Given that each chromosome requires at least one crossover event to ensure proper meiotic disjunction (Page and Hawley 2003), it follows that shorter chromosomes exhibit higher per base crossover rates, resulting in more pronounced signatures of LD breakdown, as exemplified in avian recombination (Kawakami, Smeds, et al. 2014). Furthermore, *C. reinhardtii* exhibits moderate rates of LD decay across all 17 chromosomes. Our estimates of the distance (≤ 10 kb) at which LD (r^2) decays to baseline LD levels are similar to estimates in *Ostreococcus tauri*, another unicellular green alga (~10 kb, Blanc-Mathieu et al. 2017) as well as estimates in *A. thaliana* (~10 kb, Kim et al. 2007), but shorter than in flycatchers (~17 kb, Kawakami, Backström, et al. 2014). We also note that our estimates are shorter than those obtained in a previous study of *C. reinhardtii* that reported a decay to baseline within ~20 kb (Flowers et al. 2015). The difference between our estimates may be caused by genetic structure in *C. reinhardtii*, where we sequenced isolates all from nearby localities, Flowers et al. utilized a mix of lab strains alongside isolates from a variety of populations across Quebec and Eastern

United States. This disparity in our respective estimates is consistent with barriers to recombination across the geographic range of *C. reinhardtii* in North America, with the resulting population structure increasing LD among variants (Craig et al. 2019).

We find numerous recombination hotspots across the genome, similar to observations in mammals, angiosperms, and yeast (Stapley et al. 2017). On the other hand, this recombination profile is unlike that of both *C. elegans*, which has a comparatively homogenous fine-scale recombination landscape (Rockman and Kruglyak 2009; Kaur and Rockman 2014) as well as *D. melanogaster*, which displays some degree of fine-scale heterogeneity but little evidence for highly localized elevations in recombination rate (Chan et al. 2012; Manzano-Winkler et al. 2013). We see elevated recombination and an enrichment of hotspots within regions immediately flanking genes in *C. reinhardtii*, similar to other taxa lacking the PRDM9 protein that determines hotspot locations in humans and mice (Choi and Henderson 2015). Specifically, recombination hotspots upstream of genes have been observed in fungi (Berchowitz et al. 2009; Tsai et al. 2010; Lam and Keeney 2015), finches (Singhal et al. 2015), as well as angiosperms, such as wheat (Saintenac et al. 2009), maize (Li et al. 2015), monkeyflower (Hellsten et al. 2013), and *A. thaliana* (Wijnker et al. 2013; Choi et al. 2013). In addition, the same pattern is observed in dogs (Canidae family), where PRDM9 was lost relatively recently (Axelsson et al. 2012; Auton et al. 2013). In these PRDM9-lacking organisms, chromatin structure is often invoked as an explanation of recombination hotspot conservation upstream of genes (Wu and Lichten 1994; Lichten 2008; Berchowitz et al. 2009). Nucleosome occupancy is depleted in regions where the DNA needs to be accessible, such as for the purposes of transcription. Promoter regions upstream of genes thus tend to display greater nucleosome depletion, which may in turn allow for recombination machinery to more easily induce double strand breaks in these regions (Pan et al. 2011; Lam and Keeney 2015). Our observations of elevated recombination rate immediately flanking genes suggest a similar mechanism acting in the *C. reinhardtii* genome and show that this trend is even more widespread, extending to green algae.

However, we observe that despite the trend above, recombination rate drops in shorter intergenic tracts. The *C. reinhardtii* genome is remarkably gene dense, with ~17,700 intron-rich genes (Merchant et al. 2007), together spanning 85.8% of the genome. Over 90% of intergenic tracts are <2 kb, but the elevation of recombination rate near genes is primarily driven by the ends of longer intergenic tracts, whereas shorter tracts display far lower recombination rates despite being equally gene proximate. There are a few possible explanations for this trend: First, that shorter intergenic tracts may be the space between functionally related gene clusters, where suppressed recombination has evolved to maintain coadapted gene complexes (Pál and Hurst 2003;

Poyatos and Hurst 2007); second, these gene complexes may share common regulatory sequence upstream of the clusters, such that short intergenic tracts between genes may not be in open chromatin and therefore are less susceptible to double-strand break formation; last, these short intergenic tracts may be dense with conserved functional sequence, and if recombination is mutagenic, historical crossovers may have been purged through purifying selection, reducing the effect of recombination in these tracts. Regardless of the underlying mechanism, in species where genes are tightly packed, there may not be sufficient intergenic space or suitable conditions to localize crossovers in the regions immediately flanking genes as is otherwise observed in other species lacking PRDM9.

We find a positive correlation between GC content and local recombination rate at fine scales. Our result is consistent with a trend seen in other organisms such as yeast (Gerton et al. 2000), mouse (Jensen-Seaman et al. 2004), and humans (Fullerton et al. 2001). There are several possible explanations for this trend: First, that GC-biased gene conversion is leading to increased GC substitutions at gene conversion tracts; second, that recombination preferentially initiates in GC-rich regions; third, that there is more efficient selection for GC content in regions with higher recombination rates (Kliman and Hey 1993; Campos et al. 2012). A recent study revealed weak effects of GC-biased gene conversion from the genome sequences of 27 C. reinhardtii tetrads, in concert with a low overall rate of gene conversion, thus indicating a minor role for biased gene conversion in the evolution of the C. reinhardtii genome (Liu et al. 2018). However, at the 2-kb scale, we obtain a stronger correlation than Liu et al., who report correlations in window sizes ranging from 10 to 200 kb, but we also do not obtain a significant correlation at broader (1 Mb) scales. A stronger GC-recombination correlation when considering historical recombination events suggests that the effects of weak forces governing fine-scale base composition may be more apparent over longer evolutionary timescales.

Using a crossover data set, we find that recombination correlates with nucleotide diversity across the genome of C. reinhardtii, indicating the action of selection at linked sites (fig. 3B). Theory predicts that the correlation of recombination and diversity arises as a consequence of background selection and/or selective sweeps reducing diversity in regions of low recombination (Charlesworth et al. 1993; Cutter and Payseur 2013; Campos et al. 2017). Our result suggests that selection on linked sites is a strong determinant of standing genetic variation in C. reinhardtii. Given that C. reinhardtii is likely to have a very high effective population size (Ness et al. 2012), it is expected that many mutations will be effectively selected (i.e., $N_e s > 1$) (Kimura 1962). However, although the effective population size is very large, the relatively infrequent rate of sex (see below) means that the effective recombination rate is not particularly high relative to obligately sexual species. The interaction of a large N_e facilitating efficient selection alongside reduced recombination due to a facultatively sexual life cycle means that the influence of selection at linked sites may be pronounced in the genome and modulated less by recombination rate per se than would be the case in obligate outcrossers. This principle may be more widespread in unicellular eukaryotes, which live in large populations that are only periodically sexual.

Finally, by integrating lab- and population-based measures of recombination and mutation, we have estimated the rate of sex in C. reinhardtii to be one meiosis approximately every 840 asexual generations. The frequency is higher than estimates in two yeast species, Saccharomyces cerevisiae (~50,000 generations, Ruderfer et al. 2006) and S. paradoxus (~1,000–3,000 generations, Tsai et al. 2008), and is substantially more than the estimated rate of sex in O. tauri (~94,000 generations, Blanc-Mathieu et al. 2017). However, the method we used to estimate the frequency of sex is subject to numerous assumptions, especially neutrality and demographic equilibrium; furthermore, if estimates of ρ are reduced by the effects of reduced diversity due to selection, it may downwardly bias our estimate of the frequency of sex, meaning that the true rate of sex in nature may be higher. Our estimate of sex occurring every 840 generations may point toward a seasonal ecology in C. reinhardtii. Although the precise rate of cell division in nature is unknown, lab cultures exhibit 2–3 doublings every 24 h (Bernstein 1964; Jones 1970; Harris et al. 2009) which means that 840 generations would take 336 days on average. Considering the fact that sex is induced when conditions worsen and zygotes are resistant to freezing, desiccation, and other environmental stressors (Morris et al. 1979; Suzuki and Johnson 2002; Harris et al. 2009), it is plausible that populations of C. reinhardtii in Quebec overwinter as zygotes, undergoing a sexual cycle approximately once per year.

Taken together, our results show that recombination in C. reinhardtii is punctuated, in terms of both the many recombination hotspots within the genome as well as the relative infrequency of sexual reproduction. Within-genome patterns in recombination rate patterns are consistent with other PRDM9-lacking organisms in localizing around genes. Interestingly, recombination in intergenic DNA is concentrated in the large intergenic regions between clusters of tightly packed genes. As with many microbial species, the effective population size of C. reinhardtii is very large, but this does not translate to a very high population recombination rate due to the fact that most generations are asexual. This leads to strong effects of selection at linked sites, which may be a scenario common in microbial eukaryotes more broadly. The periodic nature of recombination in microbial eukaryotes is likely a key difference in their population genetics which will need to be accounted for in future studies.

Data Availability

Strains used in this study (supplementary table S1, Supplementary Material online) are available from the Chlamydomonas Resource Center (chlamycollection.org). Short read data are available at the European Nucleotide Archive under study accession ERP109393. All scripts used in this work are available at https://www.github.com/aays/reinhardtii-ld-rcmb (main analyses) and https://www.github.com/aays/ldhelmet-sims (validation of recombination estimation method via simulations). All statistical tests in this work were implemented using R 3.5.2 (R Core Team 2017).

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

Literature Cited

Auton A, et al. 2013. Genetic recombination is targeted towards gene promoter regions in dogs. PLoS Genet. 9(12):e1003984.

Axelsson E, Webster MT, Ratnakumar A, LUPA Consortium, Ponting CP, Lindblad-Toh K. 2012. Death of PRDM9 coincides with stabilization of the recombination landscape in the dog genome. Genome Res. 22(1):51–63.

Berchowitz LE, Hanlon SE, Lieb JD, Copenhaver GP. 2009. A positive but complex association between meiotic double-strand break hotspots and open chromatin in *Saccharomyces cerevisiae*. Genome Res. 19(12):2245–2257.

Bernstein E. 1964. Physiology of an obligate photoautotroph (*Chlamydomonas moewusii*). I. Characteristics of synchronously and randomly reproducing cells and an hypothesis to explain their population curves. J Protozool. 11(1):56–74.

Blake DP, Oakes R, Smith AL. 2011. A genetic linkage map for the apicomplexan protozoan parasite *Eimeria maxima* and comparison with *Eimeria tenella*. Int J Parasitol. 41(2):263–270.

Blanc-Mathieu R, et al. 2017. Population genomics of picophytoplankton unveils novel chromosome hypervariability. Sci Adv. 3(7):e1700239.

Booker TR, Ness RW, Keightley PD. 2017. The recombination landscape in wild house mice inferred using population genomic data. Genetics 207(1):297–309.

Burki F, Shalchian-Tabrizi K, Pawlowski J. 2008. Phylogenomics reveals a new 'megagroup' including most photosynthetic eukaryotes. Biol Lett. 4:366–369.

Campos JL, Charlesworth B, Haddrill PR. 2012. Molecular evolution in nonrecombining regions of the *Drosophila melanogaster* genome. Genome Biol Evol. 4(3):278–288.

Campos JL, Zhao L, Charlesworth B. 2017. Estimating the parameters of background selection and selective sweeps in *Drosophila* in the presence of gene conversion. Proc Natl Acad Sci U S A. 114(24):E4762–E4771.

Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. PLoS Genet. 8(12):e1003090.

Chang CC, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4(1):7.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134(4):1289–1303.

Charlesworth D, Wright SI. 2001. Breeding systems and genome evolution. Curr Opin Genet Dev. 11(6):685–690.

Chen GK, Marjoram P, Wall JD. 2009. Fast and flexible simulation of DNA sequence data. Genome Res. 19(1):136–142.

Choi K, Henderson IR. 2015. Meiotic recombination hotspots: a comparative view. Plant J. 83(1):52–61.

Choi K, et al. 2013. *Arabidopsis* meiotic crossover hot spots overlap with H2A.Z nucleosomes at gene promoters. Nat Genet. 45(11):1327–1336.

Coop G. 2016. Does linked selection explain the narrow range of genetic diversity across species? BioRxiv. doi: 10.1101/042598.

Coop G, Ralph P. 2012. Patterns of neutral diversity under general models of selective sweeps. Genetics 192(1):205–224.

Coop G, Wen X, Ober C, Pritchard JK, Przeworski M. 2008. High-resolution mapping of crossovers reveals extensive variation in fine-scale recombination patterns among humans. Science 319(5868):1395–1398.

Corbett-Detig RB, Hartl DL, Sackton TB. 2015. Natural selection constrains neutral diversity across a wide range of species. PLoS Biol. 13(4):e1002112.

Craig RJ, et al. 2019. Patterns of population structure and complex haplotype sharing among field isolates of the green alga *Chlamydomonas reinhardtii*. Mol Ecol. 28(17):3977–3993.

Cutter AD, Payseur BA. 2013. Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet. 14(4):262–274.

D'Souza TG, Michiels NK. 2010. The costs and benefits of occasional sex: theoretical predictions and a case study. J Hered. 101(Suppl 1):S34–S41.

Dacks J, Roger AJ. 1999. The first sexual lineage and the relevance of facultative sex. J Mol Evol. 48(6):779–783.

de Massy B. 2013. Initiation of meiotic recombination: how and where? Conservation and specificities among eukaryotes. Annu Rev Genet. 47(1):563–599.

Felsenstein J. 1974. The evolutionary advantage of recombination. Genetics 83(4):845–859.

Filatov DA. 2019. Extreme Lewontin's paradox in ubiquitous marine phytoplankton species. Mol Biol Evol. 36(1):4–14.

Flowers JM, et al. 2015. Whole-genome resequencing reveals extensive natural variation in the model green alga *Chlamydomonas reinhardtii*. Plant Cell. 27(9):2353–2369.

Fullerton SM, Bernardo Carvalho A, Clark AG. 2001. Local rates of recombination are positively correlated with GC content in the human genome. Mol Biol Evol. 18(6):1139–1142.

Gaunt TR, Rodríguez S, Day INM. 2007. Cubic exact solutions for the estimation of pairwise haplotype frequencies: implications for linkage disequilibrium analyses and a web tool 'CubeX'. BMC Bioinformatics. 8(1):428.

Gerton JL, et al. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 97(21):11383–11390.

Grimsley N, Pequin B, Bachy C, Moreau H, Piganeau G. 2010. Cryptic sex in the smallest eukaryotic marine green alga. Mol Biol Evol. 27(1):47–54.

Harris EH, Stern DB, Witman GB. 2009. The Chlamydomonas sourcebook. Vol. 1. San Diego (CA): Elsevier.

Hartfield M. 2016. Evolutionary genetic consequences of facultative sex and outcrossing. J Evol Biol. 29(1):5–22.

Heesch S, et al. 2010. A sequence-tagged genetic map for the brown alga Ectocarpus siliculosus provides large-scale assembly of the genome sequence. New Phytol. 188(1):42–51.

Hellsten U, et al. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population shotgun sequencing. Proc Natl Acad Sci U S A. 110(48):19478–19482.

Henderson IR. 2012. Control of meiotic recombination frequency in plant genomes. Curr Opin Plant Biol. 15(5):556–561.

Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. Genet Res. 8(3):269–294.

Hill WG, Weir BS. 1988. Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol. 33(1):54–78.

International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437(7063):1299.

Jensen-Seaman MI, et al. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Res. 14(4):528–538.

Jones RF. 1970. Physiological and biochemical aspects of growth and gametogenesis in *Chlamydomonas reinhardtii*. Ann N Y Acad Sci. 175(2):648–659.

Kaback DB, Guacci V, Barber D, Mahon JW. 1992. Chromosome size-dependent control of meiotic recombination. Science 256(5054):228–232.

Kathir P, et al. 2003. Molecular map of the *Chlamydomonas reinhardtii* nuclear genome. Eukaryot Cell. 2(2):362–379.

Kaur T, Rockman MV. 2014. Crossover heterogeneity in the absence of hotspots in *Caenorhabditis elegans*. Genetics 196(1):137–148.

Kawakami T, Backström N, et al. 2014. Estimation of linkage disequilibrium and interspecific gene flow in *Ficedula* flycatchers by a newly developed 50k single-nucleotide polymorphism array. Mol Ecol Resour. 14(6):1248–1260.

Kawakami T, Smeds L, et al. 2014. A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. Mol Ecol. 23(16):4035–4058.

Kim S, et al. 2007. Recombination and linkage disequilibrium in *Arabidopsis thaliana*. Nat Genet. 39(9):1151–1155.

Kimura M. 1962. On the probability of fixation of mutant genes in a population. Genetics 47:713–719.

Kliman RM, Hey J. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol Biol Evol. 10(6):1239–1258.

Lam I, Keeney S. 2015. Nonparadoxical evolutionary stability of the recombination initiation landscape in yeast. Science 350(6263):932–937.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760.

Li X, Li L, Yan J. 2015. Dissecting meiotic recombination based on tetrad analysis by single-microspore sequencing in maize. Nat Commun. 6:6648.

Lichten M. 2008. Meiotic chromatin: the substrate for recombination initiation. In: Egel R, Lankenau D-H, editors. Recombination and meiosis: models, means, and evolution. Heidelberg (Germany): Springer Berlin. p. 165–193.

Liu H, et al. 2018. Tetrad analysis in plants and fungi finds large differences in gene conversion rates but no GC bias. Nat Ecol Evol. 2(1):164–173.

Manzano-Winkler B, McGaugh SE, Noor MAF. 2013. How hot are *Drosophila* hotspots? Examining recombination rate variation and associations with nucleotide diversity, divergence, and maternal age in *Drosophila pseudoobscura*. PLoS One 8(8):e71582–e71586.

Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favourable gene. Genet Res. 23(1):23–35.

McVean GAT, et al. 2004. The fine-scale structure of recombination rate variation in the human genome. Science 304(5670):581–584.

Merchant SS, et al. 2007. The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318(5848): 245–250.

Morris GJ, Coulson G, Clarke A. 1979. The cryopreservation of *Chlamydomonas*. Cryobiology 16(4):401–410.

Nachman MW. 2002. Variation in recombination rate across the genome: evidence and implications. Curr Opin Genet Dev. 12(6):657–663.

Ness RW, Kraemer SA, Colegrave N, Keightley PD. 2016. Direct estimate of the spontaneous mutation rate uncovers the effects of drift and recombination in the *Chlamydomonas reinhardtii* plastid genome. Mol Biol Evol. 33(3):800–808.

Ness RW, Morgan AD, Colegrave N, Keightley PD. 2012. Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*. Genetics 192(4):1447–1454.

Ness RW, Morgan AD, Vasanthakrishnan RB, Colegrave N, Keightley PD. 2015. Extensive de novo mutation rate variation between individuals and across the genome of *Chlamydomonas reinhardtii*. Genome Res. 25(11):1739–1749.

Page SL, Hawley RS. 2003. Chromosome choreography: the meiotic ballet. Science 301(5634):785–789.

Pál C, Hurst LD. 2003. Evidence for co-evolution of gene order and recombination rate. Nat Genet. 33(3):392–395.

Pan J, et al. 2011. A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. Cell 144(5):719–731.

Pedersen BS, Quinlan AR. 2017. cyvcf2: fast, flexible variant analysis with python. Bioinformatics 33(12):1867–1869.

Poyatos JF, Hurst LD. 2007. The determinants of gene order conservation in yeasts. Genome Biol. 8(11):R233.

R Core Team. 2017. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available from: http://www.R-project.org/.

Rockman MV, Kruglyak L. 2009. Recombinational landscape and population genomics of *Caenorhabditis elegans*. PLoS Genet. 5(3):e1000419.

Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L. 2006. Population genomic analysis of outcrossing and recombination in yeast. Nat Genet. 38(9):1077–1081.

Saintenac C, et al. 2009. Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). Genetics 181(2):393–403.

Schurko AM, Neiman M, Logsdon JM Jr. 2009. Signs of sex: what we know and how we know it. Trends Ecol Evol (Amst). 24(4): 208–217.

Singhal S, et al. 2015. Stable recombination hotspots in birds. Science 350(6263):928–932.

Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philos Trans R Soc Lond B Biol Sci. 372(1736):pii:20160455.

Suzuki L, Johnson CH. 2002. Photoperiodic control of germination in the unicell *Chlamydomonas*. Naturwissenschaften 89(5):214–220.

Tsai IJ, Bensasson D, Burt A, Koufopanou V. 2008. Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. Proc Natl Acad Sci U S A. 105(12):4957–4962.

Tsai IJ, Burt A, Koufopanou V. 2010. Conservation of recombination hotspots in yeast. Proc Natl Acad Sci U S A. 107(17):7847–7852.

Weedall GD, Hall N. 2015. Sexual reproduction and genetic exchange in parasitic protists. Parasitology 142(Suppl 1):S120–S127.

Weisse T. 2008. Distribution and diversity of aquatic protists: an evolutionary and ecological perspective. Biodivers Conserv. 17(2): 243–259.

Wijnker E, et al. 2013. The genomic landscape of meiotic crossovers and gene conversions in *Arabidopsis thaliana*. eLife 2(2):e01426.

Wright SI, Ness RW, Foxe JP, Barrett SCH. 2008. Genomic consequences of outcrossing and selfing in plants. Int J Plant Sci. 169(1):105–118.

Wu TC, Lichten M. 1994. Meiosis-induced double-strand break sites determined by yeast chromatin structure. Science 263(5146): 515–518.

Associate editor: Gwenael Piganeau