nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-018-0503-9

Altered access to beneficial mutations slows adaptation and biases fixed mutations in diploids

Daniel A. Marad, Sean W. Buskirk and Gregory I. Lang*

Ploidy varies considerably in nature. However, our understanding of the impact of ploidy on adaptation is incomplete. Many microbial evolution experiments characterize adaptation in haploid organisms, but few focus on diploid organisms. Here, we perform a 4,000-generation evolution experiment using diploid strains of the yeast *Saccharomyces cerevisiae*. We show that the rate of adaptation and spectrum of beneficial mutations are influenced by ploidy. Haldane's sieve effectively alters access to recessive beneficial mutations in diploid populations, leading to a slower rate of adaptation and a spectrum of beneficial mutations that is shifted towards dominant mutations. Genomic position also has an important role, as the prevalence of homozygous mutations is largely dependent on their proximity to a recombination hotspot. Our results demonstrate key aspects of diploid adaptation that have previously been understudied and provide support for several proposed theories.

Understanding the impact of ploidy on adaptation is a central challenge in evolutionary biology. Ploidy varies considerably in the natural world[1] and all sexual organisms alternate between ploidy states through gamete fusion and meiosis[2]. Despite its importance, we have incomplete knowledge of how ploidy affects the rate of adaptation and the spectrum of beneficial mutations. In principle, how ploidy affects adaptation depends mostly on assumptions regarding the dominance of new beneficial mutations. If new beneficial mutations are mostly dominant, then diploids, with twice the mutational target size compared to haploids, will be twice as likely to acquire beneficial mutations, and will have greater evolutionary potential. Alternatively, if beneficial mutations are recessive, then haploids will have access to new beneficial mutations that have no selective benefit as heterozygotes in diploid populations[3]. It is probable that the filtering of recessive beneficial mutations in diploids—Haldane's sieve[4]—is not the only factor that affects the impact of ploidy on adaptation. Other factors, such as deleterious load and physiological differences (for example, protein expression), may also contribute to differences in adaptive potential[5]. If deleterious mutations are mostly recessive, negative phenotypes will be masked in diploids. However, if deleterious mutations are dominant (for example, haploinsufficient mutations), diploids would have twice the deleterious load compared to haploids. It has been shown in yeast that different ploidy states experience differential regulation of some genes[6], which may further affect the mutations that are selectively accessible in haploids and diploids.

The budding yeast, *S. cerevisiae*, can be propagated as haploids or diploids, making it an ideal system for studying the effect of ploidy on adaptation. Early work appeared to demonstrate that diploids evolved faster[7]; however, most recent studies find that haploid yeast adapt more quickly. This result holds across strain backgrounds and environments[8–11], with possible exceptions in small populations[11] or in high concentrations of fungicide[8]. Despite these exceptions, there is a general consensus that haploids adapt faster than diploids across a range of environments. However, the reasons why haploids adapt more quickly remain unclear.

Evolve-and-resequence studies have focused almost exclusively on haploid rather than on diploid organisms[12–15]. Collectively, these studies have shown that the spectrum of beneficial mutations in haploid populations is skewed towards loss-of-function mutations. Although there is less information regarding the spectrum of beneficial mutations in higher ploidies, one recent study has used whole-genome sequencing and gene expression analysis to explore adaptation of polyploid yeast to raffinose media and has observed a broader spectrum of beneficial mutations in tetraploids[16]. Some work in diploids indicate that Haldane's sieve is a filter for recessive beneficial mutations. For example, one study found that all haploid-evolved nystatin-resistant mutations were recessive[17]. Recent theory also suggests that beneficial mutations in diploids may be overdominant[18] and there is some experimental evidence suggesting that this is the case[11,19]. There is therefore a need to determine how diploidy changes the spectrum of beneficial mutations and the dynamics of adaptation.

Here we measure the rate of adaptation for 48 diploid populations through 4,000 generations and compare these results to previously evolved haploid populations[14]. We sequence two clones each from 24 populations after 2,000 generations and we perform whole-genome whole-population time-course sequencing on two populations. We show that diploids adapt more slowly than haploids, that ploidy alteres the spectrum of beneficial mutations, and that the prevalence of homozygous mutations depends on their genomic position. In addition, we validate haploid-specific, diploid-specific and shared mutational targets by reconstructing individual mutations in the ancestral background.

Results

Previously, we have determined the spectrum of mutations in haploid populations evolved for 1,000 generations[14]. We have observed a large number of nonsense and frameshift mutations, suggesting that adaptation in haploids is mostly driven by loss-of-function mutations. Many of these mutations are likely to be recessive in diploids given that only 3% of gene deletions are haploinsufficient or haploproficient[20].

Haploids adapt mostly by recessive beneficial mutations. We hypothesize that haploid adaptive mutations are recessive. To test this, we isolated twelve individual clones containing diverse mutations from ten evolved haploid populations and crossed them to the

Department of Biological Sciences, Lehigh University, Bethlehem, PA, USA. *e-mail: glang@lehigh.edu

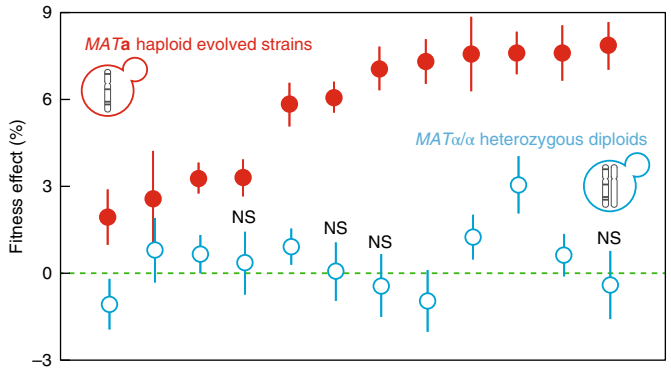

Fig. 1 | Beneficial mutations in evolved haploids are recessive. Filled red circles indicate fitness of twelve evolved *MAT***a** clones from the haploid experiment[14]. Open blue circles indicate the fitness of the same twelve strains as *MAT***a**/**a** diploids, in which all of the evolved mutations are heterozygous. Data are mean ±s.e.m. of 28 replicates. The four diploid populations that are not significantly different from zero are indicated with NS (two-tailed, one-sample *t*-test, $\alpha = 0.05$). Haploid fitness data are from ref. [31]. In order from left to right, the data points correspond to populations (and known haploid-beneficial mutations) RMS1-G02-545 (*mot3*), BYS1-E03-745 (*ymr102C, say1, ira1*), BYS2-C06-1000, BYS1-A08-545 (*gas1, ste12*), BYS2-D06-910 (*ira2, ste5, gas1*), BYS1-A09-1000 (*gpb2, cne1*), RMS1-G02-825 (*ira1, yur1*), RMS1-H08-585 (*ira1, kre6*), BYS-D08-1000 (*ste4, yur1*), BYS2-E01-745 (*kel1, hsl7*), RMS1-H08-585 (*ste11, kre6*) and RMS1-D12-910 (*mid2, ira1, rot2*) in refs [14,31].

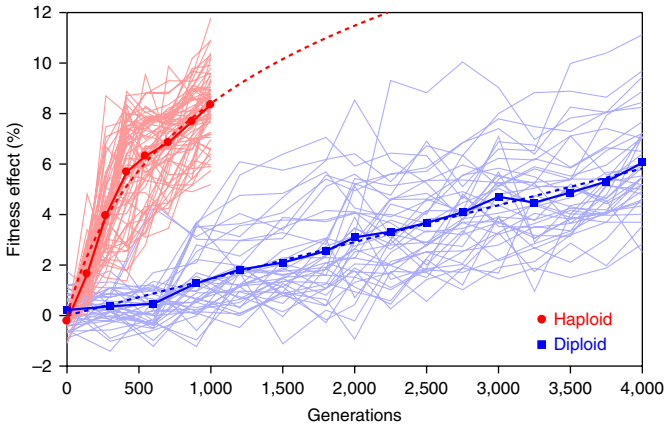

Fig. 2 | The rate of adaptation of haploid and diploid populations. Over the course of 4,000 generations of laboratory evolution, diploids adapt more slowly than haploids. Fitness effects for haploids and diploids are plotted as means represented by red circles and blue squares, respectively. Light-red and light-blue traces indicate individual trajectories of haploid and diploid populations, respectively. The red dotted line represents the fit of the haploid average to the power law equation $y = (bx + 1)^a - 1$. The blue dotted line represents the linear fit of the diploid average data. Haploid data are from ref. [14]. Diploid fitness data points below 0 are likely owing to measurement error.

*MAT*α version of the ancestor to generate diploids that are heterozygous for all of the evolved mutations. To control for differences in gene expression and physiology, all heterozygous diploids were converted to *MAT***a**/**a**. Fitness of the haploid evolved strains ranged from 2% to 8% (Fig. 1). Heterozygous diploid fitness is significantly different from zero in 8 out of 12 populations, but decreases compared to the haploid fitness (Fig. 1). On the basis of this result, we expect Haldane's sieve to filter out most of these mutations if they occur in a diploid background, leading to a slower rate of adaptation in diploids and a spectrum of beneficial mutations that is shifted away from recessive loss-of-function mutations.

Diploids evolve more slowly than haploids. To test our hypothesis that the rate of adaptation and mutational spectrum differs between ploidies, we evolved 48 diploid populations in rich glucose medium for 4,000 generations under conditions identical to those used for the haploid experiment[14]. On average, diploid populations increased in fitness by only 5.8% over 4,000 generations, whereas haploid populations increased in fitness by 8.5% over 1,000 generations (Fig. 2). As expected given the stochastic nature of evolution, we observed a

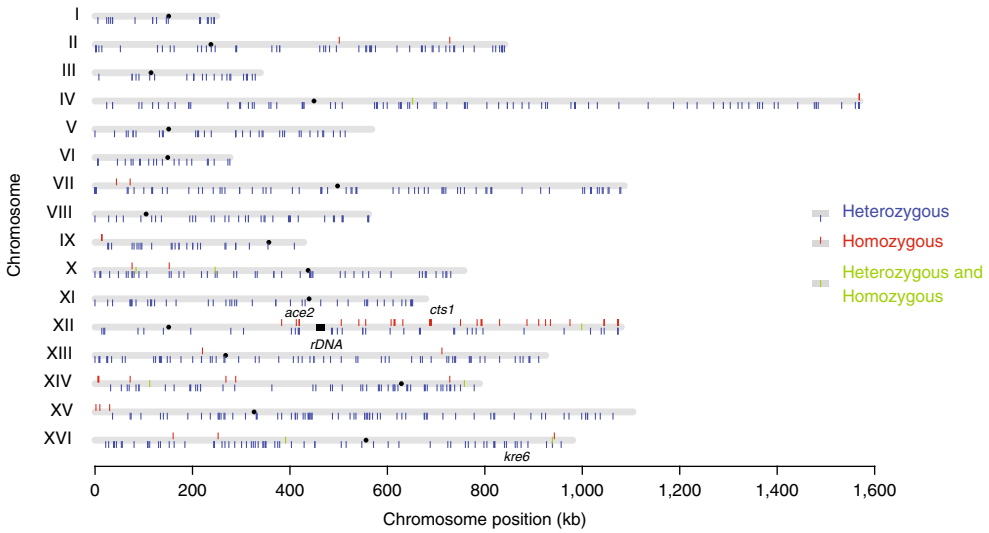

Fig. 3 | Genome-wide distribution of mutations in diploids evolved for 2,000 generations. Sequencing data for 831 individual nuclear mutations (both shared and clone-specific) across all 24 diploid populations show an overrepresentation of homozygous mutations on the right arm of chromosome XII ($P < 0.001$, χ^2 test). Chromosomes are represented as grey bars. Homozygous mutations are represented by red lines at the top of the chromosome. Heterozygous mutations are represented by blue lines at the bottom of the chromosome. Mutations that were found both as homozygous and heterozygous mutations in paired samples are shown as green lines in the middle of the chromosome.

Table 1 | Common targets of selection in diploid populations

Gene	Mutation			Population	Biological process
	Homozygous	Heterozygous	Total		
CTS1	6	0	6	A07, D05, E06, G08, H08, F05	Cell separation after cytokinesis
ACE2	2	3	5	C04, F09, A05, F04, F04	Cell separation after cytokinesis
FLO1	0	3	3	H08, A05, F04	Flocculation
AGP1	0	2	2	E04, D04	Amino acid transmembrane transport
BCK1	0	2	2	C05, E04	Regulation of cell wall organization
BEM2	0	2	2	C04, H08	Actin cytoskeleton organization
BPH1	0	2	2	C05, A08	Cell wall organization
BST1	0	2	2	F09, H04	Negative regulation of COPII vesicle formation
CLA4	1	1	2	D09, H05	Negative regulation of sterol import
CLB2	1	1	2	E04, H08	Regulation of mitotic spindle elongation
EDE1	0	2	2	D05, C05	Endocytosis
FKS1	2	0	2	B09, E06	1,3-beta-glucan biosynthesis
HAP5	0	2	2	E06, B09	Regulation of respiration
KRE6	0	2	2	F04, G06	1,6-beta-glucan biosynthesis
MHP1	0	2	2	C04, B04	Cell wall organization
MSS116	0	2	2	D04, B04	Group I and II intron splicing
NPL6	0	2	2	C05, A07	Nucleosome disassembly
NUM1	0	2	2	F07, H05	Microtubule cytoskeleton organization
PDR1	0	2	2	D09, G08	Regulation of pleiotropic drug response
PEX1	0	2	2	F07, G06	Protein import into peroxisome matrix
PSE1	0	2	2	F07, G06	Protein import into nucleus
PUF2	0	2	2	A05, A07	Nuclear-transcribed mRNA decay
RKR1	0	2	2	H08, E04	Ribosome quality control complex
RPL33A	0	2	2	E04, H05	Cytoplasmic translation
RSF2	0	2	2	F07, A05	Regulation of transcription
SNT2	0	2	2	D04, E06	Histone ubiquitination
TRA1	0	2	2	F07, G06	Histone acetylation
UBP12	0	2	2	H08, F07	Ubiquitin recycling
UBP2	0	2	2	B09, F05	Ubiquitin recycling
URC2	0	2	2	B05, B09	Uracil catabolic process
USA1	0	2	2	E06, H08	Ubiquitin-dependent ERAD pathway
UTR2	0	2	2	B04, G08	Cell wall organization
YDR090C	0	2	2	C05, E06	Unknown
YPR202W	0	2	2	A07, H04	Unknown
YUR1	1	1	2	E04, D05	Cell wall mannoprotein biosynthesis

Mutations are counted as a single mutation in this list whether they are shared or unique to paired clones. We expect 465 genes to be mutated once by chance, 20 genes to be mutated twice by chance and one to be mutated three or more times by chance. Population F04 contains one shared and one unique mutation to *ACE2*. The list of biological processes has been manually curated to select the most relevant gene ontology (GO) terms from entries on the *Saccharomyces* Genome Database (www.yeastgenome.org). There are no statistically significant GO terms for biological processes in this gene list.

large variation in fitness gains and trajectories across replicate populations. Diploids increased in fitness by between 2% and 11% over 4,000 generations, whereas haploids increased in fitness between 5% and 12% over 1,000 generations (Fig. 2). Variance also changed over time among both haploid and diploid populations (Supplementary Fig. 1). We fitted the averaged data to both a power law (which has been shown to accurately model haploid evolution[21]) and a linear fit. The haploid data are described better by a power law fit than by a linear fit ($P < 0.001$, F-test) but the diploid data are not ($P = 1$, F-test).

The spectrum of beneficial mutations is different between diploids and haploids. Having confirmed that diploids adapt more slowly than haploids, we next determined the spectrum of mutations in diploids by sequencing 48 clones (two clones each from 24 randomly selected diploid populations) from generation 2,000. We identify 850 individual de novo mutations across the 24 populations. Of these mutations, 416 are shared between both clones in the population and 434 are unique to only one clone (Supplementary Table 1). Heterozygous mutations outnumber homozygous mutations 1,135 to 105 (Supplementary Fig. 2). The number of mutations per population ranged from 20 in population C04 to 55 in population G08. Of the 850 mutations, 342 are intergenic, 114 are synonymous, 336 are missense, 23 are nonsense and 28 are frameshift. In addition, we identified two conservative in-frame deletions, two

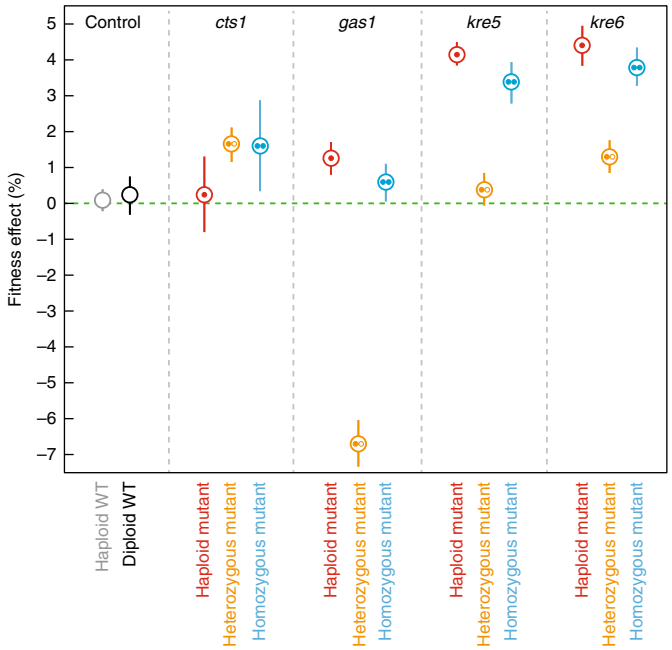

Fig. 4 | Evolved alleles show varying degrees of dominance and ploidy dependence. Average fitness effects of haploid (single filled circle in red), heterozygous diploid (one filled circle and one open circle in yellow) and homozygous diploid (two filled circles in blue) mutants with mutations in cell wall genes are compared to wild-type (WT) haploids (light-grey empty circle) and wild-type diploids (dark-grey empty circle). The corresponding gene mutated in each strain is listed at the top. The *cts1* mutation (Glu298*) is from the F05 population of our diploid data (Supplementary Table 2). The *gas1*, *kre5* and *kre6* mutations are from the haploid populations BYS2-D06 (F259L), RMB2-B10 (T872K) and RMS1-H08 (I478L), respectively[14]. Each point is a mean of at least sixteen replicates. Error bars are the s.e.m.

disruptive in-frame deletions, two stop codon readthroughs, and one complex frameshift and missense mutation (Supplementary Table 2). Homozygous mutations are overrepresented on the right arm of chromosome XII ($P < 0.001$, χ^2 test), with 19 out of 24 populations containing a homozygous mutation on the right arm of chromosome XII. Excluding chromosome XII, the genome-wide distribution of homozygous and heterozygous mutations is not significantly different from random ($P = 0.06$, χ^2 test; Fig. 3). In addition to point mutations, we observed trisomy XIV in both clones from populations G08 and H08 and trisomy VII in clone B from population D04 (Supplementary Fig. 3).

We identified candidate targets of selection as those genes in which mutations were seen in at least two populations (Table 1). We observed notable differences in the biological processes that are common targets of selection in diploids compared to haploids. Genes involved in the Ras pathway and the mating pathway are common targets of selection in haploid populations[14]. However, no mutations in either of these two pathways were among our candidate targets of selection in the evolved diploids. Cell wall biogenesis, and assembly and cytokinesis are shared targets of selection in both haploid and diploid populations. However, even within these biological processes, there are both shared (for example, *KRE6*) and ploidy-specific target genes, such as the haploid-specific *GAS1* and *KRE5* and the diploid-specific *CTS1*. Furthermore, the most common haploid target of selection, *IRA1* (a negative regulator of Ras, mutated in 21 out of 40 haploid populations), was not mutated in any of our diploid populations.

Evolved alleles show varying degrees of dominance and ploidy-dependence. Given that mutations in genes involved in cell wall biosynthesis and assembly represent a diverse subset of beneficial mutations, including haploid-specific (*GAS1*, *KRE5*), diploid-specific (*CTS1*), and common (*KRE6*) beneficial mutational targets, we focused on these four targets of adaptation. We generated these alleles as haploids, heterozygous diploids and homozygous diploids and performed competitive fitness assays on all strains to test whether the relative fitness effects of evolved alleles are contingent on the ploidy in which they arose (Fig. 4). For these reconstructions, we used alleles of *KRE6*, *GAS1*, and *KRE5* from previous haploid evolution[14] and the *CTS1* allele from our diploid population F05. Consistent with our classification of these mutations as candidate drivers of adaptation, we found that each is beneficial in the background in which it arose. Mutations in the shared target of selection, *KRE6*, are beneficial as haploids, heterozygous diploids and homozygous diploids, although the benefit for the heterozygote is less than that for the homozygote, showing that it is partially dominant (coefficient of dominance, $h = 0.34$). The diploid-specific target of selection, *CTS1*, is equally beneficial for the heterozygote and homozygote ($h \approx 1$); however this mutation is neutral in a haploid background. Mutations in the haploid-specific targets, *GAS1* and *KRE5*, are beneficial in haploids and homozygous diploids. In the case of *KRE5*, the heterozygous diploid fitness is not statistically different from zero ($h = 0.10$; Fig. 4). In the case of *GAS1*, the heterozygote is strongly deleterious ($h < 0$). Notably, this may indicate an incompatibility between the ancestral and evolved alleles of *GAS1*.

Loss-of-heterozygosity occurs more rapidly for *CTS1* than *ACE2*. Although our reconstruction experiments show no significant fitness difference between a heterozygous and homozygous *CTS1* mutation, all six *CTS1* mutations are homozygous in the diploid populations. Another common diploid-specific target, *ACE2*, is observed as a homozygous mutant in two of the four populations in which it arose. Both *ACE2* and *CTS1* are located on chromosome XII. This frequent observation of homozygous mutations on chromosome XII could potentially be explained by unbalanced structural variation resulting from whole or partial loss of one copy of chromosome XII in these populations. We found that coverage across chromosome XII was consistent with coverage across the rest of the genome, verifying that our strains had two copies of chromosome XII. Furthermore, tetrad dissections of *CTS1* mutant strains produced four-spore viable tetrads, which indicates that it is unlikely that all or part of one copy of chromosome XII is missing. These homozygous mutations therefore probably arose from balanced loss-of-heterozygosity (LOH) events, which maintain chromosome copy number.

With this in mind, we investigated the dynamics of genome sequence evolution by performing whole-genome whole-population time-course sequencing on two populations, one with a homozygous *CTS1* mutation, and one with a homozygous *ACE2* mutation. In population A05, the *ace2* allele fixes as a heterozygote before undergoing LOH and fixing as a homozygote (Fig. 5b). In population F05, however, the *cts1* homozygote establishes before the heterozygote fixes (Fig. 5d). We can use the information from the dynamics of additional mutations in these populations to inform our theory of balanced LOH. When the *ace2* allele fixes in the A05 population, an ancestral heterozygosity in an *LTR* located approximately 387 kb away from *ACE2* on chromosome XII also loses heterozygosity (Fig. 5a). In the F05 population, the same ancestral *LTR* heterozygosity and a mutation in *DUS4*, which is also located on chromosome XII, lose heterozygosity at the same time as *CTS1* (Fig. 5c).

The F05 allele of *CTS1* is a nonsense mutation (Glu298*) and is likely to be a loss-of-function mutation. The other five *CTS1*

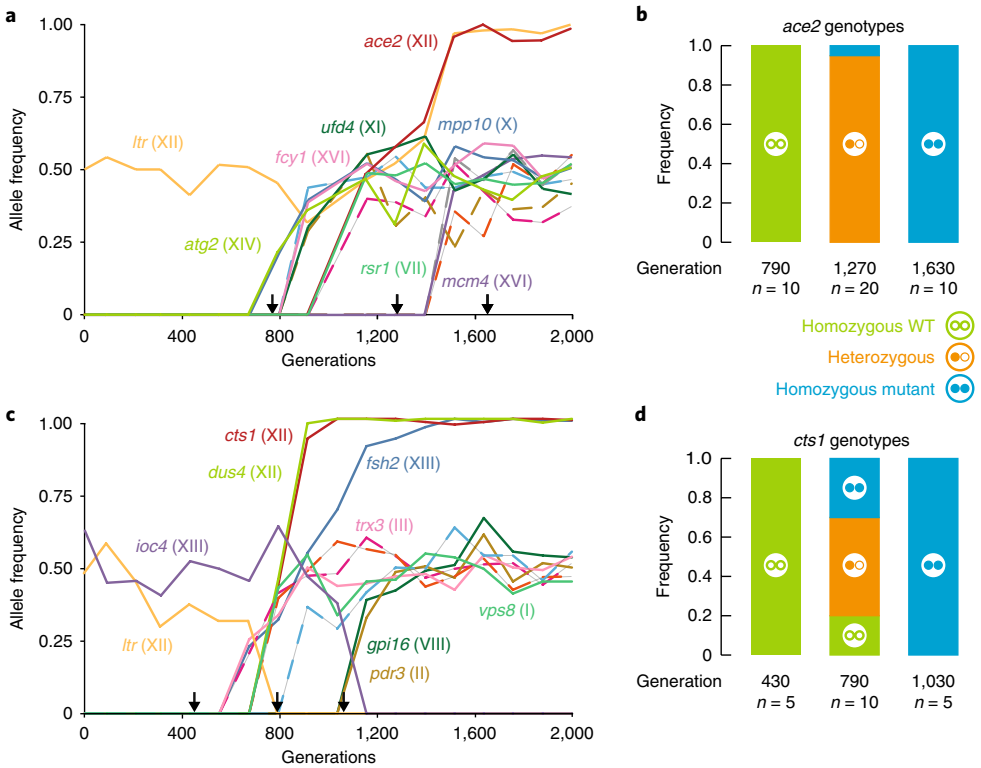

Fig. 5 | Dynamics of adaptation and LOH. a, Time-course sequencing of population A05 containing an *ace2* mutation. There is a decrease in slope around an allele frequency of 0.5, suggesting heterozygote fixation before homozygous mutant establishment. Mutations are shown with the chromosomes on which they occur. Black arrows represent time points at which the populations were sampled for Sanger sequencing. **b**, *ACE2* genotype by Sanger sequencing of clones isolated at three time points are shown as homozygous wild type (two open circles), heterozygous (one filled and one open circle) and homozygous mutant (two filled circles). Homozygous mutants only appear after the homozygous wild type is eliminated by the sweep of the heterozygous mutant. Number of clones subjected to Sanger sequencing is listed for each time point. **c**, Time-course sequencing of population F05 shows multiple mutations fixing as a cohort with *cts1*. The *cts1* mutant fixes in the population quickly, without a pause around 0.5 signifying heterozygote fixation. Mutations are shown with the chromosomes on which they occur. Dashed lines represent intergenic mutations. Black arrows represent time points at which the populations were sampled for Sanger sequencing. **d**, *CTS1* genotype by Sanger sequencing of clones isolated at three time points are shown as homozygous wild type (two open circles), heterozygous (one filled and one open circle) and homozygous mutant (two filled circles). Homozygous mutants appear before homozygous wild type is eliminated. Number of clones subjected to Sanger sequencing is listed for each time point.

mutant alleles in our diploid populations are missense mutations. Nevertheless, all six alleles of *CTS1* show similar dynamics and rates of LOH (Supplementary Fig. 4). On the basis of these results, we hypothesize that the difference in rates of LOH events between *CTS1* and *ACE2* mutations is probably not due to the effects of individual mutations, but rather is due to the location of the *CTS1* and *ACE2* genes themselves. A high rate of LOH on the right arm of chromosome XII is further supported by examining the fate of the *LTR* polymorphism on chromosome XII, which is one of six loci that was heterozygous in the founding strain. The other five, on chromosomes I, II, XII, XIII and XIV, maintain heterozygosity in approximately 90% of the clones at generation 2,000. By contrast, the chromosome XII polymorphism loses heterozygosity in all 24 populations, with 11 populations fixing one allele and 13 populations fixing the other. This is consistent with a high rate of LOH on the right arm of chromosome XII and suggests that selection is not necessary for LOH in this region.

Discussion

We show that our diploid populations adapt more slowly than haploids. We attribute this result to an altered spectrum of beneficial mutations selectively available to diploids. In support of this, we show that, while recessive loss-of-function mutations are common in haploid evolution, their effect is limited in diploid evolution. The rate of haploid adaptation starts off higher compared to diploids but decreases over time, whereas the diploid rate of adaptation is lower but constant (Supplementary Fig. 5). This is similar to a previous study of haploids and diploids evolved for 5,000 generations, which finds that haploids initially adapt faster, but that the average fitness increase for haploids and diploids converges by 5,000 generations[11]. The declining rate of adaptation in haploids may be due to the exhaustion of available recessive beneficial mutations; however, our data do not rule out other possibilities such as stronger diminishing-returns epistasis among beneficial mutations in haploids.

We identify 1,266 total de novo mutations across 48 sequenced clones. We find no correlation between the number of mutations in a population and the fitness of that population (Supplementary Fig. 6). We identify candidate targets of selection as genes that were mutated more often than expected by chance across all populations (Table 1). Two populations (A09 and H06) did not have any mutations in common targets of selection. Through our comparison of the mutations gained during haploid and diploid adaptation, we observed that both the mating pathway and the negative regulation of Ras were prominent targets of adaptation in haploids, but these pathways were not common targets of selection in diploids. This makes sense for the mating pathway, as it is repressed in diploids (and therefore sterile mutations are not selectively advantageous). With regards to why negative regulation of Ras is not a target of

adaptation in diploid populations, there are three possibilities. The first is that all or most of the spectrum of beneficial mutations that can be made to genes involved in the negative regulation of Ras are recessive in their fitness effects, and thus unlikely to pass through Haldane's sieve. Another possibility is that the Ras pathway is regulated differently in haploids and diploids, such that some genes are not functionally redundant in haploids and diploids. Some yeast genes are known to have ploidy-specific regulation[6]. Interestingly, *CTS1* is included among these differentially regulated genes, increasing in expression with ploidy. This physiological difference between haploids and diploids may explain why mutations in *CTS1* are diploid-specific. Lastly, it is possible that the Ras pathway is a target of selection in diploids but that gain-of-function mutations (which are accessible to diploids) are simply less common than the loss-of-function mutations accessible to haploids. Indeed, at least one mutation in this pathway (a mutation to *CYR1*) has been previously seen in evolved diploids[22]. This lends further support to an altered spectrum of mutations that are selectively accessible to haploids and diploids.

By examining our diploid evolved mutations, we found that Haldane's sieve is effective at filtering out recessive beneficial mutations from evolving diploid populations, but it is not the only factor affecting the rate of adaptation and spectrum of beneficial mutations in diploids. We also see that genomic position is an important factor, largely owing to varying rates of recombination throughout the genome. This is particularly visible on the right arm of chromosome XII, which contains the *rDNA* locus, a known recombination hotspot in yeast[23]. This locus has also been shown to be the site of frequent LOH in natural populations[24]. Furthermore, we find one ancestral heterozygosity on the right arm of chromosome XII, which loses heterozygosity in all 24 sequenced populations, supporting the hypothesis that frequent LOH is initiated at this locus. We find that homozygous mutations are rare (only 10% of diploid mutations), but are largely concentrated on the right arm of chromosome XII (Fig. 3), particularly in the *CTS1* and *ACE2* genes (Supplementary Fig. 7). This implies that the ability for beneficial diploid mutations to become homozygous, and thus to escape from Haldane's sieve, will depend strongly on local rates of LOH events. Our results further validate the idea that mitotic recombination is an important factor in the spread of beneficial alleles in evolving asexual populations[25].

We reconstructed evolved haploid-specific, diploid-specific and shared alleles in isolation as haploids, heterozygous diploids and homozygous diploids, and found that the degree of heterozygosity differs between the alleles. Our reconstructed allele of *CTS1* is dominant ($h \approx 1$), yet all six of the *cts1* alleles present in our diploid populations rapidly converted and fixed as homozygotes. One possible explanation is that the homozygote does have an advantage over the heterozygote but that we cannot detect this difference with flow-cytometry-based fitness assays. One complication is that homozygous mutations in *CTS1* result in a cell aggregation phenotype. This aggregation phenotype has previously been shown to occur during yeast laboratory evolution[26–29]. Our evolved *CTS1*-mutant diploid populations and the reconstructed *cts1* haploid and homozygous diploid strains form aggregates, but the reconstructed heterozygous mutant does not (Supplementary Fig. 8). It is possible that this phenotype complicates fluorescence-based fitness measurements of these strains.

Furthermore, from the dynamics of adaptation of all six *CTS1*-mutant populations (Fig. 5a, Supplementary Fig. 4), we have shown that *CTS1* mutants become homozygous both very frequently and quickly. Additionally, from the dynamics of the F05 population containing our *CTS1* mutation of interest, we see a mutant *dus4* allele that travels to fixation with the *cts1* allele, and one ancestral heterozygosity in a long terminal repeat that loses heterozygosity at the same time (Fig. 5a). All three of these genetic loci are located on chromosome XII telomeric to the yeast *rDNA* locus. We propose that a single recombination event caused a LOH event for all three of these loci, establishing the *cts1* homozygous mutant. Our dynamics of adaptation also show LOH of the same ancestral heterozygous long terminal repeat paired with the rise of the *ace2* allele to fixation, suggesting that a similar LOH event may have established the *ace2* homozygote. In this same population, there is another case where LOH in *IOC4* co-occurs with the fixation of an *FSH2* mutant. Both of these loci are on chromosome XIII (Fig. 5c). Because of these examples and the concentration of homozygous mutations and lack of heterozygous mutations on the right arm of chromosome XII, we also propose that when LOH events occur, they often involve a single point of recombination leading to LOH from the point of origin through the telomere. This is similar to the types of lesions observed due to break-induced replication[30].

Although LOH allows us to explain the location-specific enrichment of homozygous mutations, these are a minority of our candidate driver mutations (Table 1) and are therefore not sufficient to explain the fitness gains that we see in diploids. There must also be some adaptive heterozygous mutations, which remain heterozygous. It is possible that some of our heterozygous candidate driver mutations are maintained as heterozygotes because they are overdominant as suggested previously[18,19]. Overall, we have shown that over the course of 4,000 generations of evolution, diploid populations adapt more slowly than haploid populations, and that diploid populations have an altered spectrum of beneficial mutations compared to haploid populations. In addition, the majority of adaptive diploid mutations are heterozygous and the prevalence of adaptive diploid homozygous mutations depends on the position of the mutations in the genome. Collectively, this work fills a gap in our understanding of how ploidy impacts adaptation, and provides empirical support for the hypothesis that diploid populations have altered access to beneficial mutations.

Methods

Strain construction. The strains used in this experiment are derived from the base strain, yGIL432, a haploid yeast strain derived from the W303 background with genotype *MAT***a**, *ade2-1*, *CAN1*, *his3-11*, *leu2-3,112*, *trp1-1*, *URA3*, *bar1*Δ::*ADE2*, *hmlα*Δ::*LEU2*, *GPA1*::*NatMX*, *ura3*Δ::P_{FUS1}-*yEVenus*. This strain was previously reported as DBY15105[27]. A nearly isogenic *MAT*α version of yGIL432 was constructed by introgressing the *MAT*α allele through three backcrosses into the yGIL432 background. This *MAT*α strain, yGIL646, was crossed to yGIL432 to generate the diploid strain yGIL672. Haploid *MAT***a** evolved strains used in Fig. 1 have been described previously[14], and their corresponding *MAT***a**/**a** heterozygous diploids have also been described previously[31]. For the reconstruction experiments, the haploid *cts1* mutation in the yGIL432 background was generated using Cas9 allele replacement[32]. In brief, we retargeted the guide-RNA (gRNA)-expressing plasmid (Addgene 43803) to *CTS1* using the gRNA sequence 5′-TTCTTCAAAATCTCAACATA-3′. We co-transformed the gRNA plasmid (pGIL083), a constitutive Cas9 plasmid (Addgene 43802), and a plasmid (pGIL089) containing the *cts1* mutant allele into our *MAT***a** strain, yGIL432. We screened for plasmid retention, PCR-screened single colonies for allele replacement, and cured the strains of the plasmids. Allele replacements of *gas1*, *kre5* and *kre6* in the yGIL432 strain were obtained from M. Remillard (Princeton University), who constructed these strains using alleles from our haploid populations BYS2-D06, RMB2-B10 and RMS1-H08, respectively. Each of the haploid *MAT***a** strains was crossed to yGIL646 to generate heterozygous diploids. The heterozygous diploids were then sporulated, tetrads were dissected, and haploid spores were mating-type tested and genotyped. Appropriate haploid spores were then crossed to each other to create homozygous diploid strains for each mutation.

Long-term evolution. To set up the long-term evolution experiment, a single clone of yGIL672 was grown to saturation in YPD (yeast extract, peptone, dextrose) medium, was diluted 1:1,000, and was used to seed 48 replicate populations in a single 96-well plate. This initial plate was duplicated and then frozen for future use. The cultures were evolved through 4,000 generations (400 daily cycles) of growth and dilution in YPD at 30 °C. Every 24 h, the populations were diluted 1:1,024 by serial diluting 1:32 (4 µl into 125 µl) × 1:32 (4 µl into 125 µl) into new YPD medium containing ampicillin (100 mg l^{-1}) and tetracycline (25 mg l^{-1}). All dilutions were performed using the Biomek Liquid Handler equipped with a Pod96. Approximately every 50 generations, populations were mixed with 50 µl of 60% glycerol and archived at −80 °C.

Competitive fitness assays. Flow-cytometry-based competitive fitness assays were performed as described previously[14,27,31]. In brief, experimental and reference strains were grown to saturation in separate 96-well plates. If the strains to be tested were coming from the freezer, the strains were passaged once by diluting 1:1,024 to re-acclimate the strains to the appropriate medium. Experimental and reference strains were mixed 50:50 using the Biomek Liquid handler and were propagated for 40 generations under identical conditions to the original evolution experiment. We excluded 10 populations from Fig. 2, because we could not accurately measure fitness for populations that formed large cell aggregates or abnormal cell pellets. In this project, we used two reference strains: diploid *MAT***a**/**a**, and diploid *MAT***a**/α. These reference strains are derived from the yGIL432 base strain, but contain a constitutive ymCitrine integrated at the *ura3* locus. Between time points 1,200 and 2,000, we measure 8 out of 152 (5%) points at or below zero. The measurement error of our flow-cytometry-based fitness assay is 0.5%. Given a 0.5% or 1% fitness advantage, there is an 18% or 3% chance, respectively, that a single measurement will fall at or below zero, based on measurement error alone.

We used previously collected competitive fitness data from evolved *MAT***a** haploid strains[14]. We performed a competitive fitness assay as described below on previously generated heterozygous *MAT***a**/**a** diploid strains[31]. To perform this assay, we generated a *MAT***a**/**a** diploid version of the fluorescently labeled reference strain using a plasmid containing *LEU2* under a *MAT***a**-specific promoter to select for LOH at the mating-type locus. These data were normalized to their respective references and compiled into Fig. 1.

Whole-genome sequencing. For sequencing clones, we used single colonies grown on YPD and picked two colonies from each population to sequence. These individuals were grown to saturation in liquid medium and total genomic DNA was isolated for each sample. For whole-genome whole-population time-course sequencing, we thawed each population at 18 time points from generation 0 to 2,000 and transferred 10 µl into 5 ml of YPD. We made genomic DNA preparations as described previously[31].

We followed a modified version of the Nextera sequencing library preparation protocol[33], according to the previously described protocol[31]. We used the Nextera sequencing library preparation kit and protocol to isolate total genomic DNA and add the unique Nextera library barcodes to all 48 samples. We measured the concentration of DNA in each sample using a NanoDrop spectrophotometer and confirmed these values using a Qubit fluorometer. We equalized the DNA concentration of each sample via dilution and mixed all 48 samples into a single pool. We used a BioAnalyzer High-Sensitivity DNA Chip (BioAnalyzer 2100, Agilent) to confirm that the pool contained DNA fragments of the appropriate length, and performed gel extraction on the pool to remove short fragments. The pool was run on an Illumina HiSeq 2500 sequencer with 157-nucleotide single-end reads by the Sequencing Core Facility within the Lewis-Sigler Institute for Integrative Genomics at Princeton University. After an initial sequencing run provided us with the number of reads from each sample, we remixed the pool to better represent underrepresented samples and it was resequenced.

Sequencing analysis pipeline. The raw sequencing information was first merged from three lanes of sequencing via concatentation. This single file was demultiplexed into 48 FASTQ files by clone-specific barcodes using a custom Perl script (barcode_splitter.py) from L. Parsons (Princeton University) and clipped using fastx_clipper from the FASTQ Toolkit. Each sample was aligned to the complete and annotated W303 genome[34] using Burrows–Wheeler Aligner (BWA, v.0.7.12), using default parameters except 'Disallow an indel within INT bp towards the ends' set to 0 and 'Gap open penalty' set to 5. Variants were called using FreeBayes, using default parameters (v.0.9.21-24-g840b412). To remove spurious calls, variants common to more than 30 individual samples were removed using the VCFtools 'vcf isec' function (v.0.1.12b). Individual VCF files were annotated using SnpEff (v.4_3o). Each annotated variant was then manually confirmed using Integrated Genome Viewer (Broad Institute) and all were merged into one TEXT file (Supplementary Table 2). Each clone was sequenced to an average depth of approximately 30× coverage, and we determined whether mutations were unique to individual clones or shared between both clones within a population.

For whole-genome whole-population time-course sequencing, we used the same Nextera sequencing library preparation protocol as described above, with the following changes. Instead of two clones from each of 24 populations, we isolated genomic DNA from whole populations at 18 time points for two populations for a total of 36 samples. During the sequencing analysis, after splitting the reads based on the Nextera barcodes, we used FASTX clipper to remove any Nextera adapter sequences introduced by sequencing short fragments. We aligned these FASTQ files to an annotated version of the S288C genome. We used a previously described set of scripts (allele_counts.pl and composite_scores.pl) to call real mutations[14].

Five ancestral heterozygosities were detected through the identification of variants called in at least 30 of the 48 sequenced clones, most of which maintained the variants as heterozygotes. A sixth ancestral heterozygosity was first called as a homozygous mutation in 22 of the 48 evolved clones. The variant was later shown to be an ancestral heterozygosity following whole-genome time-course sequencing of the A05 and F05 populations.

Evolved genomes were evaluated for aneuploidies by comparing the median coverage across each chromosome to the median genome-wide coverage for each clone. Estimates of chromosome copy number are shown in Supplementary Fig. 3.

Sanger sequencing. We isolated multiple clones from populations A05 and F05 at three time points each, grew up liquid cultures of each clone, and isolated total genomic DNA. For population A05, these time points (and number of clones isolated) were generations 790 (10), 1,270 (20) and 1,630 (10). For population F05, these time points were generation 430 (5), 790 (10) and 1,030 (5).
We performed PCR on these samples to amplify the *ACE2* gene from population A05 and the *CTS1* gene from population F05. To amplify the *ACE2* gene, we used the forward primer 5′-ACACTGAACCCATCCACATG-3′ and the reverse primer 5′-TCCGCATGGGCAGATGTATT-3′. To amplify the *CTS1* gene, we used the forward primer 5′-GCAGTGTGAGTGGTCAATTC-3′ and the reverse primer 5′-TAGCTGTTTGAATTGGGGCC-3′. These samples were sent for Sanger sequencing (GenScript) using the gene-specific forward primers above, except for the *CTS1* gene in population A07, which was sequenced using the reverse primer. The frequency of the evolved mutation is reported as the frequency of the peak height as measured in 4peaks (Nucleobytes).

Reconstruction experiments. Reconstruction of evolved mutations in the ancestral backgrounds was performed as described above. Fitness of each of these twelve total samples was measured via competitive fitness assays across seven replicates along with control haploids and diploids with no mutations, all against the appropriate fluorescent ancestor strains.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. The sequences reported in this paper have been deposited in the BioProject database (accession no. PRJNA418180). All strains and reagents are available from the corresponding author upon request.

Received: 4 August 2017; Accepted: 14 February 2018;
Published online: 26 March 2018

References

1. Hummer, K. E., Nathewet, P. & Yanagi, T. Decaploidy in *Fragaria iturupensis* (Rosaceae). *Am. J. Bot.* **96**, 713–716 (2009).
2. Goodenough, U. & Heitman, J. Origins of eukaryotic sexual reproduction. *Cold Spring Harb. Perspect. Biol.* **6**, a016154 (2014).
3. Orr, H. A. & Otto, S. P. Does diploidy increase the rate of adaptation? *Genetics* **136**, 1475–1480 (1994).
4. Haldane, J. B. S. A mathematical theory of natural and artificial selection, part V: selection and mutation. *Proc. Camb. Philos. Soc.* **23**, 838–844 (1927).
5. de Godoy, L. M. et al. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254 (2008).
6. Galitski, T., Saldanha, A. J., Styles, C. A., Lander, E. S. & Fink, G. R. Ploidy regulation of gene expression. *Science* **285**, 251–254 (1999).
7. Paquin, C. & Adams, J. Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature* **302**, 495–500 (1983).
8. Anderson, J. B., Sirjusingh, C. & Ricker, N. Haploidy, diploidy and evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. *Genetics* **168**, 1915–1923 (2004).
9. Gerstein, A. C., Cleathero, L. A., Mandegar, M. A. & Otto, S. P. Haploids adapt faster than diploids across a range of environments. *J. Evol. Biol.* **24**, 531–540 (2011).
10. Selmecki, A. M. et al. Polyploidy can drive rapid adaptation in yeast. *Nature* **519**, 349–352 (2015).
11. Zeyl, C., Vanderford, T. & Carter, M. An evolutionary advantage of haploidy in large yeast populations. *Science* **299**, 555–558 (2003).
12. Kryazhimskiy, S., Rice, D. P., Jerison, E. R. & Desai, M. M. Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014).
13. Kvitek, D. J. & Sherlock, G. Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLoS Genet.* **9**, e1003972 (2013).
14. Lang, G. I. et al. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013).
15. Tenaillon, O. et al. The molecular diversity of adaptive convergence. *Science* **335**, 457–461 (2012).
16. Scott, A. L., Richmond, P. A., Dowell, R. & Selmecki, A. M. The influence of polyploidy on the evolution of yeast grown in a sub-optimal carbon source. *Mol. Biol. Evol.* **34**, 2690–2703 (2017).
17. Gerstein, A. C., Kuzmin, A. & Otto, S. P. Loss-of-heterozygosity facilitates passage through Haldane's sieve for *Saccharomyces cerevisiae* undergoing adaptation. *Nat. Commun.* **5**, 3819 (2014).

18. Sellis, D., Callahan, B. J., Petrov, D. A. & Messer, P. W. Heterozygote advantage as a natural consequence of adaptation in diploids. *Proc. Natl Acad. Sci. USA* **108**, 20666–20671 (2011).
19. Sellis, D., Kvitek, D. J., Dunn, B., Sherlock, G. & Petrov, D. A. Heterozygote advantage is a common outcome of adaptation in *Saccharomyces cerevisiae*. *Genetics* **203**, 1401–1413 (2016).
20. Deutschbauer, A. M. et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. *Genetics* **169**, 1915–1925 (2005).
21. Wiser, M. J., Ribeck, N. & Lenski, R. E. Long-term dynamics of adaptation in asexual populations. *Science* **342**, 1364–1367 (2013).
22. Wenger, J. W. et al. Hunger artists: yeast adapted to carbon limitation show trade-offs under carbon sufficiency. *PLoS Genet.* **7**, e1002202 (2011).
23. Keil, R. L. & Roeder, G. S. *cis*-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of *S. cerevisiae*. *Cell* **39**, 377–386 (1984).
24. Magwene, P. M. et al. Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **108**, 1987–1992 (2011).
25. Mandegar, M. A. & Otto, S. P. Mitotic recombination counteracts the benefits of genetic segregation. *Proc. R. Soc. B* **274**, 1301–1307 (2007).
26. Koschwanez, J. H., Foster, K. R. & Murray, A. W. Sucrose utilization in budding yeast as a model for the origin of undifferentiated multicellularity. *PLoS Biol.* **9**, e1001122 (2011).
27. Lang, G. I., Botstein, D. & Desai, M. M. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics* **188**, 647–661 (2011).
28. Oud, B. et al. Genome duplication and mutations in *ACE2* cause multicellular, fast-sedimenting phenotypes in evolved *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **110**, E4223–E4231 (2013).
29. Ratcliff, W. C., Denison, R. F., Borrello, M. & Travisano, M. Experimental evolution of multicellularity. *Proc. Natl Acad. Sci. USA* **109**, 1595–1600 (2012).
30. Kraus, E., Leung, W. Y. & Haber, J. E. Break-induced replication: a review and an example in budding yeast. *Proc. Natl Acad. Sci. USA* **98**, 8255–8262 (2001).
31. Buskirk, S. W., Peace, R. E. & Lang, G. I. Hitchhiking and epistasis give rise to cohort dynamics in adapting populations. *Proc. Natl Acad. Sci. USA* **114**, 8330–8335 (2017).
32. Mans, R. et al. CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **15**, fov004 (2015).
33. Baym, M. et al. Inexpensive multiplexed library preparation for megabase-sized genomes. *PloS ONE* **10**, e0128036 (2015).
34. Matheson, K., Parsons, L. & Gammie, A. Whole-genome sequence and variant analysis of W303, a widely-used strain of *Saccharomyces cerevisiae*. *G3 (Bethesda)* **7**, 2219–2226 (2017).

Acknowledgements

We thank M. Remillard (Princeton University) for providing strains; A. Nguyen and M. Desai (Harvard University) for providing the plasmid with mating-type-specific selectable markers; A. Selmecki, K. Fisher and R. Vignogna for their comments on the manuscript. This work was supported by the Charles E. Kaufman Foundation of The Pittsburgh Foundation.

Author contributions

G.I.L. designed the project. D.A.M. performed the experiments. S.W.B. and D.A.M. performed the sequencing analysis pipeline. D.A.M. and G.I.L. analysed the data and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-018-0503-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to G.I.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.