UNIVERSITY OF
BATH

University of Bath

Tetrad analysis in plants and fungi finds large differences in gene conversion rates but no GC bias

Haoxuan Liu[1,4], Ju Huang[1,4], Xiaoguang Sun[1,4], Jing Li[2], Yingwen Hu[1], Luyao Yu[1], Gianni Liti[2], DachengTian[1], Laurence D. Hurst[3,*], Sihai Yang[1,*]

[1] State Key Laboratory of Pharmaceutical Biotechnology, School of Life Sciences, Nanjing University, Nanjing 210023, China; [2] Institute for Research on Cancer and Aging, Nice (IRCAN), CNRS UMR7284, INSERM U1081, University of Nice Sophia Antipolis, Nice, France; [3] The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Bath, U.K, BA2 7AY

[4] These authors contributed equally to this work.

[*] Correspondence should be addressed to S.Y. (sihaiyang@nju.edu.cn) or L.D.H. (bssldh@bath.ac.uk)

ABSTRACT

GC-favoring gene conversion enables fixation of deleterious alleles, disturbs tests of natural selection and potentially explains both the evolution of recombination as well as the commonly reported intra-genomic correlation between G+C content and recombination rate. In addition, gene conversion disturbs linkage disequilibrium, potentially affecting the ability to detect causative variants. However, the importance and generality of these effects is unresolved, not simply because direct analyses are technically challenging but also because prior within- and between-species discrepant results can be hard to appraise owing to methodological differences. Here we report results of methodologically uniform whole-genome sequencing of all tetrad products in *Saccharomyces*, *Neurospora*, *Chlamydomonas* and *Arabidopsis*. The proportion of polymorphic markers converted varies over three orders of magnitude between species (from 2% of markers converted in yeast to only ~0.005% in the two plants) with at least 87.5% of the variance in per tetrad conversion rates being between-species. This is largely owing to differences in recombination rate and median tract length. Despite three of the species showing a positive GC-recombination correlation, there is no significant net AT->GC conversion bias in any, despite relatively high resolution in the two taxa (*Saccharomyces* and *Neurospora*) with relatively common gene conversion. The absence of a GC bias means: 1) that there should be no presumption that gene conversion is GC biased, nor 2) that a GC-recombination correlation necessarily implies biased gene conversion, 3) that K_a/K_s tests should be unaffected in these species and 4) it is unlikely that gene conversion explains the evolution of recombination.

INTRODUCTION

Meiotic double strand breaks (DSB) are resolved as crossover (CO) or non-crossover (NCO) events[1]. While the population genetical impact of allelic re-assortment owing to crossing over is well studied[2–4], both CO and NCO are associated with gene conversions, the non-reciprocal exchange of alleles that typically lead to 3:1 segregation (when first identified in yeast[5] and *Neurospora*[6,7], gene conversion was considered 'aberrant recombination'). With gene conversion rates relatively poorly resolved compared, for example, to crossing over rates, here we provide a methodologically uniform estimation of gene conversion parameters for diverse species. We consider four species: a unicellular fungus *Saccharomyces cerevisiae*, a multicellular fungus *Neurospora crassa*, a unicellular plant *Chlamydomonas reinhardtii*, and a multicellular plant *Arabidopsis thaliana*. The choice of these four species is in part motivated by the fact that they are amenable to tetrad analysis, a gold standard for gene conversion analysis. We provide parent-offspring whole genome resequencing to identify recombination events at high resolution and thus estimate rate and bias.

Gene conversion can have significant impact on gene and genome evolution. It is, for example, partially responsible for the fast evolution of major histocompatibility complex genes in animals[8] and resistance genes in plants[9]. Additionally, it reshapes patterns of local linkage disequilibrium (LD). In humans, for example, loci separated by 124bp on average would be expected to be in complete LD[10]. However, many sites show only partial LD suggesting that gene conversion has increased the apparent rate of recombination[10]. In disturbing linkage disequilibrium[11], gene conversion thus potentially affects the ability to detect causative variants[11] and signatures of selection.

Possibly more importantly gene conversion may be biased in favor of GC over AT alleles[12–16], evidenced by a non 50:50 mean rate of resolution of AT:GC mismatches at meiotic heteroduplexes in favour of the G/C allele. A neglected hypothesis[17] observes that GC-biased gene conversion (gBGC) may explain the evolution of recombination for if, as Bengtsson[17] noted, gene conversion is biased in the opposite direction to mutation

3

bias, it will, on the average, have the effect of correcting mutations. One of the few phylogenetic universals appears to be that mutation is more commonly GC->AT than AT->GC[18,19]. Is then gene conversion universally biased in the opposite direction as Bengtsson might predict? Such a bias is seen in yeast where whole genome data suggest a weak bias[20] (50.63% AT->GC), while meta-analysis[16] of HIS4 hotspot data reports a stronger distortion (55.2%). In humans (68% AT->GC)[15,21,22] and flycatchers (59% AT->GC)[23] a strong bias is also seen. Population level analysis suggest that gBGC is present in honey bee[24] but possibly absent in *Drosophila*[25]. Given that the underlying assumptions of Bengtsson's model seem to accord with the current consensus, both regarding direction of mutation and the opposing direction of conversion bias, this rather neglected theory deserves reappraisal. That results differ even in within-species analysis (e.g. HIS4 data[16] compared with whole genome analysis in yeast[20]), underlines our motivation to perform methodologically uniform analysis across taxa.

If Bengtsson's model is correct, it could in principle explain both crossover and non-crossover events. Whether CO and NCO events are both associated with gBGC is poorly resolved. Direct evidence from yeast suggests that gBGC is specific to COs and not to NCOs[15]. Conversely, evidence from human pedigree analysis and sperm typing suggests significant levels of GC-bias associated with NCO-GC[21,26] and complex CO-associated gene conversions[22], while others report no gBGC at human COs hotspots[26].

Not all mutations fixed by gBGC need be wild type alleles and indeed, as its population genetics resembles meiotic drive, it can power the fixation of deleterious alleles[13,27]. Consequentially it can mislead tests of natural selection[13,27], such as the K_a/K_s test, as it can increase the probability of fixation rate of (deleterious) non-synonymous mutations and decrease the probability for synonymous mutations. This can occur if the former are more commonly AT->GC mutations and the later more commonly GC->AT mutations.

gBGC also potentially explains taxonomically widespread intragenomic correlations

between G+C content and recombination rate[12,14,28]. Indeed, at least in birds and mammals, gBGC is thought to explain the existence of isochores (domains of high and low GC content)[12,28,29]. Despite that fact that a correlation between local GC content and the local recombination rate is taken by some as evidence for gBGC, it is also consistent with the causal arrow running in the opposite direction i.e. GC rich segments of the genome being more prone to induction of recombination events[30]. One way to resolve cause and effect is to examine motifs that initiate recombination across multiple taxa to determine whether they are GC-rich and whether SNPs promoting or reducing recombination rates are GC increasing and decreasing respectively (see e.g. [26,31]). An alternative is to ask whether species with a positive intragenomic GC-recombination correlation show evidence for gBGC while those without such a correlation show no gBGC. Such evidence would support the gBGC model suggesting that recombination causes the GC content. Three of our species show a positive correlation and so are predicted to show gBGC.

In addition to estimation of bias, to understand the potential impact on tests of selection and linkage disequilibrium, we need also to understand absolute rates which might be highly heterogeneous. In yeast, for example, there are 46 NCO gene conversion events per meiosis[20], while only 1~2 NCO gene conversions per meiosis were directly detected in *Arabidopsis* and human[21,22,32]. However, due to the low resolution and high stringency in these studies the actual number of NCOs may be more like 50~200 per meiosis[21,32]. Indeed, high resolution evidence from several loci suggest that the number of DSBs per meiosis is 150~400 events per meiosis in mammals[33–37]. Intra-specific variation in rates needs to be addressed, not least because of an order of magnitude variation in recombination rates between oocytes from the same human female[35,38–40]. Difference between taxa might be expected as there are differences in the recombination rate. However, a second big unknown is whether tract lengths differ. While the median tract length of gene conversion is ~2 kb in yeast[20], it could be 50~500 bp in mouse[41] and human[42] and as short as 20 bp in *Arabidopsis*[32].

RESULTS

1. Fine-scale mapping of crossover and non-crossover events

We made ten different crosses in four species, collected their meiotic products, and whole-genome sequenced 505 samples (17 parental strains, 452 haploid and 36 diploid members from 122 tetrads). Between 1 and 4 different crosses were made in each species (Table 1): two in yeast, S1 and S2; four in *Neurospora*, N1-4, three in *Chlamydomonas*, C1-3; one in *Arabidopsis*, cross A. In *Arabidopsis*, single meiotic tetrad pollen from F1 was picked and planted on the flower of Columbia ecotype (Col) to make backcrosses. Siliques, which contain four seeds, are tetrads. Crosses S1 and A replicate prior efforts[20,32]. While it would be desirable to consider tetrad analysis in different strains in *Arabidopsis*, tetrads can only be separated in *qrt1* mutants, which are only available in Ler and Col ecotypes. Of the 122 tetrads, 29 are from yeast, 57 from *Neurospora*, 27 from *Chlamydomonas* and 9 from *Arabidopsis*. The parental strains in cross N4 are meiotic products from a single meiosis in cross N3. Whole genome re-sequencing was carried out on the Illumina HiSeq 4000 platform. Each sample was sequenced with an average depth of 40× and 96% coverage (Supplementary Table 1 and Fig. 1).

While a heteroduplex may form in the absence of marker polymorphism, for gene conversion to occur polymorphic markers are required. In our ten crosses marker density ranges from 0.04% to 1.53% (Table 1 and Supplementary Fig. 1).

By comparing a tetrad's genotype with the parental genotype, we can identify three types of recombination events: crossovers (CO), crossover-associated gene conversions (CO-GC) and non-crossover gene conversions (NCO-GC). In short, reciprocal changes of genotypes are COs while non-reciprocal changes are gene conversions (Methods, Supplementary Fig. 2). Recombination events are generated through two major pathways, Double Strand Break Repair (DSBR) and Synthesis Dependent Strand Annealing (SDSA)[43] (Fig. 2). The former predominately leads to CO, while SDSA leads to NCO[44]. While the majority of events we identified are classical COs or NCOs (Type I and type IV in Fig. 2), we also identified two types of non-classical events. Non-classical events

comprise 0.5%-12.3% of recombination events (Type II and type III in Fig. 2; Supplementary Discussion). Recombination events vary over two orders of magnitude between our species but are largely consistent within species (Supplementary Discussion).

2. Large between-species variation in rates of gene conversion

Compared with what was known about the rate of CO, the rate of gene conversion is poorly described but possibly heterogeneous[21,22,32]. We find evidence for three orders of magnitude variation between species. The pattern of variation is very similar between CO-GC and NCO-GC (Supplementary Fig. 3 and Table 1). Yeast crosses exhibit the highest rate of gene conversion, averaging 1.9%±0.47% (1.1%±0.30% for CO-GC and 0.8%±0.25% for NCO-GC) of all markers converted per tetrad per meiosis. No significant difference was identified among crosses S1, S2, and the previous yeast study[20] (ANOVA, p=0.18). While this rate is almost two orders of magnitude lower in *Neurospora* (ANOVA, $p<10^{-7}$), averaging 0.031%±0.018% (0.018%±0.015% for CO-GC and 0.013%±0.011% for NCO-GC) per tetrad per meiosis, again, no significant differences were identified within the species among N2~N4 (ANOVA, $p=0.41$).

The rate of gene conversion is even lower in the two plant species, *Chlamydomonas*, and *Arabidopsis*, significantly so compared with *Neurospora* (ANOVA, $p<10^{-3}$). This rate is 0.0043%±0.0014% (0.0040%±0.0014% for CO-GC and 0.0003%±0.0002% for NCO-GC) for *Chlamydomonas* and 0.0068%±0.0034% (0.0056%±0.0036% for CO-GC and 0.0012%±0.0011% for NCO-GC) for *Arabidopsis*, in which no significant differences were identified between two *Chlamydomonas* crosses (C2 and C3, t-test, $p=0.12$). Between species, this rate is significantly lower in *Chlamydomonas* than in *Arabidopsis* (ANOVA, $p=0.011$, sensitive to multitest correction). These results suggest that disruption of linkage disequilibrium (LD) is less likely in the two plants.

We estimate that at least 87.5% of the variation is between-species variation, thus at most 12.5% of the overall variation is within-species. Analysis of the *Neurospora* strains indicates that only 12.3% of within-species variation is between-strain variation. Overall,

7

a net 1.1-1.5% of all tetrad variation is explained by within-species between-strain variation, 7.6-11% is explained by within-strain within-species variation and 87.5-91.3% is explained by between-species variation (Supplementary Discussion).

It is worth noting that, in all these four species, the rate of CO-GC is higher than NCO-GC, and this difference is lesser in yeast and *Neurospora*, larger in *Chlamydomonas* and *Arabidopsis*. The rate of CO-GC is almost 5 times that of NCO-GC in *Arabidopsis* and as much as 13 times in *Chlamydomonas*. In other words, the total rate of gene conversion is largely owing to CO-GC in *Chlamydomonas* and *Arabidopsis*. Although the rate of CO and the rate of NCO (percentage of markers converted) vary a lot among these four species, even with only four data points we detect a tendency to positively correlate (Pearson's r, Log(NCO markers converted) v Log(cM/Mb) =0.92; p=0.07).

3. Between-species variation in conversion rates are explained by recombination rates and tract length

The between-species differences in CO rates reflect in part the fact that each chromosome needs at least one chiasmata. Indeed, the CO rates are more uniform when expressed as CO events per chromosome with about an order of magnitude difference in cM/Mb per chromosome: from ~5.3 COs/Chr in yeast (=22 cM/Mb per chromosome, N=16) to ~2.3 COs/Chr in *Neurospora* (=2.8 cM/Mb/Chr, N=7), to ~2.2 COs/Chr in *Arabidopsis* (=0.9 cM/Mb/Chr, N=5) to ~1.4 COs/Chr in *Chlamydomonas* (=0.7 cM/Mb/Chr, N=17).

Chromosome number and intra-chromosomal variation in the CO rate is, however, only part of the explanation for the between-species differences in the percentage of markers converted. Tract lengths also differ by orders of magnitude. Employing the midpoint method[20], we found that 63% of both NCO-GCs and CO-GCs are < 2 kb, 92% are < 5 kb, and 99% of them < 10 kb. The longest gene conversion (in yeast) spanned ~18 kb and converted 74 markers. These data hide within and between genome differences (Fig. 3a and 3b). In yeast, CO-GCs are significantly longer than NCO-GC events (Wilcoxon

test, p=0.016: median size for NCO-GC = 1681 bp, for CO-GC = 1841 bp). The tract length of both NCO-GC and CO-GC in *Neurospora* are much shorter than that in yeast. *Neurospora*'s NCO-GC's median size is 950 bp, while the CO-GC's median size is only 284 bp and, unlike what we observed in yeast, the tract length of NCO-GCs is longer than CO-GCs (Wilcoxon test, p<2.2e-16). In *Chlamydomonas*, the tract length for NCO-GCs is even shorter, with a median size of only 73 bp, while the median tract length for CO-GC is significantly longer at 364 bp (Wilcoxon test, p=2.6e-12). The median sizes of NCO-GCs and CO-GCs in *Arabidopsis* are 627 bp and 1067 bp, respectively (Wilcoxon test, p=0.067).

Above we have employed the midpoint method. This is, however, sensitive to marker density, being less accurate in low marker density crosses (such as cross A). To overcome this problem we also used simulations[32] to estimate the average tract length. We randomly simulated 10,000 gene conversion events at various lengths in each cross, and calculated the average number of markers converted at each length. The closest value to the observed number of markers converted is chosen as the average tract length (Fig. 3c). The estimation from simulation is close to the midpoint method estimate in higher marker density crosses (yeast, *Neurospora,* and *Chlamydomonas*), but much shorter in *Arabidopsis*. Here the simulated tract length is 80 bp for NCO-GC and 700 bp for COGC (Fig. 3d), rather different from the 627bp and 1067bp derived from the midpoint method.

Despite the discrepancy in low marker density crosses, a combination of low recombination rate and small tract size is needed to explain why the percentage of markers converted in *Chlamydomonas* is an order of magnitude lower than in *Neurospora*. *Chlamydomonas* and *Arabidopsis* have approximately the same net gene conversion rate but for different reasons: *Chlamydomonas* has double the recombination events per bp but much smaller tract sizes. Yeast has both a very high recombination rate and long tracts, explaining why it converts ~2% of markers per meiosis. Marker density variation doesn't explain between-taxa differences in these crosses (Supplementary Discussion).

4. No evidence for universal GC-biased gene conversion

In gene conversion events, if the donor converting allele and acceptor allele are randomly chosen, equal numbers of AT -> GC and GC -> AT conversions are expected. Gene conversions can, however, favor AT -> GC conversions[12,20,]. Correlations between recombination and GC content are assumed to reflect this process[14,26], although the causal arrow could run in the opposite direction[30].

In our study, no net conversion bias was identified in 7 of 8 crosses (Table 2, N.B. here we exclude N1 and C1 as events are too rare to be informative). The one exception is an AT-bias in S2 (50.97% GC -> AT, binomial test, p=0.024, Table 2), not significant after multitest correction. Combining S1 and S2 we recover a 50.7% bias in favor of GC->AT (binomial test, p=0.013, not significant after multitest correction), i.e. in the same direction as mutation bias. This contrasts with the prior yeast analysis[15,20] which identified an equal magnitude AT->GC bias (50.63%). If we combine our and the prior data we see no significant overall bias (53262 AT->GC and 52691 GC->AT; binomial test, p=0.08). With a sample size of this dimension we could have detected a bias of the order of 50.305% (without multitest correction). The overall bias in our study and the prior one are significantly different (chi squared test, df=1, p=0.005).

The discrepancy between our data and prior data is magnified when we compare NCO-GC and CO-GC events. When NCO-GC and CO-GC are treated separately, a weak and significant AT bias is identified in CO-GCs of our cross S1 and in NCO-GCs of cross S2 (none of which survive multitest correction). This contrasts with the prior report of a 50.70% AT-> GC bias observed exclusively at CO-GC events in the prior S1 cross (numbers are different from ours: chi squared test, df=1, p=0.0002). We find no evidence that the discrepancy between our data and prior data is owing to technical artifacts in our data. Indeed, via Sanger resequencing of a sample of gene conversion events we verify 100% (40/40) of them and fail to detect any new events (Supplementary Information).

In our other fungus, *Neurospora*, no bias was identified in any cross, nor when all events were considered *en masse* (AT->GC bias 50.49%; binomial test, p=0.49). With a net sample size of 5150 GC<->AT events we could have detected a bias greater than 51.37% (without multitest correction).

While these results fail to support the claim[15,20] of an AT->GC conversion bias, a repeatable GC bias was identified in NCO-GCs in *Chlamydomonas* (64.91% AT -> GC in C2, 75.86% AT -> GC in C3, Table 2, net effect bias = 68.6%, p<0.01), this being of a magnitude also inferred in humans[21]. However, in the algae the great majority of gene conversion events are CO-GC events, which are apparently unbiased, and overall we see no significant net bias (AT->GC bias 52.5%; binomial test, p=0.10). With 1097 GC<->AT conversion events we could have resolved a 53.01% net bias (prior to multitest correction). Similarly, no significant bias was identified in *Arabidopsis* combining our and prior data (AT->GC bias 53.3%; binomial test, p=0.39), although by necessity, given the low rates, our resolution is limited (limit to detection with N=197, 57.11%).

5. Three species show a positive GC-recombination rate intragenomic correlation

A positive correlation between GC-content and recombination rate has been reported in many species[45-49], and a prevailing explanation for this correlation is GC biased gene conversion[12]. Given this lack of evidence for GC biased gene conversion do we then not see a strong positive correlation between local GC and recombination rate? At the 10kb window scale, we observe both CO and NCO (and CO+NCO) to be more common in GC rich domains in yeast, *Neurospora*, and *Chlamydomonas*, while a negative correlation[32] is seen in *Arabidopsis* (not significant for NCO) (Table 3). These correlations are robust to larger window sizes (Table 3) and to control for covariates (CO/NCO rate, heterozygosity and gene density (Supplementary Tables 2 and 3).

As a check that our CO measures have population genetical relevance, we also test the

hypothesis that the CO correlates with high diversity[50] (owing to reduced selective non-independence between loci). Heterozygosity is indeed positively correlated with CO (Supplementary Table 2) in all species and is robust to control for gene density and other covariates (Supplementary Table 3). Heterozygosity is sometimes predicted by the NCO rate before and after covariate control (Supplementary Tables 2 and 3). This may reflect an ascertainment bias as local heterozygosity is needed to detect NCO events.

We conclude from the above that our direct measure of local recombination rates has relevance to predicting longer-term evolutionary trends. Why then do the species appear to differ in the strength of the GC – recombination correlation? Of the three species with a positive correlation, the correlation is weakest in *Chlamydomonas* and strongest in yeast. This might reflect the possibility that gene conversion really is weakly GC biased and that a species with much recombination (i.e. yeast) thus has a stronger correlation. However, a high recombination rate also enables accuracy of estimation. If we randomly sample to reduce the number of CO events in yeast to that seen in *Chlamydomonas*, with both species forced to have the same number of 10kb bins, in 69% of randomizations the GC-CO correlation is stronger in *Chlamydomonas*. This suggests that the dominant cause of differences in the strength of the correlations is sampling noise.

An assumption of the above tests is that in each species there is some repeatability in where recombination events occur as such repeatability would reinforce troughs and peaks. We found evidence for such repeatability (Supplementary Discussion).

DISCUSSION

Lack of gBGC renders Bengtsson's theory implausible

Given prior large-scale analyses in yeast[20], and meta-analyses[16] of small-scale experiments, we are surprised that we couldn't detect any significant net deviations (even in yeast with N> 100,000; Table 2). This suggests that Bengtsson's theory[17], that the main function of recombination is the purging of deleterious new mutations via gene conversion, is

unlikely to hold, at least in the taxa that we have examined. Moreover, that in *Chlamydomonas* there appears to be a difference between unbiased CO-GC and biased NCO-GC events, argues strongly that the major utility of crossover events is not to enable the direct removal of GC->AT mutational events. Were this the case, then why is CO-GC not biased? However, at least in this species, one could argue that the function of NCO-GC events is the purging of GC->AT mutations. But if so important, why then is the rate of NCO events so very low? If our observations in yeast are robust, the possibility of gene conversion bias in the same direction as the mutation bias (GC->AT) would render Bengtsson's model highly implausible. In this instance, gene conversion appears to be favoring new (deleterious) mutations over the wild type.

No single model can explain the GC-recombination correlation

Three models can in principle account for an intragenomic correlation between local GC content and the local recombination rate. First, recombination associated with gBGC, causes a GC biased fixation bias in domains of high recombination, while in low recombination domains the GC->AT mutation bias is the dominant force[12,28,29]. Second, if selection favors increased GC-content, more efficient selection in high recombination domains could lead to a GC-recombination correlation[51]. Third, GC rich domains might enable initiation of recombination events[30]. It has, for example, been proposed that GC-rich regions may be more likely to be targeted by recombination machinery or both GC-content and recombination covaries with a third factor[46,52]. We argue that, presently, while gBGC is a parsimonious explanation for isochores in mammals and birds, there is no single model that appears to be compatible with the accumulated data as an explanation for the GC-recombination correlation in our taxa (a correlation that is positive in all species other than *Arabidopsis*).

While in mammals the gBGC model is parsimonious as an AT->GC fixation bias[53] and gBGC[21,22] are both witnessed, for the taxa we are examining the case is much less clear. Naturally, the low sample size in the two plant species doesn't permit us to make a strong assertion of a lack of gBGC. However, this is owing to the very low rate of gene

conversion and hence one must also question the relevance of any bias to their genome evolution in the face of a consistent and opposing mutation bias. The lack of both an obvious gBGC in yeast and the lack of a GC fixation bias in domains of high recombination[30] both also argue against the gBGC model for this species.

We cannot fully exclude the possibility that, if extended over a long enough time period, even a tiny bias (<0.3%) could result in a GC-recombination correlation, especially in a species that is outbred and regularly sexual. Such a model could explain why a highly recombining species, such as yeast, might have the strongest correlation (although sexual reproduction and outcrossing might both be rare events[54]). However, this is not strong evidence, as the variation in GC-recombination strength between species appears in part to be owing to noise in estimation when the recombination rate is low. Our evidence accords with recent analysis in bacteria[55] that concluded that between-species variation in GC content could not be attributed to gBGC, despite earlier claims[56].

Both the selection and gBGC models are potentially consistent with a negative correlation in *Arabidopsis* as it is a near obligatory selfer, hence selection (or gene conversion) will be less efficient[57]. The model correctly predicts the correlation between local diversity and local recombination rate seen in all species. However, this model would also predict an excess of AT->GC fixation events associated with domains of high recombination, which, at least in yeast, isn't observed[30]. The argument that GC causes high recombination fails to explain why *Arabidopsis* is the exception, as here we see a negative GC-recombination correlation. This model is, however, consistent with the lack of an AT->GC fixation bias correlating with recombination in yeast[30]. Confirming fixation biases and their association (or lack thereof) with recombination will be an important next step to testing the three models.

With no model appearing especially parsimonious at present for our taxa, we caution against assuming gene conversion is universally AT->GC biased and in the interpretation of the commonly observed[14] GC-recombination correlation as evidence for gBGC.

Why is our data different from prior genome scale data?

Our estimates of rates of gene conversion accord well with prior genome-scale tetrad based estimates in yeast and *Arabidopsis*. However, our estimate of the AT<->GC bias in yeast is different to that seen in the prior analysis[15] (Table 2). In deriving a net lack of bias in yeast, we presumed that the prior estimates were correct, as were ours, despite the two having significantly different biases. If we presume ours alone is accurate then the conclusion of a lack of GC bias is all the more robust as, if anything, we see a GC->AT conversion bias.

How then to explain such a discrepancy? Strain differences appear not to be the explanation. There may be subtle but important method differences between our and the prior approach. One possibility is that any bias is subtly dependent on growth conditions. It is known that aspects of meiosis in yeast are susceptible to external environment[58]. A potentially important difference is that while we employ whole genome sequencing, the prior report employed arrays to detect conversion events. This use of array technology is worth further scrutiny, as it is known to be associated with missing values (e.g. owing to dye bias). By contrast, from Sanger sequencing we have confirmed 100% of a sample of our observations and additionally find no evidence for missing events. Moreover, when comparing our data with a further prior study, which also employed whole genome sequencing on the same hybrid strain[59], we find very good reproducibility as regards markers identified and on the numbers of events (Supplementary Discussion). Likewise, a GC bias as regards coverage of NGS data appears not to be important (Supplementary Discussion). Resolving the causes of such a subtle discrepancy will take a large body of further analysis, but is of importance given the potentially far reaching implications of the reversal of the direction of bias.

Methods

The four species in this study are all easy to incubate and cross under laboratory conditions, and their meiotic products (tetrad or ascospore) can be dissected and

genotyped individually. Three of them are haploid, which makes the genotyping more reliable. They also possess extensive nucleotide diversity among different strains, enabling detection of gene conversions. Analysis of yeast (cross S1) and *Arabidopsis* (Ler × Col) enables direct comparison with prior analysis.

1. Source of samples

Yeast strains S96, YJM789, and diploid hybrid strain S96/YJM789 were provided by Lars Steinmetz from European Molecular Biology Laboratory, Heidelberg, Germany; strain YPS128 and diploid hybrid strain YPS128/YJM789 were provided by Gianni Liti from Institute for Research on Cancer and Aging, Nice, University of Nice Sophia Antipolis, France. *Neurospora* strains FGSC2489 and FGSC4200 were provided by Chaoguang Tian from Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, China; strain FGSC2225 was provided by Shaojie Li from State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, China; strains FGSC3246 and FGSC1363 were purchased from Fungal Genetics Stock Center (www.fgsc.net). *Chlamydomonas* strains CC124, CC1010, CC2935, CC2936, and CC408 were purchased from Chlamydomonas Resource Center, University of Minnesota, U.S. (www.chlamycollection.org). *Arabidopsis thaliana* mutant strains Col (*qrt1*) and Ler (*qrt1*) were purchased from Arabidopsis Biological Resource Center (ABRC, abrc.osu.edu).

2. Crossing and tetrad dissection

Yeast hybrid strains S96/YJM789 and YSP128/YJM789 were induced by transferring overnight cultures from liquid YEPD to 2% potassium acetate to sporulation[60]. Then each tetrad was dissected under a dissection microscope. Crosses in *Neurospora* were made based on the protocol provided on Fungal Genetics Stock Center (www.fgsc.net). In short: i) the *Neurospora* strains with opposite mating types were planted on two ends of a crossing plate, and kept at 25 degrees centigrade and in complete darkness for 3 weeks for crossing and growing ascospores; ii) each ascospore was separated under a microscope and stored on a storage plate for at least a week; iii) the stored ascospores were heat shocked for 30 min at 65 degrees centigrade; iv) each ascospore was dissected

into 8 individual spores under a microscope. Mating and tetrad dissection in *Chlamydomonas* were carried out based on David Stern's protocol[61] (Fig. 1). The crossing and tetrad dissection in *Arabidopsis* is similar to a previous study[32]. First Col (*qrt1*) is crossed with Ler (*qrt1*), then a single meiotic tetrad pollen from F1 was picked and planted on the flower of Col to make backcrosses. Siliques, which contain four seeds, were collected. Each seed was planted individually for the subsequent genotyping.

3. DNA extraction and genome sequencing

Each dissected spore (or seed) was cultivated individually until at least 3 micrograms of DNA could be harvested. The DNA of each dissected spore (or seed) was extracted and sequenced individually. The DNA of yeast, *Chlamydomonas*, and *Arabidopsis*, was extracted using phenol/chloroform method, while the DNA of *Neurospora* was extracted using phenol/chloroform/isoamyl alcohol method.

Whole genome re-sequencing was carried out at BGI-Shenzhen and Novogene (www.novogene.com) using the same procedure: Paired-end sequencing libraries with insert size of 350 bp were constructed for each sample, then 2 × 150 bp paired-end reads were generated on Illumina HiSeq 4000 platform. In total, 505 samples were sequenced, including 119 yeast samples (3 parental strains and 29 tetrads), 233 *Neurospora* samples (5 parental strains and 57 tetrads), 113 *Chlamydomonas* samples (5 parental strains and 27 tetrads), and 40 *Arabidopsis* samples (4 hybrid F1 plants and 9 tetrads). The average depth and genome coverage for yeast samples are 72× and 99.32%, for *Neurospora* samples are 35× and 95.95%, for *Chlamydomonas* samples are 27× and 94.33%, and for *Arabidopsis* samples are 22× and 95.87% (Supplementary Table 1).

4. Genotyping and marker filtering

Saccharomyces cerevisiae reference genome (version R64) and *Neurospora crassa* reference genome (version 12) were downloaded from Ensembl (www.ensembl.org), *Chlamydomonas reinhardtii* reference genome (version 5.5) was downloaded from Joint Genome Institute (jgi.doe.gov), *Arabidopsis thaliana* reference genome (TAIR10) was downloaded from The

Arabidopsis Information Resource (www.arabidopsis.org/). For each sample, the Illumina reads were mapped onto the reference genome by Burrows-Wheeler Aligner (BWA)[62] then duplicates marking and realignment around indels were carried out by Genome Analysis Toolkit (GATK)[63] with variants called by GATK HaplotypeCaller.

The sequenced samples in yeast, *Neurospora*, and *Chlamydomonas* are all haploid. For each cross, the "homozygous" SNPs between parental strains with quality score ≥30 were first called as marker candidates, then for each tetrad were permitted if this candidate site meets the following criteria: i) called as "homozygous" in all four spores; ii) genotyped with high quality (≥30) in all four spores; iii) the genotypes of the four spores at the candidate site agree with their parental strains. This site is then used as a marker to identify recombination events.

The sequenced samples in *Arabidopsis* are diploid. First we analyzed the deeply sequenced Col (*qrt1*) and Ler (*qrt1*) from a previous study[32] and identified the SNPs that were homozygous within the parents but different between them, thus potential marker candidates. We then also deep sequenced four F1 plants in our study, and applied the following criteria in marker screening: i) each marker candidate must be heterozygous in these four F1 plants; ii) each marker should be covered with at least 10 reads in each sample and genotyped with high quality (≥30); iii) the genotypes of the four seeds from a tetrad should be either homozygous Col genotype or heterozygous Col/Ler genotype; iv) for heterozygous SNPs called in each sample, each allele should be supported by at least three reads. Marker candidates in TE (transposon element) regions and CNV (copy number variation) regions were excluded. Marker candidates that fit the criteria above were used as markers.

Any mitotic mutations arising in each sample would have a minimal effect in our analysis and should be filtered out. Mitotic mutations would be polymorphic within each sample of cells grown from an individual spore. For haploid organisms mutations arising in the very first mitotic division would be seen in 50% of reads (25% in diploid *Arabidopsis*),

mutations in the second division associated with 25% of reads (12.5% in *Arabidopsis*), *etc.*.

Three of four organisms we sequenced are haploid, so technically all variants in these samples should not be polymorphic and should be "homozygous". However, due to multi-copy regions or mapping errors or mitotic mutations, apparently "heterozygous" variants can be identified. In this study, all these "heterozygous" variants were removed. Whether these are mutations is hard to say, as it is not easy to discriminate true mitotic mutations from other types of errors. Furthermore, in each cross, we sequenced two parental stains and multiple tetrads. The genotypes of both parents are compared with each other and with multiple offspring spores in the identification of markers. We first identified the homozygous SNPs between two parental strains and only when the genotypes of all four products from a tetrad agree with parental genotypes was this SNP used as a marker. In *Arabidopsis*, only if a mutation takes place on the marker site, and changes the genotype from one parental allele to the other, would it emerge as a false gene conversion. Given a mutation rate of 7×10^{-9} per base per generation, a marker density of 0.26%, and a sample size of 36 (9 tetrads), the total number of such events roughly equals $7 \times 10^{-9} \times 1.2 \times 10^{8} \times 0.0026 \times 36 \times 1/3 = 0.0262$. Compared with the 170 SNPs being converted in *Arabidopsis*, the effect of mitotic mutations is thus miniscule.

5. Identification of recombination events

Three types of recombination event can be identified in this study: COs, CO-GCs, and NCO-GCs. For each tetrad, the genotype information of all four spores was used collectively to identify these events. Reciprocal changes of genotype in 2 of 4 spores are identified as COs. If the genotype switch point is the same in these 2 spores (as shown in Supplementary Fig. 2a), this is a simple CO event; if the genotype switch point is not the same in these 2 spores, as shown in Supplementary Fig. 2b, this is a CO event associated with a continuous CO-GC event, and in this case, the CO-GC events simply extend the genotype switch point of a CO event. Sometimes CO-GCs also emerge as a discontinuous conversion tract, as shown in Supplementary Fig. 2c, where a short non-reciprocal genotype change appears near a CO event (distance <10 kb). This is identified

as a CO event associated with a discontinuous CO-GC event.

Non-reciprocal changes of genotype that are not associated with CO events are identified as NCO-GCs. Most NCO-GC events are simple as shown in Supplementary Fig. 2d. However, we also found 2 types of non-classical recombination events. The first type is where a CO event and a NCO event occur on the same genomic position but in different spores (Supplementary Fig. 2e). The second type is where two short blocks with opposite genotype are identified at the same genomic position in 2 different spores (Supplementary Fig. 2f). Cases like this are identified as 2 NCO events. Compared with prior studies using the same material, but a different genotyping method[20,32], highly similar results (rates of CO and NCO in crosses S1 and A) were obtained.

6. Estimation of conversion tract length

Two methods, midpoint and simulation, were used to estimate the tract length of gene conversions. First we used the midpoint method from a previous yeast study[20]. On both ends of a gene conversion event, the midpoints between the end marker and nearest flanking non-converted marker are determined as the edge of the conversion tract and the distance between two edges is estimated as the length of the tract. However, this method is sensitive to marker density, which makes it less accurate at low marker density. To circumvent this, another simulation based method is employed[32]. We randomly simulated 10,000 gene conversion events at various lengths in each cross, and calculated the average number of markers converted at each length (Fig. 3c). Then the closest value in the simulation to the observed number of markers converted is chosen as the average tract length. For example, CO-GC events convert 4 markers per tract in *Arabidopsis*, and 4 markers per tract correspond to a tract length of 700 bp in the simulation (illustrated in Fig. 3c). In this way the average tract length for CO-GC is estimated to be 700 bp in *Arabidopsis*.

7. Event verification by PCR and Sanger sequencing

Some events in this study were also verified by PCR and Sanger sequencing. For CO

events, a single PCR must be designed to include markers on both ends of the CO breakpoint. PCR and Sanger sequencing must be performed on both parental strains and all 4 spores. Only if the PCR and Sanger sequencing is successful and agree with NGS results was an event deemed 'verified by PCR and Sanger sequencing'. For NCO-GC events, 1 or 2 overlapping PCRs must be designed to include converted markers and flanking non-converted markers on both ends of the conversion tract. PCR and Sanger sequencing must be performed on both parental strains and all 4 spores. Only if the PCR and Sanger sequencing are successful and agree with NGS results was an event deemed 'verified by PCR and Sanger sequencing'.

8. Data availability

The sequencing reads have been submitted to NCBI under submission number PRJNA373800 and PRJNA374752. The genotypes of each tetrad at all marker sites can be publicly accessed at http://gattaca.nju.edu.cn/pub_data.html.

References:

1. Baudat, F. & Massy, B. de. Regulating double-stranded DNA break repair towards crossover or non-crossover during mammalian meiosis. *Chromosome Res.* **15,** 565–577 (2007).

2. Muller, H. J. Some Genetic Aspects of Sex. *Am. Nat.* **66,** 118–138 (1932).

3. Otto, S. P. & Lenormand, T. Resolving the paradox of sex and recombination. *Nat. Rev. Genet.* **3,** 252–261 (2002).

4. Coop, G. & Przeworski, M. An evolutionary view of human recombination. *Nat. Rev. Genet.* **8,** 23–34 (2007).

5. Lindegren, C. C. Non-Mendelian Segregation in a Single Tetrad of Saccharomyces Ascribed to Gene Conversion. *Science* **121,** 605–607 (1955).

6. Mitchell, M. B. Aberrant recombination of pyridoxine mutants of neurospora*. *Proc. Natl. Acad. Sci. U. S. A.* **41,** 215–220 (1955).

7. Case, M. E. & Giles, N. H. Evidence from tetrad analysis for both normal and aberrant recombination between allelic mutants in Neurospora crassa*. *Proc. Natl. Acad. Sci. U. S. A.* **44,** 378–390 (1958).

8. Martinsohn, J. T., Sousa, A. B., Guethlein, L. A. & Howard, J. C. The gene conversion hypothesis of MHC evolution: a review. *Immunogenetics* **50,** 168–200 (1999).

9. Mondragon-Palomino, M. & Gaut, B. S. Gene Conversion and the Evolution of Three Leucine-Rich Repeat Gene Families in Arabidopsis thaliana. *Mol. Biol. Evol.* **22,** 2444–2456 (2005).

10. Ardlie, K. *et al.* Lower-Than-Expected Linkage Disequilibrium between Tightly Linked Markers in Humans Suggests a Role for Gene Conversion. *Am. J. Hum. Genet.* **69,** 582–589 (2001).

11. Wall, J. D. Close look at gene conversion hot spots. *Nat. Genet.* **36,** 114–115 (2004).

12. Duret, L. & Galtier, N. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* **10,** 285–311 (2009).

13. Galtier, N., Duret, L., Glémin, S. & Ranwez, V. GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* **25,** 1–5 (2009).

14. Pessia, E. *et al.* Evidence for Widespread GC-biased Gene Conversion in Eukaryotes. *Genome Biol. Evol.* **4,** 675–682 (2012).

15. Lesecque, Y., Mouchiroud, D. & Duret, L. GC-biased gene conversion in yeast is specifically associated with crossovers: molecular mechanisms and evolutionary

significance. *Mol. Biol. Evol.* **30,** 1409–1419 (2013).

16. Birdsell, J. A. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* **19,** 1181–1197 (2002).

17. Bengtsson, B. O. Biased conversion as the primary function of recombination. *Genet. Res.* **47,** 77–80 (1986).

18. Lynch, M. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci.* **107,** 961–968 (2010).

19. Hershberg, R. & Petrov, D. A. Evidence That Mutation Is Universally Biased towards AT in Bacteria. *PLOS Genet.* **6,** e1001115 (2010).

20. Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L. M. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454,** 479–485 (2008).

21. Williams, A. L. *et al.* Non-crossover gene conversions show strong GC bias and unexpected clustering in humans. *eLife* **4,** e04637 (2015).

22. Halldorsson, B. V. *et al.* The rate of meiotic gene conversion varies by sex and age. *Nat. Genet.* **48,** 1377–1384 (2016).

23. Smeds, L., Mugal, C. F., Qvarnström, A. & Ellegren, H. High-Resolution Mapping of Crossover and Non-crossover Recombination Events by Whole-Genome Re-sequencing of an Avian Pedigree. *PLOS Genet.* **12,** e1006044 (2016).

24. Wallberg, A., Glémin, S. & Webster, M. T. Extreme Recombination Frequencies Shape Genome Variation and Evolution in the Honeybee, Apis mellifera. *PLOS Genet.* **11,** e1005189 (2015).

25. Robinson, M. C., Stone, E. A. & Singh, N. D. Population Genomic Analysis Reveals No Evidence for GC-Biased Gene Conversion in Drosophila melanogaster. *Mol. Biol. Evol.* **31,** 425–433 (2014).

26. Odenthal-Hesse, L., Berg, I. L., Veselis, A., Jeffreys, A. J. & May, C. A. Transmission Distortion Affecting Human Noncrossover but Not Crossover Recombination: A Hidden Source of Meiotic Drive. *PLOS Genet.* **10,** e1004106 (2014).

27. Berglund, J., Pollard, K. S. & Webster, M. T. Hotspots of Biased Nucleotide Substitutions in Human Genes. *PLOS Biol.* **7,** e1000026 (2009).

28. Weber, C. C., Boussau, B., Romiguier, J., Jarvis, E. D. & Ellegren, H. Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome Biol.* **15,** 549 (2014).

29. Eyre-Walker, A. Recombination and Mammalian Genome Evolution. *Proc. R. Soc. Lond. B Biol. Sci.* **252,** 237–243 (1993).

30. Marsolier-Kergoat, M.-C. & Yeramian, E. GC Content and Recombination: Reassessing the Causal Effects for the Saccharomyces cerevisiae Genome. *Genetics* **183,** 31–38 (2009).

31. Jeffreys, A. J. & Neumann, R. Factors influencing recombination frequency and distribution in a human meiotic crossover hotspot. *Hum. Mol. Genet.* **14,** 2277–2287 (2005).

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

33. Cole, F. *et al.* Homeostatic control of recombination is implemented progressively in mouse meiosis. *Nat. Cell Biol.* **14,** 424–430 (2012).

34. Chicheportiche, A., Bernardino-Sgherri, J., de Massy, B. & Dutrillaux, B. Characterization of Spo11-dependent and independent phospho-H2AX foci during meiotic prophase I in the male mouse. *J. Cell Sci.* **120,** 1733–1742 (2007).

35. Lenzi, M. L. *et al.* Extreme heterogeneity in the molecular events leading to the establishment of chiasmata during meiosis i in human oocytes. *Am. J. Hum. Genet.* **76,** 112–127 (2005).

36. Carofiglio, F. *et al.* SPO11-Independent DNA Repair Foci and Their Role in Meiotic Silencing. *PLOS Genet.* **9,** e1003538 (2013).

37. Holloway, J. K., Booth, J., Edelmann, W., McGowan, C. H. & Cohen, P. E. MUS81 Generates a Subset of MLH1-MLH3–Independent Crossovers in Mammalian Meiosis. *PLOS Genet.* **4,** e1000186 (2008).

38. Tease, C., Hartshorne, G. M. & Hultén, M. A. Patterns of meiotic recombination in human fetal oocytes. *Am. J. Hum. Genet.* **70,** 1469–1479 (2002).

39. Cheng, E. Y. *et al.* Meiotic recombination in human oocytes. *PLoS Genet.* **5,** e1000661 (2009).

40. Ottolini, C. S. *et al.* Genome-wide maps of recombination and chromosome segregation in human oocytes and embryos show selection for maternal recombination rates. *Nat. Genet.* **47,** 727–735 (2015).

41. Cole, F. *et al.* Mouse tetrad analysis provides insights into recombination mechanisms and hotspot evolutionary dynamics. *Nat. Genet.* **46,** 1072–1080 (2014).

42. Jeffreys, A. J. & May, C. A. Intense and highly localized gene conversion activity in human meiotic crossover hot spots. *Nat. Genet.* **36,** 151–156 (2004).

43. Paigen, K. & Petkov, P. Mammalian recombination hot spots: properties, control and evolution. *Nat. Rev. Genet.* **11,** 221–233 (2010).

44. Allers, T. & Lichten, M. Differential Timing and Control of Noncrossover and Crossover Recombination during Meiosis. *Cell* **106,** 47–57 (2001).

45. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295,** 1306–1311 (2002).

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

47. Jm, C., R, R. & S, B. The many landscapes of recombination in Drosophila melanogaster., The Many Landscapes of Recombination in Drosophila melanogaster. *PLoS Genet. PLoS Genet.* **8, 8,** e1002905–e1002905 (2012).

48. Liu, H. *et al.* Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. *Genome Biol.* **16,** (2015).

49. Liu, H. *et al.* Direct Determination of the Mutation Rate in the Bumblebee Reveals Evidence for Weak Recombination-Associated Mutation and an Approximate Rate Constancy in Insects. *Mol. Biol. Evol.* **34,** 119–130 (2017).

50. Yang, S. *et al.* Parent-progeny sequencing indicates higher mutation rates in heterozygotes. *Nature* **523,** 463–467 (2015).

51. Charlesworth, B. Genetic Recombination: Patterns in the genome. *Curr. Biol.* **4,** 182–184 (1994).

52. Petes, T. D. Meiotic recombination hot spots and cold spots. *Nat. Rev. Genet.* **2,** 360–369 (2001).

53. Lercher, M. J., Smith, N. G. C., Eyre-Walker, A. & Hurst, L. D. The Evolution of Isochores: Evidence From SNP Frequency Distributions. *Genetics* **162,** 1805–1810 (2002).

54. Tsai, I. J., Burt, A. & Koufopanou, V. Conservation of recombination hotspots in yeast. *Proc. Natl. Acad. Sci.* **107,** 7847–7852 (2010).

55. Bobay, L.-M. & Ochman, H. Impact of Recombination on the Base Composition of Bacteria and Archaea. *Mol. Biol. Evol.* **34,** 2627–2636 (2017).

56. Lassalle, F. *et al.* GC-Content Evolution in Bacterial Genomes: The Biased Gene Conversion Hypothesis Expands. *PLOS Genet.* **11,** e1004941 (2015).

57. Marais, G., Charlesworth, B. & Wright, S. I. Recombination and base composition: the case of the highly self-fertilizing plant Arabidopsis thaliana. *Genome Biol.* **5,** R45 (2004).

58. Cotton, V. E., Hoffmann, E. R., Abdullah, M. F. F. & Borts, R. H. Interaction of Genetic and Environmental Factors in Saccharomyces cerevisiae Meiosis: The Devil is in the Details. in *Meiosis* 3–20 (Humana Press, 2009). doi:10.1007/978-1-59745-527-5_1

59. Rockmill, B. *et al.* High Throughput Sequencing Reveals Alterations in the Recombination Signatures with Diminishing Spo11 Activity. *PLOS Genet.* **9,** e1003932 (2013).

60. Rockmill, B., Sym, M., Scherthan, H. & Roeder, G. S. Roles for two RecA homologs in promoting meiotic chromosome synapsis. *Genes Dev.* **9,** 2684–2695 (1995).

61. Jiang, X. & Stern, D. Mating and tetrad separation of Chlamydomonas reinhardtii for genetic analysis. *J. Vis. Exp. JoVE* (2009). doi:10.3791/1274

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

63. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing

next-generation DNA sequencing data. *Genome Res.* **20,** 1297–1303 (2010).

64. Wang, J., Fan, H. C., Behr, B. & Quake, S. R. Genome-wide Single-Cell Analysis of Recombination Activity and De Novo Mutation Rates in Human Sperm. *Cell* **150,** 402–412 (2012).

65. Oh, S. D. *et al.* BLM ortholog, Sgs1, prevents aberrant crossing-over by suppressing formation of multichromatid joint molecules. *Cell* **130,** 259–272 (2007).

66. Oh, S. D., Lao, J. P., Taylor, A. F., Smith, G. R. & Hunter, N. RecQ helicase, Sgs1, and XPF family endonuclease, Mus81-Mms4, resolve aberrant joint molecules during meiotic recombination. *Mol. Cell* **31,** 324–336 (2008).

Author contributions

S.Y., L.D.H., D.T., and H.L. designed the experiments. H.L., J.H., X.S., J.L., Y.H., L.Y., S.Y, L.D.H. and G.L. performed the experiments and analyzed the data. L.D.H., H.L., S.Y. and D.T. wrote the paper.

Competing interests

The authors declare no competing interests.

Tables

Table1. Rate of COs and NCOs in each cross.

Organisms	Parental strains	Marker density	Tetrads sequenced	CO Events*	cM/Mb	CO-GC Events*	CO-GC Converted markers*&	NCO-GC Events*$	NCO-GC Converted markers*&
S. cerevisiae	YJM789 × S96 (S1)	0.44%	14	94.4	391	67.5	1.3×10^{-2}	47.8 (53.9)	9.0×10^{-3}
	YJM789 × S96[20]	0.44%	46	90.5	375	62.7	1.1×10^{-2}	46.2	8.4×10^{-3}
	YJM789 × YPS128 (S2)	0.50%	15	76.5	317	63.2	9.8×10^{-3}	46.4 (49.9)	7.4×10^{-3}
N. crassa	FGSC2489 × FGSC3246 (N1)	0.04%	10	11.5	14	0.3	3.8×10^{-5}	0.1	3.2×10^{-5}
	FGSC4200 × FGSC1363 (N2)	0.66%	20	16.1	20	8.1	2.9×10^{-4}	6.0 (11.1)	2.0×10^{-4}
	FGSC2225 × FGSC3246 (N3)	1.06%	18	15.7	19	8.7	1.2×10^{-4}	7.3 (10.0)	1.0×10^{-4}
	N3-14-2 × N3-14-4 (N4)^	0.61%	9	17.1	21	7.7	6.0×10^{-4}	4.3 (8.8)	1.8×10^{-4}
C. reinhardtii	CC124 × CC1010 (C1)	0.14%	6	16.5	8	0.2	1.0×10^{-5}	0.2	1.3×10^{-6}
	CC2935 × CC2936 (C2)	1.53%	12	25.3	12	12.3	4.6×10^{-5}	3.1 (8.1)	3.7×10^{-6}
	CC408 × CC2936 (C3)	1.17%	9	23.2	11	10.1	3.2×10^{-5}	2.2 (7.3)	3.2×10^{-6}
A. thaliana	Col × Ler (A)	0.26%	9	11.0	4.6	3.9	5.1×10^{-5}	2.0 (16.7)	1.1×10^{-5}
	Col × Ler[32]	0.12%	5	10.4	4.3	2.0	3.5×10^{-5}	1.4	1.0×10^{-5}
H. sapiens[21,64]	-	0.001%~0.05%	-	43/23#	1.5/0.8#	-	-	-	6×10^{-6}

*These numbers are counted as per tetrad per meiosis.

&These numbers are proportion of markers converted.

$Numbers in the parenthesis are corrected number based on simulated tract length and probability of detection.

^These two strains are products of a single meiosis in cross N3 (FGSC2225 × FGSC3246).

#43 COs per cell and 1.5 cM/Mb for oocytes and 23 COs per cell and 0.8 cM/Mb for sperms.

[20, 21, 32, 64] The four numbers correspond to the reference numbers, the grey shaded data are from these references.

Table 2. Nucleotide direction of gene conversion events. The ratio (number) of AT->GC and GC->TA conversions were listed in each category, p-values are from binomial test. AT-biased trend is highlighted in green while GC-biased trend is highlighted in red.

Organisms	GC-content of whole genome	Parental strains	GC-content at SNP sites	NCO-GC + CO-GC			NCO-GC			CO-GC		
				AT->GC	GC->AT	p-value	AT->GC	GC->AT	p-value	AT->GC	GC->AT	p-value
S. cerevisiae	38.30%	YJM789	48.79%	49.47% (7153)	50.53% (7307)	0.2	50.63% (2914)	49.37% (2842)	0.34	**48.70% (4239)**	**51.30% (4465)**	**0.016**
		S96	48.82%	**50.63%[15,20] (39445)**	**49.37% (38456)**	**0.0004**	50.52% (15238)	49.48% (14922)	0.07	**50.70% (24207)**	**49.30% (23534)**	**0.002**
		YJM789	47.85%	**49.03% (6664)**	**50.97% (6928)**	**0.024**	**48.52% (2816)**	**51.48% (2988)**	**0.025**	49.41% (3848)	50.59% (3940)	0.3
		YPS128	49.66%									
N. crassa	48.50%	FGSC4200	49.59%	50.13% (1133)	49.87% (1127)	0.92	49.46% (461)	50.54% (471)	0.77	50.60% (672)	49.40% (656)	0.68
		FGSC1363	50.90%									
		FGSC2225	50.43%	51.18% (735)	48.82% (701)	0.38	52.32% (361)	47.68% (329)	0.24	50.13% (374)	49.87% (372)	0.97
		FGSC3246	50.19%									
		C14-2	50.32%	50.34% (732)	49.66% (722)	0.81	50.75% (169)	49.25% (164)	0.83	50.22% (563)	49.78% (558)	0.9
		C14-4	50.01%									
C. reinhardtii	64.08%	CC2935	52.67%	52.95% (404)	47.05% (359)	0.11	**64.91% (37)**	**35.09% (20)**	**0.033**	51.98% (367)	48.02% (339)	0.31
		CC2936	52.62%									
		CC408	52.47%	51.50% (172)	48.50% (162)	0.62	**75.86% (22)**	**24.14% (7)**	**0.008**	49.18% (150)	50.82% (155)	0.82
		CC2936	52.63%									
A. thaliana	36.03%	Col	46.14%	54.17% (65)	45.83% (55)	0.41	50.00% (9)	50.00% (9)	1	54.90% (56)	45.10% (46)	0.37
		Ler	45.79%									
		Col[32]	45.98%	51.95% (40)	48.05% (37)	0.82	33.33% (1)	66.67% (2)	1	52.70% (39)	47.30% (35)	0.73
		Ler	44.89%									

[15, 20, 32] The three numbers correspond to the reference numbers, the grey shaded data are from these references.

Table 3. Correlations between COs, NCOs and GC-content at different scales. The genome is divided into 10 kb (20 kb, 100 kb, and 200 kb) non-overlapping windows. The number of COs, number of NCOs, and GC-content (percentage of identifiable bases that are G or C) are calculated for each window. The raw correlations between these variables at each window size are shown in table below. Significant correlations are highlighted in red.

Variables	Block size	Yeast		Neurospora		Chlamydomonas		Arabidopsis	
		Spearman's ρ	*p*-value	Spearman's ρ	*p*-value	Spearman's ρ	*p*-value	Spearman's ρ	*p*-value
CO and GC-content	10k	0.252	2.2E-16	0.077	8.4E-7	0.032	0.00077	-0.0387	2.5E-5
	20k	0.207	2.5E-7	0.129	5.5E-9	0.035	0.0102	-0.0458	0.0004
	50k	0.144	0.023	0.167	1.7E-6	0.048	0.026	-0.0621	0.0024
	100k	0.0588	0.51	0.154	0.0019	0.053	0.078	-0.117	5.1E-5
	200k	-0.149	0.22	0.167	0.016	0.075	0.076	-0.162	6.6E-5
NCO and GC-content	10k	0.195	6.8E-12	0.049	0.0018	0.036	0.00019	-0.0129	0.16
	20k	0.167	3.4E-5	0.062	0.00499	0.044	0.0012	-0.0194	0.13
	50k	0.167	0.0084	0.089	0.011	0.054	0.0114	-0.0362	0.077
	100k	0.1	0.26	0.152	0.0021	0.087	0.0038	-0.0368	0.2
	200k	-0.05	0.68	0.209	0.0025	0.175	3.5E-5	-0.0171	0.68
(CO+NCO) and GC-content	10k	0.270	<2.2E-16	0.0939	2.2E-9	0.0414	1.8E-5	-0.0407	8.9E-6
	20k	0.229	1.1E-8	0.145	5.0E-11	0.0455	0.0008	-0.0505	9.6E-5
	50k	0.201	0.0015	0.190	5.0E-8	0.0646	0.0025	-0.0709	0.0005
	100k	0.151	0.0898	0.192	9.5E-5	0.0783	0.009	-0.125	1.5E-5
	200k	-0.100	0.42	0.206	0.0029	0.142	0.0008	-0.157	0.0001

Figures legends

Figure 1. Schematic illustration of experiment design. After a cross is made between two strains, the meiotic products are dissected and whole-genome re-sequenced individually. The meiotic products in yeast, *Chlamydomonas,* and *Arabidopsis* are tetrads. Each tetrad can be dissected into four spores and each spore is sequenced after being grown into a clonal colony. The meiotic products in *Neurospora* are ascospores, which are products of a meiosis followed by a mitosis and thus have eight spores, every two spores originate from a mitosis. Four of the eight spores in each ascospore are selected for sequencing.

Figure 2. Outcomes of DSBR and SDSA pathways in meiotic recombination. Four kinds of outcomes of recombination events are shown in this figure: Type I: A single crossover event; Type II: A crossover and a non-crossover event occur at the same locus indicating that the two crossing over chromatids invaded a third chromatid during a recombination event[65,66]; Type III: Two non-crossover events occur at the same locus. This can be explained by resolution of a double Holiday junction in a NCO fashion[65,66] or two chromatids breaks at the same locus. Type IV: A single non-crossover event. The rates (count per tetrad per meiosis) of each outcome in these four organisms are shown. As shown in black dotted frame, for type II events, the CO and the NCO event do not share the same genotype switching point, and for type III events, the two NCO events mostly do not share the same conversion tract, and they always lead to 2:2 or 4:0 segregation.

Figure 3. Estimation of tract length for gene conversion events. Two methods were used for the estimation. **a.** and **b.** show the violin plot for estimated tract length for NCO-GCs (a) and CO-GCs (b) by the midpoint method. **c.** is the simulation of average markers converted at different tract lengths in each cross. e.g. in cross A, the observed number of markers converted by CO-GC is 4, and in this simulation this averages at 700 bp tract length. **d.** Summary of tract lengths estimated from the two methods.