Trait Variation in Yeast Is Defined by Population History

Jonas Warringer[1,2]*, Enikö Zörgö[2], Francisco A. Cubillos[3], Amin Zia[4], Arne Gjuvsland[5], Jared T. Simpson[6], Annabelle Forsmark[1], Richard Durbin[6], Stig W. Omholt[2], Edward J. Louis[3], Gianni Liti[3], Alan Moses[4], Anders Blomberg[1]

1 Department of Cell and Molecular Biology, University of Gothenburg, Gothenburg, Sweden, 2 Centre for Integrative Genetics (CIGENE), Animal and Aquacultural Sciences, Norwegian University of Life Sciences (UMB), Ås, Norway, 3 Centre for Genetics and Genomics, Queen's Medical Centre, University of Nottingham, Nottingham, United Kingdom, 4 Department of Cell and Systems Biology, University of Toronto, Toronto, Canada, 5 Centre for Integrative Genetics (CIGENE), Mathematical Sciences and Technology, Norwegian University of Life Sciences (UMB), Ås, Norway, 6 Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

Abstract

A fundamental goal in biology is to achieve a mechanistic understanding of how and to what extent ecological variation imposes selection for distinct traits and favors the fixation of specific genetic variants. Key to such an understanding is the detailed mapping of the natural genomic and phenomic space and a bridging of the gap that separates these worlds. Here we chart a high-resolution map of natural trait variation in one of the most important genetic model organisms, the budding yeast *Saccharomyces cerevisiae*, and its closest wild relatives and trace the genetic basis and timing of major phenotype changing events in its recent history. We show that natural trait variation in *S. cerevisiae* exceeds that of its relatives, despite limited genetic variation, and follows the population history rather than the source environment. In particular, the West African population is phenotypically unique, with an extreme abundance of low-performance alleles, notably a premature translational termination signal in *GAL3* that cause inability to utilize galactose. Our observations suggest that many *S. cerevisiae* traits may be the consequence of genetic drift rather than selection, in line with the assumption that natural yeast lineages are remnants of recent population bottlenecks. Disconcertingly, the universal type strain S288C was found to be highly atypical, highlighting the danger of extrapolating gene-trait connections obtained in mosaic, lab-domesticated lineages to the species as a whole. Overall, this study represents a step towards an in-depth understanding of the causal relationship between co-variation in ecology, selection pressure, natural traits, molecular mechanism, and alleles in a key model organism.

Introduction

An overall aim in modern biology is to achieve an in-depth understanding of how selection for certain traits in the context of an organism's ecology favors specific mechanistic changes in the cell machinery and establish distinct genetic variants within populations. The budding yeast *Saccharomyces cerevisiae* and its closest relatives are expected to become the first eukaryotic organisms for which such a coherent understanding of causal relationships is to be achieved. This is primarily thanks to the uniquely detailed molecular knowledge that have been accumulated on this organism's internal cellular machinery, which in turn is due to the ease with which artificial genetic manipulation can be conducted [1]. Through the establishment of vast collections of mutants carrying single and double gene knockouts [2,3], temperature sensitive alleles [4] and promoter constructs that allow controlled gene expression [5–7], budding yeast has become the flagship of reverse genetics. However, the fact that the bulk of our knowledge on yeast gene-trait relationships derives from studies on artificial gene constructs in a few genetically mosaic and partially lab-domesticated strains is cause for concern. Mosaic lab

strains constitute artificial combinations of alleles that never jointly have been exposed to natural selective pressure and thus poorly reflect the natural state of the species. In addition, individual gene knockouts only to a very limited extent reflect the bulk of natural genetic variation which arises in the form of promoter, missense and gene duplication mutations rather than complete gene loss. Awareness of the limitations of artificial variation in lab domesticates has motivated an increasing number of endeavors focusing on the genetic basis for natural trait variation [8–13]. However, the lack of coherent charts of natural trait variation in *S. cerevisiae* has impeded a deeper understanding of the relation between ecology, selective pressure, traits and molecular mechanisms in this key model organism. In its natural ecological setting, baker's yeast has a peculiar life history dominated by clonal reproduction; it only completes one meiotic cycle for every 1000 mitotic divisions and 99% of these sexual cycles correspond to self-fertilization [14,15]. In fact, it has been suggested that natural yeasts are remnants of repeated population bottlenecks in essentially clonal lineages [1]. Such a life history dominated by mitotic proliferation implies a strong evolutionary influence of genetic drift and predicts that trait variation is largely defined by

Author Summary

An overall aim in modern biology is to achieve an in-depth understanding of an organism's physiology in the context of its ecology and historic selective pressures that have been acting on its genome. The baker's yeast, *Saccharomyces cerevisiae*, has a peculiar life history completely dominated by clonal reproduction and self-fertilization, prompting the suggestion that natural yeasts are remnants of repeated population bottlenecks in essentially clonal lineages. Such a life history dominated by mitotic proliferation purports a strong evolutionary influence of genetic drift and predicts trait variation to be high and largely defined by the genetic history of each population. Here we chart a highly resolved map of natural trait variation in *S. cerevisiae* and its closest non-domesticated relative, *Saccharomyces paradoxus,* and confirm this prediction. We found that trait variation in budding yeast is indeed high and largely defined by population rather than source environment. In particular, the West African population was found to be phenotypically unique with an extreme abundance of low-performance alleles. Our findings support the idea of population bottlenecks in the recent yeast evolutionary history and a large influence of genetic drift.

the genetic history of each population. To test this prediction, we charted a highly resolved map of natural trait variation in *S. cerevisiae* and its closest non-domesticated relative *Saccharomyces paradoxus*, species that are separated by about 2 billion generations [1]. This design allowed the quantification of trait variation in relation to source environment and genetic history and the tracing of the molecular basis for between population variations down to individual alleles. We found that trait variation in budding yeast is largely defined by population rather than source environment, in support of recent population bottlenecks and a large influence of genetic drift.

Results

Co-variation of the rate and efficiency of proliferation in natural yeast isolates

To exhaustively survey natural trait variation in the partially domesticated *S. cerevisiae* and its closest wild relative *S. paradoxus*, we compared 39 sequenced *S. cerevisiae* isolates from a wide variety of geographic locations and sources to an equal number of sequenced *S. paradoxus* isolates. These strains represented all known populations in *S. cerevisiae* (Malaysian, West African, North American, European, Sake) and *S. paradoxus* (American, Far East, European) [16]. To provide an overview of the complete *Saccharomyces sensu stricto* clade, a smaller number of isolates from *Saccharomyces kudriavzevii, Saccharomyces bayanus, Saccharomyces mikatae* and the recently identified *Saccharomyces arboricolus* [17] was also included. Isolates were subjected to high-resolution quantification of mitotic proliferative ability across 200 environments; these were selected to represent components of natural yeast habitats, such as commonly found carbon and nitrogen sources, plant and microbial toxins and metabolites, and shifting availability of vitamins and minerals (Table S1). From high density mitotic growth curves, the fitness components lag of proliferation, rate of proliferation (population doubling time) and efficiency of proliferation (population density change) were extracted, providing 600 distinct measures of organism-environment interactions (Figure 1A, Table S2) [18,19]. Overall, we found the rate and efficiency of

yeast mitotic proliferation to be strongly correlated (r = 0.77) (Figure 1B). For 84% of the environments, the Pearson correlation exceeded 0.5, and for only two milieus, proliferation in absence of either copper or magnesium, was the correlation negative. This refutes assumptions of an evolutionary trade-off between rate and efficiency [20] but is in line with reports on artificial loss-of-function mutants in *S. cerevisiae* [21]. No evidence of a general adherence to a 1:1 correlation between rate and the square root of the efficiency was found, which would be expected if the rate was maximized in each environment and solely restricted by biomass yield [22]. Rather, the slope of the correlation was highly environment dependent (Figure S1). Taken together, this suggests that the rate and efficiency of proliferation in yeast have similar underlying genetic structures. In contrast, the time to initiate proliferation showed clear evidence of being physiologically and evolutionary distinct since no correlation (r<0.02) was observed to any of the two other variables (Figure 1B).

Two phenetic groups in the *Saccharomyces sensu stricto* clade

From species trait averages, it was evident that the major phenetic divide within *Saccharomyces sensu stricto* is the clear separation between *S. paradoxus, S. cerevisiae* and *S. mikatae* on the one hand and *S. kudriavzevii, S. bayanus* and *S. arboricolus* on the other - with the non-*sensu stricto S. castellii* as an expected outlier (Figure 1C). These two clades differed significantly (clade average; Bonferroni corrected Student's t-test, p<0.1) over 42% of all traits. Remarkably, 99% of these deviating traits corresponded to inferior performance of the *S. kudriavzevii, S. bayanus* and *S. arboricolus* group (Figure 1D). Most (62%) of these low performance traits remained significant (clade average; Bonferroni corrected Student's t-test, p<0.1) when normalizing for proliferation in basal conditions, suggesting that they are not the exclusive consequence of a general proliferation defect (Figure S2). Interestingly, although the deficiencies encompassed all types of ecologically relevant features (Figure 1E), they were clearly underrepresented among carbon sources (Fisher's exact test, p = 7E-8). This suggests early optimization of proliferation on a wide range of carbon substrates in the common ancestor of *sensu stricto* and that this selective pressure has been mostly maintained through the recent evolutionary history. Among the traits with the largest phenotypic variances between *sensu stricto* species was tolerance to citric acid (a dominant fruit organic acid), growth at high temperatures and growth in synthetic wine must, suggesting that these traits may have corresponded to important habitat differences during species divergence (Figure S3).

Between *S. kudriavzevii, S. bayanus* and *S. arboricolus*, overall pairwise trait similarity was extremely high (Pearson correlation, r = 0.82–0.85) despite that these species diverge genetically as much from each other as from the remaining *sensu stricto* species [16,23]. Hence, surprisingly much of the accumulated sequence variation in this clade appears to be neutral for the studied traits. Given that these species are mostly allopatric [24], the most attractive interpretation of their trait similarity is occupation of similar habitats in different geographic locations. *S. paradoxus, S. cerevisiae* and *S. mikatae* are genetically more similar; nevertheless, trait divergence between these species was much higher (Pearson correlation, r = 0.55–0.69). In fact, *S. cerevisiae* and *S. paradoxus* diverged significantly (species average; Bonferroni corrected Student's t-test, p<0.1) for more than 18% of all traits, in contrast to conventional wisdom that marks the two species as phenotypically indistinguishable [25]. The vast majority of environment-dependent deviations between the two species placed *S. cerevisiae* as a superior performer (Figure 1F), even though no difference in
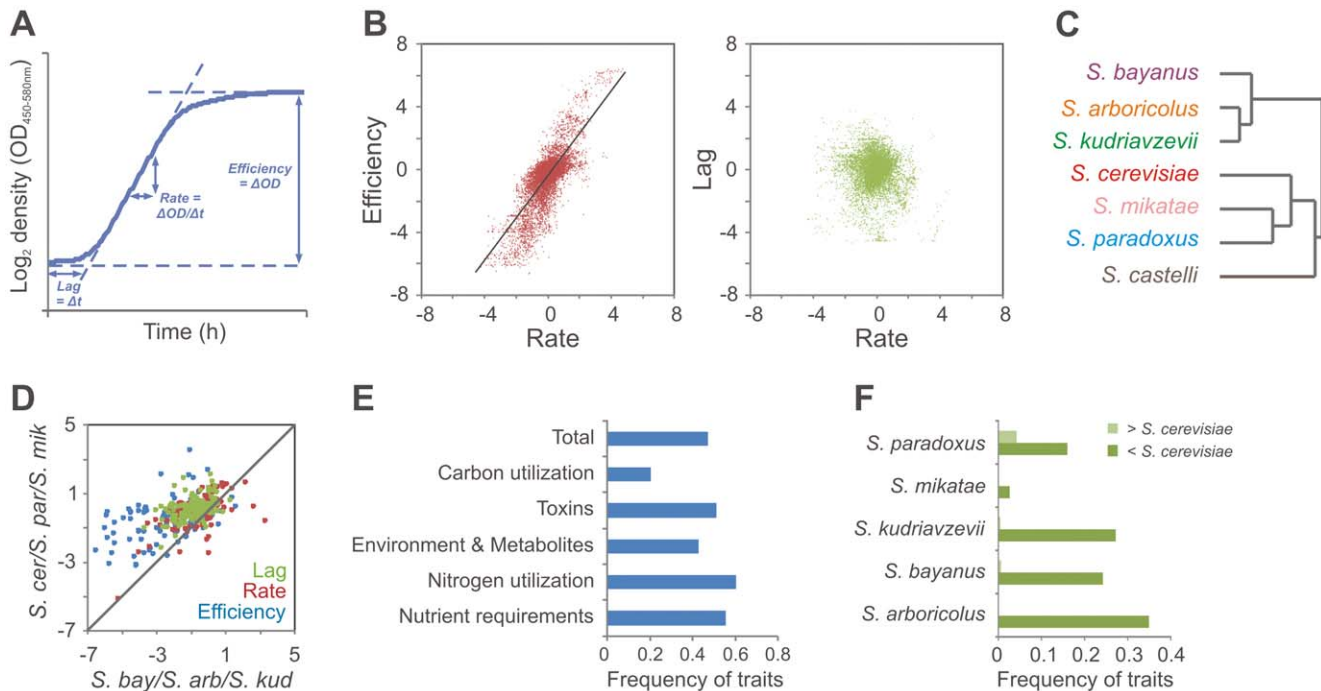
Figure 1. Trait variation between *Saccharomyces sensu stricto* species. A) The proliferative lag (time to initiate proliferation), proliferative rate (population doubling time) and proliferative efficiency (change in population density) were extracted from high density growth curves (n = 2) of 86 *Saccharomyces sensu stricto* isolates over ~200 environments. B) The fitness components proliferative lag, rate and efficiency were compared over all isolates and environments. Black line depicts linear correlation between the proliferative rate and efficiency (Pearson, r = 0.77). C) Hierarchical clustering of species trait averages. D) Comparing trait averages of the *S. bayanus/S. kudriavzevii/S. arboricolus* clade versus the *S. cerevisiae/S. paradoxus/S. mikatae* clade over all traits. Line indicates 1:1 correlation. E) Frequency of significant (Student's t-test, Bonferroni correction, p<0.1) trait differences between the *S. bayanus/S. kudriavzevii/S. arboricolus* clade and the *S. cerevisiae/S. paradoxus/S. mikatae* clade, considering each class of environmental variation individually. F) Frequency of significant (Student's t-test, Bonferroni correction, p<0.1) trait differences for each species in comparison to *S. cerevisiae*.
doi:10.1371/journal.pgen.1002111.g001

proliferation was observed in basal conditions (Figure S4A). This superior performance included better ability to utilize the sugar maltose and higher tolerance to both temperature and a wide range of plant and microbial toxins (Figure S4B). The two unique examples where *S. paradoxus* performed better than *S. cerevisiae* were superior utilization of the sugar mannitol and extreme tolerance to oxalic acid, an organic acid prevalent in oak bark, a typical *S. paradoxus* habitat (Figure S4C, D). However, it should be noted that medium and test conditions have been tailored to *S. cerevisiae* and thus the generally inferior performance of *S. paradoxus* may reflect overall differences in physiology that confer media or test condition dependent defects manifested only under stress.

Trait variation within *S. cerevisiae* is defined by the population history

S. cerevisiae differed from *S. paradoxus* not only in its on average better tolerance to environmental stress but also in the degree of intra-species trait variability which exceeded variance in *S. paradoxus* by more than 60% despite lower genetic diversity (Figure S5A, S5B). The difference was consistent over the tested populations (Figure S5C). However, it is unknown to what extent we have sampled the existing global variation in each species and it cannot be excluded that future sampling of *S. paradoxus* will increase its trait variability. Utilization of less common carbon substrates (melibiose and maltose), growth in absence of particular vitamins (biotin and pantothenate) and tolerance to high concentrations of certain metabolites (ethanol, methanol and formaldehyde) showed the strongest variations within *S. cerevisiae*,

suggesting niche variations with a large impact on trait divergence (Figure S6). The overall correlation between genetic and phenotypic similarity within *S. cerevisiae* was substantial (Pearson r = 0.52) and even higher (Pearson r = 0.66) when excluding mosaic isolates for which genetic distances have no straightforward evolutionary interpretation. Hierarchical clustering of all yeast isolates based on the complete array of traits also provided groups that essentially followed species and population boundaries, with mosaic traits resembling those of the parent population donating the major part of the genome (Figure 2, Figure S7A–S7C). In fact, 45% of all traits were distinct for one individual population (FDR = 2%). In contrast, only 1.2% of traits were distinct (FDR = 2%) for an individual source, i.e. neither clinical isolates nor industrial strains from processes such as baking or fermentation showed trait coherence within groups (Figure 2). Principal component analysis confirmed the stronger influence of population than source on *S. cerevisiae* traits (Figure S8). Notably, commonly used *S. cerevisiae* lab strains were extremely diverse, reflecting that lab domestication either has proceeded along very different routes or has had little overall influence (Figure S9). The few source-linked phenotypes figured among wild *S. cerevisiae* isolates, which showed low respiratory capability (glycerol, ethanol and arabinose utilization) and low tolerance to copper, suggesting that vigorous respiration and copper resistance may have been principal domestication traits (Figure S10). Disconcertingly, the yeast universal type strain, S288C, constituted the most atypical *S. cerevisiae* strain in the screen (Figure 3), diverging strongly from the average *S. cerevisiae* strain by e.g. showing superior proliferative rate
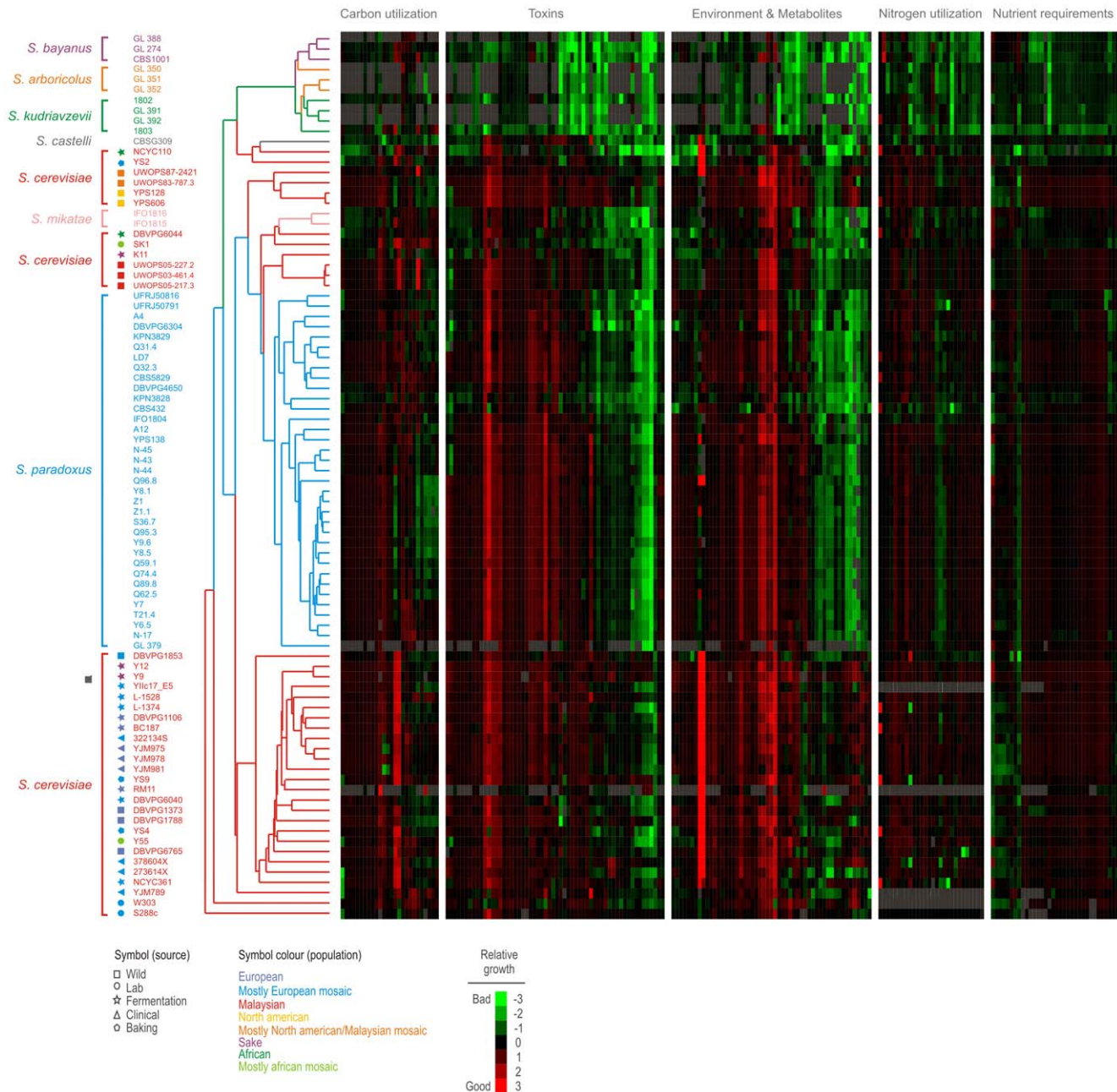
Figure 2. Trait variations within *Saccharomyces sensu stricto* species. Hierarchical clustering of *Saccharomyces sensu stricto* isolates based on trait profiles over 600 traits (all proliferation variables) performed using a centered Pearson correlation metric and average linkage mapping. Species are indicated by line color. For *S. cerevisiae* isolates, source habitats (shape) and populations (color) are indicated with symbols. Heat map depicts the relative proliferation rate (Log$_2$ [BY4741/strain]); relative proliferation lag and efficiency are given in Figure S7. Green = inferior proliferation, red = superior proliferation, black = BY4741 proliferation, grey = missing data.
doi:10.1371/journal.pgen.1002111.g002

on rich medium, superb maltose utilization but lower than average ability to tolerate very high ethanol concentrations (Figure S11). Given the wide-spread use of S288C as a norm for *S. cerevisiae*, this phenotypic uniqueness is a concern.

The West African *S. cerevisiae* population is phenotypically unique

From an evolutionary perspective, the most interesting feature of an organism-wide phenotype map is the variation between phylogenetically distinct groups of individuals, i.e. in *S. cerevisiae* the

Malaysian, West African, Sake, European and North American populations. Mapping our phenotypic information on the recently established population genomics tree [16], we can reconstruct the phenotypic history of *S. cerevisiae*. The most dramatic event in the *S. cerevisiae* trait history relates to the West African population, which deviated significantly (FDR = 2%) from other *S. cerevisiae* clean populations over 35% of all traits (Figure 4). Remarkably, all but one of the West African population traits represented severe reproductive deficiencies, including extremely poor utilization of galactose and hypersensitivity to high temperatures (Figure S12).
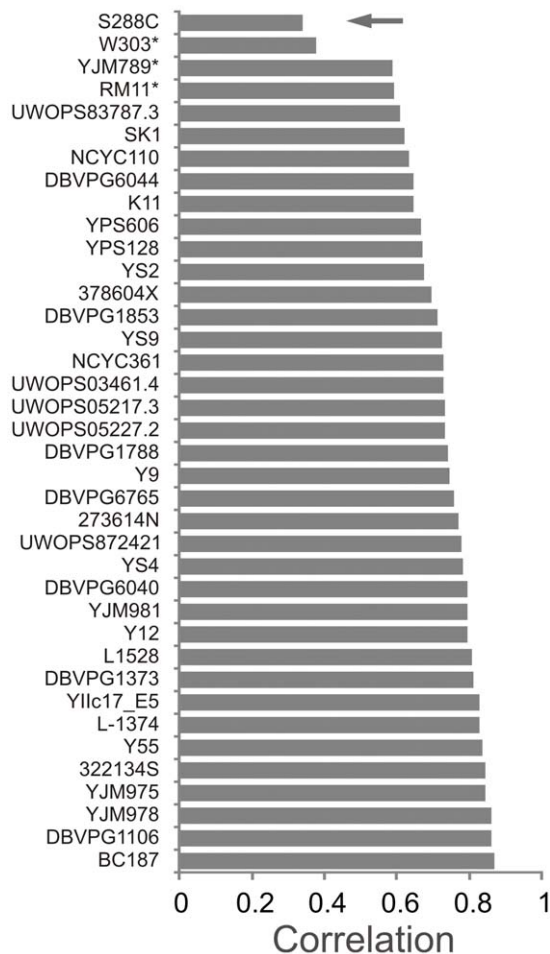
Figure 3. The *S. cerevisiae* type strain S288C is a phenotypic extreme. A *S. cerevisiae* mean trait profile was calculated and the similarity (Pearson correlation) between the trait profile of each individual *S. cerevisiae* isolate and the *S. cerevisiae* mean trait profile was obtained. Strains were ranked according to degree of similarity. The universal reference strain S288C (arrow) was found to be a phenotypic extreme. * = Isolates with auxotrophic markers; these were excluded from calculations of the mean trait profile.
doi:10.1371/journal.pgen.1002111.g003

This suggests extensive genome degeneration following relaxation of ancestral selective pressures. Our previous linkage analysis on crosses of a West African isolate to the other clean lineages [26] showed that its many proliferative deficiencies were due to environment-specific large effect QTLs widely dispersed over the genome (Figure S13). A less pronounced phenotypic burst distinguishes the branching-off of the European population (Figure 4). The European population traits encompassed phenotypes that may reflect man-directed selection for industrial production, such as high respiratory capability (ethanol growth), good proliferation in synthetic wine must, tolerance to tartaric acid (highly concentrated in grapes) and tolerance to copper, the latter presumably deriving from early brewery cultivation in copper containers and anti-parasitic spraying of vineyards with copper sulphate [27] (Figure S12). Not surprisingly, industrial yeasts are mostly European isolates or derivates thereof (Table S1). The European strains were also uniquely tolerant to Na^+ and Li^+, a trait which because of pleiotropy of the causal locus may not reflect adaptation to a saline niche (see below). The other clean populations were phenotypically less distinct (Figure 4). The

Malaysian population displayed superb utilization of melibiose and mannitol, storage carbohydrates common in many tropical plants [28] (Figure S12), whereas the North American oak isolates were essentially unable to metabolize maltose and showed supreme tolerance to oxalic acid. The phenotypically more diverse Sake population, where the K11 isolate deviates both genetically and phenotypically from Y9 and Y12 (Figure 2), featured good utilization of glycerol and excellent proliferation in absence of biotin, the latter of which has been recently reported [29].

Trait variation patterns differ between *S. paradoxus* and *S. cerevisiae*

As for *S. cerevisiae*, trait variation in *S. paradoxus* agreed well with recently reported population boundaries based on genome sequences [16]. However, although unique phenotypes were observed for each of the American, European and Far East populations, no phenotypic event as dramatic as the one distinguishing the West African *S. cerevisiae* population was detected in *S. paradoxus* (Figure S14). Moreover, traits that were specific for individual *S. paradoxus* populations did not agree with traits that were specific for individual *S. cerevisiae* populations (Fisher's exact test, p = 0.11). For example, whereas differences in the ability to utilize different carbon sources were prevalent between individual isolates in *S. cerevisiae*, such variations were virtually absent in *S. paradoxus* (Figures S11, S14). In contrast, *S. paradoxus* populations varied more between each other in the ability to tolerate the absence of important vitamins and minerals. Overall, the American population, which covers a vast geographical latitude range, was the phenotypically most distinct, featuring superior tolerance to the toxin canavanine and the metabolite selenomethionine, both affecting the protein folding machinery, but reduced tolerance to superoxide anions as generated by the insecticide paraquat. The genetically distinct South American sub-population, previously regarded as a separate species and referred to as *S. cariocanus* and here represented by the isolates UFRJ50791 and UFRJ50816, was unique also phenotypically, suggesting either adaptation to a different environmental niche or genetic drift due to genetic isolation from other American isolates (Figure 2). The European population primarily diverged in its metal ion tolerance profile, with low tolerance to lithium and cadmium but high tolerance to arsenic in the form of arsenite, which may reflect soil composition. The Far East population had extremely low performance in absence of biotin, suggesting occupation of a biotin rich habitat (Figure S14). Interestingly, neither traits separating the *S. paradoxus* populations, nor traits separating the *S. cerevisiae* populations, overlapped more than chance expectation (Fisher's exact test, p>0.12) with traits separating the two species from each other, implying that different selective pressures have driven trait divergence at different time points in the history of these two species.

Genetic basis for population specific traits in *S. cerevisiae*

The considerable divergence between *S. paradoxus* populations resulted in that offspring from crosses over population borders suffered drastic reductions in proliferative ability, as exemplified by on average 42% slower rate of proliferation in optimal conditions of F1 recombinants deriving from a cross between the American YPS138 and the European CBS432. This population incompatibility prevented linkage analysis of population specific traits in *S. paradoxus*. Such limitations do not hamper genetic dissection of *S. cerevisiae* population traits, with the exception of the Malaysian population which is effectively reproductively isolated [28]. However, the number of SNPs detected in the 36 partially sequenced *S. cerevisiae* isolates exceeds 230.000 [16], preventing a
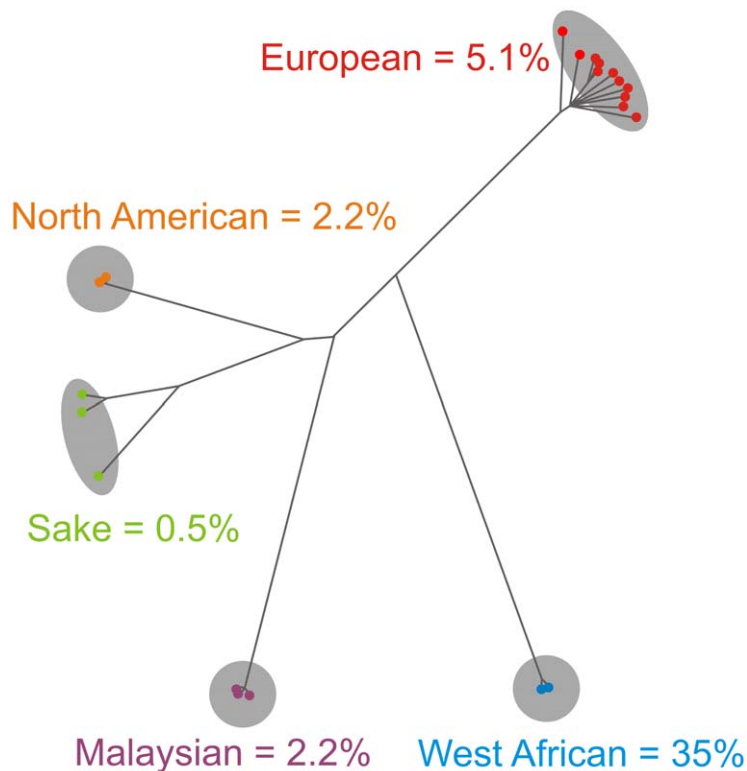
Figure 4. Reconstructing the phenotypic history of *S. cerevisiae*. The frequency of population specific traits for each *S. cerevisiae* population mapped onto a recently established population genomics tree based on low coverage genome sequence data [16]. Population specific traits were defined as traits for which the performance of one population deviated significantly from the performance of isolates in all other populations (FDR = 2%), Percentages indicate frequency of population specific traits. Total number of population specific traits: West African = 190, European = 30, North American = 13, Malaysian = 13 and Sake = 3.
doi:10.1371/journal.pgen.1002111.g004

full scale genome-wide association analysis over all genetic variations. To circumvent this "large number of hypotheses, small number of samples" situation, we first performed a genome-wide association analysis limited to genetic variation suspected of having strong phenotypic consequences: i.e. gene presence variations, non-sense mutations (premature stop codons), and gene Copy Number Variations (CNV) (Tables S3, S4, S5). The scored gene-phenotype associations were cross-referenced to a global linkage map covering crosses between representatives of the clean lineages examined in 24 environments where strong population specific phenotypes were scored [26]. Where associations agreed with QTL regions, candidate genetic variations were retained as high confidence hits.

One of the most distinct West African traits was the extremely poor utilization of galactose, a sugar that is ubiquitous in natural gum and plant mucilage of e.g. cactus, but not present in an accessible form in grapes or other fruits. The poor utilization of galactose in the West African strain associated to a premature stop codon in the transcriptional regulator *GAL3*, which acts as a positive regulator to control galactose-induced expression of galactose utilization enzymes [30] (Figure 5A, 5B). Linkage analysis of the galactose utilization ability in crosses involving the West African DBVPG6044, as well as poor proliferation of the hemizygote BY4741 (*gal3Δ*)×DBVPG6044 on galactose, confirmed this association (Figure 5C, 5D). In *S. kudriavzevii* and in some non *sensu stricto* yeasts, loss of the complete galactose system has been observed [31,32]. Multiple mutations throughout the *GAL* system in West African strains, notably frameshifts in *GAL2* and *GAL4*, suggested a similar loss of selective constraints following emergence of the *GAL3* stop codon (Figure 5A). However,

unperturbed galactose growth of the hemizygotes (*galxΔ*)× DBVPG6044 showed that, except for *GAL3*, all *GAL* genes, including *GAL2* and *GAL4*, were fully functional in DBVPG6044 (Figure S15A, S15B). Interestingly, the *GAL3* stop codon was not found outside the West African population and the West African derived mosaics SK1 and Y55. Rather, the complete inability of some mosaic strains, such as 273614N, to utilize galactose was due to selective loss of *GAL1* (Figure 5E), whereas the severe galactose deficiency of S288C agreed with a well-known point mutation in the galactose permease *GAL2* [33]. Hence, inactivation of the galactose utilization pathway has occurred as multiple, independent events in the recent *S. cerevisiae* history.

The association analysis also showed that the ability to utilize the related, but rare, sugar melibiose, a disaccharide of galactose and glucose, was unique for the Malaysian lineage, and was completely associated with the presence of a single gene, the melibiase encoding *MEL1* (Figure 5F). The poor melibiose utilization in the West African population, which contains a *MEL1* with no non-synonymous variations but only performed marginally better than strains missing *MEL1*, was in this context surprising. The hemizygotes BY4741 (*galxΔ*)×DBVPG6044 showed normal melibiose proliferation, demonstrating that the melibiose utilization inability is not due to defects in the *GAL* system which is known to regulate *MEL1* expression [34] (Figure S15C). Interestingly, the *MEL1* of the West African and Malaysian populations have diverging evolutionary origins, with the West African *MEL1* deriving from a recent introgression of the *S. paradoxus MEL1* (Figure 5G), potentially explaining the inferior melibiose utilization of the West African population.
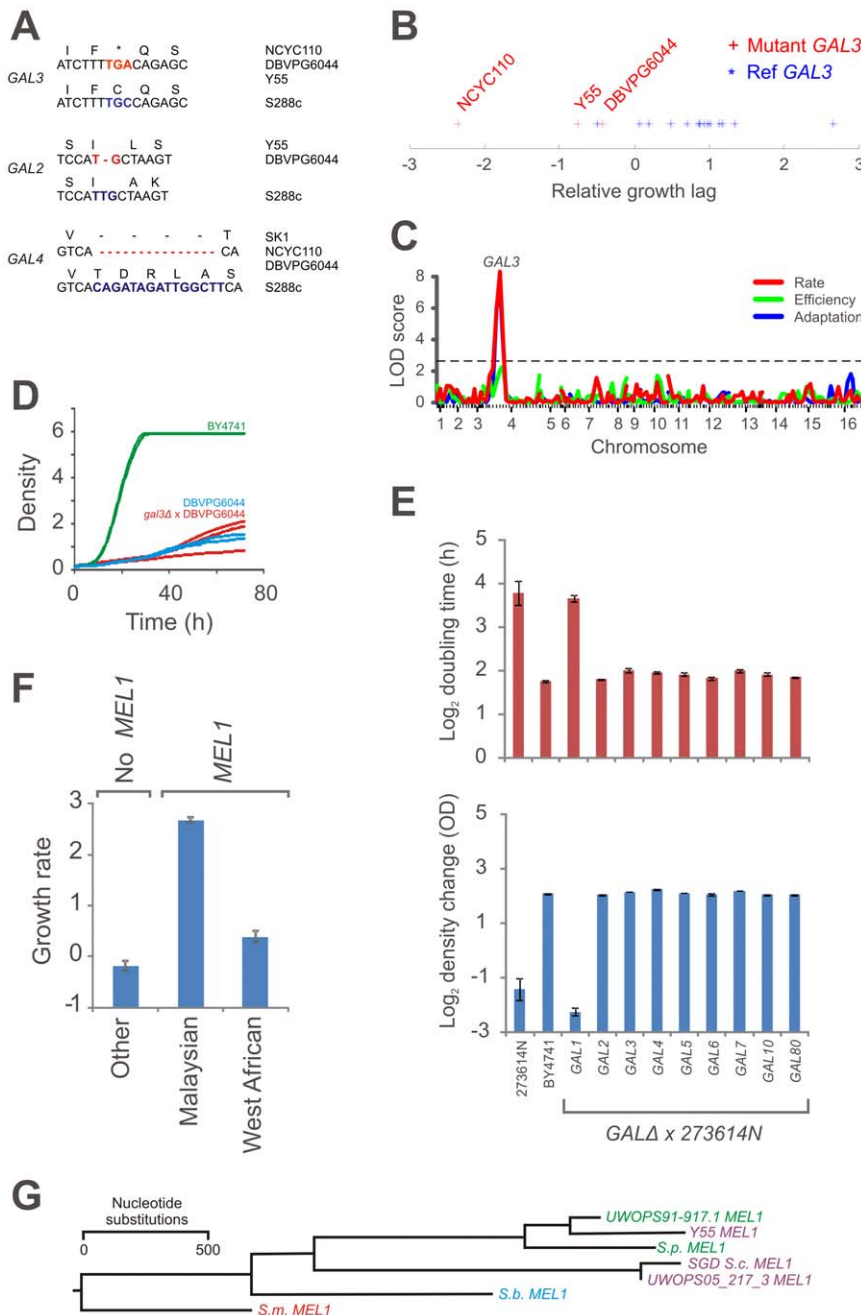
Figure 5. Recent, parallel inactivation of galactose utilization in multiple *S. cerevisiae* lineages. A) Potential loss-of-function polymorphisms in the galactose utilization pathway in the West African population; a premature stop codon in *GAL3* and out of frame deletions in *GAL2* and *GAL4*. B) A TGC→TGA (C→Stop) mutation in *GAL3* associates (Bonferroni corrected Student's t-test p<0.015 and Kolmygorov-Smirnov p<0.15) to the proliferative lag during initiation of galactose proliferation in natural *S. cerevisiae* strains. Three West African derived genomes contain the variant mutation, 18 other strains contain the reference sequence. C) Linkage analysis of a cross between the West African DBVPG6044 and the North American YPS128 supports a link between *GAL3* and poor galactose utilization. 96 haploid F1 offspring were genotyped at 130 chromosomal markers and the degree of galactose growth was determined. Chromosome numbers indicate centromere position and tick marks indicate marker position. D) Proliferation of hemizygote BY4741 (*gal3Δ*)×DBVPG6044 (n = 3), and the parentals DBVPG6044 (n = 2) and BY4741 (n = 2) using galactose as carbon source. E) Population doubling time (rate) and total change in population density (efficiency) of the hemizygote BY4741 (*galxΔ*)×273614N as compared to the parentals BY4741 and 273614 (n = 3) using galactose as carbon source. Error bars = standard errors. F) Average proliferative rates of strains in the Malaysian population (including the partially Malaysian UWOPS87-2421), the West African population (including the partially West African Y55) and all other *S. cerevisiae* strains, using melibiose as carbon source. Error bars = standard errors. G) Rooted N-J tree based on a multiple alignment of *MEL1* from the *S. cerevisiae* Malaysian (UWOPS05_217_3) and West African (Y55) populations, a *S. paradoxus* Hawaiian isolate (UWOPS91-917.1), the *S. cerevisiae MEL1* sequence stored in SGD, and previously reported *MEL1* sequences from *S. mikatae* (S.m), *S. bayanus* (S.b) and *S. paradoxus* (S.p) [53].
doi:10.1371/journal.pgen.1002111.g005

The drastically elevated copper tolerance of the European and Sake lineages associated strongly with a copy number variation of the copper binding metallothionein *CUP1* with limited contribution from other loci: the more *CUP1* genes, the higher the copper tolerance (Figure 6A). This association was confirmed by linkage analysis of crosses between populations with high and low performance on copper (Figure 6B). Tandem amplifications of *CUP1* is well known to mediate copper tolerance in adapting yeast populations during artificial selection for this trait [35]. The emergence of the *CUP1* CNV in both the European and Sake lineages, but not in other *S. cerevisiae* populations nor in *S. paradoxus*, suggested that the *CUP1* CNV may be a true case of convergent evolution due to man-directed selection for industrial production. Such a model postulates independent amplification events which most likely would result in variable amplification breakpoints. We compared amplification breakpoints in the Sake strain Y12 and the mostly European S288C and W303, and found the amplified *CUP1* segments in these lineages to be clearly distinct, providing

strong support for independent events and thus convergent evolution (Figure 6C, Figure S16). *S. paradoxus* isolates consistently showed low *CUP1* copy numbers and also low tolerance to copper.

The European population's extreme tolerance to Li$^+$ and Na$^+$ associated to a CNV of the *ENA* Na$^+$ exporter (Figure 7A), supporting our own (Figure 7B) and recently published linkage data [36]. Sequence comparison to *S. paradoxus* showed that the single *ENA* gene in the non-European *S. cerevisiae* lineages, recently found in a mosaic strain and referred to as *ENA6* [37], and the three *ENA* copies in the laboratory strain S288C, have diverging evolutionary origins. The phylogeny strongly supports a recent introgression of one of the two *S. paradoxus ENA* variants into the *S. cerevisiae* European lineage (Figure 7C, Figure S17), as suggested [38]. Transfer of *ENA6* to a European mosaic in which the three native *ENA* genes *(ENA1,2,5)* had been deleted clearly demonstrated the inferior performance of the *ENA6* allele. In fact, *ENA6* did not confer any tolerance to low concentrations of Na$^+$ or to Li$^+$, and at high Na$^+$ concentrations only alleviated the efficiency,
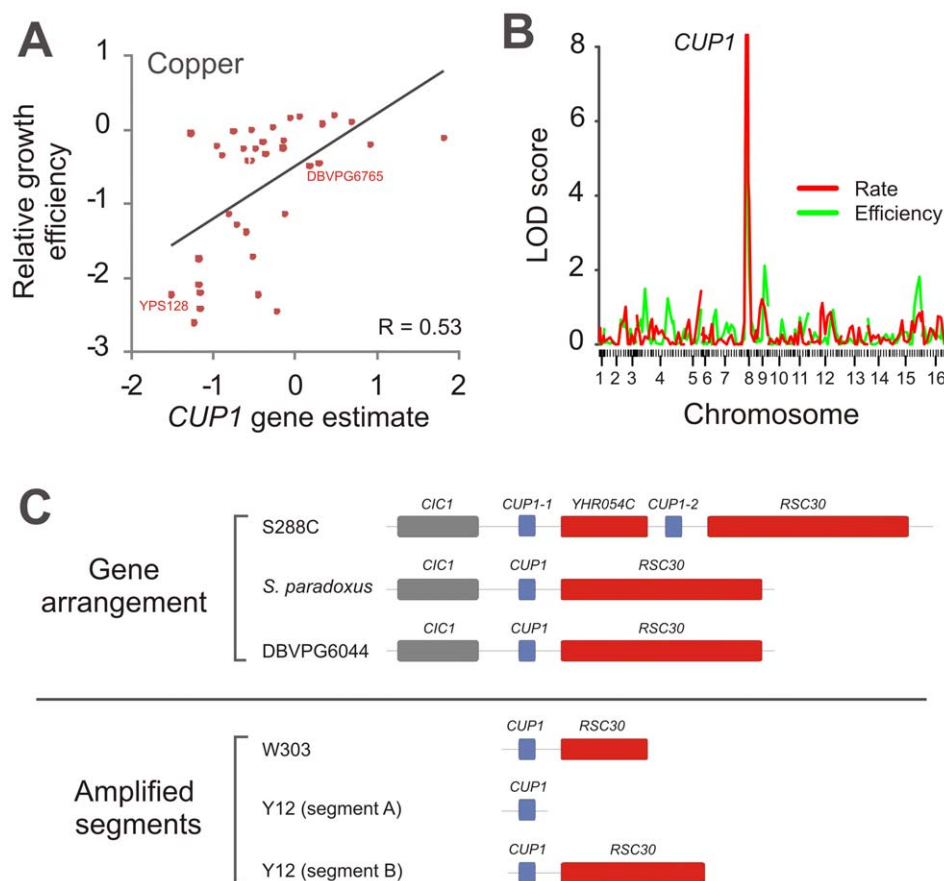

Figure 6. Parallel amplifications of *CUP1* in *S. cerevisiae* populations reflect convergent evolution for copper tolerance. A) A gene amplification of *CUP1* associates strongly to copper tolerance in *S. cerevisiae*. The proliferative efficiency of each *S. cerevisiae* isolate in presence of 0.75 mM CuCl$_2$ versus the *CUP1* gene number estimate for each strain is displayed. The gene number estimate was defined as log[(1+observed reads)/(1+expected reads)]. Linear correlation is displayed. B) To verify a causative link between the *CUP1* CNV and copper tolerance, the European DBVPG6765 was crossed to the North American YPS128, 96 haploid offspring were obtained after meiotic recombination and co-inheritance of the proliferation performance during 0.75 mM CuCl2 exposure and 130 chromosomal markers was investigated using linkage analysis [26]. Chromosome numbers indicate centromere position on each chromosome and tick marks marker position. C) The *CUP1* region in the West African DBVPG6044, the Sake Y12 and the European derived mosaic W303, was determined by high coverage (25–31×) sequencing. The gene arrangement in DBVPG6044 agreed with the *S. paradoxus* reference assembly and no amplifications were detected. For Y12 and W303, amplified segments, but not gene arrangements, could be unambiguously determined. Two independent segment amplifications were detected in Y12, encompassing only *CUP1* and *CUP1* plus 1624 bp of the neighboring *RSC30* respectively. One amplified segment, encompassing *CUP1* plus 1069 bp of *RSC30*, was detected in W303. Breakpoints were completely strain specific: no Y12 reads matched the W303 breakpoint and no W303 reads matched the Y12 breakpoints. The W303 amplification matched the known amplification in the S288C reference genome.
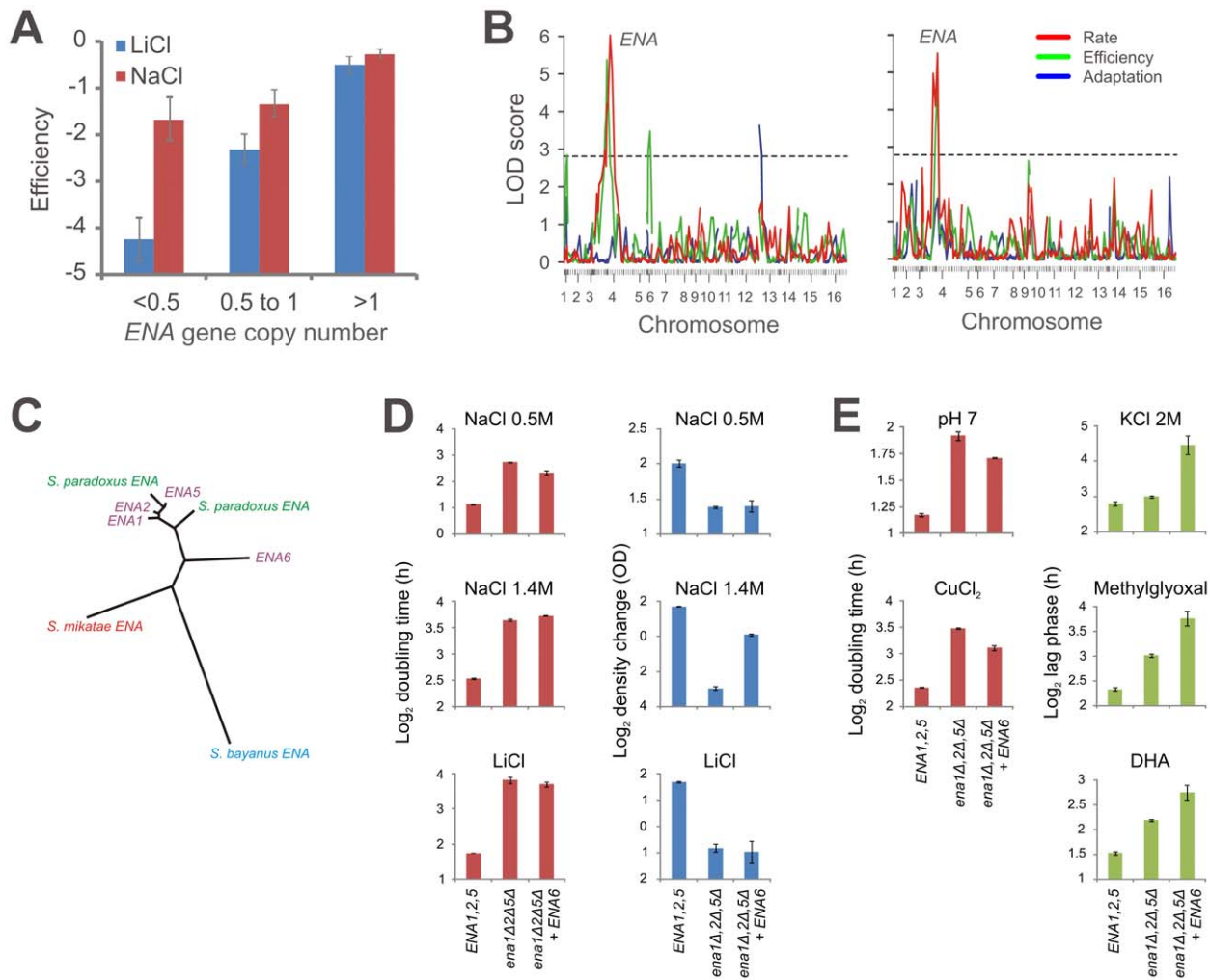doi:10.1371/journal.pgen.1002111.g006

Figure 7. An introgressed and serially amplified *S. paradoxus ENA* gene causes pleiotropy in *S. cerevisiae*. A) A copy number variation in *ENA* gene locus associates to Na$^+$ and Li$^+$ tolerance in natural *S. cerevisiae* isolates. Strains were divided into equally sized bins on the basis of an estimate of *ENA* copy numbers, log[(1+observed reads)/(1+expected reads)], and the average proliferative efficiency in presence of Na$^+$ and Li$^+$ was determined in each bin. Error bars = standard errors. B) Linkage analysis of crosses between representatives of natural yeast populations supports co-segregation of the marker closest to the *ENA* locus and proliferation during Li$^+$ (left panel) and Na$^+$ (right panel) exposure. Left panel: a cross between the West African DBVPG6044 and the European DBVPG6765, right panel: a cross between the Sake Y12 and the European DBVPG6765. For each cross, 96 haploid F1 offspring were obtained, genotyped for parental genotype at 130 chromosomal markers loci and the proliferative ability in presence of 1 M NaCl and 0.225 mM LiCl was determined. Chromosome numbers indicate centromere position and tick marks indicate marker position. C) Unrooted N-J tree based on a multiple alignment of *ENA1, 2* and *5* from the *S. cerevisiae* reference genome, *ENA6* obtained from the mosaic CEN.PK2 strain [37], and all *ENA* genes detected in *S. paradoxus*, *S. bayanus* and *S. mikatae*. D) *ENA6* from the West African derived SK1 was transferred to an *ENA* triple deletion (*ena1Δ2Δ5Δ*) in the mostly European BY4741. The proliferative rate (doubling time, h) and efficiency (density change, OD units) of the WT (*ENA1,2,5*), *ena1Δ2Δ5Δ*, and *ena1Δ2Δ5Δ+ENA6* strains (n = 4–8, error bars = standard errors) in 0.5 M NaCl, 1.4 M NaCl and 0.3 M LiCl were measured. E) The proliferative rate (doubling time, h) and lag (adaptation time, h) of the WT (*ENA1,2,5*), *ena1Δ2Δ5Δ*, and *ena1Δ2Δ5Δ+ENA6* strains (n = 4–8, error bars = standard errors) in pH 7, 1 mM CuCl$_2$, 2 M KCl, 10 mM methylglyoxal and 80 mM dihydoxyacetone (DHA).
doi:10.1371/journal.pgen.1002111.g007

but not the rate, proliferation defect (Figure 7D). However, the selective pressure maintaining the *ENA* introgression and copy number variation need not necessarily have been for Na$^+$ or Li$^+$ tolerance; both association and linkage analysis [26] connected the *ENA* CNV with multiple other traits. Poor performance of the *ENA1,2,5* triple deletion during exposure to K$^+$, Cu$^+$, alkali stress and the toxic metabolites methylglyoxal and DHA confirmed that the *ENA* introgressed from *S. paradoxus* confers a distinctly pleiotropic advantage (Figure 7E). The *ENA6* allele had either substantially lower or adverse effects on these traits, supporting the notion that it represents a selectively weaker allele. This degree of pleiotropy of the

recently emerged *ENA* amplification is surprising, given that adaptive mutations in pleiotropic genes are expected to be rare [39].

Hydroxyurea tolerance associated to an inactivating stop codon early in the reading frame of the functionally unknown *HUR1* in the West African and Malaysian populations (Figure S18). *HUR1* overlaps the ion exporting ATPase *PMR1* (coding on opposite strands) and deletion mutants of both *HUR1* and *PMR1* are highly hydroxyurea sensitive [40]. However, the early stop codon in *HUR1* with only a neutral amino acid change, F947P, in the poorly conserved C-terminal region of *PMR1* suggests that the genetic variation in *HUR1* is indeed the causative allele, providing

a rare example of two functional overlapping genes that potentially are maintained by diverging selective pressures.

Discussion

A fundamental shortcoming of modern biology is that our understanding of model organisms, such as baker's yeast, rests heavily on studies of a few lab domesticated isolates with mosaic genomes, i.e. natural trait variation in these species remains a largely unmapped territory. To address this issue, we charted an exhaustive map of the landscape of mitotic proliferation traits in *S. cerevisiae* and its closest relatives, revealing astounding trait diversity between and within these species. Overall, trait variation in *S. cerevisiae* exceeded that of its never domesticated cousin *S. paradoxus*, despite substantially lower genetic variation. This situation is reminiscent of what has been observed for morphological and behavioral traits in *Canis lupus*, where a large variation due to recent population bottlenecks exists among domesticated dogs relative their wolf cousins with no corresponding increase in overall genetic variability [41]. Periods of small population sizes promote trait variation between populations by allowing deleterious, pleiotropic and large effect alleles, which rarely are adaptive in longer evolutionary perspectives, to reach high frequency by genetic drift [39]. The idea that natural yeast genomes are remnants of repeated bottlenecks in essentially clonal populations has recently been contemplated [1]. In fact, the natural life cycle of budding yeasts with one sexual cycle only every 1000 mitotic divisions and only 1% of these sexual cycles corresponding to non-self-fertilization [14,15] makes these unicellular eukaryotes especially susceptible to genetic drift. The high trait variation in *S. cerevisiae* reinforces the idea of *S. cerevisiae* populations being remnants of recent bottleneck events. Such a bottleneck hypothesis is also supported by trait variation in *Saccharomyces sensu stricto* largely following phylogenetic boundaries. First, the six species were phenotypically clearly delineated, an observation that is surprising given wide-spread sympatry [13,24,42,43] and the traditional view that at least *S. paradoxus* and *S. cerevisiae* are phenotypically indistinguishable [25]. Secondly, trait variation within *S. cerevisiae* essentially followed the population structure with little apparent influence from source environments. This is in line with a strong impact of genetic drift, such as imposed by recent bottlenecks, but contrasts against earlier reports that describe links between sources and the traits of specific yeast lineages [10,24,44–46]. The here reported strong effect of population genetic history on traits suggests that these source-to-trait links may either be rare exceptions or due to a confounding influence from a shared population history.

A predicted consequence of recent population bottlenecks is the prevalence of large effect and deleterious alleles that is otherwise typically weeded out by long-term evolution [39]. Hence, the recent allele frequency based estimate that as much as 12% of coding SNPs and 7% of non-coding SNPs in *S. cerevisiae* are deleterious provides strong support for recent bottlenecks in this species [38]. We found abundant experimental evidence for the prevalence of conditionally deleterious alleles in *S. cerevisiae*, most notably in the phenotypically extreme West African population. This lineage showed drastically reduced mitotic performance in 35% of environments, many of which are ubiquitous in a range of biotopes. These low performance traits mapped to a variety of large effect loci, two of which, premature translational termination signals in *GAL3* and *HUR1*, could be traced to loss-of-function mutations in individual genes. The *GAL3* linked galactose utilization defect is especially interesting; although recent loss of the ability to utilize galactose has occurred in multiple *S. cerevisiae* lineages, the genetic basis for this loss diverge between lineages in

the form of loss-of-function mutations in either of *GAL1*, *GAL2* or *GAL3*. This is in line with that even when traits are similar between populations, the underlying genetic structures typically vary between lineages with the vast majority of QTLs being private to one individual population [26], as expected given a bottleneck hypothesis. Here, population specific, large effect alleles were also detected in the other four *S. cerevisiae* populations. Notably, yeast domestication traits, traditionally associated with wine/beer production, were specific for the European population and mapped to individual large affect loci. Hence, it is conceivable that many domestication traits are consequences of genetic drift in the European population and pre-dates, or coincides with, the emergence of large scale wine and beer production. However, there are cases of apparent man-enforced adaptive evolution, suggesting that genetic drift does not completely define the yeast genotype-phenotype map. Notably, we show that *CUP1* amplifications, which underlie natural variation in yeast copper tolerance, have arisen separately in the European and Sake populations. It cannot be excluded that a smaller number of pleiotropic alleles, such as the well known amino acid polymorphism in *MKT1* that affects a wide variety of traits [47], contribute disproportionately to the strength and frequency of population specific traits. However, we found the major QTLs that define population specific traits to be essentially distinct for specific environments, suggesting that the overall influence of pleiotropic alleles is minor [26].

Regardless of underlying ecological and population history determinants, the massive trait variation in *S. cerevisiae* is disconcerting, given the implicit assumption that studies in one or a few laboratory strains of a model organism may provide a general understanding of gene- environment interactions in the species as a whole. Even the set of genes that are essential for viability in *S. cerevisiae*, the most fundamental of traits, was recently reported to vary by more than 5% between two related isolates [48], underscoring this concern. The problem is exacerbated by the universal *S. cerevisiae* type strain S288C proving to be a highly atypical representative of the species. Although it was previously known that S288C constitutes an outlier for certain traits, e.g. with regards to its low sporulation capacity [49] and high transposon content [16], the abnormality of this strain in relation to wild and industrial isolates over a wide array of traits was unexpected. The mosaic genome structure of S288C may contribute to its abnormal traits; although predominantly derived from European stock, it also contains genetic material from a plethora of other progenitors [16,50]. The astounding diversity of yeast traits and the limited representativeness of the universal reference strain calls for care in extrapolating gene-phenotype links derived from a single lab-domesticated and mosaic individual to general statements on the ecology and physiology of the organism. This stresses the need for application of the powerful yeast molecular toolbox to a wider diversity of genetic backgrounds.

Materials and Methods

Strains

Diploid isolates of *S. cerevisiae*, *S. paradoxus*, *S. arboricolus*, *S. kudriavzevii*, *S. bayanus*, *S. mikatae* and *S. castellii* were collected from diverse sources and geographical locations as described [16] and as detailed in Table S1. Hybrid strains for tests of hemizygote growth were prepared by mating a haploid parental strain to the single deletion strain of interest in the BY4741 background; hybridization was verified by the ability to sporulate. The *ENA1,2,5* triple deletion strain BYT5 (*ena1Δ2Δ5Δ::loxP*, BY4741 derivative) was a kind gift from Hana Sychrová of the Institute of Physiology Academy of Sciences of the Czech Republic. The West African

ENA6 was amplified by PCR using primers as described (Table S6) and incorporated into the *URA3* locus of BYT5 using standard molecular biology methods.

Phenotyping

Strains were subjected to high throughput phenotyping by micro-cultivation (n = 2) in an array of environments (Table S2) as described [19]. Briefly, strains were inoculated in 350 µl of Synthetic Defined (SD) medium (0.14% yeast nitrogen base, 0.5% ammonium sulfate and 1% succinic acid; 2% glucose; 0.077% Complete Supplement Mixture (CSM, ForMedium), pH set to 5.8 with NaOH or KOH) and incubated for 48 h at 30°C. Experiment-dependent variations in this protocol are described in Text S1. Pre-cultures were diluted 35× to an OD of 0.03–0.15 in 350 µL of SD medium and cultivated (n = 2) at 30.0° for 72 h in a Bioscreen analyzer C (Growth curves Oy, Finland). Optical density was measured every 20 minutes using a wide band (450–580 nm) filter. Flocculation, which is a serious problem in liquid cultivations of wild yeast cells in higher cultivation volumes, was not observed. The mitotic proliferation rate (population doubling time), lag (population adaptation time) and efficiency (total change in population density) were extracted from high density growth curves and $\log_2$ transformed. Relative fitness variable for each strain and trait, LSC_{ij}, were calculated as:

$$LSC_{ij} = \frac{\sum_{r=1}^{2}\left[\left[\frac{1}{10}\sum_{k=1}^{10}\log\left(wt_{kj}^{r}\right)\right] - \log\left(x_{ij}^{r}\right)\right]}{2}$$

wt_{kj} is the fitness variable of the *k*:th measurement of the wildtype for trait *j*, x_{ij} is the measure of strain *i* for trait *j* and *r* indicates the run. The measure for proliferation efficiency was inverted to maintain directionality between fitness components. Derived $\log_2$ relative proliferation variables were used for all statistical analysis, except where otherwise stated. The average coefficient of variation between replicates, considering all variables and environments, equalled 11.6%, (see Dataset S1).

Statistical analysis

Strains W303, YJM789, RM11 and YIIc17_E5, which contained known or suspected auxotrophies with potential confounding effects on traits, were excluded from all statistical comparisons and calculations of trait and species averages. A two-tailed Student's t-test with equal variance assumptions was used for two-group comparisons. Pearson correlation coefficients were used for correlation analysis, Fisher's exact test was used for tests of enrichments and tests of hypergeometric distributions; Bonferroni corrections were applied in multiple hypotheses testing situations. Population and source specific traits were defined using the False Discovery Rate [51]. A threshold of 2% was applied; however, the relation between the number of population and source specific traits was essentially constant in the FDR interval 0.5–5%. Hierarchical clustering, as outlined in [52], was performed using data centred over each trait and a Pearson correlation coefficient similarity metric. Group clustering was achieved using group averages. Some of the *S. kudriavzevii*, *S. bayanus* and *S. arboricolus* isolates were not available at the start of the experimental series; consequently, data for these strains is missing in 30% of environments. Missing measurements were treated as "missing data" throughout the analysis. To ensure that this did not affect conclusions, analyses relating to these species were repeated also only considering environments for which all isolates were tested. No impact on conclusions was observed.

Associating phenotypes to CNVs, stop codons, and novel genes

For all *S. cerevisiae* isolates, copy number variations, stop codons and novel genes were extracted using previously published genomic sequence information [16]. Gene copy number variations, stop codons and novel genes were extracted and tested for associations to individual traits as detailed in Text S1. Candidate associations were compared to a systematic linkage mapping of crosses between the clean lineages [26] and were retained only if a significant QTL was detected in the corresponding chromosomal region in at least one of the crosses.

Supporting Information

Dataset S1 Mitotic proliferation fitness variables. The mitotic proliferative variables proliferation rate (population doubling time), lag (population adaptation time) and efficiency (total change in population density) were extracted from high density growth curves and $\log_2$ transformed. Relative fitness variables for each strain and trait were calculated by normalization to the corresponding measures of the WT, BY4741 or S288C, as indicated in the Materials and Methods. Natural strain isolates were grouped on the basis of species, population (*S. cerevisiae* and *S. paradoxus*) and source environment (*S. cerevisiae*), whereas environmental variations were divided into classes based on overall physiological impact.
(XLS)

Figure S1 The correlation between proliferative rate and efficiency is environment dependent. The linear correlation (Pearson correlation coefficients, r) between the proliferative rate and the square root of the proliferative efficiency was plotted for each environment separately. Three sample environments are displayed. Red = clotrimazole (3 µM) exposure, blue = maltose (8%) growth, green = 1,2,4-aminotriazole (300 mM) exposure. No evidence of a general adherence to a 1:1 correlation was observed, which would be expected if the rate was maximized in each environment and restricted only by biomass yield.
(PDF)

Figure S2 The *S. cerevisiae*, *S. mikatae* and *S. paradoxus* clade shows superior performance to the *S. kudriavzevii*, *S. bayanus* and *S. arboricolus* clade in a wide range of environments. Trait averages within the *S. cerevisiae*/*S. mikatae*/*S. paradoxus* clade and the *S. kudriavzevii*/*S. bayanus*/*S. arboricolus* clade were calculated for each trait separately, and the clade difference was calculated as (*S. kudriavzevii*/*S. bayanus*/*S. arboricolus*) - (*S. cerevisiae*/*S. mikatae*/*S. paradoxus*). Lag, rate and efficiency traits are displayed separately. Dotted lines indicated with "No stress" correspond to differences in basal conditions.
(PDF)

Figure S3 Traits varying between *Saccharomyces sensu stricto* species. The trait variance between *Saccharomyces sensu stricto* species was calculated for each trait separately using species trait averages. The top ten environments with highest between-species trait variance for each proliferative measure are displayed.
(PDF)

Figure S4 *S. cerevisiae* shows superior performance to other *Saccharomyces sensu stricto* species in a wide range of environments. A) Average proliferative ability in basal conditions in *S. paradoxus* and *S. cerevisiae* in relation to the *S. cerevisiae* reference strain (BY4741, $\log_2$ scale). Error bars correspond to standard errors. No significant (Bonferroni corrected Student's t-test, p>0.25) differences were observed. B) The difference in proliferation between

each of *S. arboricolus* (n = 3), *S. bayanus* (n = 3), *S. kudriavzevii* (n = 4), *S. mikatae* (n = 2) and *S. paradoxus* (n = 35) and *S. cerevisiae* (n = 35) was determined separately for each proliferative measure. Significant differences at p<0.1 (Student's t-test, Bonferroni correction) are displayed. Green indicates proliferation inferior to that of *S. cerevisiae*, red indicates proliferation superior to that of *S. cerevisiae*. * pre-culture performed using the same nitrogen source as in the experiment (see Text S1) C) Proliferation of the *S. paradoxus* Far East (N-43), European (CBS432) and American (UFRJ50791) populations as compared to that of *S. cerevisiae* (BY4741), using mannitol as carbon source. D) Proliferation of the *S. paradoxus* Far East (N-43), European (N-17) and American (A12) populations as compared to that of *S. cerevisiae* (BY4741) during oxalic acid (12.5 mg/mL) exposure.
(PDF)

Figure S5 Trait variation in *S. cerevisiae* exceeds trait variation in *S. paradoxus*. A) Trait variance within *S. cerevisiae* and within *S. paradoxus* was calculated for each trait separately and a mean over all traits was formed. Error bars represent standard errors. B) Genetic variation within *S. cerevisiae* and within *S. paradoxus* [16], represented by the total number of SNPs within each species. C) To determine whether the difference in within-species trait variation between *S. cerevisiae* and *S. paradoxus* is due to population sub-structure effects, trait variance calculations, systematically and repeatedly excluding one individual population at a time, were performed. The difference between species was clear regardless of population excluded.
(PDF)

Figure S6 Traits varying within *S. cerevisiae*. The trait variance over all *S. cerevisiae* isolates was calculated for each trait separately. Environments were ranked separately for each proliferative measure according to degree of variance. Top ten environments are displayed.
(PDF)

Figure S7 Clustering of trait profiles of *Saccharomyces sensu stricto* isolates. Hierarchical clustering of all traits was performed using a centered Pearson correlation metric and average linkage mapping as described in the Materials and Methods. Species are indicated by line color. The heat map reflects the proliferation of each isolate in relation to either BY4741 or S288C. Green = inferior proliferation, red = superior proliferation, black = BY4741/S288C performance, grey = data not available. For *S. cerevisiae*, source habitats (symbols) and population structure [1] are indicated (symbol color). A) Proliferative rate traits B) Proliferative efficiency traits C) Proliferative lag traits.
(PDF)

Figure S8 Trait variation in *S. cerevisiae* is defined by population structure rather than source. To stringently evaluate the effect of population and source on trait variation within *Saccharomyces cerevisiae*, principal component analysis (PCA) followed by ANOVA was performed. Traits with one or more missing values were removed leaving 523 traits, and 21 clean lineage isolates, the latter independently categorized into five population and five source categories. Before PCA, strains were centered and scaled to unit variance. A) Scree plot of phenotypic variance (%) explained by each principal component. The plot shows that the phenotypic variation is high dimensional, with only the first principal component having substantial explanatory power on its own (42%). B) ANOVA was performed on the five first principal components individually, using population and source as regressors. The plot shows the fraction of variance in each principal component that is explained by population and source respective-

ly. Filled bars = significant (p<0.01) effect of population/source, empty bars = non-significant effect of population/source. The principal component analysis shows that the first and overwhelmingly dominant principal component, explaining 42% of the trait variation, was very strongly influenced by population (ANOVA, r = 0.87, p = 1.1E-4) with no effect of source, whereas the following four components, together accounting for 31% of the variation, showed minor and comparable effects of population and source. Hence, the overall effect of population was substantially larger than the overall effect of source.
(PDF)

Figure S9 *S. cerevisiae* lab strains have diverging trait profiles. The proliferative efficiency of the four commonly used *S. cerevisiae* lab strains, S288C, W303, SK1 and Y55, in relation to that of the reference strain BY4741, Log_2 (isolate/BY4741), in ~200 environments. Environments where a strong difference of the respective lab strain to BY4741 was observed are indicated with names.
(PDF)

Figure S10 Source dependent traits in *S. cerevisiae*. Traits that differ significantly (FDR = 2%) between isolates from one source classes and isolates from all other source classes.
(PDF)

Figure S11 Traits unique for the universal reference strain S288c. Traits for which S288C deviate significantly (p<0.1, Student's t-test, Bonferroni correction) from all other *S. cerevisiae* isolates.
(PDF)

Figure S12 Reconstructing the phenotypic history of *S. cerevisiae*. Population specific traits in *S. cerevisiae* were mapped onto a recently established population genomics tree based on low coverage genome sequence data [16]. Population specific traits were defined as environments where the performance of isolates in one population deviated significantly from isolates in other populations (FDR = 2%). Percentages indicate frequency of population specific traits in each population. Inset bar diagrams show a subset of population specific traits for each population. Bars represent trait averages with bar color indicating population and error bars representing standard errors. Total number of population specific phenotypes: West African = 190, European = 30, North American = 13, Malaysian = 13 and Sake = 3.
(PDF)

Figure S13 Low performance traits of the West African population map to different QTLs. To map the chromosomal location of the many phenotypes of the West African phenotypic burst, the West African DBVPG6044 was crossed to the North American YPS128, the Sake Y12, and the European DBVPG6765 as described [26]. For each cross, 96 haploid F1 offspring were obtained after meiotic recombination, and co-inheritance of the West African DBVPG6044 proliferation defects and 130 chromosomal markers was investigated using linkage analysis. Figures depict LOD score plots for the co-inheritance of each chromosomal marker and four West African traits: low heat tolerance (the cross DBVPG6044×YPS128), low copper tolerance (the cross DBVPG6044×DBVPG6765), low cobalt tolerance and low paraquat tolerance (both in the cross DBVPG6044×Y12). Chromosome numbers indicate centromere position on each chromosome and tick marks indicate the position of each marker.
(PDF)

Figure S14 Reconstructing the phenotypic history of *S. paradoxus*. Population specific traits in *S. paradoxus* were mapped

onto the *S. paradoxus* population genomics tree [16]. Population specific traits were defined as traits where the proliferative performance of one population deviated significantly from other *S. paradoxus* isolates (FDR = 2%). Percentages indicate the frequency of population specific traits in each population. Inset bar diagrams show a subset of population specific traits. Bars represent trait averages with bar color indicating population and error bars representing standard errors. Total number of population specific phenotypes: American = 36, European = 19, Far East = 8. Bar color indicates population, error bars indicate standard errors of population averages.
(PDF)

Figure S15 The West African galactose utilization defect is due to a defect in *GAL3*. A) Deletion strains for all components in the galactose utilization pathway were crossed to the galactose defect West African strain DBVPG6044 (WA) to form diploid hemizygotes, BY4741 (*galxΔ*)×WA. Hemizygotes as well as their haploid parents and the hybrid BY4741×WA were micro-cultivated for 72 h in glucose and galactose medium respectively and the population doubling time was quantified. Only the BY4741 (*gal3Δ*)×WA hemizygote shows a galactose growth defect. B) Population doubling time of heterozygote deletion strains (BY4743) corresponding to galactose utilization pathway components, in galactose and glucose respectively. All heterozygotes showed unperturbed growth, including the *gal3Δ* heterozygote, demonstrating that *GAL* gene hemizygosity *per se* does not affect galactose growth. C) Deletion strains for all components in the galactose utilization pathway were crossed to the galactose defect West African strain DBVPG6044 (WA) to form diploid hemizygotes, BY4741 (*galxΔ*)×WA. Hemizygotes as well as their haploid parents and the hybrid BY4741×WA were micro-cultivated for 72 h in glucose and galactose medium respectively and the population doubling time was quantified. The Malaysian strain UWOPS05_217_3 is shown for comparison. The WA melibiose utilization defect is not suppressed by presence of any of the BY *GAL* genes.
(PDF)

Figure S16 Independent amplifications of the *CUP1* locus in the *S. cerevisiae* Sake and European populations. To test the hypothesis of independent amplifications of *CUP1* in the Sake isolate Y12 and the European derived W303, we investigated the breakpoints of the *CUP1* amplification in Y12 and W303 by *de novo* assembly. Y12, W303 and the West African isolate DBVPG6044 was sequenced to 25, 31 and 29-fold coverage respectively with 108 bp reads from a 300 bp insert using the Illumina Genome Analyzer. Reads for each strain were independently assembled using the SGA (string graph assembly) algorithm (Simpson and Durbin, in preparation). SGA is a graph-based assembler which derives the relationship between sequence reads using the FM-index data structure [54]. The graph is traversed to find unambiguous walks which are output as sequence contigs and the topology of the graph indicates the presence and structure of amplified sequence. To identify *CUP1* amplification breakpoints, the structure of the assembly graph around the *CUP1* locus was analyzed for each strain separately. We mapped the sequence of *CUP1* to each assembly to identify matching contigs. A) Dotplot of the *CUP1* contig in DBVPG6044 vs. the *CUP1* region in *S. paradoxus*. In DBVPG6044, *CUP1* is located in a single contig without breaks, strongly indicating that *CUP1* is not duplicated in this strain. We compared the sequence of the matched DBVPG6044 contig to the corresponding locus in *S. paradoxus*, confirming that the layout of the *CUP1* locus in DBVPG6044 is identical to that of *S. paradoxus* and consistent with a single *CUP1*. B) Assembly graph of *CUP1*

sequence in W303 contigs. The *CUP1* sequence had three partial matches to W303 contigs. The region of the assembly graph containing these contigs was manually inspected and a simple cycle (black arrows) was found, indicating duplication. C) Dotplot of the *CUP1* region in W303 as compared to the *CUP1* region in the *S. cerevisiae* reference genome (S288C). To find the breakpoint of the amplification in W303, we constructed a putative assembly of the region where the cycle is traversed a single time. This assembly was aligned to the corresponding region in S288C revealing that this amplification is the same as the *CUP1* tandem duplication present in S288C. D) Dotplot of the *CUP1* region in W303 (single cycle assembly) as compared to the *CUP1-1* and *RSC30* ORFs in S288C. The amplified sequence in W303 corresponds to *CUP1* and the first 1,069 bp of *RSC30*. Alignment of the W303 sequence reads in the 150 bp region surrounding the breakpoint in *RSC30* revealed six reads supporting the break. No Y12 reads matched this breakpoint, strongly suggesting that it does not exist. E) Assembly graph of *CUP1* sequence in Y12 contigs. *CUP1* had three partial matches to Y12 contigs. Manual inspection revealed two cycles (blue and black edges respectively), which suggests two independent amplifications of *CUP1* in Y12. F) Dotplot of the *CUP1* region in Y12 (assembly of the blue cycle) as compared to the *CUP1-1* and *RSC30* ORFs in S288C. The blue cycle corresponded to an amplification of *CUP1* alone; 32 Y12 reads supported this break but no reads from W303. G) Dotplot of the *CUP1* region in Y12 (assembly of the black cycle) as compared to the *CUP1-1* and *RSC30* ORFs in S288C. The black cycle represented an amplification of *CUP1* and the first 1,624 bp of *RSC30*, i.e. 555 bp more than the amplification in W303; 25 reads from Y12 supported this breakpoint but no reads from W303. Although we are unable to reconstruct the exact number and sequence of traversals through the cycles in the graphs using the short read data, the three breakpoints found (one in W303, two in Y12) are unique to the particular strains, clearly demonstrating that the amplifications of *CUP1* in Y12 and W303 are independent events.
(PDF)

Figure S17 The history of *ENA* genes in the *S. cerevisiae* European population diverge from that of the *ENA* genes in other *S. cerevisiae* populations. The *ENA1* gene sequence from the *S. cerevisiae* reference genome contained in the Saccharomyces Genome Database (http://www.yeastgenome.org/) was BLASTed (BlastN) against *S. cerevisiae* genomes [16]. The top hit in each genome was retained and a multiple alignment using ClustalW was performed. The *ENA2* and *ENA5* reference strain paralogs and the two *S. paradoxus ENA* orthologs were included for comparison. The *ENA* genes of the European population and of mosaics predominantly of European origin show an evolutionary history that diverges from that of the *ENA* gene (*ENA6*) in the other populations. An *ENA6* ortholog was also found in the North American YPS128 and YPS606; however, sequences were fragmented and a high confidence assembly could not be made.
(PDF)

Figure S18 A stop codon in *HUR1* explains low tolerance to hydroxyurea in West African and Malaysian *S. cerevisiae* populations. A premature AAA→TAA (K→Stop) stop codon in *HUR1* shows significant association (Student's t-test $p < 0.015$, Kolmogorov-Smirnov $p < 0.15$) to the proliferative efficiency during exposure to 15 mg/mL hydroxyurea. West African and Malaysian derived genomes (DBVPG6044, UWOPS03.461.4, UWOPS05.217.3, UWOPS05.227.2) contain the variant mutation, 14 other strains contain the reference sequence.
(PDF)

Table S1 Strains used in the study. ID refers to the strain ID number in the Gianni Liti collection. For further details on *S. cerevisiae* and *S. paradoxus* strains, see Liti *et al* [16].
(DOC)

Table S2 Environments used in the screen. The classification "carbon utilization" indicates that 2% glucose in these experiments was substituted with the indicated amounts of the indicated carbon source, the classification "nitrogen utilization" indicates that 0.5% ammonium sulfate was substituted with the indicated nitrogen sources at nitrogen limiting concentrations (see Text S1). # = two consecutive pre-cultures were performed, the first with nitrogen limiting amounts of ammonium sulfate (low ammonium sulfate), the second with nitrogen limiting amounts of the indicated nitrogen source. * = pre-cultures were performed in medium identical to the experimental medium.
(DOC)

Table S3 Novel genes whose presence varies between strains. Novel genes whose presence varies between strains, as previously defined [16]. 1 = gene present, 0 = gene not detected.
(DOC)

Table S4 Nonsense mutations in natural yeast isolates. Premature stop codon mutations with a minor allele frequency $\geqq 3$ in *S. cerevisiae*. Stop codons in dubious genes, as well as in genes with a high number of premature stop codons (>2), were not included. In all cases, the reference sequence contains an amino acid encoding codon whereas the variation is a stop codon mutation. Strains carrying a premature stop codon are represented by "0", whereas strains carrying the reference sequence are represented by "1". Empty cells correspond to missing data for the associated strain.
(DOC)

Table S5 Copy Number Variations in natural yeast isolates. CNVs in *S. cerevisiae* with product P-value of less than e-50 (see Text S1). For each gene in each strain, the 'copy number ratio' was computed as $\log[1+\text{observed reads}] - \log[1+\text{expected reads}]$. Dubious genes and transposable elements, which represented the most highly variable genes, were excluded.
(DOC)

Table S6 Primers used in strain construction. Primers used for inserting the SK1 *ENA6* into the URA3 locus of strain BYT5 – a BY4741 carrying a *ENA1,2,5* triple deletion. 5′ overhangs correspond to *URA3* flanking regions. Gene specific sequences for *URA3* or *ENA6* at the 3′ end of the primers according to the table. See Text S1 for further details.
(DOC)

Text S1 Supplementary Materials and Methods. Detailed experimental procedures and protocols.
(DOC)

References

1. Dujon B (2010) Yeast evolutionary genomics. Nat Rev Genet 11: 512–524.
2. Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, et al. (2010) The genetic landscape of a cell. Science 327: 425–431.
3. Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418: 387–391.
4. Li Z, Vizeacoumar FJ, Bahr S, Li J, Warringer J, et al. (2011) Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat Biotechnol.
5. Mnaimneh S, Davierwala AP, Haynes J, Moffat J, Peng WT, et al. (2004) Exploration of essential gene functions via titratable promoter alleles. Cell 118: 31–44.
6. Schuldiner M, Collins SR, Thompson NJ, Denic V, Bhamidipati A, et al. (2005) Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. Cell 123: 507–519.
7. Sopko R, Huang D, Preston N, Chua G, Papp B, et al. (2006) Mapping pathways and phenotypes by systematic gene overexpression. Mol Cell 21: 319–330.
8. Brem RB, Kruglyak L (2005) The landscape of genetic complexity across 5,700 gene expression traits in yeast. Proc Natl Acad Sci U S A 102: 1572–1577.
9. Brem RB, Storey JD, Whittle J, Kruglyak L (2005) Genetic interactions between polymorphisms that affect gene expression in yeast. Nature 436: 701–703.
10. Gerke JP, Chen CT, Cohen BA (2006) Natural isolates of Saccharomyces cerevisiae display complex genetic variation in sporulation efficiency. Genetics 174: 985–997.
11. Liti G, Haricharan S, Cubillos FA, Tierney AL, Sharp S, et al. (2009) Segregating YKU80 and TLC1 alleles underlying natural variation in telomere properties in wild yeast. PLoS Genet 5: e1000659. doi:10.1371/journal.pgen.1000659.
12. Perlstein EO, Ruderfer DM, Roberts DC, Schreiber SL, Kruglyak L (2007) Genetic basis of individual differences in the response to small-molecule drugs in yeast. Nat Genet 39: 496–502.
13. Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, et al. (2002) Dissecting the architecture of a quantitative trait locus in yeast. Nature 416: 326–330.
14. Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L (2006) Population genomic analysis of outcrossing and recombination in yeast. Nat Genet 38: 1077–1081.
15. Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proc Natl Acad Sci U S A 105: 4957–4962.
16. Liti G, Carter DM, Moses AM, Warringer J, Parts L, et al. (2009) Population genomics of domestic and wild yeasts. Nature 458: 337–341.
17. Wang S, Bai F (2008) Saccharomyces arboricolus sp. nov., a yeast species from tree bark. Int J Syst Evol Microbiol 58: 510–514.
18. Warringer J, Anevski D, Liu B, Blomberg A (2008) Chemogenetic fingerprinting by analysis of cellular growth dynamics. BMC Chem Biol 8: 3.
19. Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A (2003) High-resolution yeast phenomics resolves different physiological features in the saline response. Proc Natl Acad Sci U S A 100: 15724–15729.
20. Novak M, Pfeiffer T, Lenski RE, Sauer U, Bonhoeffer S (2006) Experimental tests for an evolutionary trade-off between growth rate and yield in E. coli. Am Nat 168: 242–251.
21. Bell G (2010) Experimental genomics of fitness in yeast. Proc Biol Sci 277: 1459–1467.
22. Wong WW, Tran LM, Liao JC (2009) A hidden square-root boundary between growth rate and biomass yield. Biotechnol Bioeng 102: 73–80.
23. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423: 241–254.
24. Replansky T, Koufopanou V, Greig D, Bell G (2008) Saccharomyces sensu stricto as a model system for evolution and ecology. Trends Ecol Evol 23: 494–501.
25. Barnett JA, Payne RW, Yarrow D (2000) Yeasts: Characteristics and identification. Cambridge, UK: Cambridge University Press.
26. Cubillos FA, Billi E, Zorgo E, Parts L, Fargier P, et al. (2011) Assessing the complex architecture of polygenic traits in diverged yeast populations. Mol Ecol 20: 1401–1413.
27. Fay JC, McCullough HL, Sniegowski PD, Eisen MB (2004) Population genetic variation in gene expression is associated with phenotypic variation in Saccharomyces cerevisiae. Genome Biol 5: R26.
28. Naumov GI, Serpova EV, Naumova ES (2006) [A genetically isolated population of Saccharomyces cerevisiae in Malaysia]. Mikrobiologiia 75: 245–249.
29. Wu H, Ito K, Shimoi H (2005) Identification and characterization of a novel biotin biosynthesis gene in Saccharomyces cerevisiae. Appl Environ Microbiol 71: 6845–6855.
30. Lohr D, Venkov P, Zlatanova J (1995) Transcriptional regulation in the yeast GAL gene family: a complex genetic network. Faseb J 9: 777–787.

31. Hittinger CT, Goncalves P, Sampaio JP, Dover J, Johnston M, et al. (2010) Remarkably ancient balanced polymorphisms in a multi-locus gene network. Nature.

32. Hittinger CT, Rokas A, Carroll SB (2004) Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proceedings of the National Academy of Sciences of the United States of America 101: 14144–14149.

33. Winston F, Dollard C, Ricupero-Hovasse SL (1995) Construction of a set of convenient Saccharomyces cerevisiae strains that are isogenic to S288C. Yeast 11: 53–55.

34. Johnston M, Davis RW (1984) Sequences that regulate the divergent GAL1–GAL10 promoter in Saccharomyces cerevisiae. Mol Cell Biol 4: 1440–1448.

35. Fogel S, Welch JW (1982) Tandem gene amplification mediates copper resistance in yeast. Proc Natl Acad Sci U S A 79: 5342–5346.

36. Kim HS, Fay JC (2007) Genetic variation in the cysteine biosynthesis pathway causes sensitivity to pharmacological compounds. Proc Natl Acad Sci U S A 104: 19387–19391.

37. Daran-Lapujade P, Daran JM, Luttik MA, Almering MJ, Pronk JT, et al. (2009) An atypical PMR2 locus is responsible for hypersensitivity to sodium and lithium cations in the laboratory strain Saccharomyces cerevisiae CEN.PK113-7D. FEMS Yeast Res 9: 789–792.

38. Doniger SW, Kim HS, Swain D, Corcuera D, Williams M, et al. (2008) A catalog of neutral and deleterious polymorphism in yeast. PLoS Genet 4: e1000183. doi:10.1371/journal.pgen.1000183.

39. Stern DL, Orgogozo V (2009) Is genetic evolution predictable? Science 323: 746–751.

40. Zewail A, Xie MW, Xing Y, Lin L, Zhang PF, et al. (2003) Novel functions of the phosphatidylinositol metabolic pathway discovered by a chemical genomics screen with wortmannin. Proc Natl Acad Sci U S A 100: 3345–3350.

41. Wayne RK, Ostrander EA (2007) Lessons learned from the dog genome. Trends Genet 23: 557–567.

42. Sampaio JP, Goncalves P (2008) Natural populations of Saccharomyces kudriavzevii in Portugal are associated with oak bark and are sympatric with S. cerevisiae and S. paradoxus. Appl Environ Microbiol 74: 2144–2152.

43. Sniegowski PD, Dombrowski PG, Fingerman E (2002) Saccharomyces cerevisiae and Saccharomyces paradoxus coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEMS Yeast Res 1: 299–306.

44. Kvitek DJ, Will JL, Gasch AP (2008) Variations in stress sensitivity and genomic expression in diverse S. cerevisiae isolates. PLoS Genet 4: e1000223. doi:10.1371/journal.pgen.1000223.

45. Landry CR, Townsend JP, Hartl DL, Cavalieri D (2006) Ecological and evolutionary genomics of Saccharomyces cerevisiae. Mol Ecol 15: 575–591.

46. Schacherer J, Shapiro JA, Ruderfer DM, Kruglyak L (2009) Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. Nature 458: 342–345.

47. Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, et al. (2010) Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464: 1039–1042.

48. Dowell RD, Ryan O, Jansen A, Cheung D, Agarwala S, et al. (2010) Genotype to phenotype: a complex problem. Science 328: 469.

49. Deutschbauer AM, Davis RW (2005) Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nat Genet 37: 1333–1340.

50. Mortimer RK, Johnston JR (1986) Genealogy of principal strains of the yeast genetic stock center. Genetics 113: 35–43.

51. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society 57: 289–300.

52. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–14868.

53. Naumova ES, Turakainen H, Naumov GI, Korhola M (1996) Superfamily of alpha-galactosidase MEL genes of the Saccharomyces sensu stricto species complex. Mol Gen Genet 253: 111–117.

54. Simpson JT, Durbin R (2010) Efficient construction of an assembly string graph using the FM-index. Bioinformatics 26: i367–373.